Journal of Official Statistics vol. 38, 1 (March 2022)

Journal of Official Statistics, Vol. 38, No. 1, 2022, pp. 1–4, http://dx.doi.org/10.2478/JOS-2022-0001

Preface

Introduction to the Special Issue on Price Indices in Official Statistics

There has been an increasing demand for more timely and detailed price statistics in the past decade or so, partly motivated by and partly in response to the increasing availability of alternative data sources beyond designed price sample surveys. Information from scanner data, web-scraping, and various administrative data sources has raised new theoretical and practical challenges and spurred methodological developments. This special issue of Journal of Official Statistics provides the readers with an opportunity to take stock of the most recent developments in this topic area of Official Statistics.

The special issue contains several contributions on housing and property price indices, always of interest to the public not least because of the large share of housing in total private household expenditure in most countries globally. It also includes articles on export and import price indices, a topic for which previous literature is quite scarce. Index formula choice, always a central topic to price statisticians, also constitutes an overall theme of the special issue.

Finally, and not surprisingly, the COVID-19 pandemic is also a main topic. When this special issue was first announced, the pandemic had not yet started to spread around the world. Quite soon afterwards, however, it became clear that COVID-19 has had a paramount impact on personal health and the economy, and it is doing so continuously. How to produce reliable measures of inflation in the presence of such a shock has become a real problem for National Statistical Offices. Several contributions to the special issue discuss price index compilation in the context of the pandemic.

All in all, the issue contains 13 articles and three comments, covering a wide range of topics within the area of price statistics. A summary of the contributions is given below.

In this Issue

There is a growing practice of harnessing web-scraped price data at National Statistical Offices, either to supplement or potentially to replace manually collected price quotes. One shortcoming of web-scraped data (e.g., compared to scanner data) is the lack of quantity information in many cases, which would enable the construction of weighted item-level price indices. In their contribution, **Ayoubkhani and Thomas** propose to predict product sales quantities from their ranks (for example, when products are sorted ‘by popularity’ on consumer websites) by assuming appropriate statistical distributions. Such an approach to data collection and index calculation related to web-scraped price data may be explored further.

Laureti and Polidoro discuss how to use scanner data to compile indices comparing prices between different geographical areas. They review previous literature on the estimation of subnational spatial price indices and estimate such indices for grocery

© Statistics Sweden

products in Italy using detailed data on prices and quantities. Their results indicate quite a large variability in prices across Italian regions.

In another contribution, **Benedetti et al.** present a methodology for computing spatial housing rent indices, also applied to Italy's regions and districts within regions. They use a hedonic regression model that is an extension of the Country Product Dummy method. Since housing costs are crucial for estimating the cost of living and vary substantially, such a methodology is essential, for example, for poverty analyses. The results show big differences in rents both between regions and between locations (center/outskirt/rural) within regions.

Fast et al. discuss the use of unit values from trade data for export and import price indices, noting that administrative trade data underlying the unit values have become more detailed and thereby more homogeneous. For 123 product categories, they use a number of tests to classify unit value indices into good, undecided or poor. The coefficient of variation (an indicator of homogeneity) proves to be a key statistic for the classification.

Von Auer and Shumskikh start from the fact that export and import price indices are both computed using the Laspeyres formula, whose substitution bias can give rise to a distortion of the terms of trade (ratio of export to import price indices). In their article, they provide a retroactive correction approach that addresses this bias and gives long-run time series of import and export price levels. An empirical study is conducted that demonstrates the practical advantages of the approach.

Hill et al. propose a method for rolling-time-dummy house price indices that requires less data and is thus applicable to higher-frequency indices, such as monthly or weekly, but also in a situation where transactions are scarce. Using real data for Sydney and Tokyo, they test their model and the sensitivity of the results to the choice of the length of the estimation window and the linking method, and benchmark against a quarterly series. The (lower-frequency) benchmark is used to infer the optimal choice for the window length and linking method on the higher-frequency index. With a view to alleviate the structural effects in periods with low number of transactions, they show that an estimation window and linking method that excludes the problematic periods performs well. These are important practical insights in dealing with scarce data.

Silver uses property price indices as a motivation for the hedonic approach, particularly the underlying regression model. He surveys a wide range of regression diagnostics, which are organised in separate sections for basic diagnostics, such as R squared, multicollinearity, specification, and omitted variable bias, and heteroscedasticity and the normality of residuals – including some R code in the footnotes. This article will be very useful to the practitioners because the topic is typically missing in the various guidelines to price index production, although a good regression model is of course critical to any successful application of the hedonic approach, essentially forming the basis of the thus derived quality-adjusted price indices. Due to the limit of space, however, the diagnostics are not illustrated numerically; the readers are invited to work through the article with their real data sets.

Bentley investigate a multilateral time-product-dummy method based on the property fixed-effects model, asking the question of the optimal length of the estimation window over which the property effects can be held fixed. This model is applied to rentals for housing, an important component of consumer expenditures. Bilateral methods based on a matched sample, on the other hand, cannot accommodate the dynamics of the universe of rental objects. The conceptual issues related to the choices are highlighted, and the pragmatic choice is illustrated and discussed empirically based on a dataset spanning over 25 years. It is found that the choice of window length can have a significant impact on estimates of cumulative inflation; indeed, the longer the data window, the greater is the estimated inflation rate. This points to an important problem for future research.

Dawber et al. outline a framework for the UK regional CPI, aiming at a methodological basis that is as close as possible to the national CPI. They analyse various aspects of the data scarcity empirically and conclude that estimating the expenditure weights directly from the within-region data is the most important cause to the volatility of the resulting regional CPI. To a certain extent, applying a multilevel model commonly used in Small Area Estimation can smooth the indirectly estimated weights and stabilize the corresponding CPIs. Further developments on data availability and suitable estimation methods are needed to satisfactorily produce the UK regional CPI.

The contribution by **Martin et al.** is a welcome addition to price index literature; while the choice of elementary level index formula for CPIs has been discussed extensively, the same cannot be said for Producer Price Indices (PPIs). In their article, Martin et al. discuss elementary level formulas for PPIs and show empirically that formula choice can have economically significant effects on overall PPI inflation figures.

Diewert and Fox discuss methodological issues for the CPI arising from effects of the COVID-19 pandemic. The unavailability of some goods and services because of lockdowns has posed a considerable challenge to statistical offices worldwide. Except for analytical series, expenditure weights in consumer price indices have remained unchanged. Instead, imputations for the missing prices using comparable products (or a higher-level index) were proposed in internationally agreed guidance. In their article, Diewert and Fox discuss various theoretical and practical methods to impute the missing prices. Notably, they claim that following the official guidance might lead to an understatement of inflation and instead reservation prices would yield the “true” index. Further consideration is also given to more practical measurement problems other than unavailable products.

The article by Diewert and Fox is followed by a discussion, including contributions by **Boldsen, Goldhammer** and **Abe**. The discussion highlights a major controversy about the use of reservation prices per se, with the discussants criticizing reservation prices and instead consenting to the international guidance – and statistical practice – of imputing prices based on observations for similar products. Further debated are the linking methods proposed by Diewert and Fox, the effects of stockpiling, and the use of monthly expenditure weights. All of these issues – and more related ones – will likely be subject to debate and further scrutinization in the prices community.

In another contribution, **Abe et al**. investigate price index numbers under large-scale demand shocks, focusing on the case of face masks during the first months of the COVID-19 outbreak in Japan. In particular, this includes the cost-of-living index (COLI) recently proposed by Redding and Weinstein (2020), which explicitly accommodates effects of changing product variety. The empirical results show that the COLI behaved quite differently to the superlative Fisher index and other alternatives, in a manner that more appropriately reflects the demand shock according to the authors. It would be intriguing to extend this line of research both theoretically and empirically.

The special issue ends with a contribution by **Pfeifer and Steurer**, also related to the COVID-19 pandemic. It addresses the effects of the pandemic on the real estate market in two major European cities – London and Vienna. The authors look at both prices and rents for apartments and find different relative behaviours of the sales and the rental markets in the two cities. In Vienna all indicators show positive market developments during the first eight months of the pandemic whereas in London only the sales market was positive while the rental market became weaker.

Acknowledgments

On behalf of the Editorial Board, we would like to thank all the authors who have submitted manuscripts for our consideration and all the referees who have taken the time to review these manuscripts and to make comments and suggestions which have in many cases greatly improved the articles.

Jörgen Dalén
Jens Mehrhoff
Olivia Ståhl
Li-Chun Zhang
Guest Editors

References

Redding, S.J., and D.E. Weinstein. 2020. “Measuring Aggregate Price Indices with Taste Shocks: Theory and Evidence for CES Preferences.” *Quarterly Journal of Economics*, 135: 503–560. DOI: https://doi.org/10.1093/qje/qjz031.

Journal of Official Statistics, Vol. 38, No. 1, 2022, pp. 5–21, http://dx.doi.org/10.2478/JOS-2022-0002

Estimating Weights for Web-Scraped Data in Consumer Price Indices

Daniel Ayoubkhani[1] *and Heledd Thomas*[1]

In recent years, there has been much interest among national statistical agencies in using web-scraped data in consumer price indices, potentially supplementing or replacing manually collected price quotes. Yet one challenge that has received very little attention to date is the estimation of expenditure weights in the absence of quantity information, which would enable the construction of weighted item-level price indices. In this article we propose the novel approach of predicting sales quantities from their ranks (for example, when products are sorted 'by popularity' on consumer websites) via appropriate statistical distributions. Using historical transactional data supplied by a UK retailer for two consumer items, we assessed the out-of-sample accuracy of the Pareto, log-normal and truncated log-normal distributions, finding that the last of these resulted in an index series that most closely approximated an expenditure-weighted benchmark. Our results demonstrate the value of supplementing web-scraped price quotes with a simple set of retailer-supplied summary statistics relating to quantities, allowing statistical agencies to realise the benefits of freely available internet data whilst placing minimal burden on retailers. However, further research would need to be undertaken before the approach could be implemented in the compilation of official price indices.

Key words: Index numbers; price index; alternative data sources; web scraping; expenditure weights.

1. Introduction

The Consumer Prices Index including owner occupiers' housing costs (CPIH) is produced by the Office for National Statistics (ONS), the UK's national statistical institute (NSI). It is the headline measure of consumer price inflation in the UK, and its users and uses are both varied and important: the Bank of England for inflation targeting; central government for setting and monitoring economic policies, and for indexing state pensions and benefits; businesses for making decisions over future investment and staff pay; and the general public, often via the media, for household budget and consumption planning.

The CPIH's methodology conforms to international standards (ILO 2020). Around 700 items of interest are specified in an annually updated basket of goods and services (ONS 2020a); each month, price changes for these items are aggregated to estimate the overall rate of inflation, using a Laspeyres-type index with weights derived from expenditure shares. The item-level price changes are calculated from around 180,000 product-level

[1] Office for National Statistics, Government Buildings, Cardiff Road, Newport, NP10 8XG, UK. Emails: daniel.ayoubkhani@ons.gov.uk and heledd.thomas@ons.gov.uk
Acknowledgments: We thank colleagues in the Methodology and Prices divisions of the Office for National Statistics and members of the Technical Advisory Panel on Consumer Prices for their valuable input into this work. We are also grateful to three anonymous reviewers for their helpful comments and suggestions.

© Statistics Sweden

price quotes, the majority of which are collected manually. Many of these quotes are sourced by human price collectors, who physically visit retail stores of all types and sizes up and down the UK each month. Whilst some internet price quotes are used in the index, these are collected in a manual rather than an automatic fashion.

Recent years have seen a proliferation of research by NSIs into using routinely collected electronic data in the compilation of official measures. This has been driven by an ever-increasing availability of such data sets, coupled with advancements in computer storage and processing capabilities and 'big data' analysis methods. Though not collected for statistical purposes, these 'alternative', non-physically collected sources offer many benefits compared with more traditional methods of data collection, including: increased sample coverage; more frequent data collection; improved publication timeliness; reduced production costs; and the potential to publish regional inflation measures. In particular, *scanner* data involves retailers supplying the statistical agency with transactional information generated at the point of sale, while *web-scraping* makes use of robots to automatically collect price quotes from consumer-facing websites.

Aside from being a means by which to realise the statistical and financial benefits listed above, utilising web-scraped data may be viewed as a necessary adaptation to an evolving marketplace. Internet sales comprise a large and growing component of total retail sales, particularly during the coronavirus disease 2019 (COVID-19) pandemic. Following the introduction of national lockdown measures in Great Britain, including the closure of non-essential retail businesses, internet sales in April 2020 contributed 30.7% of total retail sales (ONS 2020b), a record high at the time and a large increase from 18.7% a year earlier (ONS 2019a). This potentially represents a permanent switch in shopping behaviour for many consumers.

Studies around the globe have demonstrated the use of web-scraped data to estimate price indices (see Nygaard 2015; Polidoro et al. 2015; Ten Bosch and Griffioen 2016; Van Loon and Roels 2018 for a description of the Norwegian, Italian, Dutch, and Belgian experiences, respectively). In the UK, the ONS has recently published its intentions to incorporate web-scraped data in the CPIH by 2023 (ONS 2019b) and has acknowledged the considerable challenges in doing so. Accordingly, the ONS has committed to undertaking a significant period of methodological research to enable these data sources to be fully exploited in the CPIH (ONS 2020c), particularly around aspects such as product classification, outlier detection, quality adjustment, index number construction, and the estimation of product-level weights; the last of these has received very little attention in the price statistics literature, and is therefore the focus of the present article. To date, investigators have generally made use of unweighted index number techniques to construct elementary aggregates from web-scraped data, for example Cavallo and Rigobon (2016) calculated geometric means of product-level price relatives to evaluate the inflation experience of numerous countries using web-scraped data. Unweighted geometric means were also calculated by Hull et al. (2017) to aggregate web-scraped fruit and vegetable prices in Sweden. Willenborg (2017) outlined and evaluated various approaches to constructing elementary price indices from web-scraped data, though weighted aggregation to higher levels remained beyond the scope of the review.

At present, the lack of expenditure weights beneath the elementary aggregate level of the UK CPIH is somewhat mitigated against by the fact that prices are collected for only a sample of products for each item in the basket. The sampling is performed largely purposively based on local popularity, ensuring approximately equal importance between sampled products within each item. However, the advent of alternative data sources means that price quotes are available for *all products* that are sold by a retailer each month, and not just a sample of them, so the prospect of differential importance and the need for product-level weights are much greater. Unlike transactional data, web-scraped data sets include only price quotes and not the associated quantities in which products were sold in each month, such that product-level expenditure weights cannot be directly calculated and incorporated into the price index. An unweighted approach may be adopted in the absence of expenditure weights, but this would result in all products having an equal influence on the index, irrespective of their popularity, potentially introducing a bias into the price index.

This article describes exploratory research into the estimation of product-level weights for web-scraped data in the UK CPIH, by translating observed product rankings (for example, when products are sorted 'by popularity' on consumer-facing websites) into estimated expenditure shares that can be incorporated into price indices. Section 2 provides an overview of the data set and index number formula used in our analysis; Section 3 summarises the results of applying a heuristically chosen formula for deriving product-level weights; Sections 4 and 5 respectively outline the approach and present the results of using classical statistical distributions to predict products' sales quantities from their ranks; and Section 6 offers some concluding remarks and points for further discussion.

2. Study Data and Index Number Construction

2.1. Description of the Analysed Data

We analysed monthly transactional data covering the 24-month period May 2011 to April 2013 for two consumer items, shampoo and toothpaste. The data were supplied to the ONS by a single UK retailer. Although our interest lies in web-scraped data, the availability of both prices and quantities on the transactional data set meant that estimated expenditure weights and the resulting indices could be compared against corresponding observed values and assessed for their accuracy. In a real-world setting, product rankings would be scraped from retailers' websites alongside the associated prices. For this study we therefore assumed that quantity ranking (as observed in the analysed transactional data set) would be reasonably well approximated by popularity ranking (as would be observed in a web-scraped data set).

The data set comprised 736 shampoo products and 310 toothpaste products (Table 1). Both items are characterised by having a small number of market-leading products, and 'long tails' consisting of many products with relatively low sales. The top selling shampoo and toothpaste products contributed 2.6% and 3.1% of sales value respectively (under perfect competition, each product would contribute 0.14% and 0.32% respectively). The top 50 selling products made up nearly half of the sales value for shampoo and nearly two-thirds of the sales value for toothpaste. This highly nonlinear relationship between sales and their ranks is illustrated in Figure 1 (only the first period in the data set, May 2011, is shown for illustrative purposes).

Table 1. Summary statistics for shampoo and toothpaste, may 2011 to april 2013.

Statistic	Shampoo	Toothpaste
Number of products analysed	736	310
Products available in every month (%)	45	35
Mean number of months products were available	18.0	16.4
Products available less than 6 months of the year (%)	9.2	17.7
Sales (expenditure) contributed by the highest expenditure product	2.6	3.1
Sales (expenditure) contributed by the top 5 products (%)	8.6	13.8
Sales (expenditure) contributed by the top 10 products (%)	14.1	24.3
Sales (expenditure) contributed by the top 50 products (%)	43.0	64.5

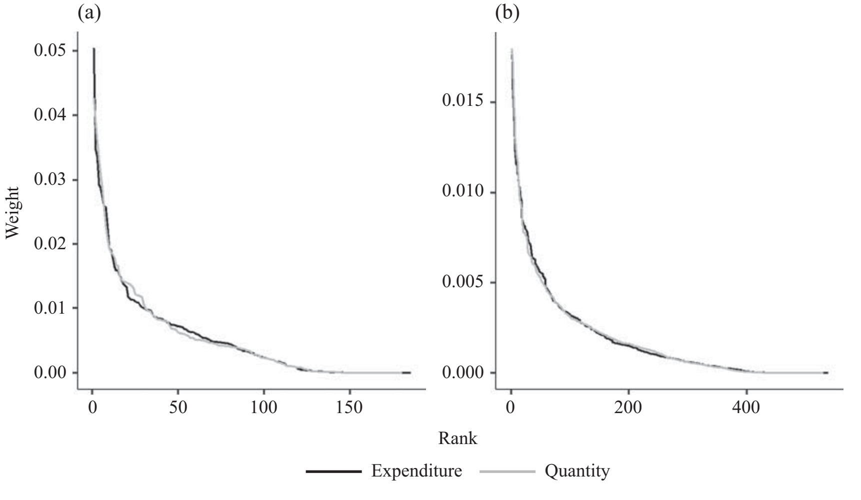


Fig. 1. Relationship between sales and corresponding ranks for (a) toothpaste and (b) shampoo, may 2011.

2.2. Choice of Index Number Formula

When selecting an appropriate price index number formula with which to conduct the analysis, our starting point was the Jevons approach, which is used to construct the majority of elementary aggregates (groups of products for which sufficiently granular expenditure data are not available) in the UK CPIH; that is, the price index series for shampoo and toothpaste were calculated as vectors of unweighted geometric means of price-relatives amongst the constituent products:

$$P_J^{0,t} = \prod_{i=1}^{n} \left(p_i^t / p_i^0\right)^{\frac{1}{n}} \tag{1}$$

where: $P_J^{0,t}$ is the value of the Jevons index representing price changes amongst n products between base month 0 and current month t; and p_i^t and p_i^0 are the price levels of product i in months t and 0, respectively.

For duality with the unweighted Jevons approach, and to draw direct comparisons with a product-weighted version of it, we next constructed a benchmark price index series for each item using the geometric Laspeyres formula, with weights equal to expenditure shares observed in the transactional data set:

$$P_{GL}^{0,t} = \prod_{i=1}^{n} \left(p_i^t / p_i^0\right)^{w_i^0} \tag{2}$$

where: $P_{GL}^{0,t}$ is the value of the geometric Laspeyres index representing price changes amongst n products between base month 0 and current month t; p_i^t and p_i^0 are the price levels of product i in months t and 0, respectively; and w_i^0 is the weight (expenditure share) of product i in month 0.

Finally, different approaches to estimating expenditure weights could be assessed by replacing the observed w_i^0 in Equation 2 with its estimate and comparing the resulting price index series to the benchmark.

Unlike a traditional data collection of price quotes, the sample of products on which transactional data are available is dynamic, reflecting real-world consumer preferences rather than statistical design choices. Only 45% of shampoo products and 35% of toothpaste products were available in all 24 analysed months and, on average, shampoo products appeared in the data set for 18.0 months and toothpaste products for 16.4 months (Table 1).

3. A Heuristic Formula for Estimating Product-Level Weights

We first applied a pragmatic formula to transform product rankings to base-period weights, simply by calculating 'rank shares' instead of the usual expenditure shares. Given the shape of the expenditure-rank relationships illustrated in Figure 1, substituting ordinal ranks in place of cardinal expenditures – thereby assuming a linear relationship between expenditure and rank – is clearly an unsatisfactory approach to replicating the benchmark weight distribution, so we increased the flexibility of the formula by raising the base-period quantity ranks r_i^0 to a power x (optimised on the observed data):

$$w_i^0 = \left(n - r_i^0 + 1\right)^x / \sum\nolimits_{i=1}^{n} \left(n - r_i^0 + 1\right)^x \tag{3}$$

We compared the resulting geometric Laspeyres index series to the benchmark series (using weights based on observed expenditure) over the 12-month period May 2012 to April 2013. We evaluated unit increments of x, finding that differences between the rank- and expenditure-weighted series decreased until $x = 6$, after which the series began to diverge. The index series incorporating the rank-based weights closely mirrored the benchmark series in terms of month-on-month changes, albeit generally at higher levels and with increasing differences over the year (Figure 2).

The results above demonstrate that by simply using the sixth power of quantity rank in place of expenditure (which would not be observed in a web-scraped data set), weighted price indices can be obtained with reasonable accuracy for the two items under analysis. Though despite its intuitive appeal, this heuristic approach to translating product ranks to weights lacks statistical foundation and its functional form is essentially arbitrary.

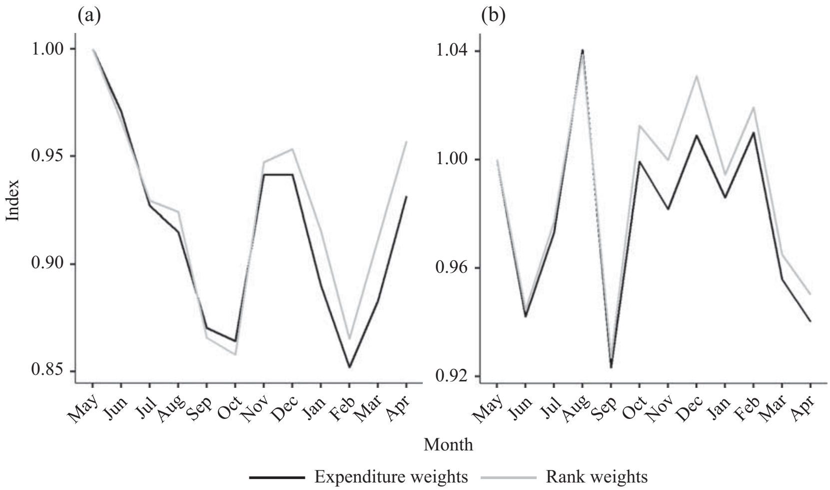


Fig. 2. Directly calculated expenditure- and rank-weighted price indices for (a) toothpaste and (b) shampoo, may 2012 to april 2013.

Moreover, we were only able to optimise this functional form (that is, set $x = 6$) and evaluate its performance as we had access to a transactional data set containing both price and quantity information, and could therefore calculate expenditure weights; in real-world applications, a web-scraped data set would provide only price and *rank* information, and thus there would be no way to construct a benchmark index series with which to optimise and evaluate the weighting formula. The approach described above – taking the sixth power of product rank and applying it as in Equation (3) – is unlikely to be generalisable beyond the items and time periods in our sample.

4. Estimating Sales Quantities from Their Ranks Using Statistical Distributions

4.1. Motivation for Using Statistical Distributions

Developing a weighting methodology that produces sufficiently accurate estimates of product-level expenditure shares for *all* items in *all* time periods, not just those for which we have transactional data, necessitates the incorporation of item- and period-specific data. At this point we emphasise that our objective is to estimate products' expenditure shares based on their ranks (as indicated by the products' ordering when sorted by their popularity on consumer-facing websites from which the corresponding price data are scraped), motivated by the constraint that retailers may be unwilling or unable to share their product-level transactional data with the NSI; indeed, if this information was readily available then there would be no need to scrape internet-based price data at all. We now explore the possibility that retailers may be willing and able to share item- and period-specific *summary statistics*, such as the mean and standard deviation of sales quantities amongst all products within a particular item for a particular month. Aggregate summaries such as these are likely to be treated with less sensitivity by retailers and to present less of a technical challenge to collate and transmit compared with the more granular product-level microdata.

For both shampoo and toothpaste, we translated base-period sales quantity ranks to quantiles of the cumulative distribution $F(q_i) = 1 - r_i/n$. In this formulation there are r_i products with sales quantities greater than or equal to that of product i (that is, q_i). The goal of the analysis was then to find a statistical distribution that suitably approximated the observed quantiles, and to use this distribution to predict sales quantities from their ranks.

4.2. Candidate Distributions

The choice of candidate theoretical distributions was informed by the fact that the empirical quantity distributions exhibit long tails, with a very small number of products having very large sales quantities and the majority of products making up the rest of the distribution. An additional practical consideration when selecting the candidate distributions was the ease with which their parameters can be obtained; requiring retailers to perform burdensome calculations and conduct numerical optimisation procedures to obtain aggregate summary statistics is unlikely to be any more palatable than asking them to provide the underlying product-level microdata.

Given the shape of the above frequency distributions, and the relative ease with which their parameters can be estimated by retailers, we considered the log-normal, truncated log-normal and Pareto (power-law) distributions as candidates for predicting sales quantities from their ranks. These distributions have previously been fitted to sales data from a variety of sectors and time periods, often with reasonable success.

Stanley et al. (1995) demonstrated that manufacturing firms' sales can be well approximated by the log-normal distribution. However, a Zipf plot of log-sales against log-ranks illustrated that the upper tail of the empirical distribution was too thin relative to the theoretical log-normal distribution. Online book sales on Amazon.com and BN.com during 2001 were analysed by Chevalier and Goolsbee (2003), who modelled sales quantities as being Pareto distributed and used this distribution to translate observed ranks into predicted quantities. The resulting sales-weighted average book prices were notably different to the corresponding unweighted average prices for both retailers. Daily data on the sales of digital cameras in Japan between 2004 and 2008 were analysed by Hisano and Mizuno (2010), who concluded that the observed sales distribution could be well approximated by the log-normal distribution for some periods while a power-law distribution provided a better fit for others. Touzani and Buskirk (2015) used transactional data from US retailers between 2004 and 2011 to model the quantity distributions of refrigerators, freezers and washing machines, finding that quantities could be predicted from their ranks with reasonable accuracy by modelling them with a log-normal distribution. Truncation of the distribution led to further improvements in goodness-of-fit, particularly among the top selling products.

4.3. Parameter Estimation

We estimated monthly parameters for the log-normal distribution by maximum likelihood estimation; that is, the sample mean and standard deviation of the natural logarithm of sales quantities:

$$\hat{\mu}_{LN} = \frac{1}{n}\sum\nolimits_{i=1}^{n} \ln q_i; \;\; \hat{\sigma}_{LN} = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n} (\ln q_i - \hat{\mu}_{LN})^2}$$

We estimated the lower and upper bounds of the truncated log-normal distribution as $\min(q_i)$ and $\max(q_i)$, respectively, and used maximum likelihood estimation to estimate the parameters of the Pareto distribution; for the scale parameter this is $\min(q_i)$, and for the shape parameter:

$$\hat{\alpha}_P = n \times \left(\sum\nolimits_{i=1}^{n} \ln\left[q_i/\min(q_i)\right]\right)^{-1}$$

4.4. Assessing Performance

We assumed that discrete sales quantities could be sufficiently well approximated by continuous distributions because of the relatively large sample of products and range of quantities within each of the items. Before assessing their accuracy, we discretised the predicted sales quantities by rounding them down to the nearest integer:

$$\hat{q}_i^* \begin{cases} \lfloor \hat{q}_i \rfloor & if\ \hat{q}_i \geq 1 \\ 1 & \text{otherwise} \end{cases} \tag{4}$$

For each of the two items, we assessed the predictive performance of each candidate distribution in each month using the mean absolute percentage error (MAPE) of predicted log-quantities ln $(\hat{q}_i^*)$:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\left[\ln q_i - \ln \hat{q}_i^*\right]}{\ln q_i} \times 100\right| \tag{5}$$

We evaluated the out-of-sample performance of the statistical distributions by fitting them to the first 12 months of the data (the ‘training’ set, May 2011 to April 2012) and then calculating the MAPE of their predictions over the next 12 months (the ‘holdout’ set, May 2012 to April 2013). This simulates the situation whereby the models would be periodically re-estimated using a given year’s data, and the fitted models would then be applied to make predictions for one or more subsequent years. This evaluative approach avoids overstating the accuracy of the distributions that might be observed in their real-world application; there we might reasonably expect a lag between the reference period and receipt of the necessary parameter estimates from retailers. To assess the cost of this lag in terms of predictive accuracy, we also calculated in-sample MAPEs by re-estimating the distributional parameters and re-evaluating predictive performance over the same period, May 2012 to April 2013.

The out-of-sample predicted quantities were multiplied by observed prices to estimate product-level expenditures and, in turn, expenditure shares. We compared the resulting geometric Laspeyres price index series over the holdout period, calculated as per Equation (2), to that obtained using observed rather than estimated expenditure shares as weights.

5. Results from the Fitted Statistical Distributions

5.1. Fitting the Distributions to the Training Set

The parameter estimates over the 12 months of the training set, May 2011 to April 2012, are summarised for each item in Table 2. On average, the monthly mean, standard

Table 2. Monthly parameter estimates for the fitted distributions, may 2011 to april 2012.

Item	Month	$\hat{\mu}_{LN}$	$\hat{\sigma}_{LN}$	$\max(q_i)$	$\hat{\alpha}_P$
Shampoo	May	6.511	2.778	38,470	0.154
	June	6.736	2.500	30,159	0.148
	July	6.767	2.587	41,992	0.148
	August	6.514	2.799	58,219	0.154
	September	6.709	2.686	44,906	0.149
	October	6.609	2.714	59,613	0.151
	November	6.560	2.606	42,008	0.152
	December	6.511	2.809	56,758	0.154
	January	6.357	2.741	68,773	0.157
	February	6.275	2.851	36,689	0.159
	March	6.485	2.835	39,200	0.154
	April	6.396	2.878	58,729	0.156
Toothpaste	May	7.160	3.343	85,216	0.140
	June	6.835	3.348	102,216	0.146
	July	7.219	3.179	100,259	0.139
	August	7.184	3.145	97,522	0.139
	September	7.257	3.023	80,124	0.138
	October	7.189	3.033	126,425	0.139
	November	6.883	3.259	55,224	0.145
	December	7.100	3.087	73,437	0.141
	January	6.618	3.276	116,674	0.151
	February	7.153	2.974	64,602	0.140
	March	6.846	3.252	80,741	0.146
	April	6.964	3.192	76,972	0.144

deviation and maximum sales quantities of toothpaste were greater than those of shampoo ($\min[q_i]$ was 1 for both items in all 12 months). Both items had similar estimates of the Pareto shape parameter, with values less than unity indicating relatively slow decay in the probability density function, and rendering the mean of the distribution infinite and the variance undefined. The parameter estimates were generally stable over the year, with $\max[q_i]$ exhibiting the greatest month-to-month variation.

5.2. *Out-Of-Sample Performance*

Across all months of the holdout period (May 2012 to April 2013), the truncated log-normal distribution achieved an out-of-sample MAPE of 30.7% for shampoo and 28.5% for toothpaste, with the monthly MAPEs ranging from 22.2% to 40.1%, and 19.5% to 44.1%, respectively (Table 3). However, truncation did not result in a systematic improvement in predictive accuracy; monthly differences in out-of-sample MAPEs between the truncated and non-truncated variants of the log-normal distribution ranged from -0.8 to 0.6 percentage points for shampoo, and -1.6 to 2.5 percentage points for toothpaste.

As expected, the out-of-sample MAPEs were notably larger than those calculated in-sample for both variants of the log-normal distribution, with differences extending to 11.0 and 12.6 percentage points for shampoo and toothpaste, respectively. Differences between out-of-sample and in-sample performance were negligible for the Pareto distribution, but the MAPEs

Table 3. Mean absolute percentage error (MAPE) of quantity predictions from the statistical distributions, may 2012 to april 2013.

Item	Month	In-sample MAPE (%)			Out-of-sample MAPE (%)		
		Log-normal	Truncated log-normal	Pareto	Log-normal	Truncated log-normal	Pareto
Shampoo	May	26.6	27.9	48.6	31.6	31.2	48.9
	June	25.8	27.3	50.4	36.8	36.0	50.7
	July	24.7	25.8	51.0	33.0	32.3	49.1
	August	29.4	31.8	45.4	40.5	40.1	45.5
	September	26.4	27.5	50.1	34.4	33.9	49.1
	October	24.9	25.5	50.3	28.7	28.4	49.5
	November	30.2	30.9	50.2	35.9	35.7	50.4
	December	22.0	22.1	52.3	21.8	22.2	52.6
	January	24.3	26.1	47.9	28.6	29.2	48.0
	February	26.3	26.7	50.2	25.8	26.4	50.2
	March	26.2	26.8	48.6	28.2	28.5	48.7
	April	22.7	23.2	49.7	23.3	23.7	49.8
	All months	*25.8*	*26.8*	*49.5*	*30.8*	*30.7*	*49.3*
Toothpaste	May	21.6	21.6	52.9	18.9	19.5	52.9
	June	27.3	29.4	49.5	26.1	28.6	49.4
	July	22.2	22.5	54.0	19.9	20.1	54.1
	August	19.6	21.2	50.7	20.8	21.0	51.0
	September	16.4	17.6	47.6	22.3	20.7	48.1
	October	20.2	22.8	47.0	31.4	30.6	45.6
	November	23.1	28.2	42.6	34.0	34.1	42.4
	December	24.0	28.0	43.7	36.6	35.7	44.3
	January	20.5	25.1	46.3	23.9	25.6	46.3
	February	21.2	24.2	47.1	32.1	31.5	47.4
	March	26.1	29.7	45.5	31.2	32.8	45.4
	April	32.3	36.3	38.9	44.4	44.1	39.0
	All months	*22.8*	*25.4*	*47.2*	*28.3*	*28.5*	*47.2*

Note: Out-of-sample MAPEs were calculated by estimating parameters over the period May 2011 to April 2012 and evaluating the performance of the fitted distributions over the period May 2012 to April 2013; in-sample MAPEs were calculated by estimating parameters and evaluating performance over the same period, May 2012 to April 2013.

themselves were notably larger than for the other two distributions: 49.3% and 47.2% for shampoo and toothpaste, respectively, across all 12 months when calculated out-of-sample.

To shed further light on the differences and similarities between the distributions in terms of their overall out-of-sample performance, we compared observed product-level quantities with 100 simulated draws from each of the fitted candidate distributions; for illustrative purposes, Figure 3 shows the observed and simulated log-quantities against their log-ranks for a single month, January 2013, with parameters estimated on data from the previous January. The observed log-quantities were generally within the range of those simulated by the truncated log-normal distribution, but they did not exhibit the 'signature' linear trend that would be expected if the data followed a power-law distribution such as the Pareto distribution. Furthermore, the log-normal distribution tended to overestimate sales for the top-ranking products.

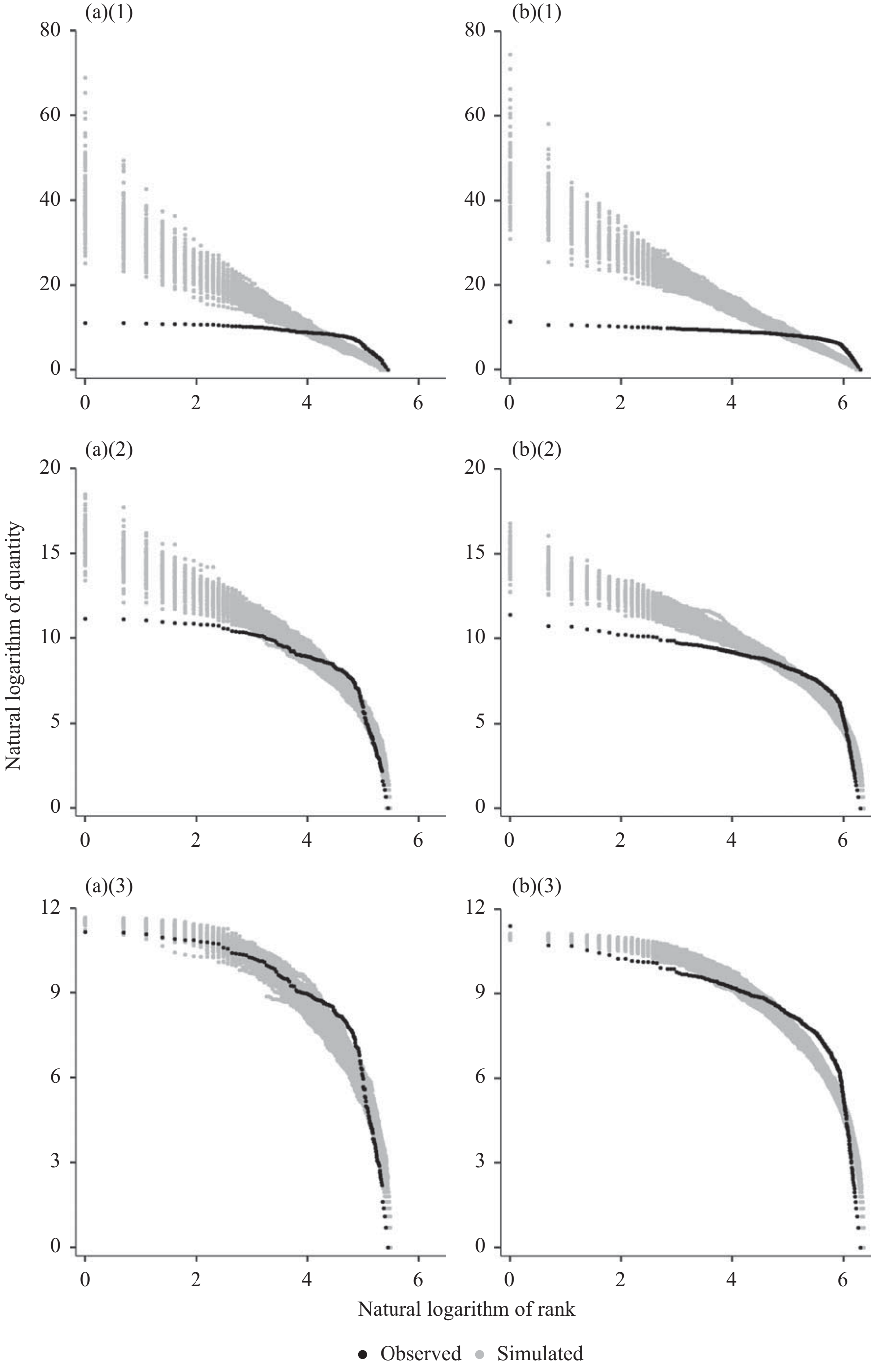


Fig. 3. Observed quantities and ranks of (a) toothpaste and (b) shampoo versus simulations from the (1) Pareto, (2) log-normal and (3) truncated log-normal distributions (log scale), january 2013.

Note: Distributional parameters were estimated using data from January 2012.

These findings are reinforced by directly comparing observed and predicted log-quantities (Figure 4), which demonstrate that most of the improvement in goodness-of-fit that arose by truncating the log-normal distribution was due to a reduced tendency to over-predict sales at the top end of the quantity distribution. However, truncation did not appear to substantially improve predictive accuracy in the middle and bottom end of the distribution.

5.3. Comparison of Index Series

Despite the apparent similarity in overall MAPE between the log-normal and truncated log-normal distributions, the improved predictive accuracy achieved by truncation for the most popular products (that is, those carrying the most weight) resulted in a holdout-set index series that was notably closer to the expenditure-weighted benchmark (Figure 5). As with the rank-weighted indices discussed in section 3, both variants of the log-normal distribution led to index series that were generally above and increasingly divergent from the benchmark series.

6. Discussion

In this article we have demonstrated the value of fitting statistical distributions to predict product-level sales quantities from their ranks for two common consumer items – a notable addition to the analyst's toolkit when considering web-scraped data for price inflation measurement. Of the three candidate statistical distributions we evaluated, the truncated log-normal distribution provided the closest approximation to the observed quantities of both shampoo and toothpaste, with truncation reducing the propensity for over-prediction amongst higher ranked products. (Of course, whether the predictions from the truncated log-normal distribution are sufficiently accurate in absolute terms to be considered for implementation in the official UK CPIH remains an open question.) The observed quantities did not exhibit characteristic power-law behaviour for either item, which was reflected in the predictive performance of the Pareto distribution.

The choice of statistical distributions included in our analysis was driven not only by the characteristics of the data, but also by the requirement for the parameter estimates of the distributions to be readily calculable by retailers (that is, estimated using straightforward summary statistics with closed-form solutions rather than numerical optimisation techniques). This approach means that the NSI can estimate expenditure weights to accompany web-scraped price quotes, while the retailer need not supply product-level microdata. Given our finding that parameter estimates other than $\max\left[q_i\right]$ were very stable over the year, the exercise may be further simplified by requiring that the retailer need only supply this single value each month, with the other parameter estimates being fixed over the year. Furthermore, statistical distributions other than the three we assessed may provide greater predictive accuracy, though at the cost of increased complexity to parameterise them.

Our out-of-sample validation paints a reasonably realistic picture of the prediction error with which the ONS might be faced in a production setting, when timeliness constraints mean that retailers may only be able to supply monthly summary statistics relating to the previous year. A longer span of data than that available to us in this study may lead to

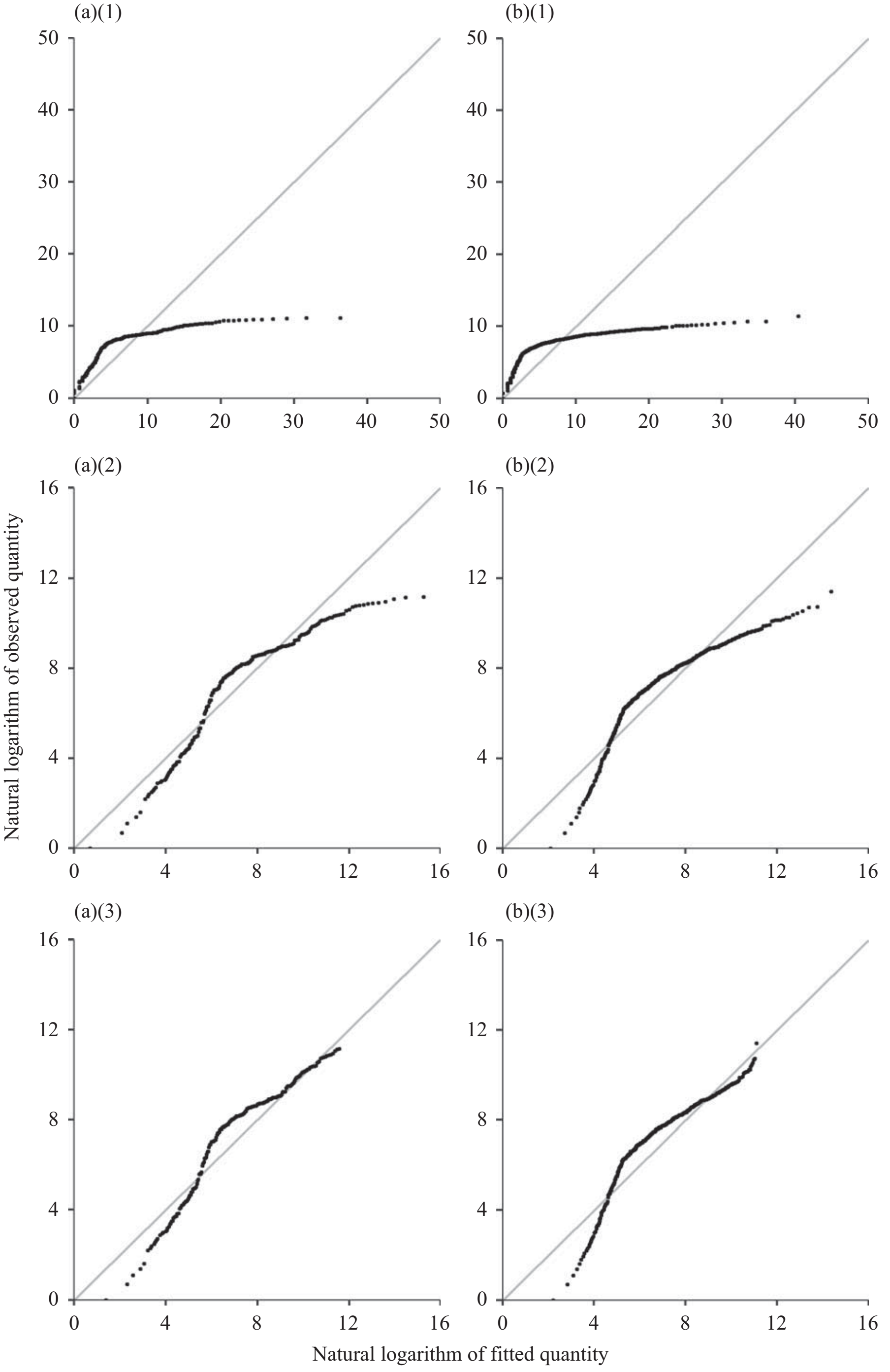


Fig. 4. Observed quantities of (a) toothpaste and (b) shampoo versus those predicted by the (1) Pareto, (2) log-normal and (3) truncated log-normal distributions (log scale), january 2013.

Note: Distributional parameters were estimated using data from January 2012.

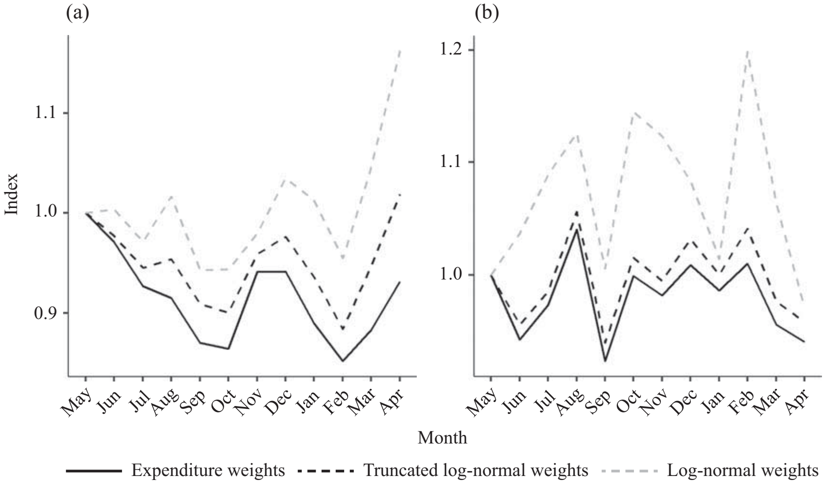


Fig. 5. Directly calculated price indices for (a) toothpaste and (b) shampoo using weights derived from observed and predicted quantities, may 2012 to april 2013.

reduced error associated with using year-old data to parameterise the distributions, for example by using time series models to predict the upper truncation point of the truncated log-normal distribution. Moreover, a notable limitation of our study is that we had access to only two years of historical data for a single retailer for two very similar items (in terms of their product numbers and sales distributions). It is therefore unlikely that our results can reliably be generalised to other time periods, retailers and items, and certainly external validation would need to take place before the methods described in this article could be credibly implemented in the official UK CPIH; for example, an analysis of more recent data for items with very different product numbers and sales distributions to those included in our study would likely prove illuminating.

As stated previously, although the motivation for our study was the absence of quantity data in *web-scraped* data sets, we analysed *transactional* data so that we could calculate an expenditure-weighted index series, against which we could benchmark our results. This necessitated the assumption of the quantity rankings in the analysed transactional data set being coherent with the popularity rankings that would be observed in an analogous web-scraped data set. The validity of this assumption relies on the popularity rankings that appear on retailers' websites being reflective of quantities sold; in reality, this will entirely depend on the algorithms used to define popularity by individual retailers, which may comprise facets other than, or in addition to, sales quantities (for example, the number of page views a product attracts). This uncertainty regarding product rankings scraped from retailers' webpages – a possible source of measurement error which is not captured in our analysis – is one of the most notable limitations of our study. In a sample of 15 of the 'most popular' shampoo products scraped from a particular retailer's website, Auer and Boettcher (2017) found that only two were present in all nine months of the study period, and suggest that this may be attributable to the popularity ranking being used by retailers as a marketing tool to promote certain products.

More broadly, the degree to which changes in web-scraped prices align with those observed in transactional data is unclear, as it is not necessarily the case that a retailer's online prices match those on their physical shelf-edge for an identical product. Only an analysis of both sources of data provided by a single retailer covering a common sample of products over the same time period would facilitate such an inference; however, these data were not available to us at the time of analysis. Cavallo (2017) compared corresponding transactional and web-scraped prices for 24,000 products sold by 56 retailers across ten different countries, finding online and offline prices to be identical 72% of the time and non-significant differences in the mean frequency ($p = 0.07$) and magnitude ($p = 0.10$) of online and offline price changes. In any case, with the growing popularity of internet shopping and the online price becoming increasingly relevant for a large proportion of consumers, it is unclear whether the in-store or online price quote should be viewed as the 'gold standard' in the event of a divergence between the two. The question of how best to account for channel mix in the price index remains beyond the scope of this article.

This article has addressed just one challenge with using web-scraped data for price inflation measurement, that is the lack of expenditure data with which to estimate product-level weights. Of course, a range of other challenges would need to be confronted before such data could be incorporated into the UK CPIH. For instance, not all retailers (or more importantly, not all *types* of retailers) sell their products online. Furthermore, the NSI has no control over what the retailer sells from month to month and human price collectors are not on hand to make on-the-spot substitution decisions, so the missing-data issue of 'product churn' becomes more pronounced when working with data sets obtained from alternative sources; the transitory nature of product availability also means that rankings will not be constant through time, and further investigations into how this phenomenon should be handled in our approach are warranted. The technological challenges posed by web-scraped data are also non-trivial: retailers can put technical measures in place to prevent prices being scraped from their websites, interrupting the supply of data; automatic classification tools must be developed to assign products to items using free-form product descriptions, which will vary in structure and detail between products and retailers; and the quality of the data is limited by the quality of the technology used to obtain it, for example its ability to scrape potentially differing prices for different varieties (sizes, colours, weights, and so on) of the same item.

In conclusion, we believe that using statistical distributions to translate web-scraped sales ranks to expenditure weights using retailer-supplied summary statistics offers an innovative solution to an as-yet largely neglected issue. However, while the approach has demonstrated considerable potential, further research would need to be undertaken before it could be implemented in the UK CPIH. Such research may focus on assessing predictive performance across a broader range of items, using more up-to-date data supplied by a number of different retailers, and investigating the correspondence of product rankings between web-scraped and transactional data sources.

7. References

Auer, J., and I. Boettcher. 2017. "From Price Collection to Price Data Analytics: How New Large Data Sources Require Price Statisticians to Re-think their Index Compilation Procedures. Experiences from Web-Scraped and Scanner Data." In International Working Group on Price Statistics: Fifteenth Meeting of the Ottawa Group, May

10–12, 2017. Eltville am Rhein: Ottawa Group. Available at: https://www.ottawa-group.org/Ottawa/ottawagroup.nsf/4a256353001af3ed4b2562bb00121564/1ab31c25-da944ff5ca25822c00757f87/$FILE/From%20price%20collection%20to%20price%20-data%20analytics%20-Josef%20Auer,%20Ingolf%20Boettcher%20-Paper.pdf (accessed February 2020).

Cavallo, A. 2017. "Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers." *American Economic Review* 207(1): 283–303. DOI: https://doi.org/10.1257/aer.20160542.

Cavallo, A., and R. Rigobon. 2016. "The Billion Prices Project: Using Online Prices for Measurement and Research." *Journal of Economic Perspectives* 30(2): 151–178. DOI: https://doi.org/10.1257/jep. 30.2.151.

Chevalier, J., and A. Goolsbee. 2003. "Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com." *Quantitative Marketing and Economics* 1(2): 203–222. DOI: https://doi.org/10.1023/A:1024634613982.

Hisano, R,. and T. Mizuno. 2010. "Sales Distribution of Consumer Electronics." *Physica A: Statistical Mechanics and its Applications* 390(2): 309–318. DOI: https://doi.org/10.1016/j.physa.2010.09.033.

Hull, I., M. Lof and M. Tibblin. 2017. "Price Information Collected Online and Short-Term Inflation Forecasts." In IFC – Bank Indonesia Satellite Seminar on Big Data: International Statistical Institute Regional Statistics Conference, March 20–24, 2017. Bali: International Statistical Institute. Available at: https://www.bis.org/ifc/publ/ifcb44e.pdf (accessed February 2020).

ILO, International Labour Office. 2020. *Consumer Price Index Manual: Concepts and Methods, Draft January 2020*. Geneva: International Labour Office. Available at: https://www.imf.org/~/media/Files/Data/CPI/cpi-manual-concepts-and-methods.ashx?la = en (accessed May 2020).

Nygaard, R. 2015. "The Use of Online Prices in the Norwegian Consumer Price Index." In International Working Group on Price Statistics: Fourteenth Meeting of the Ottawa Group, May 20–22, 2015. Urayasu City: Ottawa Group. Available at: https://www.ottawagroup.org/Ottawa/ottawagroup.nsf/4a256353001af3ed4b2562bb00121564/d012f001b8a1cf6cca257eed008074c9/$FILE/Ragnhild%20Nygaard%20(Statistics%20Norway-%20The%20use%20of%20online%20prices%20in%20the%20Norwegian%20Consumer%20Price%20Index.pdf (accessed February 2020).

ONS, Office for National Statistics. 2019a. *Retail Sales, Great Britain: April 2019*. Newport: Office for National Statistics. Available at: https://www.ons.gov.uk/businessindustryandtrade/retailindustry/bulletins/retailsales/april2019 (accessed December 2020).

ONS, Office for National Statistics. 2019b. *Introducing Alternative Data Sources into Consumer Price Statistics*. Newport: Office for National Statistics. Available at: https://www.ons.gov.uk/economy/inflationandpriceindices/articles/introducingalternativedatasourcesintoconsumerpricestatistics/2019-05-09 (accessed May 2020).

ONS, Office for National Statistics. 2020a. *Consumer Price Inflation Basket of Goods and Services: 2020*. Newport: Office for National Statistics. Available at: https://www.ons.gov.uk/releases/consumerpriceinflationbasketofgoodsandservices2020 (accessed May 2020).

ONS, Office for National Statistics. 2020b. *Retail Sales, Great Britain: April 2020*. Newport: Office for National Statistics. Available at: https://www.ons.gov.uk/businessindustryandtrade/retailindustry/bulletins/retailsales/april2020 (accessed December 2020).

ONS, Office for National Statistics. 2020c. *Consumer Prices Development Plan: Updated March 2020*. Newport: Office for National Statistics. Available at: https://www.ons.gov.uk/economy/inflationandpriceindices/articles/consumerpricesdevelopmentplan/updatedmarch2020 (accessed May 2020).

Polidoro, F., R. Giannini, R.L. Conte, S. Mosca and F. Rossetti. 2015. "Web Scraping Techniques to Collect Data on Consumer Electronics and Airfares for Italian HICP Compilation." *Statistical Journal of the IAOS* 31(2): 165–176. DOI: https://doi.org/10.3233/sji-150901.

Stanley, M.H.R., S.V. Buldyrev, S. Havlin, R.N. Mantegna, M.A. Salinger and H.E. Stanley. 1995. "Zipf Plots and the Size Distribution of Firms." *Economics Letters* 49(4): 453–457. DOI: https://doi.org/10.1016/0165-1765(95)00696-D.

Ten Bosch, O., and R. Griffioen. 2016. "On the Use of Internet Data for the Dutch CPI." In Meeting of the Group of Experts on Consumer Price Indices: United Nations Economic Commission for Europe, May 2–4, 2016. Geneva: United Nations. Available at: https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2016/Session_2_Netherlands_on_the_use_of_internet_data_for_the_Dutch_CPI.pdf (accessed February 2020).

Touzani, S., and R.V. Buskirk. 2015. "Estimating Sales and Sales Market Share from Sales Rank Data for Consumer Appliances." *Physica A: Statistical Mechanics and its Applications* 451(1): 266–276. DOI: https://doi.org/10.1016/j.physa.2016.01.030.

Van Loon, K., and D. Roels. 2018. "Integrating Big Data in the Belgian CPI." In Meeting of the Group of Experts on Consumer Price Indices: United Nations Economic Commission for Europe, May 7–9, 2018. Geneva: United Nations. Available at: https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2018/Belgium.pdf (accessed February 2020).

Willenborg, L. 2017. *Elementary Price Indices for Internet Data*. The Hague: Statistics Netherlands. Available at: https://www.cbs.nl/en-gb/background/2017/25/elementary-price-indexes-for-internet-data (accessed December 2020).

Received May 2020
Revised December 2020
Accepted April 2021

Journal of Official Statistics, Vol. 38, No. 1, 2022, pp. 23–56, http://dx.doi.org/10.2478/JOS-2022-0003

Using Scanner Data for Computing Consumer Spatial Price Indexes at Regional Level: An Empirical Application for Grocery Products in Italy

Tiziana Laureti[1] *and Federico Polidoro*[2]

The importance of constructing sub-national spatial price indexes (SPIs) has been acknowledged in the literature for over two decades. However, systematic attempts to compile sub-national SPIs on a regular basis have been hampered by the labour-intensive analyses required for processing traditional price data. In the case of household consumption expenditures, the increasing availability of big data may change the current approach for estimating sub-national SPIs by considering the use of weighted index formulae. The aim of this paper is twofold: firstly, to review previous literature on sub-national SPIs and secondly to estimate Italian consumer SPIs. To this aim we use scanner data referring to grocery products sold in a random sample of approximately 1,800 Italian outlets belonging to the most important retail chains and including information on prices, quantities and quality characteristics of products at barcode level. Various weighted index formulas are used for calculating consumer SPIs at detailed territorial level and at the lowest aggregation level. Our results show an interesting territorial variability of consumer prices of products sold in large-scale retail outlets across the Italian regions. Overall, the Southern regions appear to have price levels below the national average both for food and non-food products with some interesting exceptions.

Key words: Regional price levels; spatial price indexes; sub-national PPPs; official statistics; scanner data.

1. Introduction

Over the last decade, National Statistical Institutes (NSIs) have been dealing with a variety of new sources in the big data domain which may provide opportunities for delivering a more efficient statistical service and offering new insights into important topics, such as collecting online job advertisement data and evaluating the presence of firms on social media in the context of ICTs used by enterprises. In this framework, web-scraped price data and business transaction data obtained from retailers have proved to be useful for improving official price statistics (see, for example Mehrhoff 2019). Switching from traditional surveys to electronic data delivery reduces the burden for both statistical agencies and retailers while the availability of expenditure data in scanner data sets can be exploited to increase the accuracy of Consumer Price Index (CPI) figures. However, the transition from traditional survey data to scanner data may be a lengthy process, which entails different stages from establishing the first contact with a retail chain to evaluating a method for calculating price indexes in a CPI production environment. When transaction

[1] University of Tuscia, Via del Paradiso 47, 01100, Viterbo (VT), Italy. Email: laureti@unitus.it
[2] Istat, Via Cesare Balbo 16, Roma 00184, Italy. Email: polidoro@istat.it

© Statistics Sweden

data are not available, NSIs are considering web scraping online prices of consumer goods as a feasible alternative. In this case, Chessa and Griffioen (2019) addressed the issue of whether web-scraped data are suitable for price index calculation. Nevertheless, statistical agencies are becoming aware of the potential issues that may be encountered when using scanner data, which is aided by the recent Eurostat guidelines for scanner data acquisition and processing and by holding annual scanner data workshops (Eurostat 2017).

A growing number of EU countries have been using scanner data for computing CPIs by adopting different approaches given that, for compiling the Harmonised Index of Consumer Prices (HICP), Eurostat allows NSIs to develop their own methods both for processing scanner data and for obtaining price indexes for elementary aggregates.

Contrastingly, to the authors' knowledge, only the Italian National Statistical Institute (Istat) has started an official research project for computing sub-national price indexes using scanner data (Laureti and Polidoro 2017), while French National Statistical Institute (Insee) experts conducted a pilot experiment with the aim of calculating indices that measure differences in consumer price levels between different areas of metropolitan France. Using scanner data obtained from 1,833 stores in April 2013, representing approximately 30% of the potential field, Léonard et al. (2019) estimated sub-national Spatial Price Indexes (SPIs) related to industrial food, that is, food products and beverages sold in supermarkets located in 1,330 municipalities in 707 urban areas of metropolitan France. The authors found that dispersion of price levels between regions is limited. The highest prices are found in the Paris region and Corsica, with a magnitude of differences in the order of a few percentage points.

When it comes to SPIs, web-scraped and scanner data have been mainly explored for making international price comparisons (Heravi et al. 2003; Feenstra et al. 2017; Cavallo et al. 2018). Cavallo et al. (2018) showed that online prices can be used to construct quarterly Purchasing Power Parities (PPPs) published in real-time, using a closely-matched basket of goods and identical methodologies referring to a variety of developed and developing economies.

It is worth nothing that PPPs are essentially spatial price index numbers. The concept of purchasing power parity is used to measure the price level in one location compared to that in another location. More specifically, at international level, purchasing power parities of currencies are defined as the number of currency units of a country that can purchase the same basket of goods and services that can be purchased with one unit of currency of a reference currency. PPPs are calculated for product groups and for each of the various levels of aggregation up to and including Gross Domestic Product (GDP).

Heravi et al. (2003) used scanner data for providing estimates of intercountry price parities at basic heading (BH) level, which is defined as a group of similar products. The application was based on approximately one million transactions for television sets in three countries over a two-months period. Feenstra et al. (2017) examined prices and varieties of products at barcode level in various cities in China and the United States (US) and it was observed that product prices tend to be lower in larger Chinese cities which is not the case in the US.

When constructing spatial consumer price indexes, scanner data may enable countries to measure price level differences across regions which is essential for assessing regional disparities in the distribution of real incomes and supporting regional policy making

(Rokicki and Hewings 2019). The importance of constructing consumer SPIs within a country, or sub-national PPPs, plays a crucial role especially in the case of EU member states where regional economic analyses have become essential due to the implementation of EU Cohesion Policy promoting more balanced and sustainable territorial development. In this context, nominal GDP has been conventionally adjusted using a national deflator which could bias the comparison of regional GDP figures and per-capita income in the presence of spatial price differential (Costa et al. 2019). In countries characterised by large territorial differences in prices and quality of products and household characteristics, such as Italy, it is essential to calculate sub-national SPIs in order to assess inequality in the distribution of real incomes and consumption expenditures.

Although sub-national price comparisons are not as widespread as international price comparisons through the International Comparison Program (ICP), which is administered by the World Bank with the collaboration of the OECD, Euostat and other organizations (World Bank 2013, 2020), several NSIs and individual researchers have conducted interesting research studies on the compilation of sub-national SPIs in various countries (Laureti and Rao 2018). In Italy, official experimental regional SPI computations were carried out by Istat in 2010 using CPI price data and *ad-hoc* surveys. Various research studies have focused on measuring consumer price differences across Italian geographical areas, using both traditional CPI data (Cannari and Iuzzolino 2009; De Carli 2010; Biggeri et al. 2017; Montero et al. 2019) and alternative data sources (Menon et al. 2019).

However, systematic attempts to compile sub-national SPIs on a regular basis have been hindered by the labour-intensive analyses required for processing traditional price data, that is, data used for compiling CPIs, and by the costs involved for carrying out *ad-hoc* surveys for collecting price data.

In this context, the use of big data is both a challenging yet feasible solution for solving the difficulties NSIs face when making spatial price comparisons worldwide. Indeed, as already mentioned, after having adopted scanner data in official CPI computation, Istat has been exploring the possibility of using this new source of price data for compiling sub-national SPIs.

This article marks a departure from previous literature on sub-national SPIs in Italy by demonstrating the feasibility of using scanner data for comparing consumer prices at different territorial levels, which are representative of local consumption patterns and comparable on the basis of a set of price determining characteristics. The article also shows how detailed information on the product prices and quantities sold provided by scanner data enables us to use more complex methodological approaches which involve weighting item prices according to their economic importance.

However, in order to produce sub-national SPIs for the total household consumption expenditure, other sources should be considered in addition to scanner data since in this case they only refer to grocery products.

The aim of this article is twofold: firstly, to carry out a review on previous literature regarding sub-national SPIs, with the aim of emphasizing how our research study fits into the broader picture of recent research activities within the area and secondly to estimate Italian SPIs for 2017 using a scanner data set constructed for experimental CPI computation. This data set refers to a stratified random sample of approximately 1,800 hypermarkets and supermarkets belonging to the most important retail chains (95% of large-scale retail trade distribution) and covers 55.4% of grocery products considered in the total retail trade distribution. Since data are available at detailed territorial level, we compared consumer prices for products sold in the large-scale retail outlets among Italian provinces within a region, thus obtaining within-region SPIs. Moreover, we compared product prices among Italian regions by aggregating the group of products included in the scanner data set into food and non-food consumption aggregates.

The article is organized as follows. Section 2 provides a review of the literature on the compilation of sub-national SPIs and PPPs using unit-value prices derived from household expenditure survey and CPI price data. Section 3 discusses the need for computing sub-national SPIs for Italy. Section 4 describes the scanner data and introduces their use for computing SPIs while Section 5 illustrates the methodological approach adopted for comparing consumer prices for products sold in the large-scale retail outlets in Italy. Section 6 shows the results obtained and the main issues concerning future developments regarding official sub-national SPIs for Italy. In Section 7 some concluding remarks are drawn.

2. Sub-National Spatial Price Indexes: A Review

2.1. *Spatial Comparisons Using Unit-Value Prices*

Over the last years, a growing body of research on spatial price comparisons has been conducted in various countries using data from a variety of sources. In this framework, a widely applied approach uses unit-value prices obtained from household expenditure survey data.

Since the early-2000s several studies have been carried out using this source of data for compiling SPIs in Brazil, India, Indonesia, Australia, Vietnam and Italy.

Aten (1999) and Aten and Menezes (2002) focused on spatial differences in food prices in ten Brazilian cities. These measures are based on unit value prices obtained from household expenditure surveys. Besides reviewing practices for price collection, the authors employed various methods, including, Geary-Khamis, GEKS and Fisher indexes, for compiling SPIs. The initial analysis focusing on food prices was extended to include product prices obtained from surveys. The authors found that poor consumers often face different prices than the average consumer. Moreover, price levels for food in some of the poorest cities are higher than those in higher income cities. Aten and Menezes (2002) concluded that the variation in prices and expenditures across cities and across income groups is substantial, thus suggesting the use of income-specific price levels or a set of regional price levels when analyzing poverty levels across Brazilian cities.

Spatial price differences in India were measured and analysed by Coondoo et al. (2004) who estimated regional price index numbers using data from the National Sample Survey on household expenditures which is conducted in India on a regular basis. The method used, based on a specific demand system, is closely related to the country-product-dummy (CPD) model, that is currently considered to be the principal aggregation method under the stochastic approach for spatial index number construction. Spatial price indexes covering 45 food items for 25 states/provinces in India, were computed using variants of the CPD model and standard multilateral methods. The authors estimated regional consumer price indexes separately for three categories of rural and urban households finding robust results which demonstrated the potentialities of the suggested methodology. Majumder, Ray and Sinha have further refined the methodology for compiling regional consumer price indexes for India, Indonesia and Vietnam, thus broadening and improving the analyses with the aim of proposing a unified framework for estimating intra and inter-country spatial price indexes (Majumder et al. 2012, 2015a, 2015b). Referring to the traditional concept of the "true cost of living index", Mishra and Ray (2014) proposed a method for estimating preference-based sub-national SPIs that measure the extent of price variation between regions both in a given time period and over time. The results showed that during the previous two decades spatial price variation had increased steadily in Australia thus demonstrating a high level of heterogeneity in price levels across Australian states.

Majumder et al. (2015b) proposed a uniform analytical framework for calculating both sub-national SPIs and inter-national SPIs. The methodology is based on the use of adjusted unit values, obtained taking into account quality and demographic effects, thus allowing for regional variation in preferences. The authors illustrated the usefulness of preference consistent methods by applying them for comparing living standards between India and Vietnam.

Deaton and Dupriez (2011) also used unit values from household survey data to investigate spatial differences in prices for India and to obtain more reliable urban and rural price differentials to be used in official poverty estimates (Deaton and Tarozzi 2005). The authors suggested to adjust the unit values for quality differences in order to avoid the overstatement of the urban/rural price difference especially in the case of poor countries.

Recently, Menon et al. (2019) estimated Italian regional price parities based on household budget data and "pseudo" unit values to compare living standards between Italian regions. The authors found that average differences in the "true" cost of living between North and South is approximately 30–40% depending on the regions selected for comparison.

However, it is worth noting that the use of unit value prices from household expenditure surveys shows limitations; the main problem being that meaningful unit values are only available for broadly classified food items that reduce actual comparability. Moreover, spatial differences in food prices need to be augmented to include other consumption goods and services in order to provide a comprehensive measure of price level differences.

2.2. *Spatial Comparisons Using Data Collected for CPI Computation*

The main approach for making sub-national price comparisons adopted by NSIs and researchers is based on data collected for the purpose of compiling CPIs.

Early research on sub-national price comparisons was mostly conducted in the US. After the first official measures of sub-national differences in cost of living represented by the U.S. Bureau of Labor Statistics (BLS) standard budgets, developed in the 1940s (Sherwood 1975), in the early 1990s several studies were carried out by the BLS and the Bureau of Economic Analysis (BEA) in order to determine whether CPI price data could be used for making inter-area comparisons across the US (Kokoski 1991) by employing the CPD method in order to deal with item heterogeneity. Törnqvist binary indexes and the Gini-Elteto-Koves-Szulc (GEKS) method were subsequently used to aggregate price relatives over items, thus obtaining inter-area price indexes for higher level commodity aggregates. These pioneering efforts at the BLS and BEA, further improved by Kokoski et al. (1994) and Kokoski et al. (1999) using hedonic CPD models, were later continued by Aten (2006, 2008) finally leading to the regular compilation of spatial price differences in the US through the computation of Regional Price Parities (RPPs) for metropolitan and urban areas of the US. Since 2014, the RPPs and the price-adjusted estimates of regional personal income have become official statistics of the BEA and are being published annually (Aten 2017; Aten et al. 2012, Aten et al. 2015).

The Australian Bureau of Statistics started a research project for compiling SPIs. In 2003, experimental indexes of cost of living in eight Australian capital cities using existing CPI price data and GEKS method were disseminated by also acknowledging that improvements were planned in the future (Waschka et al. 2003). Similarly, Statistics New Zealand has been evaluating the possibility of carrying out spatial price comparisons of prices since 2005 and two experts have been assigned to develop a methodology for constructing spatial cost of living indexes. However, to the authors' knowledge, estimates of sub-national SPIs have never been disseminated (Tam and Clarke 2015).

As regards Europe, several attempts have been made to calculate differences in consumer price levels in various countries, including the United Kingdom (UK), Germany, the Czech Republic, Poland and Italy.

In the UK, using a database constructed by considering the Reward Group's Cost of Living Comparisons and Family Expenditure Surveys, Johnston et al. (1996) found that price levels diverged during the 1980s with London's price level increasing from 5% above the UK average in 1980 to around 7.5% above average a decade later.

At official level, the UK Office for National Statistics (ONS) carried out a one-off exercise in 2000 for obtaining indicative figures regarding price level differences across regions. Subsequently, in 2004, the ONS published estimates of regional price level comparisons for the year 2003 (Fenwick and O'Donoghue 2003). A year later, the ONS produced consumer price level comparisons at regional level based on CPI price data mainly for food items, tobacco, and drinks; supplemented with administrative data sources and more importantly a purpose-designed price level survey for items such as clothing, furniture, electrical goods and travel (Wingfield et al. 2005). The ONS then produced Relative Regional Consumer Price Levels (RRCPLs) for 2010 following more suitable procedures (ONS 2011) which involved data on price observations from the existing monthly CPI collection and regional price surveys which were required for computing Spatial Adjustment Factors for the Eurostat-OECD PPP Program. The methodology employed by ONS for constructing RRCPLs is consistent with the approach used by Eurostat for calculating the PPPs (ONS 2011).

Various research studies have focused on the construction of sub-national SPIs in Germany. Roos (2006) reported on intra-country regional price level differences using a sample of product price levels in 50 German cities in 1993. Blien et al. (2008) analysed whether wage differences between cities and rural areas in western Germany are due to unobserved differences in regional price levels.

Multiple imputation was used for generating prices for all regions since regional prices were available for only 10% of the regions. Recently, Weinand and Von Auer (2020) introduced a multi-stage CPD approach based on German CPI data and on the so-called perfect matches precept. The authors demonstrated that the regionalised structure of German CPI data allows for the computation of an accurate regional price index. Their results reveal considerable price differentials across the 402 regions where the prices in the most expensive region, Munich, are approximately 27% higher than those in the cheapest region.

Several researchers focused on comparing price levels across regions in the Czech Republic and found significant regional differences in market prices of goods, services, as well as housing and rentals. Kocourek et al. (2016) estimated regional price indexes using raw data from the period 2011–2013 at the LAU-1 territorial unit level. Their findings revealed that when indexes for the Czech NUTS-3 regions were aggregated, the variability in price levels was lower than that reported by previous studies. Indeed, Čadil et al. (2014) using data from various sources, such as the monthly consumer price survey published by the Czech Statistical Office and the Institute for Regional Information for data on rents, showed a high level of heterogeneity among Czech regions, with Hlavni mesto Praha being the most expensive region (119.7) and Ustecky kraj the cheapest (94.9). Recently, Musil and Kramulová (2019) used price statistics (collected regional average prices, detailed weights of representatives) and national accounts (housing services, structure of expenditure in division by COICOP) for updating previously estimated regional price levels in 14 Czech NUTS-3 regions by the same authors. The results covered therefore the years 2007, 2012, and 2017 confirming spatial heterogeneity in consumer price levels across Czech regions. The highest price level remained in the capital city Praha while the Săedočeský and Jihomoravský regions showed prices above the average in all years. Contrastingly, lower price levels were observed in structurally affected regions.

In Poland, Rokicki and Hewings (2019) analyzed regional price differentials using unique raw-price data for calculating regional PPP deflators for the 16 NUTS-2 regions. In accordance with expectations, the authors found that the price levels are significantly higher than average in regions with large metropolitan areas (e.g., Mazowieckie with the capital Warsaw) and lower in regions without large cities (e.g., Eastern border regions such as Lubelskie or Podlaskie).

Using data from the Spanish National Statistical Institute, Alberola and Marques (2001) analyzed relative consumer prices for the Spanish regions and found that deviations in relative prices tend to be small but persistent.

As regards the Italian sub-national SPI calculation, at official statistical level, Istat launched the project for estimating sub-national PPPs in the National Statistical Program 2008–2010. The Commission for the Guarantee of Statistical Information (COGIS), voiced its opinion on the 2008–2010 National Statistical Program in the meeting of 5 July 2007 by stating that the most relevant innovation in the field of price statistics was represented by the survey of regional PPPs.

In accordance with the National Statistical Program 2008–2010, Istat in cooperation with *Unioncamere* and "Tagliacarne" Institute launched a research project aimed at exploring the possibility of using and integrating the statistical information currently supplied by CPI surveys.

In its first attempt to estimate regional SPIs, Istat found considerable 2006 price differences for a subset of CPI products. Indeed, using the GEKS formula, sub-national SPIs for the 20 Italian regional chief towns were computed for three expenditure divisions

(Food and Beverages, Clothing and Footwear, Furniture), which represented approximately 34% of the total consumer expenditure (De Carli 2010). Moreover, *ad-hoc* surveys were designed and carried out for "Clothing and Footwear" and "Furniture" product groups. Subsequently, the Bank of Italy extended price comparisons to other categories of goods and services estimating the difference in price levels between Northern and Southern Italy at approximately 16–17% (Cannari and Iuzzolino, 2009).

In its second attempt, Istat estimated regional PPPs referred to the entire 2009 CPI basket, using the same source of data and methodology with the exception of rents. For this category spatial comparisons were carried out using CPD models and Household Budget Survey data that includes detailed information on dwelling characteristics (Istat 2010). The results obtained confirmed the significant differences across Italian regions and in particular between Northern and Southern Italian regions although difficulties emerged regarding the comparability of elementary data.

Further research studies based on CPI data have been conducted and have provided interesting results for Italy. Biggeri et al. (2017) considered CPI data and focused on seven BHs belonging to the food and non-alcoholic CPI group with the aim of understanding whether to use elementary price quotes or average outlet prices and investigating the performance of various CPD models estimated considering data obtained using various levels of aggregation.

Montero et al. (2019) proposed a spatially-penalized country-product-dummy model for making spatial price comparisons while accounting for the presence of spatial dependencies in consumer prices using official Italian CPI data for constructing regional spatial price indexes for 2014. The results showed that price levels are higher in the Northern-Central regions of Italy than in the South.

The results obtained and the growing demand by users prompted Istat to confirm the official project for producing sub-national PPPs on a regular basis and exploit the possibilities offered by the new data sources.

3. The Need for Computing Sub-National Spatial Consumer Price Indexes for Italy

Several players in the economic and social debate have acknowledged the need for compiling sub-national household consumption purchasing power parities for Italy due to the high socio-economic heterogeneity across its geographical areas.

Compared to other OECD countries, the Italian regions vary greatly in terms of household economic conditions. Indeed, Italy has an almost dualistic economy with all Southern Italian households attaining, on average, lower level of income than households residing in Central and Northern regions. This situation is clearly demonstrated by some key indicators. Looking at the NUTS-1 geographical level on the side of expenditures, the differences amongst the Italian macro-areas are remarkable. In the North-Western Italy the average monthly expenditure of households in 2017 (net of imputed rents) amounted to 2,249.19 euros (the highest monthly expenditure) whereas in the Islands it was 30% less at a value of 1,558.04 euros (the lowest monthly expenditure) while the values in the North East, Centre and the South amounted to 2,186.11; 1,970,56 and 1,638.86 euros respectively.

At regional level (NUTS-2), in 2017 the monthly expenditure of households residing in the autonomous province of Bolzano/Bozen was, on average, almost double of the monthly expenditure of Calabrian households (Istat 2018a). It is worth noting that Italy is divided into 21 NUTS-2 regions (Trentino Alto Adige is split in two autonomous provinces of Trento and Bolzano) each of which has its own regional capital (specified in brackets): Aosta Valley (Aosta), Piedmont (Turin), Liguria (Genoa), Lombardy (Milan), Trentino-Alto Adige (Trento), Trentino-Alto Adige (Bolzano), Veneto (Venice), Friuli-Venezia Giulia (Trieste), Emilia-Romagna (Bologna), Tuscany (Florence), Umbria (Perugia), Marche (Ancona), Lazio (Rome), Abruzzo (L'Aquila), Molise (Campobasso), Campania (Naples), Apulia (Bari), Basilicata (Potenza), Calabria (Catanzaro), Sicily (Palermo), Sardinia (Cagliari).

Indicators concerning absolute poverty (measured according to household expenditure) confirmed the differences observed across the Italian territory. Before the two economic crises, in 2007, the absolute poverty rate in terms of households was 3.1% in the North and 4.6% in the South and Islands. In 2017, the absolute poverty rate in Italy increased as well as the gap between the two geographical macro-areas reaching the value of 10.3 in the South and Islands and 5.4 in Northern Italy (Istat 2018b). Even if in 2019 this gap appeared to decrease (8.6% versus 5.8%) it still indicates a considerable disparity (Istat 2020). The North-South gap widens if we consider the relative household poverty rate (in expenditure terms). In 2017, this rate was 5.9% in the North and 24.7% in the South while in 2019 it increased to 6.8% in the Northern regions and decreased to 21.1% in the South, thus indicating that in 2019 Southern households in relative poverty remained three-fold higher than those recorded in the North whereas households in absolute poverty were approximately 1.5 times higher than those in the North.

These differences between the absolute and relative poverty rates together with household expenditure level data provide further proof that measuring consumer price differences in Italy is essential. Indeed, Istat computes absolute poverty indicators taking into account thresholds that differ by municipality typology and by geographical macro-areas. In terms of geographical areas, the differences amongst the thresholds are mainly due to the consumer price component that was considered in their estimation, thus the comparison of these thresholds between different geographical areas represents a proxy of "poverty purchasing power parities". As regards relative poverty, the threshold is uniform for the entire Italian territory and therefore does not take the differences of prices amongst regions into consideration.

It is clear that these indicators combined with the evidence coming from the current survey on consumer prices for estimating inflation, prove that household purchasing power differs across Italian regions and that consumer prices are mainly responsible for this difference.

This situation is confirmed by the aforementioned estimates of SPIs carried out by Istat that showed significant differences in consumer price levels across the regional capitals (Istat 2010). Consumer price levels in the Northern cities were generally higher than those recorded in the Centre and especially in the South. Bolzano (105.6) and Milan (104.7) showed the highest prices compared with the Italian average (100) while the least expensive city proved to be Naples (93.8).

An extra motivation to study this topic was prompted by Istat use of scanner data for compiling official CPIs. At the same time, Istat launched a research project for exploring the possibility of using them for compiling sub-national SPIs.

4. Using Scanner Data for Computing Sub-National SPIs: The Case of Italy

Since 2014 scanner data have been regularly provided to Istat by the market research company ACNielsen which is authorized by large-scale retail chains thanks to an agreement between Istat and ADM, the Association of large-scale retail trade distribution.

Scanner data are transmitted by ACNielsen on a weekly basis. The itemized information contained in these data, that is, turnover and quantities for each well-specified item code, may be fruitfully used in order to compile weighted spatial price indexes at detailed territorial level and on a regular basis.

As already mentioned, over the last decade there has been a growing interest in using scanner data for constructing official CPIs thus increasing the availability of this new data source. Several EU countries have acquired scanner data either directly from retail chains or indirectly from market research companies, such as Nielsen and GfK. Regardless of which provider scanner data come from, NSIs must reclassify them in order to make them suitable for constructing official CPIs and various procedures are followed to achieve this. Since the primary use of scanner data is not to measure temporal and spatial price differences, methodological and empirical issues regarding scanner data quality need to be addressed. This reclassification and data cleaning phases are essential yet IT resource intensive.

Nevertheless, scanner data may help to overcome the issue of price data availability in the various areas involved in spatial price comparisons thus fulfilling the requirements of representativeness and comparability that emerge when compiling sub-national SPIs. Due to the high territorial coverage which characterizes scanner data, it is possible to compare price levels at different territorial levels within a country (NUTS-3; NUTS-2; NUTS-1). Moreover, it is worth noting that GTIN codes uniquely distinguish products and they are generally the same for each item at national level. In this way, the issue of comparability can be solved. As detailed information on turnover and quantities for each item code in every city are available, it is possible to account for the economic importance of each item in its own market, thus fulfilling the representativeness requirement. Moreover, as different large-scale retail distribution chains can sell products of different quality and offer additional services, information on the type of outlet and retail chain can be included in order to account for the "quality" characteristics that may influence the price of a product. Finally, using scanner data to carry out spatial comparisons will increase cost efficiency since price data collection may be limited to traditional outlets thus lowering data collection costs for the NSIs.

The main disadvantage of using scanner data is that they usually only cover products sold in large-scale retail distribution chains. Therefore, they must be combined with data obtained from other sources to cover the entire consumption basket.

From a methodological point of view, when using scanner data, it is possible to consider a wide range of methods for calculating spatial price indexes due to the availability of quantities and expenditure information (Heravi et al. 2003). In this way, we can improve the quality of sub-national SPIs since we can also use the Geary-Khamis method and weighted CPD methods at the lowest level of aggregation (BH level), at which expenditure and expenditure share data are not usually available from traditional data sources (World Bank 2020). In addition, scanner data can provide a basis for further statistical developments in the estimation of sub-national SPIs. Indeed, detailed price and quantity data are available on a monthly basis thus allowing us to add the time dimension to multilateral price comparisons.

In this article we use a scanner data set constructed for experimental CPI computation after having carried out data cleaning and outlier trimming processes (Bernardini et al. 2019). The data set used refers to the year 2017 and to a stratified random sample of outlets and a cut-off sample of items within each outlet.

More specifically, the universe of 9,000 outlets belonging to the 16 most important retail chains (95% of large-scale retail trade distribution) has been stratified by province, distribution chains and type of outlets, thus obtaining 888 strata. Outlets are selected within each stratum, with probabilities proportional to the 2016 turnover.

Therefore, the sample is composed of 510 hypermarkets and 1,271 supermarkets for a total of 1,781 outlets located in the 107 Italian provinces.

The identification of the items is based on barcodes (GTINs), which, as already mentioned, univocally classify the products across the entire national territory. In each outlet, items were selected using cut-off sampling: the most sold (in terms of 2016 turnover) GTINs (sampling units) were selected in order to cover up to 60% of the total turnover of the product aggregate (BH) to which the GTINs belong. Although this data set was constructed for CPI compilation, it can also be used for making spatial price comparison among Italian regions bearing in mind that only the best selling products, which are typically consumed in each Italian province and region, may have be included according to the CPI selection procedure. Therefore, these products may not be strictly comparable across different provinces and regions. Nevertheless, we found a reasonable overlap among the different geographical areas which is also characterized by a chain structure. It is worth noting that perishables and seasonal products such as vegetables, fruit and meat were excluded from the scanner data because these products are sold at price per quantity and are not pre-packaged with GTIN codes.

Since data are available for the 107 Italian provinces, within-region and between-region SPIs were estimated both for each group of products and for higher product aggregates, that is, food and non-food consumption aggregates which are well covered by our scanner data set (55.4% of total retail trade distribution for this category of products). To this aim we referred to a list of grocery products derived from the data set that specifically covers 54 BHs which corresponds to product aggregates according to ECOICOP classification (8 digit), belonging to five expenditure divisions of the ECOICOP (01, 02, 05, 09, 12).

Taking into account the results obtained by Laureti and Polidoro (2017), we decided to use the finest classification of item that is available within each BH, that is, the product barcode (GTINs) and compute weighted average of weekly unit values for each outlet. Indeed, Laureti and Polidoro (2017) focused on aggregation issues when comparing consumer prices across space as the choices made may reflect different implicit assumptions regarding consumer behaviour. In order to understand how best to aggregate the detailed information contained in the Italian scanner data for constructing sub-national SPIs, several analyses were carried out. Results showed significant differences in prices of the same item across outlets thus suggesting product differentiation which is embodied in the range and/or quality of services offered by different retailers, both across chains and across stores within the same chain. This computation involves the use of the formula

$$p_y = \sum_{w \in y} p_w \left(\frac{p_w q_w}{\sum_{w \in y p_w q_w}} \right)$$

where p_w is the weekly price of the specific GTIN, calculated as the total turnover for that item code divided by the total quantities sold over the week, and $p_w q_w$ is the outlet turnover for GTIN product in each week. Therefore, we obtain an annual GTIN unit value for each outlet by using:

$$p_y = \sum_{w \in y} \frac{p_w q_w}{q_w} \left(\frac{p_w q_w}{\sum_{w \in y p_w q_w}} \right) = \sum_{w \in y} p_w \left(\frac{p_w q_w}{\sum_{w \in y p_w q_w}} \right).$$

Provincial averages of the results obtained in the previous step were then calculated as a weighted arithmetic mean according to the aforementioned sampling design and the probability proportional to size (PPS) sampling approach within each stratum. The inclusion probability of each outlet in the strata is proportional to the potential of the outlet, as measured by the 2016 turnover (Bernardini et al. 2019), so that larger outlets are more likely to be selected than smaller units. Our choice was to follow the probabilistic sampling procedures adopted by Istat for the selection of outlets and products for the purpose of CPI computation. More specifically, we estimated the provincial average price of a specific item across different outlets and strata as a weighted mean by using outlet weights that reflect the probability of selection for outlets in the different size-strata. Therefore, within each stratum, by giving relatively smaller weights to larger outlets we can compensate for the higher probability of inclusion of larger outlets (due to PPS) in the sample (as in the case of CPI compilation). However, the final weight by which the outlet contributes to the provincial average price also considers the size of the stratum (defined by considering province/type of outlet/chain). Moreover, since the number of strata considered in the sampling design is very high, equal to 888, in several cases the outlets selected in the strata (in total 1,781 in 2017) are auto-representative. As a result, 487,094 annual provincial GTIN prices were obtained.

It is worth noting that the structure of the Italian food market is quite complex and, as shown in Table 1 the large-scale retail trade distribution is not uniformly distributed across Italian territory in terms of types of retail chains and market share, which may influence the results.

As already mentioned, we are aware that our data do not cover the complete household monetary consumption expenditure and only refer to products sold in large-scale retail outlets, thus the estimated SPIs cannot be generalized to the entire consumption basket. Nevertheless, by considering the results obtained from an analysis carried out in 2019 by Istat and referring to the Italian Households Budget Survey, it is possible to draw some interesting considerations. Since 2015, Istat has asked households to indicate the type of outlets where they have purchased a list of 25 broadly defined products (including bread, pasta, biscuits, milk, eggs, toys, medicine, and so on). The results confirm that even if consumers tend to buy certain categories of products outside supermarkets (i.e., fresh

Table 1. Scanner data: % market shares (hypermarket + supermarket) – year 2016.

	NORTH - WEST				NORTH - EAST				CENTER				SOUTH AND ISLANDS								
RETAIL CHAINS	PIEDMONT	AOSTA VALLEY	LIGURIA	LOMBARDY	TRENTINO-ALTO ADIGE	VENETO	FRIULI-VENEZIA GIULIA	EMILIA-ROMAGNA	TUSCANY	UMBRIA	MARCHE	LAZIO	ABRUZZO	MOLISE	CAMPANIA	PUGLIA	BASILICATA	CALABRIA	SICILY	SARDINIA	ITALY
COOP ITALIA	18.2	-	42.2	7.9	18.0	9.1	21.3	41.2	51.2	30.8	18.5	14.3	10.0	-	4.4	18.6	6.9	-	6.3	-	18.5
CONAD	4.3	22.3	17.0	3.3	13.8	3.6	7.7	26.5	14.8	29.9	12.6	24.5	29.8	30.9	20.5	9.6	10.3	30.2	19.5	30.6	13.3
ESSELUNGA	12.4	-	3.9	31.3	-	1.2	-	9.9	22.1	-	-	0.9	-	-	-	-	-	-	-	-	12.1
SELEX COMMERCIALE	17.9	8.6	4.8	9.9	-	32.3	9.4	6.6	1.1	22.1	18.2	3.4	2.7	23.4	7.6	29.1	6.0	3.3	4.4	12.8	11.1
GRUPPO AUCHAN	7.0	-	0.7	8.2	-	6.3	1.1	1.5	1.9	2.7	25.8	10.7	11.1	-	8.1	17.2	10.4	17.3	20.1	12.6	7.8
GRUPPO CARREFOUR ITALIA SPA	16.4	45.1	8.8	9.9	-	2.1	4.2	1.8	2.8	0.7	0.9	13.3	5.7	1.6	9.2	-	0.9	8.9	1.5	5.6	7.1
FINIPER	1.5	-	-	6.4	-	1.6	2.9	1.4	-	-	4.1	-	8.3	-	-	-	-	-	-	-	2.3
GRUPPO VEGE	-	-	1.5	1.1	-	6.2	-	0.2	0.1	0.2	-	0.7	2.6	5.7	20.7	1.2	5.0	4.0	19.8	13.8	3.2
GRUPPO SUN	1.4	-	3.2	2.6	-	2.0	1.2	0.3	-	2.4	9.8	14.4	18.2	27.6	-	-	-	-	-	-	3.1
AGORA' NETWORK SCARL	2.5	-	13.5	6.1	34.4	0.4	-	0.2	0.2	-	-	-	-	-	-	-	-	-	-	-	2.8
GRUPPO PAM	3.7	-	2.7	0.9	0.6	3.1	8.0	1.8	5.4	3.1	-	8.5	0.7	-	0.2	1.4	-	-	-	3.8	2.7
ASPIAG	-	-	-	-	32.4	12.7	29.9	1.8	-	-	-	-	-	-	-	-	-	-	-	-	2.7
BENNET SPA	8.7	-	1.3	5.2	-	1.2	4.0	1.9	-	-	-	-	-	-	-	-	-	-	-	-	2.5
SIGMA	0.1	-	-	1.1	-	2.8	2.6	3.0	0.3	0.3	7.0	0.8	3.2	6.4	2.8	6.9	5.3	1.6	1.1	5.0	1.8
CRAI	1.6	-	0.3	0.2	-	2.6	2.1	0.5	0.0	-	0.4	1.7	0.7	0.9	2.3	0.2	5.4	3.5	7.5	9.7	1.4
DESPAR SERVIZI	-	-	-	0.6	-	-	-	-	-	-	-	0.0	-	-	1.8	7.1	17.6	18.4	6.2	4.3	1.2
TOTAL	**95.9**	**76.0**	**99.8**	**94.8**	**99.1**	**87.0**	**94.3**	**98.5**	**99.9**	**92.2**	**97.4**	**93.2**	**92.9**	**96.6**	**77.5**	**91.3**	**67.9**	**87.2**	**86.4**	**98.0**	**93.7**

fish, bread, medicine and toys), the most suitable places for shopping are always super and/or hypermarkets with percentage (across households that have declared to buy that specific product) ranging from 45.4% for bread and 50% for fresh fish to 75% for pasta and 79% for yogurt. Therefore, turnovers can be considered reasonable indicators of consumption baskets, especially for the product categories considered in our study.

5. Methodological Approach

5.1. Overview of the Procedure for Calculating the Sub-National SPIs

In order to estimate regional spatial indexes for products sold in large-scale retail trade chains by using data for the 107 Italian provinces, a two-step procedure similar to the one used in the ICP was adopted whereby provinces are grouped into regions (World Bank 2020). In the first step, within-region SPIs are computed by comparing price and quantity data referring to products sold in the various provinces within each region, while in the second step, between-region SPIs are obtained for each region by using provincial prices adjusted for differences among provinces within the region.

Moreover, as in international practice, sub-national SPI compilation is undertaken at two levels, viz., at BH level and at a more aggregated level (food and non-food products, the latter including all the personal and home care goods belonging to grocery products sold in large-scale retail chains). The methods selected for making multilateral comparisons is based on several axiomatic properties, including two basic properties: transitivity and base region invariance. Transitivity simply means that the SPI between any two regions should be the same whether it is computed directly or indirectly through a third region. The second requirement is that the SPIs be base region–invariant, which means that the SPIs between any two regions should be the same regardless of the choice of base region.

Figure 1 illustrates the overall process we followed for constructing regional SPIs. Price comparisons begin with the identification of products in the scanner data set for which

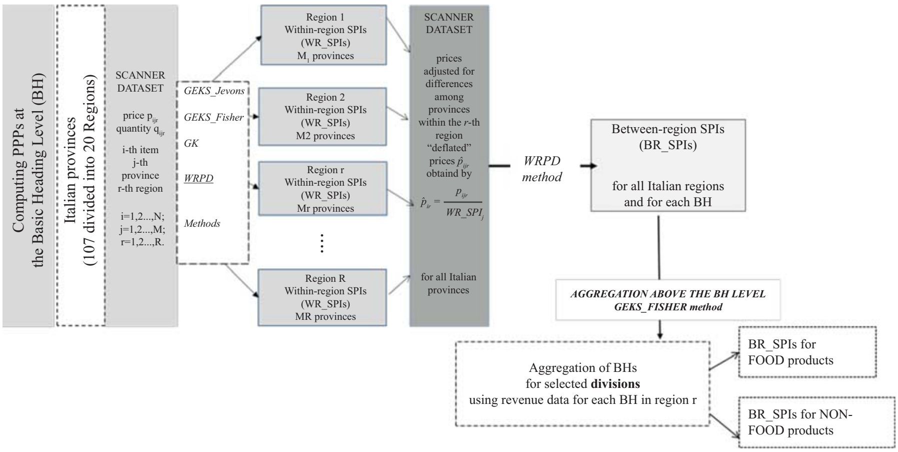


Fig. 1. Methodology for computing SPIs for food and non-food products.

price and quantity data are available in the various provinces that make up each region. These individual products are grouped into product groups (ECOICOP 8 digit) which are called basic headings. Even if it is not necessary to price all the listed products in all of the provinces, reliable regional price comparisons can be made as long as there is a reasonable overlap in the items priced in different provinces. However, a significant overlap of products can be obtained when comparisons are made in a region in which all the provinces are fairly similar.

With the aim of compiling regional SPIs for each BH while taking into account the variability of price levels among provinces within a region, firstly provincial prices and expenditure should be adjusted for differences among provinces within the region ("deflated" prices) using within-region SPIs. To this aim we compute within-region SPIs for each Italian region using various methods and the weighted-expenditure regional product dummy model were selected on the basis of the results obtained in the empirical study which confirmed its good theoretical properties.

As a result of the first step, "deflated" prices and expenditures may be used for computing between-region SPIs for each BH using the weighted region product dummy model.

Finally, we aggregated the SPI results for each BH using the GEKS Fisher method.

5.2. *Aggregation Methods at BH Level*

5.2.1. First Step: Within-Region SPIs

Let us assume that we are attempting to make a spatial comparison of prices between M_r provinces in each region r. In the first stage of aggregation of price data at item level, which leads to price comparisons at BH level, p_{ij} and q_{ij} represent price and quantity of *i-th* item in *j-th* province in *r-th* region with $i = 1,2 \ldots, N_j$; $j = 1,2, \ldots, M_r$; $r = 1,2, \ldots R$.

In order to compute within-region SPIs, we explored different spatial index formulae with the aim of demonstrating the feasibility of implementing various aggregation methods given that expenditure share weights can be used for constructing SPIs at elementary level when scanner data are available. Starting from methods based on binary comparisons between provinces, it is important to note that any binary comparison between provinces j and k can be made by overlapping price data referring to commonly priced items. If a commodity is not priced in one of the two provinces, that item cannot be included in the SPI computation. It is worth noting that gap-filling prices even when quantities are zero in certain regions may be useful in the case of Fisher with the aim of using more information contained in the scanner data set. To this aim, we also computed the Fisher index using the gap-filled price matrix. We decided to select the Fisher based on commonly priced items to compare the results with those obtained from the unweighted Jevon index. We considered the unweighted GEKS-Jevons index since it is used within the Eurostat-OECD PPP Program and the GEKS-Fisher index to analyse the effect of using expenditure share weights. Turning to the multilateral index number methods, we considered the GK method since it is not usually applied at the basic heading level, due to the lack of weighs, even if SPIs based on GK indexes are commensurable, base invariant and transitive (Eurostat-OECD 2012). Finally, we used the RPD (the regional version of the CPD) and the WRPD as they directly refer to the stochastic approach to index numbers and are adopted within the ICP.

Unweighted Jevons Index

As already mentioned, we firstly used a similar method to the one adopted by the OECD-Eurostat for making international price comparisons based on binary comparisons between countries (Eurostat-OECD 2012). Therefore, we calculated the unweighted Jevons indexes, based on commonly priced items between provinces j and k:

$$WR\_SPI_{jk}^{Jevons} = \prod_{i \in N_{jk}} \left[\frac{p_{ik}}{p_{ij}}\right]^{1/N_{jk}} \tag{1}$$

where N_{jk} represents the number of items in the BH that are commonly priced in provinces j and k. In this case, all items in the BH are treated as equally important. The Jevons index yields transitive comparisons only when all provinces price all products in the basic heading (a complete tableau). In the most general case in which not all items are priced in all provinces (incomplete price tableau) the GEKS procedure must be adopted in deriving transitive comparisons from a matrix of binary comparisons based on the Jevons formula, thus obtaining $WR\_SPI_{jk}^{GEKS\_Fisher}$.

Since the best practice index number theory typically involves weighting prices according to their economic importance, we computed the Fisher binary index using price and quantity data for commonly priced items in order to avoid those prices representing large expenditures are given the same weight as those representing small expenditures.

Fisher Index

The Fisher price index is known to have a range of good axiomatic and economic theoretic properties (Diewert 1976; Balk 2008). It satisfies all the axiomatic properties expected of a price index number formula with the exception of the circularity test and it is also known to be a superlative index number. The Fisher index is the geometric average of the Laspeyres and Paasche index numbers and is given by:

$$WR\_SPI_{jk}^{Fisher} = \sqrt{P_{jk}^{Laspeyres} \cdot P_{jk}^{Paasche}} \tag{2}$$

where

$$P_{jk}^{Laspeyres} = \frac{\sum_{i \in N_{jk}} p_{ik} q_{ij}}{\sum_{i \in N_{jk}} p_{ij} q_{ij}} \tag{3}$$

$$P_{jk}^{Paache} = \frac{\sum_{i \in N_{jk}} p_{ik} q_{ik}}{\sum_{i \in N_{jk}} p_{ij} q_{ik}} \tag{4}$$

In order to construct transitive multilateral comparisons from a matrix of binary comparisons derived using Fisher indexes, which do not satisfy the transitivity property, the GEKS procedure must be adopted.

The GEKS procedure

This method suggested by Gini, Eltetö and Köves and Szulc can be used for aggregating price data (that is, at the BH level) and above the BH level. Details of the GEKS procedure can be found in Balk (2009) and Rao (2009). Using the GEKS approach it is possible to obtain a transitive index that deviates the least from a given matrix of binary comparisons. If non-transitive binary indexes are indicated by $SPI_{jk}\ (j, k = 1, 2, \ldots, M_r)$ then the GEKS transitive indexes are obtained by minimizing:

$$\sum_{j=1}^{M_r} \sum_{k=1}^{M_r} \left[lnSPI_{jk}^{GEKS} - lnSPI_{jk} \right]^2$$

$$\text{subject to } SPI_{jk}^{GEKS} = SPI_{jl}^{GEKS} \cdot SPI_{lk}^{GEKS} \, for\, all\, j, k.$$

The GEKS-based SPIs are obtained as an unweighted geometric average of the linked (or chained) comparisons between provinces (or regions) j and k using each of the provinces (or regions) in the comparisons as a link.

$$SPI_{jk}^{GEKS} = \prod_{j=1}^{M_r} \left[SPI_{jl}^{GEKS} \cdot SPI_{lk}^{GEKS} \right]^{1/M_r} \tag{5}$$

Using Equation (5) we obtained the GEKS based on Jevons index**,** $WR\_SPI_{jk}^{GEKS\_Jevons}$ and the GEKS based on Fisher index, $WR\_SPI_{jk}^{GEKS\_Fisher}$.

In addition to GEKS based on bilateral Jevons and Fisher indexes, we used the Geary-Khamis and RPD methods for making multilateral comparisons (for a review of the methods see Weinand 2020; World Bank 2020; Laureti and Rao 2018). All methods were implemented using *ad-hoc* R script codes.

(A) Geary-Khamis Index

The Geary-Khamis (GK) method, proposed by Geary (1958) and Khamis (1972) has been the principal aggregation method used for cross-country price comparisons since the inception of the ICP in 1968. The GK method was only replaced by Fisher-based GEKS formula in the 2005 and 2011 ICP comparisons. At international level, the GK method uses the twin interdependent concepts of purchasing power parities of currencies and average international prices of commodities.

Theoretically, the original GK method provides a way of calculating international PPPs from price and quantity data for individual products when such data are available. In practice, the input data are not prices and quantities for individual products, but estimated prices and quantities for basic headings that comprise sets of products (Eurostat-OECD 2012).

However, in the case of scanner data and for computing sub-national SPIs, the GK method can be applied at the lowest level of aggregation since price and quantity data are available for individual products.

At sub-national level, the GK method generates values of the unknown spatial price indexes and "regional" prices from the solutions obtained from a system of linear homogeneous equations that define the regional prices and spatial price indexes as functions of the observed price and quantity data across provinces (Rao and Salvanathan 1992). In other words, the GK method entails valuing a matrix of quantities using a vector of "regional" prices which is obtained by averaging provincial prices, after they have been adjusted to a uniform price level by spatial price indexes, across the group of provinces compared.

The GK method works as long as each commodity is consumed in at least one province and each province has at least one commodity with positive quantities.

Technically, the method is defined through the system of interrelated linear equations that are used for all products and all provinces in the comparison:

$$\begin{aligned} WR\_SPI_j^{GK} &= \frac{\sum_{i=1}^{N_j} p_{ij}q_{ij}}{\sum_{i=1}^{N_j} P_i q_{ij}} \qquad j = 1, 2, \ldots, M_r \\ P_i &= \frac{\sum_{j=1}^{M_r} \left(p_{ij}q_{ij}\right)/WR\_SPI_j^{GK}}{\sum_{j=1}^{M_r} q_{ij}} \qquad i = 1, 2, \ldots, N_j \end{aligned} \tag{6}$$

The solution is unique up to a factor of proportionality; therefore, an additional normalization is required in order to uniquely determine the spatial indexes. To this aim, SPI for regional capital is set equal to 1. These equations are solved using an iterative procedure for obtaining $WR\_SPI_j^{GK}$.

(B) Regional Product Dummy model (RPD)

Unweighted RPD

The unweighted RPD is the regional version of the CPD method used in international comparisons. The basic statistical model underlying the RPD method can be stated as $p_{ij} = SPI_j \cdot P_i \cdot u_{ij}$ $i = 1, 2 \ldots, n; j = 1, 2, \ldots, M_r; r = 1, 2, \ldots R$, n represents the number of items in the BH that are priced in the various provinces included in the comparison, SPI_j is the spatial price index of the *j-th* province relative to the other provinces within the region; P_i is the "regional" average price of the *i-th* commodity; and u_{ij}'s are independently and identically distributed random variables. In this study, these disturbances are assumed to be lognormally distributed. The RPD model can be best described as a hedonic regression model in which the characteristics used are the province and the commodity specifications. By taking natural logs on both sides, price levels are estimated by regressing logarithms of prices on provinces and product dummy variables. The model is given for each BH and region r by:

$$\begin{aligned} lnp_{ij} &= lnSPI_j + lnP_i + lnu_{ij} \\ &= \sum_{j=1}^{M_r} \pi_j D^j + \sum_{i=1}^{n} \gamma_i D^i + v_{ij} \end{aligned} \tag{7}$$

where D^j is a provincial-dummy variable that takes value equal to 1 if the price observation is from *j-th* province in the *r-th* region; D^i is a dummy variable that takes value equal to 1 if the price observation is for *i-th* commodity and v_{ij} are normally distributed with a zero mean and a constant variance.

Parameters of this model can be estimated once one of the parameters of the model is set at a specified value. For example, if province 1 is taken as the reference or numeraire region then π_1 is set at zero and remaining parameters are estimated. If $\hat{\pi}_j$ $\left(j = 2, \ldots, M_r\right)$ are estimated parameters, the within-region SPI for the province j in region r is given by:

$$WR\_SPI_j^{RPD} = e^{\hat{\pi}_j} \tag{8}$$

The RPD model produces a transitive set of SPIs taking into account all the price information in a single step.

Furthermore, it is easy to demonstrate that when all the items are priced in all the provinces then the RPD based price comparisons are identical to the Jevons-based comparisons in Equation (1) (see Rao 2004 for a proof of this statement).

Weighted Regional Product Dummy (WRPD)

With the aim of taking into account the economic importance (representativeness) of each product, we used a weighted regional product dummy model (Rao 2001). As underlined by Diewert (2004) the economic importance may be measured adopting different methods, including the use of quantities transacted or expenditures pertaining to each component. However, since we are comparing prices across large and small provinces (or regions), we decided to weight the importance of each commodity price by its share in the province's expenditures in the class of commodities under consideration thus avoiding to give too much weight to the larger provinces (Rao 2001). The choice of using expenditure share instead of quantity share weights is also based on the results reported by Silver (2003) who concluded that expenditure share weights are the most appropriate set of weights for hedonic regressions and the RPD can be viewed as a very simple type of hedonic regression model where the only characteristic of the commodity is the commodity itself (Diewert 2004). The weighted RPD model is expressed by:

$$\sqrt{w_{ij}}lnp_{ij} = \sum_{j=1}^{M_r} \pi_j \sqrt{w_{ij}} D^j + \sum_{i=1}^{n} \eta_i \sqrt{w_{ij}} D^i + v_{ij} \tag{9}$$

Where D^j and D^i are, respectively, province and commodity dummy variables which take values 1 for province j and commodity i respectively and 0 otherwise, v_{ij} are independently and identically distributed random disturbances with mean 0 and variance σ^2, w_{ij} is the expenditure share value of *i-th* item in the *j-th* province of the region r over the total number of items in the *l-th* BH:

$$w_{ij} = \frac{p_{ij}q_{ijr}}{\sum_{i=1}^{N_l} p_{ij}q_{ij}}$$

With $\sum w_{ij} = 1$ for each province and BH.

The within-region SPI for the province j in region r obtained by the weighed expenditure RPD model (9) is defined as

$$WR\_SPI_j^{WR\_WRPD} = e^{\hat{\pi}_j} \tag{10}$$

5.2.2. Second Step: Between-Region SPIs

Let us assume that we are attempting to make a spatial comparison of prices between R regions, $r = 1, \ldots, R$. On the basis of the properties of the weighted RPD method, discussed in Rao (2009), we decided to use this method for estimating between-region SPIs. The use of RPD method is also referred to as the stochastic approach to multilateral comparisons. Hajargasht and Rao (2010) and Rao and Hajargashat (2016) discussed properties of the weighted CPD model and show how most of the multilateral index numbers can be derived using weighted CPD model.

In order to use provincial prices adjusted for differences among provinces within the *r-th* region, item prices in each province j located in region r, denoted by p_{ijr}, are converted by using:

$$\hat{p}_{ir} = \frac{p_{ijr}}{WR\_SPI_j} \tag{11}$$

The adjusted prices (in log form) were used for estimating a weighted RPD (WRPD) model with regional dummies:

$$\sqrt{\hat{w}_{ir}} ln \hat{p}_{ir} = \sum_{r=1}^{R} \pi_r \sqrt{\hat{w}_{ir}} D^r + \sum_{i=1}^{N} \eta_i \sqrt{\hat{w}_{ir}} D^i + \upsilon_{ir} \tag{12}$$

where $\hat{w}_{ir}$ are weights defined using adjusted expenditure share for each item i in region r, υ_{ir} are independently and identically distributed random disturbances with mean 0 and variance σ^2. In order to estimate parameters of this model, i.e., π_r and η_i, we impose normalization $\sum_{r=1}^{R} \pi_r = 0$ thus treating all regions in a symmetric manner.

The WRPD method produces a transitive set of between-region SPI given by $BR\_SPI_r = e^{\hat{\pi}_r}$ and comparison between region r and s are given by:

$$P_{rs}^{WRPD} = \frac{exp(\hat{\pi}_s)}{exp(\hat{\pi}_r)} \quad for\, all\, s, r = 1, 2, \ldots R \tag{13}$$

5.3. Method for Aggregation Above BH Level

The next and final step for compiling regional price indexes is to aggregate the results from BH level comparisons to higher level aggregates. We decide to use the Fisher price index since it has a range of good axiomatic and economic theoretic properties. Suppose there are L basic headings ($l = 1, \ldots, L$) and let p_l^r and e_l^r represent price and expenditure for *l-th* basic heading in region r, respectively. The Fisher index given in Equation (2) becomes:

$$P_{rs}^{Fisher} = \sqrt{P_{rs}^{Laspeyres} \cdot P_{sr}^{Paasche}} \tag{14}$$

where

$$P_{rs}^{Laspeyres} = \frac{\sum_{l=1}^{L} p_l^s q_l^r}{\sum_{l=1}^{L} p_l^r q_l^r} = \sum s_l^r \left(\frac{p_l^s}{p_l^r}\right)$$

$$P_{rs}^{Paasche} = \frac{\sum_{l=1}^{L} p_l^s q_l^s}{\sum_{l=1}^{L} p_l^r q_l^s} = \left[\sum_l s_l^s \left(\frac{p_l^s}{p_l^r}\right)^{-1}\right]^{-1}$$

$$\text{with } s_l^r = \frac{e_l^r}{\sum_{l=1}^{L} e_l^r} = \frac{p_l^r q_l^r}{\sum_{l=1}^{L} p_l^r q_l^r}$$

As the Fisher binary index is not transitive, as usual the GEKS index is used to generate transitive multilateral price comparisons both within region, *WR_SPIs* and across different regions *BR_SPIs*. The GEKS-Fisher based formula is used in cross-country comparisons made within the ICP at the World Bank (2013, 2020) and the OECD-Eurostat comparisons. In order to obtain a set of BR_SPIs that refer to the group of regions (Italy) we standardized the GEKS-Fisher based SPIs (S-GEKS).

6. Sub-National SPI Results

6.1. Within-Region Results

The within-region SPIs show that price levels may vary within a region, therefore two Italian regions were compared as an example: Tuscany and Lombardy.

In Tuscany, as reported in Table 2 and Figure 2, the most expensive province for Food products was Siena while Prato was the cheapest compared to Tuscany as a whole.

Table 2. Within-region SPIs for Tuscany by product aggregates (Tuscany = 100).

	Food products WR_SPIs	Non-food products WR_SPIs
AR-Arezzo	98.94	98.44
FI-Florence	98.38	97.83
GR-Grosseto	101.98	103.07
LI-Livorno	102.23	104.01
LU-Lucca	99.48	99.24
MS-Massa Carrara	100.59	101.12
PI-Pisa	98.76	98.43
PO-Prato	98.29	97.37
PT-Pistoia	98.61	97.54
SI-Siena	102.87	103.24

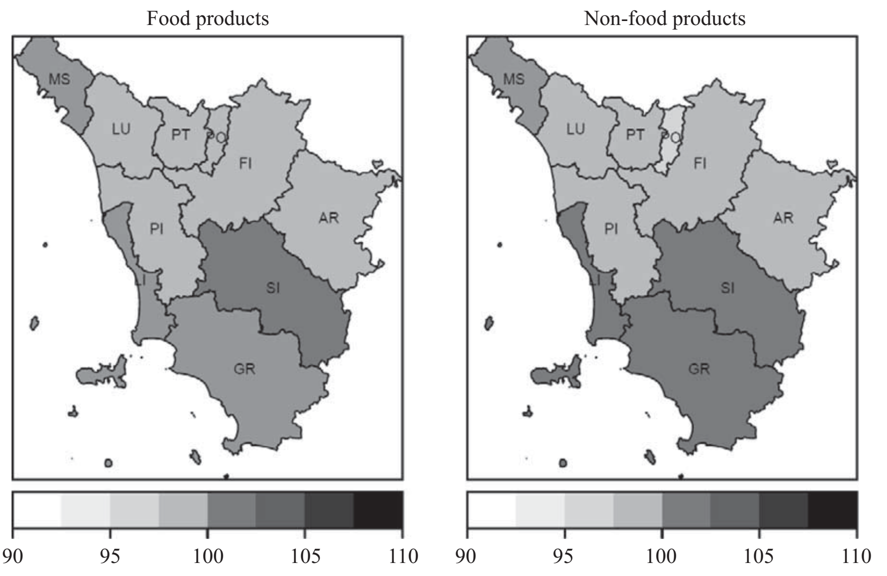


Fig. 2. Within-region SPIs based on scanner data for Tuscany (Tuscany = 100).

A higher heterogeneity in price levels was observed for Non-Food products Livorno being the most expensive province.

In Lombardy fewer price level differences among provinces were observed for both product aggregates as illustrated in Table 3 and Figure 2. More specifically, for Food Products, the coefficient of variation is equal to 0.46% for Lombardy and 2.79% for Tuscany. A higher price level heterogeneity is found for Non-Food Products with a coefficient of variation equal to 1.84% for Lombardy and 6.03% for Tuscany. Various factors can influence the observed difference in price level heterogeneity between these two regions. The different income level may play an important role in explaining this effect and provinces within the two regions are characterized by heterogeneous socio-economic conditions. However, it is not possible to fully explore the relationship between per capita

Table 3. Within-region SPIs for Lombardy by product aggregates (Lombardy = 100).

	Food products WR_SPIs	Non-food products WR_SPIs
BG-Bergamo	99.12	98.45
BS-Brescia	101.20	100.98
CO-Como	100.75	101.54
CR-Cremona	99.96	101.29
LC-Lecco	99.74	99.36
LO-Lodi	100.14	100.53
MB-Monza e della Brianza	99.49	100.13
MI-Milano	99.54	100.36
MN-Mantova	98.99	98.18
PV-Pavia	101.06	101.91
SO-Sondrio	100.16	97.37
VA-Varese	99.87	100.00

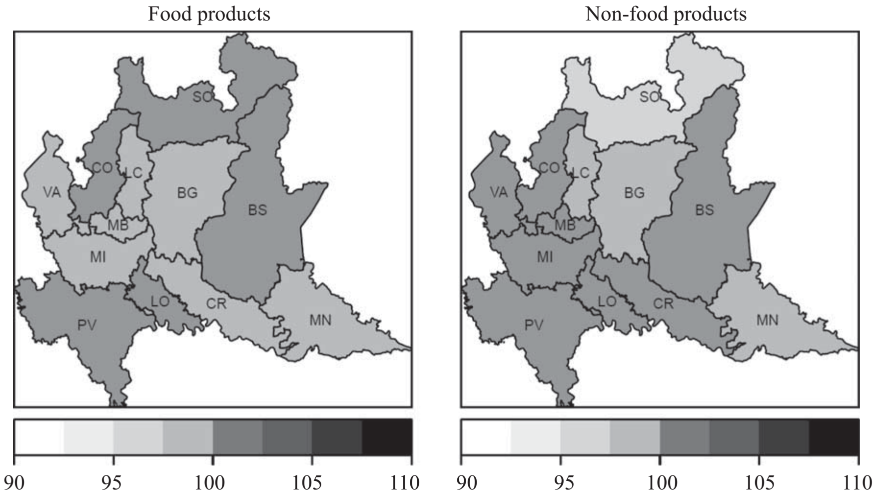


Fig. 3. Within-region SPIs based on scanner data for Lombardy (Lombardy = 100).

income heterogeneity and price level heterogeneity because the scanner data set we used only cover part of the household consumption expenditure and only refer to the large-scale retail trade distribution. The spatial component may also provide useful insights for understanding the heterogeneity in price level differences observed in these two Italian regions. Indeed, it is reasonable to assume that product prices are spatially autocorrelated and also exhibit spatial heterogeneity (the so-called spatial effects) especially when comparing consumer prices across provinces within a region. The spatial effect may be stronger in one region and in specific product aggregates (Montero et al. 2019).

Lombardy's low level of heterogeneity in consumer price differences is not confirmed when considering specific food products, that is, Pasta and Couscous as illustrated in Table 4.

Observing the results provided in Table 4, some considerations can be made regarding the figures obtained from different spatial price index formulae.

Overall, in the case of Lombardy, similar results are obtained using various methods. However, using weights in index formulas lead to different results compared to equal

Table 4. WR_SPIs using different methods: pasta products and couscous (Milan = 100).

	GEKS_Jevons	GEKS_Fisher	GK	RPD	WE_RPD
Bergamo	100.85	101.19	100.14	100.34	100.96
Brescia	103.90	104.98	101.71	103.70	104.60
Como	101.49	101.23	100.53	101.17	101.07
Cremona	103.33	101.99	101.76	103.77	102.08
Lecco	100.88	101.63	100.38	100.85	101.57
Lodi	100.09	100.24	99.77	99.61	99.52
Monza-Brianza	99.63	99.97	99.78	99.63	99.73
Milan	100.00	100.00	100.00	100.00	100.00
Mantova	102.45	103.06	100.87	102.10	102.74
Pavia	101.63	101.90	100.66	101.43	102.17
Sondrio	105.37	106.69	102.08	104.59	105.89
Varesa	100.67	101.06	100.38	100.82	100.90

weights methods. This effect is evident when comparing WR_SPIs obtained using GEKS-Jevons and GEKS-Fisher, for which quantity data for commonly priced items are used. The effect of introducing weights become more evident for some provinces with the highest differences observed for Cremona and Sondrio. Since the use of weights that reflect the economic importance of each item involved in the price comparison is a well-established recommendation in the index number literature (e.g., Diewert 2004; Rao 2001) we preferred to adopt methods which use this information available in scanner data.

As already mentioned, we applied the Geary-Khamis method to analyse its performance when aggregating prices at elementary level. As expected lower price levels were observed for household goods in relatively poorer provinces when the GK method was used (Ackland et al. 2013). Indeed, the GK method weights the *i-th* item in the commodity group of province *j* by the province's share of commodity group in the entire region so that large province in the region will get large weight in this average.

Multilateral comparisons based on commonly priced items do not capture all the information contained in the scanner database, unless prices and quantities are imputed using for example a regression based method. It is worth noting that we also computed the Jevons and Fisher indexes using gap-filled price matrices with the aim of using more information from the scanner data set. Price data for the other items are indirectly used through the GEKS extension of the binary Jevons and Fisher indexes. Since price imputations in the first stage slightly affected the WR_SPIs we decided to not report those results which are available upon request. Therefore, in order to take advantage of the itemized information provided by our scanner data set, the unweighted and the weighted RPD method are used.

Since in our empirical study product overlaps show a chain structure, demonstrated by graph analyses, the weighted RPD method exhibits some aspects of spatial chaining and therefore we selected this method for computing within-region SPIs for product aggregates (Hajargasht et al. 2017). More specifically, we developed a graph analysis for each BH and found connected graph for all product aggregates with the exception of Whole Milk and Low-Fat Milk which were therefore excluded from the comparison. It is worth noting that a graph is said to be connected if any pair of geographical areas can be linked through a sequence of areas that are all connected by edges. Indeed Rao (2004) provided a comparison of CPD and GEKS methods and demonstrated the clear advantages in using the CPD method. In previous studies we illustrated the advantage of using the weighted CPD method for dealing with various data-related issues (Biggeri et al. 2017; Laureti et al. 2017). As shown in Table 4, by using the RPD methods, the quality of price comparisons seems to be improved thanks to stronger interconnections and overlaps in the priced items across different provinces. The weighted RPD variant is preferable since it may allow us to weight commodities according to their economic importance, using expenditure share weights, thus obtaining more accurate estimates of price levels (Diewert 2005).

Within-region findings could be considered for obtaining spatial adjustment factors to be used when prices for other product categories are collected using traditional survey (on-field price collection) or ad-hoc surveys carried out in the regional capital.

6.2. Between-Region Results

As shown in Figure 4 and Table 5, the Southern regions appear to have lower price levels than the national average for both food and non-food products, with the exception of

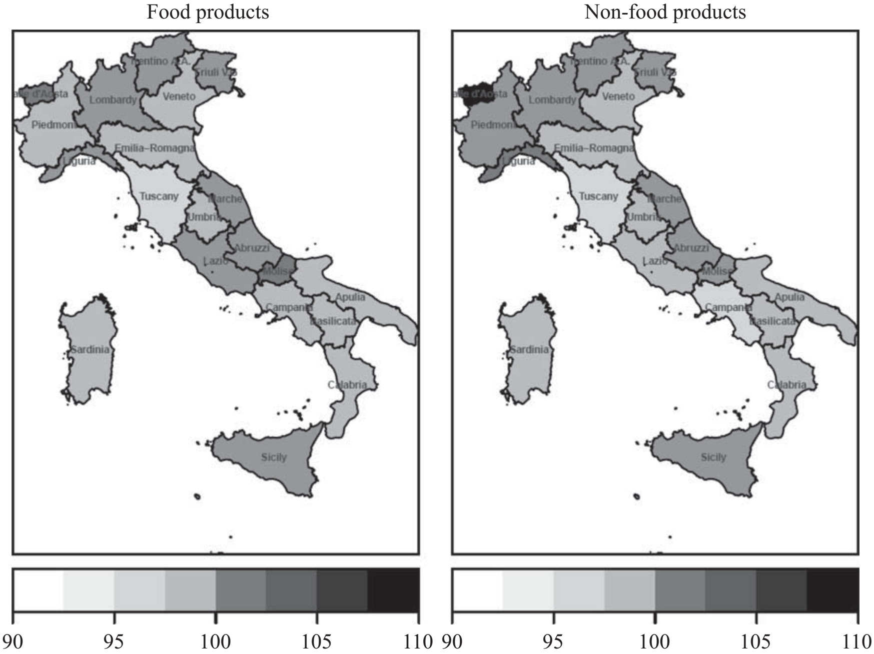


Fig. 4. Between-region SPIs based on scanner data for Italy (Italy = 100).

Table 5. Between-region estimation results for food and non-food products (Italy = 100) using scanner data.

	Food products	Non-food products
North-Center		
PIEDMONT	99.80	100.35
AOSTA VALLEY	104.95	107.86
LIGURIA	102.44	102.65
LOMBARDY	100.18	100.59
TRENTINO	101.56	102.41
VENETO	99.09	98.48
FRIULI	100.77	100.70
EMILIA-ROMAGNA	98.31	98.40
TUSCANY	96.24	95.17
UMBRIA	98.53	98.02
MARCHE	101.08	101.44
LAZIO	100.23	99.82
South and Islands		
ABRUZZO	101.90	101.33
MOLISE	102.90	101.24
CAMPANIA	98.65	97.20
APUGLIA	97.74	97.78
BASILICATA	97.53	99.54
CALABRIA	98.02	98.22
SICILY	101.93	101.57
SARDINIA	98.61	97.88

Abruzzo (101.90 and 101.33, respectively), Molise (102.90 and 101.24) and Sardinia (101.93 and 101.57).

However, it is worth noting that some Northern regions also show lower price levels than the national averages, such as Emilia-Romagna (98.31 and 98.40), Veneto (99.09 and 98.48) and Piedmont for Food products (99.80). On average, Tuscany proved to be the least expensive region for both product aggregates (96.24 and 95.17).

As regards the large-scale retail trade outlets, the results obtained partially change the expected relationship between Northern and Southern Italian price levels and suggest that investigating the influence of the various distribution channels when defining sub-national SPIs might be an interesting line for future research.

Caution is required when interpreting these results since: a) they may be influenced by the characteristics of large-scale retail trade distribution which is not uniformly distributed across the Italian territory in terms of types of retail chains and market share (see Table 1); b) these results are based on data selected for CPI compilation and on a sample of hypermarkets and supermarkets (e.g., hard discounts offering products at more affordable prices to consumers by reducing marketing and merchandising costs, are excluded).

By considering various group of products, it is interesting to note that the usual divide between North and South is not confirmed in some BHs. Regional spatial price indexes for two specific groups of products, that is "Pasta products" (BH1), which belongs to the aggregate Food products, and "Non-electrical appliances" (BH2, e.g., razors, scissors, hairbrushes, toothbrushes, etc.) included in the Non-Food aggregate are illustrated in Table 6.

Our findings confirm large differences in price levels among Italian regions even if BH2 shows a higher territorial heterogeneity than BH1 (range is equal to 18.38 and 13.02 respectively).

In the case of BH1, five out of eight regions located in the South and Islands and three out of eleven Northern-Central regions show lower prices than the national average while higher price indexes are observed in six out of eight Southern regions and in five out of 12 regions in Northern and Central Italy for BH2.

As already mentioned, this different territorial pattern of consumption SPIs is not confirmed when aggregated regional SPIs are computed for Food and Non-food products (Italy $=$ 100).

6.3. Using Scanner Data in the Italian Regional SPI Estimation Project

The results obtained show that the expected price variability across regions is explained by large-scale retail distribution which has to take into account local purchasing power in order to fine-tune its pricing policies. These findings may be strengthened by analysing other scanner data referring to hard discount, large-scale distribution in small outlets, shop selling detergents and toiletries, that Istat has started to use for estimating inflation since January 2020. The weight of hard discounts varies across regions and this will generate variation in spatial price comparisons which may be detected using the methods described in this article.

Nevertheless, as aforementioned, scanner data, which was used for making spatial comparison in this article, only focus on one area of the retail trade market: grocery products sold in large-scale retail outlets, that are hypermarkets and supermarkets, covering approximately 10–11% of household final consumption expenditures.

Table 6. WRPD estimation results for "pasta products" and "non-electrical appliances" Italy = 100.

	Pasta products (BH1)				Non-electrical appliances (BH2)			
Region	Coeff.	SE..error	p-value	BR_SPI	Coeff.	SE.error	p-value	BR_SPI
North-Center								
PPIEDMONT	0.0028	0.0027	0.3071	100.28	−0.0550	0.0056	0.0000	94.65
AOSTA VALLEY	0.0367	0.0028	0.0000	103.74	0.0528	0.0059	0.0000	105.43
LIGURIA	0.0323	0.0034	0.0000	103.28	−0.0061	0.0056	0.2829	99.40
LOMBARDY	0.0104	0.0027	0.0001	101.05	−0.0402	0.0056	0.0000	96.06
TRENTINO	0.0557	0.0029	0.0000	105.73	0.0268	0.0057	0.0000	102.71
VENETO	0.0188	0.0027	0.0000	101.89	−0.0133	0.0056	0.0183	98.68
FRIULI	0.0276	0.0026	0.0000	102.80	−0.0079	0.0057	0.1611	99.21
EMILIA-ROMAGNA	0.0068	0.0031	0.0270	100.68	−0.0386	0.0056	0.0000	96.22
TUSCANY	−0.0209	0.0028	0.0000	97.93	−0.1205	0.0057	0.0000	88.65
UMBRIA	−0.0254	0.0029	0.0000	97.50	0.0027	0.0056	0.6357	100.27
MARCHE	0.0398	0.0031	0.0000	104.06	0.0258	0.0056	0.0000	102.61
LAZIO	−0.0159	0.0026	0.0000	98.42	0.0075	0.0056	0.1823	100.75
South and Islands								
ABRUZZO	0.0401	0.0030	0.0000	104.09	0.0036	0.0057	0.5254	100.36
MOLISE	0.0311	0.0031	0.0000	103.16	0.0354	0.0058	0.0000	103.60
CAMPANIA	−0.0256	0.0029	0.0000	97.47	0.0348	0.0057	0.0000	103.54
APULIA	−0.0547	0.0029	0.0000	94.68	−0.0071	0.0057	0.2132	99.29
BASILICATA	−0.0570	0.0029	0.0000	94.46	0.0236	0.0057	0.0000	102.39
CALABRIA	−0.0445	0.0029	0.0000	95.65	0.0270	0.0057	0.0000	102.74
SICILY	−0.0758	0.0034	0.0000	92.70	0.0679	0.0057	0.0000	107.03
SARDINIA	0.0176	0.0036	0.0000	101.78	−0.0192	0.0057	0.0007	98.10

Therefore, the encouraging results obtained using grocery scanner data may help Istat to achieve the aim of producing sub-national SPIs for Italy on a regular basis by adopting a multi-source approach in order to cover all the retail trade channels and product baskets. To this aim it is essential to replicate what has already been done in the field of temporal comparison. More specifically, price collection should be based on:

A. An appropriate use of CPI data by carrying out an in-depth analysis of the basket as well as the microdata. The analysis of the basket is aimed at selecting all the products which, by definition, are comparable and do not need further specifications in additions to those ones already present in the Italian CPI basket. This is the case of fresh fish, of all the different varieties of fresh fruit and vegetables that are well specified in the CPI basket, of some services related to the maintenance of dwellings, and so on., for a total weight on the consumer price basket of approximately 20%. The analysis of microdata is therefore aimed at selecting all the elementary products that are present in remarkable amounts in a sufficient number of regions in order to be both representative and comparable,
B. Scanner data for grocery products and, in the future, for other groups of products whose prices could be detected by acquiring electronic transaction data that could cover over 10–11% of information derived from this type of source,
C. Administrative data that are already used for CPI aims (in particular those concerning automotive and heating fuels) and other administrative data that will be used in the near future (i.e., the data base of rents already provided by Italian Tax Office),
D. Excluding those products for which it is unnecessary to make a comparison among different geographical areas (because the prices are the same or differences are not relevant as in the case of tobacco products or telecommunications, for a total weight of about 8%), and
E. A specific price data collection survey for the remaining CPI products or for those for which the microdata analysis has not provided or will not provide satisfactory results.

7. Concluding Remarks

The main objective of this article is to provide an estimate of spatial price differences in consumer prices for grocery products across Italian regions using scanner data obtained from the large-scale retail trade distribution.

With respect to cross-country comparisons of prices and real incomes within the ICP, work on compilation of sub-national SPIs is at a preliminary stage even if regional price comparisons are gathering some momentum in various countries.

The increased availability of high-frequency electronic-point-of-sale "scanner data" has the potential to significantly change how to compile spatial price indexes thanks to the very detailed specifications concerning the quality characteristics of products at barcode level and to the availability of information regarding turnover and quantities associated with each GTIN.

It is clear that for the time being scanner data only cover part of the universe of household monetary consumption expenditure and only refer to the large-scale retail trade distribution.

Measuring spatial variability of prices for all of the products purchased by households for consumption and covering all the distributional channels require price data obtained from multiple sources and outlets, which are representative of local consumption patterns and comparable on the basis of a set of characteristics.

However, the results obtained for grocery products sold in large-scale retail trade outlets using scanner data generate groundbreaking ideas for compiling sub-national SPIs. A strategy (as it is illustrated in Subsection 6.3) is now available for achieving this aim thanks to the results of these analyses.

This article has also shed light on which methodological approach is more appropriate by demonstrating the feasibility of implementing various aggregation methods. Indeed, there are several price index methods which can be used for scanner data given that expenditure data are available at elementary level. Taking into account the results of our empirical study, which confirm theoretical expectations, we suggested to use the weighted RPD model since product overlaps demonstrated to exhibit a chain structure, therefore regions with close geographic and economic links tend to have more products in common than regions that are more distant (Rao 2004).

Scanner data enable NSIs to compute sub-national SPIs at local level to be used for adjusting regional economic indicators. This new source of data may also improve the accuracy of spatial price indexes, thus contributing to the advancement of spatial price index literature. The issue of evaluating the uncertainty associated to point estimates of sub-national SPIs using scanner data will be an important future line of research. With the aim to obtaining variance estimation of the calculated SPIs, both “direct” methods, which rely on analytic variance and “resampling” methods, such as the Jackknife Repeated Replication and Boostrap, which consist of resampling a high number of “replications” from the original sample, may be used in order to compare the results obtained. Due to its widespread use, we selected the Jackknife replication method that can be performed by setting the outlet as Primary Sample Unit (PSU) and strata to provinces. In the standard ‘delete one-PSU at a time Jackknife’ version, each JRR replication is formed by eliminating one sample PSU from a particular stratum at a time and increasing the weight of the remaining sample PSU’s in that stratum appropriately so as to obtain an alternative but equally valid estimate to that obtained from the full sample. Interesting results have been already obtained from empirical applications on various basic headings, which have encouraged us to continue this line of research by also considering using the deleted Jackknife.

The possibility of producing regional PPPs on a regular basis could improve our knowledge of the real economic territorial differentiation in Italy, thus improving the measurement of relative poverty by taking into account the price dimension which is neglected given that, at present, a uniform threshold at national level is only considered and by measuring the inequalities more accurately by using income and expenditures adjusted for local purchasing power.

In order to obtain sub-national SPIs covering the entire universe of household expenditure both in terms of products and retail trade channels, it is essential to compile within-region and between-region SPIs by using these new data sources and then grouping them (by identifying the appropriate methodological solutions) with other SPIs coming from other sources of data which is the main challenge to be addressed in the near future.

8. References

Ackland, R., S. Dowrick, S., and B. Freyens. 2013. “Measuring global poverty: Why PPP methods matter”. *Review of Economics and Statistics*, 95(3): 813–824. Available at: https://www.jstor.org/stable/43554796 (accessed February 2020).

Alberola, E., and J.M. Marqués. 2001. “On the evolution of relative prices and its nature at the regional level: The case of Spain”. *Journal of Regional Science*, 41(3): 451–474. DOI: https://doi.org/10.1111/0022-4146.00226.

Aten, B.H. 1999. “Cities in Brazil: An inter-area price comparison”. In International and inter-area comparisons of income, output, and prices, edited by A. Heston and R.E. Lipsey, 211–229. Chicago: University of Chicago Press.

Aten, B.H. 2006. “Inter-area rice levels: An experimental methodology”. *Monthly Labor Review*, 129(9): 47–61. Available at: https://www.jstor.org/stable/23805663.

Aten, B. 2008. *Estimates of state and metropolitan price parities for consumption goods and services in the United States, 2005*. Bureau of Economic Analysis. Available at: https://bea.gov/papers (accessed February 2020).

Aten, B.H. 2017. “Regional price parities and real regional income for the United States”. *Social Indicators Research*, 131(1): 123–143. DOI: https://doi.org/10.1007/s11205-015-1216-y.

Aten, B.H., E.B. Figueroa, and T.M. Martin. 2012. “Regional price parities for states and metropolitan areas, 2006–2010.” *Survey of Current Business*, 92(8): 229–242. Available at: https://apps.bea.gov/scb/index.htm (accessed February 2020).

Aten, B.H., Figueroa, E.B., and B.M. Vengelen. 2015. *Real personal income and regional price parities for states and metropolitan areas, 2008–2012*. Available at: https://apps.bea.gov/scb/pdf/2014/06%20June/0614_real_personal_income_and%20_regional_price_parities_for_states_and%20metrpolitan_areas.pdf.

Aten, B., and T. Menezes. 2002. “Poverty price levels: an application to Brazilian metropolitan areas”. In World Bank ICP Conference, March, Washington, D.C.: 11–15. Available at: https://www.rug.nl/ggdc/productivity/pwt/related-research-papers/batatwb.pdf.

Balk, B.M. 2008. *Price and Quantity Index Numbers*. Cambridge Books.

Balk, B.M. 2009. “Aggregation Methods in International Comparisons: An Evaluation.” In *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, edited by D.S. Prasada Rao: 59–85. Cheltenham, U.K.: Edward Elgar.

Bernardini A., M.C. Casciano, C. de Vitiis, A. Guandalini, F. Inglese, G. Seri, M.D. Terribili, and F. Tiero. 2019. “Re-design project of the Istat consumer price survey: use of probability samples of scanner data for the calculus of price indices”. Rivista di Statistica Ufficiale, 2–3: 67–14. Available at: https://www.istat.it/it/files//2021/03/RSU_2-3_2019_Article-3.pdf (accessed February 2020).

Biggeri, L., T. Laureti, and F. Polidoro. 2017 “Computing sub-national PPPs with CPI data: an empirical analysis on Italian data using country product dummy models”. *Social Indicators Research*, 131(1): 93–121. DOI: https://doi.org/10.1007/s11205-015-1217-x.

Blien, U., H. Gartner, H. Stüber, and K. Wolf. 2009. "Regional price levels and the agglomeration wage differential in western Germany". *The Annals of Regional Science*, 43(1): 71–88. DOI: https://doi.org/10.1007/s00168-007-0205-8.

Čadil, J., P. Mazouch, M. Musil, and J. Kramulová. 2014. "True regional purchasing power: Evidence from the Czech Republic". *Post-Communist Economies*, 26(2): 241–256. DOI: https://doi.org/10.1080/14631377.2014.904109.

Cannari, L., and G. Iuzzolino. 2009. *Le differenze nel livello dei prezzi al consumo tra Nord e Sud. Questioni di Economia e Finanza*, Bank of Italy, Rome Available at: https://www.bancaditalia.it/pubblicazioni/qef/2009-0049/index.html (accessed February 2020).

Cavallo, A., W.E. Diewert, R.C. Feenstra, R. Inklaar, and M.P. Timmer. 2018. "Using Online Prices for Measuring Real Consumption across Countries", *AEA Papers and Proceedings*, vol 108: 483–487. DOI: https://doi.org/10.1257/pandp. 20181037.

Chessa, A.G., and R. Griffioen. 2019. "Comparing Price Indexes of Clothing and Footwear for Scanner Data and Web Scraped Data". *Economie et Statistique*, 509(1): 49–68. DOI: https://doi.org/10.24187/ecostat.2019.509.1984.

Coondoo, D., A. Majumder, and R. Ray. 2004. "A method of calculating regional consumer price differentials with illustrative evidence from India". *Review of Income and Wealth*, 50(1): 51–68. DOI: https://doi.org/10.1111/j.0034-6586.2004.00111.x.

Costa, A., J. Garcia, J.L. Raymond, and D. Sanchez-Serra. 2019. *Subnational purchasing power of parity in OECD countries: Estimates based on the Balassa-Samuelson hypothesis.* OECD Regional Development Working Papers 2019/12. DOI: https://doi.org/10.1787/3d8f5f51-en.

Deaton, A., and O. Dupriez. 2011 *Spatial price differences within large countries*, Manuscript, Princeton University. Available at: https://www.princeton.edu/~deaton/downloads/spatial_price_differences_in_large_countries%20_10%20jul_2011_complete.pdf (accessed January 2020).

Deaton, A., and A. Tarozzi. 2005 "Prices and poverty in India", In *The great Indian poverty debate*, edited by A. Deaton and V. Kozel, New Delhi, Macmillan: 381–411, DOI: https://doi.org/10.1093/wbro/lki009.

De Carli, R. 2010. "Intra-national Price Level Differentials: The Italian experience". In *Price indexes in time and space: Methods and Practice*, edited by L. Biggeri and G. Ferrari: 115–130. Berlin: Springer.

Diewert, W.E. 1976. "Exact and superlative index numbers". *Journal of econometrics*, 4(2): 115–145. DOI: https://doi.org/10.1016/0304-4076(76)90009-9.

Diewert, W.E. 2004. *On the Stochastic Approach to Linking the Regions in the ICP*. Department of Economics, University of British Columbia. Available at: http://papers.economics.ubc.ca/legacypapers/dp0416.pdf.

Diewert, E. 2005. "Weighted country product dummy variable regressions and index number formulae". *Review of Income and Wealth*, 51(4): 561–570. DOI: https://doi.org/10.1111/j.1475-4991.2005.00168.x.

Eurostat. 2017. *Practical Guide for Processing Supermarket Scanner Data*, European Commission, Directorate C: National Accounts; prices and key indicators. Unit C-4: Price statistics; Purchasing power parities; Housing statistics. Luxemburg, Available at:

https://circabc.europa.eu/sd/a/8e1333df-ca16-40fc-bc6a-1ce1be37247c/Practical-Guide-Supermarket-Scanner-Data-September-2017.pdf (accessed March 2020).

Eurostat-OECD. 2012. *Methodological manual on purchasing power parities.* Luxembourg: Publications Office of the European Union, Available at: https://ec.europa. eu/eurostat/web/products-manuals-and-guidelines/-/KS-RA-12-023 (accessed January 2020).

Feenstra, R.C., M. Xu, and A. Antoniades. 2017. *What is the Price of Tea in China? Towards the Relative Cost of Living in Chinese and US Cities*, National Bureau of Economic Research. No. w23161.

Fenwick, D., and J. O'Donoghue. 2003. "Developing estimates of relative regional consumer price levels". *Economic Trends*, 599: 72–83. Available at: https://webarchive. nationalarchives.gov.uk/ukgwa/20160120195958/http:/www.ons.gov.uk/ons/rel/cpi/regional-consumer-price-levels/developing-estimates-of-relative-regional-consumer-price-levels/developing-estimates-of-relative-regional-consumer-price-levels-.pdf (accessed January 2020).

Geary, R.C. 1958. "A Note on the Comparison of Exchange Rates and Purchasing Power between Countries", *Journal of the Royal Statistical Society, Series A (General)*, 121(1): 97–99. DOI: https://doi.org/10.2307/2342991.

Hajargasht, G., and. D.S.P. Rao. 2010. "Stochastic Approach to Index Numbers for Multilateral Price Comparisons and their Standard Errors", *Review of Income and Wealth*, Series 56: S32–S58. DOI: https://doi.org/10.1111/j.1475-4991.2010.00387.x.

Hajargasht, G., R. Hill, S. Shankar, and D.S.P. Rao. 2017. "Spatial Chaining in International Comparisons of Prices and Real Incomes", *Mimeographed*. Available at: https://thedocs.worldbank.org/en/doc/637701510177729968-0050022017/original/ICPTFPCC01DocSpatialchainingHilletal2017.pdf.

Heravi, S., A. Heston, and M. Silver. 2003. "Using scanner data to estimate country price parities: A hedonic regression approach". *Review of Income and Wealth*, 49(1): 1–21. DOI: https://doi.org/10.1111/1475-4991.00071 (accessed January 2020).

Khamis, S.H., 1972 "A New System of Index Numbers for National and International Purposes". *Journal of the Royal Statistical Society*, Series A, 135(1): 96–121. DOI: https://doi.org/10.2307/2345041.

Kocourek, A., J. Šimanová, and J. Šmída. 2016.37 "Estimation of regional price levels in the districts of the Czech Republic". *Statistika*, 96(4): 56–70. Available at: https://www.czso.cz/documents/10180/32912820/32019716q4.pdf (accessed December 2019).

Kokoski, M. 1991. "New Research on Interarea Consumer Price Differences", *Monthly Labor Review*, 114(31). Available at: https://www.jstor.org/stable/41843608 (accessed December 2019).

Kokoski, M., P. Cardiff, and B. Moulton. 1994. *Interarea price indices for consumer goods and services: a hedonic approach using CPI data.* U.S. Department of Labor, Bureau of Labor Statistics, Working Paper: 256.

Kokoski, M., B. Moulton, and K. Zieschang. 1999 "Interarea Price Comparisons for Heterogenous Goods and Several Levels of Commodity Aggregation", In *International and Interarea Comparisons of Income, Output and Prices*, edited by A. Heston and R. Lipsey: 123–166. University of Chicago Press.

Istat. 2010. *La differenza nel livello dei prezzi al consumo tra i capoluoghi delle regioni italiane*, Anno 2009, Comunicato Stampa, Roma. Available at: https://www.istat.it/it/archivio/6279 (accessed September 2019).

Istat. 2018a. *Household consumption expenditure-Year 2017*, Press release, Rome, Available at: https://www.istat.it/en/archive/consumption+expenditure (accessed January 2020).

Istat. 2018b. *Poverty in Italy-Year 2017*, Press release, Rome, Available at: https://www.istat.it/en/archivio/217653 (accessed January 2020).

Istat. 2020. *Poverty in Italy-Year 2019*, Press release, Rome, Available at: https://www.istat.it/en/archivio/246098 (accessed January 2020).

Johnston, R., M. McKinney, and T. Stark. 1996. "Regional price level variations and real household incomes in the United Kingdom, 1979/80–1993". *Regional Studies*, 306: 567–578. DOI: https://doi.org/10.1080/00343409612331349868.

Laureti, T., C, Ferrante, and B. Dramis. 2017. "Using scanner and CPI data to estimate Italian sub-national PPPs", In Proceeding of Scientific Meeting of the Italian Statistical Society: Statistics and Data Science: new challenges, new generations, 28–30 June 2017, Firenze, Italy: 581–588. DOI: https://doi.org/10.36253/978-88-6453-521-0.

Laureti, T., and F. Polidoro. 2017. "Testing the use of scanner data for computing sub-national Purchasing Power Parities in Italy", In Proceeding of 61st ISI World Statistics Congress, 16–21 July 2017, Marrakech, Morocco. Available at: https://www.isi-web.org/publications/proceedings (accessed January 2020).

Laureti, T., P.S.D. Rao. 2018. "Measuring Spatial Price Level Differences within a Country: Current Status and Future Developments". *Estudios de economía aplicada*, 361: 119–148. DOI: https://doi.org/10.25115/eea.v36i1.2520.

Léonard, I., P. Sillard, G. Varlet, and J.P. Zoyem. 2019. "Spatial differences in price levels between French regions and cities with scanner data". *Economie et Statistique*, 5091: 69–82. DOI: https://doi.org/10.24187/ecostat.2019.509.1983.

Majumder, A., R. Ray, and K. Sinha. 2012. "Calculating rural-urban food price differentials from unit values in household expenditure surveys: a comparison with existing methods and a new procedure." *American Journal of Agricultural Economics*, 945: 1218–1235. DOI: https://doi.org/10.1093/ajae/aas064.

Majumder, A., R. Ray, and K. Sinha. 2015a. "Spatial comparisons of prices and expenditure in a heterogeneous country: Methodology with application to India". *Macroeconomic Dynamics*, 195: 931. DOI: https://doi.org/10.1017/S1365100513000576.

Majumder, A., R. Ray, and K. Sinha. 2015b. "Estimating Purchasing Power Parities from Household Expenditure Data Using Complete Demand Systems with Application to Living Standards Comparison: India and Vietnam". *Review of income and wealth*, 612: 302–328. DOI: https://doi.org/10.1111/roiw.12073.

Mehrhoff, J. 2019. "Introduction–The Value Chain of Scanner and Web Scraped Data". *Economie et Statistique*, 5091: 5–11, DOI: https://doi.org/10.24187/ecostat.2019.509.1980.

Menon, M., F. Perali, R. Ray, and N. Tommasi. 2019. *The Tale of the Two Italies: Regional Price Parities Accounting for Differences in the Quality of Services*, Working Paper Series Department of Economics University of Verona, No. 20/2019: 2036–4679

Available at: https://dse.univr.it/home/workingpapers/wp2019n20.pdf (accessed November 2019).

Mishra, A., and R. Ray. 2014. "Spatial variation in prices and expenditure inequalities in Australia." *Economic Record*, 90289: 137–159. DOI: https://doi.org/10.1111/1475-4932.12109.

Montero J.M., T. Laureti, R. Mínguez, and G. Fernández-Avilés. 2019 "A Stochastic Model with Penalized Coefficients for Spatial Price Comparisons: An Application to Regional Price Indexes in Italy", *Review of Income and Wealth*, 66(3): 512–533 DOI: https://doi.org/10.1111/roiw.12422.

Musil, P., and J. Kramulová. 2019. Development of regional price levels in the Czech Republic." In 22nd International Scientific Conference on Applications of Mathematics and Statistics in Economics AMSE 2019, 28 August–1 September 2019, Nižná, Slovakia: 30–38. Atlantis Press. DOI: https://doi.org/10.2991/amse-19.2019.4.

ONS. 2011. *UK relative regional consumer price levels for goods and services for 2010*, Office for National Statistics, UK: 1–15. Available at: https://webarchive.nationalarchives.gov.uk/20160108054525/http://www.ons.gov.uk/ons/rel/cpi/regional-consumer-price-levels/2010/index.html.

Rao P.D.S. 2001. "Weighted EKS and generalised CPD methods for aggregation at basic heading level and above basic heading level". In Joint World Bank-OECD seminar on Purchasing Power Parities, 30 January 2 February 2001, Washington D.C. Available at: https://www. oecd.org/sdd/prices-ppp/2424825.pdf.

Rao P.D.S. 2004. *The Country-Product-Dummy Method: A Stochastic Approach to the Computation of Purchasing Power Parities* in the ICP", Working Paper Series No. 03/2004, Centre for Efficiency and Productivity Analysis, School of Economics University of Queensland, Australia. Available at: https://economics.uq.edu.au/files/5343/WP032004.pdf (accessed November 2019).

Rao P.D.S. 2009 "Generalised Elteto-koves-Szulc EKS and Country-Product-Dummy CPD Methods for International Comparisons" In *Purchasing Power Parities: Recent Advances In Methods And Applications*, edited by Prasada D.S. Rao: 86–120. Edward Elgar Publishing Company.

Rao P.D.S., and G. Hajargasht. 2016. "Stochastic approach to computation of purchasing power parities in the International Comparison Program ICP". *Journal of Econometrics*, 1912: 414–425. DOI: https://doi.org/10.1016/j.jeconom.2015.12.012.

Rao P.D.S., and E.A. Selvanathan. 1992. "Computation of standard errors for Geary–Khamis parities and international prices: a stochastic approach". *Journal of Business and Economic Statistics*, 101: 109–115. DOI: https://doi.org/10.2307/1391809.

Rokicki, B., and G.J. Hewings. 2019. "Regional price deflators in Poland: Evidence from NUTS-2 and NUTS-3 regions". *Spatial Economic Analysis*, 141: 88–105. DOI: https://doi.org/10.1080/17421772.2018.1503705.

Roos, M. 2006. "Regional price levels in Germany". *Applied Economics*, 3813: 1553–1566. DOI: https://doi.org/10.1080/00036840500407207.

Sherwood, M.K. 1975. "Family budgets and geographic differences in price levels". *Monthly Labour Review*, 98: 8–15. Available at: https://www.jstor.org/stable/41839590 (accessed January 2020).

Silver, M. 2003. "The Use of Weights in Hedonic Regressions: The Measurement of Quality-Adjusted Price Change," In 7th Meeting of the U.N. International Working Group on Price Indexes Ottawa Group, INSEE, May 27–29, Paris: 175–212. Available at: https://www.ottawagroup.org/Ottawa/ottawagroup.nsf/51c9a3d36edfd0dfca256ac-b00118404/aee68df6fad39f4fca257272001959e9?OpenDocument (accessed January 2020).

Tam, S.M., and F. Clarke. 2015. "Big data, official statistics and some initiatives by the Australian Bureau of Statistics". *International Statistical Review*, 833: 436–448. DOI: https://doi.org/10.1111/insr.12105.

Waschka, A., W. Milne, J. Khoo, T. Quirey, and S, Zhao, S. 2003. "Comparing living costs in Australian capital cities. A progress report on developing experimental spatial price indexes for Australia". 32nd Conference of Economists, 29th September – 1st October, Australian National University, Australian Bureau of Statistics. Available at: file:///C:/Users/laure/Downloads/Comparing_Living_Costs_in_Australian_Capital_Citie.pdf (accessed January 2020).

Weinand, S. 2020. Measuring Spatial Price Differentials: A Comparison of Stochastic Index Number Methods, *Deutsche Bundesbank Discussion Paper No. 12/2020*, Available at: https://ssrn.com/abstract = 3581148.

Weinand, S., and L. von Auer. 2020. "Anatomy of regional price differentials: evidence from micro-price data". *Spatial Economic Analysis*: 1–28. DOI: https://doi.org/10.1080/17421772.2020.1729998.

Wingfield, D., D. Fenwick, and K. Smith. 2005. "Relative regional consumer price levels in 2004". *Economic Trends* ONS, UK, 615: 36–46. Available at: https://webarchive.nationalarchives.gov.uk/ukgwa/20160105160709/http://www.ons.gov.uk/ons/rel/elmr/economic-trends–discontinued-/no–615–february-2005/relative-regional-consumer-price-levels-in-2004.pdf (accessed January 2020).

World Bank. 2013. *Measuring the Real Size of the World Economy: The Framework, Methodology, and Results of the International Comparison Program (ICP)*. The World Bank. Available at: https://openknowledge.worldbank.org/handle/10986/13329.

World Bank. 2020. *Purchasing Power Parities and the Real Size of the World Economies Results from the 2017 International Comparison Program*, Washington D.C. Available at: https://openknowledge.worldbank.org/handle/10986/33623.

Received August 2020
Revised January 2021
Accepted August 2021

Journal of Official Statistics, Vol. 38, No. 1, 2022, pp. 57–82, http://dx.doi.org/10.2478/JOS-2022-0004

Sub-National Spatial Price Indexes for Housing: Methodological Issues and Computation for Italy

Ilaria Benedetti[1], *Luigi Biggeri*[2], *and Tiziana Laureti*[1]

It is essential to measure within-country differences in housing costs in order to evaluate costs of living, assessing and comparing poverty levels, quantifying salaries and disposable income of families and finally for designing housing policies at local level. To the authors knowledge, no studies have yet been carried out on the computation of Space Price Indexes for Housing Rents (SPIHRs). In this article we computed preliminary estimates of sub-national SPIHRs by using hedonic regression model, which is an extension of the Country Product Dummy method, for all the Italian regions. The hedonic regression is generally used to obtain multilateral spatial indexes, thus allowing us to obtain multilateral SPIHRs for the Italian regions. The estimates have been done using 2017 data from the Real Estate Market Observatory which is a part of the Italian Agency of Revenue and Tax. This data source is the most comprehensive source of information on Italian houses price rents with a wide geographical coverage, including data for each Italian municipality. The obtained results show significant differences across the Italian regions, thus highlighting the importance of calculating SPIHR in Italy on a regular basis and the need to continue researches in this field.

Key words: Spatial rental price indexes; hedonic regression; country product dummy method; symbolic data analysis.

1. Introduction

It is essential to measure within-country differences in housing costs in order to evaluate costs of living, assessing and comparing poverty levels, quantifying salaries and disposable income of families and finally for designing housing policies at a local level (Bishop et al. 2017).

It is well known that when making comparisons, the sub-national (regional) values of the economic indicators should be adjusted for sub-national price differences in order to avoid misleading regional analyses and the consequent policy implications and outcomes, which is done by computing sub-national Spatial Price Indexes (SPIs) as illustrated by Biggeri et al. (2017) and Laureti and Rao (2018).

In this context, the computation of sub-national SPIs for housing costs is particularly important since housing costs can be a substantial financial burden to households, especially for low-income families.

These sub-national SPIs have proven to be useful for assessing and comparing poverty levels and for implementing housing policies for the poor at local level, however they can also be used as proxies of general sub-national household consumption SPIs or in

[1] University of Tuscia, Via del Paradiso 47, 01100, Viterbo (VT), Italy. Emails: i.benedetti@unitus.it and laureti@unitus.it
[2] The University of Florence, Viale Morgagni, 59, 50134 Firenze, 50134, Italy. Email: Luigi.biggeri@unifi.it

© Statistics Sweden

combination with them, for adjusting poverty thresholds (Bishaw 2009; Bishop et al. 2017; Biggeri and Pratesi 2017).

To the authors' knowledge, official estimates of sub-national Spatial Price Indexes for Housing Rent (SPIHRs), useful for the estimation of the poverty indicators in real terms, are mainly provided in the United States (US).

During the 1970's and 1980's several researchers worked on the computation of SPIs for housing costs which, however, were based on limited data and could not be used for countrywide poverty measures.

In the early 1990's, the National Academy of Science (NAS) Panel on Poverty and Family Assistance proposed that, as "a first and partial step", poverty thresholds must be adjusted for differences in the costs of housing across geographical areas, taking into account that housing is the most expensive cost of the household budget especially for poor households, and further research is required in order to develop more refined methods and data (Citro and Michael 1995).

The U.S. Bureau of Census computed a housing cost index using data from the decennial census and the method used by the U.S. Department of Housing and Urban Development for computing Fair Market Rents using data from a specific type of houses. After this first step, the NAS panel suggested improving the method and using better data in order to gain a better understanding of poverty levels which are known as Supplementary Poverty Measures (Jolliffe 2006; Renwick 2009; Bishop et al. 2017; Dalaker 2017).

The turning point in the computation of spatial housing cost differences occurred in 2007 when the American Community Survey (ACS) was conducted by U.S. Census Bureau. The ACS provides detailed annual data on housing costs and a set of interarea price indexes compiled by the researchers of the U.S. Bureau of Economic Analysis (BEA) and by the U.S. Bureau of Labor Statistics (BLS).

In the beginning, interarea price indexes were used to develop an index based on 2007 gross rental costs considering all rental units (Bishaw 2009). Finally, after numerous experimental estimates had been conducted in collaboration with other US federal agencies (Martin et al. 2013), the BEA computed Regional Price Parities (RPPs) which combine the CPI price data with rental data. Both spatial indexes regarding the prices of goods and services and housing rents, were elaborated using hedonic regression models and following the methods suggested by the International Comparison Program (ICP) in order to obtain multilateral indexes.

Since 2014, the RPPs and the price-adjusted estimates of regional personal income have become official statistics (Aten and Figueroa 2015).

Italian researchers have recently conducted a series of experiments for computing sub-national SPIs (Laureti and Rao 2018; Laureti and Polidoro 2018; Biggeri and Laureti 2018; Ferrante et al. 2019).

However, no complete studies have yet been carried out on the computation of SPIHRs, even if Italian households pay housing costs equal to approximately 35% of their total expenditures (Istat 2017) and significant differences in average housing costs were observed across the various Italian regions as highlighted in the report published by Istat in 2010. A first estimate of the territorial distribution of house prices for Italy in 2002 has been done by Cannari and Faiella (2008) by using three different data sources.

In order to fill this gap, we believe that it is important to compute preliminary estimates of sub-national SPIHRs by using hedonic regression models for all the Italian regions.

To this aim, we used data from OMI (Osservatorio del Mercato Immobiliare; Real Estate Market Observatory) which is a part of the Italian Agency of Revenue and Tax. The OMI database is a comprehensive source of information on house prices and rents in Italy with a wide geographical coverage, which provides quotations on a semi-annual basis regarding minimum – maximum market prices (sale or rental) per unit area (square meters), and information on the location and the characteristics of the neighbourhood of the dwellings, which are considered the most important determinants of housing prices and rents. It is worth noting that the OMI data have been previously used by Ghirardo (2013) and Ghirardo et al. (2014) in order to analyse the role of dwelling characteristics in determining housing prices and rents.

The rest of the article is organized as follows: Section 2 reviews the literature on using hedonic regression models for temporal and spatial price index construction. Section 3 focuses the main characteristics of the hedonic approach for estimating spatial price indexes for housing rents. Section 4 describes the characteristics of the OMI database used. Section 5 shows the specification of the hedonic models in our study and the approaches employed to estimate regression models with symbolic interval data referring to the minimum and maximum values of the rental prices. Section 6 discusses the estimated results and the sub-national SPIHRs obtained. Finally, Section 7 presents some concluding remarks and future research suggestions.

2. Literature Review on Using Hedonic Regression Models for Price Indexes: A Focus on Spatial Price Indexes for Housing Costs and Rents

Hedonic pricing regression models have long been used for conducting empirical analyses during the 1920's; subsequently they have been brought to light through the theoretical works written by Griliches (1971) and Rosen (1974), whose conceptual foundations were derived from Lancaster (1966), and from the application in the automobile industry made by Griliches (1961) and by Ohta and Griliches (1975).

Hedonic models are a useful tool for computing quality-adjusted "temporal" price indexes, in particular for housing prices or rents and, in fact, numerous theoretical and application works have been written on this topic, including some reviews and surveys of the literature (Malpezzi 2003; Herath and Maier 2010; Hill 2013; De Haan and Diewert, 2013). This is also highlighted in the "Handbook on Residential Property Price Indices" drafted by six international organizations (Eurostat 2013). Various National Statistical Institutes computed and published these indexes for house or rent prices (see for example Behrmann and Goldhammer 2017; Istat 2020).

As highlighted by various authors, there are both advantages and disadvantages by using hedonic model since the results of the estimates may be sensitive to the method used (Silver 2011). However, Hill (2013) concluded that the advantages of the hedonic approach outweigh its disadvantages.

The first study on using hedonic regression methods for constructing spatial price indexes for housing costs, was published by Gillingam (1975) who made inter-area comparisons of rent levels by estimating hedonic equations for ten major cities. He combined survey microdata on individual rental units on multiple unit dwellings with information on neighbourhood quality characteristics taken from the 1960 census. The estimations of the hedonic equations for type of rental units for each of the ten cities, were computed using alternative functional forms. Two summary indexes were constructed, the first representative of all multi-unit apartments and the second limited to five room apartments.

The results revealed a substantial variation among place-to-place rent indexes, and that the specification of the group of units to be represented by the index is a crucial aspect of the index design.

After Gillingam's study, several economists at the BLS and Urban Institute focussed on producing an experimental interarea price index (Primont and Kokoski 1990; Kokoski et al. 1994) and more specifically estimates of interarea differences in the shelter cost (Follain and Malpezzi 1980; Ozanne and Thibodeau 1983; Cobb 1984; Can 1992). Moulton (1995) made several novel contributions. He used detailed rent and imputed owners' equivalent rental estimates for approximately 60,000 housing units included in the housing survey within the CPI program data collection, that matched with the neighbourhood data from the decennial Census. The hedonic bilateral price indexes were computed and then multilateral indexes were developed using GEKS.

As previously mentioned in the introduction, BEA continued these experiments in collaboration with other US federal agencies. The Regional Price Parities were computed using the same updated data sets and various methods were used for constructing multilateral. Only indexes for 38 metropolitan and nonmetropolitan urban areas were estimated and then the estimates of the indexes for 50 states were included, plus District of Columbia, and 383 Metropolitan Statistical Areas (Kokoski et al. 1999; Aten, 2005, 2008; Martin et al. 2013).

A substantial improvement was achieved when it was possible to use data obtained from the new American Community Survey (ACS) conducted by the U.S. Census Bureau which provided 500,000 rent price observations per year. As illustrated by Aten (2017, 2019), the estimation of the RPPs is now obtained using the CPI microdata collected by the BLS which are not publicly available, while the ACS rental data are accessible users as Public Use Microdata Samples.

Outside of the US, during a consultation project in Cambodia concerning the estimation of poverty lines (thresholds), Dalén (2006) computed the housing component of the nonfood poverty line over time and space using household data from the Cambodian Socio-Economic Surveys funding interesting results.

The literature review reveals that: (1) the results of the spatial price index for housing cost estimations depend on the model used, as well as the availability of data characteristics; (2) it is essential to have detailed territorial information especially regarding neighbourhood quality characteristics; (3) using the rental price as dependent variable is proved to be a good choice because it well represents the cost of shelter for homeowners; (4) all the researchers concluded that the log transformation of the rental prices has numerous advantages for obtaining valid SPIHRs.

3. Main Features of the Hedonic Approach for Estimating Spatial Price Indexes for Housing Rents

The main characteristic of the hedonic approach to estimate spatial price indexes for housing rents can be summarized by referring to the above-mentioned papers and in particular to those published by Malpezzi (2003) and Herath and Maier (2010).

3.1. The General Model

The hedonic pricing method for housing costs is basically a regression of the price of the house (rent or value) against known relevant determinants (characteristics of the unit) that indirectly affect the price. A classical hedonic equation is as follows:

$$R = f(S, N, L, E, t)$$

where

R = rental price (or value) of the housing unit;
S = structural characteristics of the unit (characteristics of the building: age, type of building as number of units in the building, number of floors, structural materials, presence of garages and gardens, etc; characteristics of the unit: size or floor area, number of rooms, bedrooms and bathrooms, type of plumbing and heating systems, and so on);
N = neighbourhood (access to services as shops, schools and other essential amenities; perhaps an overall neighbourhood review, quality of schools, socio-economic characteristics of the neighbourhood);
L = locational variables (distance from the city center, business district and sub-centres of employment);
E = environmental characteristics (air quality, proximity to parks, beaches, landfills, an so on) and t is an indicator of time.

For the specification of the model a semi log linear form is mainly used, where most of the characteristics and their classifications are dummy variables. There are advantages and disadvantages in using this kind of model: (1) it enables to estimate coefficients which are easy to interpret, as the proportions of the price that are attributable to the respective characteristics; (2) the independent variables (characteristics) can be sub-divided into more specific variables giving flexibility to the specification of the model (for example, the structural characteristics of the building and the dwellings), (3) it often mitigates the common statistical issue known as heteroscedasticity.

For the time t, the general hedonic model can therefore be:

$$\ln R = \beta_0 + \beta_1 S + \beta_2 N + \beta_3 L + \beta_4 E + \varepsilon$$

where ε is the error term.

3.2. The Hedonic Model for Estimating Spatial Price Indexes for House Rents (SPIHRs)

For estimating spatial price indexes using the hedonic method, it is essential to adopt the "space dummy method", that is to include a geographical area variable (A) among the independent variables. The variable A is a set of area dummies that takes into account the number of the areas to be estimated. In this case, considering a sample of *n* independent observations of house' prices, the semi-log formulation can be expressed as follows:

$$lnr_{ij} = \sum_{j=1}^{M} \alpha_j A_j + \sum_{k=1}^{K} \beta_k C_{ik} + \varepsilon_{ij} \tag{1}$$

Where:

- r_{ij} is a vector of price of property i (with $i = 1,2, \ldots .i, \ldots, n)$ in the different areas (regions or provinces) j with $j = 1,2, \ldots, j, \ldots, M;$
- A_j is a vector of geographical areas dummies;
- C_{ik} is the matrix of the house characteristics (with $k = 1, \ldots, K$);
- β_k is the vector of hedonic regression coefficients called also characteristic shadow prices;
- ε_{ij} is the error term, that satisfies the standard assumption of a multiple regression model.

C_{ik} can be subdivided into the above-mentioned sub-groups of characteristics, such as S = structural characteristics of the building, N = locality/neighborhood and E = environmental characteristics.

The hedonic approach for estimating SPIHRs expressed by model (1), refers to a relatively new strand of the stochastic approach for price index number construction. According to Rao and Hajargasht (2016), model (1) in its multiplicative form postulates that the observed rent prices, r_{ij}, can be expressed as the product of three components: the general price level in area j relative to reference base area (denoted by $SPIHR_r$), the price level of the k-th characteristic to a base characteristic (denoted by P_k) and a random disturbance term u_{ij},

$$r_{ij} = SPIHR_j \cdot P_{ik} \cdot u_{ij}$$

The additive form of the model is obtained by taking logarithms of both sides Equation (2):

$$lnr_{ij} = a_j + b_{ik} + \varepsilon_{ij}$$

This model can be expressed as a regression equation for each rent where the independent variables are dummy variables, thus obtaining model (1). In order to estimate SPIHRs by model (1), it is essential to establish the reference or base area that could be the specific area j* (region or province) or the average of areas (regions or provinces). Moreover, for each set of h classifications, one βhk is equal to zero so that the equation is not overidentified. The remaining parameters were estimated using $j*$ and $\alpha*$ data as reference. In this case α_j is the difference of (fixed) effects associated with the geographical areas compared with the base area j*. After having estimated the α_j parameters, the SPIHR of the area j with respect to the base area is given by SPIHR j = $e^{\alpha j}$.

4. The Characteristics of the OMI Data-Base and the Specific Data Used

In order to compute SPIHRs for Italy at sub-national level we used freely available user data from OMI (which is a part of the Italian Agency of Revenue and Tax) focused on residential buildings (OMI 2017). This data source is the most comprehensive source of information on Italian houses price rents with a wide geographical coverage, including data for each Italian municipality (8,046 municipalities). It provides rental value estimates on a half-yearly basis for different types of dwellings, including the maintenance and conservation status of the building, according to positional factors.

The OMI's objective is to estimate the value of possible rents for all residential building stock in each municipality, on the basis of the Cadastre of Real Estate databases. The OMI data production system stands out both for the survey design and for the rental evaluation of the dwellings which are explained in depth in a specific Manual (OMI 2017).

4.1. Survey Design

The survey design adopted for each municipality takes into account the complexity and heterogeneity that characterize the housing market and the fact that the positional factor is the factor that strongly affects property and rental value.

In order to consider the “locational” factor and the economic and socio-environmental characteristics of the areas, a hierarchical territorial classification for each Italian municipality is adopted. Firstly, each municipal territory is divided into districts, which represents a territorial area with a precise geographical location in the municipality and reflects a consolidated urban location. Subsequently, to obtain adequate estimation of the rents, each district is further sub-divided into homogeneous zones defined by referring to the housing market characteristic. These homogeneous zones are known as OMI zones in Italy.

More specifically, the homogeneous zone is identified through a price and quality evaluation process, according to homogeneous conditions in terms of economic and socio-environmental characteristics, such as presence and accessibility to public and private facilities and services, the quality of urban and suburban transport services, road connections, the presence of educational, health, sports, commercial and tertiary facilities. It was decided to establish ex-ante territorial segmentation due to the fact that the heterogeneity of dwellings is often expressed using territorial clusters that takes into account the location and neighbourhood of the dwellings. This implies that the segmentation of the housing market based on homogeneous zones can significantly reduce the variability of the rental values.

The identification of the districts, which are equal to five specified as B = central area, C = semi-centre, D = outskirts, E = suburbs and R = rural and extra-urban and the specification of the homogeneous zones are strongly affected by the size of the municipality. In Italy there are many small municipalities: approximately 2,000 municipalities have less than 1,000 inhabitants and 5,500 municipalities have less than 5,000 inhabitants. Therefore, frequently, in the small municipalities only one or two districts are identified and within the districts sometimes only one OMI zone is identified.

In fact, from the OMI report, we know that 5,667 municipalities have at maximum three of OMI zones; 1,883 municipalities have at maximum from four to six OMI zones; 358 municipalities have at maximum from seven to ten OMI zones, 106 municipalities have at maximum from 11 to 20 OMI zones, 19 municipalities have at maximum 21 to 30 OMI zones, while 13 municipalities have more than 30 OMI zones. Bearing in mind that not all the OMI zones are included in each municipality, the frequency distribution of the number of OMI zones in the 8,046 Italian municipalities produces a total of occurrences equal to 27,426.

In conclusion, referring to the classification of independent variables mentioned in the Section 3, surely the districts can be considered as a location variable, while the OMI zones capture both the neighbourhood variable and the environmental characteristics, whose effects on price rent have almost never been captured in the previous studies mentioned in the literature review. However, it is necessary to point out that OMI zones are also distinct from each other according to their location in the various districts.

Therefore, the OMI zone classification contains detailed information regarding the location of the dwellings. In order to verify the validity and coherence of the definition of the zones in terms of their characteristics and homogeneity of prices for the different type of dwellings, OMI also conducts an additional stratified sample survey on a regular basis.

For this reason, the survey is only conducted in approximately 1,500 municipalities in order to follow the evolution of the market and to support the computation of temporal House Price Index (Istat 2020). The sample survey also aims to support the estimations of the appraisal process and to obtain more detailed information on the characteristics of the residential buildings and dwellings which would increase the number of quality characteristics to be taken into account when applying the hedonic model. However, this kind of data is not publicly available, but it is only accessible after signing a research protocol with the OMI Agency. The coherence between the confidence interval of the estimates of the average rental values and the interval min-max quotations estimated by the appraisal process are verified..

As a consequence, from an econometric point of view these two variables are collinear and cannot be included in the same model specification. As we will illustrate in Section 5, we decided to use these two different variables in order to better understand their influence on rental prices.

4.2. *The Process for Estimation of Housing Rents*

The rental valuation should be carried out for each dwelling within each OMI zone. However, this is not a practical method as the *average rent* for the different building characteristics of the dwellings is estimated by taking into account the conservation status of the building.

The following typologies of dwellings are defined according to their homogeneous distribution, organizational and functional characteristics, as follows: well-finished apartments; economic apartments, detached houses, luxury dwellings, houses and dwellings typical of the area. Not all dwellings types are located in each OMI zone and the most commonly found dwelling are detached houses, well-finished apartments and economic apartments.

Detached houses generally have gardens while well-finished apartments are dwellings made with expensive materials and finishing (floors, walls, fixtures, plant accessories) of high architectural value. These dwellings have average sized-rooms with at least two bathrooms. Economic apartments, that also includes cheap dwellings, are made with low-cost buildings materials and finishing. The surface area of the building is exploited maximally, with small sized-rooms, only one bathroom and are located in working-class areas.

It is worth noting that social housings are included in the economic apartments. For this dwelling type and for dwellings located in rural districts a corrective coefficient is applied in order to take into account the specific characteristics of the property, both intrinsic and extrinsic, which determine the different value of these properties with respect to properties with average characteristics of the area.

The conservation status of the building, which include the quality of the building materials and workmanship and the general atmosphere of the property, is divided into three categories: excellent, normal and poor quality.

As regards the survey design, there are two methods for estimating rents: the indirect or appraisal method (based on the appraisal expertise of the local housing market experts) and the direct method (based on the estimations of rents through a sample survey of the housing market rents).

In order to estimate the average rents for all municipalities and zones, an appraisal process is carried out by local technicians and verified by specific technical committees at provincial and central level, which use both data included in the cadastre data base and specific rental data obtained from rental agencies and from specific surveys.

The valuation process is based on the suggestions and methods specified in the "International Valuation Standards: A guide to the Evaluation of Real Property Assets", established by a specific international Council (Parker 2016).

The average rental values are expressed per square meter, that is they are standardized according to the dimension of the dwelling. However, it is difficult to carried out a point estimation of the rental value and therefore, the OMI committees set a *minimum* and *maximum* values for each typology of dwelling and establish that the deviation between the two values must not be higher than 50%.

It is worth noting that constructing representative samples of rental values is statistically challenging due to the heterogeneity of the market and to the fact that transaction data are often incomplete (BIS 2019). Indeed, the estimation may prove to be unreliable due to a reduced number of observations (Lopez and Hewings 2018). However, the OMI carries out quality control of sample survey data following the international standards established in the Italian Code of Official Statistics.

In conclusion, by taking into account the general survey design and the process used for the estimation of rental values, we believe that the OMI data are a useful tool for calculation of the SPIHRs.

4.3. Available and Used Data

As already underlined, the OMI data base contains two quotations (minimum–maximum) for each type of dwelling (detached houses, well-finished apartments and economic apartments) and their concerning conservation status (excellent, normal, poor quality) for each homogenous zone in each municipality for a total of 54,214 observations. In order to enhance the quality of price comparisons, we modified the OMI zone classification by aggregating specific zones for various municipalities characterized by similar socio-economic features and rental price distributions, thus obtaining 35 different OMI zones with a reliable overlap of the rental prices among geographical areas. Regarding the type of dwellings, it is worth noting that for the year 2017, 28% of the 54,214 residential buildings are well-finished apartments, 30% are economic apartments and 42% are detached dwellings.

Economic apartments are mainly present in the Southern Italian Regions (Basilicata, Calabria, Molise and Sicilia), while well-finished apartments are mainly found in Trentino-Alto Adige, Valle d'Aosta and Lazio. Two observations can be made from the available quotations. Firstly, very few quotations consider poorly conserved dwellings (only 131). Secondly, in some regions there are few quotations for excellently conserved apartments, especially regarding economic apartments in Trentino Alto Adige. Table 1 and 2 show the descriptive statistics of the rents per m^2 (euro per month) based on the

Table 1. Descriptive statistics based on the quotations available for all the municipalities. Italy, 2017.

	Conservation status					
	Total		Excellent		Normal	
Type of dwelling	Min	Max	Min	Max	Min	Max
Detached						
Mean	3.60	4.86	4.72	6.12	3.36	4.58
Standard deviation	2.04	2.78	2.42	3.39	1.86	2.55
Coefficent of variation	0.57	0.57	0.51	0.55	0.55	0.56
Skewness	2.41	2.65	2.50	2.85	2.36	2.45
Well-finished						
Mean	3.99	5.37	4.90	6.37	3.83	5.19
Standard deviation	2.09	2.94	2.54	3.76	1.96	2.73
Coefficent of variation	0.52	0.55	0.52	0.59	0.51	0.53
Skewness	2.39	2.94	2.89	3.89	2.13	2.32
Economic						
Mean	2.74	3.73	3.50	4.38	2.68	3.68
Standard deviation	1.59	2.13	1.97	2.53	1.54	2.08
Coefficent of variation	0.58	0.57	0.56	0.58	0.58	0.57
Skewness	2.41	2.25	3.43	3.12	2.19	2.09
Total						
Mean	3.46	4.67	4.58	5.91	3.27	4.47
Standard deviation	2.00	2.73	2.44	3.46	1.85	2.53
Coefficent of variation	0.58	0.59	0.53	0.59	0.57	0.57
Skewness	2.36	2.68	2.66	3.26	2.20	2.31

quotations for all of the Italian municipalities and on the average quotations of the for the 20 Italian regions according to type of dwellings and conservation status. Quotations for poorly conserved dwellings are very few to compute adequate descriptive statistics.

Table 1 shows the variability of the rental quotations observed for the 8,046 municipalities and across the Italian regions and how type of dwelling and conservation status affect the distribution of the minimum and maximum rental prices. As illustrated in Table 1, the highest rental value is observed for well-finished apartments, for which the rental price ranges from 3.99 to 5.37 EUR/m^2. In particular, the mean rental price ranges from 4.90 to 6.37 EUR/m^2 for excellently conserved apartments (the maximum rental price shows as well the highest level of variability). Contrastingly, the lowest rental value is observed for economic apartments, for which the rental prices range from 2.73 to 3.72 EUR/m^2.

From the second part of the Table 1, we can see that also the variability of the average rental values across regions is very high and this surely justify the importance to compute the SPIHRs for the Italian Regions.

Table 2. Descriptive statistics based on averages computed for the 20 Italian regions. Italy, 2017.

	Conservation status					
	Total		Excellent		Normal	
Detached	Min	Max	Min	Max	Min	Max
Min	2.51	3.44	3.00	3.00	2.50	3.42
Max	6.04	8.14	10.09	13.00	6.60	8.97
Standard Deviation	2.04	2.78	2.42	3.39	1.86	2.55
Well-finished						
Min	1.86	2.43	2.46	3.04	1.68	1.23
Max	5.59	7.46	9.41	12.58	5.24	7.00
Standard Deviation	2.09	2.94	2.54	3.76	1.96	2.73
Economic						
Min	1.25	1.67	1.62	2.18	1.23	1.66
Max	7.50	11.65	6.30	7.97	6.30	7.97
Standard Deviation	1.59	2.13	1.97	2.53	1.54	2.08
Total						
Min	1.64	2.16	2.51	3.10	2.51	3.10
Max	5.29	7.22	9.18	12.28	9.18	12.28
Standard Deviation	2.00	2.73	2.44	3.46	1.85	2.53

5. The Specification of the Hedonic Models and the Estimation Procedure

As regards the estimation strategy, it is worth noting that, as already mentioned in Section 4, the OMI database provides minimum–maximum range of rental prices according to OMI zones, type of dwelling and conservation status. Therefore, for each observation two rental prices are specified: the minimum and the maximum rental value. As a consequence our dependent variable in model (1) is an interval-valued variable which is defined for a given set of N entities $S = \{s_1, \ldots, s_N\}$, by an application $Y\colon S \rightarrow B$ such that $s_i \mapsto Y(s_i)$, $Y(s_i) = [l_i,\ u_i]$, B is the set of intervals of an underlying set $O \subseteq \mathbb{R}$, In other words, the value of the interval-valued variable Y for each $s_i \in S$ is usually defined by the lower and upper bounds l_i and u_i of $I_i = Y(s_i)$, In order to estimate the hedonic regression model [1] with $r_{ij} = [l_{ij},\ u_{ij}]$, we used the Symbolic Data Analysis approach (SDA) which provides a framework for the representation and analysis of data that comprehends inherent variability (Brito 2014).

Several approaches have been introduced for estimating regression models with symbolic interval data, Billard and Diday (2000) proposed a first approach which consisted in fitting a regression line to the centres of the intervals or midpoints defined as $c_{ij} = (l_{ij} + u_{ij})/2$. Therefore, model (1) can be expressed as:

$$lnc_{ij} = \sum_{j=1}^{M} \alpha_j A_{jC} + \sum_{k=1}^{K} \beta_k C_{ikC} + \varepsilon_{ijC} \tag{2}$$

The classical least squares estimator may be used and standard statistical inference will apply under the standard assumptions regarding the error term of the regression. Although this method can provide information on the average centrality of the intervals, it does not take into account the range of the intervals. There are several proposals aimed to incorporate the length of the interval into the analysis (Neto and De Carvalho, 2008; 2010). However, these methods will produce the same estimates obtained by using the midpoint to method.

Billard and Diday (2002) also proposed an alternative method which consisted of estimating two different regression lines, one for the minimum and another for the maximum value of the intervals with no restrictions across lines as follows:

$$\begin{aligned} lnr_{ijL} &= \sum_{j=1}^{M} \alpha_j A_{jL} + \sum_{k=1}^{K} \beta_k C_{ikL} + \varepsilon_{ijL} \\ lnr_{ijU} &= \sum_{j=1}^{M} \alpha_j A_{jU} + \sum_{k=1}^{K} \beta_j C_{ikU} + \varepsilon_{ijU} \end{aligned} \tag{3}$$

The OLS estimator may be used for minimizing: $\min \sum_i (\varepsilon_{ijL}^2 + \varepsilon_{ijU}^2)$ which entails performing two separate minimizations. However, the OLS estimator would not be the most efficient estimator because it is likely that the two error terms may be correlated given that $r_{ijL} < r_{ijU}$.

In this article we propose expressing models (3) as a system of seemingly unrelated regression Equations (SURE) which take into account the properties of the error terms (Zellner 1963).

Indeed, this set of equations has a contemporaneous cross-equation error correlation (i.e., the error terms in the regression equations are correlated). At first glance, the equations seem unrelated, but the equations are related through the correlation in the errors which is tested using the Breush-Pagan χ-statistics (Cameron and Trivedi 2009). The generalized least-squares algorithm, which produces asymptotically efficient and feasible estimates (Greene 2012) was used.

In order to better understand the influence of the locational factors and the neighbourhood and environmental characteristics on rental prices, we estimated two different model specifications: the first model considers a broader classification for the locational characteristics as specified by the "district variable"; the second model considers mainly the neighbourhood and environmental characteristics as specified by the "OMI zones variable". Since the OMI zone classification is embedded into the district classification and include also information on the location, substantially this variable can be considered as a more detailed definition of the locational characteristics that contains the specific information above mentioned. It is obvious that, for this reason, from an econometric point of view these two variables are collinear and cannot be included in the same model specification.

6. Results: Presentation and Some Comments

The SPIHRs were estimated using both models (2) and (3) by considering first all the dwelling types included in the OMI database and then only the subgroup of the economic apartments. In this way useful information may be provided to economic and social policy makers. The results of the estimates are reported in Tables 3A and 3B respectively. The mid-point included in model estimation (2) is an arithmetic mean of the min and max rental values. Estimation results of model (2) computed with the geometric mean are not significantly different from the estimated results obtained using the arithmetic mean.

Table 3A. Hedonic regional models for all dwelling types with different territorial classifications, first semester (january–june), year 2017.

	Hedonic model with districts			Hedonic model with homogeneous zones		
	Model [2]	Model [3]		Model [2]	Model [3]	
	Midpoint	Min	Max	Midpoint	Min	Max
Regions (*ref. Italy*)						
North						
Piemonte	−0.079	−0.082	−0.075	−0.065	−0.070	−0.061
	(0.005)	*(0.005)*	*(0.005)*	*(0.005)*	*(0.005)*	*(0.005)*
Valle d'Aosta	0.216	0.240	0.197	0.227	0.251	0.209
	(0.022)	*(0.022)*	*0.023*	*(0.022)*	*(0.022)*	*(0.022)*
Lombardia	0.173	0.191	0.159	0.174	0.192	0.161
	(0.005)	*(0.004)*	*(0.005)*	*(0.004)*	*(0.004)*	*(0.004)*
Liguria	0.426	0.436	0.419	0.423	0.433	0.416
	(0.008)	*(0.008)*	*(0.008)*	*0.007*	*0.007*	*(0.008)*
Trentino-Alto Adige	0.412	0.422	0.403	0.415	0.425	0.406
	(0.008)	*(0.008)*	*(0.008)*	*(0.008)*	*(0.008)*	*(0.008)*
Veneto	0.102	−0.176	−0.174	−0.176	−0.176	−0.174
	(0.009)	*(0.009)*	*0.010*	*(0.009)*	*(0.009)*	*(0.009)*
Friuli-Venezia Giulia	0.020	−0.022	0.048	0.028	−0.014	0.056
	(0.009)	*(0.009)*	*(0.009)*	*(0.009)*	*(0.009)*	*(0.009)*
Emilia-Romagna	0.222	0.235	0.213	0.218	0.231	0.209
	(0.007)	*(0.007)*	*(0.007)*	*(0.007)*	*(0.007)*	*(0.007)*
Centre						
Toscana	0.435	0.436	0.435	0.433	0.434	0.432
	(0.007)	*(0.007)*	*(0.007)*	*(0.007)*	*(0.007)*	*(0.007)*
Umbria	−0.071	−0.074	−0.068	−0.080	−0.083	−0.078
	(0.013)	*(0.013)*	*(0.013)*	*(0.013)*	*(0.013)*	*(0.013)*
Marche	0.024	0.021	0.026	0.024	0.022	0.027
	(0.008)	*(0.008)*	*(0.008)*	*(0.008)*	*(0.008)*	*(0.008)*
Lazio	0.387	0.388	0.386	0.386	0.388	0.385
	(0.008)	*(0.008)*	*(0.008)*	*(0.008)*	*(0.008)*	*(0.008)*
Abruzzo	−0.179	−0.210	−0.157	−0.182	−0.212	−0.159
	(0.008)	*(0.008)*	*(0.008)*	*(0.008)*	*(0.008)*	*(0.008)*
South and Islands						
Molise	−0.308	−0.351	−0.276	−0.284	−0.328	−0.251
	(0.023)	*(0.023)*	*(0.023)*	*(0.023)*	*(0.023)*	*(0.023)*
Campania	−0.174	−0.168	−0.179	−0.195	−0.188	−0.201
	(0.005)	*(0.005)*	*(0.005)*	*(0.005)*	*(0.005)*	*(0.005)*
Puglia	−0.122	−0.104	−0.134	−0.150	−0.131	−0.163
	(0.008)	*(0.008)*	*(0.008)*	*(0.008)*	*(0.008)*	*(0.008)*

Table 3A. Continued

	Hedonic model with districts			Hedonic model with homogeneous zones		
	Model [2]	Model [3]		Model [2]	Model [3]	
Basilicata	−0.639	−0.623	−0.651	−0.626	−0.611	−0.637
	(0.012)	*(0.012)*	*(0.012)*	*(0.012)*	*(0.012)*	*(0.012)*
Calabria	−0.362	−0.359	−0.364	−0.372	−0.369	−0.374
	(0.008)	*(0.008)*	*(0.008)*	*(0.007)*	*(0.007)*	*(0.008)*
Sicilia	−0.333	−0.346	−0.323	−0.336	−0.349	−0.326
	(0.007)	*(0.007)*	*(0.007)*	*(0.007)*	*(0.007)*	*(0.007)*
Sardegna	−0.149	−0.147	−0.151	−0.135	−0.132	−0.136
	(0.010)	*(0.010)*	*(0.010)*	*(0.010)*	*(0.010)*	*(0.010)*
Type of dwelling (*ref. well-finished apartment*)						
Economic apartment	−0.175	−0.176	−0.174	−0.175	−0.176	−0.174
	(0.004)	*(0.004)*	*(0.004)*	*(0.004)*	*(0.004)*	*(0.004)*
Detached	0.125	0.130	0.121	0.130	0.136	0.126
	(0.004)	*(0.004)*	*(0.004)*	*(0.004)*	*(0.004)*	*(0.004)*
Status of the dwelling (*ref. Normal*)						
Excellent	0.194	0.226	0.170	0.196	0.228	0.172
	(0.005)	*(0.005)*	*(0.005)*	*(0.005)*	*(0.005)*	*(0.005)*
District (*ref. Central*)				No	No	No
Semi-central	0.326	0.316	0.333			
	(0.007)	*(0.007)*	*(0.007)*			
Outskirts	0.116	0.114	0.118			
	(0.005)	*(0.005)*	*(0.005)*			
Suburban	0.031	0.033	0.030			
	(0.005)	*(0.005)*	*(0.005)*			
Rural and extra-urban	−0.255	−0.255	−0.255			
	(0.005)	*(0.005)*	*(0.005)*			
Homogeneous zones	No	No	No	Yes	Yes	Yes
Intercept	1.058	1.058	1.374	1.189	1.020	1.332
	(0.004)	*(0.004)*	*(0.004)*	*(0.004)*	*(0.004)*	*(0.004)*
Correlation		0.975			0.974	
Breusch-Pagan test		51565			51467	
Adjusted R2	0.463	0.478	0.478	0.485	0.471	0.471
RMSE	0.375	0.374	0.381	0.368	0.367	0.372
AIC	47.719	−67.210	−67.210	45.462	−69.667	−69.667
BIC	47.977	−66.644	−66.644	45.969	−68.564	−68.564
N	54,214	54,214	54,214	54,214	54,214	54,214

Source: our elaboration from OMI data 2017.

Note: standard error in brackets; all the coefficients are significant at $p < 0.01$.

Table 3B. Hedonic regional models for economic apartments with different territorial classifications. First semester (january–june), year 2017.

	Hedonic model with districts			Hedonic model with homogeneous zones		
	Model [2]	Model [3]	Model [3]	Model [2]	Model [3]	Model [3]
	Midpoint	Min	Max	Midpoint	Min	Max
Regions (*ref. Italy*)						
North						
Piemonte	−0.044	−0.064	−0.028	−0.033	−0.054	−0.016
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Valle d'Aosta	0.146	0.174	0.125	0.161	0.189	0.141
	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)
Lombardia	0.196	0.213	0.183	0.190	0.207	0.178
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Liguria	0.452	0.445	0.457	0.450	0.443	0.455
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Trentino	0.673	0.691	0.659	0.641	0.666	0.623
Alto Adige	(0.177)	(0.179)	(0.178)	(0.175)	(0.175)	(0.175)
Veneto	0.238	0.246	0.153	0.237	0.250	0.158
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Friuli Venezia	0.020	−0.027	0.052	0.031	−0.017	0.064
Giulia	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Centre						
Emilia-	0.213	0.223	0.205	0.211	0.221	0.203
Romagna	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Toscana	0.390	0.399	0.384	0.389	0.397	0.382
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Umbria	−0.070	−0.065	−0.074	−0.076	−0.071	−0.079
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Marche	0.014	0.012	0.016	0.016	0.013	0.018
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Lazio	0.433	0.427	0.438	0.436	0.430	0.441
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Abruzzo	−0.217	−0.249	−0.193	−0.217	−0.249	−0.193
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
South and Islands						
Molise	−0.360	−0.402	−0.329	−0.339	−0.382	−0.307
	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)
Campania	−0.212	−0.214	−0.211	−0.230	−0.231	−0.230
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Puglia	−0.140	−0.117	−0.156	−0.158	−0.135	−0.175
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Basilicata	−0.734	−0.720	−0.743	−0.724	−0.712	−0.733
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Calabria	−0.368	−0.363	−0.371	−0.373	−0.368	−0.375
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Sicilia	−0.415	−0.427	−0.405	−0.419	−0.431	−0.409
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Sardegna	−0.215	−0.212	−0.218	−0.193	−0.190	−0.197

Table 3B. Continued

	Hedonic model with districts			Hedonic model with homogeneous zones		
	Model [2]	Model [3]	Model [3]	Model [2]	Model [3]	Model [3]
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Status of the dwelling (*ref. Normal*)						
Excellent	0.193	0.246	0.153	0.198	0.250	0.158
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
District (*ref. Central*)						
Semi-central	0.312	0.305	0.317			
	(0.012)	(0.012)	(0.012)			
Outskirts	0.117	0.116	0.117			
	(0.008)	(0.008)	(0.008)			
Suburban	0.009	0.010	0.008			
	(0.009)	(0.009)	(0.009)			
Rural and extra-urban	−0.294	−0.295	−0.293			
	(0.008)	(0.008)	(0.008)			
Homogeneous zones	No	No	No	yes	yes	yes
Intercept	1.078	0.906	1.222	1.046	0.876	1.189
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Correlation		0.978	0.978		0.977	0.977
Breusch-Pagan test		15,345	15,345		15,322	15,322
Adjusted R2	0.457	0.463	0.449	0.477	0.482	0.470
RMSE	0.373	0.376	0.374	0.365	0.369	0.366
AIC	13.873	−22.242	−22.242	13.275	−22.888	−22.888
BIC	14.080	−21.781	−21.781	13.698	−21.958	−21.958
N	16,040	16,040	16,040	16,040	16,040	16,040

Source: our elaboration from OMI data 2017.
Note: standard error in brackets; all the coefficients are significant at $p < 0.01$.

No evidence of functional form misspecification was observed from the Ramsey test (1969). The estimation results of models (2) and (3) at regional level for all the dwelling types are reported in Table 3A by considering two model specifications. The first three columns refer to districts, while the last three columns refer to homogeneous territorial zones.

As regards the estimation results, the best fitting-models are obtained when homogeneous zones are included. Moreover, model (3) is preferable to model (2) as it allows for the correlation between ε_{ijL} and ε_{ijU} and the estimation of the full variance–covariance matrix of the coefficients. As shown in the last rows of Table 3, the correlation of the residuals in the minimum and maximum rent equations is equal to 0.975. We can reject the hypothesis that this correlation is zero on the basis of the Breush-Pagan test of independence results. The correlation is strong, so the efficiency gains with the SUR estimation is great (Cameron and Trivedi 2009).

Table 3A shows different distributional patterns as rental prices vary greatly across the Italian regions, thus confirming territorial heterogeneity. The lowest rental prices were observed for Basilicata and Molise while the highest rental prices were observed in Trentino Alto-Adige and Liguria (both for midpoint and minimum- maximum models).

As expected, the localization factor plays an important role in determining rental prices. However, it is interesting to note that the rents of houses located in central areas of the municipalities cost 11.4% – 11.8% less than the rents of dwellings on the outskirts as shown in Table 3A while houses located in rural areas cost 25.5% less than houses located in central areas of the municipalities. This effect may be due to the heterogeneity of houses included in the district classification which does not take into account the socio-economic characteristics of the territorial area. More specifically, the district variable captures an average effect of the locational characteristics since it is obtained by considering rental prices related to all the municipalities included in each region. Yet, Italy is characterized by the presence of numerous small municipalities in which the rent values are smaller than those observed in the big municipality and the more expensive houses (well-finished dwelling and detached houses) are located in semi-central and outskirts districts. Contrastingly, we have found that Italian regional capitals, such as Florence (Tuscany), Rome (Lazio) or Milan (Lombardy), show a different behaviour of housing rental prices with more expensive houses located in the central districts.

When considering a detailed specification for the localization factors (OMI zones), the effects of type-of-dwelling and property status do not vary: detached houses cost 13.6% and 12.6% more than rented well-finished apartments, while excellently conserved houses cost 22.8% and 17.2% more than those with normal conservation status.

Table 3B reports hedonic regional model estimation results for economic apartments by using models (2) and (3) specifications. As regards the estimation results, the best fitting-models are obtained when homogeneous zones are included, and model (3) is preferable to model (2). The number of observations is equal to 16,040 and estimated results confirm high territorial heterogeneity across the Italian regions, highlighting disparities between the Northern and Southern regions. As regards the localization factors, the coefficients associated to the districts dummy continue to assume the same signs but are lower than the coefficients estimated in Table 3A for all the dwelling type thus confirming the influences of houses rental prices in small municipalities.

Table 4 reports the SPHIRs obtained when homogeneous territorial zones are included in the models (2) and (3) if we consider the 20 Italian regions and compare every region to the overall mean (Italy = 100). The SPHIRs are calculated both for all dwelling type and the economic apartments.

On observing regional models reported in Table 4, our results highlight high territorial heterogeneity and confirming significant rental price differences across the Italian regions thus demonstrating that rents are higher in the Northern-Central regions than in the South in respect to the whole of Italy. Basilicata proved to be the less expensive region for rental properties while Tuscany, Liguria while the highest rents were found in Trentino-Alto Adige. Significant differences were observed in the estimated SPHIRs, when all of the dwelling type were included with the exception of the estimation for economic apartments.

Table 4. Regional SPIHRs (Italy = 100), year 2017.

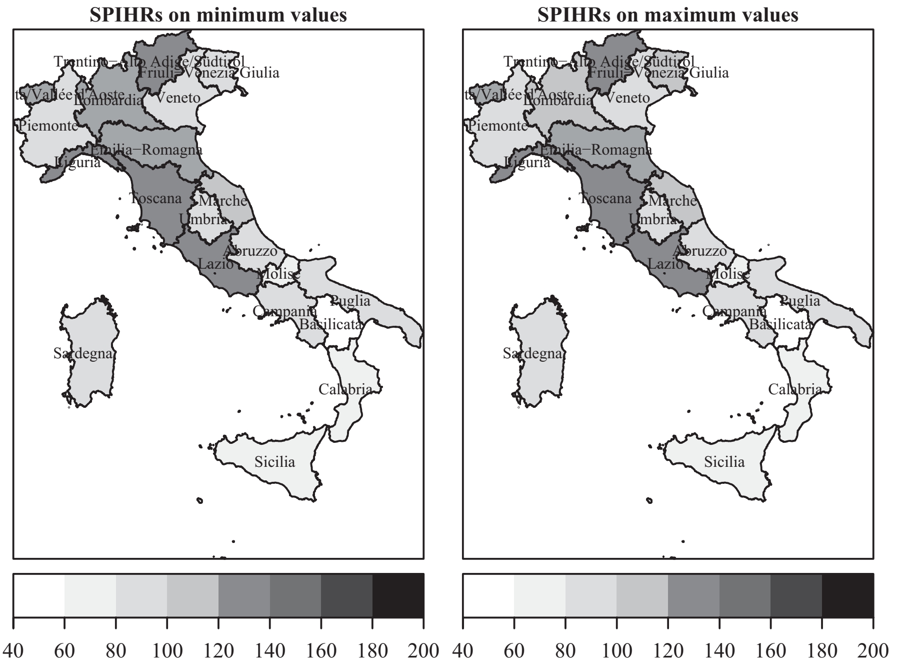
	All dwelling types			Economic apartments		
	Model [2]	Model [3]		Model [2]	Model [3]	
Region	Midpoint	Min	Max	Midpoint	Min	Max
North						
Piemonte	93.72	93.28	94.09	96.76	94.70	98.37
Valle d'Aosta	125.48	128.58	123.26	117.52	120.78	115.18
Lombardia	119.00	121.12	117.45	120.98	122.99	119.48
Liguria	152.73	154.18	151.58	156.87	155.81	157.66
Trentino-Alto Adige	151.39	152.95	150.06			
Veneto	110.09	83.82	84.02	126.69	128.45	117.14
Friuli-Venezia Giulia	102.83	98.61	105.81	103.15	98.28	106.59
Emilia-Romagna	124.37	126.00	123.23	123.43	124.76	122.52
Centre						
Toscana	154.13	154.31	154.08	147.48	148.78	146.58
Umbria	92.29	92.01	92.53	92.66	93.14	92.37
Marche	102.45	102.19	102.70	101.59	101.35	101.82
Lazio	147.11	147.34	147.01	154.65	153.66	155.47
Abruzzo	83.36	80.89	85.26	80.49	77.94	82.45
South						
Molise	75.31	72.05	77.80	71.27	68.28	73.55
Campania	82.29	82.90	81.80	79.43	79.36	79.46
Puglia	86.09	87.72	84.94	85.38	87.33	83.98
Basilicata	53.47	54.30	52.89	48.48	49.09	48.06
Calabria	68.96	69.16	68.83	68.90	69.20	68.70
Sicilia	71.44	70.53	72.15	65.79	64.96	66.45
Sardegna	87.41	87.62	87.26	82.41	82.73	82.15

Source: our elaboration from OMI data 2017.

As regards economic apartments, the SPHIRs estimated for Trentino-Alto Adige are not significant due to the reduced number of observations. In this case, the SPHIRs estimated for Veneto and Lazio are higher than the SPHIRs estimated for all types of dwelling, while the SPHIRs estimated for Toscana, Abruzzo, Molise, Sicilia and Sardegna are lower than the SPHIRs estimated for all dwelling types. These findings may be influenced by two contrasting effects which emerge when estimating SPHIRs at regional level: the inclusion of tourist destinations characterized by higher rents and small municipalities with lower rents. Taking into account the various housing characteristics provided by the OMI data set, it is clear that the role played by tourism is stronger than the effect of the prevalence of low-cost houses with economic apartments.

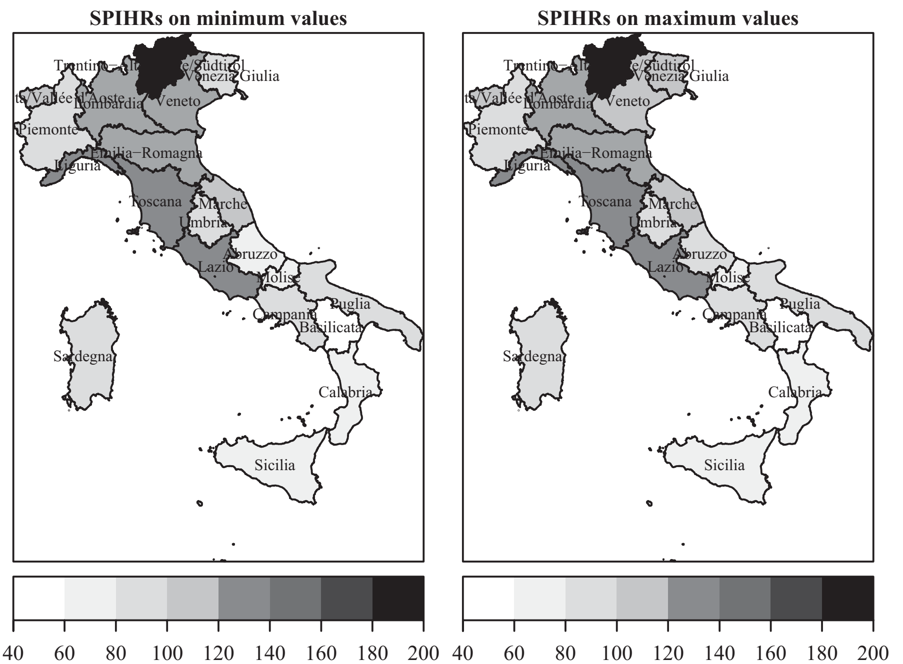
In order to explore the territorial heterogeneity of rents within each Italian region, Figures 1–4 show sub-national SPHIRs estimate with model (3) if we consider homogeneous territorial zones for the Italian regions (Italy = 100).

Figures 3 and 4 show sub-national SPHIRs with the corresponding intervals at 95%. Although most regions have small confidence intervals, few regions have overlapping confidence intervals. Overlapping confidence intervals were observed between Lombardia and Liguria SPIHRs and Calabria and Sicilia SPIHRs for the model estimated for all dwelling types and for the economic apartments. With the aim of making multiple



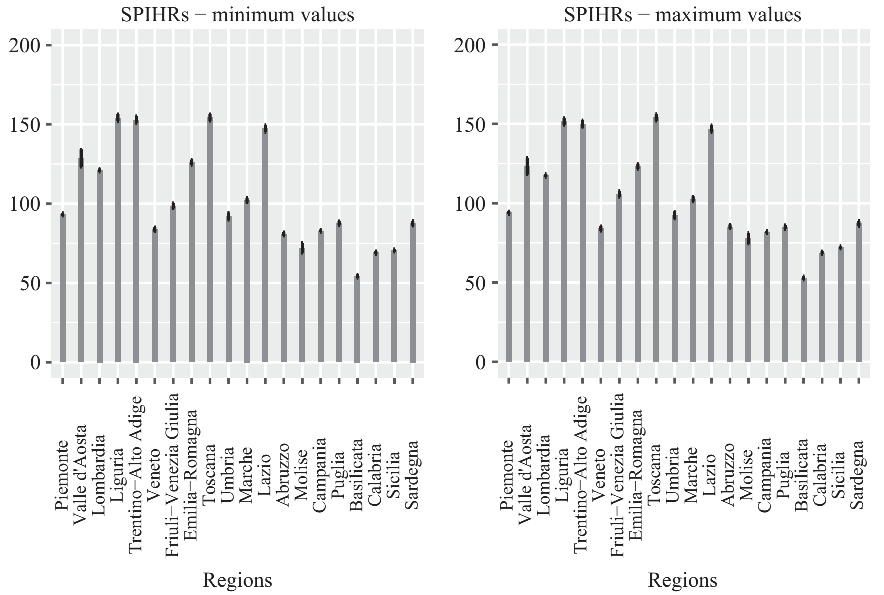
Source: our elaboration from OMI data 2017

Fg. 1. SPIHRs for Italian regions for all dwellings types, year 2017.



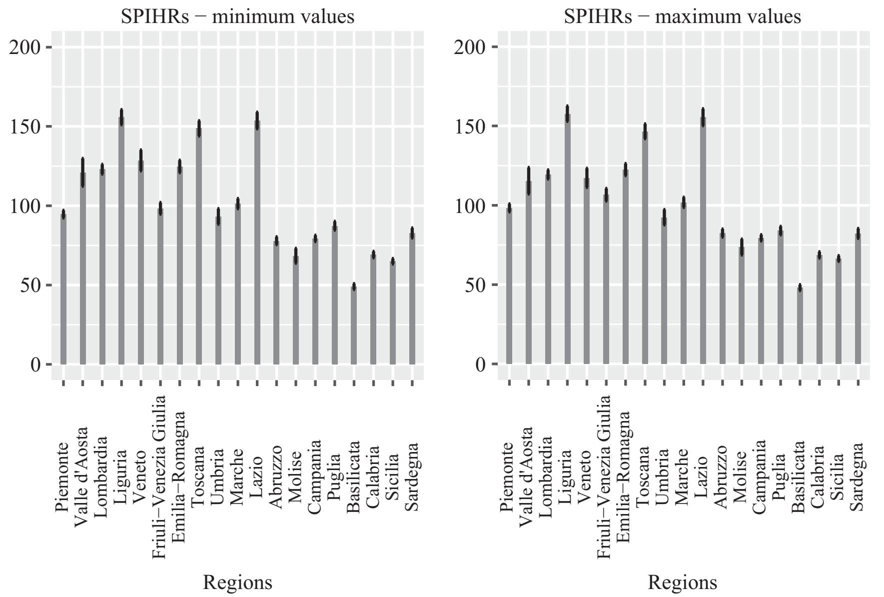
Source: our elaboration from OMI data 2017

Fig. 2. SPIHRs for Italian regions for economic apartments, year 2017.



Source: our elaboration from OMI data 2017

Fig. 3. SPIHRs for Italian regions with confidence intervals for all dwellings types, year 2017.



Source: our elaboration from OMI data 2017

Fig. 4. SPIHRs for Italian regions with confidence intervals for economic apartments, year 2017.

comparisons based on interval confidence, it would be advisable to use simulation methods as suggested by Marshall and Spiegelhalter (1998) and Leckie and Goldstein (2011) and applied by Biggeri et al. (2017) in the field of price statistics.

7. Concluding Remarks

The main objective of this article was to explore the feasibility of using the OMI data set for compiling SPIHRs for the Italian regions on a regular basis. Spatial price indexes measuring differences in price levels across regions within a country are essential for comparing real income, standards of living and consumer expenditure patterns. Housing is the most expensive cost of the household budget, and several economic and social players have highlighted the importance of providing spatial rental price indexes for countries like Italy.

With the aim of comparing rental prices across the Italian territory using hedonic regression models, various classifications of "positional factors" are available on OMI database.

The results of our computations are well-suited for the hedonic models used and for the significant differences observed across the various Italian regions. The SPIHRs for economic apartments are particularly interesting for policy makers and confirm the usefulness of calculating SPIHRs on a regular basis, every two years. However, it is important to carry out further analysis by computing the SPIHRs for other years from 2010 up to now considering also spatial dependencies inherent to rental prices will be considered by estimating spatial models.

Second, if the access to additional data is granted, further analyses will be carried out for the year 2019 by using detailed information on all of the dwelling and building characteristics for a sample of approximately 1,500 municipalities collected by OMI in order to compare the results.

Finally, it is worth noting that the strong differences found in the regional estimates of the SPIHRs surely reshuffle in some way the territorial distribution of the income and poverty indicators.

However, to assess to what extent our findings impact the regional differences in the standard of living and poverty levels in the Italian context, other sources of data on income and living conditions should be used and analysed by considering different poverty indicators computed using adjusted poverty thresholds. This important topic will constitute a further line of research.

8. References

Aten, B. 2005. *Report on Interarea Price Levels, 2003*, working paper 2005–11, U.S. Bureau of Economic Analysis. Available at: https://www.bea.gov/system/files/papers/WP2005-11.pdf (accessed May 2020).

Aten, B. 2008. *Estimates of State and Metropolitan Areas Price Parities for Consumption Goods and services in the United States, 2005*', U.S. Bureau of Economic Analysis. Available at: https://www.bea.gov/system/files/papers/P2008-1.pdf (accessed May 2020).

Aten, B.H. 2017. "Regional price parities and real regional income for the United States", *Social Indicators Research*, 131(1): 123–143 DOI: https://doi.org/10.1007/s11205-015-1216-y.

Aten, B.H. 2019. "Case Study of Subnational PPPs: United States", paper presented at the 4th Task force meeting the ICP, World Bank, in May, Paris. Available at:

https://pubdocs.worldbank.org/en/148641591895276884/pdf/ICP-Annual-Report-2019.pdf (accessed May 2020).

Aten, B.H. and E. Figueroa. 2015. "Real Personal Income and Regional Price Parities for States and Metropolitan Areas 2009-2013", *Survey of Current Business*, 95(7): 1–4.

Behrmann, T., and B. Goldhammer. 2017. "New developments in the field of house and rental price indices in German price statistics in the light of the hedonic method." In: Deutshe Bundesbank-Ottawa Group (2017), 15th Meeting of the Ottawa Group on Price Indices, Hosted by the Deutshe Bundesbank, 10–12 May, Eltville am Rhein, Germany. Available at: https://www.bundesbank.de/resource/blob/635276/865b648d8fb3d6a23-d327e59c2f48040/mL/2017-05-10-ottawa-group-02-2-abstract-data.pdf (accessed May 2020).

Biggeri, L., T. Laureti, and F. Polidoro. 2017. "Computing sub-national PPPs with CPI data: an empirical analysis on Italian data using country product dummy models", *Social Indicators Research* 131(1): 93–121. DOI: https://doi.org/10.1007/s11205-015-1217-x.

Biggeri, L., and M. Pratesi. 2017. "Monetary poverty indicators at local level: definitions, methods of estimations and comparisons in real terms." In Proceedings of the 60th World Statistics Congress of the International Statistical Institute, 16–21 July, Marrakech, Marocco. Available at: http://www.centrodagum.it/wp-content/uploads/2018/07/Biggeri-Pratesi-ISI2017.pdf (accessed June 2020).

Biggeri, L., and T. Laureti. 2018. "Publications, experiments and projects on the computation of spatial price level differences in Italy", paper presented at the 3rd Task force meeting the ICP, World Bank, 27th September, Country case studies: Italy. Available at: www.worldbank.org/en/programs/icp/brief/tf03-doc03 (accessed May 2020).

Billard, L, and E. Diday. 2000. *Regression analysis for interval-valued data, In Data Analysis, Classification, and Related Methods*. Springer: 369–374. DOI: https://doi.org/10.1007/978-3-642-59789-3_58.

Billard, L., E. Diday. 2002. "Symbolic regression analysis, In: Classification, Clustering and Data Analysis", In Proceedings of the Eighteenth Conference of the International Federation of Classification Societies (IFCS'02), July 16–19, 2002. Krakow, Poland, Springer: 281–288. DOI: https://doi. org/10.1007/978-3-642-56181-8_31.

BIS. 2019. *Mind the data gap: commercial property prices for policy*, Irving Fisher Committee on Central Bank Statistics, Bank for International Settlements. Available at: https://www.bis.org/ifc/publ/ifc_report_cb_trade_rep_deriv_data.pdf (accessed May 2020).

Bishaw A. 2009. "Adjusting Poverty Thresholds Based on Differences in Housing Cost: Application of American Community Survey", Poster Presentation Prepared for the Population Association of America April 30–May 2, 2009. Available at: https://www.census.gov/content/dam/Census/library/working-papers/2009/demo/bishaw-paa-col-3-4.pdf (accessed May 2020).

Bishop J.A., J. Lee, and L.A. Zeager. 2017. "Incorporating spatial price adjustments in US public policy analysis", *ECINEQ WP*, 438. Available at: http://www.ecineq.org/milano/WP/ECINEQ2017-438.pdf (accessed May 2020).

Brito, p. 2014. “Symbolic Data Analysis: another look at the interaction of Data Mining and Statistics”, *Data Mining and Knowledge Discovery* 4(4): 281–295. DOI: https://doi.org/10.1002/widm.1133.

Cameron, A.C, and P.K. Trivedi. 2009. *Microeconometrics with STATA*. College Station, TX: StataCorp LP.

Can, A. 1992. “Specification and estimation of hedonic housing price models”, *Regional Science and Urban Economics* 22: 453–474. DOI: https://doi.org/10.1016/0166-0462(92)90039-4.

Cannari, L. and I. Faiella. 2008. *House Prices and Housing Wealth in Italy.* Bank of Italy Temi di Discussione (Working Paper) No. A4. Available at: http://dx.doi.org/10.2139/ssrn.1162109 (accessed June 2020).

Citro C.F., T.T. Michael. 1995. *Measuring Poverty: A New Approach*, National Academy Press, Washington, D.C. Available at: https://library.econ.upd.edu.ph/ir/handle/123456789/6535 (accessed May 2020).

Cobb, S.A. 1984. “The impact of site characteristics on housing cost estimates”, *Journal of Urban Economics*, 15: 26–45. DOI: https://doi.org/10.1016/0094-1190(84)90021-4.

Dalaker, J. 2017. P*overty in the United States in 2016*: In Brief, U.S. Congressional Research Service, 7–5700. Available at: https://crsreports.congress.gov/product/pdf/R/R45397.

Dalén, J. 2006. “Spatial price comparisons in poverty measurement, An example from Cambodia”. Ninth Meeting of the Ottawa Group-International Working Group on Price Indices, April 22, London. Available at: https://www.ottawagroup.org/Ottawa/ottawagroup.nsf/home/Meeting + 9/$file/2006%209th%20Meeting%20-%20Jorgen%20Dalen%20-%20Spatial%20price%20comparisons%20in%20poverty%20measurement.pdf.

De Haan, J. and E. Diewert. 2013. “Hedonic Regression Methods”, In OECD et al. *Handbook on Residential Property Price indices*, Eurostat, Luxembourg: chap. 5. Available at: https://ec.europa.eu/eurostat/documents/3859598/5925925/KS-RA-12-022-EN.PDF (accessed May 2020).

Eurostat. 2013. *Detailed Technical Manual on Owner-Occupied Housing for Harmonised Index of Consumer Prices*, Eurostat, Luxembourg. Available at: https://ec.europa.eu/eurostat/documents/7590317/0/Technical-Manual-OOH-HPI-2017/ (accessed May 2020).

Ferrante C., T. Laureti, and F. Polidoro. 2019. “Combining data coming from scanner, traditional CPI data collection and other sources to compile sub-national PPPs in Italy”. poster presented at the Sixtheenth Meeting of the Ottawa Group, 8–10 May, Rio de Janeiro. Brazil. Available at: https://www.ottawagroup.org/Ottawa/ottawagroup.nsf/home/Meeting+16/$FILE/Combining%20data%20coming%20from%20scanner%20abstract.pdf.

Follain, J.R., and S. Malpezzi. 1980. “Dissecting housing Value and rent: Estimates of hedonic indexes for thirty-nine large SMSAs”, *Urban Institute Press*, Washington D.C.

Gillingam, R. 1975. “Place to Place Rent Comparisons”, *Annals of Economic and Social Measurement*” 4(1): 153–173. N.B.E.R. Available at: http://www.nber.org/books/aesm75-1 (accessed May 2020).

Ghirardo, E. 2013. *Aggregazione di coefficient di un modello edonico, un approccio di regressione geografica ponderata*, Quaderni OMI, Roma. Available at: https://www.agenziaentrate.gov.it/portale/documents/20143/263967/Quaderni+dell+Osservatorio+-dicembre+2013_Quaderni_Anno_II_N_2_Dic_2013.pdf/7aefea40-c028-fd8d-a893-821b8d7e3aa0 (accessed May 2020).

Ghirardo, E., R. Ciula, and M. Festa. 2014. *Effetti di non linearità nei modelli edonici: un esercizio di regressione quantilica*, Quaderni OMI, Roma. Available at: https://www.agenziaentrate.gov.it/portale/documents/20143/263967/Quaderni+dell+Osservatorio+-dicembre+2014_Quaderni+Dicembre+2014.pdf/80b0ae7a-84a5-cfda-12d4-a5a7f6e71f52 (accessed May 2020).

Greene, W.H. 2012. "Econometric Analysis", 7th edition, Upper Saddle River, NJ: Prentice Hall.

Griliches, Z. 1961. "Hedonic price indexes for automobiles: an econometric analysis of quality changes", In *The Price Statistics of Federal Government*, Washington D.C: Government Printing Office. DOI: https://doi.org/10.4159/harvard.9780674592582.c4.

Griliches Z. 1971. "Introduction: Hedonic Price Indexes Revisited", In *Price Indexes and Quality Change*, edited by Z. Griliches: 3–15, Cambridge MA: Harvard University Press.

Herath, S. and G. Maier. 2010. *The hedonic price method in real estate and housing market research: a review of the literature*, SRE-Discussion Papers, 2010/03, WU Vienna University of Economics and Business, 31 March, Vienna, Austria. Available at: https://ro.uow.edu.au/cgi/viewcontent.cgi?article=1977&context=buspapers (accessed May 2020).

Hill, R.J. 2013. "Hedonic price indexes for residential housing: A survey, evaluation and taxonomy", *Journal of economic surveys*, 27(5): 879–914. DOI: https://doi.org/10.1111/j.1467-6419.2012.00731.x.

Istat. 2010. *La differenza nel livello dei prezzi al consumo tra i capoluoghi delle regioni italiane, Anno 2009*, Roma.

Istat. 2017. *Spese per consumi delle famiglie, Anno 2017*, Statistiche report 2018.

Istat. 2020. *Prezzi delle abitazioni, 1° trimestre 2020*, Statistiche Flash, Roma. Available at: https://www.istat.it/it/archivio/245204 (accessed July 2020).

Jolliffe, D. 2006. "Poverty, prices and place: How sensitive is the spatial distribution of poverty to cost of living adjustments?" *Economic Inquiry*, 44(2): 296–310. DOI: https://doi.org/10.1093/ei/cbj016.

Kokoski, M.F., P. Cardiff, and B.R. Moulton. 1994. *Interarea price indices for consumer goods and services: An hedonic approach using CPI data*, BLS working paper 256, U.S. Bureu of Labor Statistics, Washington D.C. DOI: https://doi.org/10.1016/0304-4076(94)01648-J.

Kokoski, M.F., Moulton, Moulton, B.R. Zieschang, K. 1999. "Interarea Price Comparisons for Heterogenous Goods and Several Levels of Commodity Aggregation". In *International and Interarea Comparisons of Income, Output and Prices*, edited,by A. Heston and R. Lipsey: 123–166, University of Chicago Press.

Lancaster K.J. 1966. "A New Approach to Consumer Theory", *Journal of Political Economy* 74-13257. DOI: https://doi.org/10.1086/259131-.

Laureti, T., and D.P. Rao. 2018. "Measuring Spatial Price Level Differences within a Country: Current Status and Future Developments", *Estudios de Economía Aplicada*, 36(1): 119–148. DOI: http://dx.doi.org/10.25115/eea.v36i1.2520.

Laureti, T., and F. Polidoro. 2018. "Big data and spatial comparisons of consumer prices Testing the use of scanner data for computing sub-national Purchasing Power Parities in Italy", In proceeding of 49th Scientific Meeting of the Italian Statistical Society, 20–22 June. Palermo, Italy. Available at: https://meetings3.sis-statistica.org/index.php/sis2018/50th/paper/view/1599/46 (accessed May 2020).

Leckie, G., and H. Goldstein. 2011. "A note on The limitations of school league tables to inform school choice". *Journal of the Royal Statistical Society*: Series A, 174(3): 833–836. Available at: https://www.jstor.org/stable/23013524 (accessed May 2020).

Lopez, E., and G.J.D. Hewings. 2018. "Housing price indices for small spatial units", *Regional Science and Urban Economics*. DOI: 10.1016/j.regsciurbeco.2018.02.005.

Malpezzi S. 2003. "Hedonic Pricing Models: A Selective and Applied Review." In *Housing Economics: Essays in Honor of Duncan Maclennan*, edited by A. O'Sullivan and K. Gibb: 67–89, Blackwell: Malder MA. DOI: https://doi.org/10.1002/9780470690680.ch5.

Marshall, E.C., and D.J. Spiegelhalter. 1998. "Reliability of league tables in in vitro fertilisation clinics: Retrospective analysis of live birth rates". *British Medical Journal*, 316: 1701–1704. DOI: https://doi.org/10.1136/bmj.316.7146.1701.

Martin, T., B. Aten, and E. Figueroa. 2013. "Estimating the Price of Rents in Regional Price Parities", BiblioGov. Available at: https://citeseerx.ist.psu.edu/viewdoc/download? doi = 10.1.1.440.3550&rep = rep1&type = pdf.

Moulton, B.R. 1995. "Interarea Indexes of the cost of shelter using hedonic quality adjustment Techniques", *Journal of Econometrics* 68: 181–204. DOI: https://doi.org/10.1016/0304-4076(94)01648-J.

Neto, E.D.A.L., and F.D.A. de Carvalho. 2008. "Centre and range method for fitting a linear regression model to symbolic interval data", *Computational Statistics and Data Analysis*, 52(3): 1500–1515. DOI: https://doi.org/10.1016/j.csda.2007.04.014.

Neto, E.D.A.L., and F.D.A. de Carvalho. 2010. "Constrained linear regression models for symbolic interval-valued variables", *Computational Statistics & Data Analysis*, 54(2): 333–347. DOI: https://doi.org/10.1016/j.csda.2009.08.010.

Ohta, M., and Z. Griliches. 1975. "Automobile prices revisited: Extensions of the hedonic price hypothesis" In *Household Production and Consumption, Studies in Income and Wealth*, edited by N.E. Terleckyi, 40. University of Chicago Press, for NBER, Chicago. Available at: http://www.nber.org/books/terl76-1 (accessed May 2020).

OMI (Osservatorio del mercato immobiliare). 2017. *Manuale della Banca Dati Quotazioni dell'Osservatorio del Mercato Immobiliare Istruzioni tecniche per la formazione della Banca Dati Quotazioni OMI.* Available at: https://www.agenziaentrate.gov.it/portale/documents/20143/265514/Il+manuale+della+banca+dati+OMI_-Manuale+2016_vers_2.0+modificata+in+data+02-01-2018.pdf/c2073cce-b58f-cc2d-4827-d52a8b5eac95 (accessed April 2020).

Ozanne, L., and T. Thibodeau. 1983. "Explaining metropolitan housing price differences", *Journal of Urban Economics* 13: 51–66. DOI: https://doi.org/10.1016/0094-1190(83)90045-1.

Parker, D. 2016. "International Valuation Standards. A Guide to the evaluation of Real Property Assets", Wiley Blackwell, London.

Primont, D.F., and M.F. Kokoski. 1990. *Comparing Prices across cities: A hedonic approach*, BLS working paper 204, U.S. Bureau of Labor Statistics, Washington, D.C. Available at: https://www.bls.gov/osmr/research-papers/1992/ec921210.htm (accessed May 2020).

Rao, P.D., and G. Hajargasht. 2016. "Stochastic approach to computation of purchasing power parities in the International Comparison Program (ICP)". *Journal of Econometrics*, 191(2): 414–425. DOI: https://doi.org/10.1016/j.jeconom.2015.12.012.

Ramsey, J.B. 1969. "Tests for specification errors in classical linear least-squares regression analysis". *Journal of the Royal Statistical Society: Series B (Methodological)*, 31(2): 350–371. DOI: https://doi.org/10.1111/j.2517-6161.1969.tb00796.x.

Renwick, T. 2009. *Alternative Geographical Adjustments of U.S. Poverty Thresholds: impact on State Poverty Rates*, U.S. Census Bureau. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi = 10.1.1.227.4438&rep = rep1&type = pdf (accessed May 2020).

Rosen S. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition", *Journal of Political Economy* 82(1): 34–55. DOI: https://doi.org/10.1086/260169.

Silver, M. 2011. "House Price Indices: Does Measurement Matter?" *World Economics* 12(3): 69–86. Available at: https://www.researchgate.net/profile/Mick-Silver/publication/227641359_House_Price_Indices_Does_Measurement_Matter/links/53eb9d620cf24f 241f129611/House-Price-Indices-Does-Measurement-Matter.pdf (accessed May 2020).

Zellner, A. 1963. "Estimators for seemingly unrelated regression equations: Some exact finite sample results", *Journal of the American Statistical Association* 58: 977–992. DOI: 10.1080/01621459.1963.10480681.

Received August 2020
Revised December 2020
Accepted April 2021

Journal of Official Statistics, Vol. 38, No. 1, 2022, pp. 83–106, http://dx.doi.org/10.2478/JOS-2022-0005

Unit Value Indexes for Exports – New Developments Using Administrative Trade Data

Don Fast[1], *Susan E. Fleck*[1], *and Dominic A. Smith*[1]

U.S. import and export price indexes replaced unit value indexes forty years ago, given quality concerns of mismeasurement due to unit value bias. The administrative trade data underlying the unit values have greatly improved since that time. The transaction records are now more detailed, available electronically, and compiled monthly with little delay. The data are used by academic researchers to calculate price measures, and unit value indexes based on trade data are used by other national statistical offices (NSOs). The U.S. Bureau of Labor Statistics is now evaluating whether replacing price indexes with unit value indexes for homogeneous products calculated from administrative trade data could expand the number of published official import and export price indexes. Using export transactions, the research calculates detailed unit value indexes from 200+ million trade records from 2012–2017 for 123 export product categories. Results show that 27 of the 123 unit value indexes are homogeneous and closely comparable to published official price indexes. This article presents the concepts and methods considered to calculate and evaluate the unit value indexes and to select the product categories that are homogeneous. Compared to official price indexes, export unit value indexes for the 27 5-digit BEA (U.S. Bureau of Economic Analysis) end-use product categories would deflate real exports of these goods by 13 percentage points less over the period. Incorporating these 27 indexes into the top-level XPI would increase the value of real exports of all merchandise goods by 2.6 percentage points at the end of 2017.

Key words: Unit values; trade; large data sets.

1. Introduction

The U.S. Bureau of Labor Statistics (BLS) official Import and Export Price Indexes (MXPI) measure price changes of U.S. imports and exports of goods and a limited number of services. Other national statistics that calculate trade and output depend on the quality and detail of MXPI to adjust current-dollar measures to constant dollars (Cerritos 2015; Moulton 2018). Over time, the number of publishable detailed MXPI has declined as budget-related sample size reductions shrink the number of items in the market basket and the number of prices that support index quality. Nearly half of the import and export price

[1] U.S. Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Rm. 3955 Washington, DC 20212, U.S.A. Emails: fast.don@bls.gov, fleck.susan@bls.gov and smith.dominic@bls.gov
Acknowledgments: We recognize and thank these BLS staff for their research and data support for this paper: Brett Edgerton, Matthew Fisher, Michael Havlin, Laurence Lang, Christina Qiu, Daryl Slusher. Additional support was received from Antonio Caraballo, Ara Khatchadourian, Aric Schneider, and Praveenkumar Yerramareddy. Use of the export trade data are subject to Agreement No. 2067-2018-001, Memorandum of Understanding (MOU) between the U.S. Census Bureau and the Bureau of Labor Statistics (BLS). The BLS has received prior approval from the U.S. Census Bureau, which affirms that the research results do not present disclosure risks and approve the publication of the results. All results have been reviewed to ensure that no confidential information is disclosed.

© Statistics Sweden

indexes comprising detailed BEA end-use product categories for merchandise goods did not meet publication criteria in 2020. Furthermore, data collection has become more difficult as business respondents are less willing to participate in the voluntary business survey. Challenges to collecting data from sampled businesses are compounded by social distancing recommendations based on public health guidelines.

The BLS is considering an alternative data source to calculate monthly import and export price changes and is piloting the approach with export transactions. The data source is the official administrative trade data set collected by the U.S. Customs and Border Protection agency that are cleaned and edited by the U.S. Census Bureau for statistical purposes. This data set comprises nearly all importers' and exporters' self-reported detailed shipment records by the international Harmonized System (HS) product classification for the United States.

Unit values have been and are used in both import and export price indexes for the most homogeneous products in which the law of one-price holds, such as commodities traded on a worldwide exchange. For example, the U.S. import price index for crude petroleum is currently calculated using unit values derived from the U.S. Department of Energy (DOE) petroleum transaction import records. The DOE administrative data source is more reliable than survey data in the face of low company response rates and the price volatility of this heavily traded product.

However, using unit values, in general, has been criticized for quality concerns. Even though values and quantities are available in the administrative trade data set and can be used to calculate average prices and unit value indexes (UVIs), previous research has shown that unit value indexes often differ significantly from the price indexes that they are meant to approximate (Alterman 1991; Bradley 2005; Silver 2009). The source of concern is that differences in the composition and characteristics of transactions comprising a unit value can lead to movement within and between unit values that are unrelated to price movements and thus can mismeasure prices. Nonetheless, other research describes how UVIs can and even should be used in place of traditional methods when homogeneous items can be defined and when units of measure are consistent (Balk 1998; Dalton and Fissel 2018; Diewert and Von der Lippe 2010; Silver 2010, 2011).

This research identifies product categories where unit value indexes (UVIs) can potentially replace price indexes for U.S. exports. The homogeneity of product categories is ranked using a coefficient of variation test, which identifies items with less price dispersion. We then calculate UVIs for all product categories and show that UVIs for the more homogeneous product categories are typically consistent with the comparable official export price index (XPI). A subset of UVIs are identified as being consistent with official XPIs.

We compare official export price indexes against all UVIs for 123 5-digit BEA end-use product categories. The approach identifies a fifth of the UVIs, accounting for USD407 billion in trade (2015), to be of sufficient quality to replace published or augment otherwise unpublished price indexes, and another fourth, accounting for another USD228 billion in trade (2015), that are potentially useable but are of indeterminate quality. On the other hand, slightly more than one half of the UVIs, accounting for USD525 billion in trade (2015), were of poor quality showing a wide range of price variability and unit value bias common among heterogeneous goods.

We find that replacing current sources with administrative data for the 27 good UVIs would have a moderate impact on XPI between January 2012 and December 2017. The export indexes created with administrative data would deflate real exports by 13 percentage points less at the end of 2017, compared with current sources. Incorporating these 27 indexes into the top-level goods XPI would increase the value of real exports by 2.6 percentage points at the end of 2017.

UVIs' current prices and quantities potentially improves the current matched model approach that uses a fixed-basket of weights that is known to overestimate price changes. Moulton (2018) describes multiples biases in the MXPI-including lower-level sourcing substitution bias and new product bias-that result from insufficient observations and lagged weights, issues that can be mitigated using the administrative data. Price experts describe how unit values could actually mitigate outlet and sourcing substitution bias (Nakamura et al. 2015, 43–45). In reviewing Alterman's (1991) critique of unit value indexes Nakamura et al. (2015) re-evaluate Alterman's conclusions. They hypothesize that the slower increase of the UVI compared to official MPI that Alterman observed for some product categories could have been an appropriate adjustment for sourcing substitution bias, rather than a mismeasurement of prices (Nakamura et al. 2015, 52). Hottman and Monarch's (2018) research validates Nakamura et al. (2015). When an all-goods import unit value index is calculated with the administrative trade data it shows a flatter price trend than the modified Laspeyres formula used in the official import price index, at a rate of about 2% yearly from 1998 to 2014. Reinsdorf and Yuskavage (2018) also calculate unit values to evaluate sourcing substitution bias of MPIs. All of the research described above supports the use of unit values to mitigate substitution bias and provides impetus to our research to calculate UVIs to replace official import and export price indexes in homogeneous product areas.

Conceptually the 'unit value indexes' in our research differ from the U.S. Census' unit value indexes criticized by Alterman, and the unit values are more narrowly defined, approximating Diewert and Von der Lippe's (2010) classifications of 'reasonably' homogeneous products. While traditional price collection methods in the international price survey typically record a dozen prices per importer or exporter for a few thousand companies, the U.S. electronic trade transaction data set contains tens of thousands of transactions for detailed product classifications by company.

Section 2 first describes the constraints and new frontiers faced when blending alternative data, and particularly the administrative trade data, into official statistics. Section 3 describes the coefficient of variation test for price dispersion. Section 4 describes the building blocks of the methodology. Section 5 describes how the quality of the UVIs was evaluated. Section 6 describes the results and how replacing official XPI with UVIs would affect measures of real exports. Section 7 concludes with a discussion of the steps to operationalize the new approach and potential future work.

2. Requirements for Integrating Alternative Data into Official Statistics

Consideration of detailed U.S. trade transaction data to measure import and export price indexes is part of an effort by U.S. Bureau of Labor Statistics' official price statistics programs to actively pursue alternative data to support the following operational objectives-measure price change more accurately, improve BLS management of respondent burden, expand item and geographic coverage, publish new products, and achieve cost savings. To move forward with alternative data, the new source is expected to meet the core measurement objective of the target population by being timely, accurate, and representative, as well as being consistent with legal and data directives. The methodology applied to the new data source should be comparable or complementary to the methodology used for the official price index. Such methods would address quality change, new and replacement products, sample representativeness, and minimum publishability standards to ensure representativeness and protect against respondent identification. Alternative data sources for price indexes would mitigate loss as we maintain, improve, or expand the components of the relevant market basket. These components include replacement or expansion of current products, introduction of new products, weights for integrating the data into the larger market basket, or more frequent prices.

The completeness of the U.S. electronic administrative trade data set and new constraints on data collection with social distancing in place provide opportunity and impetus, respectively, to set up statistical production in a way that blends unit values from the administrative trade data with the directly collected data, and is similar to other efforts to blend data sources (Reid et al. 2017). This section describes the constraints and new frontiers to meet these statistical obligations, and looks at how limiting price dispersion likely improves homogeneity.

2.1. Constraints on Using Administrative Trade Data

The alternative data proposed to replace or augment direct data collection from businesses must meet core measurement objectives of the target population by being timely, accurate, and representative and must use a methodology that should be comparable or complementary to the methodology used for the official price statistics. These requirements act as constraints or limitations on the options we have to compute unit values from the administrative trade data. The high priority constraints described here relate to (1) timeliness and availability of data, (2) accuracy and representativeness of homogeneous unit values, (3) consistency of published data for data users, (4) ability to aggregate lower level UVIs to upper level published indexes.

First, the administrative trade data must meet timeliness and availability constraints to fit the production schedule of data processing. Sufficient data must be received at the beginning of each month for the previous month's data to assure timely and accurate publication of the preliminary price indexes two weeks after month's end. The first publication of the preliminary price index is updated with additional data in the subsequent three months, but the preliminary index must meet quality standards of representativeness. Collaboration with other government agencies is ongoing to evaluate how much data can be made available in time to meet current publication deadlines. Research includes assessing the quality difference between preliminary and final indexes based on partial and full month trade data, respectively. This work is ongoing and is not described in detail here.

Second, the accuracy and representativeness of unit value indexes rests on the homogeneity of the items and their unit values. We find that there is such variability of product categories that no one item key, or group of transaction characteristics, could define homogeneity across all product categories. Each product category has a different mix of goods, and different patterns of trade. The need to mitigate unit value bias while maintaining item continuity leads us to use the coefficient of variation to evaluate price variability as a key indicator of homogeneity, or at the minimum, of heterogeneity.

The common definition of unit value bias is the compositional effect on indexes when a unit value, or average price, is not adequately narrowly defined to represent one item or product. The choice of index framework will be discussed in the methodology section. As the Export and Import Price Index Manual advises, direct data collection is preferable to unit values based on Customs data (International Monetary Fund 2009). Yet of the six reasons to prefer price survey indexes over unit value indexes, only three pertain to the BLS MXPI-biases due to changes in the mix of the heterogeneous items in customs transaction records, poor quality of recorded data on quantities, and infrequent trade of 'unique' goods, such as ships and large machinery." (International Monetary Fund 2009, xiv). Of these three, the first is the issue that is addressed in our research. The second issue of missing quantities affects only 10% of U.S. trade transactions, providing millions of transactions with quantities and total value, compared to the 45% response rates for exports in the international price survey reported in July 2020 (U.S. Bureau of Labor Statistics 2020a); and the last concern regarding unique high priced products is not adequately addressed in price surveys either, given that sampling and pricing methods exclude infrequent trade of unique items. The compositional differences of product codes in scanner data have been widely discussed for consumer prices, and we borrow from this research to consider how U.S. trade transaction records could measure import and export price changes.

The literature on scanner data informs our approach to defining homogeneous items. There are differences in the way to bundle transactions to calculate unit values for trade data, which Nakamura et al. (2015, 56) think bodes well for combining producer products. Given the production- and contract-driven aspect of international trade, transactions with shared characteristics will likely be tailored to one supplier or market and will represent larger dollar amounts than consumer purchases. The broader homogeneity is also discussed by Diewert and von der Lippe when they describe the tradeoff between homogeneity and continuity. That is, the narrower the classification of the items, the more likely prices and quantities will be zero across periods given the sporadic nature of shipments (Diewert and Von der Lippe 2010, 20). This tradeoff is described mathematically by Chessa (2019), who evaluates scanner data for the Dutch CPI. A homogeneous product is defined by balancing two measures: one measure quantifies the homogeneity of product bar codes and other attributes, while the second measure expresses the degree to which products can be matched each month to ascertain continuity of trade and minimize item churn (Chessa 2019).

Because there is such fluidity between all aspects that define homogeneity – price dispersion, number of grouped characteristics to define an item, and item continuity, we depend on a qualitative and quantitative comparison of the UVIs with the official price indexes to evaluate whether a UVI is equivalent enough to the official price index to replace it.

The third constraint–that the administrative trade data should assure consistency of published data for data users–goes hand in glove with the accuracy and representativeness of both UVIs and the official MXPI. The official import and export price indexes are based on a methodical and statistically sound approach to sample, interview, and collect data from U.S. importers and exporters. As a major economic indicator, all aspects of the survey process are scrutinized with quality performance measures on a quarterly basis. Nearly a thousand price indexes are systemically reviewed for quality and publishability as part of monthly publication, and variance statistics assure representativeness.

Since most of the import and export BEA end-use price indexes are used to deflate net trade in the U.S. GDP, any major difference would need to be explained clearly as breaks in series could affect the usefulness of the indexes. While it is expected that known upward biases of lower level substitution should be dampened with the UVIs, any differences that cannot be explained would reduce the likelihood of acceptance of the new indicator. The trust of data users in the current data product limits us to refine and define homogeneity in a way that assures as much continuity as feasible with our directly collected MXPI. It is not ex-ante clear that unit value indexes are always worse, given that official price indexes may be based on a small number of price quotes and are constructed using less desirable index formulas. Matching directly collected data both gives us confidence in the quality of the UVIs and lets us publish a consistent time series when incorporating UVIs in XPI.

Finally, for consistency of methodology, the data source must be integrated into the directly collected survey data systemically. Direct data collection begins with elementary level items whose prices are reported by the respondent for each sampled product the company recently traded. So too can transaction prices be averaged for individual items, aggregated at the classification group level and then aggregated to the upper level index.

Each transaction in the data set is a shipping record that is primarily categorized by the 10-digit HS classification and includes a few dozen other transaction characteristics. We select a subset of these transaction characteristics, including the HS classification, as an item key. Each unique combination of values for the selected characteristic constitutes an item. We establish four item keys, each of which uses a different subset of characteristics to create different items and unit values. The most broadly defined item key is for the 10-digit HS classification, which produces one item and one unit value composed of all transactions for that product classification. To specify the other three item keys, we heed the Export and Import Price Manual to select price-determining characteristics of export items (International Monetary Fund 2009). These include the foreign destination, U.S. point of origin, domestic producer (U.S. principle party of interest), and other characteristics, like unit of measure and related vs arms-length trade. The variables and their abbreviations used in the figures are domestic/re-export goods (F), employer ID (E), state of origin (S), zip code of export (Z), country of destination (C), U.S. Port (D), quantity units (Q), related trade indicator (R), and 10-digit Harmonized system classification (H). Together these variables provide significant detail about trading relationships, but they provide less detail about the good being traded than would be available in the XPI survey. We use this information to define items at a level of detail that was not studied in previous applications. This approach is consistent with the suggestion of Von der Lippe and Mehrhoff (2010, 7) for improving unit values in German price indexes.

Final price indexes are created through multiple layers of aggregation. Items are combined to form classification group indexes, which are similar to HS classifications. The classification group provides continuity when HS classifications change and itself is mapped through product and industry concordances to the three different classification systems for which the MXPI are published-product indexes based on the Harmonized System and BEA end-use, and industry indexes based on the NAICS. Upper level indexes are aggregated from classification group indexes using a Lowe formula, which the BLS refers to as a modified Laspeyres formula. The trade weights used in this aggregation are updated once yearly, lagging the index by two years. A full explanation of the different classification schemes, the upper level index formula, and the trade weights used by the MXPI is available in the International Price Program Handbook of Methods (U.S. Bureau of Labor Statistics 2020b).

2.2. New Frontier of Administrative Trade Data

From constraints to frontiers, there are potential improvements to be gained from using alternative data sources. The primary reason for revisiting unit value indexes is to reinstate publication of detailed indexes that no longer meet quality standards due to insufficient prices or company coverage. The administrative trade data provides much better representativeness of trade in homogeneous goods, as well as a volume of transactions that dwarfs in-person data collection. Furthermore, the cost savings realized when administrative data replace direct data collection would be used to expand the currently thin coverage of service import and export indexes.

The data source can also mitigate biases that are difficult to address with existing practices. Access to current prices and quantities provides an opportunity to account for lower level substitution bias as described by Moulton, which for export price indexes can be described as producer-side product replacement bias, in which "price changes that occur at the time of product replacements tend to be dropped," (Nakamura and Steinsson 2012, 3278).

As Nakamura et al. (2015) show in appendix 2b solutions to export product substitution and import sourcing substitution bias both necessarily involve averaging unit prices for different products. They, like we, look to balance homogeneity with substitution, to define an item that is not so broad as to have unit value bias, but not so narrow that price and quantity changes of two similar items are not reflected in the index. We consider groupings of items with broadly and narrowly defined item keys. We find that there is no one item key for all product categories for which the price index best matches the official price index.

3. First Steps to Evaluate Price Dispersion and Homogeneity – Coefficient of Variation

Similar to the proof of concept for unit values described in Fast and Fleck (2022) we propose using a coefficient of variation test to select homogeneous items as a first step to evaluate UVIs and the item keys underlying them for unit value bias. The coefficient of variation test identifies whether within-month price variation of items is small relative to the mean price of that item, which we take as an indicator that the transactions for each item involve relatively homogeneous goods. The coefficient of variation is defined for an item as the standard deviation of the unit value of an item within a month divided by the mean unit value of the item within a month multiplied by 100. We calculate the coefficient of variation for each item's unit value with multiple transactions within a month. A small coefficient of variation for an item within a month suggests that the physical goods traded in those transactions are similar. Thus, the transactions comprising the item are more likely to represent the same product. Calculating the coefficient of variation within each month minimizes the role of inflation in causing variation across transactions within an item, but it also understates variation that may be due to quality changes over time.

We calculate the coefficient of variation for each item in each 5-digit BEA end-use product category and calculate the distribution of these coefficients across items by category on a weighted basis for each month. Then, we plot the average of these monthly distributions. Figure 1 plots the cumulative distribution function (CDF) of the distribution of item level coefficient of variations by product category. More homogeneous product categories have many items with small coefficients of variation, indicating less price variability within a month, which results in a CDF that is above the other lines.

In some cases the lines corresponding to different product categories cross indicating that one product category has many items that have a small coefficient of variation, but also some items that are more variable. This makes it difficult to strictly rank the product categories according to their homogeneity. In the next section, we construct indexes for each product category and compare them to XPI indexes as a benchmark. We find that UVIs with CDFs that are generally above the CDF of the seasonal product category 'vegetables, vegetable preparations, and juices' almost always are close to the corresponding XPI index. Thus, we will refer to any product category with a CDF above that of 'vegetables, vegetable preparations, and juices' as homogeneous.

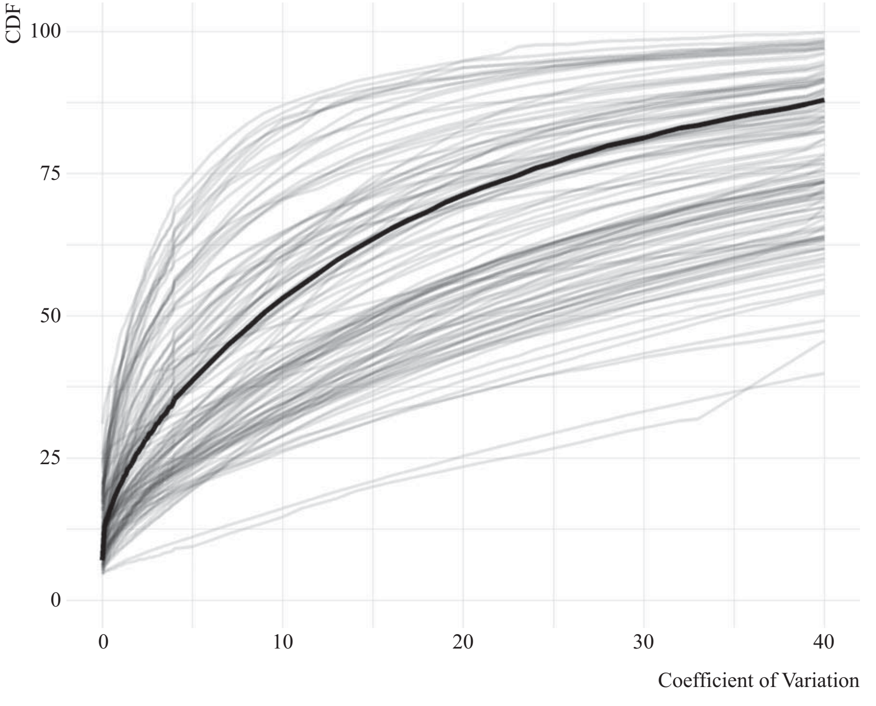


Fig. 1. CDF of Item Coefficient of Variation for 123 BEA end-use product categories. A lower C.V. value indicates less price dispersion, and the steeper and more concave lines represent product categories with more items that have less price dispersion. The dark line is the CDF for 'vegetables, vegetable preparations, and juices.' CDFs are calculated for individual months and then averaged across months. The x-axis is truncated at 40 for clarity.

Our results suggest that the coefficient of variation test could be used to identify indexes with well-defined items for unit values if a published index benchmark is not available. However, establishing precise thresholds may be challenging because the expected amount of variation of homogeneous items may vary across settings. Gopinath and Itshoki (2010) and Gopinath and Rigobon (2008) demonstrate that homogeneous goods experience both more frequent and larger price changes than differentiated goods in their analysis of the directly collected import and export price microdata. For our relative threshold, we choose a detailed item key for vegetables, which results in a relatively flat but concave cumulative distribution. Vegetable export prices are widely variable, even at detailed item levels. Both supply and demand are variable, as yield, quality, and availability depend on growing seasons and weather locally and worldwide. There is also significant competition in the sector; 2015–2016 records report more than 4,000 exporters monthly, and nearly 1,000 large regular exporters (Fast and Fleck 2019). Missing prices are filled in by imputation, and this can also cause variability. The combination of competition, short horizon from field to market, and imputation results in price variability, even though many of the 161 HS product categories are of nondifferentiated products, like light red kidney beans, or certified organic asparagus. The index encompasses nearly all conditions of variability that indexes face. There are 36 product categories that have CDFs that are generally above that of vegetables. We will refer to these product categories as homogeneous. There are an additional 18 product categories with CDFs that are similar to that of vegetables.

Our primary evidence of unit value bias comes from comparing UVIs to indexes created using directly collected data. The different data sources are used to create indexes that are blended using a modified Laspeyres formula to calculate overall U.S. Import and Export Price Indexes. This approach to blending data sources is currently in place for imports of crude petroleum and exports of grain, sourced from the Energy Information Agency and Department of Agriculture, respectively.

4. Computing Unit Value Indexes

This section describes our approach to constructing UVIs. The construction takes place in three steps. First, unit values for each trade transaction are computed from value and quantity. Second, unit values for transactions are combined to form item unit values using a weighted geometric mean. Items are defined by shared characteristics of transactions. Third, the price relatives of these items are aggregated to a classification group using a Tornqvist index formula. Finally, the classification group indexes are aggregated to create 123 5-digit BEA end-use indexes using a modified Laspeyres formula. We consider four different "item keys" which are sets of characteristics. Each set of items related to these item keys results in different unit values and UVIs. Thus, there are four options for each 5-digit BEA end-use index.

4.1. Calculating Unit Values

The administrative data contain information on the value of shipments and the associated quantity that can be used to form unit values. We calculate the unit value for a transaction by dividing the dollar value by the quantity. The unit value P of a transaction s for any given month t is the value of the transaction V divided by the quantity of units Q in the transaction,

$$P_{s,t} = \frac{V_{s,t}}{Q_{s,t}}. \tag{1}$$

The transactions, S_k, associated with some item k, defined as a unique combination of characteristics, are aggregated using a weighted geometric mean to form a unit value for each item. The unit value of an item k at some month t is

$$P_{k,t} = \left[\prod_{s \in S_k} P_{s,t}^{V_{s,t}}\right]^{\frac{1}{\sum_{j \in S_k} V_{j,t}}}. \tag{2}$$

4.2. Classification Group Unit Value Indexes

The items are then aggregated to the classification group using a Törnqvist formula. The unit value index for classification group c is calculated by aggregating the unit value indexes for the set of items K_c that belong to classification group c. The weights, $w_{k,t} = \frac{V_{k,t}}{\sum_{j \in K_c} V_{j,t}}$, are created using the value of trade associated with an item in some month, $V_{k,t} = \sum_{s \in S_k} V_{s,t}$.

$$R_{c,t} = \prod_{k \in K_c} \left[\frac{P_{k,t}}{P_{k,t-1}}\right]^{\frac{W_{k,t-1} + W_{k,t}}{2}} \tag{3}$$

Equation (3) describes the month-to-month changes in the classification group UVI. This is converted to an index level by starting the index at 100 in January of 2012 and advancing it using R_{ct}. Therefore, the index level for a classification group in a month is

$$P_{c,t} = P_{c,t-1} R_{c,t} \tag{4}$$

with $P_{c,0} = 100$.

4.3. Forming 5-Digit BEA End Use Indexes

For this article we focus on aggregating the indexes to the 5-digit BEA end-use product category because that is the level at which decisions about whether to use unit value indexes for a given set of classification group indexes will be made. Eventually the classification group unit value indexes will be used as the building blocks for HS, BEA end-use, and NAICS based indexes. These indexes will be combined with price indexes from directly collected data using a modified Laspeyres index formula. The weights used to aggregate from classification group to BEA end-use product category are lagged values for consistency with directly collected data. The weights used in the MXPI are lagged two years due to delays in the availability of real-time weights. The formula for the unit value index for BEA end-use product category e is

$$R_{e,t} = \sum_{c \in C_e} \frac{V_{c,b}}{\sum_{i \in C_e} V_{i,b}} \frac{P_{c,t}}{P_{c,t-1}} \tag{5}$$

This is converted to an index level by starting the index at 100 in January of 2012 and advancing it using $R_{e,t}$. Therefore, the index level for a product category in a month is

$$P_{e,t} = P_{e,t-1} R_{e,t} \tag{6}$$

with $P_{e,0} = 100$.

4.4. *Missing and New Items*

Our approach to missing and new items is similar to the approach used in the official MXPI. When an item is missing, we impute a price for the item using cell mean imputation for up to three months before discontinuing the series. For directly collected data, there are attempts to replace discontinued items and contact respondents for missing prices, whereas for administrative data a missing price for an item indicates that no trade occurred. An item naturally occurs as trade transactions begin, and we pre-impute a price in the month before the item enters using cell mean imputation.

4.5. *Possibilities for Improvement and Caveats*

There are a number of ways in which unit value indexes with administrative data may improve on the official XPI. The administrative data eliminate issues with representativeness of trade and response rate concerns. The data also allow us to address changes in trade in a timelier manner.

The administrative data could potentially expand the number of detailed published indexes, because two major hurdles with directly collected data are avoided. Representativeness of trade and nonresponse are no longer problems when transaction records comprise the universe of trade. Administrative trade data permit detailed indexes to be published at whatever level that trade occurs, while the directly collected data are based on a voluntary survey. Indexes must meet quality review standards of sample representativeness and index quality. The number of sample units is distributed across all merchandise goods categories by their trade weight to assure representativeness for the aggregate MXPI, but this sacrifices more detailed index representativeness. Even though item prices are collected, detailed indexes are not representative of the trade if the trade dollar value is below the cut-off. For product categories meeting the trade dollar cut-off, nonresponse negatively affects publishability as measures for representativeness of companies, number of companies, and number of items are all used as publishability standards. Nonresponse is a nonissue for administrative data.

Using current values and quantities with administrative data allows for the incorporation of new goods and replacement of substitutes. New and exiting goods are accurately counted because the complete data set accounts for all current trade. Furthermore, substitution with classification groups can be addressed because the availability of current trade values allows us to use a superlative index formula.

5. Comparison of UVIs to XPI Benchmarks

After calculating UVIs for each item key and product category, we must determine which are of sufficient quality to replace the product category's XPI counterpart. Our first criteria is agreement with the relevant XPI when the XPI is of sufficient quality to be used as a benchmark. We evaluate agreement for each of the four item keys under consideration and manually choose the best key for each product category and rate the quality of the fit. This analysis used graphs of index levels, month-to-month changes, and statistics measuring the quality of the fit. Then, we evaluate the performance of this "best" item key against XPI using both the manual ratings and multiple statistics that measure the quality of fit. Our focus in evaluating the quality of fit is on long-term agreement between the two index levels. Finally, we discuss what we do when the official XPI may not be a reliable benchmark.

5.1. Identifying "Best" Item Keys and Visual Analysis

Our first step of determining which unit value indexes were potentially able to replace official price indexes was to have research team members manually review each index. Each member studied each of the BEA end-use product categories individually using graphs of index levels, month-to-month changes, and statistics comparing the fit of each index. They identified the item key that was the best fit and rated the fit of that key as either "good", "undecided", or "poor". Examples of each type of index are presented in Figures 2, 3, and 4. In each figure, we show the index levels and month-to-month changes for each item key and the XPI benchmark. All three examples involve published XPI indexes that are considered to be high quality. Figure 2 is an example of a "good" index. The best item key is HECQR, a key with a medium amount of detail. The UVI tracks the XPI index very closely. The UVI generated using just information on HS product codes, the H key, has large spikes and performs much worse. Figure 3 shows an "undecided" index. The best item key is H. In this case, the unit value index is less close to the XPI benchmark. There are multiple times where the UVI increases significantly more than the XPI benchmark, but the levels are generally close until the final year. In this case, there is some hope that with further improvements to our methodology, such as using a different item key or handling outliers different, a UVI could be used. Figure 4 shows a "poor" index where there is little reason to believe a unit value index could be used for this BEA end-use product category. The more detailed keys generally perform better, but none of the keys performs very well.

Table 1 shows the rating of the indexes and groups the number of BEA end-use product categories by rating and publication status of official XPI. When there was disagreement between team members, the more common rating was used. Disagreements were relatively rare and only involved whether an index was "good" or "undecided" or whether an index was "undecided" or "poor".

Manually reviewing the indexes allows us to consider multiple pieces of information simultaneously when rating indexes, but the process is time consuming as we evaluate the impact of alternative methodologies. Additionally, manual ratings can vary across individuals and time. In the following subsection, we describe our approach to develop acceptance rules based on statistical comparisons. We search for a set of acceptance

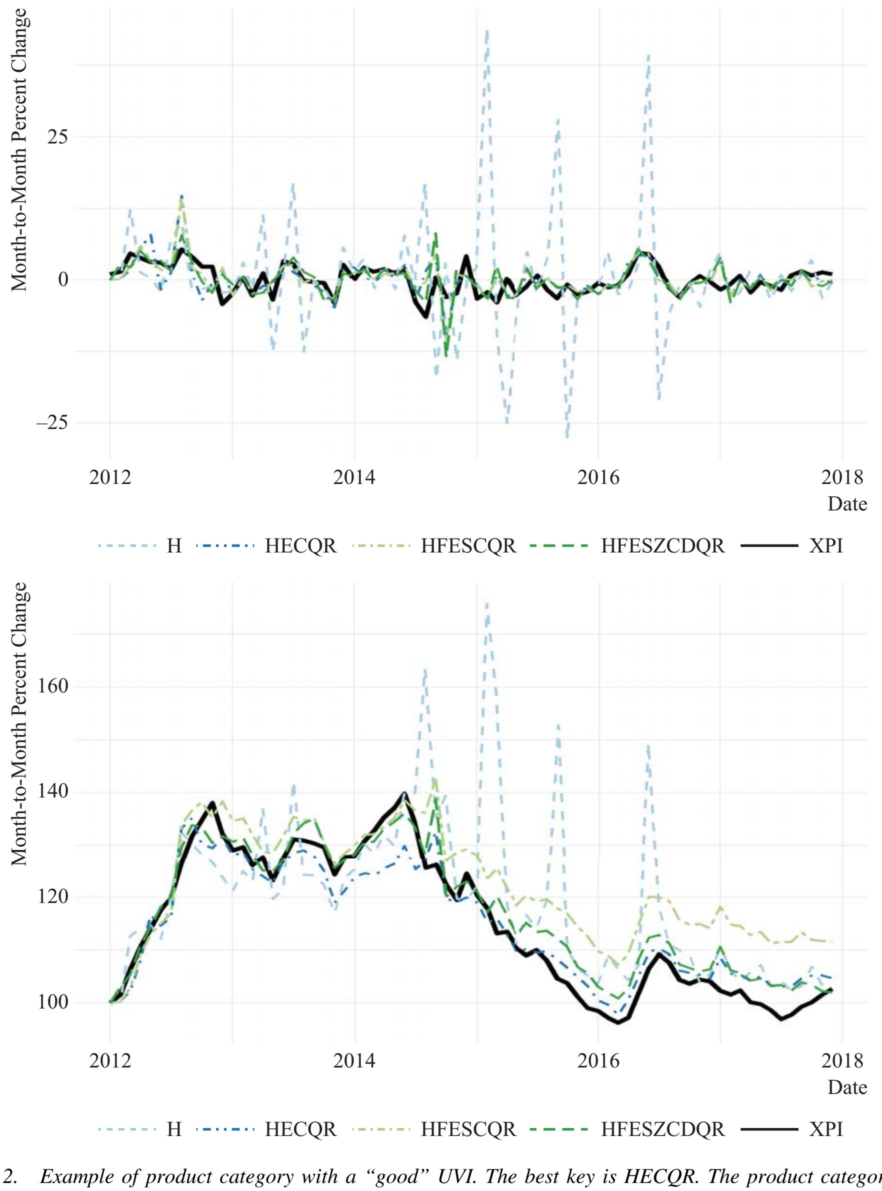


Fig. 2. Example of product category with a "good" UVI. The best key is HECQR. The product category is "Other animal feeds, not elsewhere classified." The top panel depicts month-to-month percent changes. The bottom panel depicts index levels.

criteria that can be applied to our data. The goal of these criteria is to approximate the manual review. We focus on a set of criteria that can differentiate between "good" and "poor" indexes.

5.2. Statistical Comparison

We considered multiple statistics to create a set of criteria to rate unit value indexes. The criteria use information on both index levels and month-to-month changes. From the initial

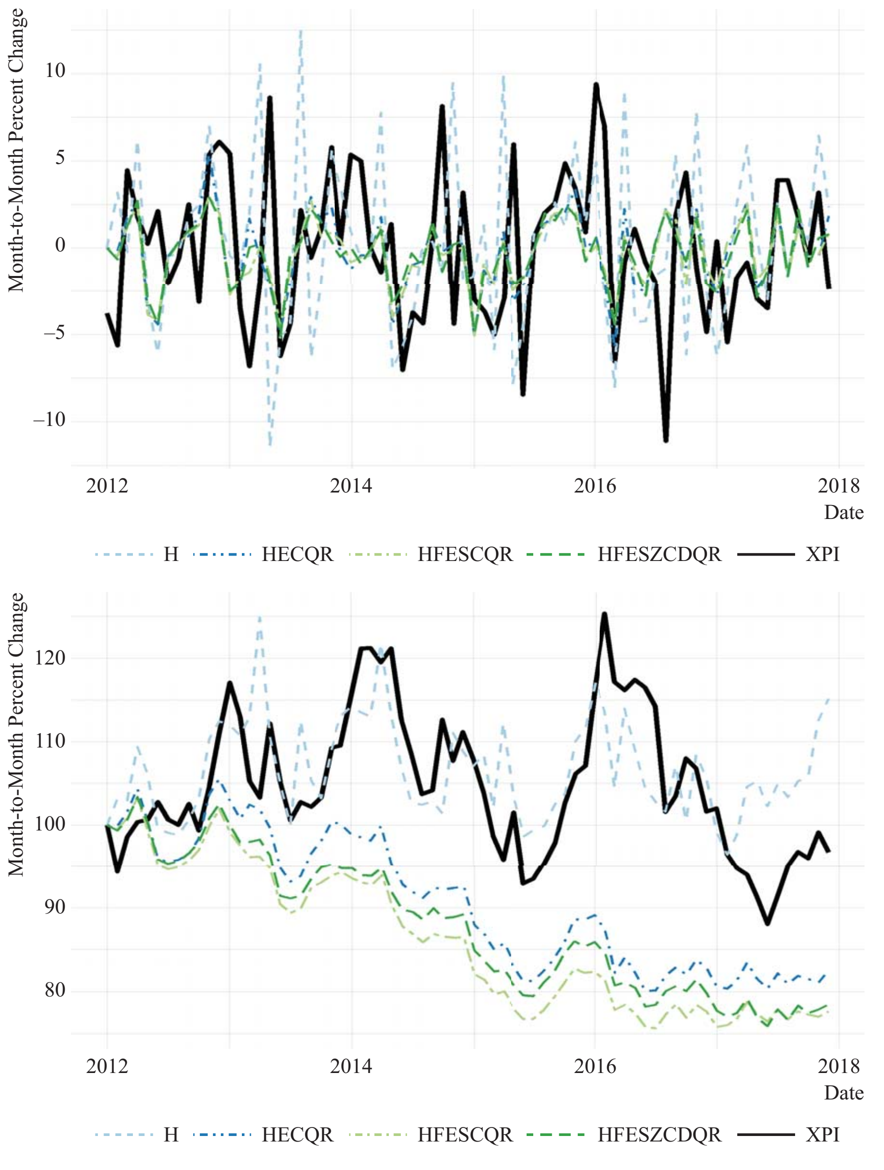


Fig. 3. Example of a product category with an "undecided" UVI. The best key is H. The product category is "Fruit and fruit preparations, including fruit juices." The top panel depicts month-tomonth percent changes. The bottom panel depicts index levels.

set of statistics we identified three that jointly had significant predictive power for whether an index would be classified as "good" by the manual review. We describe the three statistics that had significant predictive power and what values of these statistics generally indicated a good index.

Previous work has used a wide range of statistics to evaluate whether price indexes generated with alternative data are similar to benchmark indexes, but there is significant disagreement about which tests should be used and what the appropriate thresholds are (Fitzgerald and Shoemaker 2013). The appropriate threshold will depend on the needs of

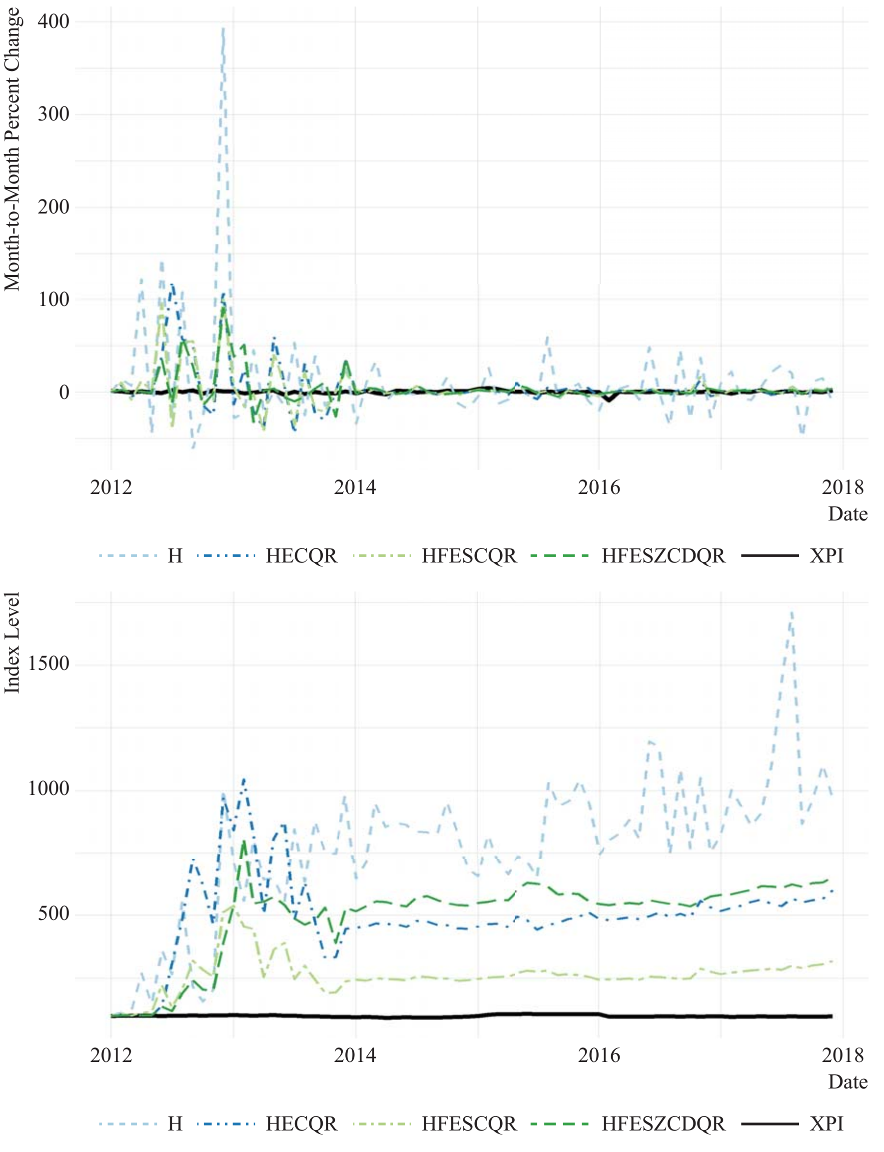


Fig. 4. Example of a product category with a "poor" UVI. The best key is HFESCQR, but that is not a close approximation to the official XPI. The product category is "Industrial inorganic chemicals." The top panel depicts month-to-month prercent changes. The bottom panel depicts index levels.

data users and potential benefits of replacing official indexes with alternative data. It will also depend on whether it is most important for the UVIs and benchmark to have similar month-to-month changes or longterm trends. Certain statistics, such as the root mean squared error between the month-to-month changes in indexes, can indicate agreement over the short term, but lead to large long-term differences if the monthly misses are all in a certain direction.

We identified three statistics that together seemed to have the most predictive power. They are the fifth largest absolute difference between index levels, the intercept of a

Table 1. Number of BEA end-use product categories receiving a rating by published status.

Rating	Published	Unpublished	Total
Good	15	12	27
Undecided	11	21	32
Poor	24	40	64

Notes: Numbers are the count of BEA end-use product categories manually assigned a rating. Columns differentiate between product categories that are published and those that are not.

regression of index levels of the unit value index on XPI price index levels, and a t-test of whether the month- to-month changes have the same mean. For each statistic, we suggest an approximate cutoff value to differentiate the "good" and "poor" UVIs. These cutoff values were chosen jointly as the values such that satisfying all three cutoffs indicated a "good" UVI and failing one or more indicated a "poor" UVI. We experimented with multiple combinations of statistics and with treating the "undecided" UVIs as either "good" or "poor". The set of three statistics proposed here and their associated cutoffs represent the choices that consistently performed well at minimizing misclassification of "good" and "poor" UVIs. They should be treated as guidelines to quickly evaluate the quality of generated indexes before a more rigorous review.

The fifth largest absolute difference tests whether index values stay close to each other over time. Using the fifth largest difference allows for temporary differences between indexes if the indexes come back together before five months have passed. We find there are cases where indexes have large differences for a month or two, but otherwise match well. Therefore, we prefer the fifth largest difference over a test such as the maximum difference. We also considered using the percentage difference accounts for differences in index levels due to different growth rates across BEA end-use product categories, but the % difference is not symmetric. We find that a fifth largest absolute difference of less than 22 index points is a necessary condition for a unit value index to receive a "good" rating.

The second test comes from running a regression of the index values of the unit value index on the index value of the XPI. If the values of the two indexes were always in agreement the estimated parameters would be an intercept of zero and a slope of one. These tests are similar to testing for a strong positive correlation between the two series. However, this test avoids cases where series are correlated, but diverge because in those cases either the slope or intercept will be far from its target value. We have generally found that an estimated intercept of less than 50 is indicative of a "good" UVI. Recall that the indexes are 100 in their base period which helps explain the rather large cutoff. Figure 5 illustrates this test using the "good" index from Figure 2. The points are generally near the 45-degree line, but the UVI tends to exceed the XPI index level when both levels are low. Then, when index levels have grown, the UVI is consistently less than the XPI index level.

Table 2 The third test is a paired t-test. The test checks whether the means of the month-to-month changes of the two indexes are different. The null hypothesis of this test is that the month-to- month changes from the two indexes come from the same distribution. We use the p-value of the paired t-test as the value we track. A large p-value indicates that the null hypothesis cannot be rejected which is the goal of this test. We find that a p-value greater than 0.5 is typically indicative of a "good" UVI.

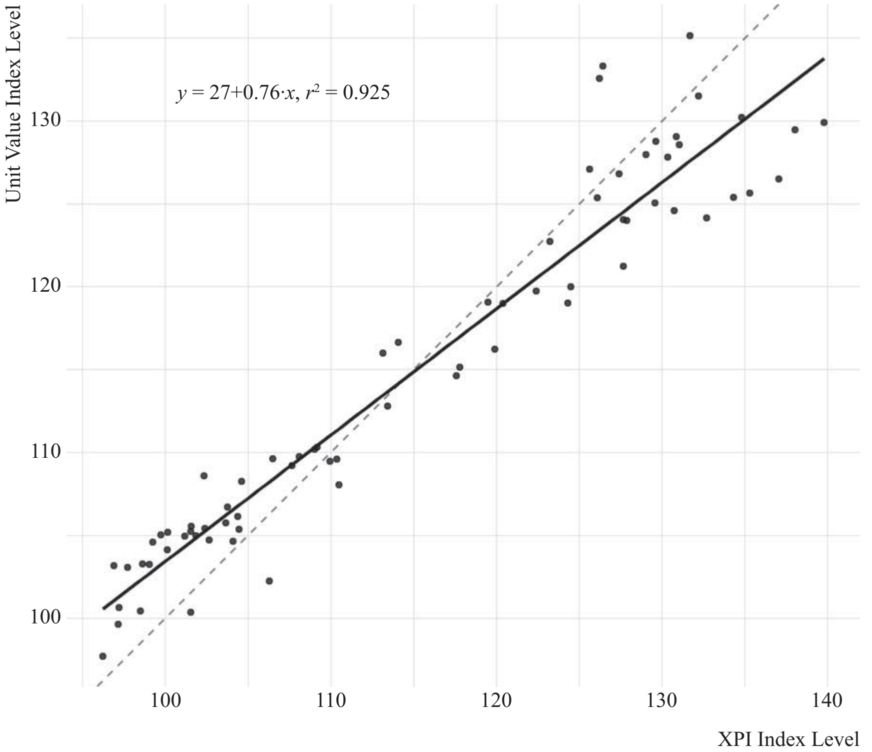


Fig. 5. Regression of Unit Value Index Levels on XPI Index Levels. Each point represents the index levels for a month. The 45 degree line is included as a dashed line. The solid line is a linear regression line and the equation for the line is at the top left of the figure along with the R-squared of the regression. The product category is "Other animal feeds, not elsewhere classified." The UVI for this product category is plotted in Figure 2.

Table 2. Average of three main statistics by manual UVI rating and published status.

	Published			Unpublished		
	Good	Undecided	Poor	Good	Undecided	Poor
Difference	8.91	23.75	172.76	13.25	26.44	492.90
Intercept	34.11	124.14	2162.52	29.97	101.37	2804.93
T-Test	0.80	0.30	0.14	0.83	0.57	0.21

Notes: Average of each statistic for indexes in each group. Difference is the fifth largest absolute difference. Intercept is the intercept of a regression of the index values of the unit value index on the index value of the XPI. T-test is the p-value from a paired t-test using the month-to-month changes of both indexes. Lower values indicate more agreement for the difference and intercept. Higher value indicate more agreement for the t-test. Published indicates the BEA end-use product category is published by the XPI, indicating it is a more reliable benchmark.

Table 3 presents results for additional statistics to show that these follow the same trends with the "good" product categories performing the best. The results of these statistics are broadly consistent with the results presented above. However, in each of these cases, we found instances where the statistic indicated significant agreement, but the manual review indicated the UVI was not a good match with the XPI benchmark. The table shows that mean absolute difference in month-to-month changes between the UVIs in published product categories that received a "good" rating and their corresponding XPI

Table 3. Average of three Main Statistics by manual UVI rating and published status.

	Published			Unpublished		
	Good	Undecided	Poor	Good	Undecided	Poor
RMSE	3.35	15.66	36.64	4.30	5.58	60.55
MAD	2.50	4.15	12.03	3.10	3.84	25.97
Sign Agreement	67.81	58.18	48.63	55.71	48.03	44.54

Notes: Average of each statistic for indexes in each group. RMSE is the root mean-squared error between the month-to-month % changes. MAD is the mean absolute difference between the month-to-month % changes. Sign agreement is the percentage of months where the sign of the month-to-month changes is the same. Published indicates the BEA end-use product category is published by the XPI, indicating it is a more reliable benchmark.

index is 2.5 percentage points. The sign of the month- to-month changes agreed 67.81% of the time. The "undecided" indexes performed slightly worse with a difference of 4.15 percentage points, agreeing only 58.18% of the time. The "poor" indexes performed the worst by far with a difference of 12.03 percentage points, but the sign of the change still agreed over 48% of the time.

6. Results

Table 4 shows that 28.7% of the share of goods export trade could be represented by good quality unit value indexes valued at USD407 billion of 2015 trade, counting both published and unpublished XPI indexes. If the undecided unit value indexes were to have a better fit with different assumptions, an additional 16% of export trade could use unit value indexes to measure price change. Of the undecided unit value indexes, 21 (representing 9% of trade) are not published. The discrepancy between unit value indexes and XPI could be because XPI is not a good benchmark. Seven of the 21 product categories are more homogeneous than "vegetables" so using unit value indexes for at least these product categories may be possible. There are 64 unit value indexes that are considered poor, representing 56.6% of the value of trade. Forty of these come from unpublished indexes, but only two of the 40 product categories are more homogeneous than "vegetables" so it appears likely that many of the poor fitting unit value indexes should not replace the XPI index. Most of the remaining product categories are heterogeneous goods such as machinery that are not suitable for UVIs.

Table 4. Manual rating by published status.

Published	Rating	Number of Indexes	Trade USD Value	Share of Indexes	Share of Trade USD
Yes	Good	15	271	12.1	19.1
No	Good	12	136	9.8	9.6
Yes	Undecided	11	100	8.9	7.1
No	Undecided	21	128	17.1	9.0
Yes	Poor	24	307	19.5	21.6
No	Poor	40	218	32.5	15.4

Notes: Published indicates whether XPI publishes a detailed index for a given product category. Rating is the consensus rating from manual review. The trade USD value is in 2015 billions.

6.1. Homogeneity Matters

Even though evaluating homogeneity was not the primary test of index comparability, the use of the coefficient of variation is supported by the results. Figures 6, 7, and 8 show the CDFs of the coefficient of variation for the product categories with good UVIs (Figure 6), undecided quality UVIs (Figure 7), and poor quality UVIs (Figure 8). The product categories with good indexes have coefficient of variation CDFs that are usually above the

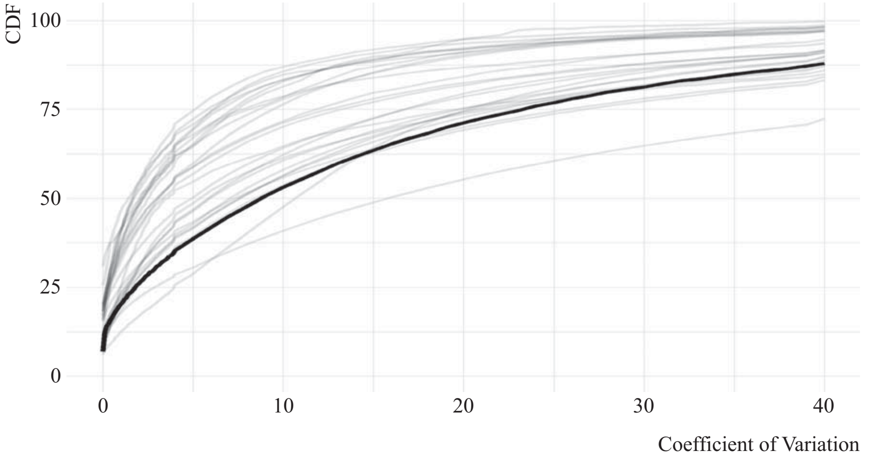


Fig. 6. CDF of Item Coefficient of Variation for 27 BEA end-use product categories that received a "good" rating. Lines represent the fraction of items in a product category that have a coefficient of variation below a given number. CDFs are calculated for individual months and then averaged across months. The darker line is the CDF for 'vegetables, vegetable preparations, and juices.' The x-axis is truncated at 40 for clarity.

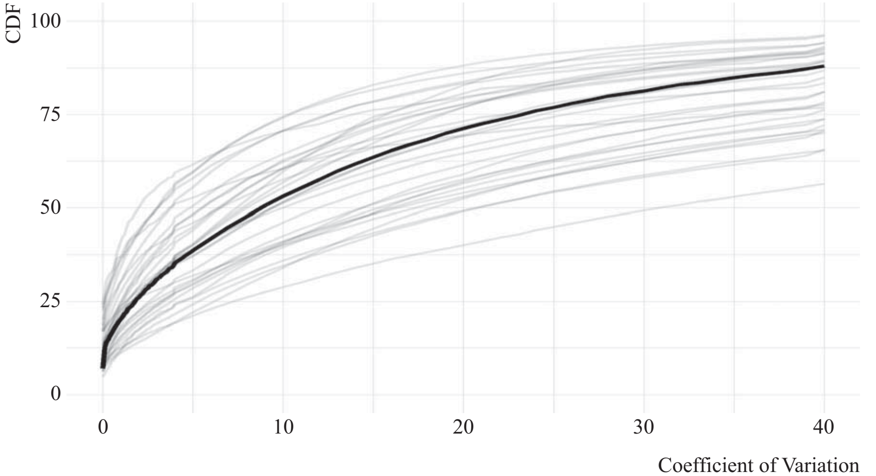


Fig. 7. CDF of Item Coefficient of Variation for 32 BEA end-use product categories that received an "undecided" rating. Lines represent the fraction of items in a product category that have a coefficient of variation below a given number. CDFs are calculated for individual months and then averaged across months. The darker line is the CDF for 'vegetables, vegetable preparations, and juices.' The x-axis is truncated at 40 for clarity.

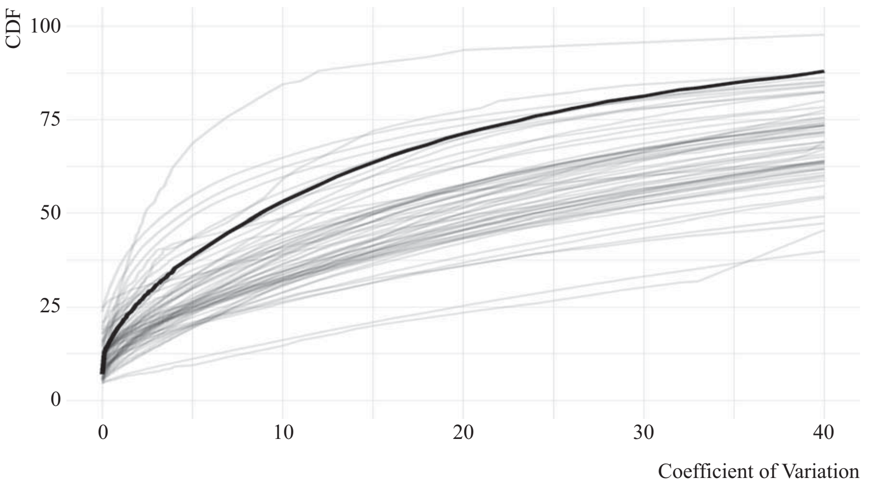


Fig. 8. CDF of Item Coefficient of Variation for 64 BEA end-use product categories that received a "poor" rating. Lines represent the fraction of items in a product category that have a coefficient of variation below a given number CDFs are calculated for individual months and then averaged across months. The darker line is the CDF for 'vegetables, vegetable preparations, and juices.' The x-axis is truncated at 40 for clarity.

vegetables CDF. The product categories with undecided quality UVIs are more mixed, while the product categories with poor quality UVIs are almost always below the CDF of vegetables.

6.2. *Impact on Gross Domestic Product*

MXPI at a detailed level are used as deflators for net exports by both the US Census and the Bureau of Economic Analysis. Net exports are a component of Gross Domestic Product. If the administrative data actually provide more and better quality statistics with the administrative trade data, they will contribute to improving the measure of GDP.

To simulate the impact of replacing the historic BEA end-use 5-digit export price indexes for 2012–2017, price indexes for each subgrouping of the "good", "undecided", and "poor" price indexes are aggregated and the difference between the trade-weighted unit value index endpoint of December 2017 is aggregated for each quality group relative to that of the corresponding official export price indexes – first for the partial aggregation and then to the top-level index, holding all else equal. As can be seen in Table 5, the 27 5-digit BEA end-use "good" unit value indexes show a December 2017 index value of 75.62. This compares to the estimate of the corresponding official XPIs of 83.87. The total dollar value of these indexes is 29% of export trade. Calculating the impact on all goods' export prices is done by assuming that the other 71% of trade value does not change prices. Applying official MXPI methodology, the 5-digit indexes are aggregated to an all-goods measure, which results in a top-level aggregate unit value index of 93.00 and the official XPI of 95.37. The price change gap in the "good" price indexes is 8.3 percentage points over the six years. For the top-level index comparison, the price gap between the "good" unit value indexes versus the historic official price indexes is 2.4 percentage points from 2012 to 2017. That is, if the "good" unit value indexes were to be incorporated into the all-

Table 5. Effect of UVIs on XPI price indexes.

Index classification	Average of detailed XPI price indexes	Average of detailed unit value indexes	Constant all-goods official price index	Constant all-goods unit value index	% of total trade USD value
Good (N=27)	83.87	75.62	95.37	93.00	29
Undecided (N=32)	100.22	119.52	100.04	103.13	16
Poor (N=64)	99.43	312.09	99.68	220.20	57

Notes: Columns 2 and 3 calculate a price index by combining the BEA end-use product categories that receive a given manual rating. Columns 4 and 5 compare the overall goods index when the UVIs with a given rating are used instead of the XPI indexes for those product categories.

goods XPI, export prices would have risen 2.4 percentage points more. In addition, when applied as deflators, real export prices would have fallen by 2.4 percentage points more. The same thought experiment comparing the "undecided" unit value indexes with their comparable 5-digit official price indexes results in a gap of 3.1 percentage points in the other direction. This gap is only slightly larger than the 2.4 percentage point different for the "good" indexes, but this fact is because the "undecided" indexes account for only 16% of the value of exports, much less than the 29% for good indexes. The "poor" unit value indexes show an extreme price gap of over 100 percentage points.

This range of price gaps and the extreme variability for the "poor" quality grouping validates the basic tenet of minimizing price variability and maximizing substitutability that guides this research and qualitatively affirms that the three quality bins are reasonable.

The effort to measure the impact on net trade and thus GDP will depend on the results from the import comparison. Hypothetically, if there were a commensurate adjustment upward to the import index prices, and thus downward impact on real goods imports, the impact on net trade may be small, and thus have a minimal impact on GDP. The direction of the impact is sensitive to the choice and number of indexes determined to be "good". The different composition of imports, with a larger share of heterogeneous products, must be calculated before the possible impact on the real trade balance and GDP can be measured.

7. Conclusion

This research on a large-scale transition to using administrative data develops an exhaustive approach to evaluate product variety, price variability, substitutability, and index comparability for potential replacement in the official MXPI. The results show that not all unit value indexes are the same and that unit value indexes most likely to replace official MXPI measures are homogeneous and should closely align with the index they replace. The ambiguity of some of the indeterminate quality indexes may lie in the fact that there is no comparable published official XPI, or that the product variety itself is broad.

The test run of unit value indexes for 123 detailed product categories over six years shows that defining homogeneity matters, and that one can develop statistical tests and create cutoffs to evaluate differences and make judgments on the consistency, reliability, and comparability of unit value indexes relative to the official export price indexes. Product categories representing 43% of the value of exports could potentially be replaced by unit value indexes if future research is able to convert all "undecided" indexes into "good" indexes. Ongoing approaches attempt to improve the quality of these unit value indexes, by evaluating the product variety key, the bias introduced by the new methodology, and/or the nonresponse bias in the official price index.

Research continues to be evaluated and refined. Some efforts to estimate hedonic linear regressions on the complete data set to develop a systematic method of identifying the best item key combination for each strata are constrained by IT capacity. Exploring time-dummy hedonic models may reduce the specification constraints of grouping data variables and calculating indexes for items. At the margins, improvements to the quality of the "good" set of indexes and additionally "undecided" indexes possibly can move to the "good" category. The Tornqvist index formula helps with substitution bias at the classification group level, but it introduces new concerns because it is a monthly-chained index. Frequent chaining has been determined to exacerbate chain drift (Ivancic et al. 2011; De Haan and Van der Grient 2011). Thus, work is being done to measure chain drift in the UVIs and investigate alternative aggregation methods, such as the base construction strategy described by Statistics Finland (Nieminen and Montonen 2018).

In addition, work will be done to analyze and compare import unit value indexes with official import price indexes and to calculate partial-month measures with low variance compared to full-month data. These two large projects must be addressed before the project to blend administrative data into official import and export price measures can begin.

8. References

Alterman, W. 1991. "Price trends in US trade: New data, new insights." In *International economic transactions: Issues in measurement and empirical research*, edited by P. Hooper and J.D. Richardson, 109–143. University of Chicago Press. Available at: https://www.nber.org/system/files/chapters/c8429/c8429.pdf (accessed April 2021)

Balk, B.M. 1998. "On the Use of Unit Value Indices as Consumer Price Subindices." Proceedings of the Fourth Meeting of the Ottawa Group, International Working Group on Price Indices. Washington, D.C., April 22–24, 1998. Available at: http://www.ottawagroup.org (accessed April 2021).

Bradley, R. 2005. "Pitfalls of Using Unit Values as a Price Measure or Price Index." *Journal of Economic and Social Measurement* 30: 39–61. DOI: https://10.3233/JEM-2005-0243.

Cerritos, G. 2015. "The Role of BLS Import and Export Price Indexes in the Real GDP," *Monthly Labor Review*. DOI: https://10.21916/mlr.2015.19.

Chessa, A. 2019. "MARS: A Method for Defining Products and Linking Barcodes of Item Relaunches." Proceedings of the Sixteenth Meeting of the Ottawa Group, International

Working Group on Price Indices. Rio de Janeiro, Brazil, May 8–10, 2019. Available at: http://www.ottawagroup.org (accessed April 2021).

Dalton, M., and B. Fissel. 2018. "A Unified Framework for Calculating Aggregate Commodity Prices from a Census Dataset," *Journal of Economic and Social Measurement* 43: 85–104. DOI: https://10.3233/JEM-180453.

De Haan, J. and H.A. van der Grient. 2011. "Eliminating Chain Drift in Price Indexes Based on Scanner Data." *Journal of Econometrics* 161: 36–46. DOI: https://10.1016/j.jeconom.2010.09.004.

Diewert, W.E., and P. von der Lippe. 2010. *Notes on Unit Value Index Bias.* Discussion Paper 10-08, Department of Economics, University of British Columbia. October. Available at: https://economics.ubc.ca/files/2013/06/pdf_paper_erwin-diewert-10-8-notes-unit-value.pdf (accessed April 2021)

Fast, D., and S.E. Fleck. 2019. *Measuring Export Price Movements with Administrative Trade Data.* Working Paper 518. U.S. Bureau of Labor Statistics. June. Available at: https://www.bls.gov/osmr/research-papers/2019/pdf/ec190080.pdf (accessed April 2021).

Fast, D., and S.E. Fleck. 2022. "Unit Values for Import and Export Price Indexes – A Proof of Concept." in *Big Data for Twenty-First-Century Economic Statistics. University of Chicago Press*, edited by K.G. Abraham, R.S. Jarmin, B.C. Moyer, and M.D. Shapiro. Available at: https://www.nber.org/books-and-chapters/big-data-twenty-first-century-economic-statistics (accessed April 2021).

Fitzgerald, J., and O. Shoemaker. 2013. *Evaluating the Consumer Price index Using Nielsen's Scanner Data.* Office of Survey Methods Research Paper. U.S. Bureau of Labor Statistics. Available at: https://www.bls.gov/osmr/research-papers/2013/st130070.htm (accessed April 2021).

Gopinath, G., and O. Itskhoki. 2010. "Frequency of Price Adjustment and Pass-Through." *The Quarterly Journal of Economics* 125 (2): 675–727. DOI: https://10.1162/qjec.2010.125.2.675.

Gopinath, G., and R. Rigobon. 2008. "Sticky Borders." *The Quarterly Journal of Economics* 123 (2): 531–575. DOI: https://10.1162/qjec.2008.123.2.531.

Hottman, C.J., and R. Monarch. 2018. "Estimating Unequal Gains across US Consumers with Supplier Trade Data." *International Finance Discussion Papers* 1220. DOI: https:// 10.17016/IFDp. 2018.1220.

International Monetary Fund. 2009. *Export and Import Price Index Manual: Theory and Practice*. Washington, D.C. Available at: https://www.imf.org/external/np/sta/xipim/pdf/xipim.pdf (accessed April 2021).

Ivancic, L., W.E. Diewert, and K.J. Fox. 2011. "Scanner Data, Time Aggregation and the Construction of Price Indexes." *Journal of Econometrics* 161 (1): 24–35. DOI: https://10.1016/j.jeconom.2010.09.003.

Moulton, B. 2018. "The measurement of output, prices, and productivity." Available at: https://www.brookings.edu/research/the-measurement-of-output-prices-and-productivity/ (accessed April 2021).

Nakamura, A.O., W.E. Diewert, J.S. Greenlees, L.I. Nakamura, and M.B. Reinsdorf. 2015. "Sourcing Substitution and Related Price Index Biases." In *Measuring Globalization:*

Better Trade Statistics for Better Policy, edited by S.N. Houseman and M.J. Michael, 1: 21–89. DOI: https://10.17848/9780880994903.

Nakamura, E., and J. Steinsson. 2012. "Lost in Transit: Product Replacement Bias and Pricing to Market." *American Economic Review* 102 (7): 3277–3316. DOI: https://10.1257/aer.102.7.3277.

Nieminen, K., and S. Montonen. 2018. "The Foundation of Index Calculation." *Helsinki: Statistics Finland*. Available at: http://www.stat.fi/static/media/uploads/meta_en/menetelmakehitystyo/foundation_of_index_calculation_nieminenmontonen_final.pdf (accessed April 2021).

Reid, G, F. Zabala and A. Holmberg. 2017. "Extending TSE to Administrative Data: A Quality Framework and Case Studies from Stats NZ." *Journal of Official Statistics* 33 (2). DOI: https://10.1515/jos-2017-0023.

Reinsdorf, M., and R. Yuskavage. 2018. "Offshoring, Sourcing Substitution Bias and the Measurement of US Import Prices, GDP and Productivity." *The Review of Income and Wealth* 64 (1): 127–146. DOI: https://10.1111/roiw.12263.

Silver, M. 2009. "Do Unit Value Export, Import, and Terms of Trade Indices Represent or Misrepresent Price Indices?" *IMF Staff Papers* 56 (2). 297–322, June. DOI: https://10.1057/imfsp. 2008.24.

Silver, M. 2010. "The Wrongs and Rights of Unit Value Indices." *Review of Income and Wealth* 56: S206–S223. DOI: https://10.1111/j.1475-4991.2010.00391.x.

Silver, M. 2011. "An Index Number Formula Problem: The Aggregation of Broadly Comparable Items." *Journal of Official Statistics* 27 (4): 1–17. Available at: https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/an-index-number-formula-problem-the-aggregation-of-broadly-comparable-items.pdf (accessed April 2021).

U.S. Bureau of Labor Statistics. 2020a. *Response Rate Report.* Office of Survey Method Research. Online only. Available at: https://www.bls.gov/osmr/response-rates/ (accessed April 2021).

U.S. Bureau of Labor Statistics. 2020b. "International Price Program." *Handbook of Methods*. Available at: https://www.bls.gov/opub/hom/ipp/home.htm (accessed April 2021).

Von der Lippe, P., and J. Mehrhoff, 2010. "Unit Value Bias Reconsidered." Contributions to the annual meeting of the Association for Social Policy. Okonomie der Familie – Session: How Reliable are German Aggregate Price Indices? No. F2-V3 Association for Social Policy, Frankfurt a.M. Available at: http://hdl.handle.net/10419/37526 (accessed April 2021).

Received May 2020
Revised January 2021
Accepted May 2021

Journal of Official Statistics, Vol. 38, No. 1, 2022, pp. 107–126, http://dx.doi.org/10.2478/JOS-2022-0006

Substitution Bias in the Measurement of Import and Export Price Indices: Causes and Correction

Ludwig von Auer[1] *and Alena Shumskikh*[1]

The import and export price indices of an economy are usually compiled by some Laspeyres type index. It is well known that such an index formula is prone to substitution bias. Therefore, also the terms of trade (ratio of export and import price index) are likely to be distorted. The underlying substitution bias accumulates over time. The present article introduces a simple and transparent retroactive correction approach that addresses the source of the substitution bias and produces meaningful long-run time series of import and export price levels and, therefore, of the terms of trade. Furthermore, an empirical case study is conducted that demonstrates the efficacy and versatility of the correction approach.

Key words: Distortion; official statistics; terms of trade; time series.

1. Introduction

Besides the consumer price index and the producer price index, the national statistical offices (NSOs) usually publish a monthly or quarterly export price index and import price index. The latter two indices are used in the indexation of various types of international contracts and they are also required in the national accounts as deflators of nominal values of exports and imports. These are necessary to derive volume estimates of GDP by the expenditure approach. In the assessment of an economy's inflationary trends, special attention is paid to the import price index, because it is considered as an early indicator of increasing or weakening inflationary pressure. Correspondingly, the export price index is an early indicator of the inflationary pressure in the destination countries of the exports.

The terms of trade index of an economy is usually defined as the ratio of the economy's export price index and import price index. Changes in the terms of trade translate into changes of the real income of the economy's population.

Bias in the measurement of the import and export indices could weaken the reliability of economic statistics used for public and private economic decision making. Therefore, the "Export and Import Price Index Manual" published by the IMF (2009) provides recommendations for the measurement of the export and import price indices. In practice, the NSOs must also ensure a cost efficient and timely publication of the newest index numbers. Therefore, most NSOs rely on some type of Laspeyres index, although these indices are suspected of (upper-level) substitution bias (e.g., Dridi and Zieschang 2004,

[1] University of Trier, Fachbereich IV-Finanzwissenschaft, Universitaetsring 15, Trier, 54296, Germany. Emails: vonauer@uni-trier.de and shumskikh@uni-trier.de

Acknowledgments: We gratefully acknowledge the extremely helpful advice of two anonymous referees.

© Statistics Sweden

169; IMF, 2009, 413–439). In chained Laspeyres type indices the substitution bias would accumulate over time.

Therefore, the central contribution of the present article is a fully worked out retroactive correction approach that, with some delay, provides more reliable index numbers than those produced by a chained Laspeyres type index. Our correction approach produces meaningful long-run time series of import and export prices and, therefore, of the terms of trade. The approach is simple and transparent. It can be applied not only to an import or export price index, but also to a consumer or producer price index.

The first pillar of our correction approach is a retroactively computed price index that compares the prices of the two latest expenditure reference periods, that is, the two latest periods for which detailed information about the relative importance of the various products is available. Because of its symmetric treatment of the two expenditure reference periods, we use the Törnqvist formula, though other index formulas that treat the two periods in a symmetric fashion would be equally appropriate (e.g., Walsh, Marshall-Edgeworth, Fisher index). Chaining consecutive Törnqvist indices instead of Laspeyres indices removes the long-run substitution bias. However, it does not correct the index numbers of the periods between the expenditure reference periods. To revise also these index numbers, we construct from the ratio of the Laspeyres and Törnqvist index a correction factor the impact of which gradually increases between the two expenditure reference periods. This correction factor is the second pillar of our approach.

The second contribution of the present article is an empirical case study that not only estimates a lower bound of the long-run substitution bias in officially published import and export price indices, but also demonstrates how the correction approach can be implemented to mitigate both the short-run bias arising between the expenditure reference periods and the long-run bias. We have opted for the trade data of the Federal Statistical Office of Germany (Destatis). These data are publicly available, they comprise price data collected from producers and wholesalers (instead of the less reliable unit values compiled from customs sources), and the official compilation procedure of Destatis is documented in accessible publications (Statistisches Bundesamt 2019, 6–7).

Theoretically, if the substitution bias in the export and import price indices was of equal magnitude, the terms of trade index would remain unbiased. However, we show that there is hardly any substitution bias in the German export price index, while the upward substitution bias in the German import price index is substantial.

The explanation of this finding is the third contribution of this article. Our analysis reveals that the difference between the bias in the export and import price index is mostly driven by the volatility of the prices of oil and gas and by the fluctuations in the exchange rates.

Our findings are relevant not only for Destatis, but for all NSOs that use infrequently chained Laspeyres type indices for their measurement of import and export prices. Without an appropriate retroactive correction of these indices, the substitution bias in the officially published index numbers accumulates over time.

The article proceeds as follows. Section 2 presents in stylized form the interpretation and official computation of the import and export price indices. Section 3 explains why they are likely to be biased. Furthermore, a retroactive correction approach is introduced. Its application to the German foreign trade data is presented in Section 4. Section 5 concludes.

2. Calculation of the Terms of Trade

Suppose that during a reference period $t = 0$ a country imports only one single good i and exports only one other good j. Let p_i^0 be the euro price of the imported good and z_j^0 the euro price of the exported good. Then, the terms of trade of the reference period (ToT^0) are defined by the ratio of these two prices: $\mathrm{ToT}^0 = z_j^0/p_i^0$. Correspondingly, the terms of trade of some comparison period $t = 1$ are $\mathrm{ToT}^1 = z_j^1/p_i^1$. The change in the terms of trade between the reference and the comparison period can be expressed by the following ratio:

$$\frac{\mathrm{ToT}^1}{\mathrm{ToT}^0} = \frac{z_j^1/p_i^1}{z_j^0/p_i^0} = \frac{z_j^1/z_i^0}{p_j^1/p_i^0}. \tag{1}$$

The right-hand side equality is the change in the export price in relation to the change in the import price.

In the case of many goods, average price changes can be measured by price index formulas. Applying the Laspeyres formula, the change in the export prices is given by

$$E_{\mathrm{La}}^{0\to1} = \frac{\sum_{j=1}^{N} z_j^1 x_j^0}{\sum_{j=1}^{N} z_j^0 x_j^0}, \tag{2}$$

where x_j^0 is the quantity of good j exported during the reference period and N is the number of exported goods. The subscript "La" stands for "Laspeyres" and the superscript "$0 \to 1$" indicates that the index measures the average price change between the reference period 0 and the comparison period 1. Analogously, the price index of the import prices is

$$I_{\mathrm{La}}^{0\to1} = \frac{\sum_{i=1}^{M} p_i^1 m_i^0}{\sum_{i=1}^{M} p_i^0 m_i^0}, \tag{3}$$

where m_i^0 is the quantity of good i imported during the reference period and M is the number of imported goods.

The change in the terms of trade between the reference and the comparison period is derived from the quotient of the price indices (2) and (3):

$$\mathrm{ToT}^{0\to1} = \frac{\mathrm{ToT}^1}{\mathrm{ToT}^0} = \frac{E_{\mathrm{La}}^{0\to1}}{I_{\mathrm{La}}^{0\to1}}. \tag{4}$$

This ratio is the terms of trade index. It can be interpreted as a generalisation of the right-hand side equality of Equation (1) to the case of many goods. Even though $\mathrm{ToT}^{0\to1}$ defined by Equation (4) measures "the *change* in the terms of trade between the reference and the comparison period", it is conventionally denoted as "*the* terms of trade of the comparison period". We follow this convention.

The terms of trade index (4) can also be written in the form that corresponds to the left-hand side equality in Equation (1):

$$\mathrm{ToT}^{0\to 1} = \frac{\left(\sum_{j=1}^{N} z_j^1 x_j^0\right) / \left(\sum_{i=1}^{M} p_i^1 m_i^0\right)}{\left(\sum_{j=1}^{N} z_j^0 x_j^0\right) / \left(\sum_{i=1}^{M} p_i^0 m_i^0\right)}. \tag{5}$$

Interpreting the exported quantities of the reference period, $\left(x_1^0, \ldots, x_N^0\right)$, as the "reference period export basket" and the imported quantities of the reference period, $\left(m_i^0, \ldots, m_M^0\right)$, as the "reference period import basket", the denominator of Equation (5) measures the purchasing power of a reference period export basket measured in units of reference period import baskets. The numerator indicates the purchasing power of the reference period export basket during the comparison period. Again, this purchasing power is measured in units of reference period import baskets. Therefore, the terms of trade index (5) indicates the change in the purchasing power of the reference period export basket, measured in units of reference period import baskets.

Figure 1 shows the officially published terms of trade of the G7-countries (Canada, France, Germany, Italy, Japan, UK, and US) from 1995 to 2018. The statistical methodologies of the NSOs these countries are described in Statistics Canada (2019), INSEE (2019), Peter (2009, 2014, 2019), Statistiches Bundesamt (2004), Istat (2019), Bank of Japan (2019), ONS (2017), and U.S. Bureau of Labor Statistics (2018a,b, 2020). Canadian terms of trade are largely driven by the price of oil and gas. The terms of trade of all other G7-countries are negatively correlated with the Canadian ones. Between 2000 and 2007, there was a strong devaluation of the Japanese yen against the euro, though not against the USD. Despite the yen's subsequent appreciation until 2012, the Japanese terms of trade remained at their lower level.

The terms of trade depicted in Figure 1 are derived from the ratio of the price index of export prices and the price index of import prices. The next section explains why such indices may exhibit substitution bias and how the problem can be mitigated.

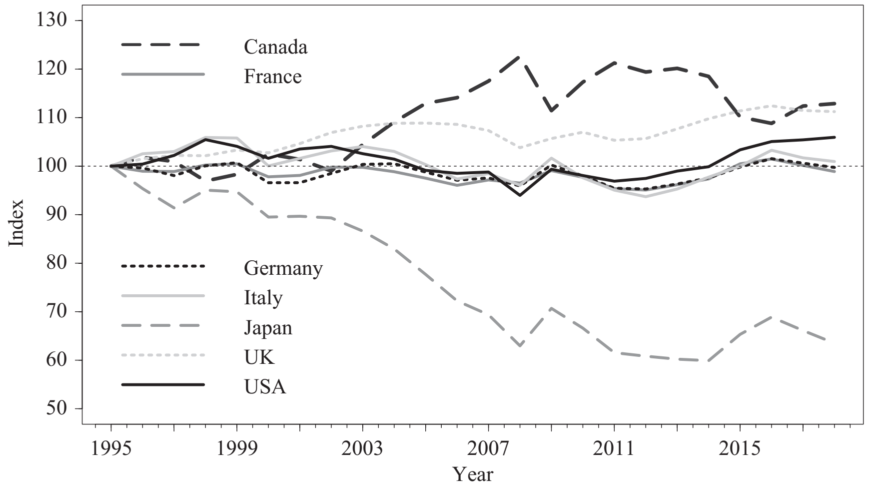


Fig. 1. Terms of trade of the G7 countries (1995 = 100) from 1995 to 2018.
Source: Own calculations based on data of OECD (2022).

3. Retrospective Correction of Substitution Bias

Bias in the terms of trade arises, when the export price index, E_{La}, or the import price index, I_{La}, or both price indices are distorted – provided that the two distortions do not cancel. Of course, such distortions can occur already during data collection and processing. These problems are well known and extensively discussed in IMF (2009, 287–297). Therefore, it is assumed here that the available price and quantity data are accurate and that the only remaining source of bias is the choice of the index formula.

In IMF (2009, 413–439) it is argued that imports and exports can be viewed from a resident's perspective or a non-resident's perspective. For each perspective, economic theory makes predictions about the direction of the measurement bias arising from the Laspeyres index formula.

When the resident's perspective is applied, the import quantities and prices collected by NSOs reflect the residents' cost minimizing consumer behavior (including firms purchasing their inputs). The consumer side increases the purchases of products that become relatively less expensive and they reduce the purchases of products that become relatively more expensive. This consumer behavior would result in a negative correlation between intertemporal price and quantity changes and, therefore, in upward substitution bias of the Laspeyres index.

Furthermore, the observed export quantities and prices reflect the revenue maximization of the residents' producer side. This side increases the output of products that become relatively more expensive and they reduce the output of products that become relatively less expensive. Therefore, the intertemporal price and quantity changes are positively correlated. In a Laspeyres index, this would lead to downward substitution bias.

In sum, in the resident's perspective the numerator in the terms of trade index (4) would understate the average change in export prices, while the denominator would overstate the average change in import prices. Therefore, the measured terms of trade would exhibit downward bias.

When a non-resident's perspective were applied, the direction of the bias would be reversed. However, most NSOs use the resident's perspective.

These considerations are conjectures that are based on economic theory. An empirical examination of these conjectures requires a measurement approach that can be expected to produce unbiased index numbers. A deviation between these unbiased index numbers and the Laspeyres numbers is an indication of the direction and extent of the actual substitution bias of the Laspeyres index.

The derivation of such unbiased index numbers starts with the Laspeyres index (3). It is often expressed in the following equivalent form:

$$I_{\mathrm{La}}^{0\to 1} = \sum_{i=1}^{M} s_i^0 \frac{p_i^1}{p_i^0}, \tag{6}$$

with

$$s_i^0 = \frac{p_i^0 m_i^0}{\sum_{j=1}^{M} p_j^0 m_j^0}.$$

Therefore, the Laspeyres index can be interpreted as a weighted arithmetic mean of price ratios where the weights are the expenditure shares of the reference period.

The Laspeyres index is not the only index formula that is prone to substitution bias. The Paasche index,

$$I_{\text{Pa}}^{0\rightarrow 1} = \frac{\sum_{i=1}^{M} p_i^1 m_i^1}{\sum_{i=1}^{M} p_i^0 m_i^1} = \left(\sum_{i=1}^{M} s_i^1 \left(\frac{p_i^1}{p_i^0} \right)^{-1} \right)^{-1}, \tag{7}$$

has a similar problem, though its substitution bias would point in the opposite direction. The right-hand side of Equaiton (7) expresses the Paasche index as the weighted harmonic mean of the price ratios where the weights are the expenditure shares of the comparison period.

A large number of price indices avoid the substitution bias of the Laspeyres index. Examples are the Walsh, Marshall-Edgeworth, Fisher, and Törnqvist index. These index formulas utilize not only the set of reference period quantities or the set of comparison period quantities, but both sets of quantities. Usually, these four formulas generate very similar index numbers. Therefore, we confine our analysis to the Törnqvist index. It is defined by

$$I_{\text{Tö}}^{0\rightarrow 1} = \exp \left(\sum_{i=1}^{M} \frac{1}{2} \left(s_i^0 + s_i^1 \right) \ln \left(\frac{p_i^1}{p_i^0} \right) \right). \tag{8}$$

The compilation of this price index requires not only the expenditures $p_i^0 m_i^0$ and the price ratios p_i^1/p_i^0, but also the expenditures $p_i^1 m_i^1$. However, the collection and compilation of the latter expenditures is a labor intensive process that would significantly delay the publication of the indices (IMF 2009, 58). Therefore, NSOs do not have the capacity to timely calculate and publish the results of such an index formula.

Nevertheless, the IMF (2009, 56) points out that a retroactive revision of the index numbers would be feasible when the requisite data on updated expenditure weights become available. The method presented here retroactively applies the updated expenditure weights to gradually transform the Laspeyres indices to Törnqvist indices, thus addressing the substitution bias with a historical revision.

For example, suppose that in January 2010 (shorthand notation 1/10) a survey was conducted providing us with the import expenditure weights for that month, $s_i^{1/10}$. Therefore, this month is our first *expenditure reference* period. For February 2010 (2/10) and all subsequent months we calculate a monthly Laspeyres import price index that measures the average price change between the *price reference* (and first expenditure reference) period January 2010 and some comparison period t:

$$I_{\text{La}}^{1/10\rightarrow t} = \sum_{i=1}^{M} s_i^{1/10} \frac{p_i^t}{p_i^{1/10}}, \quad t = 1/10,\ 2/10, \ldots. \tag{9}$$

For the comparison period $t = 1/10$, this index yields $I_{\text{La}}^{1/10\rightarrow 1/10} = 1$, that is, the reference price level of the index series is January 2010 $= 1$. Therefore, January 2010 is not only the price and weight reference period, but also the *index reference* period.

Suppose that January 2015 is the next expenditure reference period, but that the expenditure weights relating to that month become available only in July 2018. Therefore, January 2010 remains the price (and index) reference period also for the Laspeyres indices compiled between January 2015 and July 2018. In July 2018, when the expenditure weights of January 2015 become available, we can conduct a retroactive revision of past index numbers. We propose to conduct this revision in three stages. The first and second stage revise the index numbers of January 2015 to July 2018, while the third stage revises the index numbers of January 2010 to December 2014. The three stages yield a consistent time series of price levels that stretches from January 2010 to July 2018. It can be easily continued without harming its consistency.

Stage 1: We begin the revision by computing a new series of Laspeyres index numbers for January 2015 to July 2018. This new series uses as price, expenditure, and index reference period January 2015 instead of January 2010:

$$I_{\mathrm{La}}^{1/15\to t} = \sum_{i=1}^{M} s_i^{1/15} \frac{p_i^t}{p_i^{1/15}}, \quad t = 1/15,\ 2/15, \ldots, 7/18. \tag{10}$$

This new series of Laspeyres index numbers can be expected to exhibit considerably less substitution bias than the original series, because the quantity information is more up to date (January 2015 instead of January 2010). However, some substitution bias remains, because we still apply the Laspeyres index formula. Only when the results of the next expenditure reference period will become available, this bias can be addressed.

Stage 2: To rebase the new series to the index reference period January 2010, advocates of the Laspeyres index would multiply the index numbers calculated by Equation (10) by the Laspeyres index $I_{\mathrm{La}}^{1/10\to 1/15}$. However, we know that this Laspeyres index exhibits substantial substitution bias. Therefore, we recommend to apply the Törnqvist index (8) instead:

$$I_{\mathrm{Tö}}^{1/10\to 1/15} = \exp\left(\sum_{i=1}^{M} \frac{1}{2}\left(s_i^{1/10} + s_i^{1/15}\right) \ln\left(\frac{p_i^{1/15}}{p_i^{1/10}}\right)\right). \tag{11}$$

In view of the upward substitution bias of the original Laspeyres index, we expect that $I_{\mathrm{Tö}}^{1/10\to 1/15} < I_{\mathrm{La}}^{1/10\to 1/15}$. The rebased series is obtained from

$$I^{1/10\to t} = I_{\mathrm{Tö}}^{1/10\to 1/15} \cdot I_{\mathrm{La}}^{1/15\to t}, \qquad t = 1/15,\ 2/15, \ldots, 7/18. \tag{12}$$

Therefore, $I^{1/10\to 1/15} = I_{\mathrm{Tö}}^{1/10\to 1/15}$. The rebased series of price index numbers relates each of the price levels between January 2015 and July 2018 to January 2010 = 1. Overall, we expect a significant downward revision of the original price levels compiled by Equation (9).

The first two stages of revision affect only the time series of import price levels from January 2015 to July 2018. The price levels of the previous months have not yet been revised. Quite likely, the original price level of December 2014, $I_{\mathrm{La}}^{1/15\to 12/14}$, is larger than the downwards revised price level of January 2015, $I_{\mathrm{Tö}}^{1/10\to 1/15}$. To obtain a consistent time series stretching from January 2010 to July 2018, the price level of the index reference period January 2010 should remain at 1, but the price levels of February 2010 to December 2014 must be revised. This is the third and most challenging stage of the retroactive revision.

Stage 3: In the course of Stage 2, we replaced the Laspeyres index $I_{\text{La}}^{1/10\to 1/15}$ by the Törnqvist index $I_{\text{Tö}}^{1/10\to 1/15}$. Obviously, we cannot make the same replacement for the previous months ($t = 2/10$ to $t = 12/14$), because the quantities of these months are unknown. Quantity information is available only for the two expenditure reference periods January 2010 and January 2015. However, we can utilize in our retroactive revision of these earlier index numbers the ratio $\left(I_{\text{Tö}}^{1/10\to 1/15}/I_{\text{La}}^{1/10\to 1/15}\right)$ as a correction factor:

$$I^{1/10\to t} = I_{\text{La}}^{1/10\to t} \cdot \left(\frac{I_{\text{Tö}}^{1/10\to 1/15}}{I_{\text{La}}^{1/10\to 1/15}}\right)^{\lambda_t}, \qquad t = 1/10, 2/10, \ldots, 1/15. \tag{13}$$

The parameter λ_t represents the impact that we concede to the correction factor. For the comparison period $t = 1/15$, the correction factor should exert its full impact ($\lambda_{1/15} = 1$) such that Equation (13) yields $I^{1/10\to 1/15} = I_{\text{Tö}}^{1/10\to 1/15}$. However, for the initial months ($t = 2/10, 3/10,...$) the situation is different. During these comparison periods, the import basket of month $t = 1/10$ is far less outdated than in month $t = 1/15$ and, therefore, the substitution bias of $I_{\text{La}}^{1/10\to 2/10}$, say, tends to be much smaller than that of $I_{\text{La}}^{1/10\to 1/15}$. Accordingly, during the initial months, the impact λ_t should be smaller than 1. In fact, for the comparison period $t = 1/10$, the correction factor should have no impact ($\lambda_{1/10} = 0$). Otherwise, we would get $I^{1/10\to 1/10} \neq 1$. As the comparison period t moves away from the price reference period 1/10, the impact λ_t should gradually increase above 0 until, in $t = 1/15$, it finally reaches its maximum value 1.

For a formal definition of the impact λ_t we introduce the counter variable s. In the first month, $t = 1/10$, this integer has the value $s = 1$, in month $t = 2/10$ the value $s = 2$, and so on. In month $t = 1/15$ the counter reaches its maximum value $s = 61$. Denoting this maximum value by S, we can define the impact λ_t by

$$\lambda_t = \frac{s-1}{S-1}. \tag{14}$$

Equations (13) and (14) yield the desired series of revised price index numbers for the months January 2010 to January 2015. Combining this series with the revised index numbers that were compiled during Stages 1 and 2 of the revision process, yields a consistent series of revised index numbers stretching from January 2010 to July 2018. It is consistent in the sense, that, for month $t = 1/15$, formulas (12) and (13) yield the same index number, namely $I_{\text{Tö}}^{1/10\to 1/15}$.

Equations (13) and (14) are not the only conceivable method for Stage 3 of the revision process. Some alternative options are presented in Von Auer and Shumskikh (2022), including proposals by De Haan et al. (2010).

4. Application to German Foreign Trade Data

Different NSOs apply different compilation methods for their export and import price indices. Our correction approach is adaptable to a wide range of such methods. To verify this claim and to get an impression of the magnitude of the bias inherent in official compilation procedures, we adapt our approach to the method and the trade data of Destatis.

An important difference between the calculation outlined in Section 2 and the method of Destatis is the choice of the period lengths. In Section 2, all periods had a uniform length, namely one month. By contrast, the price, expenditure and index reference periods of Destatis are years, while the comparison periods are months. The resulting complications arising in the Destatis method are described in the Appendix (Section 6). There we also demonstrate how our correction approach can be adapted to these complications.

For January 1995 to May 2019, we have monthly price levels of 30 categories of German imports and 28 categories of German exports. In addition, we know the categories' expenditure weights for the years 1995, 2000, 2005, 2010 and 2015.

As documented in Pötzsch (2004) and Peter (2009, 2014, 2019), the officially published long-run import price index of Destatis (Statistisches Bundesamt, 2019, 8–9) incorporates Stage 1 of our revision process but not Stages 2 and 3. For example, in September 2018 Destatis published the index numbers for the comparison months $t = 1/10$ to $t = 7/18$. They were compiled by the Laspeyres index with expenditure and index reference year 2010, $I_{\mathrm{La}}^{2010 \to t}$. However, the index numbers for the comparison months $t = 1/15$ to $t = 7/18$ were only preliminary. As soon as the survey results of the year 2015 became available, Destatis replaced these index numbers by revised index numbers that were compiled by the Laspeyres index with price, expenditure and index reference year 2015, $I_{\mathrm{La}}^{2015 \to t}$ (Peter 2019, 37). This revision is equivalent to Stage 1 of our three-stage revision process outlined in Section 3.

Our first task is to replicate the compilation process of Destatis and to compile a consistent time series of price levels relating to the index reference year 1995. The replicated index is denoted by $I_{\mathrm{repl}}^{95 \to t}$. Following the exposition in the Appendix, we use the following formulas:

$$
\begin{aligned}
t &= 1/95, \ldots, 1/00: & I_{\mathrm{repl}}^{95 \to t} &= I_{\mathrm{La}}^{95 \to t} \\
t &= 1/00, \ldots, 1/05: & I_{\mathrm{repl}}^{95 \to t} &= I_{\mathrm{repl}}^{95 \to 1/00} \cdot I_{\mathrm{Pa}}^{1/00 \to 00} \cdot I_{\mathrm{La}}^{00 \to t} \\
t &= 1/05, \ldots, 1/10: & I_{\mathrm{repl}}^{95 \to t} &= I_{\mathrm{repl}}^{95 \to 1/05} \cdot I_{\mathrm{Pa}}^{1/05 \to 05} \cdot I_{\mathrm{La}}^{05 \to t} \\
t &= 1/10, \ldots, 1/15: & I_{\mathrm{repl}}^{95 \to t} &= I_{\mathrm{repl}}^{95 \to 1/10} \cdot I_{\mathrm{Pa}}^{1/10 \to 10} \cdot I_{\mathrm{La}}^{10 \to t} \\
t &= 1/15, \ldots, 5/19: & I_{\mathrm{repl}}^{95 \to t} &= I_{\mathrm{repl}}^{95 \to 1/15} \cdot I_{\mathrm{Pa}}^{1/15 \to 15} \cdot I_{\mathrm{La}}^{15 \to t}.
\end{aligned}
$$

Our results are depicted in Figure 2 and compared to the (rebased) official import price index published by Destatis. Even though our data only relate to rather broad categories and do not include all subcategories of the official import price index of Destatis, our replicated import price index (labelled as "repl"), is very close to the official one (labelled as "official").

Our second task is to apply our correction approach and to compute the revised import price index, $I_{\mathrm{rev}}^{95 \to t}$. Since the replicated index includes Stage 1 of the correction approach, any deviations between the revised index $I_{\mathrm{rev}}^{95 \to t}$ and the replicated index $I_{\mathrm{repl}}^{95 \to t}$ can be attributed to Stages 2 and 3.

Our compilations follow the process outlined in the Appendix:

$$t = 1/95, \ldots, 1/00: \quad I_{\text{rev}}^{95\to t} = I_{\text{La}}^{95\to t} \cdot \left(\frac{I_{\text{Tö}}^{95\to 00}}{I_{\text{La}}^{95\to 1/00} \cdot I_{\text{Pa}}^{1/00\to 00}} \right)^{\lambda_t}$$

$$t = 1/00, \ldots, 1/05: \quad I_{\text{rev}}^{95\to t} = I_{\text{Tö}}^{95\to 00} \cdot I_{\text{La}}^{00\to t} \cdot \left(\frac{I_{\text{Tö}}^{00\to 05}}{I_{\text{La}}^{00\to 1/05} \cdot I_{\text{Pa}}^{1/05\to 05}} \right)^{\lambda_t}$$

$$t = 1/05, \ldots, 1/10: \quad I_{\text{rev}}^{95\to t} = I_{\text{Tö}}^{95\to 00} \cdot I_{\text{Tö}}^{00\to 05} \cdot I_{\text{La}}^{05\to t} \cdot \left(\frac{I_{\text{Tö}}^{05\to 10}}{I_{\text{La}}^{05\to 1/10} \cdot I_{\text{Pa}}^{1/10\to 10}} \right)^{\lambda_t}$$

$$t = 1/10, \ldots, 1/15: \quad I_{\text{rev}}^{95\to t} = I_{\text{Tö}}^{95\to 00} \cdot I_{\text{Tö}}^{00\to 05} \cdot I_{\text{Tö}}^{05\to 10} \cdot I_{\text{La}}^{10\to t} \cdot \left(\frac{I_{\text{Tö}}^{10\to 15}}{I_{\text{La}}^{10\to 1/15} \cdot I_{\text{Pa}}^{1/15\to 15}} \right)^{\lambda_t}$$

$$t = 1/15, \ldots, 5/19: \quad I_{\text{rev}}^{95\to t} = I_{\text{Tö}}^{95\to 00} \cdot I_{\text{Tö}}^{00\to 05} \cdot I_{\text{Tö}}^{05\to 10} \cdot I_{\text{Tö}}^{10\to 15} \cdot I_{\text{La}}^{15\to t},$$

with

$$\lambda_t = \begin{cases} 0 & \text{for } s = 1, 2, \ldots, 12 \\ \dfrac{s - 12}{61 - 12} & \text{for } s = 13, 14, \ldots, 60. \end{cases}$$

In month $t = 1/95$, the value of the counter s is equal to 1. In the subsequent months it increases until in month $t = 12/99$ it reaches the value 60. In the following month, $t = 1/00$, the value of s is reset to 1. The same reset happens in months 1/05 and 1/10. Price changes within an expenditure reference year ($s = 1, 2, \ldots, 12$) are not revised ($\lambda_t = 0$).

The formulas generate the revised import price index $I_{\text{rev}}^{95\to t}$ depicted in Figure 2 (labelled as "rev"). The index deviates from the replicated index $I_{\text{repl}}^{95\to t}$ ("repl") and,

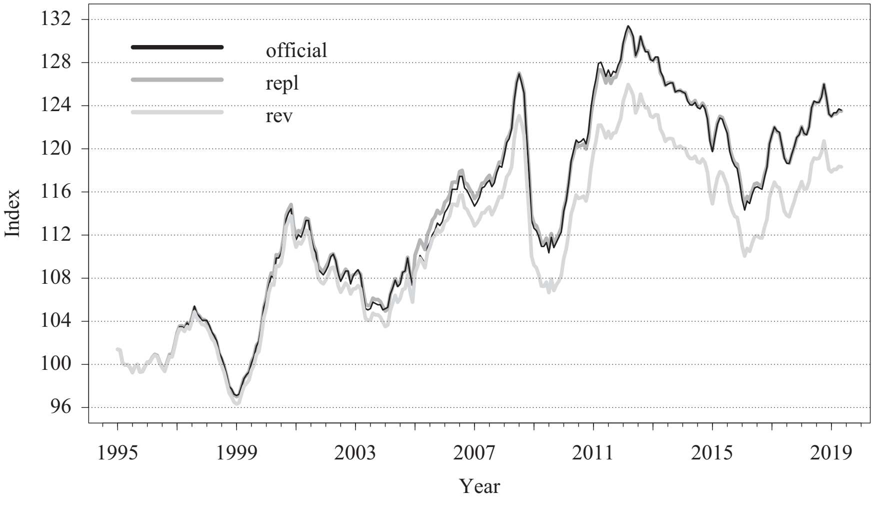


Fig. 2. German import price index (1995 = 100) for January 1995 to May 2019.
Source: Own calculations based on data of database "Genesis" of Statistisches Bundesamt (2022).

therefore, from the official index of Destatis ("official"). The deviation increases over time and, in 2019, reaches more than five percentage points. This value can be considered as a lower bound of the accumulated (upper-level) long-run substitution bias in the official price index.

This reinforces the case for a revision that does not stop at Stage 1, but includes also Stages 2 and 3. Only this comprehensive revision can avoid the accumulating long-run substitution bias inherent in chained Laspeyres indices. The revision requires no additional data. It exclusively draws on information that is used in the original price index compilation.

To this point, we were exclusively concerned with the import price index. If the export price index exhibited the same bias, the terms of trade index (ratio of the export price index and import price index) would remain unbiased. Unfortunately, such a compensatory effect is unlikely. The reason is illustrated in the upper part of Figure 3. It shows the export and import price indices of Germany as compiled by Destatis. Both indices are normalized to January 1980 = 100.

The export price index rises very evenly, while the import price index rises much more erratically. Similar patterns can be observed in France, Italy, and Japan. In the following, we explain why a more erratic index is more vulnerable to substitution bias. The line of reasoning starts with an empirical observation. Our data reveal that for a given pair of consecutive months the change in the import index (measured in percent where the sign is eliminated) is positively correlated with the coefficient of variation of the intertemporal price relatives of the 30 categories of German imports. For the time span for which we have price data on the 30 import categories (January 1995 to May 2019), the (Pearson) correlation coefficient is almost 0.54.

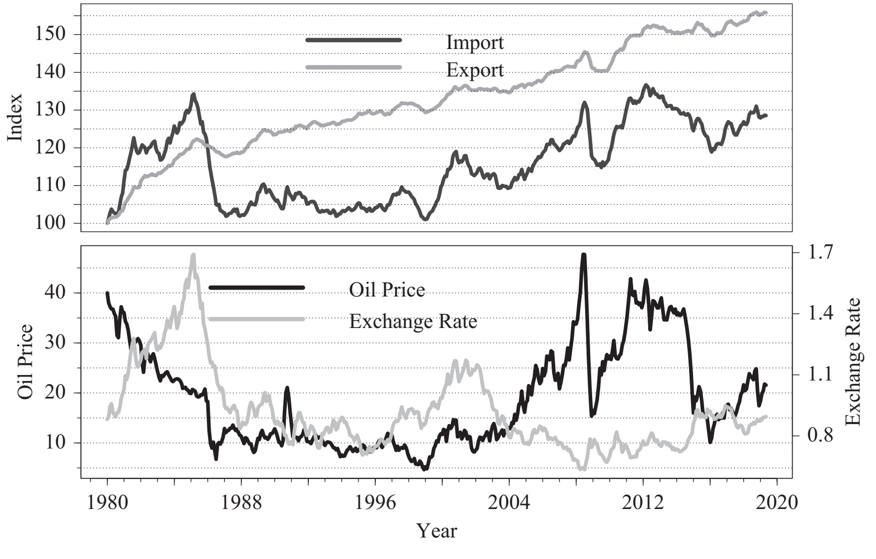


Fig. 3. German export and import price indices (January 1980 = 100), real oil prices (brent oil, in USD of 1980) and exchange rate (EUR/USD) for January 1980 to May 2019.

Source: Database "Genesis" of Statistisches Bundesamt (2022), Deutsche Bundesbank (2021), World Bank (2021) and U.S. Bureau of Labor Statistics (2021).

From the work of Von Bortkiewicz (1923, 374–376) we know that the relative divergence between the Laspeyres and Paasche index depends on three factors: the coefficient of variation of the intertemporal quantity relatives, the coefficient of variation of the intertemporal price relatives, and the coefficient of linear correlation between the price and quantity relatives. Thus, for a given correlation, the divergence between the Laspeyres and Paasche index tends to increase with the volatility of the prices and quantities. An increasing Paasche-Laspeyres spread translates into an increasing Törnqvist-Laspeyres spread, because the Törnqvist index closely approximates the geometric average of the Laspeyres and Paasche index (known as the Fisher index).

Since the Törnqvist-Laspeyres spread is interpreted as an indication of substitution bias, the previous considerations can be condensed to a simple conjecture: The larger the volatility of a Laspeyres index, the larger the risk of substitution bias. Therefore, the rather steadily evolving export price index is less likely to suffer from substitution bias than the much more volatile import price index.

To verify this conjecture, we first replicate the official export price index of Destatis, $E_{\text{repl}}^{95\to t}$, then compile the retroactively revised export price index, $E_{\text{rev}}^{95\to t}$, and finally compare the difference of the two indices to the difference that we computed in the context of the import price index. We follow the same procedures that we used for the import price index. The results are depicted in Figure 4.

We find our claim confirmed. The revision is smaller than in the import price index, indicating that also the accumulated long-run substitution bias in the official export price index is smaller. The direction of the bias is upwards, hinting at a negative correlation between price and quantity changes. This negative correlation is more in line with the behavior of purchasers than with the behavior of producers or sellers. In other words, our empirical findings related to the German export prices corresponds to the non-resident's perspective of economic theory (IMF 2009, 414).

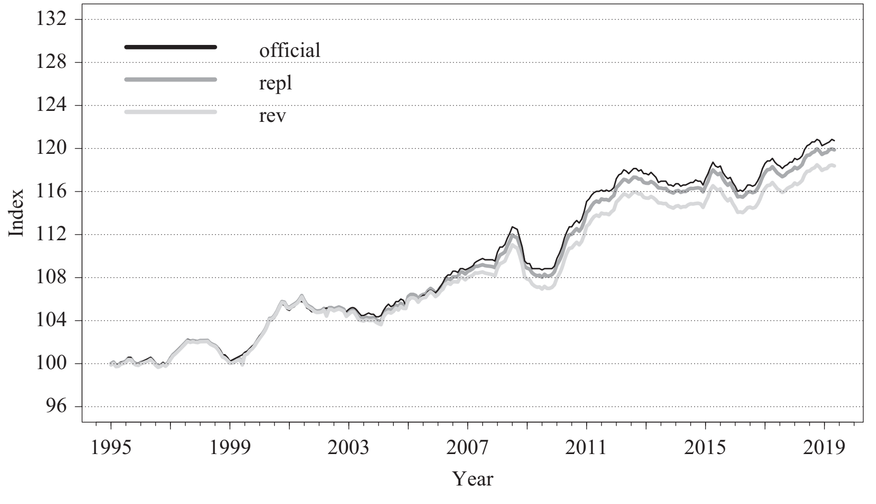


Fig. 4. German export price index (1995 = 100) from January 1995 to May 2019.
Source: Own calculations based on data of database "Genesis" of Statistisches Bundesamt (2022).

The preceding results have important implications for the German terms of trade index, that is, for the ratio of the export and import price index. Since the export price index exhibits only a minor upward bias, the substantial upward bias in the import price index translates into a substantial downward bias in the German terms of trade index. This is depicted in Figure 5. The terms of trade index compiled from the revised indices, $E_{\text{rev}}^{95\to t}$ and $I_{\text{rev}}^{95\to t}$, deviates from the terms of trade index compiled from the replicated indices, $E_{\text{repl}}^{95\to t}$ and $I_{\text{repl}}^{95\to t}$. The deviation suggests that in 2019 the accumulated bias in the official terms of trade reaches roughly four percentage points.

The bias in the terms of trade can be attributed to the difference in the volatility of the import and export price index. What causes the difference in volatility? Until the introduction of the euro in January 1999, German imports had to be converted into Deutsche Mark before they entered the import price index. As a consequence, the German import price index depended not only on the price changes in the countries of origin, but also on the exchange rate of the Deutsche Mark against the foreign currencies, most importantly against the USD. Therefore, until December 1998 we expect a strong positive correlation between the import price index and the nominal exchange rate in price notation, that is the price of one USD.

Figure 3 confirms this expectation. The gray line in the lower part of the graph (reference is the right axis) shows the euro-USD exchange rate in price notation, that is, the price of one USD expressed in euros (or in units of 1.95583 Deutsche Mark before 1999). Between January 1980 and December 1998, the correlation coefficient between the first differences of the import price index and the first differences of the exchange rate is almost 0.74.

Since January 1999 parts of the imports and exports are invoiced in Eurostat (2017) shows that in 2016 almost 50% of German imports from non-EU countries are invoiced in euro, while almost 60% of the exports in non-EU countries are invoiced in euro. This is likely to dampen the impact of the foreign exchange rate on the import price index.

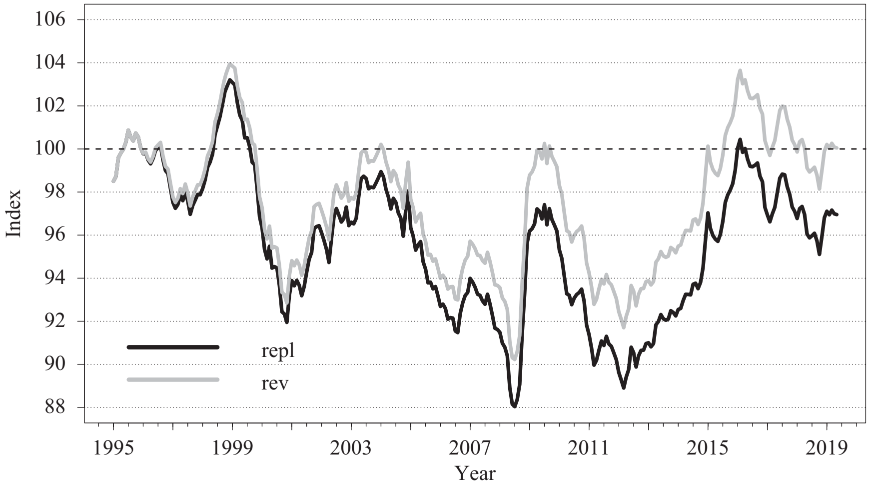


Fig. 5. German terms of trade (1995 = 100) from January 1995 to May 2019.
Source: Own calculations based on data of database "Genesis" of Statistisches Bundesamt (2022).

Figure 3 also confirms this second conjecture. For the time interval January 1999 to May 2019, the correlation between the first differences of the import price index and the first differences of the exchange rate is below 0.26. This indicates that also other factors must be responsible for the larger volatility of the import prices as compared to the export prices.

Obvious suspects are the prices of oil and gas. These two products represent almost ten percent of the German imports as compared to an export share of less than one percent (Peter 2019, 39–40). The black line in the lower part of Figure 3 depicts the evolution of the real oil prices (reference is the left axis). The strong volatility of the real oil prices and their positive correlation with the import price index are clearly visible. The correlation coefficient of the first differences of the import index and the first differences of the real oil price for the time interval January 1999 to May 2019 is almost 0.72 (before 1999 it is below 0.29).

5. Concluding Remarks

For a given pair of months, the extent of the substitution bias of Laspeyres type indices is positively correlated with the variation of the intertemporal price relatives, the variation of the intertemporal quantity relatives, and the linear correlation between the two. Chaining of such distorted indices leads to accumulated bias. In most countries, the official import and export price indices are compiled as chained Laspeyres type indices. Therefore, these indices are likely to suffer from substitution bias.

In the present study, we examined this conjecture. In a case study of the German trade statistics we compiled a lower bound for (upper-level) substitution bias in the German import price index and export price index. For the time interval January 1995 to May 2019 the accumulated upward substitution bias of the import price index is more than five percent, while the upward bias of the export price index is slightly above one percent.

Furthermore, we demonstrated that the accumulating substitution bias can be easily avoided by a three-stage retroactive revision process. This process requires no additional data and is adaptable to the specific index compilation procedures of the various national statistical offices.

The longer the intervals between the surveys providing the quantity data of the import and export price indices, the larger the expected substitution bias. Several countries rely on five-year-intervals (e.g., Canada, Germany, Italy, Japan, UK). Such countries are most likely to benefit from the implementation of a retroactive revision process. However, also in countries with shorter intervals (e.g., France, US) the proposed revision process may help to improve the reliability of the published long-run indices.

In countries like France, Germany, Italy, and Japan the import price index is considerably more volatile than the export price index. For Germany, we showed that a larger volatility of the index translates into a larger variation of the price and quantity relatives and this results in a larger substitution bias. In other words, the substitution bias of the import price index is likely to exceed the bias of the export price index. As a consequence, also the terms of trade index should be biased. Our empirical analysis confirmed this conjecture for the German case.

What causes the difference in the volatility of the import and export prices? One likely reason are the strong fluctuations in the prices of oil and gas. These two products are all but absent from the exports of France, Germany, Italy and Japan, but they represent a substantial share of these countries' imports. In some cases, exchange rate fluctuations may also play a role (e.g., Germany before the introduction of the euro).

Our retroactive correction approach can also be applied to other important areas of price measurement such as the Consumer Price Index or the Producer Price Index.

6. Appendix: Retroactive Correction in Practice

In this appendix we describe how our three-stage revision process can be applied to the trade statistics of Destatis. The monthly import and export price indices of Destatis are compiled in two variants. One is based on price data collected from exporting producers and wholesale traders, while the other is compiled from customs sources. Gehle (2013, 932) and Von der Lippe and Mehrhoff (2010) show that the two variants generate different results. Following the general recommendation of the IMF (2009, xiv), we study the variant based on price data. Until December 2004 we had to excerpt the price levels from printed publications of Destatis (Statistisches Bundesamt 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004). Later price levels we could retrieve from the online data base "Genesis" provided by Statistisches Bundesamt (2022). In addition, we could compute from the online data base the categories' expenditure weights for the years 1995, 2000, 2005, 2010, and 2015.

We begin with the monthly German import price index from January 2010 until July 2018. The official compilation method of Destatis (a monthly Laspeyres index) is sketched out in Statistisches Bundesamt (2019, 6–7). Destatis knows the yearly expenditure weights of the year 2010 (shorthand notation: 10). Therefore, the price, expenditure, and index reference period is not a month, but the year 2010, that is, 2010 = 1. The corresponding Laspeyres index is

$$I_{\mathrm{La}}^{10\to t} = \sum_{i=1}^{M} \bar{s}_i^{10} \frac{p_i^t}{\bar{p}_i^{10}}, \qquad t = 1/10, 2/10, \ldots, 7/18, \tag{15}$$

with

$$\bar{s}_i^{10} = \frac{\bar{p}_i^{10}\bar{m}_i^{10}}{\sum_{j=1}^{M} \bar{p}_j^{10}\bar{m}_j^{10}},$$

where $\bar{p}_i^{10}$ denotes the average euro price of good i in 2010 and $\bar{m}_i^{10}$ denotes the total imported quantity of good i in 2010. Therefore, $\bar{s}_i^{10}$ is the import share of good i in 2010.

In July 2018, the expenditure weights of the year 2015 become available. Therefore, Destatis revises the Laspeyres indices calculated for January 2015 to July 2018. The revised series uses the year 2015 as price, expenditure, and index reference period:

$$I_{\mathrm{La}}^{15\to t} = \sum_{i=1}^{M} \bar{s}_i^{15} \frac{p_i^t}{\bar{p}_i^{15}}, \qquad t = 1/15, 2/15, \ldots, 7/18. \tag{16}$$

This new series must be connected to the index numbers computed by the Laspeyres index (15). Direct chaining does not work here, because the comparison period of the Laspeyres index (15) is a month (for chaining, it would be January 2015, that is, $t = 1/15$), while the index reference period of the Laspeyres index (16) is the year 2015. Therefore, Destatis introduces the following Paasche index that binds the price level of January 2015 to the price level of the year 2015:

$$I_{\mathrm{Pa}}^{1/15\to 15} = \frac{\sum_{i=1}^{M} \bar{p}_i^{15} \bar{m}_i^{15}}{\sum_{j=1}^{M} p_j^{1/15} \bar{m}_j^{15}} = \left(\sum_{i=1}^{M} \bar{s}_i^{15} \left(\frac{p_i^t}{p_i^{1/5}} \right)^{-1} \right)^{-1} = \frac{1}{I_{\mathrm{La}}^{15\to 1/15}}. \tag{17}$$

This Paasche index is more convenient than the Laspeyres index $I_{\mathrm{La}}^{1/15\to 15}$, because the Paasche index requires the readily available expenditure weights of the year 2015, while the Laspeyres index would require the unknown expenditure weights of January 2015.

Based on the three Equations (15), (16) and (17), Destatis calculates a consistent series of monthly index numbers covering the time span January 2010 to July 2018. To this end, Destatis multiplies the price index numbers of January 2010 to December 2014 compiled by Equtaion (15) by the so-called "Verkettungsfaktor" $[I_{\mathrm{La}}^{10\to 1/15} \cdot I_{\mathrm{Pa}}^{1/15\to 15}]^{-1}$ (e.g., Statistisches Bundesamt, 2017, 6, 2019, 6). Note that the first factor of the "Verkettungsfaktor" is index (15) with comparison period $t = 1/15$ and the second factor is index (17). The Verkettungsfaktor is the reciprocal of the price change between the years 2010 and 2015. The index numbers of January 2015 to July 2018 are directly computed by Equtaion (16). The index reference period of the resulting series is the year 2015, that is, $2015 = 1$.

To rebase this series to the index reference year 2010, we multiply the whole series by $[I_{\mathrm{La}}^{10\to 1/15} \cdot I_{\mathrm{Pa}}^{1/15\to 15}]$, that is, by the inverse of the "Verkettungsfaktor". As a result, the index numbers of January 2010 to December 2014 are compiled by Equation (15), while the index numbers of January 2015 to July 2018 are compiled by:

$$I^{10\to t} = I_{\mathrm{La}}^{10\to 1/15} \cdot I_{\mathrm{Pa}}^{1/15\to 15} \cdot I_{\mathrm{la}}^{15\to t}, \qquad t = 1/15, 2/15, \ldots, 7/18. \tag{18}$$

For the comparison month $t = 1/15$, the Laspeyres index (15) and Equaiton (18) give $I_{\mathrm{La}}^{10\to 1/15}$ – see the last equality in Equation (17). Therefore, Equations (15) and (18) generate a consistent time series with the reference price level $2010 = 1$. This series differs from the official time series only by a constant factor, namely the "Verkettungsfaktor".

Equation (15) is a Laspeyres index and, therefore, prone to upward substitution bias. Equation (18) is a chain index comprising two upwardly biased Laspeyres indices and a Paasche index. The Paasche index is prone to downward bias. However, the bias is probably much smaller than the combined upward bias of the two Laspeyres indices because the time distance between January 2015 and the full year 2015 (Paasche index) is much smaller than that between the year 2010 and January 2015 or later months (Laspeyres indices). Therefore, not only Equation (15), but also Equation (18) is likely to exhibit severe upward bias (this was empirically confirmed in Section 4).

When the expenditure weights of the year 2015 have become available to Destatis, it is possible to reduce the substitution bias. To this end, we conduct the retroactive three-stage revision process outlined in Section 3. To adapt this process to the methodology of Destatis, only one modification is necessary. It relates to the denominator of the correction factor in Stage 3.

Stage 1: Using Equation (16), a new series of Laspeyres indices for January 2015 to July 2018 can be compiled. This part of the revision process is already implemented in the official index compilations of Destatis.

Stage 2: To rebase the new series of Laspeyres index numbers compiled in Stage 1 to the price level 2010 = 1, we compute the following Törnqvist index:

$$I_{\text{Tö}}^{10\to 15} = \exp\left(\sum_{i=1}^{M} \frac{1}{2}\left(\bar{s}_i^{10} + \bar{s}_i^{15}\right) \ln\left(\frac{\bar{p}_i^{15}}{\bar{p}_i^{10}}\right)\right). \tag{19}$$

Next, we replace in the chain index (18) the first two links by the Törnqvist index (19) and obtain the following chain index:

$$I^{10\to t} = I_{\text{Tö}}^{10\to 15} \cdot I_{\text{La}}^{15\to t}, \qquad t = 1/15, 2/15, \ldots, 7/18. \tag{20}$$

For January 2015 to July 2018, this chain index yields more reliable results than the chain index (18).

Stage 3: We want to revise the index numbers of January 2010 to December 2014 compiled by the Laspeyres index (15). The new series of index numbers must be consistent with the revised index number of January 2015 compiled by the chain index (20). Therefore, we multiply the Laspeyres index (15) by a correction factor that is constructed in analogy to the correction factor in Equation (13). The result is the following index formula:

$$I^{10\to t} = I_{\text{La}}^{10\to t} \cdot \left(\frac{I_{\text{Tö}}^{10\to 15}}{I_{\text{La}}^{10\to 1/15} \cdot I_{\text{Pa}}^{1/15\to 15}}\right)^{\lambda_t}, \qquad t = 1/10, 2/10, \ldots, 12/14, \tag{21}$$

with

$$\lambda_t = \begin{cases} 0 & \text{for} \quad s = 1, 2, \ldots, 12 \\ \dfrac{s-12}{61-12} & \text{for} \quad s = 13, 14, \ldots, 61. \end{cases} \tag{22}$$

For the twelve months of the year 2010 ($s = 1, 2, \ldots, 12$) the impact λ_t is 0. Therefore, the correction factor has the value 1, that is, no correction of the Laspeyres index $I_{\text{La}}^{10\to t}$ occurs until December 2010. This is intended because only the expenditure weights of the year 2010 and not the expenditure weights of the year 2015 should be included in the calculation of price changes *within* the year 2010. In January 2011 ($s = 13$) the impact λ_t is equal to 1/49. The correction factor then deviates minimally from 1, leading to a small correction of the Laspeyres index $I_{\text{La}}^{10\to 1/11}$. As t increases, the counter s and the impact λ_t also gradually increase. Only in January 2015, the counter s would reach its maximum value 61 and, therefore, the impact λ_t its maximum value 1. In this last month, Equation (21) would simplify to

$$I^{10\to 1/15} = \frac{I_{\text{Tö}}^{10\to 15}}{I_{\text{Pa}}^{1/15\to 15}}$$

which, in view of the last equality in Equation (17), is identical to Equation (20). Therefore, using Equation (21) for the comparison months January 2010 to December

2014, and using Equation (20) for all subsequent months, generates a consistent time series of price levels.

Quite likely, in 2023 the expenditure weights for the year 2020 will become available. Since the expenditure weights of the year 2020 add no relevant information for measuring the price changes between 2010 and 2015, there is no need for a second revision of the index numbers of January 2010 to December 2014.

The new import price index numbers of January 2020 and all subsequent months (Stages 1 and 2) are compiled by the chain index

$$I^{10\rightarrow t} = I_{\text{Tö}}^{10\rightarrow 15} \cdot I_{\text{Tö}}^{15\rightarrow 20} \cdot I_{\text{La}}^{20\rightarrow t}, \qquad t = 1/20, 2/20, \ldots . \tag{23}$$

The revised import price index numbers of January 2015 to December 2019 (Stage 3) are calculated by the chain index

$$I^{10\rightarrow t} = I_{\text{Tö}}^{10\rightarrow 15} \cdot I_{\text{La}}^{15\rightarrow t} \cdot \left(\frac{I_{\text{Tö}}^{15\rightarrow 20}}{I_{\text{La}}^{15\rightarrow 1/20} \cdot I_{\text{Pa}}^{1/20\rightarrow 20}} \right)^{\lambda_t}, \qquad t = 1/15, 2/15, \ldots, 12/19,$$

where the impact λ_t is defined as in Equation (22) with $s = 1$ in January 2015. These computations generate a consistent time series of index numbers with the index reference period 2010. The series covers the time interval from January 2010 to 2023 and beyond.

The compilation and revision of the export price index can be conducted in a perfectly analogous manner. The ratio of the revised export price index and import price index of some month t yields the revised terms of trade index of that month.

7. References

Bank of Japan. 2019. Corporate Goods Price Index (2015 base). Available at: http://www.boj.or.jp/en/statistics/pi/cgpi_2015/index.htm (accessed July 2019).

De Haan, J., B. Balk, and C. Hansen. 2010. "Retrospective Approximations of Superlative Price Indexes for Years Where Expenditure Data Is Unavailable." In: *Price Indexes in Time and Space. Contributions to Statistics*, edited by L. Biggeri, and G. Ferrari: 25–42, Physica-Verlag, Heidelberg. Available at: https://doi.org/10.1007/978-3-7908-2140-6_2 (accessed September 2021).

Deutsche Bundesbank. 2022. Deutsche Bundesbank Statistics. Available at: https://www.bundesbank.de/en/statistics/exchange-rates (accessed February 2022).

Dridi, J., and K. Zieschang. 2004. "Export and Import Price Indices." *IMF Staff Papers*, 51(1): 157–194. DOI: https://doi.org/10.2307/30035869.

Eurostat. 2017. *Euros or dollars? EU trade in goods by currency*. Available at: https://ec.europa.eu/eurostat/en/web/products-eurostat-news/-/DDN-20170627-1 (accessed November 2020).

Gehle, S. 2013. "Methodenvergleich zwischen Preisindizes und Durchschnittswertindizes im Außenhandel." *Wirtschaft und Statistik*, 1(10): 930–940. Available at: https://rn.bpb.de/system/files/dokument_pdf/destatis_Methodenvergleich_Indizes_Aussenhandel.pdf (accessed June 2021).

IMF. 2009. *Export and Import Price Index Manual: Theory and Practice*. DOI: https://doi.org/10.5089/9781589067806.069.

INSEE. 2019. *Producer price indices of french industry – all markets*. Available at: https://www.insee.fr/en/metadonnees/source/fichier/MET_PPI_BASE2015_EN.pdf (accessed January 2020).

Istat. 2019. *Foreign trade and industrial import prices*. Available at: https://www.istat.it/en/archivio/229630 (accessed January 2020).

OECD. 2022. *Terms of trade (indicator)*. DOI: https://doi.org/10.1787/7722246c-en.

ONS. 2017. *Producer price indices QMI*. Available at: https://www.ons.gov.uk/economy/inflationandpriceindices/methodologies/producerpriceindicesqmi (accessed January 2020).

Peter, F. 2009. "Die Indizes der Außenhandelspreise auf Basis 2005." *Wirtschaft und Statistik*, (7): 711–723. Available at: https://www.destatis.de/DE/Methoden/WISTA-Wirtschaft-und-Statistik/2009/07/aussenhandel-basis-2005-072009. html (accessed June 2021).

Peter, F. 2014. "Die Indizes der Außenhandelspreise auf Basis 2010." *Wirtschaft und Statistik*, (4): 262–282. Available at: https://www.destatis.de/DE/Methoden/WISTA-Wirtschaft-und-Statistik/2014/04/aussenhandel-basis-2010-042014.pdf?blob = publicationFile&v = 4 (accessed June 2021).

Peter, F. 2019. "Die Indizes der Außenhandelspreise auf Basis 2015." *Wirtschaft und Statistik*, (3): 36–53. Available at: https://www.destatis.de/DE/Methoden/WISTA-Wirtschaft-und-Statistik/2019/03/aussenhandels-grosshandelsverkaufspreise-basis-2015-032019.pdf?blob = publicationFile (accessed June 2021).

Pötzsch, K. 2004. "Die Indizes der Außenhandelspreise auf Basis 2000." *Wirtschaft und Statistik*, (5): 511–524. Available at: https://www.statistischebibliothek.de/mir/servlets/MCRFileNodeServlet/DEAusgabe_derivate_00000018/1010200041054.pdf;jsessionid = CE66647CC6224FAB686F48868950B3B7 (accessed June 2021).

Statistics Canada 2019. *International Merchandise Trade Price Index*. Available at: https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2203 (accessed January 2020).

Statistisches Bundesamt. 1995. *Preisindizes für die Einfuhr/Ausfuhr, Thematische Veröffentlichungen*, Fachserie 17, Reihe 8.1/8.2 Preise.

Statistisches Bundesamt. 1996. *Preisindizes für die Einfuhr/Ausfuhr, Thematische Veröffentlichungen*, Fachserie 17, Reihe 8.1/8.2 Preise.

Statistisches Bundesamt. 1997. *Preisindizes für die Einfuhr/Ausfuhr, Thematische Veröffentlichungen*, Fachserie 17, Reihe 8.1/8.2 Preise.

Statistisches Bundesamt. 1998. *Preisindizes für die Einfuhr/Ausfuhr, Thematische Veröffentlichungen*, Fachserie 17, Reihe 8.1/8.2 Preise.

Statistisches Bundesamt. 1999. *Preisindizes für die Einfuhr/Ausfuhr, Thematische Veröffentlichungen*, Fachserie 17, Reihe 8.1/8.2 Preise.

Statistisches Bundesamt. 2000. *Preisindizes für die Einfuhr/Ausfuhr, Thematische Veröffentlichungen*, Fachserie 17, Reihe 8.1/8.2 Preise.

Statistisches Bundesamt. 2001. *Preisindizes für die Einfuhr/Ausfuhr, Thematische Veröffentlichungen*, Fachserie 17, Reihe 8.1/8.2 Preise.

Statistisches Bundesamt. 2002. *Preisindizes für die Einfuhr/Ausfuhr, Thematische Veröffentlichungen*, Fachserie 17, Reihe 8.1/8.2 Preise.

Statistisches Bundesamt. 2003. *Preisindizes für die Einfuhr/Ausfuhr, Thematische Veröffentlichungen*, Fachserie 17, Reihe 8.1/8.2 Preise.

Statistisches Bundesamt. 2004. *Preisindizes für die Einfuhr/Ausfuhr, Thematische Veröffentlichungen*, Fachserie 17, Reihe 8.1/8.2 Preise.

Statistisches Bundesamt 2017. *Preise und Preisindizes für gewerbliche Produkte (Erzeugerpreise)*, Fachserie 17, Reihe 2. Available at: https://www.statistischebibliothek.de/mir/receive/DESerie_mods_00000239 (accessed April 2020).

Statistisches Bundesamt 2019. *Index der Ein- und Ausfuhrpreise – Zusammenfassungen, Terms of Trade*, Fachserie 17, Reihe 8.1.

Statistisches Bundesamt. 2022. *Genesis. Die Datenbank des Statistischen Bundesamtes.* Available at: https://www-genesis.destatis.de/genesis/online (accessed February 2022).

U.S. Bureau of Labor Statistics. 2018a. *Import, Export Price Indexes*. Available at: https://www.bls.gov/mxp/ippfaq.htm (accessed February 2019).

U.S. Bureau of Labor Statistics. 2018b. *Terms of Trade Indexes*. Available at: https://www.bls.gov/mxp/terms-of-trade.htm (accessed February 2019).

U.S. Bureau of Labor Statistics. 2020. *Handbook of Methods. International Price Program: Overview*. Available at: https://www.bls.gov/opub/hom/ipp/home.htm (accessed November 2020).

U.S. Bureau of Labor Statistics. 2021. *Consumer Price Index (CPI) Databases of U.S. Bureau of Labor Statistics*. Available at: https://www.bls.gov/cpi/data.htm (accessed June 2021).

Von Auer, L., and A. Shumskikh. 2022. *Retrospective Computations of Price Index Numbers: Theory and Application.* Research Papers in Economics, No. 1/22, Universität Trier. Available at: https://www.uni-trier.de/index.php?id=51049 (accessed February 2022).

Von Bortkiewicz, L. 1923. Zweck und Struktur einer Preisindexzahl, erster Artikel. *Nordisk Statistik Tidskrift*, 2: 369–408.

Von der Lippe, P., and J. Mehrhoff. 2010. Unit Value Bias Reconsidered. *Paper presented at the 11th Meeting of the Ottawa Group.* Available at: https://www.econstor.eu/bitstream/10419/37526/1/VfS_2010_pid_211.pdf (accessed June 2021).

World Bank. 2021. World Bank *Commodity Price Data* (The Pink Sheet). Available at: https://www.worldbank.org/en/research/commodity-markets (accessed June 2021).

Received June 2020
Revised November 2020
Accepted June 2021

Journal of Official Statistics, Vol. 38, No. 1, 2022, pp. 127–151, http://dx.doi.org/10.2478/JOS-2022-0007

Rolling-Time-Dummy House Price Indexes: Window Length, Linking and Options for Dealing with Low Transaction Volume

Robert J. Hill[1], *Michael Scholz*[2], *Chihiro Shimizu*[3], *and Miriam Steurer*[1]

Rolling-time-dummy (RTD) is a hedonic method used by a number of countries to compute their official house price indexes (HPIs). The RTD method requires less data and is more adaptable than other hedonic methods, which makes it well suited for computing higher frequency HPIs (e.g., monthly or weekly). In this article, we address three key issues relating to RTD. First, we develop a method for determining the optimal length of the rolling window. Second, we consider variants on the standard way of linking the current period with earlier periods, and show how the optimal linking method can be determined. Third, we propose three ways of modifying the RTD method to make it more robust to periods of low transaction volume. These modifications could prove useful for countries using the RTD method in their official HPIs.

Key words: House price index; hedonic quality adjustment; optimal window length; optimal chain linking; higher frequency indexes; low transaction volume.

1. Introduction

The housing market and the broader economy are closely connected. While it is true that economic booms and recessions can trigger booms and busts in the housing market, the causation can also run in the opposite direction. The global financial crisis of 2007–2010 was a case in point. For central banks to effectively maintain financial stability, it is therefore important to have reliable and timely house price indexes (HPIs).

To effectively distinguish between genuine price changes and compositional differences, HPIs are typically computed using hedonic methods. The hedonic approach entails estimating shadow prices on the characteristics of properties (such as floor area, age, and location) so as to ensure that quality is held fixed when measuring price changes from one period to the next. For example, Eurostat recommends that countries in Europe should compute their official HPIs using hedonic methods (Eurostat 2016).

A number of hedonic methods for constructing HPIs have been proposed in the literature (see Hill 2013, for an overview). One hedonic method that has been attracting increased attention in recent years is the rolling-time-dummy (RTD) method. It was first

[1] Department of Economics, University of Graz, Universitätsstraße 15/F4, 8010 Graz, Austria, Email: robert.hill@uni-graz.at, and miriam.steurer@uni-graz.at
[2] Department of Economics, University of Klagenfurt, Universitätsstraße 65-67, 9020 Klagenfurt, Austria. Email: michael.scholz@aau.at
[3] University of Tokyo, Center for Spatial Information Science, 5-1-5, Kashiwanoha, Kashiwa, Chiba, 277-8568, Japan. Email: cshimizu@csis.u-tokyo.ac.jp
Acknowledgments: We acknowledge financial support for this project from the Austrian Research Promotion Agency (FFG), grant #10991131. We also wish to thank Niall O'Hanlon and three anonymous referees for helpful comments.

© Statistics Sweden

proposed by Shimizu in 1998 as part of a project entitled Construction of Property Price Indexes using Big Data 1998–2002 at Reitaku University (Shimizu et al. 2003, 2010).

The RTD method has four desirable properties. First, it is relatively simple to compute and interpret. Second, it requires less data than some other hedonic methods, such as the hedonic imputation or average characteristics methods. This is because it pools the data of multiple periods when estimating the characteristic shadow prices, which allows them to be measured with greater accuracy. RTD is particularly useful for smaller countries that have less data. In Europe, it is used by Croatia, Cyprus, France, Ireland and Portugal to construct their official HPIs (Hill et al. 2018). Japan has recently decided to compute its official residential and commercial property price indexes using the RTD method (Shimizu and Diewert 2019). Also, Brunei Darussalam (https://www.ambd.gov.bn/SiteY/o20AssetsY/o20Y/o20News/RPPI-Technical-Notes.pdf), Peru and Thailand (see https://www.bot.or.th/App/BTWS_STAT/statistics/DownloadFile.aspx?file = EC_EI_008_S2_ENG.PDF) are using RTD, and Indonesia is about to start using it (see Rachman 2019). RTD's effectiveness with smaller data sets means that it is also a good candidate for computing higher frequency indexes, such as monthly or weekly.

Third, an index provider using the RTD method can choose the length of the rolling window. A longer window increases the robustness of the index, which can be important when the data set is small, while a shorter window increases the current market relevance of the index. Index providers can trade off these two aspects when choosing the window length. In Europe, France and Portugal use a two-quarter rolling window, In Europe, Cyprus and Croatia use the RTD method with a four-quarter window, and Ireland a 12-month window, while France and Portugal employ a two-quarter window (see Hill et al. 2018). Note that the two-period RTD is also referred to as the adjacent-period method (Triplett 2004) or as the chained two-period time-dummy method. These choices are consistent with the idea that smaller countries should choose longer windows.

Fourth, an RTD index is not revised when new periods are added to the data set. This avoids confusion among users. By contrast, the time-dummy method violates the non-revisability criterion.

In this article, we address three key issues relating to the RTD hedonic method. First, there is the question of how one determines the optimal window length for any given data set? We develop an approach for answering this question and use it to compute the optimal window length for weekly HPIs in Sydney and Tokyo.

Second, the standard version of the RTD method links the current period to the period directly preceding it. It turns out this is just one of many ways that the HPI can be computed from the estimated hedonic model. We compare a number of other ways of linking in the current period, and develop an approach for determining which linking method is optimal. Our approach is then again tested on Sydney and Tokyo data. As similar linking issues arise in the scanner-data literature, we then briefly discuss the parallels between the HPI and scanner-data literatures.

Third, periods of low transaction volume can generate weak links in the RTD HPI, potentially undermining the integrity of the whole time series. We propose three ways of modifying the RTD method to make it more robust to weak links. We illustrate the problem using Sydney data. These modifications could prove useful for countries using the RTD method in their official HPIs.

Our focus is on weekly indexes. A trade-off exists between reliability and timeliness when choosing the frequency of a house price index. Higher frequency indexes, such as weekly indexes, are useful when timeliness is important, for example, when central banks make their monetary policy decisions. However, it should be noted that transaction prices are often available only after a time lag of many weeks or even months (see, for example, Shimizu et al. 2016). In such cases, a weekly index has less appeal. Higher frequency indexes, such as weekly indexes, may therefore need to be computed using list price data, which are available without any lag directly from property listing websites.

2. The Rolling Time Dummy (RTD) Method

2.1. *The Standard Version*

The RTD method estimates a hedonic model that includes the data of a fixed number of time periods, with time dummies included for each period (except the base period). Price indexes are derived from the estimated coefficients on these time dummies. The model is then moved forward one-period and re-estimated. The overall RTD price index is constructed by chaining together the prices indexes from these rolling windows.

More specifically, consider the standard version of the RTD method with a window length of $k + 1$ periods, as defined in Shimizu et al. (2010) and O'Hanlon (2011). Supposing that the first period in the window is period t, the first step is to estimate a semilog hedonic model as follows:

$$\ln p_{\tau h} = \sum_{c=1}^{C} \beta_c z_{\tau ch} + \sum_{s=t+1}^{t+k} \delta_s d_{\tau sh} + \varepsilon_{\tau h}, \tag{1}$$

where h indexes the housing transactions that fall in the rolling window, $p_{\tau h}$ the transaction price of property h in time period τ (where $t \leq t \leq t + k$), c indexes the set of available characteristics of the transacted dwellings, and ε is an identically, independently distributed error term with mean zero. The characteristics of the dwellings are given by the $z_{\tau ch}$, while $d_{\tau sh}$ is a dummy variable that equals 1 when $\tau = s$, and zero otherwise.

Estimating this model using ordinary least-squares, the change in the price index from period $t + k-1$ to period $t + k$ is then calculated as follows:

$$\frac{P_{t+k}}{P_{t+k-1}} = \frac{\exp(\hat{\delta}^t_{t+k})}{\exp(\hat{\delta}^t_{t+k-1})}, \tag{2}$$

where $\hat{\delta}$ denotes the least squares estimate of δ. A superscript t is included on the estimated δ coefficients to indicate that they are obtained from the hedonic model with period t as the base (i.e., $P_t = 1$). As can be seen from Equation (2), the hedonic model with period t as the base is only used to compute the change in house prices from period $t + k-1$ to period $t + k$. The window is then rolled forward one period and the hedonic model is re-estimated. The change in house prices from period $t + k$ to period $t + k + 1$ is now computed as follows:

$$\frac{P_{t+k+1}}{P_{t+k}} = \frac{\exp(\hat{\delta}^{t+1}_{t+k+1})}{\exp(\hat{\delta}^{t+1}_{t+k})}, \tag{3}$$

where now the base period in the hedonic model is period $t+1$. The price index over multiple periods is computed by chaining these bilateral comparisons together as follows:

$$\frac{P_{t+k+1}}{P_t} = \left[\frac{\exp(\hat{\delta}^{t-k}_{t+1})}{\exp(\hat{\delta}^{t-k}_{t})}\right]\left[\frac{\exp(\hat{\delta}^{t-k+1}_{t+2})}{\exp(\hat{\delta}^{t-k+1}_{t+1})}\right] \times \cdots \times \left[\frac{\exp(\hat{\delta}^{t+1}_{t+k+1})}{\exp(\hat{\delta}^{t+1}_{t+k})}\right]. \tag{4}$$

An important feature of the RTD method is that once a price change P_{t+k}/P_{t+k-1} has been computed, it is never revised. Hence when data for a new period $t+k+1$ becomes available, the price indexes P_t, $P_{t+1}, \ldots, P_{t+k}$ are already fixed. The sole objective when re-estimating the hedonic model to include period $t+k+1$ is to compute P_{t-k+1}/P_{t+k}.

2.2. *Linking Variants on the RTD Method*

Instead of always focusing on the last two estimated δ coefficients in each hedonic model, an alternative would be to focus on the last and third last coefficients. In this case, the price change from period $t+k-1$ to period $t+k$ could be calculated as follows:

$$\frac{P_{t+k}}{P_{t+k-1}} = \left(\frac{P_{t+k-2}}{P_{t+k-1}}\right)\frac{\exp(\hat{\delta}^{t}_{t+k})}{\exp(\hat{\delta}^{t}_{t+k-2})}, \tag{5}$$

where as has been noted above both P_{t+k-1} and P_{t+k-2} are already fixed by the time the data for period $t+k$ becomes available. Another alternative is the following:

$$\frac{P_{t+k}}{P_{t+k-1}} = \left(\frac{P_{t+k-3}}{P_{t+k-1}}\right)\frac{\exp(\hat{\delta}^{t}_{t+k})}{\exp(\hat{\delta}^{t}_{t+k-3})}, \tag{6}$$

and more generally,

$$\frac{P_{t+k}}{P_{t+k-1}} = \left(\frac{P_{t+k-j}}{P_{t+k-1}}\right)\frac{\exp(\hat{\delta}^{t}_{t+k})}{\exp(\hat{\delta}^{t}_{t+k-j})}, \tag{7}$$

where $j \leq k$. In Equation (5), the hedonic model is used to link each new period with two periods earlier. In Equation (6), each new period with three periods earlier, while in Equation (7), each new period with j periods earlier. In other words, given a window length of $k+1$ periods, there are k distinct ways of linking period $t+k$ with the earlier periods. Each will give a different answer, and one cannot say ex-ante that one is better than another.

Another possibility is to compute the geometric mean of these k sets of results as follows:

$$\frac{P_{t+k}}{P_{t+k-1}} = \prod_{j=1}^{k}\left[\left(\frac{P_{t+k-j}}{P_{t+k-1}}\right)\left(\frac{\exp(\hat{\delta}^{t}_{t+k})}{\exp(\hat{\delta}^{t}_{t+k-j})}\right)\right]^{1/k}. \tag{8}$$

This method uses each single-period link in turn to generate k distinct estimates of P_{t+k}/P_{t+k-1}, and then takes the geometric mean of these estimates.

A weighted geometric mean could also be computed, with more recent periods being given more weight. For example, the weights could decline geometrically as follows:

$$\frac{P_{t+k}}{P_{t+k-1}} = \prod_{j=1}^{k} \left\{ \left[\left(\frac{P_{t+k-j}}{P_{t+k-1}} \right) \left(\frac{\exp(\hat{\delta}^{t}_{t+k})}{\exp(\hat{\delta}^{t}_{t+k-j})} \right) \right]^{\dfrac{(1-\lambda)\lambda^{j-1}}{1-\lambda^{k}}} \right\}, \tag{9}$$

where $0 < \lambda < 1$. The idea here is that price index comparisons between closer together periods may be more accurate then comparisons between further apart periods.

There are close similarities here with the literature on constructing monthly or weekly price indexes for consumer goods using scanner data. Focusing on the case of monthly indexes, price indexes in this literature are often computed using a 13-month rolling window. The price index itself in each window is not necessarily computed using a hedonic method. Nevertheless, the same issue arises regarding how the current period should be linked to earlier periods. In this literature, the standard linking method used by the RTD method in Equation (2) is sometimes referred to as a *movement* splice (see, for example, De Haan 2015, and Chessa et al. 2017). Other possibilities considered are a *window* splice which links the new month in by comparing it with the corresponding month one year earlier (Krsinich 2016), and a mean splice, which is analogous to the geometric mean in Equation (8) (Diewert and Fox 2020). These methods are all discussed in De Haan et al. (2020). In addition, Melser (2018) proposes a weighted mean splice. Melser's method is similar in spirit to our weighted mean described in Equation (9), although the context is rather different. His weighted mean splice is a solution to a logarithmic weighted least squares problem focused on transitivizing bilateral price indexes. The weights are derived from the product overlaps between adjacent periods. No such equivalent weights exist in our context.

2.3. Low Transaction Periods

Many data sets exhibit pronounced seasonal fluctuations in the number of transactions. For example, each year in Sydney the number of transactions falls very significantly in December and January (see Subsection 5.6). This period of low transaction volume could create problems potentially causing a level shift or drift in the overall RTD price index. A weekly price index over this period will contain an unusually high level of noise. For example, suppose the price index for week 51 of 2020 contains a large positive random error. This will have a permanent impact on the RTD price index, causing it to drift upwards.

We consider three ways of mitigating the effect of low transactions volume on RTD house price indexes. Our starting point is that the desired window length is known and that special action is deemed necessary for any period that has less than N transactions.

2.3.1. Method 1

When computing the price index for period $t + k$, if any of earlier periods that are supposed to be in the window have less than N transactions, then these periods are deleted and replaced by the most recent available earlier period that has at least N transactions. If period $t+k$ has less than N transactions, the RTD method is still used to compute P_{t+k}/P_{t+k-1} (or $P_{t+k}/P_{\mathrm{t+k}-2}$ if period $t+k-1$ has less than N transactions). But period $t+k$ is then not used to compute the price indexes of later periods.

An example should help clarify the rule. Suppose the window length is set at three weeks, and that 2020, weeks 51, 52, and 2021 week 1 have less than N transactions.

Current period	**Periods included in the rolling window**
2020w50:	(2020w48, 2020w49, 2020w50)
2020w51:	(2020w49, 2020w50, 2020w51)
2020w52:	(2020w49, 2020w50, 2020w52)
2021w1:	(2020w49, 2020w50, 2021w1)
2021w2:	(2020w49, 2020w50, 2021w2)
2021w3:	(2020w50, 2021w2, 2021w3)
2021w4:	(2021w2, 2021w3, 2021w4)

The linking structure in this example is graphed in Figure 1. 2020w51, 2020w52, 2021w1 and 2021w2 are all linked into the price index via 2020w50. From then on, normal chronological chaining as described in Subsection 2.1 resumes.

In the scenario described above, the three-week rolling window price indexes are calculated as follows:

$$\frac{P_{2020w51}}{P_{2020w50}}=\frac{\exp\left(\hat{\delta}^{2020w49}_{2020w51}\right)}{\exp\left(\hat{\delta}^{2020w49}_{2020w50}\right)},\ \frac{P_{2020w52}}{P_{2020w50}}=\frac{\exp\left(\hat{\delta}^{2020w49}_{2020w52}\right)}{\exp\left(\hat{\delta}^{2020w49}_{2020w50}\right)},\ \frac{P_{2020w1}}{P_{2020w50}}=\frac{\exp\left(\hat{\delta}^{2020w49}_{2020w1}\right)}{\exp\left(\hat{\delta}^{2020w49}_{2020w50}\right)},$$

$$\frac{P_{2021w2}}{P_{2020w50}}=\frac{\exp\left(\hat{\delta}^{2020w49}_{2021w2}\right)}{\exp\left(\hat{\delta}^{2020w49}_{2020w50}\right)},\ \frac{P_{2021w3}}{P_{2021w2}}=\frac{\exp\left(\hat{\delta}^{2020w50}_{2021w3}\right)}{\exp\left(\hat{\delta}^{2020w50}_{2021w2}\right)},\ \frac{P_{2021w4}}{P_{2021w3}}=\frac{\exp\left(\hat{\delta}^{2021w2}_{2021w4}\right)}{\exp\left(\hat{\delta}^{2020w2}_{2020w3}\right)}.$$

In these equations, the superscript denotes the base week in each hedonic model and the subscript denotes the week of the estimated δ parameter.

The overall price index with 2020w50 normalized to 1 is then constructed as follows:

$$1,\frac{P_{2020w51}}{P_{2020w50}},\frac{P_{2020w52}}{P_{2020w50}},\frac{P_{2020w1}}{P_{2020w50}},\frac{P_{2021w2}}{P_{2020w50}},\frac{P_{2021w2}}{P_{2020w50}},\frac{P_{2021w3}}{P_{2021w2}},\frac{P_{2021w2}}{P_{2020w50}},\frac{P_{2021w3}}{P_{2021w2}},\frac{P_{2021w4}}{P_{2021w3}},\ldots$$

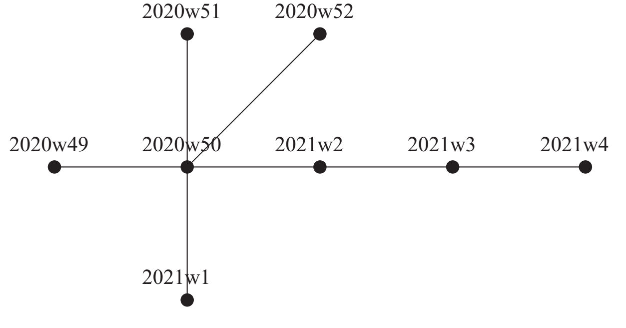


Fig. 1. The linking structure of method 1.

As can be seen the lack of data for 2020w51, 2020w52 and 2021w1 does not contaminate the longer time series RTD price index after 2021w1.

2.3.2. Method 2

When a period has fewer than *N* transactions, more transactions are drawn from the preceding period until the threshold *N* is reached. The chronologically closest transactions from the previous period are used first. If there are enough transactions in the last three days of the preceding period to reach the *N* threshold, then only these last three days are added to the current period, when computing the current period price index. These three days are still also included in the previous period.

As an example, suppose that each of 2020w51, 2020w52 and 2021w1 has more than *N*/2 and less than *N* transactions. The RTD rolling window would now be constructed as follows:

Current period	**Periods included in the rolling window**
2020w50:	(2020w48, 2020w49, 2020w50)
2020w51:	(2020w49, 2020w50, 2020w51^{+})
2020w52:	(2020w50, 2020w51^{+}, 2020w52^{+})
2021w1:	(2020w51^{+}, 2020w52^{+}, 2021w1^{+})
2021w2:	(2020w52^{+}, 2021w1^{+}, 2021w2)
2021w3:	(2021w1^{+}, 2021w2, 2021w3)
2021w4:	(2021w2, 2021w3, 2021w4)

The "+" superscripts above denote that the week is being supplemented with data from the previous week. Once each week with less than *N* transactions has been supplemented with transactions from the previous week, the RTD price index is computed in the standard way described in Subsection 2.1.

2.3.3. Method 3

Our third alternative approach is to not compute an index for a period with less than *N* transactions. Instead it is merged with the next period. If together these two periods reach the *N* transaction threshold, then they are treated as a single period. If together they still do not reach the *N* transaction threshold, then again no index is computed until the next period becomes available, d so on.

As an example, again suppose that each of 2020w51, 2020w52 and 2021w1 has more than *N*/2 and less than *N* transactions. The RTD rolling window would now be constructed as follows:

Current period	**Periods included in the rolling window**
2020w50:	(2020w48, 2020w49, 2020w50)
2020w51:	Missing
2020w51-w52:	(2020w49, 2020w50, 2020w51-w52)
2021w1:	Missing
2021w1-w2:	(2020w50, 2020w51-w52, 2021w1-w2)
2021w3:	(2020w51-w52, 2021w1-w2, 2021w3)
2021w4:	(2021w1-w2, 2021w3, 2021w4)
2021w5:	(2021w3, 2021w4, 2021w5)

All three methods ensure that periods of low transaction volume do not contaminate the index in later periods. Method 1 computes price indexes for each low transaction week using whatever transaction data are actually available. Method 2 supplements the data for low transaction weeks with data from the previous week (or more if required), while method 3 merges low transaction weeks and treats them as a single period if each individually does not contain enough transactions. Which method is best depends on the needs of users. If it is important that a price index is computed for every period (here weeks) then either method 1 or 2 should be used. Method 1 will better capture actual price movements during low transaction periods unless they become too distorted by noise arising from small sample sizes. Overall we prefer Method 1. We illustrate the impact of its use on the Sydney data set in Subsection 5.6.

3. Quarterly Benchmarks

3.1. The Hedonic Imputation Method

The hedonic imputation method is an alternative to the RTD method (see, for example Diewert 2011 and Hill 2013). We use the hedonic imputation method here as a reference index for assessing the performance of different versions of the RTD method.

The hedonic imputation approach estimates a separate hedonic model for each period:

$$\ln p_{t,h} = \beta_t \cdot z_{t,h} + \varepsilon_{t,h}, \tag{10}$$

where for convenience β_t and $z_{t,h}$ now both denote vectors. The hedonic model is then used to impute prices for individual houses. For example, let $\hat{p}_{t+1,h}(z_{t,h})$ denote the imputed price in period $t+1$ of a house with the characteristic vector $z_{t,h}$ sold in period t. This price is imputed by substituting the characteristics $z_{t,h}$, into the estimated hedonic model of period $t+1$ as follows:

$$\hat{p}_{t+1,h}(z_{t,h}) = \exp\left(\sum_{c=1}^{C} \hat{\beta}_{c,t+1} z_{c,t,h}\right). \tag{11}$$

With these imputed prices it is now possible to construct a matched sample, thus allowing standard price index formulas to be used. See Silver and Heravi (2007), Diewert et al. (2009), and Rambaldi and Rao (2013) for a discussion of some of the advantages of the hedonic imputation method.

$$\text{Geometric–Paasche Imputation}: \; P^{PI}_{t,t+1} = \prod_{h=1}^{H_{t+1}} \left[\left(\frac{\hat{p}_{t+1,h}}{\hat{p}_{t,h}(z_{t+1,h})}\right)^{1/H_{t+1}}\right] \tag{12}$$

$$\text{Geometric–Laspeyres Imputation}: \; P^{LI}_{t,t+1} = \prod_{h=1}^{H_t} \left[\left(\frac{\hat{p}_{t+1,h}(z_{t,h})}{\hat{p}_{t,h}}\right)^{1/H_t}\right] \tag{13}$$

$$\text{Törnqvist Imputation}: \; P^{TI}_{t,t+1} = \sqrt{P^{PI}_{t,t+1} \times P^{LI}_{t,t+1}} \tag{14}$$

In a comparison between periods t and $t+1$, the Geometric-Laspeyres index focuses on the H_t houses that sold in the earlier period t. Similarly the Geometric-Paasche index focuses on the H_{t+1} houses that sold in the later period $t+1$. These price indexes give equal weight to each house sold (see De Haan (2010) for a discussion on alternative weighting schemes). By taking the geometric mean of Geometric-Paasche and Geometric-Laspeyres, the Törnqvist index gives equal weight to both periods. The Geometric-Paasche, Geometric-Laspeyres and Törnqvist indexes above are of the double imputation variety, meaning that both prices in each price relative are imputed. A single imputation approach by contrast imputes only one price in each pair (since the actual price is always available for one of the two periods being compared). There has been some discussion in the literature over the relative merits of the two approaches (see, for example De Haan 2004; Hill and Melser 2008). Empirically we try both approaches. The resulting price indexes are virtually indistinguishable. Hence to simplify the presentation, we focus here only on double imputation price indexes. The hedonic imputation method allow the characteristic shadow prices to update each period.

In the context of weekly indexes, the hedonic imputation method is unlikely to work well since the sample sizes in many weeks may be too small to justify estimating a separate hedonic model each week. However, in our context, quarterly hedonic imputation will provide a useful benchmark for weekly RTD indexes.

3.2. *The Time-Dummy Method*

We also use the time-dummy index as a reference for assessing the performance of RTD weekly indexes. The time-dummy method is the limiting case of the RTD method where the window length is the same as the number of periods in the comparison. One disadvantage of the time-dummy method is that it violates non-revisability, with the effect that whenever a new period is added to the data set, all past price indexes are subject to change.

$$\ln p_{\tau h} = \sum_{c=1}^{C} \beta_c z_{\tau ch} + \sum_{t=2}^{T} \delta_t d_{\tau th} + \varepsilon_{\tau h}. \tag{15}$$

The price index for period t relative to period 1 is then calculated as follows:

$$\frac{P_t}{P_1} = \exp\left(\hat{\delta}_t\right). \tag{16}$$

3.3. *A Performance Criterion for Weekly Indexes Derived from Quarterly Indexes*

We propose a criterion here for determining the optimal window length and linking method for weekly RTD indexes, by comparing them with reference quarterly hedonic indexes. We consider two reference quarterly hedonic indexes: these are the hedonic imputation method and the time-dummy method described above. We focus on these two methods because they are quite different (i.e., one re-estimates the hedonic model every quarter while the other does not re-estimate at all). Using these quarterly indexes as benchmarks should avoid biasing the results towards any particular window length.

Empirically we find that the quarterly hedonic imputation and time-dummy methods approximate each other closely. By contrast for weekly RTD indexes, if we allow the window length to vary between two and 53 weeks, the range of possible results becomes much larger (see Subsection 5).

The greater sensitivity of weekly indexes to the choice of hedonic method makes them a more interesting focus of analysis than quarterly indexes. The choice of window length really matters for weekly RTD indexes. Furthermore, the greater robustness of quarterly indexes is a property we can exploit to discriminate between competing weekly RTD indexes.

The first step of our criterion for assessing the performance of alternative weekly RTD indexes is to construct a quarterly index from each weekly index. This can be done in the following way. Let $t = 1, \ldots, T$ indexing the quarters in the data set, and $v = 1, \ldots, V$ the 13 weeks in a quarter. A quarterly price index $P^w_{t,t+1}$ is obtained from a weekly price index as follows:

$$P^w_{t,t+1} = \prod_{v=1}^{13} \left(\frac{P_{t+1,v}}{P_{t,v}} \right)^{1/13}, \tag{17}$$

where $P_{t,v}$ denotes the level of the weekly price index in quarter t, week v. Each element $P_{t+1,v}/P_{t,v}$ in (17) is a price index comparing a particular week with another week one quarter later. In other words, each of these elements is a price index calculated at a quarterly frequency. A total of 13 such indexes can be computed in each quarter. By taking the geometric mean of these 13 quarterly frequency price indexes, we obtain an overall quarterly price index, which can be interpreted as the quarterly equivalent of the original weekly index.

Once the quarterly version of the weekly index has been constructed, its performance can be measured by comparing it with a reference quarterly index. Here we make the comparison using a metric proposed by Diewert (2002, 2009).

$$X = \frac{1}{T-1} \sum_{t=1}^{T-1} \left[\left(\frac{P^w_{t,t+1}}{P^{quart}_{t,t+1}} \right) + \left(\frac{P^{quart}_{t,t+1}}{P^w_{t,t+1}} \right) - 2 \right].$$

The smaller the value of the X metric, the more similar are the two indexes.

Given a reference quarterly index, we can then vary the length of the RTD rolling window and observe how it affects the X metric. We prefer whichever window length generates the smallest X metric. An important question then is how robust is the optimal window length to the choice of reference quarterly index? If it is reasonably robust, then the selected window length is optimal in the sense that it generates a weekly RTD index that is the most consistent with our reference quarterly indexes. Similarly, holding the window length fixed at 53 weeks, we can observe how changing the RTD linking method affects the X metric. Again, we prefer the linking method with the smallest X metric.

4. The Data Sets

4.1. The Sydney Data Set and Hedonic Model

We use a data set obtained from Australian Property Monitors that consists of prices and characteristics of houses sold in Sydney (Australia) for the years 2003–2014. For each house, we have the following characteristics: the actual sale price, time of sale, postcode, property type (i.e., detached or semi), number of bedrooms, number of bathrooms, land area, exact address, longitude and latitude. (We exclude all townhouses from our analysis since the corresponding land area is for the whole strata and not for the individual townhouse itself.)

For a robust analysis, it was necessary to remove some outliers. This is because there is a concentration of data entry errors in the tails of the distribution, caused for example by the inclusion of erroneous extra zeroes. These extreme observations can distort the results. Complete data on all our hedonic characteristics are available for 433,202 observations. To simplify the computations, we also merged the number of bathrooms and number of bedrooms into broader groups (one, two, and three or more bathrooms; one or two, three, four, five or more bedrooms).

Using weekly periods, the hedonic model for Sydney is estimated with a rolling window ranging between two weeks and 53 weeks. The window is then rolled forward one period and the hedonic model re-estimated. Hence in the case of the two-week window, a total of 711 hedonic models are estimated, covering the time interval from January 2003 to December 2014.

The hedonic model estimated for Sydney is semilog and contains the following five characteristics:

(1) number of bedrooms,
(2) number of bathrooms,
(3) log of land area,
(4) house type (detached, or semi), and
(5) postcode.

All these variables with the exception of land area take the form of dummy variables.

4.2. The Tokyo Data Set and Hedonic Model

The Tokyo data set covers the metropolitan area (621 square kilometers), and the analysis period is approximately 30 years between January 1986 and June 2016. The data set includes previously-owned condominiums published Shukan Jyutaku Joho (Residential Information Weekly) published by RECRUIT, Co. This magazine provides information on the characteristics and asking prices of listed properties on a weekly basis. Moreover, Shukan Jutaku Joho provides time-series data on housing prices from the week they were first posted until the week they were removed as a result of successful transactions. We only use the price in the final week because this can be safely regarded as sufficiently close to the contract price.

The available housing characteristics include floor space and age. The convenience of public transportation from each housing location is represented by travel time to the central business district (CBD), and time to the nearest station. City codes and a railway dummy to indicate along which railway/subway line a housing property is located are also available.

The hedonic model for Tokyo is estimated over 242,233 observations. The functional form is semilog. The explanatory variables used here are:

(1) log of floor area,
(2) age,
(3) time to nearest station,
(4) time to Tokyo central station (included as a quadratic), and
(5) city code.

5. Results

5.1. The Sensitivity of the Results to the Choice of Window Length

The spreads of the weekly RTD hedonic price indexes for Sydney and Tokyo as the window length is varied between two and 53 weeks are shown in Figures 2 and 3. It can be seen that the weekly indexes are quite sensitive to the choice of window length.

5.2. The Sensitivity of the Results to the Choice of Linking Method

Holding the window length fixed at 53 weeks, the sensitivity of a weekly RTD method to the choice of linking method is shown for Sydney and Tokyo in Figures 4 and 5. It can be seen that the variation in the RTD price indexes from varying the linking method is smaller than the variation resulting from changing the window length. However, the spread is still significant.

5.3. A Quarterly Index as a Benchmark

The hedonic imputation and time-dummy methods generate very similar quarterly price indexes. The results are shown in Figures 6 and 7. These results indicate that at a quarterly frequency, we have quite a good idea of what the right answer is. Hence, these quarterly indexes can be used as a benchmark for discriminating between competing weekly indexes.

5.4. How RTD Index Performance Depends on Window Length

Here we focus on the standard RTD linking method described in Subsection 2.1. For this case, the X metric for each RTD window length for Sydney with the hedonic imputation index as the reference quarterly index is shown in Figure 8. The X metric is minimized when the RTD

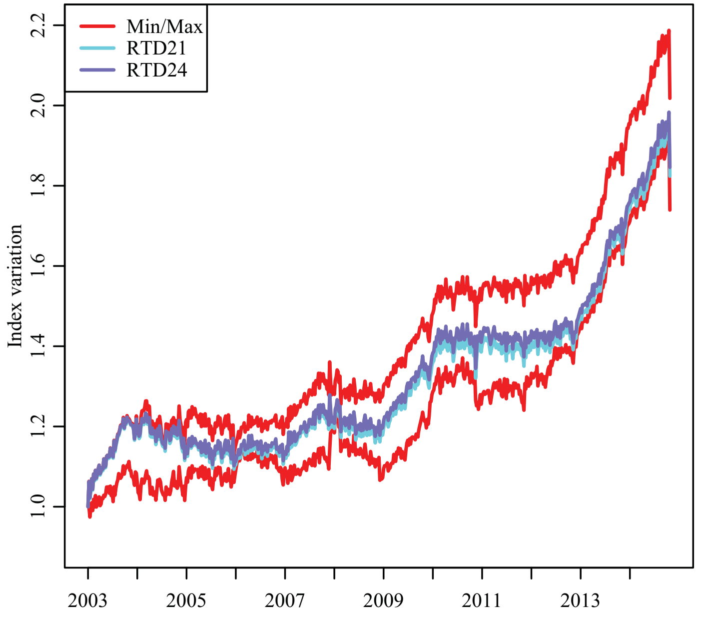


Fig. 2. The impact of varying the window length on weekly RTD house price indexes for Sydney.
Note: RTD21 and RTD24 denote the 21- and 24-week RTD price indexes for Sydney. Min and max denote the lower and upper bounds on all the RTD price indexes with windows ranging between two and 53 weeks.

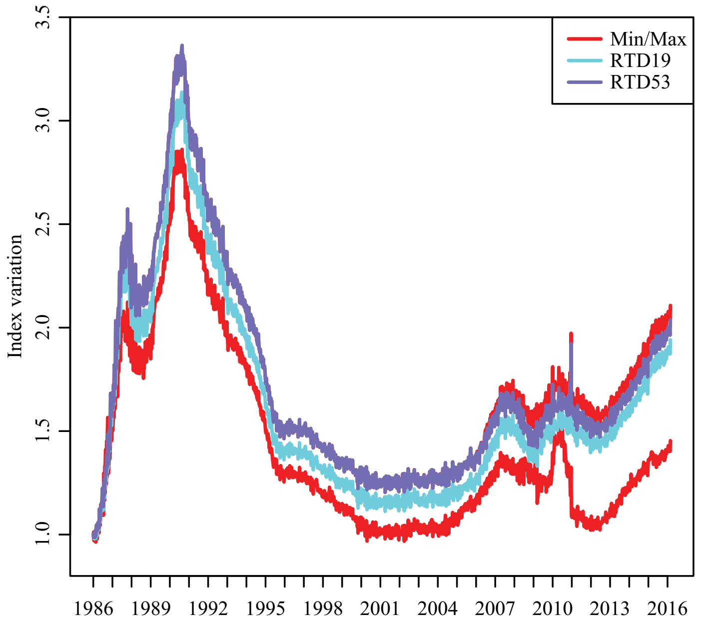


Fig. 3. The impact of varying the window length on weekly RTD house price indexes for Tokyo.
Note: RTD19 and RTD53 denote the 19- and 53-week RTD price indexes for Tokyo. Min and max denote the lower and upper bounds on all the RTD price indexes with windows ranging between two and 53 weeks.

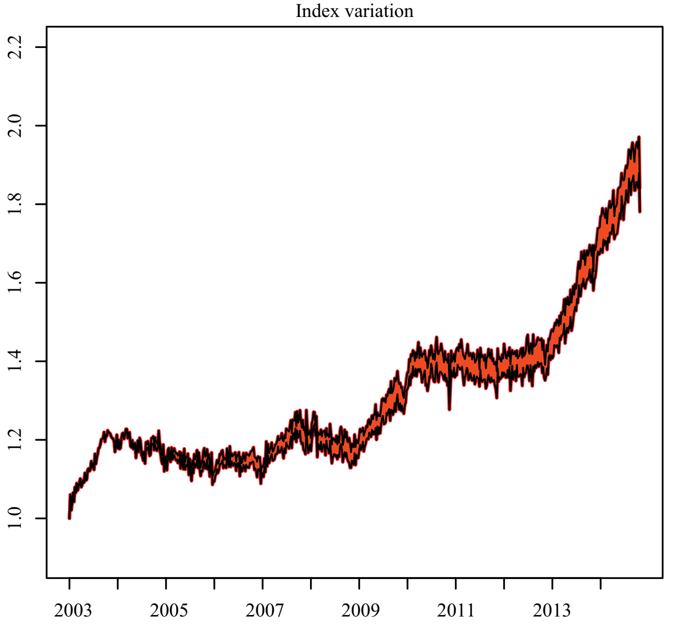


Fig. 4. The impact of varying the linking method on weekly RTD house price indexes with a 53-week window for Sydney.
Note: This graph shows the range of RTD price indexes for Sydney resulting from using different single-period linking methods. With the window length fixed at 53 weeks, there are 52 ways of doing single-period linking.

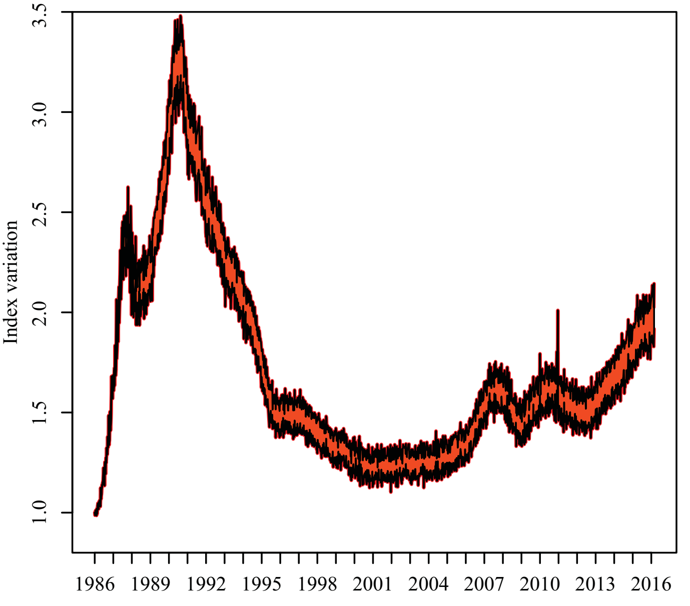


Fig. 5. The impact of varying the linking method on weekly RTD house price indexes with a 53-week window for Tokyo.
Note: This graph shows the range of RTD price indexes for Tokyo resulting from using different single-period linking methods. With the window length fixed at 53 weeks, there are 52 ways of doing single-period linking.

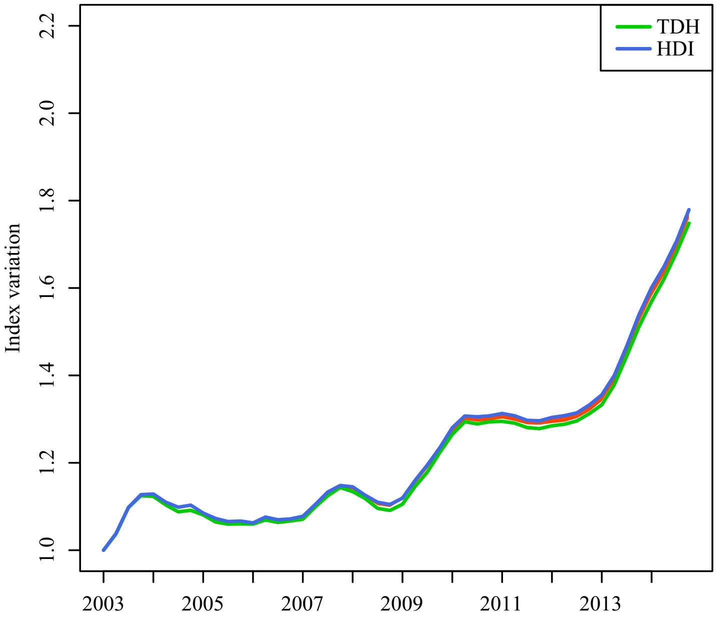


Fig. 6. Quarterly hedonic imputation and time-dummy house price indexes for Sydney.
Note: TDH and HDI here denote quarterly time-dummy hedonic and hedonic double imputation price indexes for Sydney. As can be seen, the two indexes closely approximate each other.

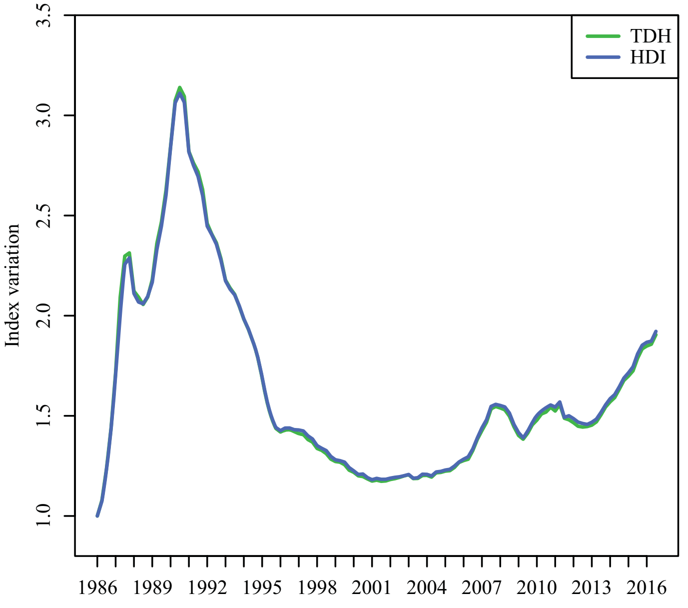


Fig. 7. Quarterly hedonic imputation and time-dummy house price Indexes for Tokyo.
Note: TDH and HDI here denote quarterly time-dummy hedonic and hedonic double imputation price indexes for Tokyo. As can be seen, the two indexes closely approximate each other.

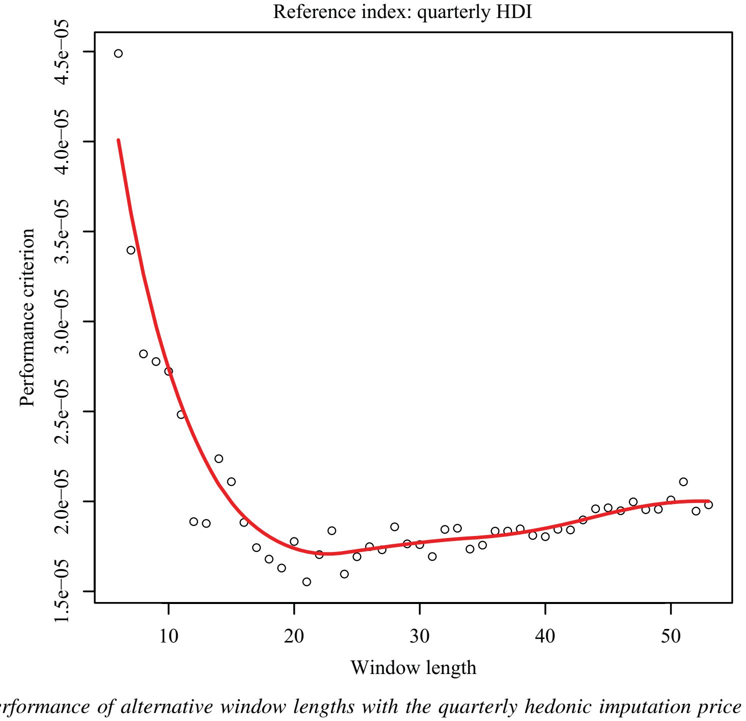


Fig. 8. Performance of alternative window lengths with the quarterly hedonic imputation price index as the reference: Sydney
Note: This graph shows which window length generates a weekly RTD price index that most closely approximates a quarterly hedonic imputation index. For Sydney the optimal window length here is 21 weeks. The red curve represents a fitted curve obtained by a local-linear smoother.

window length is 21 weeks. The same answer is obtained when the time-dummy index is used as the reference quarterly index as shown in Figure 15 in the Appendix (Section 7).

For Tokyo, when the hedonic imputation index is used as the reference quarterly index, the X metric is minimized by an RTD window length of 18 weeks, as shown in Figure 9. When the time-dummy index is used as the reference quarterly index, the results for Tokyo are not so clear, as shown in Figure 16, in the Appendix.

The optimal window length may also depend on the linking method. To illustrate this point, we recompute the optimal window length for Sydney, where now the linking is done using the geometric mean linking method as described in Equation (8). Using the quarterly hedonic imputation method as the benchmark, the optimal window length is now 19 weeks (see Figure 10). This is quite similar to the optimal window length of 21 weeks obtained using single-period linking.

In summary, we find that for Sydney the optimal RTD window length is between 19 and 21 weeks depending on the linking method used. For Tokyo, according to the quarterly hedonic imputation benchmark, the optimal window length is 18 weeks.

5.5. How RTD Index Performance Depends on the Linking Method

Now instead, we hold the RTD window length fixed at 53 weeks and compare the impact on the X metric of varying the linking method used by the RTD method. With a 53 week window, there are 52 ways of linking a new period to a single previous period, as described in Equation (6). For Sydney, the X metric corresponding to each of these 52 ways of

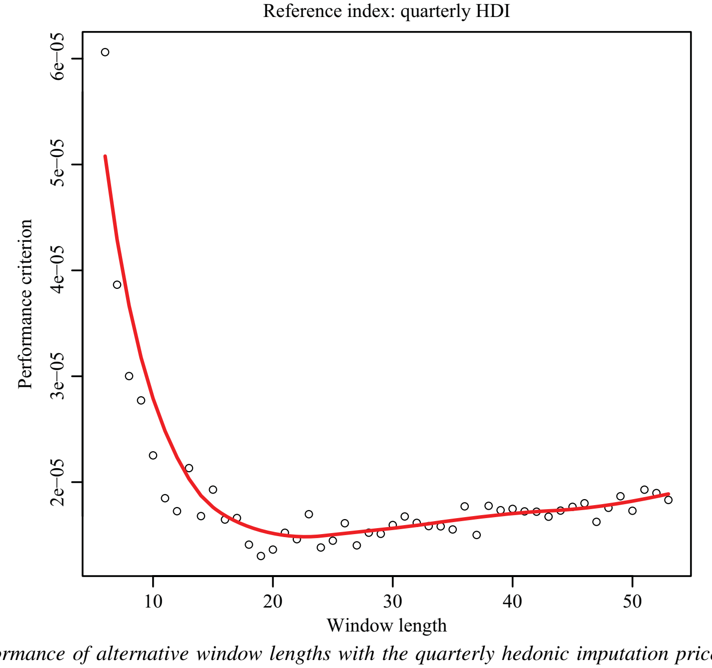


Fig. 9. Performance of alternative window lengths with the quarterly hedonic imputation price index as the reference: Tokyo.

Note: This graph shows which window length generates a weekly RTD price index that most closely approximates a quarterly hedonic imputation index. For Tokyo the optimal window length here is 18 weeks. The red curve represents a fitted curve obtained by a local-linear smoother.

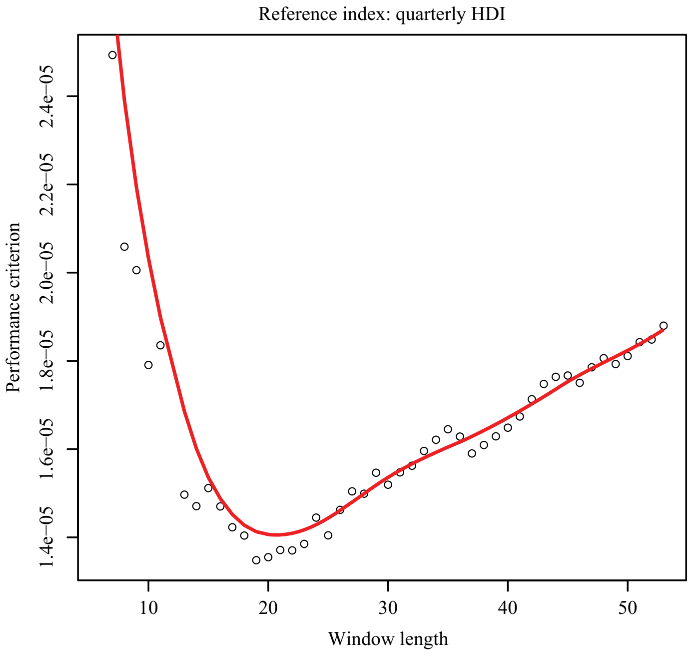


Fig. 10. Performance of geometric mean linking for different window lengths: quarterly hedonic imputation benchmark for Sydney.

Note: This graph shows which window length generates a weekly RTD price index that most closely approximates a quarterly hedonic imputation index when the current period is linked in using the geometric linking method in Equation (8). For Sydney the optimal window length here is 19 weeks. The red curve represents a fitted curve obtained by a local-linear smoother.

linking is shown in Figure 11 for the case where the quarterly hedonic imputation index is used as the reference. Corresponding results with the quarterly time-dummy index as the reference are shown in Figure 17.

In addition to these 52 single period linking methods, we also consider three average linking methods.

(1) **Geomean52** is the geometric mean of the 52 single period linking methods as described in Equation (8).
(2) **Geomean20** is the geometric mean of the 20 chronologically closest single period linking methods.
(3) **lambda = 0.95** is the weighted geometric mean method with $\lambda = 0.95$ as described in Equation (9).

When the quarterly hedonic imputation method is used as the reference index the optimal linking method links week t to week $t-16$ (see Figure 11). Linking through the period 16 weeks earlier even slightly outperforms the average linking methods (1), (2) and (3).

When the quarterly time-dummy method is used as the reference index, the optimal link for week t is with week $t-13$ (see Figure 17 in the Appendix). In this case, linking through the period 13 weeks earlier slightly outperforms methods (1) and (3), but is about equivalent to taking the geometric mean of the chronologically most recent 20 single-week links.

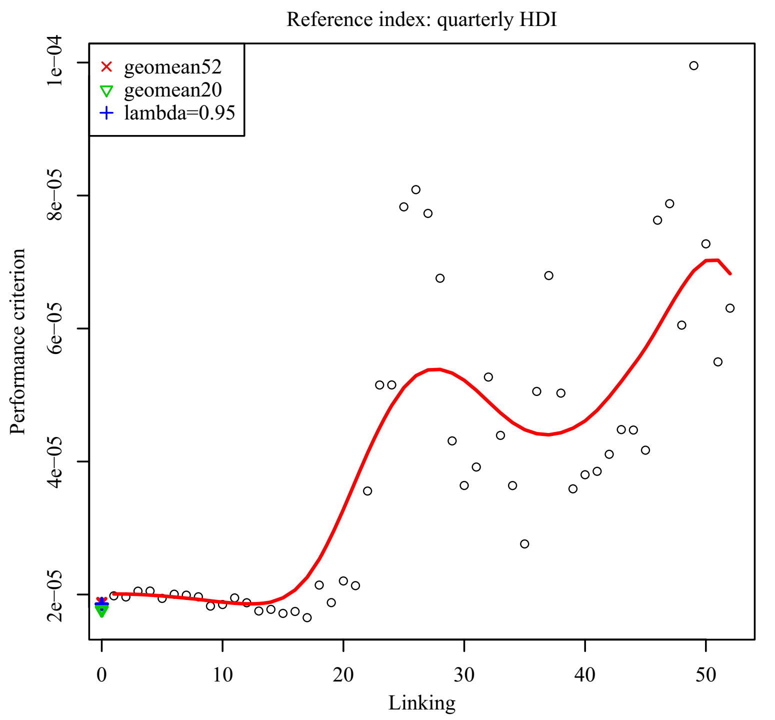


Fig. 11. Performance of alternative linking methods: quarterly hedonic imputation benchmark for Sydney.
Note: This graph shows the performance for Sydney (relative to a quarterly hedonic imputation index) of different single-period linking methods, Geomean52, Geomean20, and the weighted geomean method in (9) with A set to 0.95 (see values indicated at left-hand side of graph). The best performing method is single-period linking with 16 weeks earlier.

The corresponding results for Tokyo are shown in Figures 12 and 18. For the singleweek links, in both Figures 12, (where the quarterly hedonic imputation index is used as a benchmark) and Figure 18 in the Appendix (where the quarterly time-dummy index is used as the benchmark), the X metric is minimized by linking week t with week $t-12$ (i.e., linking the current week with 12 weeks earlier).

For Tokyo in Figure 12 the averaging methods (1), (2) and (3) perform equivalently to linking through 12 weeks earlier. In Figure 18, method (1) (i.e., the geometric mean of the 52 single-period lining methods) outperforms all the single-period linking methods.

5.6. *The Low Transaction Method Illustrated Using Weekly Sydney Data*

Here we focus specifically on low-transaction method 1, as described in Subsection 2.3. Setting the minimum number of observations per week to 250, we can see from Figure 13 that every year the last week in December and the first week in January fail to attain this threshold.

Setting the window length to seven weeks, the standard RTD method exhibits a slight upward drift compared with low transaction method 1, as can be seen in Figure 14. Towards the end of our sample, the cumulative size of this upward drift is 8.9%. The size and direction of drift will differ depending on the country or city and frequency of the index. Drift is most likely to be a problem for smaller countries without much data.

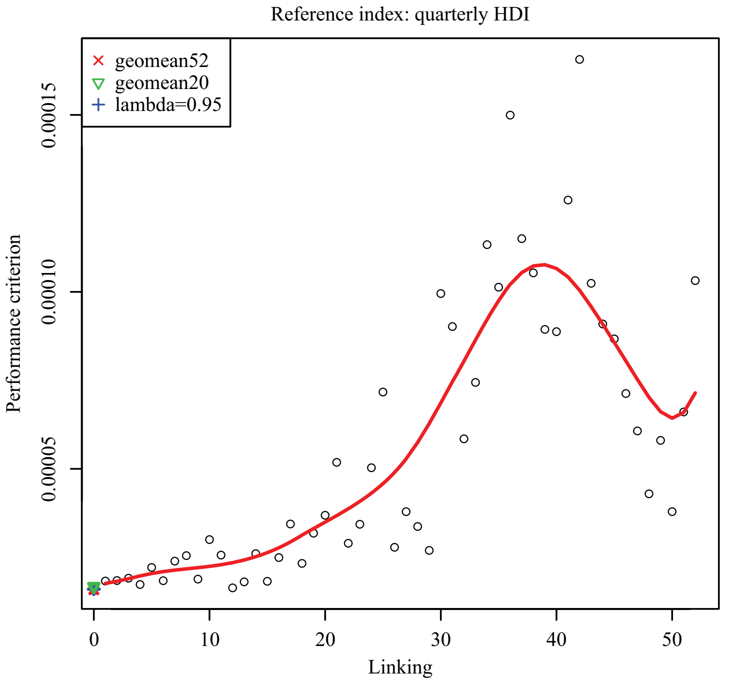


Fig. 12. Performance of alternative linking methods: quarterly hedonic imputation benchmark for Tokyo.
Note: This graph shows the performance for Tokyo (relative to a quarterly hedonic imputation index) of different single-period linking methods, Geomean52, Geomean20, and the weighted geomean method in Equation (9) with λ set to 0.95 (see values indicated at left-hand side of graph). The best performing method is single-period linking with 12 weeks earlier.

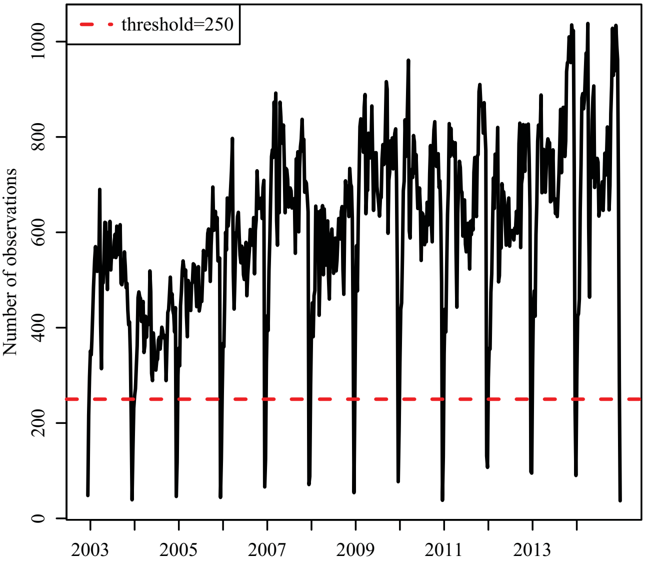


Fig. 13. Weekly observations versus the 250 observation threshold: Sydney.
Note: This graph shows the weekly number of transactions in Sydney. The low point each year is the last week in December and the first week in January. When implementing low transaction method 1, a threshold of 250 transactions per week is used.

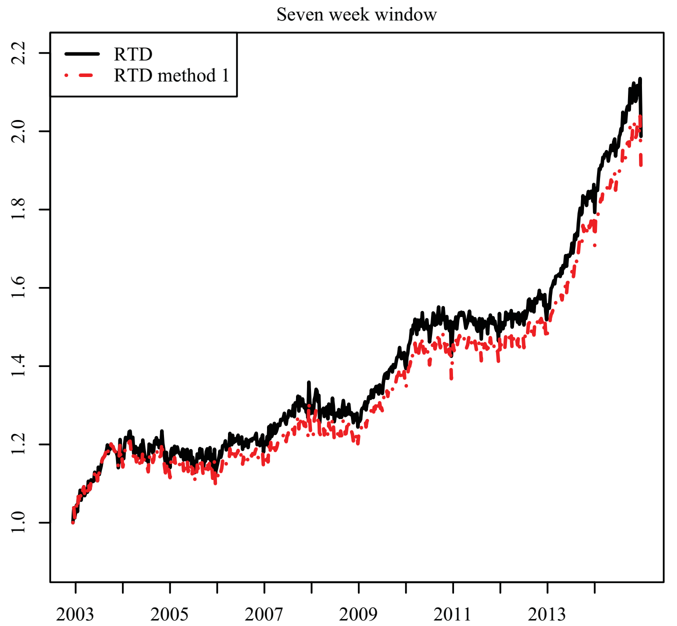


Fig. 14. Comparing the standard seven-week RTD index with the low transaction method 1 Index.
Note: RTD here is a standard weekly RTD price index computed with a seven-week window. RTD method 1 is the modified method where weeks with less than 250 transactions are treated seperatly as explained in Subsection 2.3. Failure to adjust for low transaction weeks seems to cause a slight upward drift in the RTD index.

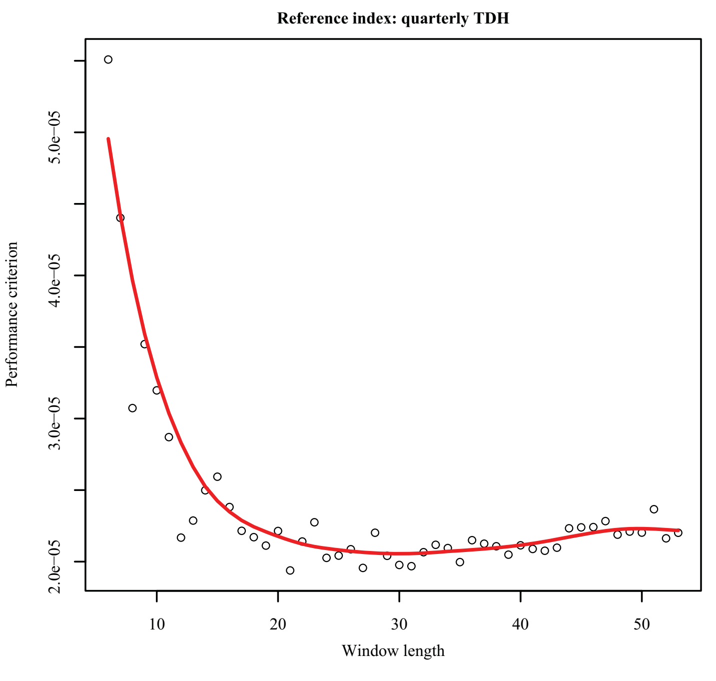


Fig. 15. Performance of alternative window lengths with the quarterly time-dummy price index as the reference: Sydney.
Note: This graph shows which window length generates a weekly RTD price index that most closely approximates a quarterly time-dummy hedonic index. For Sydney, the optimal window length here is 21 weeks.

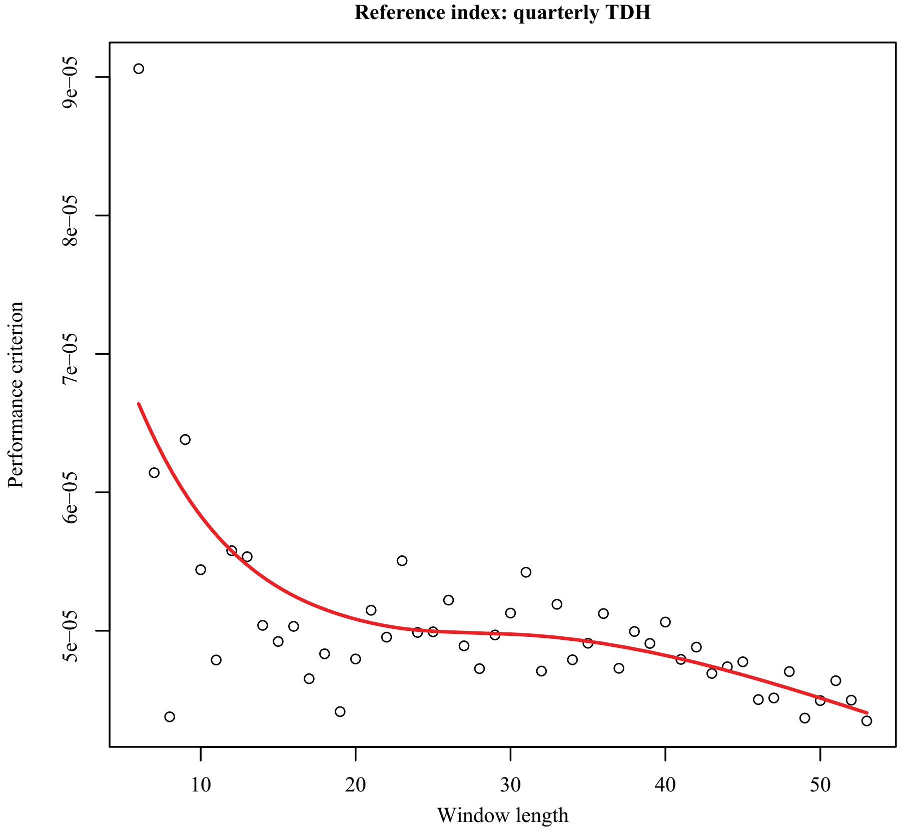


Fig. 16. Performance of alternative window lengths with the quarterly time-dummy price index as the reference: Tokyo.

Note: This graph shows which window length generates a weekly RTD price index that most closely approximates a quarterly hedonic imputation index. For Tokyo, the optimal window length here is 53 weeks.

6. Conclusion

We have considered two dimensions over which RTD HPIs can differ. These are the window length and the method used for linking the current period to earlier periods. We have proposed a new criterion for determining the optimal window length and linking method. This method, which relies on using a lower frequency index to assess the performance of higher frequency indexes, works well for weekly indexes, using quarterly indexes as a benchmark. It remains to be seen how well it will work on lower frequency indexes, such as quarterly indexes, using say yearly indexes as the benchmark.

Focusing on weekly indexes, we find that for the Sydney data set, the optimal window length is between 19−21 weeks. For the Tokyo data set the optimal window length is about 18 weeks.

We show that it is possible to improve on the standard linking method used by the RTD method. The linking method that performs best on the Sydney data set links the current week with a period between 13−16 weeks earlier. For Tokyo, linking the current week with the period 12 weeks earlier performs best. Geometric averaging of the single-period linking methods performs about the same as the best of the single-period linking methods.

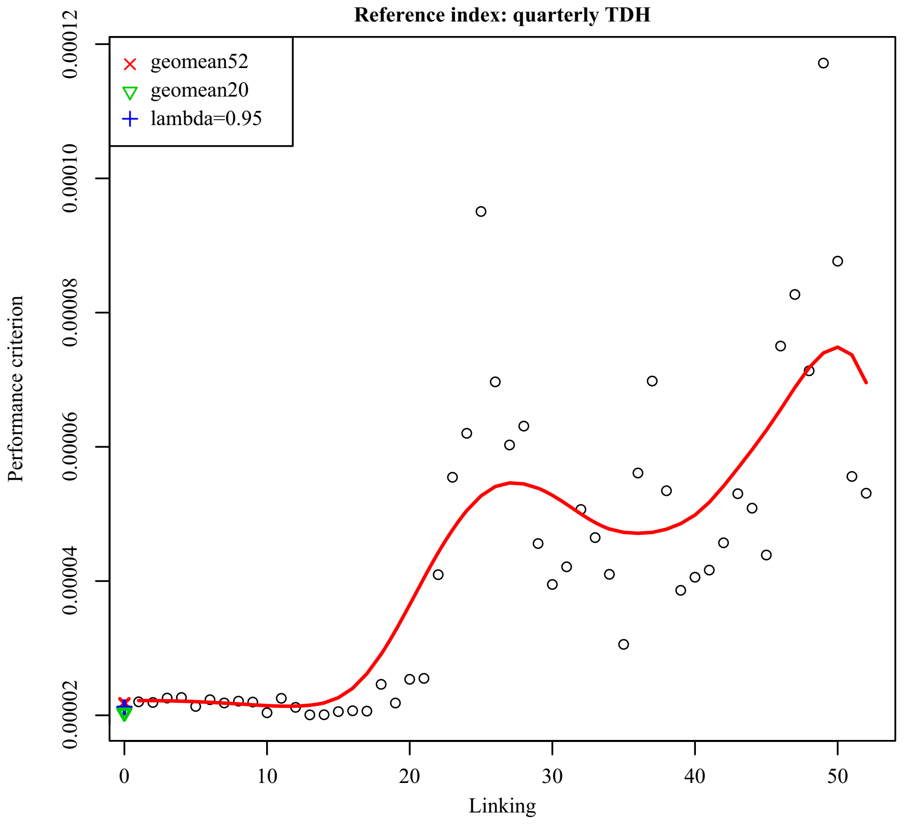


Fig. 17. Performance of alternative linking methods with standard RTD linking: quarterly time-dummy benchmark for Sydney.

Note: This graph shows the performance for Sydney (relative to a quarterly time-dummy hedonic index) of different single-period linking methods, Geomean52, Geomean20, and the weighted geomean method in Equation (9) with Α set to 0.95 (see values indicated at left-hand side of graph). The best performing method is single-period linking with 13 weeks earlier.

We have also considered how the RTD method can be adjusted to mitigate the distorting effects of low transaction periods on house price indexes. Method 1 (in Subsection 2.3) in particular could prove useful for countries in Europe and the rest of the world that compute their official HPIs using the RTD method.

Finally there is the question of whether weekly indexes are indeed useful. In our opinion the answer is yes that they are a useful complement to lower frequency indexes, as long as there are enough transactions per week to allow the construction of hedonic indexes and the time lag for obtaining the necessary price data is not too long. In many countries to avoid such long time lags it may be necessary to construct weekly house price indexes using list price data, as our Tokyo index does. What then of a daily HPI? A daily repeat-sales HPI index has been constructed by Bollerslev et al. (2016) using US data. We doubt that it would be feasible to construct a daily RTD index in most countries. A different approach that compensates for the low rate of transactions by imposing more econometric structure on the model is probably needed for daily indexes.

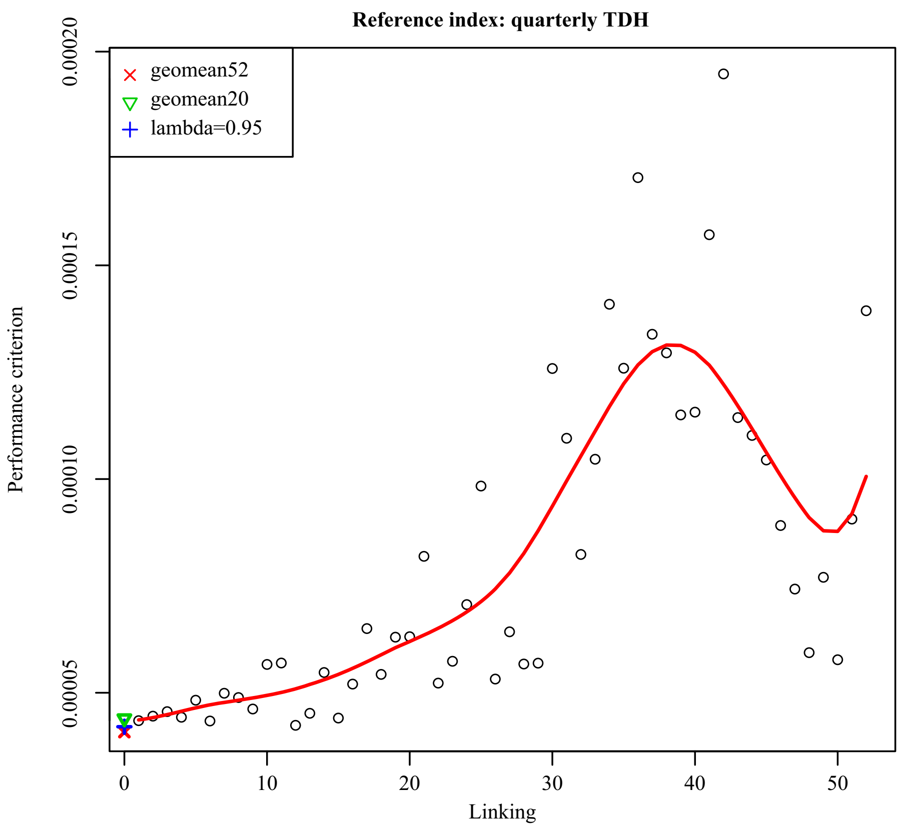


Fig. 18. Performance of alternative linking methods with standard RTD linking: time-dummy benchmark for Tokyo.

Note: This graph shows the performance for Tokyo (relative to a quarterly time-dummy hedonic index) of different single-period linking methods, Geomean52, Geomean20, and the weighted geomean method in (9) with A set to 0.95 (see values indicated at left-hand side of graph). The best performing method is Geomean52.

7. Appendix: Results Obtained Using the Quarterly Time-Dummy Index As the Reference

As a robustness check, here we recompute all the results derived using the quarterly hedonic imputation method as a benchmark. Now instead, we use the quarterly time-dummy method as the benchmark. In most cases, the results are very similar to those obtained using the hedonic imputation method.

8. References

Bollerslev, T., A.J. Patton, and W. Wang. 2016. "Daily House Price Indices: Construction, Modeling, and Longerrun Predictions," *Journal of Applied Econometrics* 31(6): 1005–1025. DOI: https://doi.org/10.1002/jae.2471.

Chessa, A.G.J. Verburg and L. Willenborg. 2017. "A Comparison of Price Index Methods for Scanner Data," Paper Presented at the Ottawa Group Meeting, International Working Group on Price Statistics, Deutsche Bundesbank, Conference Centre, 10–12 May 2017,

Eltville am Rhein, Germany. Available at: https://www.bundesbank.de/en/service/dates/fifteenth-meeting-of-the-ottawa-group-2017-634968 (accessed February 2022).
De Haan, J. 2004. "Direct and Indirect Time Dummy Approaches to Hedonic Price Measurement," *Journal of Economic and Social Measurement*, 29(4): 427–443. DOI: https://doi.org/10.3233/JEM-2004-0235 (accessed February 2022).
De Haan, J. 2010. "Hedonic Price Indexes: A Comparison of Imputation, Time Dummy and Re-Pricing Methods," *Journal of Economics and Statistics (Jahrbücher für Nationaloekonomie und Statistik)*, 230(6): 772–791. DOI: https://doi.org/10.1515/jbnst-2010-06.
De Haan, J. 2015. "Rolling Year Time Dummy Indexes and the Choice of Splicing Method," Paper Presented at the Ottawa Group Meeting, International Working Group on Price Statistics, Statistics Bureau of Japan, 20–22 May. Available at: https://www.ottawagroup.org/Ottawa/ottawagroup.nsf/51c9a3d36edfd0dfca256acb00118404/2274041837f41548ca257eed00811dad?OpenDocument (accessed February 2022).
De Haan, J., R. Hendriks, and M. Scholz. 2021. "Price Measurement Using Scanner Data: Time-Product Dummy Versus Time Dummy Hedonic Indexes," *Review of Income and Wealth*, 67: 394–417. DOI: https://doi.org/10.1111/roiw.12468.
Diewert, W.E. 2002. *Similarity and Dissimilarity Indexes: An Axiomatic Approach,* Discussion Paper 02–10, Department of Economics, University of British Columbia, Vancouver, Canada. Available at: http://papers.economics.ubc.ca/legacypapers/dp0210.pdf (accessed February 2022).
Diewert, W.E. 2009. "Similarity Indexes and Criteria for Spatial Linking," In *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, edited by D.S.P. Rao, Edward Elgar: Cheltenham, UK, chap 8: 183–216.
Diewert, W.E. 2011. *Alternative Approaches to Measuring House Price Infation*, Discussion Paper 10–10, Department of Economics, University of British Columbia, Vancouver, Canada. Available at: https://www.business.unsw.edu.au/research-site/centreforappliedecDocuments/W.%20Erwin%20Diewert%20-%20Alternative%20Approaches%20-to%20Measuring%20House%20Price%20Inflation.pdf (accessed February 2022).
Diewert, W.E., S. Heravi, and M. Silver. 2009. "Hedonic Imputation versus Time Dummy Hedonic Indexes," In, *Price Index Concepts and Measurement*, edited by W.E. Diewert, J. Greenlees, and C. Hulten, NBER Studies in Income and Wealth, University of Chicago Press, Chicago: 161–196.
Diewert, W.E., and K.J. Fox. 2022. "Substitution Bias in Multilateral Methods for CPI Construction using Scanner Data," *Journal of Business and Economic Statistics*, forthcoming. DOI: https://doi.org/10.1080/07350015.2020.1816176.
Eurostat. 2016. *Detailed Technical Manual on Owner-Occupied Housing for Harmonised Index of Consumer Prices*, Eurostat, Luxembourg. Available at: https://eu/eurostat/documents/7590317/0/Technical-Manual-00H-HPI-2017/ (accessed February 2022).
Hill, R.J. 2013. Hedonic Price Indexes for Housing: A Survey, Evaluation and Taxonomy," *Journal of Economic Surveys*, 27(5): 879–914. DOI: https://doi.org/10.1111/j.1467-6419.2012.00731.x.
Hill, R.J., and D. Melser. 2008. "Hedonic Imputation and the Price Index Problem: An Application to Housing," *Economic Inquiry*, 46(4): 593–609. DOI: https://doi.org/10.1111/j.1465-7295.2007.00110.x.

Hill, R.J., M. Scholz, C. Shimizu, and M. Steurer. 2018."An Evaluation of the Methods Used by European Countries to Compute their Official House Price Indexes," *Economie et Statistique* 500–502: 221–238. DOI: https://doi.org/10.24187/ecostat.2018.500t. 1953.

Krsinich, F. 2016. "The FEWS Index: Fixed Effects with a Window Splice," *Journal of Official Statistic* 32(2): 375–404. DOI: https://doi.org/10.1515/jos-2016-002.

Melser, D. 2018. "Scanner Data Price Indexes: Addressing Some Unresolved Issues," *Journal of Business and Economic Statistics* 36(3): 516–522. DOI: https://doi.org/10.1080/07350015.2016.1218339.

O'Hanlon, N. 2011. "Constructing a National House Price Index for Ireland," *Journal of the Statistical and Social Inquiry Society of Ireland*, 40: 167–196. Available at: http://www.tara.tcd.ie/bitstream/handle/2262/62349/o%27hanlon%20pdf.pdf?sequence=1&isAllowed=y (accessed February 2022).

Rachman, A.N. 2019. "An Alternative Hedonic Residential Property Price Index for Indonesia Using Big Data: The Case of Jakarta," Paper presented at the International Conference on Real Estate Statistics, February 20–22, Eurostat, Luxembourg. Available at: https://www.crrem.eu/crrem-at-the-eurostat-international-conference-on-real.

Rambaldi, A.N. and D.S.P. Rao. 2013. "Econometric Modeling and Estimation of Theoretically Consistent Housing Price Indexes," CEPA Working Papers Series WP042013, School of Economics, University of Queensland, Australia. https://economics.uq.edu.au/files/5157/WP042013.pdf

Shimizu, C., H. Takatsuji and H. Ono. 2003. "Conjunct method of deriving a hedonic price index in a secondhand housing market with structural change," *Reitaku International Journal of Economic Studies* 10(2): 103–134.

Shimizu, C, H. Takatsuji, H. Ono, and K.G. Nishimura. 2010. "Structural and Temporal Changes in the Housing Market and Hedonic Housing Price Indices," *International Journal of Housing Markets and Analysis* 3(4): 351–368. DOI: https://doi.org/10.1108/17538271011080655.

Shimizu, C., and W.E. Diewert. 2019. "Residential Property Price Index in Japan: Discussion in Methodology and Data Sources," Paper presented at the International Conference on Real Estate Statistics, Eurostat, Luxembourg, 2019. https://www.crrem.eu/crrem-at-the-eurostat-international-conference-on-real-estate-statistics.

Shimizu, C., K.G. Nishimura and T. Watanabe. 2016. "House prices at different stages of the buying/selling process," *Regional Science and Urban Economics* 59: 37–53. DOI: https://doi.org/10.1016/j.regsciurbeco.2016.04.001 (accessed February 2022).

Silver, M., and S. Heravi. 2007. "The Difference between Hedonic Imputation Indexes and Time Dummy Hedonic Indexes," *Journal of Business and Economic Statistics*, 25(2): 239–246. DOI: https://doi.org/10.1198/073500106000000486.

Triplett J.E. 2004. *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products*, STI Working Paper 2004/9, Directorate for Science, Technology and Industry, Organisation for Economic Co-operation and Development, Paris. DOI: https://doi.org/10.1787/643587187107.

Received July 2020
Revised January 2021
Accepted May 2021

Journal of Official Statistics, Vol. 38, No. 1, 2022, pp. 153–186, http://dx.doi.org/10.2478/JOS-2022-0008

Econometric Issues in Hedonic Property Price Indices: Some Practical Help

Mick Silver[1]

Hedonic regressions are widely used and recommended for property price index (PPI) measurement. Hedonic PPIs control for changes in the quality-mix of properties transacted that can confound measures of change in average property prices. The widespread adoption of the hedonic approach is primarily due to the increasing availability, in this digital age, of electronic data on advertised and transaction prices of properties and their price-determining characteristics. Yet hedonic PPIs are only as good as the underlying estimated hedonic regressions. Regression-based measures are unusual in official economic statistics. There is little technical support in the international Handbooks and Guides for diagnostic measures and graphical plots for estimated regression equations as applied to PPIs. These diagnostics are essential to the transparency and credibility of hedonic PPI measurement. This article seeks to remedy this.

Key words: Hedonic regressions; residential property price index; commercial property price index; house price index; regression diagnostics.

1. Introduction

Residential and commercial property price indices (RPPIs and CPPIs – hereafter PPIs) are hard to measure. Houses, never mind commercial properties, are infrequently traded and heterogeneous. Average house prices may increase over time, but the difficulty is how to separate inflation and quality change from this increase. For example, more four-plus bedroom houses in a better (more expensive) post (zip)-code transacted in a current quarter, compared with the previous quarter, would bias upwards a meaningful measure of the change in average house prices. Further, a market price can only be identified when the property is transacted, possibly leading to a relative sparseness of price observations—thin markets.

While there are alternative approaches to property price index measurement, the concern of this article is with the hedonic approach. International standards have an explicit preference for the hedonic method as the "best" method, unless there is only limited or no information on housing characteristics (Eurostat et al. (2013, 158–160) *Handbook on Residential Property Price Indices*). The European Commission (EC) (2017,

[1] Formerly, Statistics Department, International Monetary Fund. Email: 01micksilver@gmail.com
Acknowledgments: For helpful comments on this subject are due to Corinne Becker-Vermeulen (Swiss Federal Statistical Office – FSO), Manuel Brand (FSO), Jens Mehrhoff (IMF), Niall O'Hanlon (IMF), and Vincent Rappo (FSO). I also acknowledge the help of three anonymous referees. The views expressed are those of the author who remains responsible for all errors.

© Statistics Sweden

85–87, 96), *Technical Manual on Owner-Occupied Housing and House Price Indices* requires that the hedonic methodology is the preferred methodology. The IMF's *RPPI Practical Compilation Guide* (IMF 2020, 27–46) only recommends hedonic methods. The IMF (2020, 51–52) *draft* (at the time of writing) *Guide* also includes an outline of the use of "medians with stratification" but dismisses the method due to its lack of proper quality adjustments. While Eurostat (2017) *CPPI Sources, Methods and Issues* contains a valuable overview of methods, agreement could not be reached on recommending a preferred method (see, Hill and Steurer 2020; Silver 2019). Hill (2013), in his survey of methodologies for RPPIs for European countries, concluded that "Hedonic indices seem to be . . . the method of choice for constructing quality-adjusted house price indices." Hill et al. (2018) provides a valuable overview of the extensive use of hedonic RPPI adopted by European countries.

The focus here is on hedonic transaction-based indices, rather than repeat-sales (RS) indices, a methodology prevalent in the United States. RS indices overcome the quality-mix problem by constraining the sample of properties to those sold more than once over the period in question. The inherent weaknesses of the RS method lie in its deletion of single-sales data, a potential lemons bias, and failure to properly account for the depreciation/improvements to properties between sales. There is also a major problem of determining how much weight should be given to pairs of price comparisons with a long time period between sales. Leventis (2008) found that differences in the autoregressive formulation of the RS model, used to weight such paired comparisons, accounted for significant differences in the resulting indices. Shimizu et al. (2010) and Diewert et al. (2020) provide comparative studies of the results from the RS and hedonic approaches.

The cornerstone of the hedonic approach to PPI measurement is the estimation of hedonic regression equations that relate the (log of) *price* of a property to its price-determining characteristics. This is undertaken in order to estimate the partial, marginal valuation of each characteristic, for example the *size* (square footage of the interior) of a house. An increase in the average price of houses can then be adjusted for the change in quality-mix – the increased size of houses transacted – since we have an estimate of the valuation of an additional square foot.

While hedonic methods are well documented in the international *Handbooks* and *Guides* referred to above, there remains a serious omission. Regression diagnostics are well developed procedures required to assess the validity of estimated regression models. Courses in econometric methods are the mainstay of applied economics training at universities at both undergraduate and postgraduate level. Such courses have their primary focus on multiple regression and the validity of the attendant assumptions of the regression estimators, the specification of the model, and the effects of outliers, as considered by the regression diagnostics.

There is an important hole that needs to be filled. The aforementioned international standards have little to no systematic guidance on regression diagnostics. While the quite excellent RPPI Handbook, Eurostat et al. (2013) comments on R 2, heteroscedasticity, multicollinearity, and the treatment of outliers, it does not (aim to) treat such issues in a systematic fashion. The Eurostat (2017) CPPI Sources, Methods and Issues, devotes little attention to regression diagnostics, and the IMF (2020) RPPI Practical Guide has no hedonic regression diagnostics. Yet national statistical offices (NSOs), central banks, and other organisations who estimate and compile PPIs, to be transparent, should publish details of the estimated hedonic regressions upon which their measures are based. In order to validate the estimated models, the published results require not only, for each variable and the intercept, estimated coefficients, the accompanying standard errors, t-tests, and p-values, but also the relevant diagnostics for the model as a whole, as outlined in sections V-VII. More so, it may well be that the best model available fails some of the requisite tests. Econometrics is well adapted to such matters providing, in some instances, work-arounds when assumptions of the estimator are not met, insights into how the model may be improved to alleviate such problems, and to identify circumstances when the failure of a diagnostic test has little effect on the bias/precision of the estimated coefficients and predicted *prices*.

This article is not only intended to provide practical help in identifying relevant diagnostics, but to also provide the rationale, methods, and circumstances by which a PPI may still be published and used in spite of shortcomings in the diagnostic tests. There is no trickery to such methods and analysis. Applied economists need to present results from econometric models. Economic data, unlike that from controlled experiments, come with all manner of shortcomings. Econometricians have undertaken substantial technical research to identify and understand the impact of violations of assumptions in the use of particular estimators for regression models and to develop alternative procedures where possible. As will be outlined, there is much to be gained from the econometrician's tool bag to enable the provision of PPIs to be based on rigorous, well-grounded, statistical principles and practice.

One reason why the current standards on PPIs do not include advice on regression diagnostics is that it is a highly complex area. Any attempt to summarize this subject in a chapter or short paper requires hubris, and this article will have many shortcomings. The article focuses on the main methods and issues relating to regression diagnostics in a PPI context, leaving the reader to supplement it using an "introductory" econometrics text such as Wooldridge (2013, part 1), Maddala and Lahari (2009), and (a more readable) Kennedy (2008). There is an extensive literature on how to create regression-based hedonic price indices, see Triplett (2006). There is a more introductory paper/text to be written to walk the reader through the estimation of, and diagnostics for, estimated hedonic regression equations, one that may include data examples. Silver (2018) takes the reader from the estimated hedonic regression to a PPI. Silver (2018) does not adopt the approach to PPI measurement that decomposes a PPI into its land and structure components. This is approach is well developed in Eurostat et al. (2013, chap. 8) and Diewert et al. (2020), amongst others.

All of this is very accessible and doable. Statistical/econometric software (including R, Stata and EViews) are quite excellent allowing the user to readily estimate hedonic regressions using alternative specifications and estimators, and provides associated diagnostics. In this article we reference Wooldridge (2013) throughout as a consistent source to direct readers to further information as required. Wooldridge illustrates examples in R; Heiss (2020) is an econometric text also based on R and related to Wooldridge. R is a comprehensive, open-source (free) statistical/econometric software with extensive graphical facilities that can efficiently handle large data (see https://www.r-project.org.). Basic regression diagnostics, as provided in R as a default output, are outlined in Section 5 of this article.

Subsection 2.1 briefly outlines the nature of hedonic regressions for PPIs and Subsections 2.2–2.4, the three main hedonic approaches to PPI measurement: the hedonic time dummy approach, imputation, and characteristics/repricing approach. There are many alternative forms for each approach depending on (1) the functional form of the hedonic regression and aggregation; (2) the choice of reference, current or some average of the two, period(s) to estimate hedonic coefficients or hold characteristics/weights constant; (3) whether dual or single imputation is used for prices and/or weights; (4) whether a direct or indirect formulation is used; (5) the periodicity of the estimation, say monthly/quarterly/annually—chained, rolling window or fixed baskets of characteristics; and more – see De Haan and Diewert (2013) and Silver (2015, 2018). A tweaking of the methodology is outlined in Subsections 3.1 and 3.2 to help deal with instances of sparse data. Subsection 3.3 outlines issues of smoothing, Subsection 3.4 introduces weights for individual properties, and Subsections 3.5 identifies an equivalence between two of the approaches. Sections 4–8 provide the meat of the article: the basic diagnostics: sample size (n), standard error of the regression (SER), F-test, R-bar squared ($\bar{R}^2$), the standard errors, t-tests and p-values of individual estimated coefficients (Section 4); multicollinearity, specification, and omitted variable bias (Section 5); heteroscedasticity (Section 6); normality of residual (Section 7); and the detection and treatment of outliers: cleaning, leverage, influence, and robust estimators (Section 8). Section 9 provides brief conclusions.

2. The Hedonic Approach

2.1. *A Hedonic Regression*

The starting point is an estimated hedonic regression for property prices. Throughout the article hedonic PPIs are based on a semi-log (log-linear) hedonic functional form, though similar principles apply to linear, log-log, and more flexible functional forms. The semi-log form allows for curvature in the relationships, say between price and square footage, and for a multiplicative association between quality characteristics, that is, that possession of a garage and additional bathroom may be worth more than the (linear) sum of the two. The semi-log form is more practical than a log-log form since many characteristics take a zero or one value (possession or not of a characteristic) and logarithms cannot be taken of zero values. Silver (2016) provides a detailed exposition of the issues and methods for a linear functional form.

A semi-log hedonic regression equation for (the logarithm of) prices, p_i^t, of property i on $z_{k,i}^t$ price-determining characteristics for period t data is given by:

$$\ln p_i^t = \ln\beta_0^t + \sum_{k=1}^{K} z_{k,i}^t \ln\beta_k^t + \ln u_i^t. \tag{1}$$

An ordinary least squares (OLS) estimated regression for Equation (1) is given as:

$$\ln\hat{p}_i^t = \ln\hat{\beta}_0^t + \sum_{k=1}^{K} z_{k,i}^t \ln\hat{\beta}_k^t \tag{2}$$

where $\ln\hat{p}_i^t$ (and $\ln p_i^t$) are the predicted (and actual) logarithms of the price of property i in period t; $z_{k,i}^t$ are the values of each $k = 1, \ldots, K$ price-determining characteristic for property i in period t; $\hat{\beta}_0^t$ (and β_k^t) are the estimated (and actual) coefficients for each k characteristic z_k^t; and u_i^t in Equation (1) are i.i.d. errors.

The underlying functional relationship between p_i^t and $z_{k,i}^t$ in Equation (1) is $p_i^t = \beta_0 \beta_1^{z_{1,i}^t} \beta_2^{z_{2,i}^t} \ldots \beta_k^{z_{k,i}^t} u_i^t$. Taking logs of both sides of this underlying function yields Equation (1).The estimated OLS regression in Equation (2) provides us with the estimated logarithms of the coefficients. Exponents of these estimated coefficients from the output of the software have to be taken if the estimated coefficients of the original function are to be recovered, that is: $\exp(\ln\hat{\beta}_k^t) = \hat{\beta}_k^t$. In principle, the index requires an adjustment for it to be a consistent (and almost unbiased) approximation of the proportionate impact of the dummy variable on price – see Kennedy (1981), Van Garderen and Shah (2002), and the note at the end of Hill (2013). This approximation is shown by Giles (2011) to be accurate, even for quite small samples. In practice, the adjustment usually has little effect. An adjustment is also required for the imputation approach to determine the predicted prices as the exponent of $\ln\hat{p}_i^t$ – see Wooldridge (2013, 204–408).

An OLS estimator is a conventionally accepted estimator with good standard properties, as outlined in Section 4. The case for robust estimators, when circumstances require it, is outlined in Subsections 5.1 and 8.3. Of interest is the emergence of machine learning techniques. Examples comparing hedonic models using neural networks and OLS are given in Curry et al. (2001) and Oladunni and Sharma (2018).

The use of a semi-logarithmic model for RPPIs is embedded in international recommendations and practice including Eurostat et al. (2013) and Hill et al. (2018) but can be found in earlier hedonic literature including Silver and Heravi (2001) and Triplett (2006). However, Wooldridge (2013, 183–186) provides an example using house prices where the right-hand side (RHS) explanatory variables include both variables in their raw and logged forms and provides guidelines on their use. Practitioners can readily compile PPIs with and without logged terms on the RHS and identify the extent to which such choice matters. If both forms yield similar values, sticking to the semi-log form accords with standard practice and has straightforward desirable equivalences, as outlined in Section 3.5. Otherwise, comparative diagnostics outlined in this article should be evaluated.

2.2. *The Time Dummy Variable Approach*

The (logarithm of) prices of individual properties are regressed on their characteristics and dummy variables for time across properties transacted over several time periods, including the reference period 0 and successive subsequent periods $t = 1,\ldots,T$. The time dummy variables take the values of $D_i^1 = 1$ if the property is sold in period 1 and zero otherwise; $D_i^2 = 1$ if the property is sold in period 2 and zero otherwise;. . . ., and $D_i^T = 1$ if the property is sold in period T and zero otherwise – we exclude a (benchmark) period 0 time dummy variable. An estimated time dummy variable hedonic regression run over periods $t = 0,1,2,\ldots,T$ is given by:

$$\ln\hat{p}_i^t = \ln\hat{\beta}_0 + \sum\nolimits_{k=1}^{K} z_{k,i}^t \ln\hat{\beta}_k + \sum\nolimits_{t=1}^{T} D^t \ln\hat{\delta}^t \tag{3}$$

where $\exp\left(\hat{\delta}^t\right)$ are estimates of the *proportionate* change in price arising from a change between the reference period $t = 0$ and successive periods $t = 1,..,T$ having controlled for changes in the quality characteristics via the term $\sum_{k=1}^{K} z_{k,i}^t \ln\hat{\beta}_k$.

The method implicitly restricts the coefficients on the quality characteristics to be constant over time: for example, an adjacent period $t = 0,1$ time dummy hedonic regression requires $\beta_k = \beta_k^0 = \beta_k^1$. The extent of this restriction depends on the length of the time period over which the regression is run. If, for example, the regression is run over quarterly data for a ten-year window, a property price comparison between say 2010Q1 and 2020Q1 with valuations of characteristics held constant may stretch credibility, though this can be alleviated by chained shorter, moving windows, or adjacent period regressions (Silver 2016).

2.3. The Imputation Approach

The *imputation* approach works at the level of individual properties. The rational for the approach lies in the matched model method. Consider a set of properties transacted in period t. We want to compare their period t prices with the prices of the same matched properties in period 0. In this way there is no contamination of the measure of price change by changes in the quality-mix of properties transacted. However, the period t properties were not sold in period 0: there are no matched period 0 prices. The solution, in the denominator of Equation (4), is to use a (semi-log) hedonic regression estimated from period 0 data to *predict the period 0 price*, $\hat{p}_i^0$, of each individual i property, *based on its period t characteristics*, z_i^t: $\hat{p}_{i|z_i^t}^0$. This provides an answer to the counterfactual question: "what would a property with period t characteristics have sold at in period 0?"

A constant-quality *hedonic geometric mean imputation* (HGMI) price index is a ratio of the geometric means of the (predicted) prices of individual properties transacted in period t compared with the period 0 valuations of the self-same period t properties. A ratio of geometric means is equal to the geometric means of price relatives: $\prod_{i \in N^t} \left(\frac{\hat{p}_{i|z_i^t}^t}{\hat{p}_{i|z_i^t}^0} \right)^{\frac{1}{N^t}}$; a geometric mean is preferred from the axiomatic approach to an arithmetic mean – Diewert (1995) and Silver and Heravi (2007); and a ratio of double-imputations (predicted values) is in principle preferred to one with a single imputation. However, as will be explained below, from Equation (7) and the right-hand-side (RHS) of Equations (4) and (5), a single imputation is advantageous for OLS. The value in the numerators of Equation (4) is the geometric mean of the period t predicted prices based on period t price-determining characteristics, $z_{i,k}^t$. This is compared, in the denominator, with the geometric mean of the period 0 predicted prices based on the self-same period t price-determining characteristics, $z_{i,k}^t$. Where N^t is the number of properties transacted in period t:

$$P_{HGMI:z_i^t}^{0 \to t} = \frac{\prod_{i \in N^t} \left(\hat{p}_{i|z_i^t}^t \right)^{\frac{1}{N^t}}}{\prod_{i \in N^t} \left(\hat{p}_{i|z_i^t}^0 \right)^{\frac{1}{N^t}}} = \frac{\exp\left(1/N^t \sum_{i \in N^t} \ln \hat{p}_{i|z_i^t}^t \right)}{\exp\left(1/N^t \sum_{i \in N^t} \ln \hat{p}_{i|z_i^t}^0 \right)} \left[= \frac{\exp(1/N^t) \sum_{i \in N^t} \ln p_i^t}{\exp(1/N^t) \sum_{i \in N^t} \ln \hat{p}_{i|z_i^t}^0} \right]. \quad (4)$$

A constant *period 0* characteristics, z_i^0, hedonic imputation PPI, where N^0 is the number of properties transacted in period 0, is given by:

$$P^{0\to t}_{HGMI:z_i^0} = \frac{\prod_{i\in N^0}\left(\hat{p}^t_{i|z_i^0}\right)^{\frac{1}{N^0}}}{\prod_{i\in N^0}\left(\hat{p}^0_{i|z_i^0}\right)^{\frac{1}{N^0}}} = \frac{\exp\left(1/N^0\sum_{i\in N^0}\ln\hat{p}^t_{i|z_i^0}\right)}{\exp\left(1/N^0\sum_{i\in N^0}\ln\hat{p}^0_{i|z_i^0}\right)}\left[=\frac{\exp\left(1/N^0\right)\sum_{i\in N^0}\ln\hat{p}^t_{i|z_i^0}}{\exp\left(1/N^0\right)\sum_{i\in N^0}\ln p_i^0}\right]. \quad (5)$$

Neither a period 0 nor a period t hedonic imputation index can be considered to be superior. Some average or compromise solution is required. Symmetric use of period 0 and period t characteristics values leads us to:

$$P^{0\to t}_{HGMI:(\bar{z}^0+\bar{z}^t)/2} = \frac{\prod_{k=0}^{K}\left(\hat{\beta}_k^t\right)^{\bar{z}_k^\tau}}{\prod_{k=0}^{K}\left(\hat{\beta}_k^0\right)^{\bar{z}_k^\tau}} = \frac{\exp\left(\sum_{k=0}^{K}\bar{z}_k^\tau\ln\hat{\beta}_k^t\right)}{\exp\left(\sum_{k=0}^{K}\bar{z}_k^\tau\ln\hat{\beta}_k^0\right)} \text{ where } \bar{z}_k^\tau = \left(\bar{z}_k^0+\bar{z}_k^t\right)/2. \quad (6)$$

A feature of the OLS estimator is that the mean of actual prices is equal to the mean of predicted prices:

$$\frac{1}{N^0}\sum_{i\in N^0}\hat{p}^0_{i|z_i^0} = \frac{1}{N^0}\sum_{i\in N^0}p_i^0 \text{ and } \frac{1}{N^t}\sum_{i\in N^t}\hat{p}^t_{i|z_i^t} = \frac{1}{N^t}\sum_{i\in N^t}p_i^t \ \ldots \text{ which also applies to } \ln p_i. \quad (7)$$

While the denominator of the imputation index in Equation (4) must be counterfactual and use predicted prices, the numerator of Equation (4) can use actual prices. The last expression in square brackets on the right of Equation (4) shows actual period t average prices in the numerator. Use of the last expression in Equation (4) does not require the hedonic regression to be re-estimated in each current period t, a major advantage of this formulation compared to that of Equations (5) and (6) and is the formulation used in practice. Equations (5) and (6) are not advisable. The use of Equation (4) is further developed in this practical sense in Subsections 3.1 and 3.2.

2.4. The Characteristics/Repricing Approach

In contrast to the *imputation* approach, the characteristics approach works at the level of the average values of the *characteristics* of the individual properties. The average residential property in each period t is defined as a tied bundle of the averages of each price-determining characteristic, for example, 2.8 bathrooms, 3.3 bedrooms, 0.8 garages, 0.23 transactions in an up-market location, and so forth.

These average characteristics are valued using period 0 and period t hedonic regressions. This *characteristics* approach answers the question: what would be the price change of a set of average period t characteristics valued first, at period 0 hedonic valuations, and second, at period t hedonic valuations? A ratio of the results is a constant (period t) quality PPI.

A constant-quality *hedonic geometric mean characteristics* (HGMC: $\bar{z}^t$) price index from a semi-log hedonic regression equation is a ratio of geometric means with average characteristics held constant in the *current* period t, $\bar{z}_k^t$:

$$P_{HGMC:\bar{z}^t}^{0\to t} = \frac{\prod_{k=0}^{K}\left(\hat{\beta}_k^t\right)^{\bar{z}_k^t}}{\prod_{k=0}^{K}\left(\hat{\beta}_k^0\right)^{\bar{z}_k^t}} = \frac{\exp\left(\sum_{k=0}^{K}\bar{z}_k^t \ln\hat{\beta}_k^t\right)}{\exp\left(\sum_{k=0}^{K}\bar{z}_k^t \ln\hat{\beta}_k^0\right)} \quad \text{where } \bar{z}_k^t = 1/N^t \sum_{i\in N^t}^{N^t} z_{i,k}^t \tag{8}$$

In Equation (8) each (quality) characteristic, $\bar{z}_k^t$, is held constant at its period t average value in both the numerator and denominator, though valued in the numerator by its corresponding $\hat{\beta}_k^t$ in period t and in the denominator by its $\hat{\beta}_k^0$ in period 0.

A similar index could be equally justified by valuing in each period a constant *reference* period 0 average quality set, $\bar{z}_k^0$:

$$P_{HGMC:\bar{z}^0}^{0\to t} = \frac{\prod_{k=0}^{K}\left(\hat{\beta}_k^t\right)^{\bar{z}_k^0}}{\prod_{k=0}^{K}\left(\hat{\beta}_k^0\right)^{\bar{z}_k^0}} = \frac{\exp\left(\sum_{k=0}^{K}\bar{z}_k^0 \ln\hat{\beta}_k^t\right)}{\exp\left(\sum_{k=0}^{K}\bar{z}_k^0 \ln\hat{\beta}_k^0\right)} \quad \text{where } \bar{z}_k^0 = 1/N^0 \sum_{i\in N^0} z_{i,k}^0 \tag{9}$$

The numerator and denominator of Equations (8) and (9) respectively can use the (geometric) means of actual prices as argued above for imputation indices, based on Equation (7). As with Equations (4) and (5), neither a period 0 constant-characteristics index nor a period t constant-characteristic quantity basket can be considered to be superior. Symmetric use of period 0 and period t characteristics values again makes sense:

$$P_{HGMC:(\bar{z}^0+\bar{z}^t)/2}^{0\to t} = \frac{\prod_{k=0}^{K}\left(\hat{\beta}_k^t\right)^{\bar{z}_k^\tau}}{\prod_{k=0}^{K}\left(\hat{\beta}_k^0\right)^{\bar{z}_k^\tau}} = \frac{\exp\left(\sum_{k=0}^{K}\bar{z}_k^\tau \ln\hat{\beta}_k^t\right)}{\exp\left(\sum_{k=0}^{K}\bar{z}_k^\tau \ln\hat{\beta}_k^0\right)} \quad \text{where } \bar{z}_k^\tau = \left(\bar{z}_k^0 + \bar{z}_k^t\right)/2 \tag{10}$$

In practice, PPIs are not compiled using newly estimated hedonic regressions in each period. To do so would result in undue volatility of the PPIs. The imputation approach uses Equation (4) and the characteristics approach uses Equation (8), both of which require a hedonic regression only for the reference period. Equation (8), although a variant of the characteristics approach, is usually referred to as the re-pricing method, since a set of fixed average characteristics is repriced in subsequent periods – De Haan (2010) and Hill et al. (2018, 223–224). Using hedonic regressions for the repricing and imputation approaches only estimated for the reference period, in a chained link, does not result in undue volatility, and is thus advised - see Subsection 3.1. The method as applied benefits from further tweaks as outlined in Subsection 3.1 and 3.2 below.

There is an economic theory to hedonic regressions and hedonic PPI measurement. For the former, Rosen (1974), Greenstone (2017), and Triplett (2006) provide a characteristics-based supply-demand framework. The price of a property is considered as a tied bundle of its (price-determining) characteristics. The imputation and characteristics/repricing approaches can be seen to have equivalent formulations to traditional index number theory akin to Laspeyres and Paasche indices – Equations (4), (5), (8), and (9) respectively – and the more justifiable superlative Törnqvist – akin to Equations (6) and (10) – and Fisher indices. These superlative formulations are based in economic theory on Erwin Diewert's ground-breaking work – Diewert (1976, 1978) – as shown in Heravi and Silver (2007), Diewert, et al. (2009), De Haan and Diewert (2013), Hill (2013), Silver (2018), and Diewert et al. (2020), amongst others. Such formulations have also been applied at the elementary level of the individual property by Silver (2018).

3. Tweaking Hedonic PPIs for Sparse Data

For a compiler, the greatest challenge is obtaining data. The data needs to be reliable, timely, of the required coverage by location, type of property, and vintage (new vs. existing), preferably in an electronic form and include (preferably transaction) prices and price-determining quality characteristics (Eurostat et al. 2013, chap. 9). There also needs to be a sufficient sample – see Subsections 5.1 and 5.5. Countries with limited sample sizes face particularly difficult problems and PPI methodologists have responded by adapting the above hedonic methodologies to deal with it. Some of this is outlined in Subsections 3.1–3.3 below.

3.1. Use an HPI Methodology that Only Require a Hedonic Regression to be Estimated in the Reference Period

The last term in Equation (4) is a hedonic imputation PPI that only requires a hedonic regression to be estimated in the reference period 0. A requirement to estimate a new hedonic regression in each period not only opens the estimation and compilation of PPIs up to both the vagaries of rushed hedonic estimation, especially in thin markets, but also to an increased delay in publication. There is, however, a possible sample selectivity bias in just using period t characteristics, as shown in Silver (2018), but this is argued to be minimal given the advantages of not having to update the hedonic regression every period.

Hill et al. (2018, 224) rightly note that for the repricing method the reference period should be updated at regular time intervals. "For example, Italy and Luxembourg update the base year every year. However, not all the NSIs using the repricing method update this frequently. Indeed, this is the key problem with the repricing method. It provides a temptation to get lazy and not update the base year. In the empirical comparisons that follow based on Sydney and Tokyo data, we consider two versions of the repricing method. The first never updates the base year, while the second updates it every five years. Our empirical results show that failure to update the base year can lead to drift in the index." Similar issues arise from the use of the imputation index outlined in Equation (4).

In practice, the period over which the hedonic is estimated, say period b, may (partly or in its entirety) precede the index reference period, 0, which in turn precedes the current periods, t. For example, a hedonic regression might be estimated using 2020 data, with an index reference period of 2020Q4 = 100.00, and current periods of 2021Q1, 2021Q2,. . .continuing until a new hedonic regression is estimated and chain-linked to the previous series, much like the rebasing of a consumer price index (CPI) – see Appendix (Section 10).

3.2. Use an Extended Period for the Reference Period

For sparse data, the reference period should be over an extended period, rather than a quarter. First, there may not be an adequate number of observations and/or sufficient variation in the characteristics of the sample of properties transacted. An extended period 0 regression will be more likely to better encompass the characteristics of period t properties as well as basing the regression on a larger sample size providing coefficient estimates with more degrees of freedom. The repricing variant of the characteristics approach used by eight countries in Europe has an extended reference period of a year.

3.3. Smoothing and Sparse Data

Sparse transaction data can result in volatile PPIs. Volatile series are not useful to macroeconomic policy-making that seeks to identify turning points. Issues of sparse data can arise from a desire to unduly stratify the sample to benefit from enabling weights to be applied to each stratum, an issue considered in Subsections 6.4 and 6.5.1. Smoothing techniques can be applied as part of the PPI compilation, as a rolling window formulation of a time dummy approach, as recommended by O'Hanlon (2011) and IMF (2020) for volatile series. However, a four-quarter rolling window effectively centers the PPI on a period six-months in the past. This lag will be accentuated by untimely data retrieved from the data source and the time required to clean and use it. A four-quarter rolling window would further constrain the estimated coefficients to be constant over each four-quarter average. Thus, while an extended period and smoothing both serve as effective strategies against problems of small samples, they are not costless in terms of lagging and smoothing the resulting PPIs. There is a case for publishing the volatile series along with simple moving averages, something that promotes transparency. An alternative strategy is to reduce the frequency of the PPI series, from say monthly to quarterly or quarterly to bi-annually. Rambaldi and Fletcher (2014) and Hill et al. (2020) demonstrate a Kalman filter approach that optimally links and weights past information yielding a stable set of (time-varying) hedonic estimates (conditional on current and past market information) even if samples are small. There are other methods of measuring property price indices where there are sparse transactions including Geltner (1993), Bokhari and Geltner (2012) and Silver and Graf (2014).

3.4. Weights at an Elementary Level

A benefit of the matched-paired comparisons generated by the imputation approach is that weights can be readily and intuitively applied at the elementary level. The *expenditure* weight for an individual property is its relative price and data on relative prices exist in real time for each property and can be applied to each respective property's price change. Similar considerations apply to the characteristics/repricing approach; a weighted average of the characteristics is calculated: a less intuitive, though equivalent approach to that proposed for weighted hedonic imputation PPIs. The weighted version of Equation (6) is given by Equation (11) as a quasi-hedonic formulation of a Törnqvist index.

$$\begin{aligned} P^{0\to t}_{QToHGMI:z_i^t} = \prod_{i\in N^t}\left(\frac{\hat{p}^t_{i|z_i^t}}{\hat{p}^0_{i|z_i^t}}\right)^{\hat{\omega}_i^\tau} &= \exp\left(\sum_{i\in N^t}\hat{\omega}_i^\tau \ln\left(\frac{\hat{p}^t_{i|z_i^t}}{\hat{p}^0_{i|z_i^t}}\right)\right) \\ &= \exp\left(\sum_{i\in N^t}\hat{\omega}_i^\tau \ln\left(\hat{p}^t_{i|z_i^t} - \hat{p}^0_{i|z_i^t}\right)\right) \end{aligned} \tag{11}$$

where $\hat{w}_i^T = \frac{1}{2}\left(\frac{\hat{p}^t_{i|z^t_i}}{\sum_{i\in N^t}\hat{p}^t_{i|z^t_i}} + \frac{\hat{p}^0_{i|z^t_i}}{\sum_{i\in N^0}\hat{p}^0_{i|z^t_i}}\right)$, an index that has excellent properties in economic theory as a superlative index. It is "quasi" in the sense that it does not make use of the sample of period 0 transactions. It is "superlative" in the sense that the index of price changes of transactions undertaken in period 0 makes symmetric use of reference and current period weight (price – expenditure) information (Diewert 1976; Balk 2008). Silver (2018) shows how a weighted version of Equation (8) can also be formulated as can a superlative index that makes use of both base period 0 and current period t transactions. A similar formulation can be made for the characteristics approach, but not so for the time dummy approach, a major deficiency of the approach, see Silver (2002, Appendix 1, 2016 Annex 2, 2018; Silver and Heravi 2005).

Of importance is the fact that a weighted PPI at this elementary level requires no further data. The price of an individual property is its transaction expenditure, and its weight, the relative price. In principle stock weights are preferred, but these real-time transaction weights will be more suitable than an implicit equally-weighted (unweighted) index, unless a democratic-weighted index is required.

Weights at the higher level can be the relative values of transactions or stocks of properties for each stratum. This choice between the use of "transactions" or "stocks" as weights depends on the purpose of the property price index and availability of adequate data on the stock of properties. Fenwick (2013) and Mehrhoff and Triebskorn (2016) outline issues relevant to such a choice – Subsection 6.5 considers weighting between stratum in more detail.

3.5. *Some Equivalences*

The three hedonic approaches have different, yet valid, intuitions. Yet as long as the functional form of the aggregator is properly aligned to the hedonic regression in the manner shown in Table 1, the imputation and characteristics/repricing approaches yield the same results. This consolidation not only markedly narrows down the choice between approaches but validates these two approaches as ones that provide identical results but arise from quite different intuitions.

For a semi-log functional form of a hedonic regression, the requirements are that (1) for the characteristics approach, $\bar{z}_k^0$ and $\bar{z}_k^t$ are *arithmetic* means of characteristic's values, the right-hand-side (RHS) of the hedonic regression, and (2) for the imputation approach, the ratio of average predicted prices is a ratio of *geometric means*, the left-hand-side (LHS). Similar equivalences shown in Table 1 to apply to linear and log-log forms. While Hill and Melser (2008) confine the equivalences to the semi-log hedonic model, they identify the same property.

Table 1. Equivalences of hedonic approaches.

Hedonic regression: Functional form	Characteristics approach: Form of average of characteristics	Imputation approach: Form of average of predicted prices
Linear	Arithmetic mean	Arithmetic mean
Semi-log	Arithmetic mean	Geometric mean
Log-log	Geometric mean	Geometric mean

More formally: a log-linear hedonic characteristics PPI with constant reference-period average characteristics, $\bar{z}_k^0 = 1/N^0 \sum_{i \in N^0} z_{i,k}^0$, is equal to a hedonic imputation index for reference period 0 properties:

$$P_{HGMI:z_i^0}^{0 \to t} = \frac{\prod_{k=0}^{K} \left(\hat{\beta}_k^t\right)^{\bar{z}_k^0}}{\prod_{k=0}^{K} \left(\hat{\beta}_k^0\right)^{\bar{z}_k^0}} = \frac{\exp\left(\sum_{k=0}^{K} \bar{z}_k^0 \ln \hat{\beta}_k^t\right)}{\exp\left(\sum_{k=0}^{K} \bar{z}_k^0 \ln \hat{\beta}_k^0\right)} = \frac{\exp\left(\frac{1}{N^0} \sum_{k=0}^{K} \sum_{i \in N^0} z_{i,k}^0 \ln \hat{\beta}_k^t\right)}{\exp\left(\frac{1}{N^0} \sum_{k=0}^{K} \sum_{i \in N^0} z_{i,k}^0 \ln \hat{\beta}_k^0\right)}$$

$$= \frac{\exp\left(\frac{1}{N^0} \sum_{i \in N^0} \sum_{k=0}^{K} z_{i,k}^0 \ln \hat{\beta}_k^t\right)}{\exp\left(\frac{1}{N^0} \sum_{i \in N^0} \sum_{k=0}^{K} z_{i,k}^0 \ln \hat{\beta}_k^0\right)} = \frac{\prod_{i \in N^0} \left(\hat{p}_{i|z_i^0}^t\right)^{\frac{1}{N^0}}}{\prod_{i \in N^0} \left(\hat{p}_{i|z_i^0}^0\right)^{\frac{1}{N^0}}}. \tag{12}$$

and similarly, an index with average characteristics held constant in the current period t, $\bar{z}_k^t = 1/N^t \sum_{i \in N^t} z_{i,k}^t$, is equal to an imputation index for current period t properties given by:

$$P_{HGMI:z_i^t}^{0 \to t} = \frac{\prod_{k=0}^{K} \left(\hat{\beta}_k^t\right)^{\bar{z}_k^t}}{\prod_{k=0}^{K} \left(\hat{\beta}_k^0\right)^{\bar{z}_k^t}} = \frac{\prod_{i \in N^t} \left(\hat{p}_{i|z_i^t}^t\right)^{\frac{1}{N^t}}}{\prod_{i \in N^t} \left(\hat{p}_{i|z_i^t}^0\right)^{\frac{1}{N^t}}}. \tag{13}$$

4. Diagnostics: General Principles

A regression equation is fitted to data by means of an *estimator*, which is a rule for determining the estimated coefficients of the equation that best fits the sample data. The ordinary least squares (OLS) estimator minimizes the sum of squared vertical differences between the actual points and the line and is the go-to estimator for most econometric and statistical regression work. The focus of this article will be on the assumptions and data issues for an OLS estimator though consideration will also be given to the use of alternative estimators, as needs be. Each of the following subsets of diagnostics will be considered in two respects: their effect on: (1) an estimated hedonic regression equation and (2) on the use of the estimated regression for PPI measurement in terms of the three hedonic approaches outlined in section II.

The provision of a comprehensive set of diagnostic tests is important. The estimated regression model must be seen to be reliable by conventional statistical/econometric standards and should be presented in a detailed working paper. There is an extensive literature on OLS diagnostic tests. This includes many graphical devices whose albeit subjective evaluations give more insights than the one-off results of numerical measures and hypothesis tests. However, the focus here will remain on tests, rather than these graphical devices, since compilers need to have clear and transparent, though not necessarily optimal, rules by which they decide on the suitability of a model.

Graphical devices, such as for outlier detection and heteroscedastic errors, give insight into the nature of problems in the specification of the hedonic regression, that can perhaps then be rectified. For example, if standardized residuals are seen in a plot to have a larger variance at the top end of fitted *prices*, then the regression may need to include finer locational variables for high-end geographical areas or squared terms for the *size* of a house and/or interaction terms of particular geographical *areas* with *size*.

Our concern here is with the simple case of a cross-sectional OLS multiple linear regression (MLR), for which there are five (Gauss-Markov) assumptions: (1) the model in the population can be written in a linear form in the β_k coefficients, as in Equation (1); (2) a random sample of observations – data missing at random will not result in bias; (3) there is no *perfect* multicollinearity; (4) the residual error from the model has an expected value of zero – which can fail if the hedonic regression is misspecified; and (5) the residual error has the same variance (is homoscedastic) given any value of the explanatory variables. Under these assumptions the OLS estimator of the $\hat{\beta}_k$ explanatory variables is the best linear unbiased estimator (BLUE). We will examine the requirements of a satisfactory regression model: their detection (diagnostic measures), implications if not met, rectification of failure, and other issues. This will be undertaken in Section 5 for the basic diagnostics, as constitutes the default output from R; Section 6 for multicollinearity, specification, and omitted variable bias; Section 7 for heteroscedasticity and normality of residuals; and Section 8 for outlier detection and treatment. Conclusions are drawn in Section 9.

5. The Basic Diagnostics

These include the sample size (n); standard error of the regression (SER); F-test; R-bar squared ($\bar{R}^2$); and standard errors, t-tests and p-values of individual estimated coefficients.

5.1. The Sample Size (n)

For a multivariate regression, n is less important than the degrees of freedom, $df = n\text{-}k\text{-}1$, where k is the number of explanatory variables and a further 1 is deducted for the constant (intercept). A coefficient is estimated for the constant and each of the k terms in a multiple regression model. Each additional explanatory variable reduces the degrees of freedom: the observations or information content, available to estimate the coefficients. A sample size of 40, for example, for a regression equation with $k = 19$ explanatory variables reduces the effective sample, the df, to 20.

Statistical theory shows there is no rule of thumb for a required sample size. To determine the required sample size of a sample estimate it is the confidence interval, based on the standard error of the sample estimate, that matters. Our concern is that the sample size for the hedonic regression is large enough to yield precise estimates of the individual coefficients of the explanatory variables. Precise estimates are those whose confidence intervals are sufficiently narrow and statistical significance tests have sufficient power, to reject a false null hypothesis, at a given level of probability. The null hypothesis is that there is no difference between the estimated coefficient and zero (no effect). To reject it at a given level of probability is to say the magnitude of this difference is over and above that due to sampling error. The width of the confidence interval for an individual estimated coefficient in a multiple regression is based mainly on the standard error of the estimated regression coefficient. The determinants of such a standard error are not solely limited to the sample size and, in any event, are particularly problematic for small samples, as outlined in Subsection 5.5.

Many countries, may have limited sample sizes of transactions for all manner of reasons including the limited size of their population; poor records of possibly out-of-date, unreliable, transactions, advertised or loan data; relatively low stock of residential/commercial property; high proportion of properties rented; heterogeneous property; and cultural practice of passing on property within a family over generations. Problems of bias in the data source, say due to tax avoidance, can be particularly problematic.

5.2. The Standard Error of the Regression (SER)

The variance of the error of the regression is $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(p_i - \hat{p}_i)^2}{(n-k-1)} = \sum_{i=1}^{n} \frac{\hat{u}_i^2}{(n-k-1)}$, that is, the sum of squared residuals (SSR) divided by the *df*. The square root of $\hat{\sigma}^2$ is $\hat{\sigma}$, the *standard error of the regression* (SER), also referred to as the residual standard error and root mean squared error. The measure is similar to the mean squared error (MSE) which is a simple average of the squared residuals divided by n rather than *df*.

SER provides an absolute measure of the average (vertical) distance between the (log of) actual p_i in Equation (1) and their corresponding fitted values $\hat{p}_i$ from the estimated regression equation in Equation (2). It differs from the mean squared error (MSE) which has as its denominator the sample size n, rather than *df*. SER is a measure of how well the regression fits the data and benefits from being in the same units of measurement, (log) price, as the observed data. Since it is based on an OLS regression, greater emphasis is given in this measure of the average residuals to those points lying well above/below the line, the differences being squared. The measure will be seen in Subsection 5.5 to form a component of the standard error and the confidence intervals for each estimated coefficient, $\hat{\beta}_k$, in Equation (2) of the hedonic regression. The measure is essentially the standard deviation of the residuals. A smaller SER indicates the regression model is providing a better fit to the data.

5.3. F-test

The F-test provides a test of whether the difference between the estimated model and one with no explanatory variables is statistically significant: whether the model has any explanatory power. Passing an F-test is the bare bones of acceptability of a model, especially for a large sample. Our real concern is that when we reject the F-test's null hypothesis of the model has no explanatory power, how good is the fit of the model?

5.4. R-bar Squared ($\bar{R}^2$)

A commonly used measure of how well an estimated model fits the data is $\bar{R}^2$ (R-bar squared). Where $\sum_{i=1}^{n}\left(p_i - \hat{p}_i\right)^2 = \sum_{i=1}^{n}\hat{u}_i^2$ (SSR) is the sum of the squared residuals – a higher value reflects a poorer fit of the model; $\sum_{i=1}^{n}\left(\hat{p}_i - \bar{p}\right)^2$ (SSE) is the explained sum of squares: the sum of the variation in the predicted price, $\hat{p}_i$ around its mean (for OLS $\bar{p} = \bar{\hat{p}}$); and $\sum_{i=1}^{n}\left(p_i - \bar{p}\right)^2$ (SST) is the total variation in p_i. SST = SSE + SSR and $\bar{R}^2$ = [1-(SSR/(n-k-1)]/[SST/(n-1)], which is the proportion of sample variation in p_i that is explained by the OLS regression. The division of SSR by *df* (denoted by the bar above R in $\bar{R}^2$ bar squared) adjusts (penalizes) this goodness of fit measure for including additional explanatory variables. $\bar{R}^2$ is commonly used to decide whether additional explanatory variables should be included or excluded in the estimated hedonic regression. There are a number of reasons why $\bar{R}^2$ may not be a suitable measure of how well an estimated model fits the data.

First, the hedonic imputation and characteristics approaches, sections II.3 and II.4 respectively, are based on out-of-sample predicted prices. For example, in the Equation (4) imputation approach a hedonic regression is estimated for period 0 and used to predict the price of the characteristics of each individual property transacted in period t, as $\hat{p}^0_{i|z^t_i}$. Similar considerations apply to the characteristics/repricing approach. $\bar{R}^2$ is based on in-sample predicted prices in period 0 which is not wholly relevant to the model's out-of-sample predictive performance.

Second, $\bar{R}^2$ will increase only if the t-statistic on the additional explanatory variable exceeds unity, that is, even if its estimated coefficient is not statistically significant.

Third, say there was a variable in the regression relating to the income of the purchaser, or preferential loan or tax breaks associated with a characteristic of the purchaser, such as being a first-time buyer. The $\bar{R}^2$ of the regression might increase due its inclusion since it may be a contributory explanatory factor that explains price variation. However, such variables should be excluded from a hedonic regression for a PPI index since they do not relate to changes in the quality-mix of the characteristics of the property. Higher property prices due to higher incomes and/or tax incentives should not be controlled for in a hedonic regression; such effects are part and parcel of measured property price inflation.

Finally, it is possible for regression models with low $\bar{R}^2$ to have precise estimates of individual coefficients. However, models with a relatively low $\bar{R}^2$ will generally not have a high predictive power, the essence of hedonic imputation and characteristic/repricing approaches – Wooldridge (2013, 193).

5.5. *Standard Errors, t-tests and p-values*

Our concern is with the precision of an individual estimated hedonic regression coefficient. Precision is defined as the width of the confidence interval surrounding a coefficient estimate and is determined by $\mp 2 \times se\left(\hat{\beta}_k\right)$ – for a 95% the level of confidence. $se\left(\hat{\beta}_k\right)$ is the (asymptotic) standard error for the OLS estimate of the regression coefficient of the kth explanatory variable and is given by:

$$se\left(\hat{\beta}_k\right) = \frac{\hat{\sigma}}{\left[SST_k\left(1 - R_k^2\right)\right]^{1/2}} = \frac{\hat{\sigma}}{\sqrt{n}sd(z_k)\sqrt{1 - R_k^2}} \tag{14}$$

where $\hat{\sigma}$ is the standard error of the regression (SER) defined in Section 5.2 above; $SST_k = \sum_i^n (z_{ik} - \bar{z}_k)^2$ whose square root is ($\sqrt{n}$ times) the standard deviation of the explanatory variable k, say the number of *bedrooms*, $sd(z_k)$; and R_k^2 is the R^2 from an (auxiliary) regression of *bedrooms*, on the K-1 remaining explanatory variables. This is outlined in more detail in Subsection 6.1 on multicollinearity. The precision of a coefficient estimate for an individual explanatory variable is, from Equation (14), dependent not only on (1) the (square root of its) sample size n, but also on (2) the overall fit of the regression, $\hat{\sigma}$; (3) the interdependence between the explanatory variable and other variables in the regression, R_k^2; and (4) the standard deviation of the explanatory variable, $sd(z_k)$. Contributory factors to an estimated coefficient for an explanatory variable having a wide confidence interval are: low sample sizes; ill-fitting regression models; high interdependence with other explanatory variables; and, perversely, a low standard deviation for the values z_k. A concern of this section is to warn the reader not to solely judge what an adequate sample should be on the basis of its sample size, n. An adequate sample size is one that provides a precise coefficient estimate – narrow confidence intervals and a satisfactory power to the tests of significance. Since tests of hypotheses – and confidence intervals – are (asymptotically) distributed with t_{n-k-1} degrees of freedom a larger sample size is required to take account of the additional explanatory variables that make up a hedonic regression, as mentioned in Subsection 5.1 above, see also Wooldridge (2013, 167–170).

However, for large samples, as n increases, the marginal reduction in the standard error becomes smaller – as a result of $\sqrt{n}$, not n, being in the denominator – and there is often little to be gained from having very large sample sizes, in terms of reductions in $se\left(\hat{\beta}_k\right)$.

For example, if n increases by 300, from 100 to 400, $\sqrt{n}$ doubles: from 10 to 20 and can be expected from Equation (14) to have a marked effect in reducing $se\left(\hat{\beta}_k\right)$. However, if n increases by 300 from 10,000 to 10,300, $\sqrt{n}$ only increases by 1.5% – from 100 to 101.5 – leading to a much smaller reduction in $se\left(\hat{\beta}_k\right)$. Wooldridge (2013, 169–170) argues that since, aside from $\sqrt{n}$, all other terms in Equation (14) for large samples converge to constants, it is the sample size that matters. However, since estimated hedonic regressions for PPIs should be regularly updated (see Subsection 3.1) it is apparent that improving the fit of the regression over time will result in decreasing $\hat{\sigma}$ and that modelling, rather than not overly deleting, extreme values of explanatory variables results in a larger $sd(z_k)$. These are useful strategies to reduce $se\left(\hat{\beta}_k\right)$, especially since they will also likely improve the out-of-sample predictive power of the hedonic regression.

The $se\left(\hat{\beta}_k\right)$ is not only the prime determinant of the width of confidence intervals (at a given level of confidence) around the individual coefficient estimates of the explanatory variables, but also the basis of tests of a null hypothesis that $\beta_k = 0$ – the variable z_k has no explanatory power. A p-value less than 0.05 rejects at a 5% level of significance the null hypothesis that $\beta_k = 0$. The hypothesis test is phrased in a manner that should a property characteristic, say possession of a garage, have an impact on price, the null hypothesis will be rejected, at a 5% level of significance, with a p-value of less than 0.05. We would expect possession of a garage to have a positive statistically significant coefficient. Sensible signs, magnitudes, and statistically significant test results are a primary consideration in evaluating regression results. Such test results say nothing about whether the coefficient has a substantive effect. Only that the difference between its estimated value and zero is over and above that due to sampling errors which, for a large sample size, may be very small.

It is stressed that for the time dummy PPI in Subsection 2.2 the null hypotheses for a t-test and associated p-value for each time dummy is that, in Equation (3), $ln\delta^t = 0$, that is $exp\left(\hat{\delta}^t\right) = 1$; there is no change in the rate of property price inflation, having conditioned on changes in the quality-mix of the bundle of properties transacted. However, it may well be that with low property price inflation a failure to reject a null hypothesis of no change is a proper reflection of quality-mix-adjusted property price change. In using the standard error for a time dummy coefficient when inflation is low, it is better to look at the width of the confidence interval than the failure or otherwise to reject a null hypothesis test of zero price inflation.

The hedonic characteristic/repricing and (equivalent) imputation approaches, Subsections 2.3 and 2.4, both rely on predicted average values and these too have associated standard errors for which we can derive confidence intervals. These confidence intervals are for average values and differs from those for individual property prices i such as $\hat{p}^0_{i|z_i^t}$ in Equation (4), see Wooldridge (2013, 200–203) and Madalla and Lahari (2009, 154).

6. Multicollinearity, Specification, and Omitted Variable Bias

This section includes help with the detection, interpretation, and impact of multicollinearity in hedonic PPIs; the specification of hedonic regressions including omitted variable bias and inclusion of irrelevant variable bias; introduction of curvature (logs and powers); and stratification and interaction effects.

6.1. Multicollinearity

Multicollinearity is where the explanatory variables are highly intercorrelated. As the degree of multicollinearity increases, the regression model estimates of the coefficients become unstable and the standard errors for the coefficients can get wildly inflated, as explained by the inclusion of R_k^2 in Equation (14). Confidence intervals can become very wide, tests of statistical significance for individual coefficients unsatisfactory, and the signs on coefficients estimates may be counter intuitive – wrong. It is difficult to disentangle the effects of individual variables. The problem is one of the sample of data, not the estimation or specification of the regression. For example, number of *bedrooms*, square footage *size*, and number of *bathroom*s all relate to the size of a property and may be multicollinear, but there may be an added price premium on larger houses that have the same *bedrooms*, but extra *bathrooms*. If the data – the quality mix of houses transacted in a period – do not have a sufficient sample of larger properties with similarly sized *bedrooms* but different number of *bathrooms*, then the regression cannot isolate and identify the partial marginal valuation of *bathrooms*. The problem often lies with the data: the available properties transacted for that period, one that cannot be rectified.

First, we need to measure which if any coefficients are multicollinear: the *variance inflation factor* (VIF) is a satisfactory measure:

$$VIF = 1/\left(1 - R_k^2\right) \tag{15}$$

where R_k^2, as in Equation (14), is the R-squared from regressing an individual z_k on all other explanatory variables.

For each estimated coefficient in a regression, a *VIF* can be readily produced; for example, in the software R by the command *vif(mod1)* following the *lm* regression command. As a rule of thumb, a variable whose *VIF* is greater than 10 should merit further investigation. Tolerance, defined as 1/*VIF*, is also used to check on the degree of multicollinearity. A tolerance value lower than 0.1 is comparable to a *VIF* of 10. When there is a perfect linear relationship among the explanatory variables, the estimates for a regression model *cannot* be uniquely computed.

The importance of *VIF* statistics for PPI measurement is when results are presented for the hedonic regressions and intuitively important variables do not have the expected sign or magnitude. With *VIF* statistics shown alongside the individual estimated coefficients, standard errors, and *p*-values, multicollinear variables can be identified. A methodological note can explain to users that multicollinear variables are indeed valuable to the regression, it is just that their effects, as shown by the *VIF*, are entangled. The advice is to always publish *VIF*s. If the *VIF*s are low and estimated coefficients statistically significant, there is no problem. Where they are high the user is warned to not give too much credence to the individual *p*-values.

Such analysis is embodied in econometrics in the general to specific methodology: econometrics favors parsimonious models to general ones. It is elegant to have a simple model explaining a lot and, furthermore, saves on degrees of freedom which is important when you have small samples. When there are unexpected results, such as the estimated coefficient on *size* not being statistically significant, but having a high *VIF*, *size*-related variables with high *VIF*s can be individually dropped from the regression and the hedonic model re-estimated. Should the coefficient on *size* re-emerge as statistically significant the effect of multicollinearity can be explained to the user and *size* kept in the model. Thus, individual variables with high *VIF*s can be individually dropped from the regression to demonstrate to the user how their effect on price of their multicollinear counterparts are entangled. However, together such variables still usefully explain prices. All of this requires an ease of tabular layout for different hedonic models with different variables dropped. Since this is common econometric and statistical practice, R has such a facility in *library (stargazer)*. For PPI purposes a general model is preferred to include variables that are demonstrably multicollinear, even if their estimated coefficients are not statistically significant. In this way we demonstrate to users that their inclusion is not arbitrary, include entangled effects, and help with out-of-sample predictions.

Multicollinearity will occur when you add interaction terms and squared terms. There is a procedure called "centering" – subtracting the mean for each observation's values for continuous variables – that can help with the interpretation of main effects, interaction powers of the individual coefficients concerned – Iacobucci et al. (2016).

A hedonic time dummy variable approach is fine as long as the multicollinearity is with the "control" variables – the explanatory variables that control for quality characteristics – and not with the time dummy variables. Time dummy multicollinearity with other explanatory variables would arise if the sample of data, for example, comparing periods 0 and *t*, had much larger (*size*) properties in period *t* than period 0: much larger to the extent that precise estimates of whether the change in price in period 1 is due to the passage of time or (the control for) the *size* of house. The two effects are too muddled together to individually identify them. Such high multicollinearity may happen in small samples of transactions, but is most unlikely in large ones. It can be investigated if there is a worrying spike or drop in prices. (A feature of multicollinearity is that new samples can lead to unwarranted large changes in coefficient estimates from one sample (period) to the next). In such an instance, the time dummy and *size* would have high VIFs and dropping one would make a difference).

The hedonic imputation (and characteristics) approaches are based on predicted values which are unbiased when explanatory variables are multicollinear. Yet our concern is with out-of-sample prediction. For example, in the imputation approach, Equation (4), a hedonic regression is estimated for the property price transactions in period 0 and used to predict the price of the characteristics of each individual property transacted in period t, as $\hat{p}^0_{i|z_i^t}$. The hedonic regression estimated using period 0 transactions should include a wide range of explanatory variables to capture the period t sample characteristics which may differ in its balance of locational, size, and other variables, from the period 0 sample. The period 0 hedonic regression must be sufficiently general to act as a spider's web to capture and predict period t out-of-sample characteristics. It may of course be the case that the period 0 transactions data is sufficient to disentangle the multicollinearity and allow predictions for period t data even if period t data would fail to do so. In either case, the omission of seriously multicollinear variables is unwarranted for this purpose. Similar arguments apply to the characteristics/repricing approach.

6.2. *Omitted Variable Bias (OVB) and Inclusion of Irrelevant Variable Bias (IIVB)*

6.2.1. Omitted Variable Bias (OVB)

OVB can take many forms. First, it can arise because data on a pertinent price-determining variable is simply not available, for example whether a house has a *view* (of different types/standards) or whether it is located in a particular part of say Washington D.C. The "part" may be one of the eight wards that make up DC, or even finer, the 270 zip (post) codes. Prices for similar houses can vary greatly by such locational variables, never mind between larger geographical areas.

The omitted variable may also arise from the specification of the hedonic function, in that it omits say interaction terms, powers, or log transformations of existing explanatory variables. If in a linear model, for example, only *size* is included and *size*^{2} is excluded, the hedonic regression will not capture a, for example, declining marginal value for *size* (see also Subsection 7.3). There may also be interaction effects between variables, say *size* and *location*: the marginal value of an additional square foot in one particular Ward in D.C. is likely to be quite different from another, but a functional form without interaction effects is constrained to only have single coefficient estimates for *size* and (dummies for) *Wards*.

With omitted variables, the coefficients of included variables may be biased. While the nature and direction of bias can be identified for models with two explanatory variables of which one is omitted, it is more complex where there are more than two explanatory variables, as would be the case for meaningful hedonic PPIs, see Wooldridge (2013, 87–88).

Wooldridge (2013, Appendix 3A, 681–682) shows the effect on OLS estimated coefficients $\hat{\beta}_j$, using the full set of $j = 0, 1, \ldots..k$ explanatory variables, of omitting a single variable, z_k.

$$\tilde{\beta}_j = \hat{\beta}_j + \hat{\beta}_k \tilde{\gamma}_j \tag{16}$$

$\tilde{\beta}_j$ is an estimate of the partial effect, having controlled for other variables in the regression, of an individual variable z_j on *price* in a regression equation that omits a variable z_k, say due to data not being available. $\hat{\beta}_j$ is its counterpart estimate were the omitted variable z_k

included. The last term is the omitted variable bias, $\left(\tilde{\beta}_j - \hat{\beta}_j\right) = \hat{\beta}_k \tilde{\gamma}_j$: the difference between an individual estimated coefficient having excluded and included z_k. It is equal to the (unknown) estimated coefficient $\hat{\beta}_k$ on variable z_k were data available and had it been included in the regression. $\hat{\beta}_k$ is *multiplied by* $\tilde{\gamma}_0$, the estimated coefficient on the included variable z_j in a regression of the omitted variable z_k on the remaining $j = 0, 1, \ldots ..k - 1$ explanatory variables. In a hedonic regression based on a sparse variable set it is unlikely that either or both of these conditions will be even approximately met.

Using the time dummy approach of Subsection 2.2, our concern is with the individual coefficients on the time dummy – these are the $\tilde{\beta}_j$ in Equation (16). Consider $j = 1$, the first time dummy, say for quarter 2 (Q2), benchmarked on the constant, Q1.It is apparent from Equation (16) that $\tilde{\beta}_j$ is unbiased if $\hat{\beta}_k = 0$, the omitted variable has no effect on price. But this is a trivial result that ignores the real problem of the likely omission of important price determining variables.

Yet even if $\hat{\beta}_1 \neq 0$, there would be no OVB if $\tilde{\gamma}_j = 0$ (and *vice versa*). Say the omitted variable was number of *bedrooms*. $\tilde{\gamma}_j$ in Equation (16) is the estimated coefficient on the time dummy of interest in an auxiliary regression of *bedrooms* on the remaining $j = 0, 1, \ldots ..k - 1$ explanatory variables, that includes the time dummies. Thus, if we believe, for we have no data, that a shift in the data between Q1 and Q2, the time dummy, having controlled for other variables in the regression, does not explain the number of *bedrooms* transacted, then there is no OVB in this instance. Bear in mind that the sample in Q1 is drawn from the happenstance of the transactions taking place and similarly for Q2. Were we in a property price bubble, or leaving one, it might be that the sample of transactions becomes biased to a particular type of property that is not accounted for by other variables in the model. However, this is something for the individual statisticians familiar with the residential/commercial property markets in their countries to determine. If they believe $\hat{\beta}_1$ and $\tilde{\gamma}_j$ are likely to be substantial, then it becomes important to fill this data gap. If not, there is an albeit quite technical explanation as to why the excluded variable may not matter in this context.

Subject to reasonable predictive ability, part of which is the $\bar{R}^2$, a predictive model can only be reasonably assessed on its ability to obtain optimal predictions based on the data available to it. What is important is that those compiling/estimating a hedonic PPI have a longer-term strategy to ascertain which excluded variables can be feasibly, given any resource and time constraints, developed and included in the model, and a program put in place to start the process. Three areas come to mind. The first is a finer breakdown of locational dummy variables within target locations such as larger cities (Hill and Scholtz 2017). This may include in explicit locational variables using distance to facilities of individual properties, as outlined Subsection 6.4 below. The second, a merger of the PPI database with dwelling's population census data to give an indication of socio-demographic characteristics of the neighborhood dwelling in lie. Third, there are developments in the quality of properties and the environment including eco-friendly real estate with internet-of-things (IoT) capabilities (Fuerst and Shimizu 2016) and variables relating to environmental factors such as air pollution (Greenstone 2017).

6.2.2. Inclusion of Irrelevant Variable Bias (IIVB)

The counterpart to OVB is bias due to the inclusion of irrelevant variables: an over-specified regression. Yet with irrelevant variables included, the estimates of the coefficients for both included relevant and irrelevant variables are unbiased; and that includes the estimated coefficient for the irrelevant variable being zero. However, assuming large samples, there may be a cost to the inclusion of irrelevant variables.

The standard errors of the estimates for the relevant variables are higher due to the inclusion of the irrelevant variable(s). It was shown in Equation (14) that the standard error of the estimated coefficient, $se(\hat{\beta}_k)$, of a relevant explanatory variable increases as $1/\sqrt{(1-R_k^2)}$ increases. Bear in mind that R_k^2 is the R^2 from an (auxiliary) regression of the explanatory variable of interest, z_k, on the remaining explanatory variables. While included irrelevant variables will have no partial effect on price, they may be multicollinear with included relevant ones thus increasing $se(\hat{\beta}_k)$ when compared to a regression model that excluded irrelevant variables.

6.3. *Functional Form: Curvature*

Specification improvements include ensuring curvature in the relationship is properly represented. Curvature has already, to a large extent, been achieved via the use of the semi-logarithmic functional form. If necessary, it could be further developed within a semi-logarithmic model by the inclusion of squared (quadratic) explanatory variables, as outlined in Wooldridge (2013, 188–190). Alternatively, curvature could be introduced into a linear form – without logarithmic transformations – simply by including powers as explanatory variables, as illustrated below.

The interpretation of a coefficient is quite straightforward when a variable appears once as an explanatory variable. However, as is the case for the inclusion of powers of a variable and, in the next section, an interaction term, the partial (first derivative) effect of the variable appearing more than once is more complicated. Consider the following hedonic regression of *price* on *size*, number of *bedrooms*, and *bathrooms*, based on Wooldridge (2013, 183–188).

The functional form used in Subsections 6.3, 6.4, 6.5.b is the linear form. Similar reasoning applies to the semi-logarithmic form though the interpretation of the estimated β differs. Woodridge (2013, 186–190 and 230–238) uses both linear and semi-log models to explain such effects, though the evaluation of partial effects in Equations (18) and (21) is more clearly explained here, and in Wooldridge (2013, 712), via a linear model.

$$price = \beta_0 + \beta_1 size + \beta_2 bedrooms + \beta_3 size{\cdot}bedrooms + \beta_4 bathrms + \beta_5 bathrms^2 + u \tag{17}$$

The inclusion of higher powers, such as quadratic and cubic terms, introduces curvature into the relationship and is readily achievable in modern software – in R>lm(logprice ~ size + I(size)^2 + I(size)^3). Equation (17) also includes a squared term of the number of *bathrms*. The partial derivative of *bathrms* – effect of *bathrms* on *price* having controlled for other explanatory variables in the regression – is:

$$\frac{\Delta price}{\Delta bathrms} = \beta_4 + (2 \times \beta_5) bathrms \tag{18}$$

Usually evaluated at the mean of *bathrms*. Of note, the individual β_4 and β_5 may not be statistically significant, but jointly, the two effects may be so. Tests for the joint significance of variables include the Wald, Likelihood ratio, and F-tests – for example, in R: wald.test – Wooldridge (2013, 564–565).

6.4. Functional Form: Interaction Effects

A particularly important use of interaction terms arises with the modelling of location and housing types: "*apartments*" and "*houses*" for a RPPI or "*office* "and "*retail*" for a CPPI. As a simplified example, say, for a hedonic RPPI, a city only has residential apartments and is divided by a railway track into two distinct areas, A and B. An appealing strategy is to separately estimate hedonic regressions and develop RPPIs for both A and B. A weighted average of the resulting component RPPIs for A and B can be compiled as an RPPI for the city. The approach has the advantage of not restricting the coefficients for A and B to be the same. Let the two regressions be:

$$price = \beta_0^A + \beta_1^A size + \beta_2^A bedrooms + \beta_3^A bathrms + u^A \text{ and} \tag{19a}$$

$$price = \beta_0^B + \beta_1^B size + \beta_2^B bedrooms + \beta_3^B bathrms + u^B \tag{19b}$$

Alternatively, a dummy intercept variable could be created: $track = 0$ if the property was on the A side of the track, and 1 otherwise. A dummy slope interaction variable is also included in Equation (20): $size{\cdot}track = 0$ if the property was on the A side of the track, and *size* otherwise. Equations (19a) and (19b) can be phrased as a single equation:

$$\begin{aligned} price = {} & \beta_0^A + \beta_1^A size + \beta_2^A bedrooms + \beta_3^A bathrms + \beta_4^{DI} track + \beta_5^{DS} size{\cdot}track \\ & + \beta_6^{DS} bedrooms{\cdot}track + \beta_7^{DS} bathrms{\cdot}track + u^C \end{aligned} \tag{20}$$

In Equation (20) β_0^A is the coefficient on the intercept and $\beta_4^{DI} = \left(\hat{\beta}_0^B - \hat{\beta}_0^A\right)$ is the coefficient on the intercept dummy variable, *track*. If the property is on the A side of the track, $track = 0$, $\beta_4^{DI} track = 0$. This leaves the intercept as β_0^A in Equation (20), as in Equation (19a). If the property is on the B side, $track = 1$, the intercept in Equation (20) is now $\hat{\beta}_0^A + \left(\hat{\beta}_0^B - \hat{\beta}_0^A\right) = \hat{\beta}_0^B$ for B, as in Equation (19b). A similar derivation arises for the estimated coefficients on the slopes. For *size*, $size{\cdot}track = 0$ if the property is on the A side of the track and the estimated slope coefficient for *size* is $\hat{\beta}_1^A$. If the property is on the B side, $track = size$, the estimated coefficient on *size* for B is now $\hat{\beta}_1^A + \beta_5^{DS} = \hat{\beta}_1^A + \left(\hat{\beta}_1^B - \hat{\beta}_1^A\right) = \hat{\beta}_1^B$, and similarly for *bedrooms* and *bathrms*. The net result is that the estimated Equation (20) is akin to separately estimating Equations (19a) and (19b) – the proof is spelt out in Maddala and Lahari (2009, 314–319) and explained in Wooldridge (2013, 230–238).

The introduction of interaction effects in Equation (20) is a very effective way to relax the restriction of constancy of coefficients across regions and post/zip codes for the *intercept*, *size*, *bedrooms*, and *bathrms*. The only assumption made in using Equation (20) to model Equations (19a) and (19b) was that $u^c = u^A = u^B$. Were $u^A \neq u^B$ a heteroscedastic-consistent estimator could be used – see Subsection 4.4.

Slope and intercept dummies can be readily included in a regression model. Using R: say there are four geographical areas, for which there are dummy variables *South*, *East* and *West* (against a benchmark *North*) and a variable for *size* of property to interact with each.

To include intercept dummies to allow for different price levels and interaction slope dummies to model how the estimated coefficients for *size* will vary by area. In R, for model, mod1: (mod1 < -lm(logprice ~ size*(South, East, West), data = dat1)). The regression output will include estimated coefficients for *South*, *East*, *West*, with the benchmark intercept as constant, *North*, as well as for *size* and for interactions of *size* with each of *South*, *East*, *West*. Interaction effects can also be derived without having to create individual dummy variables. Say we have 80 geographical blocks under the variable name *BLOCK* and each block is called, *AA*, *AB*, and so on. Then you simply include *BLOCK* as a variable and R transforms it into 79 dummy variables with names *BLOCKAB, BLOCKAC* and so on. R chooses the first BLOCKAA as the default omitted one. The same could be done for the time dummy. Say, there were 35 observed transactions in A and 65 in B. Separate estimates of Equations (19a) and (19b) would be based on ($n - k - 1 =$) 31 and 61degrees of freedom for A and B respectively. Were A and B modelled in a single Equation (20), the estimation would benefit from ($100 - 7 - 1 =$) 92 degrees of freedom which in turn would result in more efficient coefficient estimates resulting from lower standard errors for $\hat{\beta}_k$. Indeed, formal testing of the joint effects of intercept and interaction effects using the Wald, Likelihood ratio or F-tests, based on some *a priori* knowledge of which cities, regions, post/zip codes hang together – that is whose differences in coefficient estimates are likely to be zero – may well reduce the number of terms. This would be especially useful where there is a fine classification of locations.

Locational effects can be included by means of targeted locational variables. A useful sub-set of such variables can be based on the use of spatial coordinates – latitude and longitude – to measure *proximity to city centers*, *train and bus stations*, *retail malls*, *green spaces*, *schools*, *noise intensity*, *tourist spots*, and so forth. Individual property records may have spatial coordinates or there may exist Geographic Information Systems (GIS) that allow such distances to be readily calculated (see Brand et al. 2017, Subsection 5.2 for practical examples.)

6.5. Broad Stratification and the Inclusion of Interaction Effects

The international guidelines on RPPI measurement (Eurostat et al. 2013) recommend stratification by broad groups and hedonic adjustments (which may include interaction effects) within strata. Strata are usually major locations, housing types, and vintages (old versus new properties). Within each stratum the hedonic model would have intercept and slope (interaction) dummies for sub-regions/cities, towns, post/zip codes as locational data permits. Some interaction effects will be more relevant than others for particular locations. The individual stratum PPIs would be ideally weighted together by the relative value of the total value of stock of property though relative transaction values would be a second-best.

The use of the relative value of stocks as opposed to transactions is appropriate for financial stability analysis and macroeconomic policy. This is because it is not just those who transact, but all owners of the stock of properties who can carry loans and have a wealth effect that in part determines aggregate investment and consumption. A hedonic rental price index can be employed for a consumer price index (CPI) for the compilation of quality-mix adjusted rental price indices and rental equivalent price indices for owner-occupied housing (OOH). Transaction weights would be appropriate for such indices.

Weights at the lower, individual property, level can be readily applied within the hedonic quality-mix correction as outlined in Silver (2018) and Subsection 3.4 above.

The weights should be updated as regularly as possible. Stock weights at the strata level for an RPPI can be based on population Census data that provide an estimate of the number of residential units by type and geographical areas. These numbers can be multiplied by (mid-period) estimates of the average price of such properties to give stock values. However, such estimates require updating, preferably annually. The perpetual inventory method makes use of estimates of additions to housing stock, retirement and efficiency (depreciation) losses to do so – OECD (2009 Part II, though see also Annex C for a simplified version). If stock weights do not exist, relative transaction values are readily available based on the sum of prices within a stratum.

6.5.1. Which Method? (4) Interaction Effects or (5) Stratification and Interaction Effects Within Strata

First, ease of use: interaction effects without weighted strata can be easily employed using R or any modern statistical/econometric software. While the running of separate regressions for different strata, say *West*, is also readily undertaken in R using $>$ subset(DF, $West = 1$), the selection of a subsample requires: (1) decisions on choice of strata, say location; (2) separately estimated hedonic models, model selection and PPIs specific to each location; and (3) information on weights. User attention will focus on the individual location results as well as the weighted national index requiring additional verification. There will be a separate set of diagnostics for each location. Yet with the data already set up and routines in R already written and having only to be run on subsections of data (strata), it is not extremely resource intensive.

Second, a primary advantage of stratification is the use of weights between stratum. A stratified PPI is a weighted average of locational PPIs where the weights are relative values of the stock of housing. A model with interaction effects such as Equation (20) implicitly weights by the relative number of observations and by the influence each observation has in determining the regression coefficients, Silver and Heravi (2005 Appendix) and (Silver 2016, Annex 2). Weights based on the relative number of transactions will also misrepresent relative value stock weights if, as is likely, there are differences between strata in the average prices of properties transacted against the stock – say downtown apartments more so than suburban houses in a city and (2) the turnover of properties. Weighting by location and type of property is a huge advantage.

Third, is user needs: Publication of a national PPI begs questions as to the extent to which property price inflationary trends differ for major types of property and regions/cities, to which stratification can provide answers and whose constituent parts sum to a meaningful total.

Finally and importantly, is the need for a sufficient sample size for each stratum. PPIs for strata that have a sufficient sample size and for which weights exist should be estimated. For the remaining areas there will be natural clusters that can be combined, say by similar municipality type, for example rural, or by region, or by administrative grouping that coincides with regional groupings used for other socio-economic geographical data. Again, sample size and the availability of weights permitting, stratification by type of property within a stratum should also be undertaken. Clustered strata would have a separate hedonic regression estimated but with meaningful intercept dummies and interaction terms for the (included) locations within the stratum. If such interaction terms are not statistically significant – Wald, Likelihood ratio, and F- tests – then the clustering of the included locations is fully justified. If there are differences, then these are properly modelled by the interaction terms and included in the regression. It may be that all locations where the sample size is insufficient are simply aggregated into a single stratum: “other locations.”

6.5.2. Interpreting the Estimated Coefficients of Interaction Effects

The partial effect (holding other variables constant) of *bedrooms* (which appears in the linear regression model in Equation (17) as a variable in its own right and as an interaction with *size*) on *price*, is:

$$\frac{\Delta price}{\Delta bedrooms} = \beta_1 + \beta_3 size \tag{21}$$

The partial effect of *bedrooms* would be evaluated at, say, the mean *size*. If $\beta_3 > 0$ then an additional bedroom yields a higher increase in price for larger houses.

Interaction terms and variables with higher powers can also be used to give the functional form greater flexibility, though due care is necessary in the interpretation of the partial coefficient of the effect of the explanatory variables and the need for joint testing of terms which contain the same variable, as is apparent from Equations (18) and (21).

7. Heteroscedasticity and the Normality of Residuals

7.1. On Heteroscedasticity and Robust Estimators

An OLS assumption is that the variance of the residuals should be constant: homoscedastic. A more detailed treatment of this subject is in Wooldridge (2013, chap. 8). The detection of heteroscedasticity may be via a Breusch-Pagan or White test; the null hypothesis is homoscedastic errors. A plot of standardised residuals against fitted values, using *plot(mod1)* in R, where regression model 1 is *mod* 1, will provide an insight into the nature of any heteroscedasticity and is one of several such R diagnostic plots. A plot may show, for example, whether it is higher-priced properties that have a greater variance in the residuals and require improved specification of the hedonic regression. Re-specifying the hedonic regression to better include curvature or introducing variables for up-market locations may mitigate against the heteroscedasticity. Similar considerations apply to omitted variables and interaction terms.

However, if this improved specification is not feasible, at least in the short-term, note that even when the residuals are heteroscedastic, estimated coefficients are unbiased and $\bar{R}^2$ unaffected. Standard errors, however, are incorrect and t and F statistics unreliable. The degree of the problem depends on how serious the heteroscedasticity is.

There is a further option. Bear in mind that the OLS estimator is no longer BLUE if the residuals are heteroscedastic, see Section 4. R contains a number of (robust) heteroscedastic-consistent estimators (HCE) that should be employed instead of OLS, especially for large sample sizes.The estimated regression coefficients will remain the same as those from an OLS estimator, but the presentation of the results from the hedonic regression output should include the OLS coefficient estimates and the standard errors, t-tests, and p-values from the HCE, with a footnote to this effect.

7.2. On Normality of the Residuals

Some researchers believe that regression requires normality of the individual variables. This is neither true for the left-hand-side (LHS) nor RHS explanatory variable. It is normality of *residuals* that is required for OLS in order that the $se(\hat{\beta}_k)$, the standard error of the sampling distributions of the individual $\hat{\beta}_k$ estimates, can be derived and are valid.

Thus, why house prices are unlikely to be normally distributed, say having a long positive tail of higher-priced larger homes in up-market areas, it is harder to say in theory whether the regression residual – after partialing out the effect of size and location – is normally distributed. In practice it is an empirical issue.

The $se(\hat{\beta}_k)$ forms the basis of the t-tests and F-test, and associated p-values. However, the residuals may not be normally distributed. Fortunately, for large samples the central limit theorem (CLT) applies.

The CLT is the (powerful) basis for the theory and practice of inferential statistics. For example, the confidence interval and tests of hypotheses for a single sample mean, $\bar{x}$, are based on the sampling distribution of the means, not on the distribution of the underlying variable, x. The CLT shows that the sampling distribution of the means will be normally distributed for large n – the rule of thumb is $n \geq 30$. irrespective of the shape of the distribution of the underlying variable, x. The mean of this sampling distribution is the population mean, μ, and the standard deviation – standard error – is $\sigma/\sqrt{n} \approx s/\sqrt{n}$, where σ and s are the population and sample standard deviation of the underlying variable, x. The test statistic $(x - \bar{x})/s/\sqrt{n}$ is akin to the test statistic for a coefficient estimate $(\hat{\beta}_k - \beta_k)/se(\hat{\beta}_k)$ which has an asymptotic standard normal distribution under the CLT. OLS estimators satisfy asymptotic normality, that is, they are approximately normal for large samples. We do not require MLR.6. This is good news. For large samples we can proceed with MLR.1–5 in Section 4. We still require homoscedastic residuals, or a correction thereof, but no longer the normality of residuals. For PPI purposes, even within strata, the sample sizes should be sufficiently large, see Wooldridge (2013, 110–113, 165–170).

The question is: how large should n be for the CLT to hold? The rule of $n \geq 30$ for estimating sample means, should not apply in this multivariate context. It is $df = n$-k-1 – see V.1 – that matters, not n, and for hedonic regressions k may be relatively large. There is no ready rule of thumb. Strata with small sample sizes should be combined with others and interaction terms used to model differences in strata-specific coefficient estimates, as outlined in Subsubsection 6.5.1.

Yet, the non-normality of residuals, especially when plotted, can give insights into which observations are responsible for, say, excess skewness. These same observations may be discovered by Cook's distance (see Section 8.2) and be added to the tool bag for detecting and rectifying specification problems. It is good practice to still publish test results of normality of residuals if at best, to explain that the specification is good, or at worst, to inform users via footnotes that it does not affect the estimates, which are unbiased, and tests for large samples. Tests of normality of residuals include the Jarque-Bera (JB) and Shapiro-

Wilk (SW) tests. There is a restriction in R to the application of the SW test of normality for large samples. The SW test tends to over-reject null hypotheses of normality for large samples – say $>$2,000. JB is an alternative test that is applicable for large samples. The null hypothesis is that the data (input: residuals from regression) are normally distributed. The JB test is an average of kurtosis and skewness and, when decomposed, can provide insights into why the residuals are non-normal by providing tests for (1) the null hypothesis that the kurtosis is 3; (2) the skewness is 0; and (3) the JB test combining both the kurtosis and the skewness to test for the normality of the input data. The JB test does have some problems but at least is appropriate for large samples.

8. Cleaning, Leverage, Influence, and Robust Estimators

8.1. Exclusion of Records: Cleaning

The univariate detection of extreme values of the price and explanatory variables is obligatory mainly to detect mis-recording of values. Particular attention should be given to particularly high values that may result from the use of, for example, prices wrongly entered with additional zeros or non-standard units of measurement, such as sq. ft. instead of sq. yards for *size*. Simple techniques such as boxplots, histograms, and/or stem-and-leaf diagrams on studentized residuals are useful for univariate outlier detection. The observations with say studentized residuals greater than 2 or 3 can be listed along with their characteristic values and compared with the means for the characteristic values. The aim is not to use automatic deletion but to identify and correct recording errors.

Equation (14) identified the standard error of an estimated coefficient, $se\left(\hat{\beta}_k\right)$, to be in part determined by the standard deviation of the variable, $sd(z_k)$. An undue restriction on particular values of explanatory variables, say a particular *size*, not only runs the risk of biasing the PPI, if larger properties have a different price trend than smaller ones, but of leading to imprecise estimates of the estimated coefficients, $\hat{\beta}_k$, by reducing $sd(z_k)$: the dispersion in z_k.

8.2. Multivariate Outliers

For multivariate outliers we cannot rely on univariate plots. As a second stage, the model should be estimated and influence, outlier and leverage techniques employed. It is *influence* that concerns us: the deletion of an influential observation has a marked effect on the estimated coefficients of a regression. *Outliers* are observations that have high residuals in a regression and observations with a high *leverage* are extreme values in the explanatory variables, $z_{k.}$ *Influential* observations are both *outliers* and have high l*everage*. Observations with high leverage need not by themselves affect the coefficient estimates since they may lie on the path of the estimated regression equation, albeit at extreme values. Leverage plots can be separately tabulated. Generally, we should carefully examine observations with leverage values greater than $(2k+2)/n$ - where k is the number of explanatory variables and n is the number of observations.

Cook's distance (Cook's D) is a summary measures of the combined impact of an observation on all of the estimated regression coefficients. It encapsulates the change in the model's coefficients when omitting an observation. An R routine can identify observations with high leverage and Cook's D, and list the values of their explanatory variables, $z_{i,k}$. These $z_{i,k}$ values can then be compared with the means from the rest of the sample of transactions in order to help us to identify the manner of, and reason for, the high Cook's D in a way that can be used to improve the specification of the hedonic regression, reduce the residuals of these observations, and, if successful, not have to delete the influential observations. The lowest value that Cook's D can take is zero, and the higher the Cook's D, the more influential the point. The conventional cut-off point for the absolute value of Cook's D is $> 4/(n - k - 1)$.

For example, a location variable (with a sufficient sample size) may include a sub-region on a hill that has large houses with wonderful views. *View* may be an omitted variable resulting in large residuals for these large houses in this location. The houses with large Cook's Ds, vill mainly be in specific parts of the country identified by the data. The specification of the hedonic can be improved to include dummy variables for up-market areas with interaction effects for size.

Also worthy of consideration is DFBETA_k, a more specific measures of influence that can be used to assess the extent to which each estimated coefficient, $\hat{\beta}_k$, changes by deleting an observation. The differences are scaled by dividing by the standard errors so comparison can be made. DFBETA_k can be positive (or negative) – if an observation has a positive (negative) effect, removing it decreases (increases) the estimated coefficient. A DFBETA_k value in excess of $2/sqrt(n)$ merits further investigation. DFBETA_k might be useful for the time dummy variable when using the hedonic regression time dummy approach. We obtain a direct measure of the influence of an observation on the time dummy PPI estimate. DFBETA_k is based on similar considerations as Cook's D: it is a measure of the effect of large residuals and leverage combined.

While DFBETA_k is a measure of the influence of an observation on a coefficient, DFFITS_i provides a measure of the scaled effect on predicted prices, $\hat{p}_i$, if an observation i is included and excluded from the analysis. It is useful if the hedonic imputation or characteristics/repricing approaches are used. One rule of thumb is that an observation is deemed influential if the absolute value of DFFITS_i is greater than $2\sqrt{(k+2)/(n-k-2)}$. DFFITS_i is also affected by both leverage and prediction error and is very similar to Cook's D.

8.3. Robust Estimators

A final, albeit unusual approach, is to use a robust estimator, instead of OLS, that is not only heteroscedastic consistent, but also robust to outliers. An OLS estimator minimizes the sum of squared (vertical) differences between the actual and fitted values. The squaring is a device to change negative residuals to positive ones so that the total residuals can be summed and minimized (using the calculus) to get the formula for the estimated coefficients. But this means that an observation with a high residual has an enormous effect (on squaring) on determining the coefficients. There are a range of alternative estimators. One of these is the robust estimator in *R*, *library (robustbase) – lmrob –* which is heteroscedastic consistent and robust to outliers. Not only are standard errors corrected for heteroscedasticity but there is a lesser penalty than OLS resulting from extreme residuals. Another such estimator that downplays outliers is the least absolute deviation (LAD) estimator. Instead of minimizing the sum of squared deviations it gets rid of the negative signs by summing absolute values of residuals. Its only concern is outliers and it deals (all too) harshly with them: it is a (conditional on the explanatory variables) median-based estimate. The LAD estimator is more suitable for a democratic PPI concerned with typical house prices and is not be suitable for financial stability analysis and macroeconomic policy.

9. Conclusions

In adopting a hedonic regression approach to PPI measurement, the starting point is an evaluation of data sources and selection of the (at least potentially) most reliable and comprehensive source. Preferably the source should be in electronic form, be timely, include transaction prices, and have a host of relevant price-determining characteristics. The source's potential development should include being merged with additional data bases – such as Census data – to have its variable set further enhanced. On-line property sales databases have much potential, especially when they can be merged with databases that include timely transaction records. Data from mortgage lending institutions may also have much potential.

Software such as R and Stata are well suited for data handling, the statistical/econometric estimation of hedonic regression equations, associated diagnostics covered in this article, and more. A prerequisite for a transparent and credible PPI is to experiment with different specifications of a hedonic regression and methodologies for PPIs. Use should be made of this article, Eurostat et al. (2013), Silver (2018), and other papers referenced here.

A regularly updated Working Paper distributed to the user community, any Advisory Boards, and presented at conferences, to receive feedback, provides the backbone for a reliable, credible, and transparent PPI. PPI measurement for the time dummy hedonic approach is embodied in the estimated hedonic regression equation and the reliability of the PPI directly emanates from that of the hedonic regression. Conventional diagnostics play an obvious role here. The imputation and characteristics/repricing approaches, while both based on predicted prices, also require the publication of details of the underlying hedonic regression with a full set of diagnostics. This is in order that users are not relying on a black box methodology and the effect of improvements, as time goes by, are transparent to users. A well-specified hedonic regression that complies with the methods and issues raised in this article in turn leads to more precise, unbiased, and consistent predicted values upon which these hedonic PPIs are based.

Economic statistics are rarely based on regression models and this area provides new challenges to price statisticians. There is a technically well-informed user community who will no doubt be interested in the methodology for this important economic indicator. Applied macroeconomists are only too aware of deficiencies in their model and are well versed in the issues outlined in this article. The detailed nature of econometrics arose in order to identify and, where possible surmount, such problems. Even with data deficiencies, such as omitted variables, we can still proceed with measurement with a good technical understanding of the effects of such omission, and other deficiencies, aligned with a strategy of longer-term data and methodological enhancement. We proceed with eyes wide open.

10. Appendix: Hedonic Imputation Indexes with Different Periods for the Hedonic Estimation and Index Reference = 100.0

Let b denote the period for which the hedonic regression is estimated, 0 the PPI reference period = 100.00, and t the current periods. Equation (1) is a PPI comparing period 0 with period t *divided by* a similar PPI comparing period 0 with period t-1 yielding, in Equation (2) a period t on t-1 PPI comparison. Each square bracket contains a ratio of geometric mean prices divided by – adjusted for – the change characteristic-mix, between z_i^t and z_i^0 each predicted using the period b hedonic regression. $\hat{p}^b_{i|z_i^t}$, $(\hat{p}^b_{i|z_i^0})$, and $(\hat{p}^b_{i|z_i^{t-1}})$ are the predicted prices from the period b hedonic regression of property i based on its period, $t(0)$, and $(t-1)$ characteristics.

$$PPI_l^{t-1\to t}=\left[\frac{\left(\prod_{i\in N^t}p_i^t\right)^{\frac{1}{N^t}}}{\left(\prod_{i\in N^0}p_i^0\right)^{\frac{1}{N^0}}}\div\frac{\left(\prod_{i\in N^t}\hat{p}^b_{i|z_i^t}\right)^{\frac{1}{N^t}}}{\left(\prod_{i\in N^0}\hat{p}^b_{i|z_i^0}\right)^{\frac{1}{N^0}}}\right]\div\left[\frac{\left(\prod_{i\in N^{t-1}}p_i^{t-1}\right)^{\frac{1}{N^{t-1}}}}{\left(\prod_{i\in N^0}p_i^0\right)^{\frac{1}{N^0}}}\div\frac{\left(\prod_{i\in N^{t-1}}\hat{p}^b_{i|z_i^{t-1}}\right)^{\frac{1}{N^{t-1}}}}{\left(\prod_{i\in N^0}\hat{p}^b_{i|z_i^0}\right)^{\frac{1}{N^0}}}\right] \tag{1}$$

$$\left[\frac{\left(\prod_{i\in N^t}p_i^t\right)^{\frac{1}{N^t}}}{\left(\prod_{i\in N^{t-1}}p_i^{t-1}\right)^{\frac{1}{N^{t-1}}}}\div\frac{\left(\prod_{i\in N^t}\hat{p}^b_{i|z_i^t}\right)^{\frac{1}{N^t}}}{\left(\prod_{i\in N^{t-1}}\hat{p}^b_{i|z_i^{t-1}}\right)^{\frac{1}{N^{t-1}}}}\right]. \tag{2}$$

These are quite meaningful expressions. The finding that Equation (2) follows Equation (1) follows the well-known result that a ratio of two Lowe indexes is a Lowe index:

While Equation (2) is a meaningful quality-mix adjusted PPI it does not have a clear equivalence to Equation (4). This would require period b = period $0 = 100$:

$$RPPI_l=\frac{\left(\prod_{i\in N^t}p_i^t\right)^{\frac{1}{N^t}}}{\left(\prod_{i\in N^b}p_i^b\right)^{\frac{1}{N^b}}}\div\frac{\left(\prod_{i\in N^t}\hat{p}^b_{i|z_i^t}\right)^{\frac{1}{N^t}}}{\left(\prod_{i\in N^b}\hat{p}^b_{i|z_i^t}\right)^{\frac{1}{N^b}}}=\frac{\left(\prod_{i\in N^t}\hat{p}^b_{i|z_i^t}\right)^{\frac{1}{N^t}}}{\left(\prod_{i\in N^b}\hat{p}^b_{i|z_i^t}\right)^{\frac{1}{N^b}}} \tag{3}$$

Note $\left(\prod_{i\in N^b}p_i^b\right)^{\frac{1}{N^b}}$ in the denominator of the first term is simply the geomean of prices in period b. A fortunate feature of OLS is that the mean (in this case of the logs) of the actual prices is equal to the mean of the fitted values, $\left(\prod_{i\in N^b}p_i^b\right)^{\frac{1}{N^b}}=\left(\prod_{i\in N^b}\hat{p}^b_{i|z_i^b}\right)^{\frac{1}{N^b}}$. So, on division, they cancel leaving us with the right-hand-side of Equation (3). A similar simpler formulation to Equation (1) drops out when comparing quarter t with t-1.

11. References

Balk, B.M. 2008. *Price and Quantity Index Numbers*, Cambridge: Cambridge University Press.

Bokhari, S., and D. Geltner. 2012. “Estimating real estate price movement for high-frequency tradable indices in a scarce data environment”. *Journal of Real Estate Finance and Economics*, 45 (2): 522–543. DOI: http://dx.doi.org/10.1007/s11146-010- 9261-4.

Brand, M., C. Becker Vermeulen, D. Fischbach, and Y. Carpy. 2017, "Swiss residential property price index: the use of geolocalised information for quality adjustment in location." Paper presented at the 15th Meeting of the Ottawa Group International Working Group on Price Indices, May 10–12, Eltville, Germany. Available at: https://www.ottawagroup.org/Ottawa/ottawagroup.nsf/4a256353001af3ed4b2562b-b00121564/1ab31c25da944ff5ca25822c00757f87/$FILE/Swiss%20residential%20-property%20price%20index%20-%20Becker%20Vermeulen%20paper.pdf (accessed January 2022).

Curry, B., P. Morgan, and M. Silver. 2001. "Hedonic regressions: mis-specification and neural networks". *Applied Economics*, 33(5): 659−671. DOI: http://dx.doi.org/10.1080/000368401750106261.

De Haan, J. 2010. "Hedonic price indices: a comparison of imputation, time dummy and re-pricing methods" *Jahrbücher für Nationalökonomie und Statistik*, 230(6): 772–791. DOI: http://dx.doi.org/10.1515/9783110511123-010.

De Hann, J., and W.E. Diewert. 2013. "Hedonic regression methods." In *Eurostat et al.* (2013) *op. cit*., chap. 5. DOI: http://dx.doi.org/10-2785/34007.

Diewert, W.E. 1976. "Exact and superlative index numbers". *Journal of Econometrics*, 4(2): 115–145. DOI: http://dx.doi.org/10.1016/0304-4076(76)90009-9.

Diewert, W.E. 1978. "Superlative index numbers and consistency in aggregation", *Econometrica*, 46: 883–900. DOI: http://dx.doi.org/10.2307/1909755.

Diewert, W.E. 1995. Axiomatic and economic approaches to elementary price indexes, *NBER Working Paper Series* 5104, National Bureau of Economic Research (NBER), Cambridge MA, May. Available at: https://www.nber.org/system/files/working_papers/w5104/w5104.pdf (accessed January 2022).

Diewert, W.E., S. Heravi, and M. Silver. 2009. "Hedonic imputation indices versus time dummy hedonic indices." In *Price Index Concepts and Measurement*, *Studies in Income and Wealth*, 70, edited by W.E. Diewert, J. Greenlees, and C.R. Hulten. National Bureau of Economic Research, Chicago: University of Chicago Press: 278–337.

Diewert, W.E., K. Nishimura, C. Shimizu, and T. Watanabe. 2020. *Property Price Index: Theory and Practice*, Japan: Springer Japan.

European Commission (EC), Eurostat. EC, 2017. *Technical Manual on Owner-Occupied Housing and House Price Indices*, Luxembourg: Eurostat. Available at: https://ec.europa.eu/eurostat/documents/7590317/0/Technical-Manual-OOH-HPI-2017/ (accessed January 2022).

Eurostat, European Union, International Labor Organization, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations Economic Commission for Europe, and The World Bank. 2013. *Handbook on Residential Property Prices Indices* (*RPPIs*), Luxembourg: European Union. Available at: https://ec.europa.eu/eurostat/documents/3859598/5925925/KS-RA-12-022-EN.PDF.

Eurostat 2017. *Commercial Property Price Indicators: Sources, Methods and Issues*, Luxembourg: Eurostat. Available at: https://ec.europa.eu/eurostat/documents/7870049/8545612/KS-FT-16-001-EN-N.pdf/9e4bbc9b-8c6f-44a9-b686-1083a7a8fa0f.

Fenwick, D. 2013. "Uses of residential property price indices." In *Eurostat et al. 2013 op. cit*., chap. 2. DOI: http://dx.doi.org/10-2785/34007.

Fuerst, F., and C. Shimizu. 2016. "The rise of eco-labels in the Japanese housing market", *Journal of Japanese and International Economy*, 39: 108–122. DOI: http://dx.doi.org/10.1016/j.jjie.2016.01.003.

Geltner, D. 1993. "Temporal aggregation in real estate return indices", *Journal of the American Real Estate and Urban Economics Association* 21(2): 141–166. DOI: http://dx.doi.org/10.1111/1540-6229.00605.

Giles, D.E. 2011. *Interpreting dummy variables in log-linear regression models: exact distributional results*, University of Victoria, Department of Economics, *Econometrics Working Paper* EWP 1101, January. Available at: https://www.researchgate.net/publication/228954943_Interpreting_Dummy_Variables_in_Semi-Logarithmic_Regression_Models_Exact_Distributional_Results (accessed January 2022).

Greenstone, M. 2017. "The continuing impact of Sherwin Rosen's "Hedonic prices and implicit markets: product differentiation in pure competition," *Journal of Political Economy*, 125(6): 1891–1902. Available at: https://static1.squarespace.com/static/585017d2bebafbe412f1a838/t/6033d2f812ec930bb253a046/1614009365638/The-Continuing-Impact.pdf.

Heiss, F. 2020. *Using R for introductory econometrics*, 2nd Edition, Düsseldorf, Germany: Florian Heiss, URfie.net.

Heravi, S., and M. Silver. 2007. "Different approaches to estimating hedonic indexes", In *Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches*, Studies in Income and Wealth, edited by E. Berndt and C. Hulten, 67: 235–268, National Bureau of Economic Research, Chicago: University of Chicago Press.

Hill, R.E. 2013. "Hedonic price indices for residential housing: a survey, evaluation and taxonomy." *Journal of Economic Surveys*, 27(5). DOI: http://dx.doi.org/10.1111/j.1467-6419.2012.00731.x.

Hill, R.J., and D. Melser. 2008. "Hedonic imputation and the price index problem: An application to housing", *Economic Inquiry*, 46: 593–609. DOI: http://dx.doi.org/10.1111/j.1465-7295.2007.00110.x.

Hill, R.J., and M. Scholtz. 2017. "Can geospatial data improve house price indices? a hedonic imputation approach with splines", *Review of Income and Wealth*, 64(4): 737–756. DOI: http://dx.doi.org/10.1111/roiw.12303.

Hill, R.J., A.N. Rambaldi, and M. Scholtz. 2020 "Higher frequency hedonic property price indices: a state-space approach", *Empirical Economics*. DOI: https://doi.org/10.1007/s00181-020-01862-y.

Hill, R.J., and M. Steurer. 2020. "Commercial property price indices and indicators: review and discussion of issues raised in the CPPI Statistical Report of EUROSTAT (2017)", *Review of Income and Wealth*. DOI: http://dx.doi.org/10.1111/roiw.12473.

Hill, R.J., M. Scholtz, C. Shimizu, and M. Steurer. 2018. "An evaluation of the methods used by European countries to compute their official house price indices." *Economie et Statistique*, (500, 501, 502). DOI: http://dx.doi.org/10.15396/eres2018_201.

Iacobucci, D., M.J. Schneider, D.L. Popovich, and G.A. Bakamitsos. 2016. "Mean centering helps alleviate "micro" but not "macro" multicollinearity", *Behavioral Research*, 48: 1308–1317. DOI: 10.3758/s13428-015-0624-x.

IMF, International Monetary Fund. 2020. *RPPI Practical Compilation Guide*, Washington DC: IMF. Available at: https://www.imf.org/en/Data/Statistics/RPPI-guide (accessed July 2020).

Kennedy, P. 1981. "Estimation with correctly interpreted dummy variables in log-linear equations", *American Economic Review* 71(4): 801. Available at: www.jstor.org/stable/1806207 (accessed January 2022).

Kennedy, P.E. 2008. *A Guide to Econometrics*, 6th Edition, Oxford: Wiley-Blackwell.

Leventis, A. 2008. *Revisiting the differences between the OFHEO and S&P/Case-Shiller house price indexes: new explanations*, Office of Federal Housing Enterprise Oversight (OFHEO), Available at: https://fhfa.gov/PolicyProgramsResearch/Research/PaperDocuments/20080115_RP_RevisitingDifferencesOFHEOSPCaseShillerHPI_N508.pdf (accessed 28 January, 2022).

Maddala, G.S., and K. Lahari. 2009. *Introduction to Econometrics*, 4th Edition, Chichester, England: John Wiley and Sons.

Mehrhoff, J., and E. Triebskorn. 2016. How should we measure residential property prices to inform policy makers? Narodowy Bank Polski Workshop: Recent Trends in the Real Estate Market and its Analysis, August 9, Poland. Available at: https://papers.ssrn.com/sol3/papers.cfm? abstract_id = 2842531 (accessed January 2022).

OECD. 2009. *Measuring Capital – OECD Manual 2009*, Second edition, Paris: OECD Publishing. Available at: https://www.oecd-ilibrary.org/docserver/9789264068476-en.pdf?expires = 1610123185&id = id&accname = guest&checksum = 2335110A437CEF603F6E28E38485ECCB (accessed January 2022).

O'Hanlon, N. 2011. "Constructing a national house price index for Ireland," *Journal of the Statistical and Social Inquiry Society of Ireland*, 40: 167–196. Available at: http://www.tara.tcd.ie/bitstream/handle/2262/62349/o%27hanlon%20pdf.pdf?sequence = 1&isAllowed = y (accessed January 2022).

Oladunni, T., and S. Sharma. 2018. "H2O deep learning for hedonic pricing", *International Journal of Computers and their Applications*, 25(1): 1–9. Available at: https://www.researchgate.net/publication/342914955_H2O_Deep_Learning_for_Hedonic_Pricing (accessed January 2022).

Rambaldi, A.N., and C.S. Fletcher. 2014. "Hedonic imputed price indices: the effects of econometric modeling choices", *The Review of Income and Wealth* 60: s423–s448. DOI: http://dx.doi.org/10.1111/roiw.12143.

Rosen, S. 1974. "Hedonic prices and implicit markets: product differentiation in pure competition." *Journal of Political Economy*, 82(1): 34–55. Available at: https://www.jstor.org/stable/1830899?seq = 1 (accessed January 2022).

Shimizu, C., K.G. Nishimura, and T. Watanabe. 2010. "House prices in Tokyo – a comparison of repeat-sales and hedonic measures", *Journal of Economics and Statistics*, 230(6): 792–813. DOI: http://dx.doi.org/10.1515/9783110511123-011.

Silver, M. 2002. "The use of weights in hedonic regressions: the measurement of quality-adjusted price changes." researchgate.net. Available at: https://www.researchgate.net/publication/238355850_The_Use_of_Weights_in_Hedonic_Regressions_the_Measurement_of_Quality-Adjusted_Price_Changes (accessed January 2022).

Silver, M. 2015. "The degree and impact of differences in house price index measurement", *Journal of Economic and Social Measurement*, 39: 305–328. DOI: http://dx.doi.org/10.3233/JEM-150406.

Silver, M. 2016. *How to better measure hedonic residential property price indices*, IMF Working Paper WP/16/213, Washington D.C. Available at: https://www.researchgate.net/publication/309786163_How_to_Better_Measure_Hedonic_Residential_Property_Price_Indexes (accessed January 2022).

Silver, M. 2018. "How to measure hedonic residential property price indices better", *Eurostat Review on National Accounts and Macroeconomic Statistics Indicators* (*EURONA*), volume 1: 35–66. Available at: https://www.researchgate.net/publication/326186402_How_to_Measure_Residential_Property_Price_Indices_Better#fullTextFileContent (accessed January 2022).

Silver, M. 2019. "Data Sources for CPPIs: An overview and strategy", Paper presented at the European Commission International Conference on Real Estate Statistics, 20–22 February, Luxembourg. Available at: https://www.researchgate.net/publication/342601082_Data_Sources_for_CPPIs_An_Overview_and_Strategy (accessed January 2022).

Silver, M., and B. Graf. 2014. *Commercial property price indexes: problem of spare data, spatial, spillovers, and weighting*, IMF Working Paper. WP/14/73. Washington D.C. DOI: http://dx.doi.org/10.2139/ssrn.2445456.

Silver, M., and S. Heravi. 2001. "Scanner data and the measurement of inflation", *The Economic Journal*, 111, 472: 383–404. DOI: http://dx.doi.org/10.1111/1468-0297.00636.

Silver, M., and S. Heravi. 2005. "A failure in the measurement of inflation: results from a hedonic and matched experiment using scanner data", *Journal of Business and Economic Statistics*, 23(5): 269–281. DOI: http://dx.doi.org/10.1198/07350010400000 0343.

Silver, M., and S. Heravi. 2007. "Why elementary price index number formulas differ: Evidence on price dispersion." *Journal of Econometrics*, 140(2): 874–883. DOI: http://dx.doi.org/10.1016/j.jeconom.2006.07.017.

Solon, G., S.J. Haider, and J.M. Wooldridge. 2015. "What are we weighting for?" *Human Resources*, Spring, 50, 2, 301–316. See also *National Bureau of Economic Research (NBER) Working Paper* 18859, February 2013. Available at: https://www.nber.org/system/files/working_papers/w18859/w18859.pdf (accessed January 2022).

Triplett, J.E. 2006. *Handbook on Hedonic Indices and Quality Adjustments in Price Indices: Special Application to Information Technology Products.* Paris: OECD Publishing.

Van Garderen, K.J., and C. Shah. 2002. "Exact interpretation of dummy variables in semilogarithmic equations", *Econometrics Journal* 5: 149–159. DOI: https://doi.org/10.1111/1368-423X.00078.

Wooldridge, Jeffrey M. 2013. *Introductory Econometrics: A Modern Approach*, 5th Edition, Mason OH: South Western, Cengage Learning.

Received July 2020
Revised January 2021
Accepted April 2021

Journal of Official Statistics, Vol. 38, No. 1, 2020, pp. 187–211, http://dx.doi.org/10.2478/JOS-2022-0009

Rentals for Housing: A Property Fixed-Effects Estimator of Inflation from Administrative Data

Alan Bentley[1]

Official rentals for housing (rent) price inflation statistics are of considerable public interest. Matched-sample estimators, such as that used for nearly two-decades in New Zealand (2000–2019), require an unrealistic assumption of a static universe of rental properties. This article investigates (1) a property fixed-effects estimator that better reflects the dynamic universe of rental properties by implicitly imputing for price change associated with new and disappearing rental properties; (2) length-alignment simulations and property life-cycle metrics to inform the choice of data window length (eight years) and preferred splice methodology (mean-splice); and (3) stock-imputation to convert administrative data from a 'flow' (new tenancy price) to 'stock' (currently paid rent) concept. The derived window-length sensitivity findings have important implications for inflation measurement. It was found that the longer the data window used to fit the model, the greater the estimated rate of inflation. Using administrative data, a range of estimates from 55% (window length: three-quarters) to 127% (window of 90-quarters) were found for total inflation, over the 25-years to 2017 Q4.

Key words: Price indexes; multilateral index number methods; quality change; fixed-effects mean-splice; rolling window indexes.

1. Introduction

Rent (actual rentals for housing) is one of the most important components of consumer price indices. In New Zealand, rent is about 10% of the Consumers Price Index (CPI) by expenditure weight. For households who pay rent, the proportion of their expenditure on rent is typically 30–40%. In many other countries, the contributions of rent price indexes in CPIs are amplified through their use as a proxy for owner-occupied housing costs (a 'rental equivalent' approach), as well as representing actual rent prices. Beyond their significance in aggregate price indexes, rent statistics are of considerable public policy interest, and to the public at large.

Price indexes aim to measure 'pure' price change over time by decomposing value change into price and volume (quality and quantity) components. Traditionally, this has been achieved using bilateral index numbers for products (goods and services) available in

[1] Statistics New Zealand, 8 Gilmer Terrace, PO Box 2922, Wellington 6140 New Zealand. Email: *alan.bentley@stats.govt.nz*

Acknowledgments: A large proportion of this research was undertaken whilst the author was on secondment at the New Zealand Ministry of Business, Innovation and Employment. The author would like to thank Frances Krsinich for her guidance and thought-provoking discussions, which has enriched this research. Confidence in the results has been enhanced by Donal Lynch and his independent coding of the models for ongoing production. The opinions, findings, recommendations, and conclusions expressed in this paper are those of the author. They do not represent those of Stats NZ, nor the Ministry of Business, Innovation and Employment.

© Statistics Sweden

both time periods; that is, for a *matched-sample* of products (Bentley and Krsinich 2017). Almost universally National Statistical Institutes (NSIs) have based their CPI on the Lowe (1823) price index. The validity of using such measures to be representative of all products (including new and disappearing ones) depends on the simplifying assumption that the complete universe of products is static over time. This is a serious limitation for rental properties, since the reality is a dynamic market where properties: are built and demolished; switch between owner-occupation and landlord-ownership; are temporarily unavailable during times of refurbishment. Conflated with these availability dynamics is that newly let properties are a common opportunity to increase rents (ILO et al. 2004).

Researchers and practitioners have used a variety of approaches to address *the quality change problem* when products are unmatched between periods (ILO et al. 2004; Eurostat et al. 2013; Eurostat 2018; for New Zealand examples see Krsinich 2014). The two leading methods for rent prices, are: *hedonic regression* (which dates back to Court 1939, for automobiles) and *repeat rent regression* (originating as a repeat sales regression by Bailey et al. 1963). The methods are analogous to those used for residential property prices ('house prices') since in both instances there is a need to control for temporal variation in the quality of properties. Example applications of hedonic regression for rent price indexes can be found in Germany (Behrmann and Goldhammer 2017), Canada (Keshishbanoosy and Taylor 2019) and Ireland (Private Residential Tenancies Board 2013). Eurostat (2018) consider hedonic models a grade-A approach for estimating rent price indexes. *Repeat rent regressions* have been applied in Japan (Shimizu et al. 2009) and the United States (Ambrose et al. 2015); albeit, less notably than their cousin repeat sales regressions as has been applied to house price indexes (such as those following Case and Shiller 1987).

Hedonic regression builds on revealed preferences theory to estimate the shadow prices of property characteristics to quality-adjust price indexes for unmatched properties over time. Such models can be formulated either as a bilateral or multilateral price index. An important practical consideration is that hedonic regression requires high-quality data on property characteristics that explain most of the cross-sectional variation in rental prices; inadequate property characteristics risks *omitted variable bias*. A major concern with the repeat rent method is that it may suffer from *sample selection bias* if particular types of tenancies tend to turn over more frequently than others.

In this article we look to address the quality adjustment problem by using a multilateral, *fixed-effects regression*, estimator that simultaneously considers observations from multiple time periods. This approach can be thought of as a *hedonic regression*, where unique property identifiers control for the effect of each property's 'bundle of characteristics' (Krsinich 2016). It is also a *generalisation of repeat rent regression*, broadened to encompass multiple rent changes per property. Adopting this approach makes greater use of the available data and mitigates a criticism of repeat rent regression of the potential for sample selection bias. Applying this model using rolling windows of data (explained in Section 4) addresses another criticism of the repeat rent regression approach that property characteristics remain unchanged over time.

Multilateral index number theory, originally developed for cross-country price comparisons (Gini, 1931; Eltetö and Köves 1964; Szulc 1964; Geary 1958; Khamis 1972; Summers 1973; Diewert 2004), has been applied in a temporal context by a number of authors (e.g., Ivancic et al. 2009; De Haan and Krsinich 2014; Krsinich 2009). The primary motivation to use multilateral methods for high frequency retail sales data is often to overcome chain drift bias. This can arise from oscillating prices and quantities when chaining bilateral price indexes (even superlative ones) to create longer time series (Diewert and Fox 2017).

For a rent price index, our major concern is that a matched-sample approach might bias the index (downwards) as implicit price change associated with newly rented properties are excluded. In the absence of explicit quality characteristics, Krsinich (2016) has shown that a fixed-effects index implicitly imputes price change for new and disappearing rental properties, for properties with two or more observations. This is an important observation that underpins the validity to use a property fixed-effects regression to estimate rent price inflation. Aizcorbe et al. (2003) describe this imputation, for new products (or properties, in a rent price index), as the difference between the quality-adjusted price for the new property and the average quality-adjusted price for all properties in the prior period.

The remainder of this article is structured as follows. Section 2 describes the New Zealand data sources used; a longitudinal (panel) survey of landlords, and administrative data relating to new tenancies. Section 3 explains the traditional bilateral matched-sample approach historically used to estimate a quality-adjusted rent price index, and introduces our new property fixed-effects regression estimator. Section 4 refines the model by exploring sensitivity to data window length and index-chain alignment options. Section 5 presents the empirical results; firstly, on a *flow* concept, representative of new tenancies, and secondly, by simulating a stock concept by imputing currently paid rent for all properties. Section 6 presents conclusions with the outcome being a new methodology for estimating rent price inflation.

2. Data

Many NSIs commission special purpose surveys to collect data on rent prices (e.g., an online survey is used in Germany, see Behrmann and Goldhammer 2017), or ask about rental costs in multipurpose surveys (e.g., Statistics Canada ask rental costs as part of their Labour Force Survey, see Keshishbanoosy and Taylor 2019). These can be surveys of either landlords or tenants (Eurostat 2018). In many jurisdictions administrative data is available either as part of government administration (e.g., a rent price index for Ireland uses administrative tenancy data, see Private Residential Tenancies Board 2013), or from the private sector sourced from advertisements (e.g., Shimizu, et al. 2009 use rental listings data in Japan). Careful assessment of administrative data is required to assess the suitability of coverage to the target population of interest.

For more than two decades New Zealand has had two official sources of rent price statistics; a survey of landlords run by Statistics New Zealand (Stats NZ), and administrative rent data held by the Ministry of Business, Innovation and Employment (MBIE); formerly held by the then Department of Building and Housing. The survey data covers only private-sector (market) rentals. The Consumers Price Index (CPI) class ‘actual rentals for housing’ includes private and government-owned rentals, and educational accommodation. Private-sector rentals are about 95% (by expenditure weight) of the quality-adjusted price index, published quarterly. Regional series are published for five broad-regions of New Zealand. The administrative data – generated as a byproduct of regulations to lodge tenancy bonds (deposits) with MBIE's Tenancy Services – enables calculation of summary statistics, such as median and geometric mean rental amounts, which can be disaggregated to much finer regional breakdowns and split to finer temporal frequency (monthly, rather than quarterly). The published average rental amounts from administrative data cover only private-sector rentals.

2.1. Administrative Data: Tenancy Bonds, MBIE

Landlords in New Zealand can ask tenants to pay a monetary bond as security when they move into a property. Landlords who charge a bond must lodge it with MBIE's Tenancy Services within 23 working days. The Bond lodgement form (which can be completed online or by post) includes a requirement to state the weekly rent payment. Other data captured includes the dwelling address, dwelling type (such as room, flat, house), and the number of bedrooms. A unique property identifier is created as part of the administrative process.

The data set used for this analysis covers bonds lodged 1 January 1993 – 31 December 2017; 100 quarters; 25 years. It contains 4.1 million price observations, for one million unique properties. There are a mean number of 7.5 (median of 6) price observations per property. Stats NZ (2015) explains the data set further.

Bentley and Krsinich (2017) assessed the coverage of the Tenancy Bond data and concluded that the data appears reasonable compared with the New Zealand Census of Population and Dwellings. Miller et al. (2018) found very similar distributions for weekly rent amount, number of bedrooms, and sector of landlord, using New Zealand Census compared with Tenancy Bond data. They conclude "... we see good consistency between the tenancy bond variables and the census... The concepts used in the tenancy bonds are consistent with the statistical standard used by the census for each of the housing variables. Levels of missing data for tenancy bond variables are low, and comparable with the census levels of missing data."

2.2. Survey Data: Quarterly Survey of Landlords, Stats NZ

Until 2019-Q1 data collection for rent prices included in the New Zealand CPI was from a longitudinal postal survey of landlords, run continuously since 1998-Q3. The sample size was about 1,200 landlords; 2,400 properties. The survey population was all identified non-furnished rental properties within the sampled geospatial areas (meshblocks). The survey was designed based on the 1996 Census of Population and Dwellings. A scoping questionnaire was sent to all properties in the sampled areas to identify in-scope rental properties, which were surveyed every quarter for the 20-year survey duration. The survey population was maintained using the administrative Tenancy Bond data (note that in this study we investigate using the Tenancy Bond data as a data source in its own right). Within the sampled meshblocks new rental properties were identified and the landlords 'birthed' into the survey. There was a lacuna in the birthing process for five-years between 2001–2006, which resulted in a steadily declining sample size until the process was reinstated in 2006-Q2 (Stats NZ 2008), see Appendix (Section 7).

Nonresponse in a given quarter was imputed for by carrying forward the last known rental value. For persistent non-response, the value was carried forward for five quarters before the nonrespondent was assumed to be a 'death' and removed from the sample (Krsinich 2009). The five-quarter imputation improved longitudinal property matching, and therefore increased the chance of reflecting price changes when tenancies turned over, even if there was a short gap between tenancies (a risk of the approach was erroneously imputing no price change). The data set used for this analysis covered the period 2000-Q2 – 2017-Q4; 69 quarters, survey data for two quarters (2001-Q2 and 2012-Q2) were missing. In total there were 143,000 price observations, for 7,650 unique properties.

3. Developing an Improved Model

3.1. Old Methodology: Bilateral Matched-Sample

For nearly two-decades (2000-Q1 – 2019-Q1) a matched-sample approach was used to control for the changing quality of the stock of rental properties. The price index movement was calculated from properties that existed in both the previous and current quarters; properties that were new to the sample were not included in the initial quarter that they were birthed. The property-level matched-sample average price change was calculated for region by dwelling-size strata; broad region (Auckland, Wellington, Rest of North Island, Canterbury, Rest of South Island) and by number of bedrooms (1, 2, 3, 4+). A national price index was obtained by aggregation, using expenditure weights (a Laspeyres-type Lowe (1823) price index, updated three-yearly, consistent with the rest of the NZ CPI).

In 2008 concerns were raised that the matched-sample approach might bias the index (downwards) since the approach excludes implicit price change (generally an increase) associated with newly rented properties (as these are unmatched in a bilateral index). A regression fixed-effects model (of the type proposed in this article) was used by Krsinich (2009) to investigate the likely magnitude of the bias. Using survey data for 2000–2008, she found only a small bias in the matched-sample approach compared with a fixed-effects model so concluded that "the current estimation method does well at controlling for compositional change" and "the restriction of the sample [to a bilateral matched-sample] is not biasing the price measurement to a level of any practical significance". Additional data is now available to investigate the use of longer rolling windows. Importantly, since 2006-Q2 the survey data is unblemished by the reduced longitudinal match rates caused by the mothballed birthing process.

In the Appendix we demonstrate how the lack of survey births (described in Subsection 2.2) needed to ensure the sample remained representative of the dynamic rental market, led Krsinich (2009) to conclude that the current estimator didn't appear to materially bias the index compared with a fixed-effects benchmark. The benchmark was also biased due to the mothballed survey update process.

3.2. Property Fixed-Effects Regression

There is growing use of multilateral methods to estimate inflation (Bentley and Krsinich 2017; Diewert and Fox 2017; Australian Bureau of Statistics 2016). However, there is not yet international agreement on which multilateral methods are best suited to estimating inflation from different types of data. A number of researchers have applied a hedonic regression model to construct rental price indexes (eg. Shimizu et al. 2009; Behrmann and Goldhammer 2017; Keshishbanoosy and Taylor 2019).

In New Zealand, private-sector rental prices are regulated by the Rential Tenancies Act Residential Tenancies Act (1986) so that the rent payable can not exceed market rent by a "substantial amount", and rent increases are limited to once every six months (once every 12 months from August 2020, following regulatory amendment). This 'light-handed' price regulation means the market prices can be assumed to be a sound representation of people's willingness to pay so are well suited to hedonic quality- adjustment. However, since the New Zealand data does not contain many price determining quality characteristics of properties (only location, number of bedrooms, and property type) we have instead investigated a fixed-effects regression which uses a unique property identifier to control for the effect of each property's 'bundle of characteristics'.

The Time Dummy Hedonic approach to constructing quality-adjusted price indexes is well-known and discussed in the CPI Manual (ILO et al. 2004, 382). In the absence of explicit quality characteristics fixed-effects regression (1) has been suggested for longitudinal data of prices. This approach is known as the Time Product Dummy (TPD) method, named after the Country Product Dummy (CPD) model proposed by Summers (1973), or the *fixed-effects index*, using the naming conventions of Krsinich (2016).

The estimating equation is:

$$\ln p_i^t = \alpha + \sum_{t=1}^{T} \delta^t D_i^t + \sum_{i=1}^{N-1} \gamma_i D_i + \varepsilon_i^t \tag{1}$$

where: p_i^t is the price p of property i at time t; $D_i^t = 1$ if a price for property i is observed at time t and $= 0$ otherwise; $\mathrm{D}_i = 1$ if the observation relates to property i and $= 0$ otherwise; α, δ^t and γ_i are regression estimates and ε_i^t is an error term; dummies for item N and period 0 are excluded to identify the model.

The index is derived from the estimated parameters on time; price change between period 0 and period t can be expressed as:

$$P_{fixed-effects}^{0,t} = \exp(\hat{\delta}^t) \tag{2}$$

Krsinich (2016) and Aizcorbe et al. (2003) suggest that the fixed-effects index is preferable to a Time Dummy Hedonic approach, even if detailed characteristics are available. Krsinich (2016) showed that the fixed-effects method is the same as a Time Dummy Hedonic if all time invariant quality determining characteristics (and the interaction of these) are expressed as categorical and included in the regression. Aizcorbe et al. (2003) state the advantages of the fixed-effects approach as:

- it does not impose a particular functional form,
- it does not place any restrictions on the relationship between products and characteristics (as full interactions are implicitly included),
- there is no need to choose characteristics,
- fixed-effects can provide more stable parameter estimates.

Krsinich (2009) remarks that a further advantage is that the regression controls for both observed and unobserved property characteristics. De Haan et al. (2020) argued that the fixed-effects approach is susceptible to quality-change bias when product turnover is high as this leads to a lack of matching over time. The rental market is dynamic but changes are very much less frequent than clothing, the focus of de Haan's study, which is well know for its lack of matching due to seasonality. We consider the number of price observations per property to guide our choice of data window length in Subsubsection 5.1.1. De Haan et al. (2020) cautioned that the fixed-effects approach does not immediately reflect new properties (products) since at least two observations are needed for a property to have a contribution to the fixed-effects model. In Section 4, we will consider how to deal with model revisions as additional data becomes available.

The fixed-effects regression, by definition, assumes that property characteristics are constant across time. This controls for the changing 'quality-mix' of properties available for rent, as properties enter and exit the rental market. At an individual property level, the quality of a property is assumed fixed over time. There are two important aspects to this: physical property characteristics relating to maintenance (renovations) and depreciation (ageing); and changing consumer valuations of property characteristics, including the amenity value of location.

We assume (within the estimation window) that, on aggregate, maintenance and depreciation will broadly off-set each other, so that the physical quality of rental stock age-cohorts is constant over time. Residential dwellings are assumed to have an average life of 70 years in New Zealand's National Accounts (Stats NZ 2014); across a sample of countries a median of 75 years life was reported in a 2013 Eurostat-OECD survey (Eurostat and OECD 2014). These long service lives of dwellings provides some reassurance that depreciation is not a major issue within a short-moderate temporal window (see, for example, Shimizu and Diewert 2019, for consideration of potential age adjustments).

The validity of the assumption of fixed consumer valuation of property characteristics depends on what is meant by a quality-adjusted price index; the purpose of the price index. A price index used to maintain the purchasing power of money may consider the quality of a property to be 'the same' over time, even if it is now in a more desirable location due to, say, economic growth, improved labour market prospects, better services, or enhanced local infrastructure. As noted by Silver (2016), something has to be held constant to separate price and quality change so as to estimate 'pure' price inflation. One way to consider this constraint is that the fixed property characteristics control for a counter-factual – how much *would* prices have changed *if* quality was constant – to estimate inflation, rather than making any assumption that needs to reflect reality. In reality, price and quality are inseparable. Sensitivity to data window lengths (duration of fixed property characteristics) is explored further in Subsection 5.1.

Diewert (2004), in the context of the Country Product Dummy model, suggested that Weighted Least Squares (WLS) should be used to reflect the economic importance of observed prices. Ivancic et al. (2009) applied this approach in the temporal context, weighting each observation by its expenditure share. We follow suit, noting that this was found to result in numerically similar estimates to those using Ordinary Least Squares (OLS).

4. Refining the Model: Data Window Length and Index-Chain Alignment

A natural starting point to estimating model (1) is to use all the data available, across all time periods (we did not consider data cleaning as each property is self-weighted in the model, this was confirmed to be a suitable strategy later when sensitivity to outlier treatment and editing produced very similar results). A criticism of this approach is that the estimate of the most recent period- on-period change is partly dependent on all other time periods, including the distant past. The term 'characteristicity' has been used to describe the influence of data in distant time periods on the comparison at hand (Caves et al. 1982); the less influence, the greater the characteristicity.

Characteristicity is often noted from a real-time perspective. For multilateral models, as additional periods of time occur, and are appended to the data, estimates for all period-on-period changes get updated (revised). Real-time estimation also leads to consideration of temporal sample-size equality – the number of time periods (the time sample) used to estimate each period-on-period change. Allowing the number of time observations to grow can be thought of as an expanding window of data (Chessa 2016). This leads to a non-uniform temporal sampling strategy being used for estimation.

Characteristicity can be increased by estimating the model based on a temporal subset, or 'window', of data. Using a window of fixed length ensures temporal sample equality, as the same number of time periods are used for each period-on-period estimate. Greatest characteristicity can be achieved by considering two-period windows (a bilateral method). However, to create a longer time series requires bilateral estimates to be chained together over time. These chained series have a big disadvantage of not being 'transitive' (Ivancic et al. 2009). Transitive price index methods are invariant to the choice of base period and index time path; the same result is achieved if periods are compared directly, or through their relationship with other time periods. Within a given data window the property fixed-effects estimates will be transitive. To strike a balance between characteristicity and transitivity a chained rolling window, of fixed length, can be used. This is discussed in the next section.

4.1. *Choosing a Data Window Length*

There is a lack of consensus on approaches to determining appropriate data window lengths and index-chain alignment. The *Handbook on Residential Property Prices Indices* (Eurostat et al. 2013) suggests choosing a window length that "yields 'reasonable' results", but doesn't address how best to determine what is reasonable. Drechsler (1973) noted that "characteristicity and circularity [transitivity] are always... in conflict with each other". De Haan (2015) observed that "It is likely that the quality-adjusted prices from the TPD model approximate the quality-adjusted prices from the hedonic model better as the sample period grows and the number of matches for a particular item in the data increases. On the other hand, we do not want the sample period to become very long because this conflicts with the underlying assumption of fixed characteristics parameters. So there is a trade-off, but it is difficult to tell what the optimal sample period would be." Diewert and Fox (2017) suggest, in the context of the economic approach to index number theory, that the "longer the window length is, the more likely it is that substitution bias will increase".

Arguments in favour of:

Shorter window	Longer window
Greater characteristicity Allows model parameters to change, reflecting changing quality and consumer preferences Minimises substitution bias (Diewert and Fox 2017)	Greater transitivity More infrequent price observations included; greater representivity Improved model fit (De Haan 2015)

Ivancic et al. (2009) chose a window length of 13 months "as it allows for strongly seasonal commodities to be compared". Ivancic had high frequency supermarket scanner data, and only 15 months of data which limited the choice of maximum window length. Stats NZ and the Australian Bureau of Statistics (Australian Bureau of Statistics 2017) are using nine quarter windows for consumer electronic and supermarket scanner data respectively. Silver (2016), in the context of property prices, notes that "a ten-year window... with valuations of characteristics held constant may stretch credibility".

4.2. Index-chain Alignment

To create a time series longer than the chosen window length, requires a choice of *index-chain position* (Figure 1). This is new terminology to separately distinguish (1) the conceptual choice of index-chain position and (2) a common publication constraint not to revise historical estimates (which we address in Subsection 4.3). We will first interpret the logic of several researchers' choice of *splice* position to consider analogous *index-chain* positions. Rolling multilateral windows will overlap for multiple time periods. The natural choice, from a real-time estimation perspective, is to link on the most recent overlap period. That is, the end of the time series in the previous window and the lagged one period end of the time series in the newer window – **'end' chain alignment.**

Extending the notation of White (2018): let P_{OLD} be the index computed over periods 1 to w (the window length), and let P_{NEW} be the index computed over the window rolled forward one period, from periods 2 to $w+1$.

The end chain aligned index between periods $w-1$ and $w+1$ can be expressed as:

$$P_{EndChain}^{w-1,w+1} = \frac{P_{OLD}^{w}}{P_{OLD}^{w-1}} \times \frac{P_{NEW}^{w+1}}{P_{NEW}^{w}} \tag{3}$$

However, such an approach does not allow for the effect of new products to be captured in the chained time series. Following the logic of Krsinich (2016) an index-chain alignment at the most distant overlap period would alleviate this problem – *'start' chain alignment.*

The start chain aligned index between periods 1 and 3 can be expressed as:

$$P_{StartChain}^{1,3} = \frac{P_{OLD}^{2}}{P_{OLD}^{1}} \times \frac{P_{NEW}^{3}}{P_{NEW}^{2}} \tag{4}$$

Yet, by symmetry, this may create the opposite issue of not capturing well the effect of disappearing products. Following the half-window splice suggested by De Haan (2015), this may be resolved with a '**mid' chain alignment**, or using the geometric mean of all possible chain positions – **'mean' chain alignment**, suggested by Diewert and Fox (2017) in the context of a window splice solution.

The mid chain aligned index between periods $h-1$ and $h+1$ can be expressed as:

$$P_{MidChain}^{h-1,h+1} = \frac{P_{OLD}^{h}}{P_{OLD}^{h-1}} \times \frac{P_{NEW}^{h+1}}{P_{NEW}^{h}} \tag{5}$$

where $h = (w/2) + 1$ if w is even; $h = ((w-1)/2) + 1$ if w is odd.

Let P_u^t be the index value at time t computed using data of vintage u, over the periods $1+r$ to $w+r$, where the period rolled forward r increases by 1 for each successive vintage (a generalisation of P_{OLD}, where $r=0$, and P_{NEW}, where $r=1$, to consider 3 or more vintages of data). The mean chain aligned index between periods $t-1$ and $t+1$ can be expressed as:

$$P_{MeanChain}^{t-1,t+1} = \prod_{u=1}^{V} \left(\frac{P_u^t}{P_u^{t-1}}\right)^{\frac{1}{V}} \times \prod_{u=1}^{V} \left(\frac{P_u^{t+1}}{P_u^{t}}\right)^{\frac{1}{V}} \tag{6}$$

An advantage of the mean-chain alignment is that an estimate is generated for all time periods, contained in any window. Other chain alignment positions result in shorter timeseries. End-chain aligned timeseries will be left censored (the price index for periods 2 to $w-1$, in the first window can not be end-chained); start-chain aligned will be right censored; mid-chain aligned will have some left and right censoring. For example, using

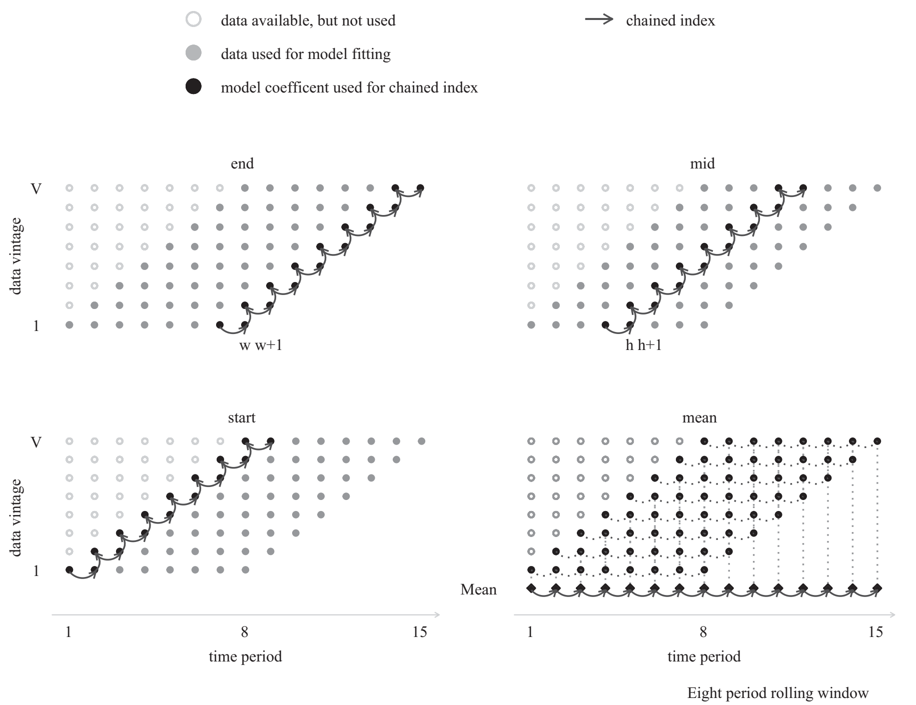


Fig. 1. Index-chain alignment options.

an eight-period rolling window (shown in Figure 1): periods 1–6 can not be end-chained; periods 10–15 can not be start-chained (using the data available at time period 15). Mean-chain alignment uses the mean of all available multilateral estimates so a consistent approach can be used to generate a chained timeseries for all periods 1–15.

A potential shortcoming of using mean-chain alignment is that it will require greater computational resource to compute. We did not find it prohibitively so.

4.3. Real-Time Estimation, With a Constraint not to Revise Historical Timeseries

An important additional consideration for a price index used for indexation of monetary payments is to add a *no revision of historical time series* constraint. That is, the first published estimate is never revised (or only in exceptional circumstances, such as a large data processing error). The constraint ensures that first published period-on-period change can be used for indexation with the confidence that the official estimate is final.

The four index-chain align options (start, mid, end, mean), can be applied as a catch-up (revision) factor to the latest period, as additional data become available (see Figure 2). Using a revision factor in the latest time period, helps to ensure that the long-run index is not biased should the model tend to be revised in a common direction (up or down) as additional data becomes available (see Krsinich, 2016). The downside to adopting this approach, is that the period-on-period change now reflects both the observed change between the periods at hand, and a bias correction (revision) factor.

Preference for **end index-chain alignment** leads to what De Haan (2015) called a **'movement splice'**, where there is no revision factor. Preference for **start index-chain alignment** leads to what Krsinich (2006) named a **'window splice'**, where the revision factor is determined by the difference between the old and new estimates for the cumulative change over the periods common both to the old and new windows. Revisions

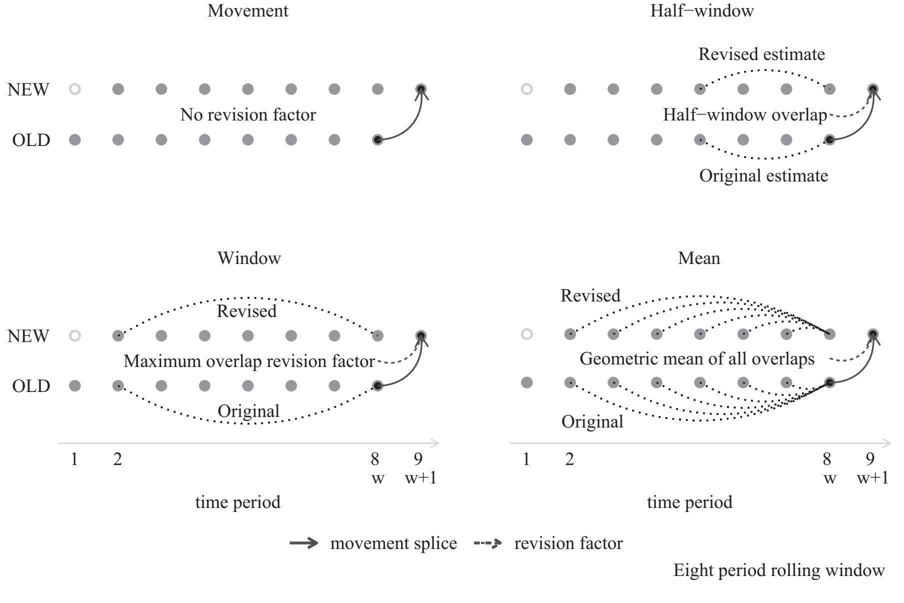


Fig. 2. Splice options.

to estimates for a particular period will affect the latest period-on-period change until sufficient real-time has elapsed to generate data to calculate the start index-chain aligned series.

The window splice index between periods w and $w + 1$ can be expressed as:

$$P^{w,w+1}_{WindowSplice} = \frac{P^{w+1}_{NEW}/P^{2}_{NEW}}{P^{w}_{OLD}/P^{2}_{OLD}} \tag{7}$$

To explicitly see the revision factor we can rewrite this as:

$$P^{w,w+1}_{WindowSplice} = \frac{P^{w+1}_{NEW}}{P^{w}_{NEW}} \times \frac{P^{w}_{NEW}/P^{2}_{NEW}}{P^{w}_{OLD}/P^{2}_{OLD}} \tag{8}$$

where the first term is the movement splice estimate and the second term is the revision factor.

Preference for **mid index-chain alignment** leads to what the Australian Bureau of Statistics (2016) called a **'half-window splice'**, where the revision factor is determined by the difference between the old and new estimates for the cumulative change for the periods belonging to second half of the common periods for old and new windows. Revisions to estimates for a particular period will affect the latest period-on-period change until sufficient real-time has elapsed to generate data to calculate the mid index-chain aligned series.

The half-window splice index can be expressed as:

$$P^{w,w+1}_{HalfWindowSplice} = \frac{P^{w+1}_{NEW}}{P^{w}_{NEW}} \times \frac{P^{w}_{NEW}/P^{h}_{NEW}}{P^{w}_{OLD}/P^{h}_{OLD}} \tag{9}$$

where $h = (w/2) + 1$ if w is even; $h = ((w-1)/2) + 1$ if w is odd.

Preference for **mean index-chain alignment** leads to what Diewert and Fox (2017) called a **'mean splice'**, where the revision factor is determined by the geometric mean of all overlaps. Defined here as:

$$P^{w,w+1}_{MeanSplice} = \frac{P^{w+1}_{NEW}}{P^{w}_{NEW}} \times \left(\prod_{t=2}^{w-1} \frac{P^{w}_{NEW}/P^{t}_{NEW}}{P^{w}_{OLD}/P^{t}_{OLD}} \right)^{\frac{1}{w-2}} \tag{10}$$

This treatment differs incrementally from the mean splice described by Diewert and Fox (2017), which for comparison can be written as Equation (11).

$$P^{w,w+1}_{MeanSplice} = \frac{P^{w+1}_{NEW}}{P^{w}_{NEW}} \times \left(\prod_{t=2}^{w} \frac{P^{w}_{NEW}/P^{t}_{NEW}}{P^{w}_{OLD}/P^{t}_{OLD}} \right)^{\frac{1}{w-1}} \tag{11}$$

The 'end' overlap, $\frac{P^{w}_{NEW}/P^{w}_{NEW}}{P^{w}_{OLD}/P^{w}_{OLD}} = 1$, in the calculation of mean splice is not included as this factor will always be equal to 1, so does not represent a bias correction factor

5. Results

5.1. Newly Rented Properties ('Flow' Concept)

It was found from the analysis that the choice of data window length has a material impact on estimates of cumulative inflation (Figure 3). For the 25-years to 2017-Q4, using the administrative data, estimates for total inflation ranged from 55% (data window length of three quarters) to 127% (window of 90 quarters). This represents a range of average annual changes of 1.8% (three quarter window) to 3.4% (90 quarter window). A similar spread was seen in the results obtained for the survey data (Appendix).

5.1.1. Property Life-Cycle Metrics

As a pragmatic solution to the choice of data window length (the variability in which was uncovered by the 25-year panel data), an eight-year period was chosen. This selected period strikes a balance between the merits of shorter and longer windows. Property life-cycle metrics (Figure 4) shows that a minimum data window length of six years results in a median number of price observations of at least three. There appears to be a trend of decreasing frequency of observations. The average length of tenancies may be increasing; partly a reflection of the data set building-up over time as each historical long tenancy is required to become compliant with the bond lodgement legislation when a new tenancy begins. Price-change frequency statistics for the survey data also displayed evidence of decreasing frequency of price change (not shown). The chosen window duration of eight years accommodates the observed trend of lengthening tenancies continuing; a median of at least three observations will likely be achieved for some years to come.

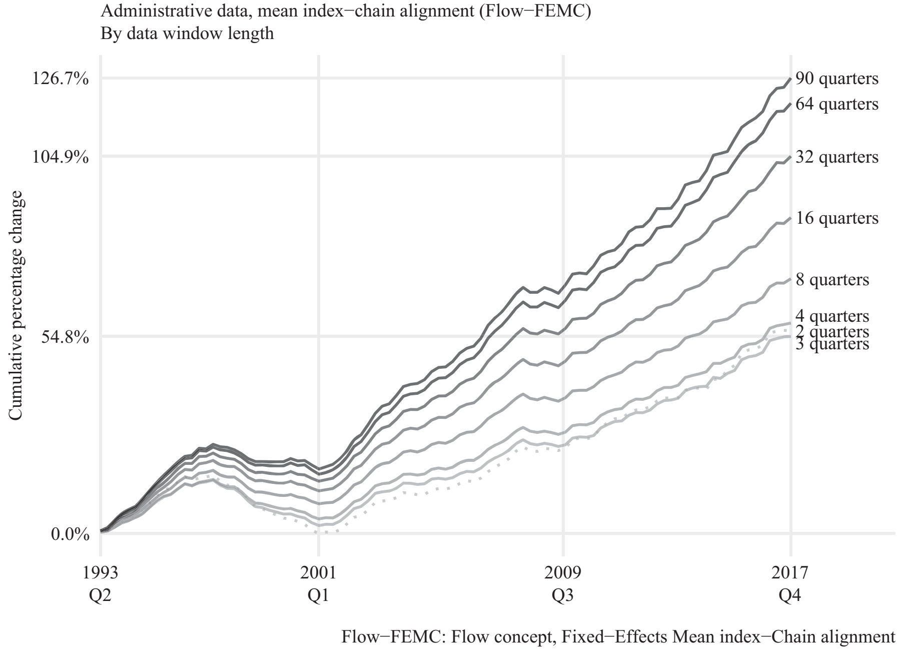


Fig. 3. Impact of data window length on estimates of rent price inflation.

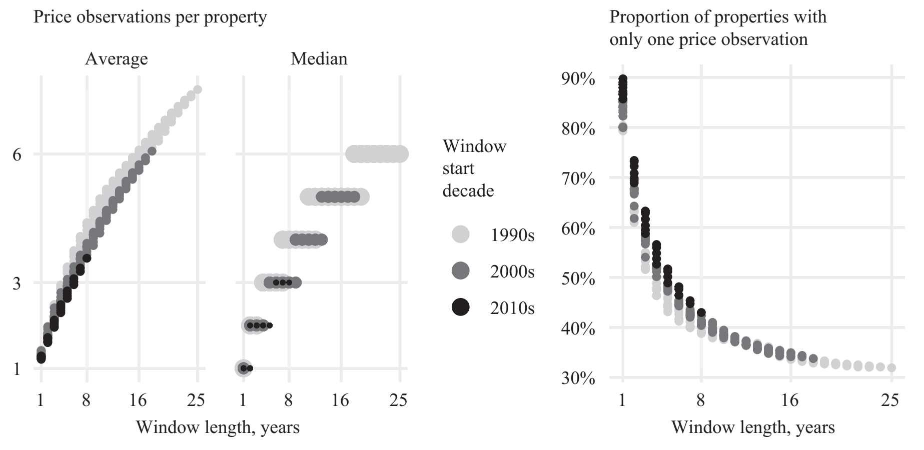


Fig. 4. Property life-cycle metrics.

Properties having only one observation within the estimation window are of particular interest as they will not be included in the fixed-effects estimator. It is necessary, therefore, to seek to minimise the proportion of properties with only one observation, in the event of a differential rate of inflation for these properties. The proportion of single-observation properties is less than half once the window length is greater than six years. The chosen eight-year window reduces the number of single-observation properties to about 40%, closer to the 30% rate observed over the full 25-year horizon. Long-term single-observation properties reflect those properties only available on the rental market once, and those with very long-term tenancies.

5.1.2. Index-Chain Alignment and Splice Position

Sensitivity analysis found only very small observed differences between the different index-chain alignments tested (Figure 5). The 'end' index-chain aligned series were found to have a higher annual percentage price change estimate than the estimates derived for the 'mean' aligned series. 'Start' aligned series were both above and below 'mean' aligned estimates. There was little discernible difference between 'mean' and 'mid' chain positions.

Given these empirical findings, and the advantages discussed in Subsection 4.2, it is proposed to adopt a mean chain alignment and, consistently, a mean splice methodology. We found little observed difference between the revisable chained series and the non-revisable spliced series (not shown). This suggests that the splicing method is working well: correcting for any long-run revision bias whilst the bias correction factor is not having a major impact on the period-on-period change.

5.2. Transformation of 'Flow' (New Tenancy Price) to 'Stock' (Currently Paid Rent) Concept

Rent prices are only observed in the bond data when tenancies begin, or a new bond lodgement form is submitted. In contrast, the survey of landlords was designed to be a

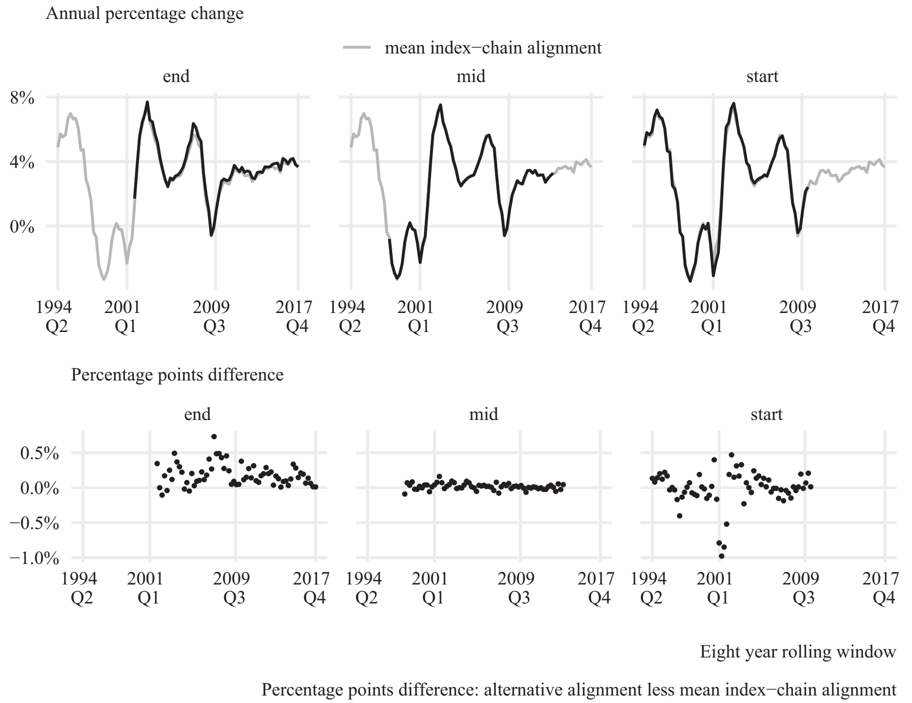


Fig. 5. Impact of index-chain alignment on estimates of rent price inflation.

nationally representative sample of the rent paid each period for all rental properties, regardless of whether these are new or existing tenancies. Lewis and Restieaux (2015) called these '*flow*' and '*stock*' measures of rents, for the administrative and survey data respectively. They noted that using stock measures in CPIs is current best practice (citing ILO et al. 2004). Johnson (2015), in the context of using rent price inflation for the rental equivalence approach to measuring owner-occupiers' housing costs, notes that arguments could be made for using the marginal (flow) of rent depending on 'the question that rental equivalence seeks to answer'.

Consultation with key price-statistics users in New Zealand, as part of this research, indicated a preference for retaining the existing stock concept for the CPI. Several users requested that a flow measure be published alongside, as a leading indicator of changes in the market price for new tenancies.

Applying the same model – FEMC(8y): Fixed-Effects, Mean index-Chain alignment, eight year rolling window – to both data sets, allows a closer look at the timing impact. Figure 6 shows the results for three major urban areas of New Zealand, alongside the country total (calculated by directly applying the model; not an aggregation of the regions). Beyond timing differences, the two series also reflect any coverage differences and, for the survey data, sampling error. In this work we have not employed the survey weighting used for the published series, as the weights were not easily obtainable. Both data sources relate to private-sector rentals (the administrative data has been limited to this coverage). The survey population excludes furnished properties whereas the administrative data includes both furnished and unfurnished properties.

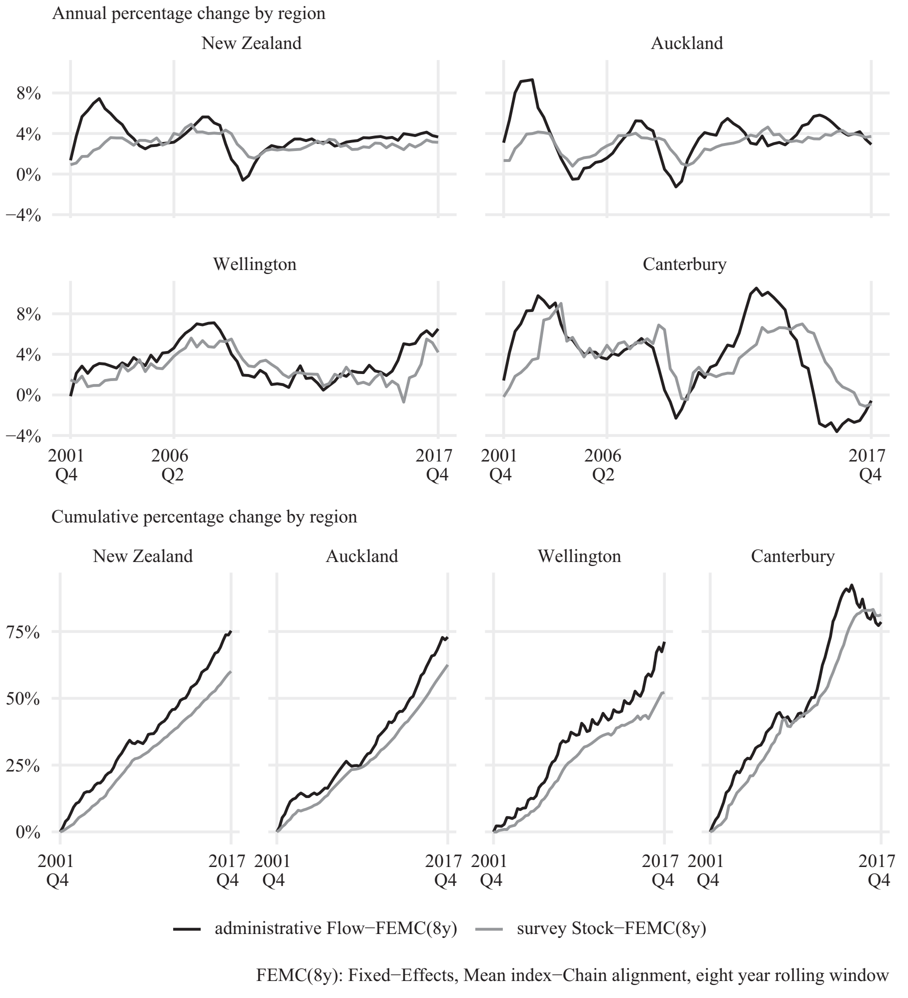


Fig. 6. Impact of data source and concept on estimates of rent price inflation.

The flow measure data shows relatively more volatility and earlier identification of turning-points in the time series than the equivalent stock data. This makes sense as the information from newly lodged bonds reflects the current market price for rental properties. The within tenancy rent price changes are likely dampened by legal or contractual obligations not to increase prices for a set time period or reflect a discount for reducing landlords' search and replacement costs (Miceli and Sirmans 1999).

Timing differences appear to be the major factor differentiating between data sources since 2006, when the survey birthing process was reestablished. This has been tested by running the model over a subset of the survey data, limited to observations that had a change in price from the previous quarter. Shown in Figure 8 (flow panel), the price-change-only survey data is quite volatile, due to the small proportion of the sample experiencing a price-change from the previous period (see Appendix); but the trend looks to track the administrative series well. This provides confidence that despite an aging survey design, and imperfect administrative data, both data sets are of reasonable, or at least similar, quality.

5.2.1. Stock Imputation

To estimate rent inflation on a 'stock' concept using the tenancy bond data, mass-imputation was used to estimate the rent currently paid between observed transactions. To reduce computational resource a 20% sample of properties was used, generated as a simple random sample of unique property identifiers. (Very similar results were later confirmed when all properties were used, once the processing was switched to a server with increased memory capacity). Rent prices tend to be quite sticky, so tend to increase in a step-wise pattern. Current rent prices were imputed each quarter by carrying forward the rent price recorded at the start of a tenancy (when the bond is lodged), for a maximum of two years or until the tenancy ends. Sensitivity to the imputation cut-off length is shown in Figure 7.

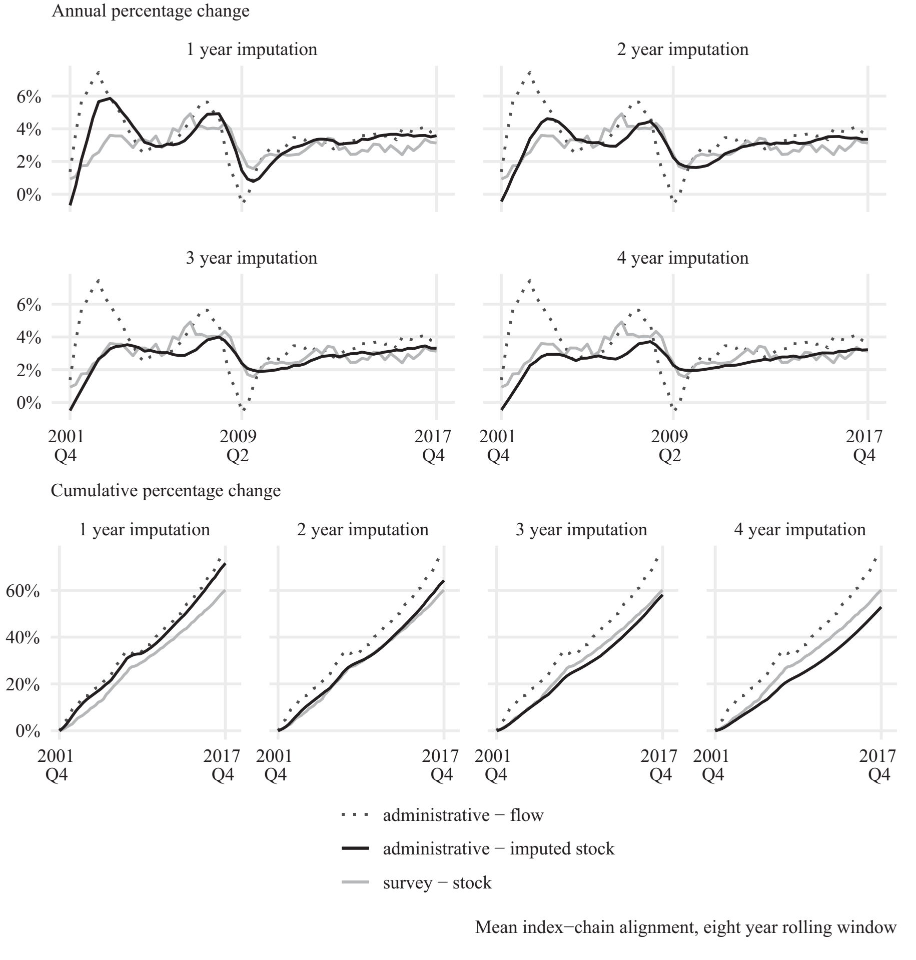


Fig. 7. Stock imputation: sensitivty to carry-forward length.

The two year cut-off is intended to cease imputation beyond the typical duration of rent prices. Investigation of the persistence of prices in survey data found an average duration of 1.8 years and a median duration of 2.1 years. The logic for the two-year cut-off is that for properties with a bond lodged in the last two years, it is most likely that there has been no price increase. For properties lodged more than tow years ago, it is most likely there has been a price increase, with the price increased assumed to be in line with the market. Properties that do not have a rent price explicitly imputed will be implicitly imputed by the fixed-effects regression model, as if the data were missing at random (Summers, 1973). In 2017, each quarter, about 9% of the rental stock (estimated as bonds that have been opened and not closed) were newly lodged, so the reported rent related to that quarter. A further 47% of properties were lodged within the past two years, so had their rent carried-forward from the lodgement date. The remaining 44% of properties had a missing rent amount, in a given quarter, which would be implicitly imputed by the fixed-effects estimator.

The survey data can be used to validate the stock-imputed geometric mean weekly rent amount estimated for the imputed admin data, compared with the survey, which is designed for a stock concept. Since 2006-Q2, once the survey new-property birthing process was re-established, the geometric mean prices are similar (see Figure 8, stock panel). The trend for geometric mean prices on a flow concept are also similar (Figure 8, flow panel), suggesting the survey and admin data have similar coverage of market rent.

6. Conclusions

The primary finding of this study is that the longer the data window used to estimate price change, the greater the estimated rate of rent price inflation. This raises fundamental questions about what we mean by a price index. Choice of window length is usually constrained by available data. Traditionally, sparse data has lead to *bilateral* index number theory. However, new administrative data sources facilitate the application of *multilateral* index number methods in a time series context. This article investigates one such method, a property fixed-effects estimator.

To the best of our knowledge, this is the first reported study to explore the impact of choosing different window lengths on a time series of 25-years. This forces one to consider the purpose of the price index, and the quality required to be held constant in order to estimate 'pure' price change. This is particularly pertinent for property characteristics, since the quality of locations varies over time due to changes in amenities and perceptions of desirability (relating to both tangible and intangible characteristics). Assessing the impact of varying data window lengths across time, using a fixed-effects estimate of price change, helps illuminate the impact of holding constant property quality characteristics.

Given relatively infrequent changes in rent prices, compared with most retail prices, a relatively long data window (such as 32 quarters; eight years) appears necessary to provide reasonable transitivity and property-level matches and therefore minimise long-run bias. The choice of window length of eight years as used here is pragmatic, balancing competing arguments for longer and shorter windows. Property life-cycle metrics (such as median number of price observations per property; proportion of properties with only a single observation) were used to inform the selected eight-year period, but we do not have a general theory to guide such choices. We expect more research will follow.

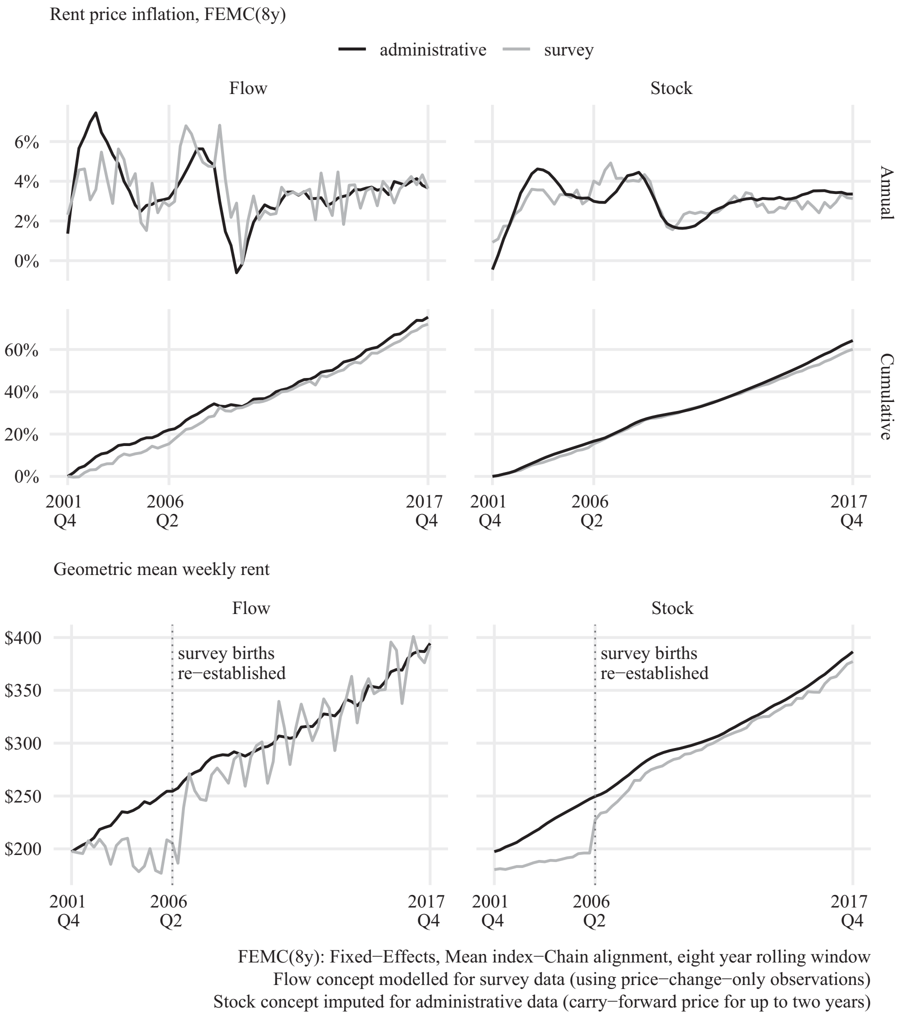


Fig. 8. Modelling stock and flow concepts.

Index-chain alignment and revisions due to additional data are important considerations for all multilateral price models. In this article we have introduced new terminology to distinguish the conceptual choice of *index-chain position* separately from *splice* options which address the combined problems of chaining and revision constraints. Yet, these design choices have been found to be of less importance for rent price data when compared with the impact of data window length.

In the New Zealand context, we have resolved a long-standing puzzle of why quality-adjusted rent price inflation (as published in the CPI) increased at a substantially lower rate than geometric mean rent prices. It was found that much of the observed differences between these series can be explained as biased quality adjustment resulting from the use of a bilateral matched-sample approach (Figure 9). Using a property fixed-effects multilateral model to perform quality-adjustment, differences between average price change and 'pure' price change are much smaller (the implied quality improvements over

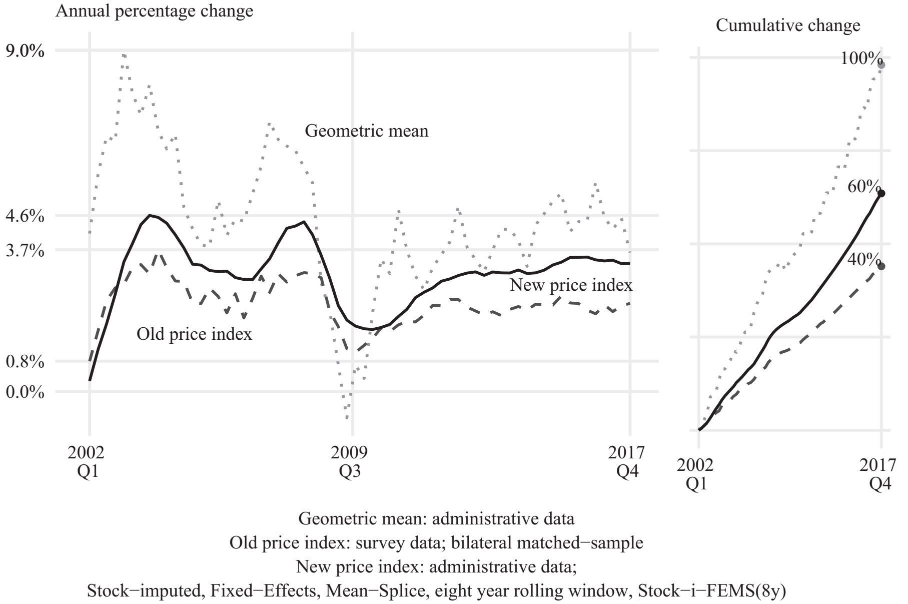


Fig. 9. Comparison of old and new price indexes.

time are less). Analysis of the impact on the all-groups CPI found that an alternative aggregate index, using the rent price index as proposed in this research, would have given an additional cumulative 0.95 percentage points increase between 2006-Q2 and 2017-Q4 (average annual change would have increased an additional 0.07 percentage points) when compared with the published CPI.

Since 2019-Q2, the 'rentals for housing' class of New Zealand's CPI has been estimated using the non-revisable model-based estimator – the FEMS(8y): Fixed-Effects, Mean-Splice, eight year rolling window – on stock-imputed administrative microdata, Stock-i-FEMS(8y). The stock-imputation estimates the currently paid rent for the entire rental '*stock*', thereby negating the need for a survey of landlords and thus reducing operational costs and eradicating survey burden. Applying the same estimator to raw flow-based administrative data, Flow-FEMS(8y) has created a new official statistic to track changes in the market price of new tenancies. Monthly series will provide more frequent price statistics than the longstanding quarterly series (Stats NZ 2019).

7. Appendix: Sensitivity to Data Window Length (Survey Data)

It seems astonishing that estimated inflation should be so sensitive to data window length; even when using the same core estimator, a fixed-effects regression. This section explores this phenomena by observing a natural experiment that happened as a consequence of the lapsed survey birthing process over five-year period, 2001–2006.

Shown in Figure 10 the achieved sample size decreased from about 1,700 properties in 2000-Q2 to just over 850 properties in 2006-Q1. The birthing process was reinstated in 2006-Q2 and the sample size was back up to about 1,700 properties one year later in 2007-Q2. A relatively small proportion of the sample has a price change in a given quarter,

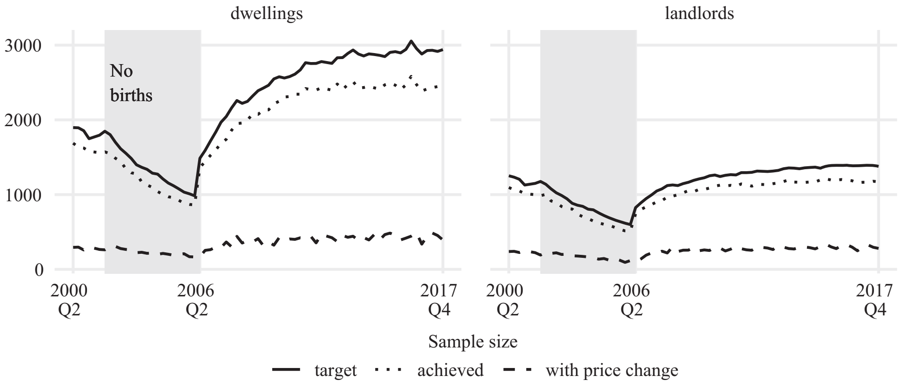


Fig. 10. Quarterly rent survey: sample size over time.

compared with the previous quarter. We note that the with-price-change sample size was below 200 properties for a time.

The fact that the survey maintained as a panel, but not updated for new (or returning) rental properties, halved in size over five-years illustrates the dynamic nature of the rental market.

In the years immediately after the return of sample maintenance, 2006–2009, the newly observed properties would likely record no price change (as the tenancies would already be priced at the market rate at the start of the tenancy). Price change associated with newly rented properties are missed in a bilaterial matched-sample index. In retrospect, using a multilateral fixed-effects estimator, price change for newly rented properties are implicitly

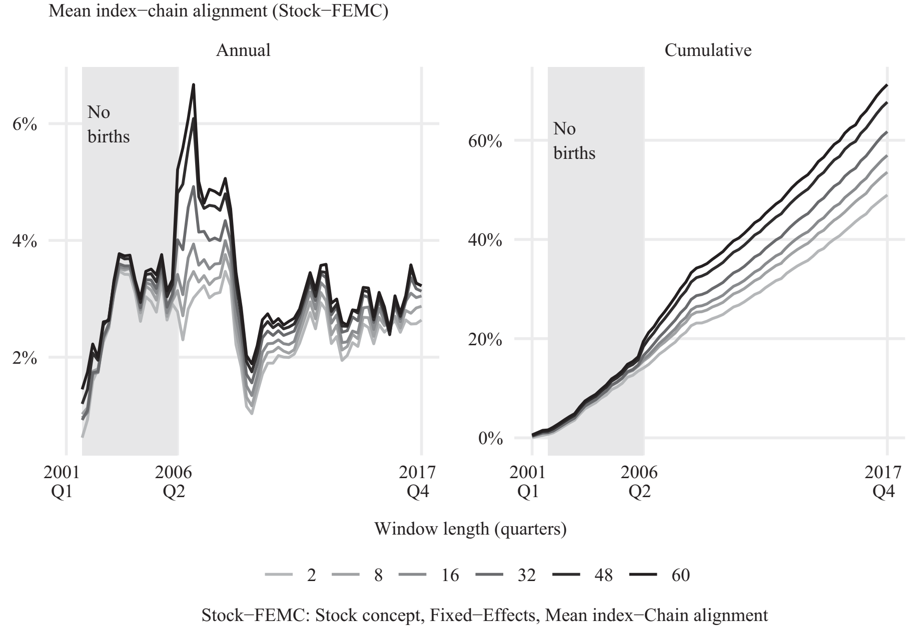


Fig. 11. Impact of data window length: survey data.

imputed, once the window length is sufficient to include multiple price change observations (Krsinich, 2016). Aizcorbe et al (2003) explain what is occurring:

> "For the turnover good, the hedonic [fixed-effects] regression imputes a price relative as the difference between the quality-adjusted price for the new variety at time t and the average quality-adjusted price for all observed models in the prior period... In contrast, [bilateral] index number methods are silent on how to handle the missing prices associated with new and exiting goods."

For rental properties, the average duration of prices was found to be 1.8 years. The distribution has a long tail with some rental prices persisting for many years. This explains why the estimated rent price inflation continues to be revised upwards in the years 2006–2009 (Figure 11) as the window length increases up to 60 quarters (15 years).

8. References

Aizcorbe, A., C. Corrado, and M. Doms. 2003. *When Do Matched-Model and Hedonic Techniques Yield Similar Price Measures?* Working Paper: 2003-14, Federal Reserve Bank of San Francisco. DOI: http://dx.doi.org/10.2139/ssrn.550421.

Ambrose, B.W., N.E. Coulson, and J. Yoshida. 2015. "The Repeat Rent Index." *The Review of Economics and Statistics.* 97 (5): 939–950, DOI: https://doi.org/10.1162/REST_a_0050

Australian Bureau of Statistics. 2016. *Making Greater Use of Transactions Data to Compile the Consumer Price Index*. Information Paper: 6401.0.60.003, Canberra: ABS. Available at: http://www.abs.gov.au (accessed October 2019).

Australian Bureau of Statistics. 2017. *An Implementation Plan to Maximise the Use of Transactions Data in the CPI.* Information Paper: 6401.0.60.004, Canberra: ABS. Available at: http://www.abs.gov.au (accessed October 2019).

Bailey, M.J., R.F. Muth, and H.O. Nourse. 1963. "A Regression Model for Real Estate Price Index Construction." *Journal of the American Statistical Association.* 58: 933–942. DOI: https://doi.org/10.1080/01621459.1963.10480679.

Behrmann, T., and B. Goldhammer. 2017. "New developments in the field of house and rental price indices in German price statistics in the light of the hedonic method." Paper presented at the Ottawa Group 2017, 10–12 May, Eltville, Germany. Available at: http://www.ottawagroup.org (accessed November 2020).

Bentley, A., and F. Krsinich. 2017. "Towards a big data CPI for New Zealand." Paper presented at the Ottawa Group 2017, 10–12 May, Eltville, Germany. Available at: http://www.ottawagroup.org (accessed October 2019).

Case, K.E., and R.J. Shiller. 1987. "Prices of Single Family Homes since 1970: New Indexes for Four Cities" *New England Economic Review*: 45–56. Available at: https://EconPapers.repec.org/RePEc:cwl:cwldpp:851.

Caves D.W., L.R. Christensen, and W.E. Diewert 1982. "Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers." *Economic Journal* 92: 73–86. Available at: https://EconPapers.repec.org/RePEc:ecj:econjl:v:92:y:1982:i:365:p:73-86.

Chessa, A.G. 2016. "A New Methodology for Processing Scanner Data in the Dutch CPI." *Eurona* 1: 49–69. Available at: https://ec.europa.eu (accessed October 2019).

Court, A.T. 1939. *Hedonic Price Indexes with Automotive Examples. In The Dynamics of Automobile Demand* General Motors, New York

De Haan, J., and F. Krsinich. 2014. "Scanner Data and the Treatment of Quality Change in Nonrevisable Price Indexes." *Journal of Business and Economic Statistics*. 32 (3): 341–358. DOI: https://doi.org/10.1080/07350015.2014.880059.

De Haan, J. 2015. "A Framework for Large Scale Use of Scanner Data in the Dutch CPI." Paper presented at the Ottawa Group 2015, 20–22 May, Urayasu, Japan. Available at: http://www.ottawagroup.org (accessed October 2019).

De Haan, J., R. Hendriks, and M. Scholz. 2020. "Price Measurement using scanner data: timeproduct dummy versus time dummy hedonic indexes." *Review of Income and Wealth*. DOI: https://doi.org/10.1111/roiw.12468

Diewert, W.E. 2004. *On the Stochastic Approach to Linking the Regions in the ICP*. Discussion Paper: 04–16. Department of Economics, University of British Columbia, Vancouver, Canada. Available at: https://ubc.ca (accessed October 2019).

Diewert, W.E., and K.J. Fox 2017. *Substitution Bias in Multilateral Methods for CPI Construction using Scanner Data.* Discussion Paper: 17-02, University of British Columbia, Vancouver, Canada. Available at: https://ubc.ca (accessed October 2019).

Drechsler, L. 1973. "Weighting of Index Numbers in International Comparisons." *Review of Income and Wealth* 19:1: 17–47. DOI: https://doi.org/10.1111/j.1475-4991.1973.tb00871.x

Eltetö, Ö., and P. Köves. 1964. " On a Problem of Index Number Computation Relating to International Comparisons" (in Hungarian), *Statisztikai Szemle* 42: 507–518. Available at: https://digitallibrary.un.org/record/750290?ln=en.

Eurostat, EU, ILO, IMF, OECD, UNECE, and The World Bank. 2013. *Handbook on Residential Property Prices Indices* (RPPIs). Luxembourg, European Union. Available from https://ec.europa.eu (accessed October 2019)

Eurostat and OECD. 2014. *Eurostat-OECD Survey of National Practices in Estimating Net Stocks of Structures*. Available from http://www.oecd.org (accessed November 2020).

Eurostat. 2018. *Harmonised Index of Consumer Prices (HICP) methodological manual: 2018 edition.* Eurostat. DOI: https://doi.org/10.2785/85564.

Geary, R.G. 1958. "A Note on Comparisons of Exchange Rates and Purchasing Power between Countries." *Journal of the Royal Statistical Society*. Series A 121: 97–99.

Gini, C. 1931. "On the Circular Test of Index Numbers." *Metron* 9: 3–24.

ILO, IMF, OECD, UNECE, Eurostat, and The World Bank. 2004. *Consumer price index manual: Theory and practice.* Geneva, International Labour Office. Available at: http://www.ilo.org (accessed October 2019).

Ivancic, L., W.E. Diewert, and K.J. Fox. 2009. *Scanner Data, Time Aggregation and the Construction of Price Indexes.* Discussion Paper 09-09, Department of Economics, University of British Columbia, Vancouver, Canada. Available at: https://ubc.ca (accessed October 2019).

Johnson, P. 2015. *UK Consumer Price Statistics: A Review.* UK Statistics Authority. Available at: https://www.statisticsauthority.gov.uk (accessed October 2019).

Keshishbanoosy, R., and L. Taylor. 2019. *New approach for estimating the rent component of the Consumer Price Index.* Statistics Canada. Available at: https://www.statcan.gc.ca/ (accessed November 2020).

Khamis, S.H. 1972. "A New System of Index Numbers for National and International Purposes." *Journal of the Royal Statistical Society.* Series A 135: 96–121. DOI: https://doi.org/10.2307/2345041.

Krsinich, F. 2009. "Using Hedonic Regression to Assess the Housing Rentals Component of the New Zealand Consumers Price Index." Paper presented at New Zealand Association of Economists 50th Anniversary conference, 1–3 July 2009. Wellington, New Zealand. Available at: https://www.nzae.org.nz (accessed October 2019).

Krsinich, F. 2014. "Quality adjustment in the New Zealand Consumers Price Index. In *The New Zealand CPI at 100: History and interpretation,* edited by S. Forbes and A. Victorio. Wellington: Victoria University Press.

Krsinich, F. 2016. "The FEWS Index: Fixed Effects with a Window Splice." *Journal of Official Statistics,* 32 (2): 375–404. DOI: https://doi.org/10.1515/jos-2016-0021.

Lewis, R., and A. Restieaux. 2015. *Improvements to the measurement of owner occupiers' housing costs and private housing rental prices.* UK Office for National Statistics. Available at: http://webarchive.nationalarchives.gov.uk (accessed October 2019).

Lowe, J. 1823. *The Present State of England in Regard to Agriculture, Trade and Finance.* Second edition, London: Longman, Hurst, Rees, Orme and Brown.

Miceli, T., and C.F. Sirmans. 1999. "Tenant Turnover, Rental Contracts, and Self Selection." *Journal of Housing Economics* 8: 301–311. DOI: https://doi.org/10.1006/j-hec.1999.0253.

Miller, S., S. Suie, and C. Bycroft. 2018. *Comparing housing information from census and tenancy bond data.* Stats NZ. Available at: https://www.stats.govt.nz (accessed October 2019).

Private Residential Tenancies Board. 2013. *The PRTB Rent Index.* Available at: https://www.rtb.ie/ (accessed November 2020).

Residential Tenancies Act (New Zealand). 1986. Available at: https://www.legislation.govt.nz/act/public/1986/0120/latest/whole.html#DLM95059.

Shimizu, C., K.G. Nishimura, and T. Watanabe. 2009. "Housing Prices and Rents in Tokyo: A Comparison of Repeat-Sales and Hedonic Measures." Paper presented at the Ottawa Group, 27–29 May 2009, Neuchatel, Switzerland. Available at: http://www.ottawagroup.org (accessed November 2020).

Shimizu, C., and W.E. Diewert 2019. "Residential Property Price Index in Japan: Discussion in Methodology and Data Sources." Paper presented at the International Conference on Real Estate Statistics, 20–22 February 2019, Luxembourg. Available at: https://www.real-estate-statistics.eu/ (accessed November 2020).

Silver, M. 2016. *How to Better Measure Hedonic Residential Property Price Indexes*. IMF working Paper 16/213. Available at: http://www.imf.org (accessed October 2019).

Stats NZ. 2008. "Rented dwellings in the CPI." *Price Index News*, April. Available from http://archive.stats.govt.nz (accessed October 2019).

Stats NZ. 2014. *Measuring capital stock in the New Zealand economy*. Fourth edition. Available at: https://www.stats.govt.nz (accessed November 2020).

Stats NZ. 2015. *IDI Data Dictionary: Tenancy bond data.* Available at: http://archive.stats.govt.nz (accessed October 2019).

Stats NZ. 2019. *New methodology for rental prices in the CPI.* Available at: https://www.stats.govt.nz (accessed October 2019).

Summers, R. 1973. "International Comparisons Based Upon Incomplete Data." *Review of Income and Wealth.* 19: 1–16. DOI: https://doi.org/10.1111/j.1475-4991.1973.tb00870.x.

Szulc, B.J. 1964. "Indices for Multiregional Comparisons", *Przeglad Statystyczny* 3: 239–254.

White, G. 2018. *IndexNumR: A Package for Index Number Calculation.* Comprehensive R Archive Network, R Foundation for Statistical Computing. Available at: https://CRAN.R-project.org/package = IndexNumR (accessed October 2019).

Received December 2019
Revised December 2020
Accepted December 2021

Journal of Official Statistics, Vol. 38, No. 1, 2022, pp. 213–237, http://dx.doi.org/10.2478/JOS-2022-0010

Experimental UK Regional Consumer Price Inflation with Model-Based Expenditure Weights

James Dawber[1], *Nora Würz*[2], *Paul A. Smith*[1], *Tanya Flower*[3], *Heledd Thomas*[3], *Timo Schmid*[4], *and Nikos Tzavidis*[1]

Like many other countries, the United Kingdom (UK) produces a national consumer price index (CPI) to measure inflation. Presently, CPI measures are not produced for regions within the UK. It is believed that, using only available data sources, a regional CPI would not be precise or reliable enough as an official statistic, primarily because the regional partitioning of the data makes sample sizes too small. We investigate this claim by producing experimental regional CPIs using publicly available price data, and deriving expenditure weights from the Living Costs and Food survey. We detail the methods and challenges of developing a regional CPI and evaluate its reliability. We then assess whether model-based methods such as smoothing and small area estimation significantly improve the measures. We find that a regional CPI can be produced with available data sources, however it appears to be excessively volatile over time, mainly due to the weights. Smoothing and small area estimation improve the reliability of the regional CPI series to some extent but they remain too volatile for regional policy use. This research provides a valuable framework for the development of a more viable regional CPI measure for the UK in the future.

Key words: CPI conceptual framework; basket of goods and services; small area estimation; Fay-Herriot models.

1. Introduction

For a long time, users of price statistics in the UK have suggested that regional indices of consumer prices would be valuable in understanding how inflation varies across the country, and whether there are important differences in regional inflation (RPI Advisory Committee 1971; Fenwick and O’Donoghue 2003; UK Statistics Authority 2013, paragraph 3.13). The official position has been that the regional numbers of price quotes are too small to support the calculation of indices, and it has not been a sufficiently high

[1] University of Southampton, Highfield, Southampton, SO17 1BJ, UK. Emails: J.P.Dawber@soton.ac.uk, P.A.Smith@soton.ac.uk and N.TZAVIDIS@soton.ac.uk
[2] Freie Universität Berlin, Department of Business and Economics, Garystraße 21, Berlin, 14195, Germany. Email: NoraUlrike.Wuerz@fu-berlin.de
[3] Office for National Statistics, Government Buildings, Cardiff Road, Newport, NP10 8XG, UK. Emails: tanya.flower@ons.gov.uk, Heledd.Thomas@ons.gov.uk
[4] Otto-Friedrich-Universität Bamberg, Feldkirchenstraße. 21, Bamberg, 96052, Germany. Email: timo.schmid@uni-bamberg.de
Acknowledgments: James Dawber and Paul Smith were funded by the ONS under contract PU-16-0031-6.009. Nikos Tzavidis and Paul Smith were supported by EU Horizon 2020 funding as part of the MAKSWELL project, grant agreement No. 770643. Nora Würz received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 730998, InGRID-2 – Integrating Research Infrastructure for European expertise on Inclusive Growth from data to policy. The opinions expressed are those of the authors, and do not necessarily reflect those of their organisations or the ONS.

© Statistics Sweden

priority to invest in additional price collection for this purpose. Some limited information from the Office for National Statistics (ONS) on variation in regional prices has been made available through publications on Relative Regional Consumer Price Levels (RRCPLs) (Baran and O'Donoghue 2002; Wingfield et al. 2005; ONS 2011), which have used information from additional price collections made every six years to adjust Purchasing Power Parity (PPP) statistics. PPP prices are collected in the capital city, and a periodic exercise is undertaken to adjust indices to represent the whole country. RRCPLs are *spatial* price indices, which show the differences in price levels between regions (relative to a reference region = 100) with a fixed basket of goods, but are not *temporal* indices designed to show price change (inflation, relative to a reference time = 100), because the basket is different each time they are produced. Because of the methodology and differences in the weights on each occasion, RRCPLs are not satisfactory even for a once every six years approximation of regional inflation. The focus of this article is on a regional index that can measure the temporal differences (inflation), rather than the spatial differences. So we allow the baskets and weights to vary by region, to best reflect expenditure patterns within regions. The resulting indices are suitable for comparing inflation rates between regions, but not for comparing price levels.

The Consumer Prices Index (CPI) is used as a national measure of inflation in many countries. Some countries have also developed inflation measures at a sub-national level. In the United States (US) the construction of the CPI includes a regional aggregation phase, and therefore component indices at the regional level are part of the standard outputs (BLS 2018); Japan also produces regional indices (e.g., Statistics Bureau of Japan 2020; Nagayasu 2011). Weber and Beck (2005) compiled a database of regional prices for Europe where they found regional or city indices in Austria, Finland, Germany, Spain and Portugal. In Germany, the Federal Statistical Office publishes regional CPIs for the 16 federal states (Statistisches Bundesamt 2020) and there are also publications with regional indices for 401 German districts relying on an econometric model (Kosfeld et al. 2009). The German national weighting pattern is updated every five years using the "Einkommens und Verbraucherstichprobe" (Statistisches Bundesamt 2018), but the sample size is insufficient for the estimation of regional weights. Other countries with regional inflation measures include Poland (Gajewski 2017), Russia (Brown et al. 2018), Indonesia (Purwono et al. 2020), South Korea (Tillman 2013) and Turkey (Yesilyurt and Elhorst 2014; Duran 2016). We did not find detailed methodological descriptions in all these cases, but of those countries where information is available, almost all have a regional aggregation stage in constructing the national price index, so that regional indices are automatically available. Other methods are the econometric model for district inflation in Germany, and an experimental investigation of small area estimation in Indonesia (Fengki et al. 2020; sec. 5).

Other countries, including the UK, do not have a regional aggregation phase, so a special exercise is needed to produce regional CPIs. There has been interest in regional indices over many years in the UK. The Chancellor of the Exchequer announced work on regional prices in 2003 (Fenwick and O'Donoghue 2003), which was translated into the development of RRCPLs. Although these have been published each time PPP data collections have taken place, there has not been any substantial development of these statistics. Fenwick and O'Donoghue (2003) also discuss the potential for regional temporal indices, but conclude that they need further development; the annex to their paper lists the issues which make such a development challenging.

Economists have an interest in regional variations in price inflation (and more widely in regional differences in the cost of living, which is not so easily defined or calculated). Borooah et al. (1996), Hayes (2005) and Rienzo (2017) have all attempted to calculate regional versions of a consumer price measure for the UK with simplified methodology and based on available data sources. Regional variations in price levels measured by PPPs are also important inputs in local economic analysis. For example, Marchetti and Secondi (2017) estimate regional household consumption expenditure adjusted for differences in regional PPP in Italy, and Marchetti et al. (2019) use regional (province)-specific prices to adjust the national poverty threshold in Italy. Both of these applications use small area modelling approaches to make predictions of the variables of interest at local levels.

We assess the feasibility of producing a regional CPI measure for comparison of inflation rates between regions of the UK, that is, regional temporal indices, based on the official data collections. The UK has twelve statistical regions, including Wales, Scotland, Northern Ireland and nine regions of England; these are the Nomenclature of Territorial Units for Statistics (NUTS) 1 statistical regions. We develop a regional CPI rather than a CPIH (CPI including owner occupiers' housing costs) despite the importance of regional variation in housing and rental costs, because of the additional complications caused by the lack of regional microdata.

The derivation of the CPI requires data on price changes of certain goods, and also on consumer expenditure, used to determine both the basket composition and the weights (the proportion of spending on the goods). Sources are needed for all these components to ensure that the national CPI can be used to calculate reliable inflation rates, with suitable sampling designs and sample sizes to ensure they are nationally representative. For the development of a regional CPI, these data for prices and expenditure are partitioned into the separate regions. The reliability of a CPI at the regional level will be reduced because the smaller sample sizes lead to lower precision of the estimates. We consider this the primary limitation to the development of reliable and temporally stable regional CPIs. Although there may be bias in the regional CPI estimates as well as high variance, we believe the latter to be the larger problem and assume the bias to be ignorable.

A second limitation is the availability of regional-level data sources. At the national level, expenditure weights are calculated using many data sources. However, not all these sources are readily available for each region. For example, at the time of researching there was no regional equivalent for the National Accounts, though since then ONS (2018) has published experimental household final consumption expenditure (HFCE) at the regional level. For future research this regional HFCE data could be investigated to give balanced regional expenditure weights. Other data sources, such as administrative data are not very accessible, but the Living Costs and Food (LCF) survey data are accessible and also have region identifiers, hence can be used to estimate regional expenditure. For this reason, we use only LCF survey data to estimate the expenditure weights. Price data do not have the same issues since the data are readily available with region identifiers, and are used directly.

The aims of this research are first to assess the feasibility of calculating an experimental regional CPI series using available data sources, and second, to investigate model-based methods to overcome the primary limitation of the reduced sample sizes. We look at smoothing and small area estimation (SAE) methods that may improve the reliability of regional CPIs without having to collect additional data. SAE methods based on composite estimation have been examined in the US in a similar context (Swanson et al. 1999), however we focus on the use of Fay-Herriot models.

The structure of this article is as follows. In Section 2, we provide more background on the conceptual framework of a regional CPI as well as background on available data sources and the Classification of Individual Consumption by Purpose (COICOP). In Section 3, we present methods for constructing an experimental regional CPI with just LCF survey data and publicly available price data. We also assess the experimental regional CPI series for 2010 – 2016. In Section 4 we investigate the use of smoothing and small area estimation approaches to estimate the regional weights, intending to improve the regional CPI series. Finally, in Section 5 we discuss the results and suggest further research.

2. Structure, Data Sources and COICOP Classification

2.1. A Conceptual Framework for Regional CPIs

To have a sound basis for the methodology of regional price indices, it is important to set out the variations in the conceptual framework from the calculation of a national index. Here we set out the target concept; then we can be clear where the available data requires us to deviate from it.

The starting point for a (temporal) regional price index should be the regional basket of goods and services, to account for differences in spending patterns between regions. A threshold is used to exclude products with only minor expenditures. The maximum threshold is set at one part per thousand (ppt) according to the EU regulation 1687/98 on HICP (EU regulation 1998), but the UK implementation uses judgement for products with between 0.14 and 0.56 ppt of expenditure. For the regional estimates, we therefore take a fixed threshold of 0.5 ppt. Products below this threshold are not considered in the derivation of the index.

For items in the regional baskets, we require regional expenditure weights. For consistency, these should come from HFCE after balancing within the national accounts. But the information is not available at this level, and therefore we approximate it using the LCF survey source. LCF data points are fuzzy, especially for households residing near regional boundaries, since the expenditure is not certainly in the region of residence; and this applies more generally for non-geographic expenditure. An alternative approach would be to obtain expenditure weights from businesses, which could be better localised, but this information does not have sufficient commodity detail for use in constructing a CPI.

The final component is the prices, and these should be operative in the region of interest. Price quotes are labelled with region identifiers, but there is a need to deal appropriately with central and nongeographic prices.

2.2. Data Sources

To develop an experimental regional CPI we use price quote data for the prices and LCF survey data for the expenditure. Monthly price quote and item index data for the UK are available from the ONS website from January 2010 (ONS 2020). The price quote data provide prices for items (goods or services) with corresponding information about the shop type, region, validity and stratum weight. Not all prices are represented in the price quote data, because many have nationally defined pricing and are collected centrally – approximately 45% of the weight of the basket is comprised of these centrally collected items. These central items are reported in the item index data sets, which report the indices and weights for the national CPI, but do not include a regional breakdown. For the regional CPI, the price quote data will need to be partitioned by region to calculate the item indices for each region, and then national level indices used for those collected centrally.

LCF survey data were obtained for the years 2008–2014 (Department for Environment, Food and Rural Affairs and Office for National Statistics 2019) for estimating expenditure, which will contribute to regional CPIs for 2010–2016 (because the CPI is a Lowe index with expenditure data from an earlier period than the reference period). The household sample sizes are shown in Table 1, and vary between regions and across years. Increases to the sample size in Northern Ireland (from 2016/7) and Scotland (from 2018) have more recently been implemented. The LCF survey data provide expenditure on products purchased by each sampled household, classified according to the COICOP classification at the COICOP-plus level. An example of the COICOP classification is shown in Table 2, which also shows the labels for the different levels, including item level, which sits below the formal COICOP hierarchy. Items are chosen by the ONS to be representative within a COICOP5 category and it is item prices that are reported in the price quote data.

We use the LCF survey data and accompanying household weights to estimate the mean (or equivalently the total) household expenditure by COICOP4 level in each of the twelve regions of the UK. For the national weights, the LCF survey data is one of multiple sources used to estimate expenditure. However, for the regional expenditure weights, we calculate directly from the LCF survey data without other sources, which are only accessible within the ONS. Relying on just one source for the expenditure weights is expected to adversely affect the reliability of the regional CPI, but it remains to be seen to what extent.

Table 1. Household sample size for LCF data, 2008 to 2014.

	2008	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**Mean**
North East	235	236	258	283	262	251	255	**254.3**
North West	592	582	596	647	623	585	588	**601.9**
Yorkshire and the Humber	491	484	485	521	521	462	459	**489.0**
East Midlands	405	393	413	455	425	424	440	**422.1**
West Midlands	469	527	470	526	513	526	470	**500.1**
East	532	499	515	543	563	497	498	**521.0**
London	472	464	476	536	490	480	407	**475.0**
South East	806	701	679	761	783	681	740	**735.9**
South West	502	518	495	507	493	429	468	**487.4**
Wales	265	272	261	251	266	246	222	**254.7**
Scotland	500	544	468	500	483	412	434	**477.3**
Northern Ireland	574	602	147	161	171	151	152	**279.7**

Table 2. *Example COICOP classification for UK national CPI framework*

Level	COICOP2	COICOP3	COICOP4	COICOP5	COICOP-plus	Item
Label	Division	Group	Class	Expenditure code	Category	Item
Example	Food and non-alcoholic beverages	Food	Bread and cereals	Bread	Buns, crispbread and biscuits	White sliced loaf branded 750g
Code	01	1.1	1.1.1	1.1.1.2	1.1.1.2.2	–
Number*	12	47	85	303	367	731

*Numbers tend to change over time, and the UK aggregates some groups as well, so these should be taken as approximate.

3. Constructing the Experimental Regional CPI

3.1. Regional CPI Price Aggregation

Ideally, the methods used to derive a regional CPI should be kept as close as possible to the national CPI. For full details of how the national CPI is calculated, see ONS (2019). Figure 1a outlines the process – elementary aggregates are calculated as geometric means of the relative prices weighted at the shop level to reflect market share of chain shops. The elementary aggregates are then aggregated into item indices using arithmetic means weighted at the stratum level, which capture variations by region and shop type (independent vs multiple). The item indices are then weighted together in proportion to the national consumption of the item, derived from the National Accounts expenditure data, LCF survey data, market research data and other sources including administrative data. Weighted arithmetic means are then calculated at the class level and then finally, these class indices are aggregated using class expenditure weights to produce the national CPI.

In adapting the methods of the national CPI to the regional level, we use the conceptual framework for regional price indices (Subsection 2.1) and follow these concepts and the national methods wherever possible.

The first step in constructing an experimental regional CPI is to collate the publicly available price quote and item indices data. Of the 713 items with non-zero weights listed in the 2016 item indices dataset, 548 items were available in the price quote dataset, so regional versions could be calculated. This leaves 165 items (23.1%) which are collected nationally, and will therefore not contribute to differences in prices between regions (though they may have different effects in different regions if they are weighted differently). For 2016, the 713 items account for 98.3% of the total weight, and the remaining 1.7% are not represented to prevent individual retailers from being identified. We ignore the undisclosed weights, and rescale the weights accordingly. The 548 items in the price quote dataset made up 53.9% of the weight. Similar percentages were observed in the other years prior to 2016.

Just as for the national CPI, the stratum indices can be calculated for each region using the weighted geometric means of the price relatives. We use the geometric mean (Jevons

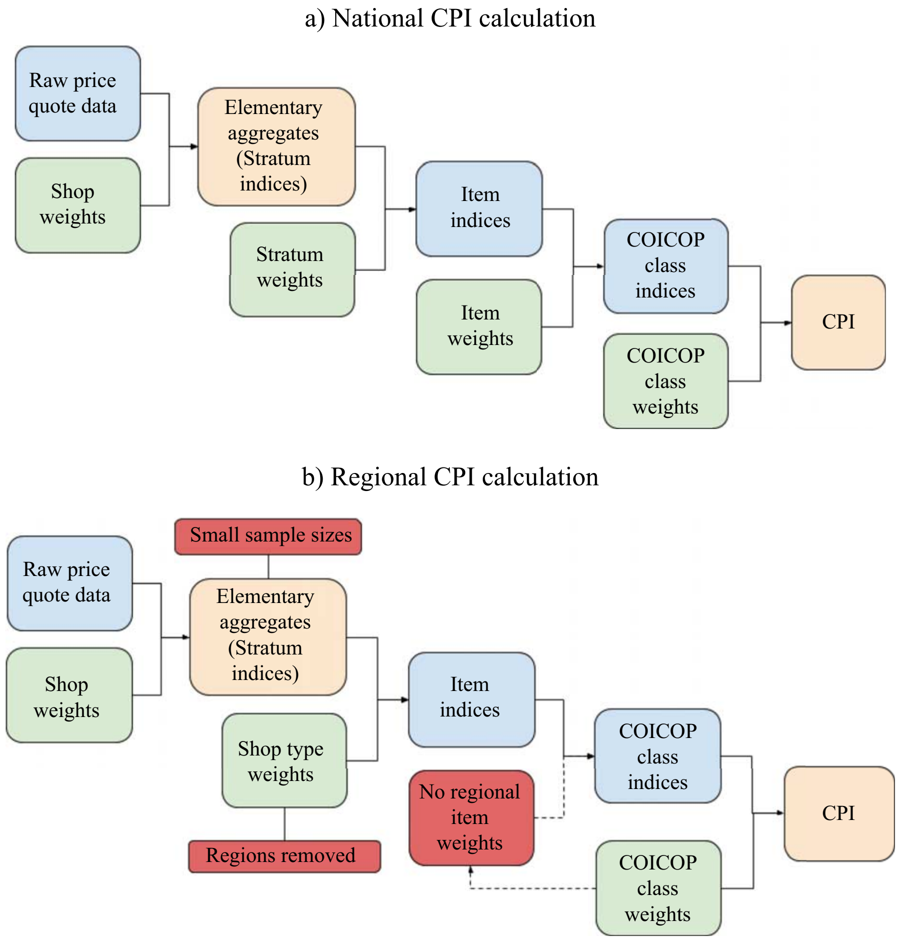


Fig. 1. Calculation of the national and regional CPI.

elementary aggregate formula) for all items, though in the national CPI the ratio of averages (Dutot formula) is sometimes used (ONS 2019). There are large numbers of strata which have extremely small sample sizes, including 5.4% with only one price measurement. The majority of the strata with small sample sizes are for independent shop prices. The geometric mean of very small samples is not very reliable, highlighting the primary limitation in constructing the regional CPI. Although it is inadvisable, for this experimental regional CPI the small sample sizes are treated as if they are satisfactory and the regional elementary aggregates are used.

With the prices aggregated into the stratum indices, the item indices can then be calculated by taking the arithmetic mean of these elementary aggregates weighted with the stratum weights. However, the stratum weights, which adjust for region and shop type in the national index, must not adjust for region here and reduce to shop type weights. Once the item indices are calculated for each region the available national item indices not represented in the price quote data can then be added to give the full set of item indices required to calculate COICOP class indices (except for the 1.7% excluded).

The next step is to aggregate item indices to give regional COICOP class indices. However, the item weights cannot be determined at the regional level, because the publicly available version of the LCF does not contain this detailed classification information (in principle the ONS could make this calculation on the full LCF data set). To overcome this problem we use a non-standard approach using the national (subscript n) item weights w_{nk} and class weights w_{nc} to first get the national proportions of item k within class c. We then multiply these national proportions by the estimated regional (subscript r) class weight w_{rc}, which gives approximations of regional item weights $\hat{w}_{rk}$:

$$\hat{w}_{rk} = w_{rc}\frac{w_{nk}}{w_{nc}}.$$

This ensures that the item weights sum up to the class weight for each region. Note that w_{rc} can be estimated using LCF data, as described in the following section.

Finally, the regional item indices and weights are used to derive the COICOP class indices and also the unchained regional CPI for each region. This process was replicated for all available years where price quote data and LCF data were readily available. The regional CPI series were then chained together using the same approach as the national CPI (ONS 2014) and indexed with January 2010 set to 100 for all regions.

The limitations encountered when constructing the regional CPI fall broadly into two categories –those due to small sample sizes, and those due to only the LCF survey being used for expenditure weights. These lead to the adjusted regional CPI framework shown in Figure 1b. These limitations require us to approximate the target concept and therefore render the regional CPIs less reliable than the national CPI. The extent of this unreliability is assessed by constructing experimental regional CPIs.

3.2. *Regional CPI Expenditure Weight Estimation*

The LCF survey data provide expenditures for a sample of households in each of the twelve regions of the UK. To get expenditure weights for the regional CPI the expenditures should be aggregated within a certain COICOP level and region. We found that the COICOP class level (COICOP4) was the most suitable, as more detailed levels had too many zero expenditures and the COICOP group level was not specific enough. Even at the class level, there were still some classes with few or no expenditures recorded at the region level, so we first inspect how well each class was represented in the LCF survey.

We use the term ‘observation’ here specifically to mean any household with non-zero expenditure recorded from a given COICOP class. Zero expenditures are still used in estimation, but for simplicity, we do not refer to these as observations. At COICOP class level, there can also be multiple observations (of more detailed expenditures) from the same household. Figure 2 shows the mean number of observations in the North East region across the available years. The North East is used as it generally has the smallest household sample in the LCF survey. A reference line at 30 observations is included in the plot, which shows that there are many classes with fewer than 30 observations. At the COICOP class level, we can see that Division 01 ‘Food and non-alcoholic beverages’ is well represented, as well as classes 5.6.1 ‘Non-durable household goods’, 11.1.1 ‘Restaurants and cafes’ and 12.1.2 ‘Appliances and products for personal care’. On the other hand, five out of 85 classes (5.9%) have no representation in the LCF data for the

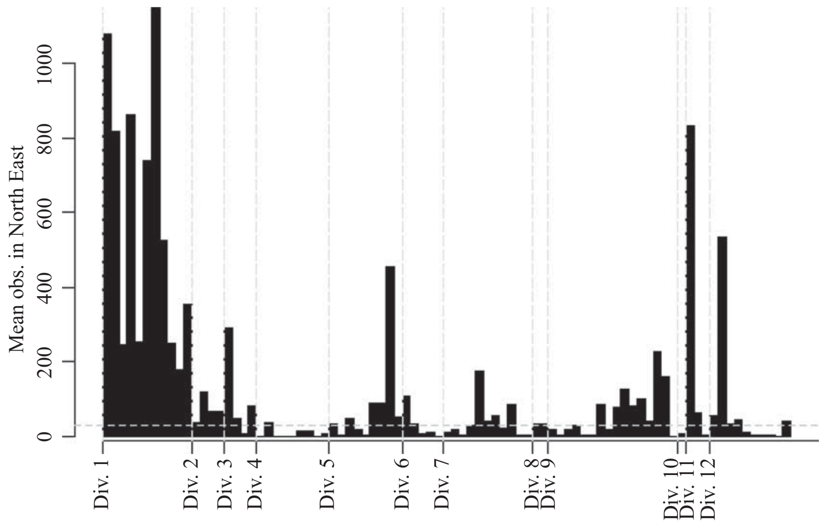


Fig. 2. North East region – mean annual number of observations from 2008–2014 for each COICOP class with a reference line at 30 observations.

available years, including water supply and sewerage, household repair services, hospital services and package holidays. These will all have zero weights in the provisional regional CPI. This is not just an issue for the North East region but for all regions, which are all reasonably comparable. This is a substantial limitation and can only be overcome through the use of additional data sources that are not publicly available at the regional level.

Although the number of observations is an important consideration, the actual expenditure must also be considered. The problem of having so few observations will be compounded when there is higher expenditure because it adds more weight to the regional CPI. The North East mean expenditures for COICOP classes are shown in Figure 3, with a

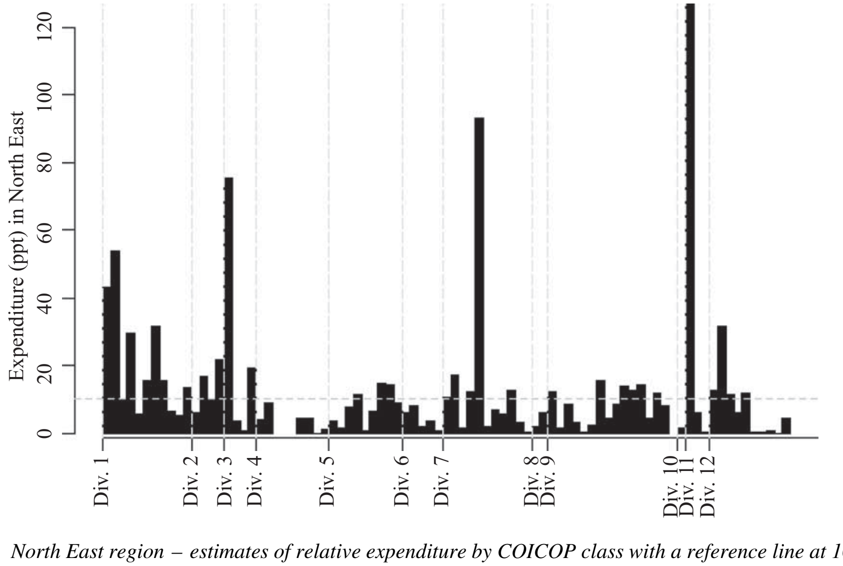


Fig. 3. North East region – estimates of relative expenditure by COICOP class with a reference line at 10 ppt.

reference line added at 10 ppt (1%). There is a lot of variation between classes, with five that make up more than 40 ppt and the remainder mostly under 10 ppt. The five largest classes, in order, are 11.1.1 'Restaurants and cafes', 7.2.2. 'Fuels and lubricants', 3.1.2. 'Garments', 1.1.2. 'Meat' and 1.1.1. 'Bread and cereals'. Note that the other regions also have highly variable relative expenditure across classes, and generally share the same largest classes.

To get direct estimates of the expenditure weights the following calculations are made. Let y_{ijct} denote the total household expenditure in pounds for COICOP class c in household j in region i and year t, and w_{ijct} be the provided household survey weight. Suppose that n_{it} is the number of sampled households in region i and year t, then the direct estimate of the mean household expenditure θ_{ict} is:

$$\hat{\theta}_{ict}^{direct} = \frac{\sum_{j=1}^{n_{it}} y_{ijct} w_{ijct}}{\sum_{j=1}^{n_{it}} w_{ijct}}. \tag{1}$$

Then the relative weights in ppt can be calculated using:

$$\hat{w}_{ict}^{direct} = 1000 \frac{\hat{\theta}_{ict}^{direct}}{\sum_c \hat{\theta}_{ict}^{direct}}. \tag{2}$$

These weights can then be used to generate the regional CPI series for 2010–2016, which is shown for the direct estimates in Figure 4b. This series can be compared to the series calculated with the national weights (but regional prices) in Figure 4a. This comparison makes it clear that the differences between regions and the general volatility of the series come primarily from the expenditure weights rather than the prices.

Due to the small sample sizes and lack of additional data, it is expected that there would be instability of the regional CPI series over time. To quantify this variability over time, we propose measuring the standard deviation of the first differences (SDFD) of the regional CPI series, that is $\frac{1}{T-2}\sum_{t=2}^{T}\left(y_t - y_{t-1}\right)^2$.

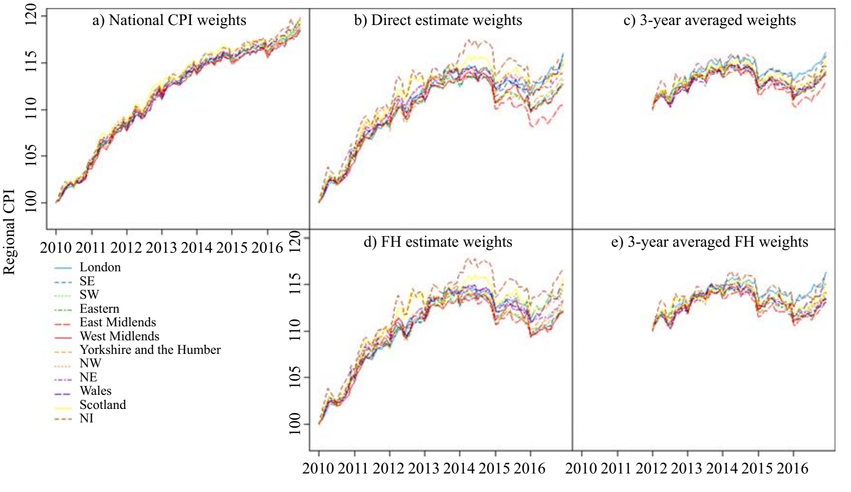


Fig. 4. Regional CPI series for different expenditure weights. Note that three-year averaged weights are rooted at 110 rather than 100.

The higher this SDFD measurement, the more variable the monthly changes in the index. As a comparison, we find that the national CPI between 2010 and 2016 has a SDFD of 0.29. This provides an approximate guide for an appropriate level of temporal variability. As shown in Figure 5 we find that the SDFD for the series with regional prices and national weights in Figure 4a ranges from 0.36 for Wales, to 0.46 for Northern Ireland. This corresponds to 1.24 and 1.60 times that of the national level of variability, from using regional prices alone. The SDFD for the regional prices and weights from the LCF data ranges from 0.48 for North East, to 0.87 for Northern Ireland, corresponding to 1.65 and 2.99 times that of the national SDFD respectively. In all but two regions, the majority of the additional variability derives from the expenditure weights rather than the prices. These two regions were the North East and South West. Averaged over regions, approximately 60% of the additional variability according to the SDFD is due to the expenditure weights. In the next section we explore smoothing and SAE methods to reduce this volatility.

4. Improving the Expenditure Weights

4.1. Smoothing and Small Area Estimation

Since direct estimation of the expenditure weights leads to substantial increases in temporal variability, we assess whether smoothing methods and small area estimation can make improvements. We test whether the three-year moving average of the expenditure weights substantially reduces the temporal instability. This serves to increase the sample sizes, as well as strengthen the temporal correlation. Regional CPIs using these smoothed weights were calculated and the series are shown in Figure 4c. Two years are removed due to the smoothing, and the resulting SDFDs range from 0.41 for the North East to 0.72 for Northern Ireland. This amounts to an 8–20% reduction in the variability. Hence there is evidence of improved stability compared to the one-year direct estimates, although only moderate improvement. This suggests that the three-year moving average approach does not eliminate all problems of volatility in the expenditure weights.

Small sample sizes are demonstrably a substantial limitation to developing reliable regional CPIs in the UK. Like smoothing, SAE can potentially improve the reliability,

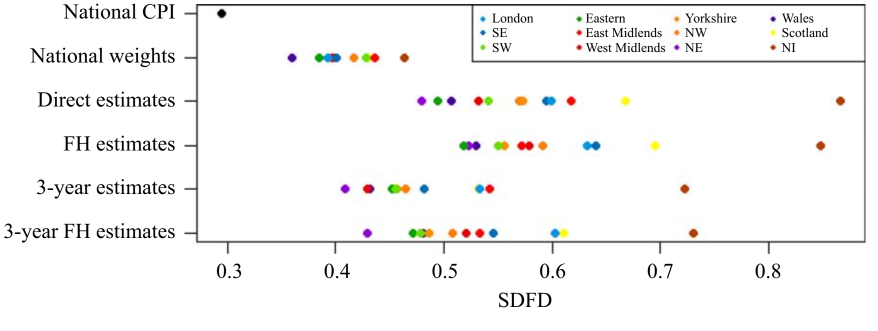


Fig. 5. Standard deviation of first differences of each regional CPI series and the national CPI.

as it utilises model-based methods to borrow strength from a wider or population-level data source to improve estimates for small domains. Although SAE methods can improve reliability by reducing the variance, this comes at the cost of introducing bias. Due to the high variance this seems a worthwhile trade-off in this case. For a general overview of SAE methodology, Pfeffermann (2013), Rao and Molina (2015) and Tzavidis et al. (2018) are highly recommended. SAE is most effective when a wider data source with high quality is available, containing strong predictors for the variables of interest within each region. In the case of the regional CPI, having region-level predictors of expenditure within COICOP classes will improve the precision of the estimates.

For the price quotes, it is difficult to utilise the beneficial aspects of SAE. This is because the sampling units are shops or prices. To effectively use SAE, comprehensive and informative data about the shop or price population would be required. This could include population numbers of shop types (independent and multiple) by region, number of supermarkets, and shop type by COICOP level. Alternative data sources from web-scraping or scanner data might also provide suitable predictors, although such sources typically have only partial coverage. However, such data sources were not available. Furthermore, the expenditure weights cause more of the variability than the price quotes, so SAE should be more effective at improving estimates of the weights.

SAE is much more feasible for the expenditure estimates, because the sampling units for expenditure data are households, about which plenty of national and regional data are available. The detailed national-level data for potential predictors (such as household types, total salary, total expenditure, head of household's age and the number of children) can be used in a model to get improved estimates of the expenditure for each class. We aim to use these small area models to give regional expenditure estimates with lower variances, and in particular look for them to be relatively smooth across years.

4.1.1. Fay-Herriot Models

The SAE method used for expenditure estimation was the Fay-Herriot (FH) model (Fay and Herriot 1979). The FH model is a commonly used region-level model for SAE (Rao and Molina 2015, sec. 4.2, chap. 6), and comprises of two stages. The first stage simply models the sampling variation of the direct estimate which was defined in Equation (1):

$$\hat{\theta}_{ict}^{direct} = \theta_{ict} + \varepsilon_{ict}$$

where the sampling errors ε_{ict} are assumed to be independent and normally distributed with $\varepsilon_{ict} \sim N(0, \sigma^2_{\varepsilon_{ict}})$. From the total household expenditure and the sampling weights we get regional direct estimators for each COICOP class and year. However, the true variance of this estimator cannot be determined from only sample data. For this reason, the variance $\sigma^2_{\varepsilon_{ict}}$ must be estimated. Possible options are the Poisson approximation or a bootstrap procedure. We used the bootstrap method of the laeken package (Alfons and Templ 2013) in R (R Core Team 2019). For each COICOP class within every year and each region i a bootstrap sample was drawn from the sample data with replacement. Using the household weights and expenditures within each bootstrap sample, the corresponding direct estimator was obtained. This was repeated many times and the variance $\sigma^2_{\varepsilon_{ict}}$ was then estimated by taking the sample variance of all bootstrap direct estimates.

The second stage of the FH model is to fit a linear model which can be used to predict θ_{ict}:

$$\theta_{ict} = x_{ict}^T \beta + u_{ict}$$

where x_{ict}^T denotes the region-level covariates for year t, β denotes the regression parameter vector, and u_{ict} represents the random effects which are assumed to be $u_{ict} \sim N(0, \sigma^2_{u_{ct}})$. The combination of the two stages of modelling leads to the combined FH model:

$$\hat{\theta}_{ict}^{direct} = x_{ict}^T \beta + u_{ict} + \varepsilon_{ict}.$$

The estimates $(\hat{\beta}, \hat{u}_{ict}, \hat{\sigma}^2_{u_{ct}})$ of these unknown parameters can be estimated using a standard linear random-effects model. From this the FH estimates can be derived as:

$$\begin{aligned} \hat{\theta}_{ict}^{FH} &= x_{ict}^T \hat{\beta} + \hat{u}_{ict} \\ &= \gamma_{ict} \hat{\theta}_{ict}^{direct} + (1 - \gamma_{ict}) x_{ict}^T \hat{\beta} \end{aligned}$$

where $\gamma_{ict} = \hat{\sigma}^2_{u_{ct}} (\hat{\sigma}^2_{u_{ct}} + \sigma^2_{\varepsilon_{ict}})^{-1}$. In cases when an area has zero observations the estimator simply becomes: $\hat{\theta}_{ict}^{FH} = x_{ict}^T \hat{\beta}$. Estimates of the precision of the FH estimates can be made using the mean squared error (MSE) which is estimated using restricted maximum likelihood (REML). Further details can be found in Rao and Molina (2015, chap. 6). Once the estimates are calculated, the weights can be created using the same adjustment as in Equation (2).

One of the improvements of SAE comes from borrowing strength from other areas, using the association between the region-level covariates and the regional expenditure. Region-level associations strengthen each region's estimate by using the region-level covariates, which are assumed to be informative. In the next section, we describe how the covariates were selected for use in the FH models.

A challenge for the FH model in this situation is due to having only twelve regions. With so few regions, model assumptions are more difficult to assess, covariate associations are more likely to occur by chance and the number of covariates that can be included in the model is restricted because there are few degrees of freedom. Another notable limitation in the application of the FH model to the expenditure weights is that some classes have no expenditure in the region. This becomes problematic, as zero expenditure for more than a few regions will lead to violations of the normality assumptions. For some classes, a zero-inflated model (Pfeffermann et al. 2008; Chandra and Sud 2012) may be beneficial, which we leave for future research.

The same FH model was also applied to the three-year averaged data to assess the collective impact of both smoothing and SAE on the regional CPI series.

4.1.2. Covariate Variable Selection

The LCF survey provides a large number of variables at the regional level which can be used to estimate expenditure. These variables relate to socioeconomic status, household composition and household features, for example, tenure type, number of adults, weekly income. These variables were aggregated to the region level. For a FH model to be effective at estimating expenditure, the region-level covariates should be predictive of the expenditure of the COICOP classes. The challenge is to select the best combination of variables that ensures the relationship is predictive but not over-fitted to the sampled data. This over-fitting is especially a concern since there are only twelve regions, and hence twelve points from which to fit a model. Furthermore, the covariates should not have high multicollinearity, as this can greatly exaggerate over-fitting. Over-fitting will lead to small area estimates with under-estimated precision, as well as overly biased point estimates. Hence the explanatory variables must be selected carefully.

The variables were chosen based on associations with the class expenditure for a pooled data set across all years from 2010 to 2016. This ensures that the covariate is predictive across all years rather than a certain year. This will also ensure consistency across time. A forward selection approach using AIC was used to select the variables for each COICOP class, with at most five selected variables. We made five the maximum since any more would be superfluous when estimating only twelve regions. At each step, the multicollinearity was assessed using the variance inflation factor (VIF). If the VIF was greater than ten then no more variables were added. This ensured that a minimal number of variables were selected and that none of the variables were highly collinear.

4.1.3. Model Assessment

A FH model relies on a strong level of prediction with explanatory variables. Figure 6 shows the R^2 values of the fitted linear models averaged over the years for each COICOP class. It shows that in Divisions 01 and 02, which include food and alcoholic beverages, the R^2 is generally high, but some classes have low R^2 values. For example, COICOP class 12.6.2 ‘Other financial services’ with a mean R^2 of just 0.03, which will be unlikely to improve the expenditure estimate in the FH model. Classes with higher R^2 may be a better reflection of the general economic differences between regions.

With the FH models fitted, it remains to be seen what the effect on the stability of the expenditure estimates is when we use the model predictions in place of the direct estimates.

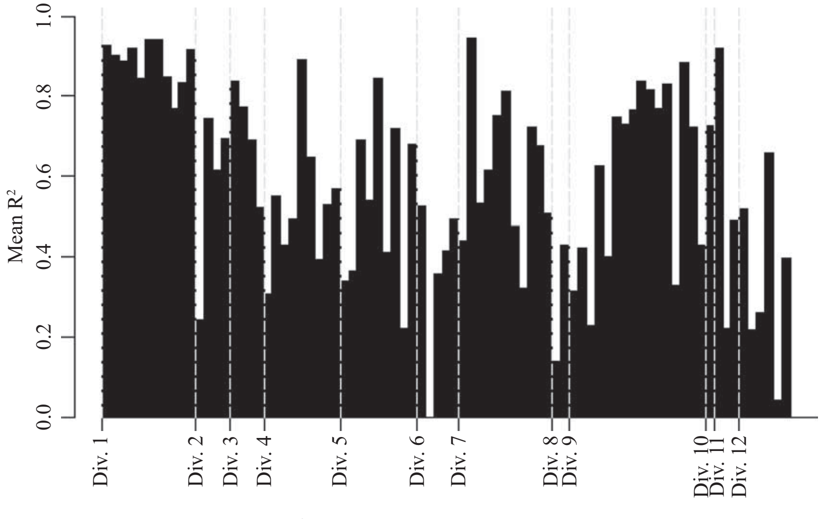


Fig. 6. Mean R^2 over 2008-2014 for each COICOP class.

As part of the model assessment, the assumptions of the models were checked, particularly the normality assumption of $\hat{u}_{ict} = x_{ic}^T\hat{\beta} - \hat{\theta}_{ict}^{FH}$. A Shapiro-Wilk test was used, and based on this test, there was evidence to reject the normal distribution for many classes for each year. Across years, between 26 and 43 (out of 80) classes were not rejected, with a median of 36. However, as the focus is on improving the temporal stability of the estimates we include all the estimates in calculating the experimental index, even if there is strong evidence to reject the normality assumptions. Although the FH model is somewhat robust to these violations (Lahiri and Rao 1995), further development of the models may provide a better basis for such estimation.

4.2. *Assessment of Fay-Herriot Estimates*

To assess the effect that FH estimation has on the expenditure we first measure how different the FH estimates are from the direct estimates. Figure 7 again uses the North East region as an example, and shows the percent difference between the FH and direct estimate, averaged over the seven years. This reveals up to a 30% difference in the estimates with many COICOP classes showing non-trivial relative differences. Clearly, FH estimation has some effect, but it is unclear what effect this is. Some classes have no percent difference because FH estimates could not be calculated, where too many regions had zero reported expenditure. In total, FH estimates could not be calculated for 16 of the 85 COICOP classes (18.8%). In these cases the direct estimates are used for the weights.

We expect that expenditure patterns are in reality rather stable between adjacent years, and change slowly as consumer spending is influenced by changes in products and their availability, particularly at higher levels of aggregation in COICOP. Therefore we judge that the more stable the estimates are over 2008–2014, the better the estimates are and the more stable the regional CPI will be. This is because the instability is likely caused by small sample sizes. Note that some expenditure patterns may vary substantially over time, so an expert opinion on what level of variability is realistic would need to be considered too.

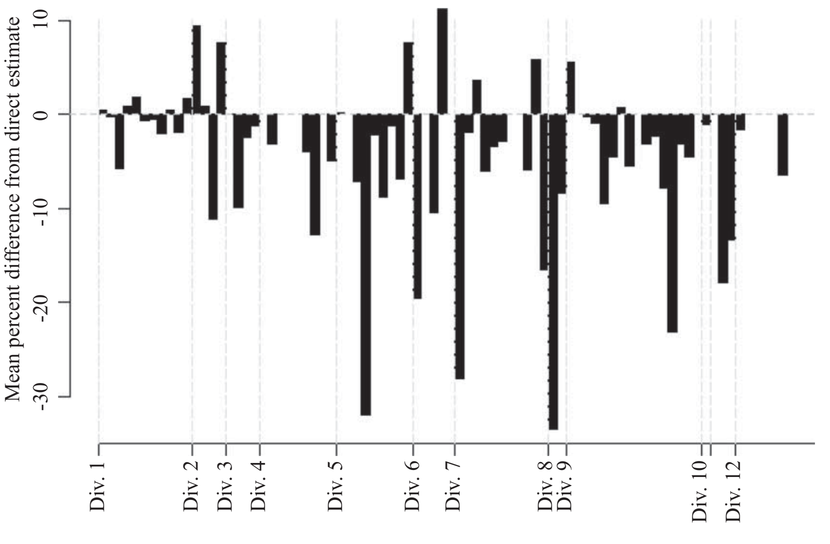


Fig. 7. North East region – mean percent difference between FH estimate and direct estimate.

To measure this stability over time we measure the variability of the yearly estimates of expenditure; this will include a small element of real change in expenditure patterns, but we expect that this is much smaller than the random variation we are trying to smooth using SAE. Typically the standard deviation or variance is used to measure variability, however this will not be appropriate in this case, because the variance is greater for COICOP classes with higher expenditure. To accommodate this, we use the coefficient of variation (CV) which is the standard deviation divided by the mean. This ensures the measure is standardised by the amount of expenditure, hence making the metric comparable across all COICOP classes. This 'temporal' CV is calculated using:

$$CV_{ic} = \frac{SD_t\left(\hat{\theta}_{ict}\right)}{Mean_t\left(\hat{\theta}_{ict}\right)}$$

where SD_t and $Mean_t$ are the standard deviation and mean of the estimates across years t respectively. We use CV_{ic} to compare the stability of the FH estimates compared to the direct estimates. The lower this temporal CV, the more stable the estimates over time.

Figure 8 compares this temporal CV for the direct and FH estimates for all twelve regions. A positive difference in temporal CV indicates that the direct estimate is less

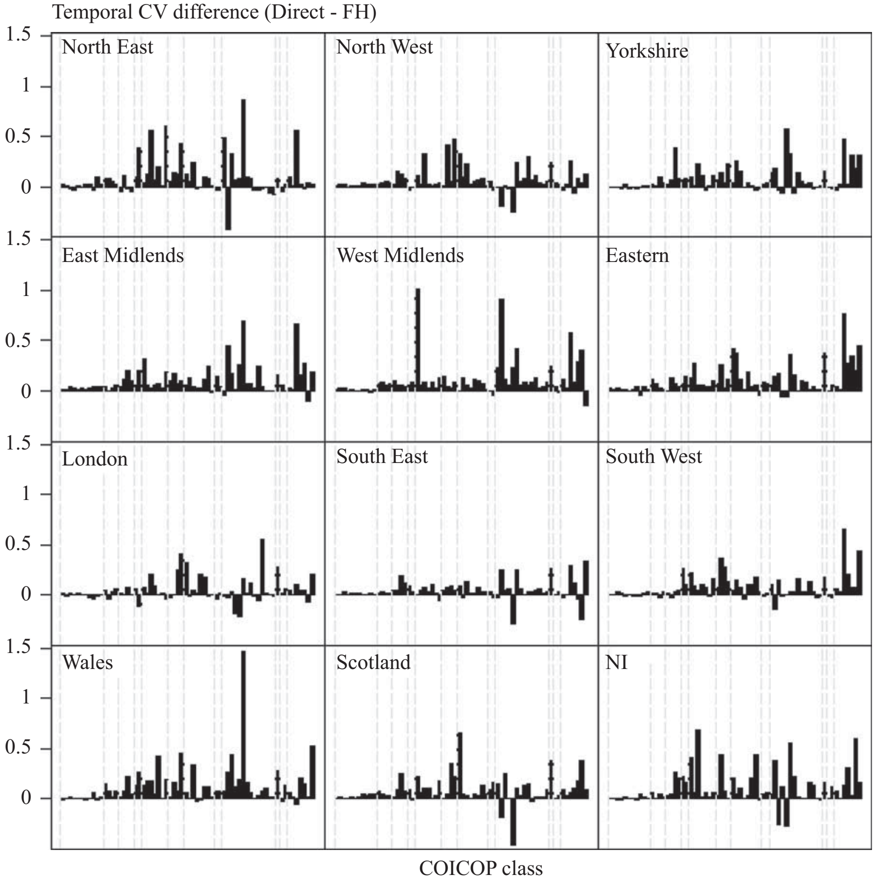


Fig. 8. All regions – difference in temporal CV between direct and FH estimates.

stable, which is the case for the vast majority of classes and regions. This suggests that FH estimation is generally improving the stability compared to the direct estimates. Notably, in well-represented classes like Division 01 the differences are very small.

Figure 9 displays the North East region estimates of γ_{ict} for each class averaged over all seven years. The higher the value of γ_{ict} the more the FH estimate utilises the data directly as opposed to the model-based component. There is a lot of variation between COICOP classes, ranging from 0 to 0.7, so there is no clear trend about what types of classes have higher values of γ_{ict}. Again, other regions showed similar patterns.

Table 3 reports the ten COICOP classes that have the greatest improvements in temporal CV due to FH estimation. These ten classes have generally few observations, ranging from 6 to 61. Interestingly, the R^2 values are not particularly high, which suggests that FH estimation does not require strongly predictive explanatory variables to provide additional stability to the estimates. Classes 6.1.2/3 'Other medical and therapeutic equipment' and 9.2.1/2 'Major durables for in/outdoor recreation' each have relatively large ppt values, which shows that it is not just the trivially small classes (such as 12.7 'Other services') that show improvements. The last column in Table 3 shows the mean values of γ_{ict} which range between 0.10 and 0.38 which is quite typical of all classes.

Figure 10 shows more broadly the effect of sample size on improved stability due to FH estimation. A smoothing spline has also been added to give an idea of the average effect for varying numbers of observations. The results show that for COICOP classes with relatively few observations, the benefit of the FH estimation is generally better although highly variable for all regions. We also see how the COICOP classes with many observations show negligible benefit from FH estimation. The improvement of FH estimation becomes reasonably small after approximately 100 household observations.

In combining all these results, we consider four attributes of the COICOP classes which relate to their suitability for SAE. These four attributes are:

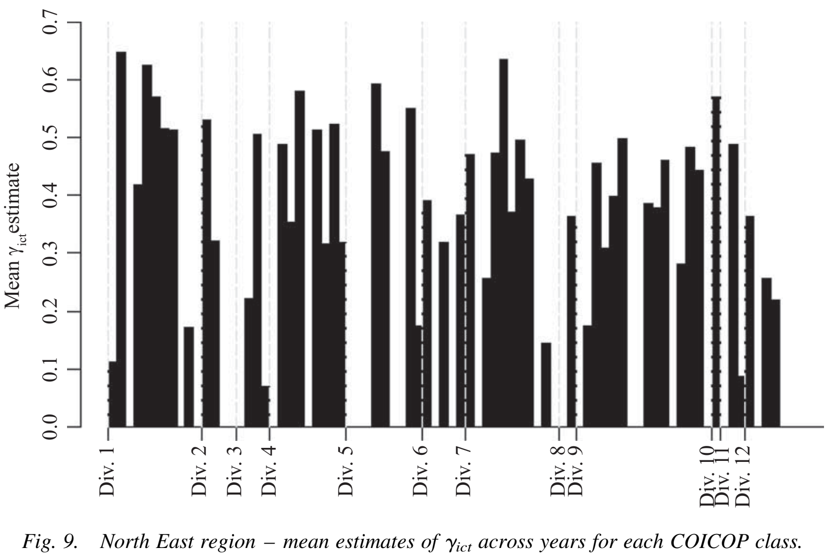


Fig. 9. North East region – mean estimates of γ_{ict} across years for each COICOP class.

Table 3. Top ten COICOP classes with the most improvement in stability due to FH estimation.

COICOP class	Temporal CV difference	Mean number of observations	Mean ppt	Mean R^2	Mean γ_{ict}
9.2.1/2	0.46	6.5	9.11	0.23	0.15
12.3.1	0.37	58.2	5.64	0.22	0.10
10	0.23	11.2	5.23	0.33	0.16
5.1.1	0.23	61.0	3.01	0.39	0.26
6.2.2	0.23	20.6	6.23	0.53	0.32
12.7	0.23	72.1	0.46	0.40	0.13
5.3.1	0.22	39.4	4.40	0.34	0.33
7.1.1A	0.20	16.6	2.91	0.36	0.38
6.1.2/3	0.19	59.7	15.03	0.22	0.21
3.1.4	0.15	15.2	1.20	0.69	0.32

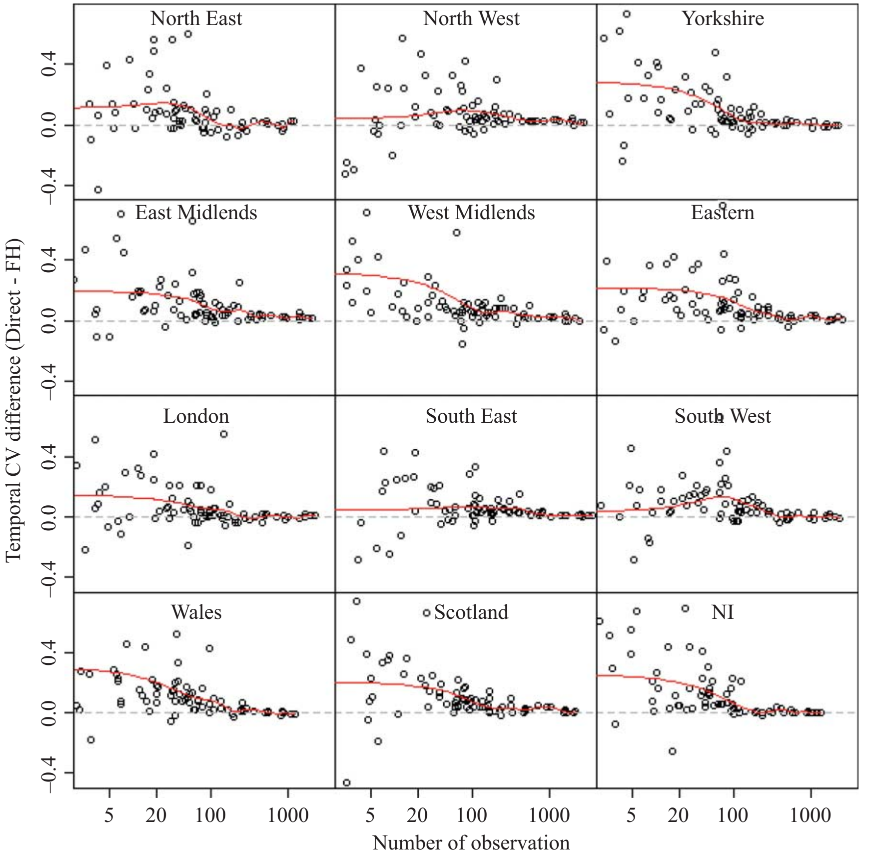


Fig. 10. Temporal CV difference by sample size within each region, with a trend line added. Each point is a COICOP class.

1. Observations recorded in all regions for all years – 63 out of 85 classes (74.1%) meet this criterion. This means at least one reported price within a class for each region and year,

2. The number of observations not being so large that SAE remains useful, chosen to be all COICOP classes where all regions have at most 100 household observations – 63 out of 85 classes (74.1%),
3. A non-negligible expenditure share in ppt, chosen to be the COICOP classes which have at least 0.5 ppt share in all regions and years (in line with the conceptual framework in section 2.1) – 58 out of 85 classes (68.2%), and
4. A non-negligible (> 0.03) decrease in temporal CV for at least one region when using FH estimation – 60 out of 85 classes (70.6%).

These four attributes are not mutually exclusive, so the total number of COICOP classes which possess all four attributes is 36 out of 85 (42.4%). Hence, based on these criteria, 42.4% of COICOP classes have distinguishable improvements in the stability of the expenditure estimates through the use of SAE.

The regional CPI series with the FH estimates for each year and also FH estimates with weights based on the three-year average are shown in Figure 4d and Figure 4e. In comparison to the series with the direct weights, all regions but Northern Ireland appear much closer together. This is expected given that SAE adds national-level information, making the estimates closer to the national value. Generally, the observable differences between series with direct and FH expenditure estimates are not large. To assess in more detail we again calculate the SDFD for the two FH-based series, these are displayed for each region in Figure 5, as well as for the other regional CPI series. The SDFD metric shows that FH estimation does not generally decrease the variability over time compared to the direct estimates. There is a slight increase in the SDFD for most regions. So it appears that while SAE appears to make inter-regional differences smaller in the series, it does so without improving the temporal stability, and appears to mildly increase variation over time.

The results show that smoothing and SAE using FH models for individual classes can improve the stability of the expenditure weights in some ways, but the smoothing appears to have a greater effect. The classes that benefit the most tended to have fewer than 100 observations, but also with enough observations for a good model to be fitted. Although FH models do ensure that the regional indices stay closer together, it does not provide reduced temporal variability for the regional CPIs.

5. Discussion

5.1. Regional Indices and Data Sources

We have shown that it is possible to construct a regional CPI series from the available data sources. Although these experimental regional CPIs are somewhat useful, the reliability of specific components of the data and procedures is generally low. Small sample sizes and absence of data sources create increased irregularities and variability over time. We show that the source of this variability is generally from the expenditure weights more than the price quotes.

Some further exploration of data sources is possible. ONS has produced experimental regional HFCE estimates, which are balanced through the national accounts (ONS 2018). These are a less detailed level of the COICOP classification than is needed for the CPI, but could be integrated into a framework of weight calculation. Alternative data sources such as web-scraped prices and scanner data may also provide useful regional data sources or predictors for either or both of the prices and weights. These could improve the expenditure weights and make the methods more similar to the national CPI. Incorporating regional estimates of owner occupiers' housing costs (OOH) and council tax to produce a regional CPIH would also be useful, and is likely to show greater regional differences, because there are large disparities in housing costs between regions.

It would also be worthwhile exploring whether it is possible to calculate a version of the regional item weights which were not available for our analysis. Marchetti and Secondi (2017) go one step further by adjusting the regional expenditure estimates for differences in PPPs, which is potentially an important extension to the conceptual framework of Subsection 2.1.

5.2. Smoothing, SAE and Bias-Variance Trade-Off

Smoothing methods like the three-year moving average were shown to reduce the variability moderately. Although FH estimation improved temporal stability of the expenditure weights there was no evidence that this reduced the temporal variability of the regional CPI series. We conclude that smoothing and SAE do generally make improvements to the series and the stability of the expenditure weight estimates, and are plausible options for improving the reliability of the regional CPI. Fengki et al. (2020) have applied FH models to estimate regional CPIs in Indonesia (using the CPI as a target, because there are city CPIs available for modelling), with some improvement over direct estimation, but in their case too, further research is needed to deal with data deficiencies.

Smoothing and SAE both involve a bias-variance trade-off, so using them in a regional CPI results in biased estimates of the regional price indices. However, since the main purpose of the index is to provide regional estimates of *inflation*, a bias may be acceptable as long as it evolves slowly, and as long as it accompanies a substantial reduction in the variability of the index. However, for long-run comparisons the bias may be important, particularly where it has different effects in regions of different size (as suggested by the estimates of γ_{ict} (Table 3, Figure 9)). A periodic benchmarking may therefore be needed to correct the path of the index if measures of long-term change are important.

It would be useful to have quality targets that a regional CPI or CPIH should aim to achieve, particularly an estimate of the variance to identify an acceptable level of volatility. How to measure the variance of a CPI is an open question, (see Smith 2021; O'Donoghue 2017; Zimmermann et al. 2020, sec. 2.a.6) for some initial assessments at national level for the UK. This work should be extended to regional measures.

5.3. Model Extensions

The FH model used does not account for the zero-inflated, longitudinal or compositional properties of the data, and these could be addressed using more advanced methods. Methods have been proposed for small area estimation with zero-inflated data (Pfeffermann et al. 2008; Chandra and Sud 2012), and it would be interesting to explore these. In particular, accounting properly for the observed zeroes in expenditure might allow modelling at a more detailed level in the COICOP classification.

Extension of the FH model to account for correlation over time should improve the small area estimates. Esteban et al. (2016) provide a comprehensive review of the literature on temporal extensions to the FH model that – among many variations – includes a model with an autoregressive structure in the sampling errors (Choudry and Rao 1989) and a model with an autoregressive structure in the random effects (Esteban et al. 2011). A further extension of the FH model simultaneously accounts for spatial and temporal effects (Marhuenda et al. 2013).

A further line of investigation is to treat the estimation of weights as the estimate of a composition. Scealy and Welsh (2017) have developed an approach to estimate expenditure proportions as compositions within small domains of a population, different from the FH models, and Esteban et al. (2020) provide a FH model for compositions. Applying similar methods may give regional expenditure weights for the regional CPI with smaller variances and less change from year to year which would lead to fewer irregularities and smoother indices.

5.4. Conclusions

The work reported here is a stepping-stone to the development of a regional CPI in the UK. We have provided a clear framework, and highlighted the deficiencies and issues faced in calculating a regional temporal CPI with available data sources. Smoothing and small area estimation offer reasonable reductions in the volatility of the weights, and could be used in a regional CPI, if the addition of other data sources does not provide the necessary level of reliability. Assessing whether the bias induced by these methods in the index affects their suitability remains an open question.

6. References

Alfons, A., and M. Templ. 2013. "Estimation of Social Exclusion Indicators from Complex Surveys: The R package laeken." *Journal of Statistical Software* 54, no. 15: 1–25. DOI: www.dx.doi.org/10.18637/jss.v054.i15.

Baran, D., and J. O'Donoghue. 2002. "Price levels in 2000 for London and the regions compared with the national average." *Economic Trends* 578: 28–38. Available at: www.escoe.ac.uk/research/historical-data/etarticles/ (accessed April 2021).

BLS. 2018. *BLS handbook of methods: Chapter 17 The Consumer Price Index*. Washington DC. Available at: www.bls.gov/opub/hom/pdf/homch17.pdf (accessed April 2021).

Borooah, V.K., P.P.L. McGregor, P.M. McKee, and G.E. Mulholland. 1996. "Cost of living differences between the regions of the United Kingdom." In *New Inequalities. The Changing Distribution of Income and Wealth in the United Kingdom*, edited by J.Hills: 103–132. Cambridge: Cambridge University Press.

Brown, M., R. de Haas, and V. Sokolov. 2018. "Regional Inflation, Banking Integration, and Dollarization." *Review of Finance* 22 8(6): 2073–2108. DOI: www.doi.org/10.1093/rof/rfx021.

Chandra, H. and U.C. Sud. 2012. "Small area estimation for zero-inflated data." *Communications in Statistics–Simulation and Computation* 41: 632–643. DOI: www.dx.doi.org/10.1080/03610918.2011.598991.

Choudry, G.H., and J.N.K. Rao. 1989. "Small area estimation using models that combine time series and cross sectional data." In Proceedings of Statistics Canada Symposium on Analysis of Data in Time, edited by A.C. Singh, and P. Whitridge: 67–74. Ottawa: Statistics Canada.

Department for Environment, Food and Rural Affairs and Office for National Statistics. 2019. Living Costs and Food Survey, 2008-2014. [data collection]. 3rd Edition. UK Data Service. SN: 7992, www.doi.org/10.5255/UKDA-SN-7992-4 (DOI for 2014 only) and also SN: 6385, 6655, 6945, 7272, 7472, 7702).

Duran, H.E. 2016. "Inflation differentials across regions in Turkey." *South East European Journal of Economics and Business* 11(1): 7–17. DOI: www.doi.org/10.1515/jeb-2016-0001.

Esteban, M.D., M.J. Lombardía, E. López-Vizcaíno, D. Morales, and A. Pérez, A. 2020. "Small area estimation of proportions under area-level compositional mixed models." *TEST* 29: 793–818. DOI: www.doi.org/10.1007/s11749-019-00688-w.

Esteban, M.D., D. Morales, A. Pérez, and L. Santamaría. 2011. "Two area-level time models for estimating small area poverty indicators." *Journal of the Indian Society of Agricultural Statistics* 66: 75–89.

Esteban, M.D., D. Morales, and A. Pérez. 2016. *Analysis of Poverty Data by Small Area Estimation, First Edition*. Edited by Monica Pratesi. Chichester: John Wiley & Sons.

EU regulation 1998. EU regulation 1687/98 on HICP available at: (www.eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri = CELEX:31998R1687&from = EN) (accessed April 2021).

Fay, R., and R. Herriot. 1979. "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data." *Journal of the American Statistical Association* 74 (366): 269–277. DOI: www.doi.org/10.2307/2286322.

Fengki, A.O., K.A. Notodiputro, and K. Sadik. 2020. "Bisakah memperoleh statistik indeks harga konsumen tingkat provinsi di Indonesia dengan ketelitian yang lebih baik? (Can provincial level consumer price indices statistics in Indonesia be obtained with better accuracy?)." *Seminar Nasional Official Statistics* 1: 297–306. DOI: www.doi.org/10.34123/semnasoffstat.v2019i1.178.

Fenwick, D., and J. O'Donoghue. 2003. "Developing estimates of relative regional consumer price levels." *Economic Trends* 599: 72–83. Available at: www.webarchive.nationalarchives.gov.uk/20160120195958/www.ons.gov.uk/ons/rel/cpi/regional-consumer-price-levels/developing-estimates-of-relative-regional-consumer-price-levels/-developing-estimates-of-relative-regional-consumer-price-levels-.pdf (accessed April 2021).

Gajewski, P. 2017. "Sources of regional inflation in Poland." *Eastern European Economics* 55(3): 261–276. DOI: www.doi.org/10.1080/00128775.2017.1284572.

Hayes, P. 2005. "Estimating UK regional price indices, 1974–96." *Regional Studies* 39: 333–344. DOI: www.doi.org/10.1080/0034340050087216.

Kosfeld, R., H.-F. Eckey, and M. Schüßler. 2009. "Ökonometrische Messung regionaler Preisniveaus auf der Basis örtlich beschränkter Erhebungen". German Council for

Social and Economic Data (RatSWD) Research Notes: 33. Available at: www.hdl.handle.net/10419/43644 (accessed April 2021).

Lahiri, P., and J.N.K. Rao. 1995. "Robust estimation of mean squared error of small area estimators." *Journal of the American Statistical Association* 90, no. 430: 758–766. DOI: www.doi.org/10.1080/01621459.1995.10476570.

Marchetti S., and L. Secondi. 2017. "Estimates of Household Consumption Expenditure at Provincial Level in Italy by Using Small Area Estimation Methods: ''Real'' Comparisons Using Purchasing Power Parities." *Social Indicators Research* 131: 215–234. DOI: www.doi.org/10.1007/s11205-016-1230-8.

Marchetti S., G. Bertarelli, L. Biggeri, G. Giusti, M. Pratesi, and F. Schirripa-Spagnolo. 2019. "Small area poverty indicators adjusted using local price indexes." Italian Conference on Survey Methodology (ITACOSM), 5–7 June, 2019, Florence. Available at: www.centrodagum.it/wp-content/uploads/2019/07/Presentation_Pratesi.pdf (accessed April 2021).

Marhuenda, Y., I. Molina, and D. Morales. 2013. "Small area estimation with spatio-temporal Fay–Herriot models." *Computational Statistics and Data Analysis* 58: 308–325. DOI: www.doi.org/10.1016/j.csda.2012.09.002.

Nagayasu, J. 2011. "Heterogeneity and convergence of regional inflation (prices)." *Journal of Macroeconomics* 33: 711–723. DOI: www.doi.org/10.1016/j.jmacro.2011.07.002.

O'Donoghue, J. 2017. "The effect of variance in the weights on the CPI and RPI." *Survey Methodology Bulletin* 77: 1–27. Available at: www.ons.gov.uk/file?uri = /methodology/methodologicalpublications/generalmethodology/surveymethodologybulletin/surveymethodologybulletinno.77autumn2017a.pdf (accessed April 2021).

ONS. 2011. *UK Relative Regional Consumer Price levels for Goods and Services for* 2010. ONS, Newport. Available at: www.webarchive.nationalarchives.gov.uk/20160108054525/www.ons.gov.uk/ons/rel/cpi/regional-consumer-price-levels/2010/uk-relative-regional-consumer-price-levels-for-goods-and-services-for-2010.pdf (accessed April 2021).

ONS. 2014. *Consumer price indices technical manual.* Available at: www.ons.gov.uk/ons/rel/cpi/consumer-price-indices--technical-manual/2014/index.html (accessed April 2021).

ONS. 2018. *Development of regional household expenditure measures.* Office for National Statistics, Newport. Available at: www.ons.gov.uk/economy/regionalaccounts/grossdisposablehouseholdincome/articles/developmentofregionalhouseholdexpendituremeasures/latest (accessed April 2021).

ONS. 2019. *Consumer prices indices technical manual.* Available at: www.ons.gov.uk/economy/inflationandpriceindices/methodologies/consumerpricesindicestechnicalmanual2019 (accessed April 2021).

ONS. 2020. *Consumer price inflation item indices and price quotes.* Available at: www.ons.gov.uk/economy/inflationandpriceindices/datasets/consumerpriceindicescpiandretailpricesindexrpiitemindicesandpricequotes/ (accessed April 2021).

Pfeffermann, D. 2013. "New important developments in small area estimation." *Statistical Science* 28(1): 40–68. DOI: www.doi.org/10.1214/12-STS395.

Pfeffermann, D., B. Terryn, and F.A. Moura. 2008. "Small area estimation under a two-part random effects model with application to estimation of literacy in developing countries." *Survey Methodology,* 34: 235–249. Available at: www.eprints.soton.ac.uk/153903/1/Pfeffermann_Moura-Terryn.pdf (accessed April 2021).

Purwono, R., M.Z. Yasin, and M.K. Mubin. 2020. "Explaining regional inflation programmes in Indonesia: Does inflation rate converge?" *Economic Change and Restructuring* 53: 571–590. DOI: www.doi.org/10.1007/s10644-020-09264-x.

R Core Team. 2019. "R: A language and environment for statistical computing". *R Foundation for Statistical Computing*, Vienna, Austria. Available at: www.R-project.org (accessed April 2021).

Rao, J.N., and I. Molina. 2015. *Small area estimation (Second Edition).* Hoboken: John Wiley & Sons.

Rienzo, C. 2017. "Real wages, wage inequality and the regional cost-of-living in the UK." *Empirical Economics* 52: 1309–1335. DOI: www.doi.org/10.1007/s00181-016-1122-4.

RPI Advisory Committee. 1971. *Proposals for retail prices indices for regions*. H.M. Stationery Office, London.

Scealy, J. and Welsh, A. 2017. "A Directional Mixed Effects Model for Compositional Expenditure Data" *Journal of the American Statistical Association,* 112, 24–36. DOI: www.doi.org/10.1080/01621459.2016.1189336.

Smith, P.A. 2021. "Estimating sampling errors in Consumer Price Indices." *International Statistical Review*. DOI: www.doi.org/10.1111/insr.12438.

Statistisches Bundesamt. 2018. *Preise Verbraucherpreisindex für Deutschland Qualitätsbericht. Statistisches Bundesamt, Wiesbaden.* Available at: www.destatis.de/DE/Methoden/Qualitaet/Qualitaetsberichte/Preise/verbraucherpreis.pdf?__blob = publicationFile (accessed April 2021).

Statistisches Bundesamt. 2020. *Regional CPIs for Germany*. Available at: www.www-genesis.destatis.de/genesis/online (search "61111-0011"), (accessed April 2021).

Statistics Bureau of Japan. 2020. *Regional CPIs for Japan*. Available at: www.e-stat.go.jp/en/stat-search/files?page = 1&layout = datalist&toukei = 00200573&tstat = 000001084976&cycle = 1&year = 20200&month = 11010303&tclass1 = 000001085955 (accessed April 2021).

Swanson, D.C., S.K. Hauge, and M.L. Schmidt. 1999. "Evaluation of Composite Estimation Methods for Cost Weights in the CPI." In *Proceedings of the Section on Survey Research Methods, American Statistical Association.* Washington D.C. American Statistical Association. Available at: www.bls.gov/osmr/research-papers/1999/st990050.htm (accessed April 2021).

Tillmann, P. 2013. "Inflation targeting and regional inflation persistence: Evidence from Korea." *Pacific Economic Review* 18(2): 147–161. DOI: www.doi.org/10.1111/1468-0106.12016.

Tzavidis, N., L.C. Zhang, A. Luna, T. Schmid, and N. Rojas-Perilla. 2018. "From start to finish: A framework for the production of small area official statistics." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 181(4): 927–979. DOI: www.doi.org/10.1111/rssa.12364.

UK Statistics Authority. 2013. *Statistics on Consumer Price Inflation – Assessment Report 257.* UK Statistics Authority, London. Available at: www.osr.statisticsauthority.gov.uk/wp-content/uploads/2015/11/images-assessmentreport257statisticsonconsumerpriceinflatio_tcm97-43135.pdf (accessed April 2021).

Weber, A.A., and G.W. Beck. 2005. *Price stability, inflation convergence and diversity in EMU: Does one size fit all?* CFS Working Paper 2005(30). Goethe University, Center for Financial Studies (CFS), Frankfurt. Available at: www.nbn-resolving.de/urn:nbn:de:hebis:30-23427 (accessed April 2021).

Wingfield, D., D. Fenwick, and K. Smith. 2005. "Relative regional consumer price levels in 2004". *Economic Trends* 615: 36–45. Available at: www.ons.gov.uk/ons/rel/elmr/economic-trends–discontinued-/no–615–february-2005/relative-regional-consumer-price-levels-in-2004.pdf (accessed April 2021).

Yesilyurt, F., and J.P. Elhorst. 2014. "A regional analysis of inflation dynamics in Turkey." *The Annals of Regional Science* 52(1): 1–17. DOI: www.doi.org/10.1007/s00168-013-0570-4.

Zimmermann, T., F. Polidoro, F. Di Leo, M. Fedeli, J. Burger, J. van den Brakel, M. Pratesi, C. Giusti, S. Marchetti, L. Biggeri, G. Bertarelli, F. Schirripa Spagnolo, T. Laureti, I. Benedetti, P.A. Smith, J. Dawber, N. Tzavidis, A. Luna, J. O'Donoghue, T. Flower, H. Thomas, N. Würz, T. Schmid, C. Articus, J.P. Burgard, C. Caratiola, H. Dieckmann, F. Ertz, J. Krause, R. Münnich, and A.-L. Wölwer. 2020. Regional poverty measurement as a prototype for modern indicator methodology: Guidelines for best practices implementation for transferring methodology. MAKSWELL. Available at: www.makswell.eu/attached_documents/output_deliverables/deliverable_3.2.pdf (accessed April 2021).

Received May 2020
Revised January 2020
Accepted May 2021

Journal of Official Statistics, Vol. 38, No. 1, 2022, pp. 239–253, http://dx.doi.org/10.2478/JOS-2022-0011

The Geometric Young Formula for Elementary Aggregate Producer Price Indexes

Robert S. Martin[1], *Andy Sadler*[1], *Sara Stanley*[1], *William Thompson*[1], *and Jonathan Weinhagen*[1]

We re-estimate historical U.S. Producer Price Indexes (PPI) using the geometric Young formula at the elementary level. The geometric Young has better axiomatic properties than the modified Laspeyres, and may better approximate a feasible economic target. We find in most cases, indexes that use the geometric Young escalate between 0.1 and 0.3 percentage points less each year than those that use the modified Laspeyres. However, for wholesale and retail trade, as well as some other services, the differences are much larger. As a result, using the geometric Young at the elementary level lowers the U.S. PPI for Final Demand by 0.55 percentage points per year during the study period, a magnitude larger than what has been previously found for the U.S. Consumer Price Index.

Key words: Inflation; Aggregation; geometric Young index; Jevons index.

1. Introduction

This article presents experimental re-estimates of historical U.S. Producer Price Indexes (PPI) using the geometric Young formula at the elementary level. Elementary aggregates are the lowest levels at which price data are combined. For most statistical agencies, revenue or expenditure weights are either unavailable or available only at low frequencies and with significant lags. The U.S. Bureau of Labor Statistics (BLS) currently uses a modified Laspeyres formula (U.S. Bureau of Labor Statistics 2015), but statistical agencies in Italy, Chile, the Netherlands, among others, use the geometric Young (or something similar) for elementary PPIs (OECD 2011). Likewise, since 1999, the BLS has used the geometric Young for most elementary aggregates underlying the U.S. Consumer Price Index (CPI) (U.S. Bureau of Labor Statistics 2019).

The issue of formula choice has broad significance. High-quality elementary indexes are critical to producing high-quality PPIs (International Monetary Fund 2004). In turn, these play an important role in economic measurement. PPIs are used to measure inflation at different points of the supply chain, to adjust procurement contracts, and to produce other economic series. In particular, PPIs serve as deflators in national accounts to convert

[1]U.S. Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Washington, DC 20212, U.S.A. Emails: Martin.Robert@bls.gov, Sadler.Andy@bls.gov, Stanley.Sara@bls.gov, Thompson.William@bls.gov and Weinhagen.Jonathan@bls.gov.

Acknowledgments: We are grateful to Jeffrey Hill, Thesia Garner, Brian Adams, Rob Cage, Erick Sager, two anonymous referees, and members of the BLS Technical Advisory Committee, in particular Alberto Cavallo and Carol Corrado, for helpful comments. All errors are our own.

© Statistics Sweden

nominal production volumes and monetary sums into constant-dollar amounts. While a number of preceding studies, for example, Boskin et al. (1996), analyzed differences between formulas for consumer prices, relatively little empirical research of this nature examines producer prices. This article aims to fill that gap.

Section 2 describes the theoretical advantages of the geometric Young relative to the modified Laspeyres. It has better axiomatic properties (International Monetary Fund 2004) and lower formula bias (McClelland 1996; Reinsdorf 1998). Unlike with consumer prices, use of the geometric Young may be less motivated by the issue of substitution bias (Waehrer 2000), but Subsection 2.2 describes how it may better approximate a feasible economic target. While an exhaustive assessment of formulas is beyond the scope of this article, the analysis of the geometric Young is relevant for PPI compilers under similar constraints to those faced by the BLS. For instance, most elementary aggregates in the U.S. PPI cover heterogeneous commodities, ruling out either the unit value index or Dutot index, which are typically employed when quantities can be meaningfully aggregated across items (International Monetary Fund 2004). While sharing some axiomatic virtues with the Jevons index, the geometric Young has the improvement of incorporating some information regarding the relative economic importance of individual commodities. Finally, theoretically attractive indexes like the Fisher and Tornqvist require simultaneous collection of quantities (or value shares) along with prices, which is infeasible for the BLS.

We use both the modified Laspeyres and geometric Young formulas to calculate approximately 7,000 elementary indexes per month covering January 2008 to December 2017. We then aggregate these into versions of 1,016 six-digit commodity (product-based) indexes, 759 six-digit industry indexes, and the headline PPIs for Final Demand and Intermediate Demand (FD-ID). For a more detailed description of PPI classifications, (see, U.S. Bureau of Labor Statistics 2015). In most cases, six-digit indexes using the geometric Young formula at the elementary level are between 0.1 and 0.3 percentage points per year lower than those that use the modified Laspeyres. However, for wholesale and retail trade, as well as some other services, the differences are much larger. As a result, re-estimating the PPI for Final Demand using the geometric Young at the elementary level lowers the index by 0.55 percentage points per year, a larger magnitude than what has been found previously for the U.S. CPI Boskin et al. (1996).

Formula choice has a greater impact on U.S. PPIs in part because the BLS uses gross margins (selling price minus acquisition price) to measure the prices received by firms that resell products, such as wholesalers and retailers. Excluding these categories, the geometric Young lowers the PPI for Final Demand by 0.24 percentage points per year, comparable to Boskin, et al.'s finding of 0.25 for the U.S. CPI. Changes in margins tend to be more highly dispersed, driving greater differences between index formulas. In addition, the geometric Young is sensitive to near-zero margins. However, our results change very little when we impose bounds on the most extreme price changes prior to index calculation.

2. Theoretical Considerations

This section describes the theoretical benefits of moving to a geometric Young formula from a modified Laspeyres. After we formally introduce the index formulas, Subsection

2.1 argues that the geometric Young has superior axiomatic and statistical properties. Subsection 2.2 then describes how from an economic perspective, the geometric Young may better approximate a feasible target.

As stated, producer price aggregation typically occurs in two stages. First, price changes within narrowly defined groupings are combined to form elementary indexes. Then, these are aggregated into broader measures like the headline PPI for Final Demand. The U.S. PPI is currently based on a modified Laspeyres formula (U.S. Bureau of Labor Statistics 2015). More precisely, the target is known as the Lowe index. Let q_i^τ and p_i^τ denote quantity and price, respectively, for an item i in some time period τ. The Lowe index is then

$$I_{Lo}^t = \left(\frac{\sum_{i=1}^N q_i^b p_i^t}{\sum_{i=1}^N q_i^b p_i^0} \right) \times 100. \tag{1}$$

The index measures the change in expenditure on a fixed basket $\{q_1^b, \ldots, q_N^b\}$ from the reference period (price reference period) 0 to the comparison period t. Period b is the base period (weight reference period) from which quantity information is drawn. If the base and reference periods happen to be the same ($b = 0$), then the index coincides with the well-known Laspeyres index, which we denote I_{Lasp}^t. If the base and comparison periods are the same ($b = t$), then it coincides with the Paasche index, I_{Paas}^t. However, in practice, it is usually true that period b precedes period 0, particularly at the elementary level, where weighting information is not collected as quickly or frequently as prices. Implementation of the Lowe index often uses its expenditure share form, given by

$$I_{Lo}^t = \left(\sum_{i=1}^N s_i^{0b} \frac{p_i^t}{p_i^0} \right) \times 100, \tag{2}$$

where the $s_i^{0b} = p_i^0 q_i^b / \sum_{j=1}^N p_j^0 q_j^b$ are hybrid expenditure weights using period 0 prices and period b quantities. The ratio p_i^t / p_i^0 is sometimes referred to as the long-term price relative.

At upper levels of aggregation, the PPI uses something close to Equation (2). At the elementary level, however, weighting data for individual products are usually available in dollar values (the products $p_i^b q_i^b$) rather than quantities (the q_i^b by themselves). As a consequence, the "modified Laspeyres" formula actually implemented is closer to the Young index, written

$$I_Y^t = \left(\sum_{i=1}^N s_i^b \frac{p_i^t}{p_i^0} \right) \times 100, \tag{3}$$

where $s_i^b = p_i^b q_i^b / \sum_{j=1}^N p_j^b q_j^b$ are the actual expenditure weights from the base period. If $b = 0$, then the Young, Laspeyres, and Lowe are all equivalent.

The geometric Young index, given in Equation (4), combines the same price and expenditure information as the Young index above, but using a geometric mean instead of an arithmetic mean.

$$I_{GY}^t = \left(\prod_{i=1}^N \left(\frac{p_i^t}{p_i^0} \right)^{s_i^b} \right) \times 100 \tag{4}$$

This formula has also been called the geometric Lowe, weighted Jevons, and Cobb-Douglas price index. The BLS uses a version of this formula for the majority of elementary CPIs, as do several other countries for their elementary level PPIs. To avoid confusion, we use "modified Laspeyres" to refer to the arithmetic form of the Young index given in Equation (3), unless otherwise specified.

A corollary to Jensen's inequality implies the geometric mean will be less than or equal to the arithmetic mean when based on the same weights, and so we should generally expect index levels to be lower when using the geometric Young. There are theoretical advantages to the geometric Young, however, which the next two subsections describe.

2.1. Axiomatic and Statistical Considerations

This subsection outlines how the geometric Young is superior to the modified Laspeyres from the axiomatic perspective. As way of background, several different sets of overlapping tests for index numbers comprise what is known as the axiomatic approach to index numbers (International Monetary Fund 2004, chap. 16). Historically, this approach has favored indexes with some sort of geometric mean, such as the Fisher and Tornqvist, both of which are infeasible for the BLS due to the detail and frequency of weight information required.

The key axiomatic shortcoming of the modified Laspeyres is that, in general, it fails the time reversal test. Formally, a generic index $I(0,t)$ satisfies the time reversal test if the following holds:

$$I(0,t) = \frac{1}{I(t,0)} \quad or \quad I(0,t)I(t,0) = 1 \tag{5}$$

The idea of this test is that the index measurement should, in some sense, be independent of which period is regarded as the reference and which is regarded as the comparison. For example, if the index says the general price level doubled from period 0 to period t, then it should also say the price level fell by half from period t to period 0. The modified Laspeyres fails because $I_Y(0,t)$ is greater than $1/I_Y(t,0)$ unless prices for all items change by the same proportion. In other words, the modified Laspeyres leads to a higher measurement than if reference and comparison periods were reversed, which can be interpreted as an upward bias. On the other hand, it is easily verified that the geometric Young satisfies the time-reversal test. Another index that is time-reversible is the Carruthers et al. (1980) and Dalén (1992) index, which is the geometric mean of the modified Laspeyres and a weighted harmonic mean analog. This index and the geometric Young approximate each other to the second-order (Dalén 1992), however, so we focus on the latter for its relative simplicity.

The modified Laspeyres also fails the transitivity (or circularity) test, which says that when chained together, indexes over adjacent intervals should equal their direct counterpart. For time periods $0 < s < t$, it should be true that:

$$I(0,t) = I(0,s) \times I(s,t). \tag{6}$$

The degree of failure, which is related to the idea of chain drift, depends on the particular patterns of the data. However, unlike time-reversal, intransitivity does not necessarily imply a bias in any particular direction, rather a user-limitation. Ratios of modified Laspeyres indexes might not yield the intended comparison, as failure of Equation (6) implies $I_Y(0, t)/I_Y(0, s) \neq I_Y(s, t)$. The Geometric Young satisfies the transitivity test, as will any geometric mean of price relatives with time-constant weights summing to one.

Reinsdorf (1998) describes another scenario in which the modified Laspeyres index systematically exceeds the Laspeyres indexes, which he refers to as "formula bias." In a statistical model of mean-reversion, where prices fluctuate around a common mean, he shows how the arithmetic Young will exceed both the Lowe and Laspeyres in expectation. The intuition for his argument is that prices that are temporarily low in the reference period will receive excess weight in the Young formula. In a mean-reverting model, these are the prices expected to rise the most from 0 to t, leading to an inflation estimate that is biased upwards. He gives indirect evidence that this bias was empirically relevant for elementary indexes used in the CPI before formula changes that took place in the 1990s. He also discusses a geometric mean index as a solution. Assuming the elementary price data in the PPI follow a similar pattern, this argument supports use of the Geometric Young over the modified Laspeyres.

2.2. *Economic Considerations*

Despite the geometric Young having superior axiomatic properties when compared to the modified Laspeyres, the economic approach to index numbers has led to questions about its appropriateness for output PPIs. The economic approach compares a formula against a theoretical target derived from a model of an optimizing agent (International Monetary Fund 2004). In the case of output PPIs, the model is of a representative, price-taking firm that chooses how much of each item to produce in order to maximize profits. Fisher and Shell (1972) and Archibald (1977) propose a class of theoretical output price aggregators called Fixed Input Output Price Indexes (FIOPIs). A FIOPI measures the firm's hypothetical change in revenue from period 0 to period t if it could change its mix of outputs, but were forced to produce using fixed levels of technology and inputs. In this context, formulas based on geometric means have been viewed with some skepticism (e.g., Waehrer 2000) because of their association with consumer substitutions toward relatively less expensive products in a cost-of-living index (COLI) or input price framework. In contrast, producer theory suggests that the representative firm has incentive to substitute toward relatively more expensive products. In this section, we argue that opposition to a geometric mean-type index on economic grounds is too narrowly focused on one index (the geometric Laspeyres, which is Equation (4) with $b = 0$) and one theoretical target (the FIOPI which fixes technology and inputs to reference period levels). We offer a plausible scenario in which a geometric Young index may better approximate a FIOPI based on average technology and inputs.

First, it is instructive to formally introduce the FIOPI concept. Let $\mathcal{S}$ denote a production possibilities set, which is a collection of output quantity vectors of the form $\boldsymbol{q} = (q_1, \ldots, q_N)$ which are feasible given a certain level of technology and inputs. For example, $\mathcal{S}^0$ is the production possibilities set based on reference period technology and inputs. We assume that in each period τ, the representative producer observes market prices $\boldsymbol{p}^\tau = \left(p_1^\tau, \ldots, p_N^\tau\right)$ and chooses a quantity vector from $\mathcal{S}$ to maximize revenues. Note that since inputs are held fixed, maximizing revenues is equivalent to maximizing profits. Define the maximized revenues as

$$R\left(\boldsymbol{p}^\tau, \mathcal{S}\right) = \max_{\boldsymbol{q} \in \mathcal{S}} \sum_{i=1}^{N} p_i^\tau q_i, \qquad \tau = 0, t. \tag{7}$$

The class of FIOPIs is then given by

$$FIOPI(\mathcal{S}) = \frac{R\left(\boldsymbol{p}^t, \mathcal{S}\right)}{R\left(\boldsymbol{p}^0, \mathcal{S}\right)}. \tag{8}$$

The precise FIOPI is determined by the chosen $\mathcal{S}$. For example, $FIOPI\left(\mathcal{S}^0\right)$ is the hypothetical revenue change from 0 to t under the reference period's productive capacity. In general, FIOPIs are infeasible to implement in official statistics. As technology and input utilization generally do change over time, this would require specifying and estimating a model of production for every industry or commodity group.

While FIOPI are generally impractical, economic theory does provide some observable bounds. Specifically, the Laspeyres index, I_{Lasp}^t, is a lower bound for $FIOPI\left(\mathcal{S}^0\right)$, while the Paasche index, I_{Paas}^t, is an upper bound for $FIOPI\left(\mathcal{S}^t\right)$ (Archibald 1977). To see the Laspeyres result, observe that under the assumption of profit maximization, $R\left(\boldsymbol{p}^0, \mathcal{S}^0\right)$ equals the (in theory) observable revenue level from the reference period, $\sum_{i=1}^N p_i^0 q_i^0$. Likewise, as prices change, the firm has incentive to shift production toward items that are relatively more profitable. This implies the maximized revenue, $R\left(\boldsymbol{p}^t, \mathcal{S}^0\right)$, must be at least as high as $\sum_{i=1}^N p_i^t q_i^0$, the revenue received if period 0's production levels are maintained. Therefore,

$$FIOPI\left(\mathcal{S}^0\right) = \frac{R\left(p^t, \mathcal{S}^0\right)}{R\left(p^0, \mathcal{S}^0\right)} \geq \frac{\Sigma_{i=1}^N p_i^t q_i^0}{\Sigma_{i=1}^N p_i^0 q_i^0} = I_{Lasp}^t. \tag{9}$$

The reference period FIOPI equals I_{Lasp}^t only in the case where the firm produces in fixed proportions, i.e., no substitutions. Otherwise, I_{Lasp}^t has a downward bias for $FIOPI\left(\mathcal{S}^0\right)$. As the geometric Laspeyres index is necessarily less than or equal to I_{Lasp}^t, it has greater bias. For this reason, Waehrer (2000) opposes the use of geometric mean indexes for output PPIs.

However, the one-way bounds concerning I_{Lasp}^t and I_{Paas}^t may be of limited empirical relevance. This is because they relate to different FIOPIs which may not be close to each other numerically. In particular, if $I_{Lasp}^t \geq I_{Paas}^t$, then we have

$$FIOPI\left(\mathcal{S}^0\right) \geq I_{Lasp}^t \geq I_{Paas}^t \geq FIOPI\left(\mathcal{S}^t\right). \tag{10}$$

While we lack data to evaluate the inequality $I_{Lasp}^t \geq I_{Paas}^t$ at the elementary level (since both indexes are infeasible), Weinhagen (2020) found $I_{Lasp}^t \geq I_{Paas}^t$ holds for broader industry and commodity aggregates. This inequality reflects negative correlation between market prices and quantities, implying that even if firms have a *ceteris paribus* incentive to substitute production towards higher relative output prices, this is outweighed in equilibrium by other factors like changes in demand or technology. If this is the case, then targeting either $FIOPI(\mathcal{S}^0)$ or $FIOPI(\mathcal{S}^t)$ might be arbitrary, seeing how the gap between them is at least as large as $I^t_{Lasp} - I^t_{Paas}$. Furthermore, the bias of the geometric Laspeyres may be different from the bias of the geometric Young, which does not (to our knowledge) have a known relationship to potential targets like $FIOPI(\mathcal{S}^0)$ or even $FIOPI(\mathcal{S}^b)$. Arguments against the geometric Young based on these bounds, therefore, are too narrowly focused.

Nevertheless, the geometric Young may be attractive if the target is a FIOPI based on average (i.e., between the reference and comparison periods) technology and inputs. From Diewert (1983), there exists such a FIOPI which, unlike $FIOPI(\mathcal{S}^0)$ or $FIOPI(\mathcal{S}^t)$, is bounded by both I^t_{Lasp} and I^t_{Paas}. Formally, for some α between 0 and 1, let $\mathcal{S}^\alpha = \alpha\mathcal{S}^0 + (1-\alpha)\mathcal{S}^t$ denote a weighted average of the reference and comparison period production possibilities sets. Diewert showed that there exists a value of α such that either $I^t_{Lasp} \leq FIOPI(\mathcal{S}^\alpha) \leq I^t_{Paas}$ or $I^t_{Paas} \leq FIOPI(\mathcal{S}^\alpha) \leq I^t_{Lasp}$. If the difference between I^t_{Lasp} and I^t_{Paas} is not too great, a symmetric average like the Fisher index, $I^t_{Fisher} = \sqrt{I^t_{Lasp} I^t_{paas}}$, is a good approximation to the "average" $FIOPI(\mathcal{S}^\alpha)$. The Tornqvist index has similar properties (Caves et al. 1982).

As stated in the previous subsection, neither the Tornqvist nor the Fisher are feasible for elementary PPIs. Nevertheless, we describe a plausible scenario under which the geometric Young may be preferred to the modified Laspeyres from the standpoint of targeting an average FIOPI. International Monetary Fund (2004, chap. 15, sec. D.3) describes conditions, reasonable for elementary aggregates, under which I^t_Y exceeds I^t_{Lasp}. The first is that the base period precedes the reference period, as with the U.S. PPI. The second is that prices are trending either up or down over the long-term, which is generally true for U.S. PPI data. Next is that changes in market quantities primarily reflect purchaser substitutions away from items with higher relative prices. This is true if $I^t_{Lasp} \geq I^t_{Paas}$, for which we interpret Weinhagen (2020) as indirect evidence. The last condition is that these substitutions are elastic, meaning revenues and prices move in opposite directions. Elastic substitution patterns are likely if the elementary product grouping contains highly similar varieties, as is the case with the retail sales data studied by Martin (2020) and others.

Under the scenario described above, $I^t_Y \geq I^t_{Lasp} \geq I^t_{Fisher}$, and it is always true that $I^t_Y \geq I^t_{GY}$. This means that switching from the modified Laspeyres to the geometric Young moves the index in the direction of the Fisher index, which approximates $FIOPI(\mathcal{S}^\alpha)$. Of course, this does not guarantee that the geometric Young will have lower bias for $FIOPI(\mathcal{S}^\alpha)$, and the conditions in the preceding paragraph may not be appropriate for all sectors. Nevertheless, it is plausible that the geometric Young better approximates a more feasible economic target than the modified Laspeyres.

3. Application

3.1. Data

The previous section enumerated reasons for which the geometric Young, Equation (4), should be preferred to the modified Laspeyres, Equation (3), from the standpoint of price index theory. To demonstrate the practical implications of formula choice, we use the price and expenditure microdata from the regular monthly production of the U.S. PPI. Using both formulas, we calculate approximately 7,000 elementary indexes per month covering January 2008 to December 2017. Roughly half of these measure output prices for industries, which are organized according to the North American Industry Classification System (NAICS). The other half measure prices for commodities (regardless of producing industry) according to an internal BLS classification system.

We then aggregate the elementary indexes to form versions of 1,016 six-digit commodity indexes, 759 six-digit industry indexes, and the headline PPIs for Final Demand and Intermediate Demand (FD-ID). Because the focus is on differences in elementary calculation, all indexes use same the Lowe formula, Equation (2), at the upper levels. Furthermore, we recalculate indexes that use the modified Laspeyres formula at the elementary level, rather than comparing to the published PPIs, in order to better hold constant other components of methodology such as imputation and item structure changes which are harder to replicate in a research environment. In 98.5% of observations, monthly percent changes of the re-estimated six-digit commodity indexes fall within 0.1 percentage points of the actual indexes from production.

3.2. *Results*

As stated, we combine each set of elementary indexes into industry and commodity aggregate indexes. The average annual change across the six-digit commodity indexes calculated using the modified Laspeyres is 1.52%, versus 1.25% for the geometric Young, a difference of 0.27 percentage points. Across industries, the modified Laspeyres indexes average 1.70%, while the geometric Young indexes average 0.36 points lower at 1.34%. There is considerable heterogeneity across commodities and industries. Figure 1 plots the frequencies of annual percentage point differences for the six-digit commodity indexes.

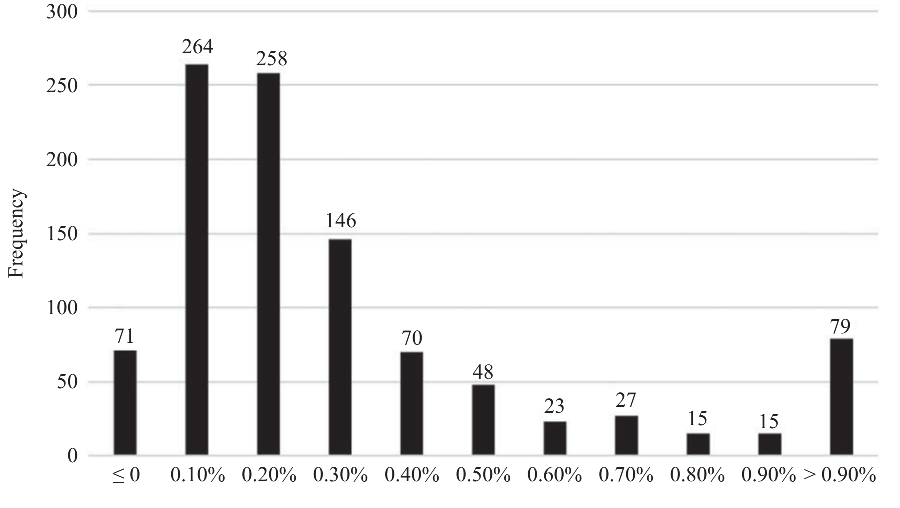


Fig. 1. Difference between modified Laspeyres and geometric Young for six-digit commodities.
Note: Observations are differences in annual percent changes for 6-digit commodities. Modified Laspeyres and Geometric Young refer to formulas used for elementary aggregation. Upper-level aggregation uses Lowe formula in all cases.

About two thirds of commodities show differences in the 0.0 to 0.3 percentage point range. Frequencies generally decline over higher values, but the right tail is long, with 79 commodities having differences exceeding 0.9 percentage points. As expected, the modified Laspeyres implies higher inflation than the geometric Young for about 95% of commodities. Qualitatively, the distribution across six-digit industries is very similar, so we omit the corresponding histogram. The modified Laspeyres implies higher inflation than the Geometric Young for about 97% of industries. Note that a geometric mean will generally result in lower index levels (i.e., reference period to comparison period measurements), but the comparison may not always hold for short-term percent changes or when the considered timeframe spans item rotations or weight updates.

Table 1 presents the average annual percent changes for seven broad commodity categories. In all but one category (Wholesale and Retail Trade), the formulas give average percent changes of the same sign, and the average differences mainly fall in the 0.2 to 0.4 percentage point range. Notable exceptions include Construction, where the average difference is only 0.05 percentage points, and Wholesale and Retail Trade, where the difference (1.35 percentage points) is more than three times that of any other category. The formulas disagree in sign for only about 3.3% of six-digit commodities overall, but within Trade, they disagree in 32% of cases. Similarly, Table 2 gives the average annual percent changes for the six-digit industry indexes within broad NAICS categories. As with the commodities, most differences average well under one percentage point per year with the exception of Wholesale Trade, Retail Trade, and Finance and Insurance, where the average differences are 1.14, 1.71, and 0.72 percentage points per year, respectively.

As a general principle, greater dispersion of the underlying elements (in this case, long-term price relatives) is associated with a greater difference between the arithmetic and geometric mean. We should then expect to see greater dispersion in industries like Trade. Because of periodic discontinuations, we can only recover the long-term relatives for items that are observed during the entire period between sample rotations, a group which we label "survivors". To check representativeness, we construct sets of industry indexes using only this subsample and present their average differences in column 4 of Table 3. The full sample differences from Table 2 have been copied to column 3 for comparison. Using the survivors only, the average differences are slightly greater in magnitude (0.42 versus 0.35 percentage points per year), but qualitatively similar to those based on the full-sample. Column 5 shows

Table 1. Commodity averages by category, 2008–2017 (Annual percent change).

Category	Mod. Lasp.	Geo. Young	Lasp. – Geo.
Food	1.66	1.45	0.21
Energy	−0.02	−0.40	0.38
Goods less food and energy	1.68	1.50	0.19
Wholesale and retail yrade	1.09	−0.26	1.35
Transportation	2.04	1.64	0.40
Services less trade and Transp.	1.08	0.64	0.44
Construction	1.60	1.55	0.05

Note: Rows are averages of six-digit commodity indexes within specified category. Mod. Lasp. (Modified Laspeyres) and Geo. Young (Geometric Young) refer to formulas used for elementary aggregation. Upper-level aggregation uses Lowe formula in all cases.

Table 2. Industry averages by NAICS category, 2008–2017 (Annual percent change).

NAICS	Description	Mod. Lasp.	Geo. Young	Lasp. – Geo.
11	Ag., forestry, fishing and hunting	1.63	1.47	0.16
21	Mining, quarrying, and oil and gas	2.92	2.46	0.47
22	Utilities	1.17	0.64	0.53
23	Construction	2.10	1.86	0.23
31–33	Manufacturing	1.78	1.55	0.23
42	Wholesale trade	1.96	0.82	1.14
44–45	Retail trade	1.05	−0.66	1.71
48–49	Transportation and warehousing	2.13	1.79	0.34
51	Information	−0.57	−1.00	0.43
52	Finance and insurance	1.83	1.11	0.72
53	Real estate and rental and leasing	1.02	0.67	0.35
54	Prof., scientific, and technical services	1.57	1.40	0.17
56	Admin., supp., waste, and rem. services	1.10	0.97	0.14
61	Educational services	1.23	0.96	0.27
62	Health care and social assistance	1.47	1.29	0.19
71	Arts, entertainment, and recreation	2.19	1.80	0.39
72	Accommodation and food services	0.81	0.61	0.20
81	Other services (ex. public admin.)	2.24	1.85	0.40

Note: Rows are averages of six-digit industry indexes within specified NAICS category. Mod. Lasp. (Modified Laspeyres) and Geo. Young (Geometric Young) refer to formulas used for elementary aggregation. Upper-level aggregation uses Lowe formula in all cases.

Table 3. Industry differences and dispersion, 2008–2017.

NAICS	Description	Full Sample Lasp. – Geo.	Survivors Lasp. – Geo.	Survivors LTR C.V.
11	Ag., forestry, fishing and hunting	0.16	0.48	0.17
21	Mining, quarrying, and oil and gas	0.47	0.42	0.18
22	Utilities	0.53	0.84	0.28
23	Construction	0.23	0.23	0.11
31–33	Manufacturing	0.23	0.30	0.13
42	Wholesale trade	1.14	1.41	0.47
44–45	Retail trade	1.71	1.68	0.35
48–49	Transportation and warehousing	0.34	0.35	0.16
51	Information	0.43	0.40	0.16
52	Finance and insurance	0.72	1.26	0.33
53	Real estate and rental and leasing	0.35	0.33	0.16
54	Prof., scientific, and technical services	0.17	0.13	0.10
56	Admin., supp., waste, and rem. services	0.14	0.12	0.10
61	Educational services	0.27	0.17	0.13
62	Health care and social assistance	0.19	0.28	0.15
71	Arts, entertainment, and recreation	0.39	0.53	0.17
72	Accommodation and food services	0.20	0.41	0.21
81	Other services (ex. public admin.)	0.40	0.41	0.22

Note: Rows are averages within specified NAICS category. Index differences are expressed as percentage points per year. “Survivors” refers to indexes calculated using only those items available during entire sample period. “LTR C.V.” denotes coefficient of variation for the long-term relatives. Lasp. and Geo. refer to formulas used for elementary aggregation. Upper-level aggregation uses Lowe formula in all cases.

the average coefficients of variation within each NAICS category. Indeed, within the Trade, Financial Services, and Insurance industries, the long-term price relatives have coefficients of variation of 0.36 on average, versus 0.14 for all other industries.

BLS views firms that resell products as providers of services rather than goods. As such, the prices used for Trade are primarily gross margins (selling price minus acquisition price). Gross margins for retailers, for example, reflect the value added by the establishment for services such as marketing, storing, displaying goods, and making the goods easily available for customers to purchase. Some indexes within financial services also use measures that are similar to margins, like bid-ask spreads or interest rate differentials. Margin prices tend to be more volatile than selling prices alone, on account of them reflecting multiple sources of variation. For example, competitive pressure may dissuade a retailer from adjusting selling prices even when acquisition costs fluctuate. Similarly, loss-leader strategies may result in margin prices that are zero or negative. These are problematic for the geometric mean — a single price relative of zero results in a mean of zero, while a single negative price relative renders the mean undefined. The BLS excludes zero or negative prices from calculation. Even so, geometric mean indexes are still sensitive to margins that are close to zero, which can cause the long-term and month-to-month price relatives to be very small or very large (International Monetary Fund 2004).

To assess potential sensitivity, we calculate the commodity indexes after imposing bounds of 0.05 and 20 on the monthly relatives, which matches the BLS procedure for the CPI. For example, if a relative is less than 0.05, we use the value 0.05 in its place. The results change very little. For the Trade category, the average percent changes for the modified Laspeyres and geometric Young increase by 0.025 and 0.04 percentage points, respectively, decreasing the gap between them by only 0.015 percentage points. Similar results hold for tighter bounds of [0.25, 4], which decrease the gap by an additional 0.026 percentage points. There are still influential outliers inside these bounds, but it does not appear that the most extreme price relatives are driving the formula differences.

Aggregation of the commodity indexes into the headline PPIs shows the importance of formula choice. Table 4 presents average annual percent changes for the FD-ID indexes over 2010–2017. We also calculate the indexes with and without Trade and Finance, which

Table 4. PPI for final and intermediate demand, 2010–2017 (Annual percent change).

	Elementary indexes		
Index	Mod. Lasp.	Geo. Young	Lasp. – Geo.
Final demand	1.56	1.01	0.55
Less trade and finance	1.47	1.23	0.24
Trade and finance	1.74	0.22	1.53
Intermediate demand	–	–	–
Processed goods	0.98	0.73	0.25
Unprocessed goods	−1.25	−1.40	0.15
Services	2.02	1.48	0.54
Less trade and finance	1.38	1.21	0.17
Trade, and finance	3.05	1.87	1.18

Note: Mod. Lasp. (Modified Laspeyres) and Geo. Young (Geometric Young) refer to formulas used for elementary aggregation. Upper-level aggregation uses Lowe formula in all cases.

include margin prices to varying degrees. The Final Demand index escalates 0.55 percentage points per year less when the elementary indexes use the geometric Young formula. This magnitude is larger, but near the range found by similar studies of CPI elementary indexes, such as Boskin et al. (1996), which found an all-items index difference of 0.25 percentage points, and Reinsdorf and Moulton (1996), which found a difference of 0.47 percentage points. In the case of the PPI for Final Demand, much of the difference is due to the Trade and Finance sectors, which collectively show 1.53 percentage points lower inflation using the geometric Young. Excluding these, the difference between Final Demand indexes is only 0.24 percentage points per year. Figures 2 and 3 illustrate the role these

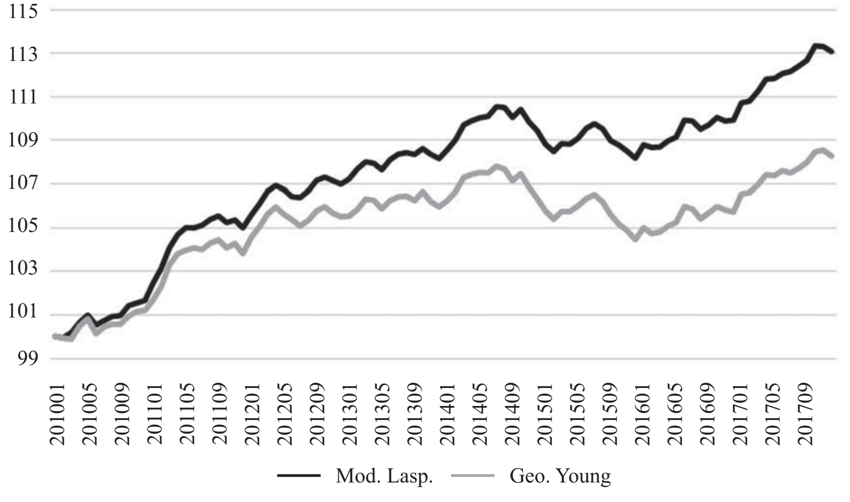


Fig. 2. PPI for final demand (Jan. 2010 = 100).
Note: Mod. Lasp. (Modified Laspeyres) and Geo. Young (Geometric Young) refer to formulas used for elementary aggregation. Upper-level aggregation uses Lowe formula in all cases.

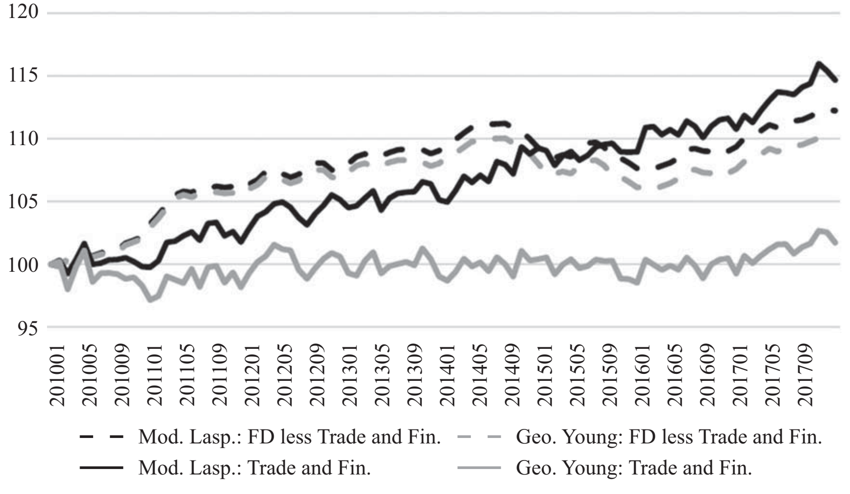


Fig. 3. PPI for final demand with and without trade and finance (Jan. 2010 = 100).
Note: Mod. Lasp. (Modified Laspeyres) and Geo. Young (Geometric Young) refer to formulas used for elementary aggregation. Upper-level aggregation uses Lowe formula in all cases.

industries play in driving the relative evolution of the experimental indexes. The Intermediate Demand indexes follow a similar pattern. The Processed and Unprocessed Goods indexes differ by 0.25 and 0.15 percentage points per year, respectively. As with Final Demand, the indexes covering Services for Intermediate Demand differ to a greater degree (0.54 percentage points per year), though this gap narrows considerably (to 0.17 percentage points per year) when excluding Trade and Finance.

4. Conclusion

The geometric Young formula has superior axiomatic properties to the modified Laspeyres, and it may better approximate a FIOPI based on an average production possibilities set. Our application to U.S. PPI data shows that these theoretical differences have economically significant consequences for elementary index aggregation. Using the geometric Young, for example, would lower the PPI for Final Demand by 0.55 percentage points per year. The effect on most industry and commodity PPIs is smaller—between 0.1 and 0.3 percentage points. For services like wholesale and retail trade, however, higher dispersion in margin prices leads to differences often exceeding one percentage point. Our main findings are little changed when bounding the price relatives, implying the formula differences are not primarily driven by outliers, for example, values close to zero. The issue of margin prices is unique to PPIs and helps explain why we find greater differences between formulas than earlier studies found using consumer prices.

This issue has broader implications for economic statistics. We expect lower PPIs, for instance, would translate into higher growth rates for measures of real output (for which PPIs often serve as deflators), and therefore higher rates of growth for real productivity. Furthermore, since formula choice has a nontrivial impact on the measurement of inflation, analysts should consider potential differences in methodologies when making international comparisons.

Finally, future research could further assess the potential benefits of alternative data sources, such as transactions records, in calculation of elementary PPIs. These may offer frequent quantity data along with prices, making theoretically preferred, variable-weight formulas like the Fisher or Tornqvist feasible for very narrowly-defined aggregates.

5. References

Archibald, R.B. 1977. "On the Theory of Industrial Price Measurement: Output Price Indexes." In *Annals of Economic and Social Measurement Vol. 6*, edited by S.V. Berg. (pp. 57–72). Cambridge: National Bureau of Economic Research. Available at: http://www.nber.org/chapters/c10502 (accessed October 2018).

Boskin, M.J., E.R. Dulberger, R.J. Gordon, Z. Griliches, D. Jorgenson. 1996. "Toward a more accurate measure of the cost of living: Final Report to the U.S. Senate Finance Committee." In *U.S. Senate Finance Committee. Senate Print 104–72, December, 104th Congress, 2nd Session*. Washington, DC: GPO.

Carruthers, A.G., D.J. Sellwood, and P.W. Ward. 1980. "Recent Developments in the Retail Prices Index." *The Statistician* 29(1): 1–32. DOI: https://doi.org/10.2307/2987492.

Caves, D.W., L.R. Christensen, and W.E. Diewert. 1982. "The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity." *Econometrica* 50(6): 1393–1414. DOI: https://doi.org/0012-9682(198211)50:6 2.0.CO;2-S.

Dalén, J. 1992. "Computing Elementary Aggregates in the Swedish Consumer Price Index." *Journal of Official Statistics* 8(2): 129–147. Available at: https://www.scb.se/-contentassets/ca21efb41fee47d293bbee5bf7be7fb3/computing-elementary-aggregates-in-the-swedish-consumer-price-index.pdf (accessed January 2022).

Diewert, W.E. 1983. "The Theory of the Output Price Index and the Measurement of Real Output Change." In *Price Level Measurement*, edited by W.E. Diewert and C. Montmarquette Diewert. 1049–1113. Ottawa: Statistics Canada.

Fisher, F.M., and K. Shell. 1972. "The Pure Theory of the National Output Deflator." In *The Economic Theory of Price Indexes*, edited by F.M. Fisher and K. Shell. 49–113. New York: Academic Press.

International Monetary Fund. 2004. In *Producer Price Index Manual: Theory and Practice*, edited by P. Armknecht. Washington, DC: IMF/ILO/OECD/Eurostat/UNE-CE/World Bank. Available at: https://www.imf.org/external/pubs/ft/ppi/2010/manual/ppi.pdf (accessed March 2019).

Martin, R.S. 2020. *Revisiting Taste Change in Cost-of-Living Measurement*. Working Paper 515, Washington, DC: U.S. Bureau of Labor Statistics. Available at: https://www.bls.gov/osmr/research-papers/2019/ec190050.htm (accessed May 2020)

McClelland, R. 1996. *Evaluating Formula Bias in Various Indexes Using Simulations.* Working Paper 289, Washington, DC: U.S. Bureau of Labor Statistics. Available at: https://www.bls.gov/osmr/research-papers/1996/ec960140.htm (accessed May 2020).

OECD, Organisation for Economic Co-operation and Development. 2011. *Producer Price Indices – Comparative Methodological Analysis.* Paris: Organisation for Economic Co-operation and Development. Available at: http://www.oecd.org/sdd/prices-ppp/48370389.pdf (accessed January 2018).

Reinsdorf, M.B. 1998. "Formula Bias and Within-Stratum Substitution Bias in the U.S. CPI." *The Review of Economics and Statistics (MIT Press)* 80(2): 175–187. DOI: https://doi.org/10.1162/003465398557375.

Reinsdorf, M.B., and B.R. Moulton. 1996. "The Construction of Basic Components of Cost-of-Living Indexes." In *The Economics of New Goods*, 397–436. University of Chicago Press. Available at: http://www.nber.org/chapters/c6073 (accessed May 2020).

U.S. Bureau of Labor Statistics. 2015. "Chapter 14: Producer Prices." In *Handbook of Methods.* Washington D.C.: U.S. Bureau of Labor Statistics. Available at: https://www.bls.gov/opub/hom/pdf/ppi-20111028.pdf (accessed March 2019).

U.S. Bureau of Labor Statistics. 2019. "Chapter 17: The Consumer Price Index." In *Handbook of Methods.* Washington, DC: Bureau of Labor Statistics. Accessed August 14, 2019. https://www.bls.gov/opub/hom/pdf/cpihom.pdf (accessed August 2019)..

Waehrer, K. 2000. *A Comparison of the Substitution Effects for Input and Output Price Indexes*. Working Paper 327, Washington, DC: Bureau of Labor Statistics. Available at: https://www.bls.gov/osmr/research-papers/2000/ec000020.htm (accessed July 2019).

Weinhagen, J. 2020. "Measuring the substitution effect in Producer Price Index goods data: 2002–2016." *Monthly Labor Review*. Available at: https://www.bls.gov/opub/mlr/2020/article/measuring-the-substitution-effect-in-ppi-data.htm (accessed July 2020).

Received May 2020
Revised September 2020
Accepted April 2021

Journal of Official Statistics, Vol. 38, No. 1, 2022, pp. 255–285, http://dx.doi.org/10.2478/JOS-2022-0012

Measuring Inflation under Pandemic Conditions

W. Erwin Diewert[1] *and Kevin J. Fox*[2]

National statistical offices have faced unprecedented circumstances in the modern history of economic measurement. There were dramatically changing consumer expenditure patterns due to pandemic conditions, with lockdowns and fear of infection making many goods and services unavailable. We examine the implications of changing relative expenditures for the construction of Consumer Price Indexes, with special reference to the treatment of prices for unavailable products. We conclude that for many purposes, it would be useful for statistical agencies to establish a continuous consumer expenditure survey. We also examine various other practical pandemic induced CPI measurement problems.

Key words: Consumer price index; disappearing products; COVID-19.

1. Introduction

The majority of national Consumer Price Indexes (CPIs) are based on pricing out a fixed basket of goods and services that people typically buy in the current period relative to a base period. For National Statistical Offices (NSOs), an implication of the 2020 Covid-19 pandemic and the associated lockdowns was a substantial *disappearing products problem* in the construction of such price indexes. During the pandemic, many goods and services became unavailable and expenditure patterns for still available products changed dramatically (see Carvallo 2020; Garcia et al. 2020; Dunn et al. (2020). Thus the fixed basket approach to the construction of a CPI led to measures of consumer price inflation that were likely to be biased because the use of a pre-pandemic basket did not reflect consumer expenditure patterns during the pandemic. This CPI credibility problem was noticed in the financial press. For example:

"Consumption patterns have changed so much that inflation indices are meaningless." Martin Wolf, *Financial Times*, May 19, 2020.

"But did you notice something about the big price drops quoted?. . .Great deals are available but no one can take advantage of them. In fact, they're available precisely because no one can take advantage of them. . .We have deflation across the basket of goods we usually buy but inflation across the much narrower range of goods we're buying now." William Watson, *Financial Post*, May 21, 2020.

[1] Department of Economics University of British Columbia #997-1873 East Mall Vancouver, BC V6T 1Z1, Canada. Email: erwin.diewert@ubc.ca
[2] School of Economics and CAER UNSW Sydney, NSW 2052, Australia Email: K.Fox@unsw.edu.au
Acknowledgments: The authors thank Paul Armknecht, John Astin, Jonathan Athow, Corinne BeckerVermeulen, Carsten Boldsen, Kate Burnett-Isaacs, Jan de Haan, François Des Rosiers, David Fenwick, Jason Furman, Daniel Hamermesh, Niall O'Hanlon, Robert Hill, Paul Konijn, Shengyu Li, Jens Mehrhoff, Leigh Merrington, James Poterba, Marshall Reinsdorf, Chihiro Shimizu, Dan Sichel, Mick Silver, Miriam Steurer, Michael Webster, Kam Yu and the referees and editors for helpful comments and discussions on an earlier version of this article.

© Statistics Sweden

How to deal with this credibility problem is a major focus of this article. We wrote the first version of this article in April 2020 with the intention of providing statistical agencies with possible methods for dealing with the absence of many goods and services due to pandemic induced lockdowns. Some of our advice has been implemented and some has not. In addition, we attempt to inform economists who have not specialized in price measurement problems on how actual CPIs are constructed.

For products which have disappeared, the advice to NSOs from Eurostat (2020), the International Monetary Fund (2020), the UNECE (2020) and the IWGPS, Intersecretariat Working Group on Price Statistics (2020) was to implement an inflation adjusted carry forward of missing prices methodology; that is, when a price is missing, the price for a commodity for the period prior to the lockdown is used in the current period, with some adjustment for inflation. We show that following this advice can lead to an understatement of inflation relative to a consumer price index that makes use of the concept of a reservation price. The IWGPS includes Eurostat, the IMF, ILO, OECD, World Bank and UNECE. It should be noted that the advice that is available on these official websites to deal with the changing weights and missing prices problems is similar to the advice that we suggest in later sections of this article. The official advice deals with many practical problems that we do not address.

In general, our article presents a broader review of the price and quantity indexes that statistical agencies could potentially produce during lockdown conditions. The options open to an NSO will depend on its access to current household expenditure data.

The article is organised as follows. In Section 2, we provide a non-technical discussion of different approaches to index number choice and highlight that this article, for the most part, uses the Fixed Basket approach. The implications of using this approach when the actual consumption basket is changing dramatically under pandemic conditions are discussed. Section 3 looks at comparisons between the Laspeyres, Paasche and Lowe price indexes when inflation adjusted carry forward prices are used for unavailable products, as recommended by international agencies. Section 4 looks at the advantages and disadvantages of using various "practical" price and quantity indexes that statistical agencies are likely to produce during lockdown conditions. We note that the way forward will depend on what types of data are available to the NSO.

Section 5 looks at the problem of a lack of matching product prices at the elementary index level; that is, we discuss the missing prices problem. Possible methods for dealing with this problem are discussed, depending on the availability of data. Section 6 takes a brief look at other practical measurement problems that an NSO may encounter when it attempts to produce a meaningful CPI under pandemic conditions. Section 7 concludes.

2. Alternative Approaches to Index Number Theory

This section presents a brief non-technical overview of the various approaches to index number theory that form the theoretical basis for a CPI. Formal definitions of the relevant indexes follow in the next section.

There are at least four main approaches to index number theory that have been put forth in the literature on bilateral price index numbers. Bilateral index number theory compares prices for two periods. These approaches are as follows:

- *Basket approaches* include the use of Lowe, Laspeyres and Paasche indexes, which will be formally defined in the following section. Informally, a basket index is an index that chooses a representative quantity "basket" and calculates the cost of purchasing this basket at the prices of the current period (numerator) and at the prices of the base period (denominator). The ratio of these two costs forms the price index that calculates inflation going from the base period to the current period. This approach to index number theory is a natural one and goes back hundreds of years. The practical problem with this approach is how exactly are we to choose the "representative" quantity weights? The Lowe index weights prices in each period using the quantities from some base period, which may not be either of the periods being compared. The Lowe index formula is used in CPI construction in most countries. Carsten Boldsen in a personal communication noted that in a 2007 ILO/UNECE survey on the 2004 CPI Manual, of the countries who replied, 65% said that they used a Lowe index while 35% said that they used a Young index. These indexes will be defined in the following section. The Laspeyres index uses the quantities from the earlier of the two periods being compared, while the Paasche index using the quantities from the later, or "current", period. From the viewpoint of index number theory, the Paasche and Laspeyres indexes have the same theoretical footing; that is, *both are equally representative for the two periods under consideration.* Thus if a single estimate of consumer inflation is to be produced, it is natural to take some sort of average of the Laspeyres and Paasche indexes as a target index for a CPI and it turns out that taking the geometric mean of these two indexes produces a "best" average that will satisfy the time reversal test from the *test approach* to index number choice; see the next dot point below. This geometric mean is the Fisher (1922) index. Thus the basket approach to index number theory (including taking averages of basket type indexes) leads to the Fisher price index as a suitable target index. (Another approach to forming a single representative estimate of inflation is to take the geometric average of the quantity weights for the two periods under consideration as the representative set of quantities. This form of averaging leads to the Walsh (1901) index as a suitable target index. We note that the Walsh and Fisher indexes will generally approximate each other very well; see Diewert (1978, 889).
- *The Test or Axiomatic approach* to index number theory also leads to the Fisher index. This approach regards the index number formula as a mathematical function of the two price vectors and two quantity vectors that represent the transactions in scope for the two periods under consideration. The approach postulates that the index number formula satisfies various "reasonable" tests or mathematical properties. If a sufficient number of tests are imposed on the index number formula, then an explicit functional form for the index can be determined. An example of an important test is the time reversal test: if the data for the two time periods is exchanged, then the resulting index is the reciprocal of the initial index that did not interchange the data for the two periods. The Fisher and Walsh indexes satisfy this test whereas the Laspeyres and Paasche indexes do not satisfy it. The Fisher index satisfies more reasonable tests than any other commonly used bilateral index number formula (Diewert 1992), and so it is generally regarded as being the "best" index number formula from the viewpoint of the test approach to index number theory.

- *The Stochastic approach* to index number theory assumes that the price index can be represented as a weighted average of the individual price ratios, say p_n^1/p_n^0 where p_n^t is the price of commodity n in period t for $t = 0$ (the base period) and $t = 1$ (the current period). The Törnqvist Theil index is generally regarded as being "best" from this perspective (Theil 1967). This index uses the arithmetic average of expenditure shares in the two periods being compared as the weight for each individual price ratio p_n^1/p_n^0. The overall index is a weighted geometric average of these price ratios using the average shares as weights. It usually approximates the Fisher index closely in empirical applications (Diewert 1978, 889).
- *The Economic Approach or the "Konus True Cost of Living Index" approach* brings economics into the picture; that is, the approach assumes that consumers either maximize utility subject to a budget constraint or they minimize the cost of achieving a certain level of utility. The Fisher, Törnqvist and Walsh indexes all receive an equally strong justification from the perspective of this approach; see Diewert (1976, 1978, 2021a). In the CPI literature, an index based on the economic approach to index number theory is often called a Cost of Living Index or a COLI. It is typically contrasted with a basket type index which is called Cost of Goods Index or COGI; see Deaton and Diewert (2002).

For more on the different approaches to bilateral index number theory (see e.g. ILO 2004; Fisher 1922; Konus 1924; Theil 1967; Diewert 1976, 1992, 2021a).

For the most part, we take the Basket Approach to index number theory throughout the article. We assume that the Laspeyres and Paasche indexes are good basket type indexes: they are intuitively plausible and easy to explain to the public. If they differ substantially and if the situation calls for a single estimate of consumer price inflation, then (from our perspective) these two indexes need to be averaged to give a single credible estimate of inflation, which leads to the Fisher index as their geometric mean. This is regarded as a good index not only from the perspective of the basket approaches to index number theory but it is also a "best" approach from the perspectives of both the economic and test approaches to index number theory.

If quantities consumed do not change much over the two periods being compared, then the actual quantity vectors that are consumed in the two periods being compared may be quite close to the reference quantity vector that is used in the definition of the Lowe index. There will be some substitution bias (as compared to the economic approach) when using the Lowe index but typically, this "bias" will be relatively small.

However, under pandemic conditions, expenditure patterns changed dramatically. A fixed basket index is very easy to explain and is perfectly reasonable under "normal" conditions. *But a fixed basket index is not particularly relevant when many commodities in the fixed basket are simply not available.* The fixed basket that is implicit in the use of the Lowe index will no longer provide an adequate approximation to actual consumption in period 1. A main message of this article is that information on actual pandemic expenditure patterns is needed so that "true" Laspeyres, Paasche and Fisher price indexes for the CPI going from pre-pandemic conditions to pandemic conditions can be computed. New estimates of current household expenditures by elementary category are required in order to measure inflation more accurately in the pandemic periods; the old basket weights are almost surely not accurate, even for categories that were not locked down. This point has been demonstrated by recent papers on how consumer expenditures on retail goods have changed due to pandemic conditions (see Cavallo 2020; Carvalho et al. 2020; Dunn et al. 2020).

We also encounter an additional significant problem in measuring consumer inflation going from the last pre-lockdown period to the first lockdown period using the Laspeyres or Lowe indexes: *what do we use for prices for the products that are no longer available due to lockdown conditions*? We address this issue in the next section.

3. Alternative Indexes and Inflation Adjusted Carry Forward Prices

Before we address the above question, it will be useful to develop the algebra for the Laspeyres, Paasche and Lowe price indexes between the pre-lockdown period 0 and a subsequent post-lockdown period 1.

We consider CPI goods and services as belonging to one of two groups: Group 1 prices and quantities are available in periods 0 and 1 and Group 2 prices and quantities are only available in period 0. (instead of considering the entire CPI, our discussion can apply to a subset of the CPI. A price index constructed over a subset of the CPI is called an elementary index. If an entire category of consumer expenditures disappears in period 1, this becomes a Group 2 category. The Group 1 category is chosen to be a category that is most closely related to the Group 2 category). Group 1 products have price and quantity vectors denoted by $p^t \equiv [p_1^t, \ldots, p_M^t] >> 0_M$ and $q^t \equiv [q_1^t, \ldots, q_M^t] >> 0_M$, respectively, for periods $t = 0, 1$. Notation: $p^t >> 0_M (\geq 0_M)$ means that all components of the M dimensional vector p^t are positive (nonnegative). The inner product of the vectors p^t and q^t is defined as $p^t \cdot q^t \equiv \Sigma_{m=1}^M p_m^t q_m^t$. The Group 2 price and quantity vectors for period 0 are $P^0 \equiv [P_1^0, \ldots, P_N^0] >> 0_N$ and $Q^0 \equiv [Q_1^0, \ldots, Q_N^0] >> 0_N$. The Group 2 quantity vector for period 1 is a vector of zero components, so that $Q^1 \equiv 0_N$. The corresponding vector of imputed prices is denoted $P^{1*} \equiv [P_1^{1*}, \ldots, P_N^{1*}] >> 0_N$ where $P_n^{1*} > 0$. It is unclear how to define the period 1 price vector P^{1*} for the products that are not available in period 1, where we use the word "product" to cover both goods and services. We will consider the implications of using standard recommended approaches for imputing these prices, but first we will define some index number formulae so it can be seen how these imputed prices come into play in index number construction.

The *Laspeyres* price index going from pre-lockdown period 0 to post-lockdown period 1, P_L^*, is defined as follows:

$$\begin{aligned} P_L^* &\equiv [p^1 \cdot q^0 + P^{1*} \cdot Q^0]/[p^0 \cdot q^0 + P^0 \cdot Q^0] \\ &= s_q^0 P_{Lq} + s_Q^0 P_{LQ}^* \end{aligned} \tag{1}$$

where the period 0 expenditure shares for always available commodities and unavailable commodities in period 1 are defined by

$$s_q^0 \equiv p^0 \cdot q^0/[p^0 \cdot q^0 + P^0 \cdot Q^0]; s_Q^0 \equiv P^0 \cdot Q^0/[p^0 \cdot q^0 + P^0 \cdot Q^0] \tag{2}$$

and the Laspeyres price indexes over always available commodities and unavailable commodities in period 1 are defined by P_{Lq} and P_{LQ} respectively:

$$P_{Lq} \equiv p^1 \cdot q^0 / p^0 \cdot q^0; P_{LQ}^* \equiv P^{1*} \cdot Q^0 / P^0 \cdot Q^0 \tag{3}$$

Recall that P^{1*} is a vector of imputed prices for the products that are no longer available in the market place in period 1. We put a superscript asterisk on P^1 to indicate that this price vector is not directly observable. Hence, since the Laspeyres index P_L* depends on the unobserved vector P^{1*}, we placed a superscript asterisk on the Laspeyres to indicate that these indexes depend on the unobserved price vector P^{1*}. In what follows, we also use this convention for other indexes that similarly depend on P^{1*}.

The *Paasche* and *Lowe* price indexes going from pre-lockdown period 0 to post-lockdown period 1, P_P and P_B*, are defined as follows:

$$\begin{aligned} P_P &\equiv [p^1 \cdot q^1 + P^{1*} \cdot Q^1] / [p^0 \cdot q^1 + P^0 \cdot Q^1] \\ &= p^1 \cdot q^1 / p^0 \cdot q^1 \qquad \text{since } Q^1 = 0_N; \\ &\equiv P_{Pq}; \end{aligned} \tag{4}$$

$$\begin{aligned} P_B^* &\equiv [p^1 \cdot q^b + P^{1*} \cdot Q^b] / [p^0 \cdot q^b + P^0 \cdot Q^b] \\ &= s_q^b P_{Bq} + s_Q^b P_{BQ}* \end{aligned} \tag{5}$$

where the *fixed basket Lowe subindexes* for continuing commodities and unavailable commodities, P_{Bq} and $P_{BQ}*$ are defined as follows:

$$P_{Bq} \equiv p^1 \cdot q^b / p^0 \cdot q^b; \quad P_{BQ}^* \equiv P^{1*} \cdot Q^b / P^0 \cdot Q^b. \tag{6}$$

The *base period hybrid shares* (prices of period 0 but quantities for a prior year b) for the continuing and disappearing commodity groups, s_q^b and s_Q^b, are defined as follows:

$$s_q^b \equiv p^0 \cdot q^b / [p^0 \cdot q^b + P^0 \cdot Q^b]; \quad s_Q^b \equiv P^0 \cdot Q^b / [p^0 \cdot q^b + P^0 \cdot Q^b]. \tag{7}$$

Note that the overall Paasche price index going from the pre-lockdown period 0 to the post-lockdown period 1, P_P, defined by the first equation in Equation (4) turns out to equal the Paasche subindex, P_{Pq}, that uses only the prices for the commodities that are available in periods 0 and 1. Thus the "true" overall Paasche price index is equal to the Paasche price index for always available commodities (and thus both indexes are observable in principle).

The situation is different for the overall Laspeyres and Lowe price indexes, P_L* defined by Equation (1) and P_B* defined by Equation (5); *these indexes require estimates for the product prices that are missing in period 1*; that is, we require an estimate for the vector of period 1 prices, P^{1*}, in order to calculate these indexes.

What is the "right" price for missing product n in period 1, P_n^{1*}? At first glance, one could argue that the "right" price is $P_n^{1*} \equiv +\infty$; that is, it is impossible to purchase product n in period at any finite price so a market price that will ensure that no one will purchase product n in period 1 is a price that is infinitely high. However, this is where economic theory can play a role. Normally, market prices are determined by the intersection of supply and demand curves. In any given period, the intersection of these curves determines P_n^1 and Q_n^1. But lockdowns of some parts of the economy mean that some products n simply become unavailable in period 1. What has happened is that the *supply curve* for product n has become a straight line that is parallel to the price axis and this line has shifted to become identical to the vertical price axis. Thus the (unobserved) market price for the product n under consideration is the price where the demand curve intersects the vertical price axis; this determines P_n^{1*} conceptually. This price will typically be less than $+\infty$; that is it does not take an infinite price to deter households from purchasing any particular product. This intersection price P_n^{1*} can be interpreted as a *Hicksian reservation price* in a Cost of Living Index (COLI) context. Hicks (1940, 114) introduced the concept of a reservation price into the economics literature and Von Hofsten (1952) introduced the term. Reservation prices have been widely used by economists to measure the benefits of new products since the pioneering work of Feenstra (1994). Reservation prices can be measured retrospectively; see Hausman (1996, 1999, 2003) and Diewert and Feenstra (2019) for empirical examples. (In addition to government mandated lockdowns, the unavailability of a commodity may be due to a lack of demand which causes supplying firms to shut down. This is where the difference between a COLI and a COGI comes into play. If the lack of supply for a commodity is due to a shift in consumer preferences (due to safety concerns), then from a COLI point of view, we need to compute two (conditional) Cost of Living Indexes: one that uses the preferences of the prepandemic period and one that uses the preferences of the post pandemic period. From a COGI point of view, changing preferences are irrelevant: what matters is the intersection of aggregate demand and aggregate supply curves to determine the market prices and quantities which should be used in a COGI. Thus we interpret the appropriate COGI price for an unavailable product to be the price where the relevant market demand curve intersects the price axis. However the determination of this unobserved market price necessarily involves some econometric modeling and hence we cannot expect statistical agencies to be able to produce estimates for these prices in real time).

Below, we will refer to an index which uses reservation prices for the missing prices as a "true" index. In the literature on the economic approach to index number theory, it is common to refer to a COLI as the "true" index. This convention dates back to a paper written by Robert Pollak for the U.S. Bureau of Labor Statistics in 1971, which was reprinted as Pollak (1983).

Currently, it is unlikely that national statistical offices can estimate such reservation prices, at least not in a timely fashion using presently available techniques. We outline below how the algebra for the Laspeyres and Lowe indexes works if inflation adjusted carry forward prices are used instead.

Define the *inflation adjusted carry forward prices* for the missing products in period 1, P^{1L}, using the Laspeyres index for always present products, P_{Lq}, as the inflation index, as follows. National Statistical Agencies do not in general use P_{Lq} as the carry forward inflation:

$$P^{1\mathrm{L}} \equiv P_{Lq}P^0 = (p^1 \cdot q^0 / p^0 \cdot q^0)P^0. \tag{8}$$

National Statistical Agencies do not in general use P_{Lq} as the carry forward inflation index; they use a wide variety of alternative indexes such as a single price ratio of a close substitute product for the missing product. See Eurostat (2020), IMF (2020), UNECE (2020) and IWGPS (2020) for lists of possible alternative indexes. These imputed prices can be used as replacement prices for the reservation prices P^{1*} in definition (1) for the overall Laspeyres price index. We called the resulting index P_L^{CL}, which is a Laspeyres inflation adjusted carry forward index using the Laspeyres index P_{Lq} as the adjusting inflation index. Thus we have:

$$
\begin{aligned}
P_L^{CL} &\equiv [p^1 \cdot q^0 + P^{1L} \cdot Q^0]/[p^0 \cdot q^0 + P^0 \cdot Q^0] && \\
&= [p^1 \cdot q^0 + (p^1 \cdot q^0/p^0 \cdot q^0)P^0 \cdot Q^0]/[p^0 \cdot q^0 + P^0 \cdot Q^0] && \text{using Definition (8)} \\
&= P_{Lq} && \\
&= P_{Lq}(s_q^0 + s_Q^0) && \text{since } s_q^0 + s_Q^0 = 1 \qquad (9) \\
&= s_q^0 P_{Lq} + s_Q^0 P_{Lq} && \\
&< s_q^0 P_{Lq} + s_Q^0 P_{LQ}* && \text{if } P_{LQ}* > P_{Lq} \\
&= P_L* && \text{using Definition (1).}
\end{aligned}
$$

The above equations tell us two things:

1. The overall Laspeyres index that is generated by the use of the carry forward prices P^{1L} defined by Equation (8) gives rise to an index P_L^{CL} which turns out to be exactly equal to P_{Lq}, the Laspeyres price index that is restricted to always available products;
2. The inflation adjusted carry forward Laspeyres price index, P_L^{CL} defined by Equation (9), will be less than the "true" overall Laspeyres price index P_L* provided that the "true" Laspeyres index defined over the Group 2 products, $P_{LQ}*$, is greater than the Laspeyres index defined over always available products, P_{Lq}.

A sufficient condition for $P_{LQ}* > P_{Lq}$ is the following one:

$$P_n^{1*} > P_{Lq} P_n^0; \qquad n = 1, \ldots, N. \qquad (10)$$

Conditions (10) are that the period 1 reservation prices for unavailable products P_n^{1*} are greater than the corresponding inflation adjusted prices from period 1, $P_{Lq}P_n^0$, for each unavailable product $n = 1, \ldots, N$. This is likely to be the case.

However, it is possible that $P_n^{1*} = P_{Lq}P_n^0$ for some products n. In a simplified scenario where we have only one unavailable product and one always available product, this case occurs if the two products are *perfect substitutes*: that is, the consumer's utility function is linear in the two products. Put another way, in terms of our simple demand equals supply partial equilibrium approach to the determination of reservation prices, this perfect substitutes case corresponds to the case where the demand curve for product n is parallel to the Q_n axis. Thus as the supply curve shifts to a vertical straight line that coincides with the Pn axis, the price of the product remains constant after adjustment for general inflation in the always available goods and services. It is unlikely that this flat demand curve scenario could approximate actual consumer behavior during a pandemic; it is much more likely that demand curves are downward sloping and in this case, we will get a downward bias in the inflation adjusted carry forward Laspeyres index relative to the "true" Laspeyres index. As indicated earlier, it is common in the COLI literature to call a COLI the "true" index. But a Laspeyres index is not a COLI except under strong assumptions (of no substitution between products). Our "true" Laspeyres index is simply a Laspeyres index that uses our concept of (unobserved) market clearing prices for unavailable products. Similarly for the "true" Lowe index.

The algebra for the Lowe index is much the same as the above algebra for the true Laspeyres index. The "true" Lowe index was defined by Equations (5)–(7) above. The inflation adjusted carry forward prices P^{1I} for the missing period 1 products are now defined as follows:

$$P^{1I} \equiv P_{Bq}P^0 = (p^1 \cdot q^b / p^0 \cdot q^b)P^0. \tag{11}$$

Substitute the inflation adjusted carry forward prices P^{1I} defined by (11) into definition (5) in order to obtain the following approximation, P_B^{CI}, to the "true" Lowe index, P_B*:

$$\begin{aligned}
P_B^{CI} &\equiv [p^1 \cdot q^b + P^{1I} \cdot Q^b]/[p^0 \cdot q^b + P^0 \cdot Q^b] \\
&= [p^1 \cdot q^b + (p^1 \cdot q^b / p^0 \cdot q^b)P^0 \cdot Q^b]/[p^0 \cdot q^b + P^0 \cdot Q^b] && \text{using Definition (8)} \\
&= P_{Bq} \\
&= P_{Bq}(s_q^0 + s_Q^0) && \text{since } s_q^0 + s_Q^0 = 1 \qquad (12) \\
&= s_q^0 P_{Bq} + s_Q^0 P_{Bq} \\
&< s_q^0 P_{Bq} + s_Q^0 P_{BQ}* && \text{if } P_{BQ}* > P_{Bq} \\
&= P_B* && \text{using Definition (5).}
\end{aligned}$$

The above equations have the following implications:

- The overall Lowe index that is generated by the use of the carry forward prices P^{1I} defined by Equation (11) gives rise to an index P_B^{CI} which turns out to be exactly equal to P_{Bq}, the Lowe price index that is restricted to always available products;
- The inflation adjusted carry forward Lowe price index, P_B^{CI} defined by Equation (12), will be less than the "true" overall Lowe price index P_B* provided that the "true" Lowe index defined over the Group 2 products, $P_{BQ}*$, is greater than the Lowe index defined over always available products, P_{Bq}.

A sufficient condition for $P_{BQ}* > P_{Bq}$ is the following one:

$$P_n^{1*} > P_{Bq}P_n^0; \qquad n = 1, \ldots, N. \tag{13}$$

As before, we think that it is extremely likely that inflation adjusted carry forward prices $P_{Bq}P_n^0$ are well below the corresponding market clearing reservation prices P_n^{1*} and thus there is a very high probability that the inflation adjusted carry forward Lowe index P_B^{CI}

defined by Equation (12) understates our suggested definition of the "true" Lowe index P_B* defined by Equation (5).

We acknowledge that statistical offices will not be in a position to calculate satisfactory approximations to the "true" Lowe index, P_B*, which measures inflation going from a prelockdown period to a post-lockdown period. In which case, it seems reasonable that NSOs notify users of their data that the price index that they put out in the first lockdown period (some version of the Lowe index that uses some version of carry forward prices) is unlikely to be an accurate measure of inflation that is comparable to previous index values. (This suggestion is somewhat moot at this time of publication. However, we wrote the first version of this article in April 2020 when it was not clear what NSOs were going to do with their CPIs). The comparison will be somewhat accurate for the subindex that is restricted to always available products. The problem is the fact that the Lowe pre-lockdown basket weights, (q^b, Q^b), may not be close to the post-lockdown weights, $(q^1, 0_N)$.

A Lowe index will approximate a Laspeyres index and a Paasche index if the quantity weights do not change much going from period to period, so that $q^b \approx q^0 \approx q^1$. Then the Lowe subindex for continuing products, P_{Bq}, will provide an adequate approximation to the Laspeyres and Paasche subindexes, P_{Lq} and P_{Pq}. However, empirical evidence suggests that expenditure shares on food and other available products did not remain approximately constant in the first lockdown period so that q^1 was substantially different from q^0. Hence, once information on current period expenditures becomes available, it would be useful to compute Fisher subindexes over always available products on a retrospective basis for the pandemic affected periods. The retrospective index could be called an analytic index or a supplementary index. We suggest that the retrospective index be a Fisher index if possible (for levels of aggregation where current period price and expenditure information is available) because of the good test properties of the Fisher formula.

The above algebra applies to indexes that are calculated using prices and quantities (or equivalently, using expenditures and unit value prices). At higher levels of aggregation, the prices become elementary price indexes for commodity categories and the quantities become *volumes*; that is, they are expenditures deflated by the relevant price indexes. The above algebra applies in both situations.

Many countries use a Young index instead of the Lowe index as a target index in the production of their CPIs at higher levels of aggregation. Recall that the Lowe index made use of the base period quantity vectors $q^b \equiv [q_1^b, \ldots q_M^b]$ and $Q^b \equiv [Q_I^b, \ldots Q_N^b]$. In order to define the Young index, we need to make use of the base period expenditure share vectors, $s^b \equiv [s_1^b, \ldots s_M^b]$ and $S^b \equiv [S_1^b, \ldots S_N^b]$ where $\Sigma_{m=1}^M s_m^b + \Sigma_{n=1}^N S_n^b = 1$. Using the above notation for observed prices p_m^t and P_n^t in each period t, the "true" Young index is defined as follows:

$$P_{Yb}^* \equiv \Sigma_{m=1}^M s_m^b (p_m^1/p_m^0) + \Sigma_{n=1}^N S_n^b (P_n^{1*}/P_n^0). \tag{14}$$

It can be seen that the Young index is a member of the class of stochastic or descriptive statistics indexes. If the base period b happens to be period 0, then it can be seen that $P_{Yb}*$ becomes $P_{Y0}* = P_L*$, the "true" Laspeyres index defined earlier by (1). This choice for the base period b share weights leads to a *relevant* Young index. Another *relevant* choice for the period b expenditure shares are the expenditure shares of period 1. The resulting Young index is equal to P_{Y1} defined as follows:

$$P_{Y1} \equiv \Sigma_{m=1}^{M} s_m^1 (p_m^1/p_m^0) + \Sigma_{n=1}^{N} S_n^1 (P_n^{1*}/P_n^0) = \Sigma_{m=1}^{M} s_m^1 (p_m^1/p_m^0) \tag{15}$$

where the second equality in Equation (15) follows since the expenditure shares for the unavailable products S_n^1 are all equal to 0. The Young index P_{Y1} defined by Equation (15) is closely related to the overall Paasche index, P_p, and the Paasche index restricted to always available products, P_{p_q}, defined earlier by Equation (4); that is, we have the following equalities and inequality:

$$P_P = P_{Pq} = [\Sigma_{m=1}^{M} s_m^1 (p_m^1/p_m^0)^{-1}]^{-1} \le \Sigma_{m=1}^{M} s_m^1 (p_m^1/p_m^0) = P_{Y1} \tag{16}$$

where the inequality follows since a weighted harmonic mean is always equal to or less than the corresponding weighted arithmetic mean; see Hardy et al. (1934, 26). Typically, the gap between P_P and P_{Y1} will not be large.

$P_{Y0}*$ and P_{Y1} are equally relevant Young indexes which could be used to measure the price change that occurred between periods 0 and 1. P_{Y1} is also known as a Palgrave price index. Thus some form of averaging of these two indexes is called for if a single measure of inflation is desired. The geometric mean of $P_{Y0}*$ and P_{Y1} is a useful average that leads to an index which satisfies the important time reversal test. However, statistical agencies will not be able to calculate $P_{Y0}*$ in real time due to the difficulty in determining the market clearing unobserved prices P^{1*}. Thus to form a real time Young index, it will be necessary for statistical agencies to use some form of inflation adjusted carry forward prices.

Recall that the vector of inflation adjusted carry forward prices for the Laspeyres index, P^{1L}, was defined above by Equation (8): $P^{1L} \equiv P_{L_q} P^0 = (p^1 \cdot q^0 / p^1 \cdot q^0) P^0$. An analogous vector of *inflation adjusted carry forward prices for the missing products* in period 1 for the Young index $P_{Yb}*$ is P^{1Y} defined as follows:

$$P^{1Y} \equiv \{[\Sigma_{m=1}^{M} s_m^b (p_m^1/p_m^0)]/[\Sigma_{m=1}^{M} s_m^b]\} P^0. \tag{17}$$

Note that the inflation index used to adjust the base period prices P^0 into imputed prices for missing commodities in period 1 is $[\Sigma_{m=1}^{M} s_m^b (p_m^1/p_m^0)]/[\Sigma_{m=1}^{M} s_m^b]$ which is essentially a Young index for the always available commodities where the base period shares for always available products s_m^b have been reweighted so that they sum to one. Now substitute the imputed carry forward prices P^{1Y} defined Equation (17) into Definition (14) in order to obtain an *inflation adjusted carry forward Young index* P_{Yb}^{CF}:

$$
\begin{aligned}
P_{Yb}^{CF} &\equiv \Sigma_{m=1}^{M} s_m^b (p_m^1/p_m^0) + \Sigma_{n=1}^{N} S_n^b (P_n^{1Y}/P_n^0) \\
&= \Sigma_{m=1}^{M} s_m^b (p_m^1/p_m^0) + \{[\Sigma_{m=1}^{M} s_m^b (p_m^1/p_m^0)]/[\Sigma_{m=1}^{M} s_m^b]\}\{\Sigma_{n=1}^{N} S_n^b] \\
&= [\Sigma_{m=1}^{M} s_m^b (p_m^1/p_m^0)]\{1 + [\Sigma_{n=1}^{N} S_n^b]/[\Sigma_{m=1}^{M} s_m^b]\} \qquad (18) \\
&= [\Sigma_{m=1}^{M} s_m^b]^{-1} [\Sigma_{m=1}^{M} s_m^b (p_m^1/p_m^0)][\Sigma_{m=1}^{M} s_m^b + \Sigma_{n=1}^{N} S_n^b] \\
&= \Sigma_{m=1}^{M} s_m^b (p_m^1/p_m^0)/\Sigma_{m=1}^{M} s_m^b.
\end{aligned}
$$

Thus using the inflation adjusted carry forward prices defined by Equation (17) for the missing prices will cause the Young index P_{Yb}^{CF} to collapse down to the reweighted Young index that measures price change for just the always available products. If the "true" market clearing period 1 prices, P_n^{1*} are greater than the inflation adjusted prices $P_n^{1Y} \equiv \{[\Sigma_{m=1}^{M} s_m^b (p_m^1/p_m^0)]/[\Sigma_{m=1}^{M} S_m^b]\}P_n^0$ for $n = 1, \ldots, N$, then the "true" Young index will be greater than the inflation adjusted Young index defined by Equation (18).

4. The Way Forward

As we have noted, statistical agencies that use a fixed basket methodology for constructing their CPI are faced with the fact that the fixed basket is no longer as relevant for pandemic periods as it was in pre-pandemic times. Thus as we have indicated in the previous section, the use of Lowe or Young indexes in pandemic times that use pre-pandemic weights will not accurately reflect changes in the cost of purchasing goods and services during pandemic periods. However, many NSOs will not have the resources to estimate representative baskets in real time. We will list a number of possible strategies that an agency could use in order to construct a CPI under pandemic conditions, depending on what kind of data they are able to collect. We will start with the assumption that very little data are available and finish with the way forward if ample data are available. For each of these cases, we will look at possible ways of addressing the lack of matching problem at the elementary index level.

4.1. Case 1: Very Little Data Availability

For this case, we suppose that the agency has only a fixed basket (q^b, Q^b) along with price data for period 0 which is the period before the lockdown, (p^0, P^0). For the pandemic periods, the agency has only price data for always available goods and services, p^t for $t = 1, 2, \ldots \tau-1$, the pandemic periods. When the pandemic is over in period τ, we assume that the agency can collect price data for always available goods, p^τ, and for commodities that were available in period 0 and become available again in period τ, P^τ. For the lockdown periods, the agency can calculate the fixed basket index for always available commodities, $p^t \cdot q^b / p^0 \cdot q^b$ for $t = 1, 2, \ldots \tau-1$. These indexes may be suitable for (partial) compensation purposes; i.e., if period 0 household expenditures on the basket q^b were equal to $p^0 \cdot q^b$, then using the index $p^t \cdot q^b / p^0 \cdot q^b$ to escalate the household's period 0 "income" (equal to $p^0 \cdot q^b$) would allow the household to purchase the bundle of commodities q^b in period t for $t = 1, 2, \ldots \tau - 1$. This index may be subject to some substitution bias. From the COLI perspective, the NSO would need to note that the CPI for these periods is not comparable to the CPI for either period 0 or period τ. (Equation (12) show that the indexes $p^t \cdot q^b / p^0 \cdot q^b$ for $t = 1, 2, \ldots \tau - 1$ could be generated by using the inflation adjusted carry forward price vectors for unavailable commodities defined as $P^{tI} \equiv (p^t \cdot q^b / p^0 \cdot q^b) P^0$. If this is done, users need to be informed that the resulting indexes are not "true" fixed basket indexes in that part of the overall fixed basket, (q^b, Q^b), is simply not available for purchase in period t. A similar comment applies to NSOs using Young indexes). To provide a useful estimate for a cost of living index relative to the standard of living in period 0 for the lockdown periods, we require estimates for reservation prices, P^{t*} for $t = 1, 2, \ldots \tau - 1$. Very few NSOs will venture to estimate reservation prices. What NSOs can do is to provide a credible CPI for goods and services which are actually available during the lockdown periods. When the lockdown ends and conditions approach "normality" in period τ, then the under-resourced statistical office can use its pre-lockdown basket, (q^b, Q^b), to calculate the price level in period τ relative to period 0 as the fixed base index $[p^\tau \cdot q^b + P^\tau \cdot Q^b]/[p^0 \cdot q^b + P^0 \cdot Q^b]$.

4.2. *Case 2: Some Data Availability*

We assume that the data availability is at least as good as in the above case. In addition, we assume that by period σ (where $1 < \sigma < \tau$), the statistical agency is able to obtain an estimate for a representative quantity vector q^σ for the always available quantities during the lockdown period. For the lockdown periods prior to period σ, the agency can calculate the fixed basket index for always available commodities, $\pi^t \equiv p^t \cdot q^b / p^0 \cdot q^b$ for $t = 1, 2, \ldots \sigma - 1$. In period σ, the new basket q^σ becomes available so it is possible to calculate the period t price index value for period t as $\pi^t \equiv \pi^{t-1}[p^t \cdot q^\sigma / p^{t-1} \cdot q^\sigma]$ for $t = \sigma, \sigma + 1, \ldots, \tau - 1$. However, the price levels $\pi^1, \pi^2, \ldots, \pi^{\sigma-1}$ may very unreliable due to the fact that the pre-lockdown quantity vector q^{b0} may be rather far from the actual consumption vectors $q^1, q^2, \ldots, q^{\sigma-1}$ over the lockdown period extending from period 1 to period $\sigma - 1$. Thus it may be preferable to define π^σ as the pseudo Fisher index comparing period 0 with period σ; that is, define $\pi^\sigma \equiv [p^\sigma \cdot q^b / p^0 \cdot q^b]^{1/2} [p^\sigma \cdot q^\sigma / p^0 \cdot q^\sigma]^{1/2}$. For lockdown periods following period σ but prior to period τ, define $\pi^t \equiv \pi^{t-1}[p^t \cdot q^\sigma / p^{t-1} \cdot q^\sigma]$ for $t = \sigma + 1, \sigma + 2, \ldots, \tau - 1$. Again, these indexes may be suitable for partial indexation purposes but they are likely to substantially understate "true" COLI-type inflation relative to the pre-pandemic period. When we get to period τ, the moderately-resourced statistical office can calculate the fixed base index relative to period 0; that is, set $\pi^\tau = [p^\tau \cdot q^b + P^\tau \cdot Q^b]/[p^0 \cdot q^b + P^0 \cdot Q^b]$.

If the statistical office has set in motion a continuous consumer expenditure survey so that a new period τ comprehensive basket (q^τ, Q^τ) can be constructed, then the office can calculate the pseudo Fisher indexes defined above. If the office has access to scanner data for some strata, then Fisher indexes can be calculated for those strata.

4.3. *Case 3: Ample Data Availability*

We assume the Case 1 data availability plus the availability of representative quantity vectors q^{bt} for all periods $t = 0, 1, \ldots, \tau$. We also assume that a representative quantity vector for the unavailable commodities is available for periods 0 and τ. Denote these vectors by Q^{b0} and $Q^{b\tau}$. The corresponding price vectors are P^0 and P^1. For period 0, define the price level as $\pi^0 \equiv 1$. For the lockdown periods, define the period 1 price index π^1 as the pseudo Fisher index $\pi^1 \equiv \{[p^1 \cdot q^{b0}/p^0 \cdot q^{b0}][p^1 \cdot q^{b1}/p^0 \cdot q^{b1}]\}^{1/2}$. For $t = 2, 3, \ldots \tau-1$ define the period t price index as $\pi^t \equiv \pi^{t-1}\{[p^t \cdot q^{bt-1}/p^{t-1} \cdot q^{bt-1}] \times [p^t \cdot q^{bt}/p^{t-1} \cdot q^{bt}]\}^{1/2}$. Thus the period to period pseudo Fisher indexes are chained together to form the period t price level. For period τ, define the price level π^τ as the comprehensive pseudo Fisher price index connecting period 0 to period τ; that is, define π^τ as follows:

$$\pi^\tau \equiv \{[p^\tau \cdot q^{b0} + P^\tau \cdot Q^{b0}]/[p^0 \cdot q^{b0} + P^0 \cdot Q^{b0}]\}^{1/2}\{[p^\tau \cdot q^{b\tau} + P^\tau \cdot Q^{b\tau}]/[p^0 \cdot q^{b\tau} + P^0 \cdot Q^{b\tau}]\}^{1/2}. \tag{19}$$

The reason for using chained pseudo Fisher price indexes for the available products during the lockdown period instead of fixed base pseudo Fisher price indexes is the likelihood that consumer purchases of available products over the lockdown periods may not be well approximated by a constant vector q^b. Initially, households will stock up on storable goods and cut back on purchases of consumer durables. If the lockdown period is long and the degree of lockdown varies, then it is quite likely that the vector of actual purchases of available commodities in period t, q^t, will be quite variable and hence a constant q^b will not provide a representative vector of household purchases over all of the lockdown periods. Note that the set of available products has varied over lockdown periods. In the early stages, food manufacturers did not produce their full line of products; they concentrated on increasing the volume of their best selling products and produced them at scale to satisfy stockpiling demands. Of course, the gold standard for the quantity vectors q^{bt} would be the actual period t consumption vectors, q^t, in which case, the pseudo Fisher indexes would become actual Fisher indexes.

5. The Lack of Matching Problem at the Elementary Index Level

A problem which has appeared as a result of country wide lockdowns is the *problem of missing products and services* in retail outlets. In many cases, the missing products and services reappeared in a later period; in some cases, they were gone for the duration of the lockdown. If the products are gone for the duration of the lockdown and the remaining products are present during the current and prior lockdown periods, then we are in position to apply the theory above to the particular elementary aggregate under consideration; that is, we need to switch from pricing out the pre-lockout basket of products to the new smaller set of products. However, real life will be more complicated than having a clear division between products present and products that have been discontinued for all lockout periods: products will be drifting in and out of scope in any particular retail outlet. This may lead to a massive lack of matching problem. We will briefly suggest possible solutions to this problem under two scenarios: (1) only web scraped data are available and (2) scanner data are available. The analysis in this section differs from the analysis that was presented in the previous section where it was known that some commodities would be unavailable for the duration of the lockdown. We now assume that the full array of pre-lockdown products is not available in the lockdown periods.

5.1. *Case 1: Only Price Data Are Available*

5.1.1. Method 1: Adapt the Section 3 Carry Forward Methodology

The adaptation here is to assume that q^0 and q^1 are equal to the vector of ones, 1_M, and Q^0 equals the vector of ones, $1_N.Q^1 = 0_N$ as in Section 4. Thus the q group of products are the *maximum overlap products* that are present in both periods and the Q products are present in the base period 0 but not in the current period 1. The given price vectors are p^0, p^1 and P^0. Applying the Section 4 methodology using the above assumptions on prices and quantities leads to the following inflation adjusted carry forward price index using Equation (12) adapted to the present situation:

$$P_B^{CI} = P_{Bq} = p^1 \cdot 1_M / p^0 \cdot 1_M = \Sigma_{m=1}^M p_m^1 / \Sigma_{m=1}^M p_m^0. \tag{20}$$

The above index is the Dutot (1738) elementary index, defined over products that are present in both periods. It has an undesirable property: it is not invariant to changes in the units of measurement of the products. It will also give a higher weight to products that are more expensive which may not be a desirable property. Nevertheless, it does approximate the theoretically more desirable Jevons index under certain conditions (Diewert 2021b).

It is a standard practice to use a sample of prices to represent price movements for the universe of commodities in an elementary category of transactions. This can be problematic during a pandemic. For example, during the pandemic, international travel by airlines fell dramatically but some flights still took place. Thus statistical agencies could use the price movements for the existing flights to represent the movement of prices over the entire air travel universe. But this practice disguises the fact that international air travel for most households was shut down for long periods of time. Thus the air travel category should be split into at least two categories in a CPI: one category for some households who are allowed to travel internationally and another category for the locked down households. The types of "bias" that we discussed in Section 3 are applicable to this situation. This potential bias problem is present whenever a sample of prices is used to represent movements in the universe of prices in scope. However, the "bias" will usually be small (from the viewpoint of the economic approach to index number theory) if the commodities in the category are highly substitutable with each other. In this case, consumers can easily substitute towards the available varieties when some varieties disappear without much overall loss of utility (Diewert 2021d). The pandemic situation is very different: the pandemic induced disappearance of many commodities surely led to large losses in utility, which were not measured. There is little that statistical offices can do to remedy this situation but it seems reasonable for NSOs to flag this problem.

5.1.2. Method 2: Use Maximum Overlap Jevons Indexes

This method simply sets the price index equal to the Jevons (1865) index for the overlapping products in the two periods under consideration. Thus using the same notation as was used to describe Method 1 above, the maximum overlap Jevons index, P_{JMO}, is equal to the geometric mean of the price ratios for the overlapping products:

$$P_{JMO} \equiv [\Pi_{m=1}^{M}(p_m^1/p_m^0)]^{1/M}. \tag{21}$$

The Jevons index has the best axiomatic properties for indexes (with no missing prices) that depend only on prices. Note in particular that the maximum overlap Jevons index is invariant to changes in the units of measurement for the products (Diewert 2021b).

5.1.3. Method 3: Use the Multilateral Time Product Dummy Method

A problem with the above two methods is that they make use of price data covering only two periods. In the situation where closely related products are moving in and out of scope, constructing maximum overlap bilateral index numbers does not make use of all of the data and hence is inefficient from a statistical point of view. For example, suppose a product is present in periods 1 and 3 and another product is present in periods 2 and 4. In a bilateral index setup, the information pertaining to these two products would not be used which is inefficient since price comparisons for product 1 between periods 1 and 3 and for product between periods 2 and 4 are perfectly valid comparisons and should be used somehow in constructing the sequence of price indexes. The way forward here is to use a *multilateral index* which utilizes the price information for all periods. For studies on the use of multilateral indexes in the time series context (see Balk 1980; Ivancic et al. 2011; De Haan and Van der Grient 2011). For a detailed discussion of these methods, see Diewert (2021c).

A widely used multilateral method is the *Time Product Dummy Method.* The method was originally devised for making price comparisons across countries and is known as the Country Product Dummy multilateral method; see Summers (1973). A weighted version of this model (with missing observations) was first applied in the time series context by Aizcorbe et al. (2000). The method can be interpreted as a special case of a hedonic regression model (see De Haan 2004, 2010; De Haan and Krsinich 2014, 2018).

We introduce some new notation in order to describe this method. We now assume that there are N products and T time periods but not all products are purchased (or sold) in all time periods. The price and quantity vectors for period t are denoted by $p^t = [p_{t1},\ldots,p_{tN}]$ and $q^t = [q_{t1},\ldots,q_{tN}]$. If product n in period t is missing, we set the corresponding price and quantity, p_{tn} and q_{tn}, equal to 0. For each period t, define the set of products n that are present in period t as $S(t) = \{n : p_{tn} > 0\}$ for $t = 1,2,\ldots,T$. It is assumed that these sets are not empty; i.e., at least one product is purchased in each period. For each product n, define the set of periods t where product n is present as $S^*(n) = \{t : p_{tn} > 0\}$. Again, assume that these sets are not empty; that is, each product is sold in at least one time period. Define the integers $N(t)$ and $T(n)$ as follows:

$$N(t) \equiv \Sigma_{n\in S(t)}1; \qquad t = 1,\ldots,T; \tag{22}$$

$$T(n) \equiv \Sigma_{t\in S^*(n)}1; \qquad n = 1,\ldots,N. \tag{23}$$

If all N products are present in period t, then $N(t) = N$; if product n is present in all T periods, then $T(n) = T$.

The economic model that is consistent with the Time Dummy Product multilateral method is the following one:

$$p_{tn} = \pi_t \alpha_n; \qquad t = 1, \ldots, T; \; n \in S(t) \tag{24}$$

where π_t is the period t price level and α_n is a quality adjustment parameter for product n. If all products were available in all periods, Equation (24) tell us that prices for the group of products in scope are moving in a proportional manner. This is consistent with purchasers of the N products having the linear utility function, $f(q) = \alpha \cdot q \equiv \Sigma_{n=1}^{N} \alpha_n q_n$ where $\alpha \equiv [\alpha_1, \ldots, \alpha_N]$ and $q \equiv [q_1, \ldots, q_N]$. It can be seen that this approach will only be adequate if the products are very close substitutes since a linear utility function implies that the products are perfect substitutes; see Diewert (2021d) for further explanation of the underlying economic model.

Now take logarithms of both sides of Equation (24), add error terms etn to the resulting equations and we obtain the following system of estimating equations:

$$lnp_{tn} = \rho_t + \beta_n + e_{tn}; \qquad t = 1, \ldots, T; \; n \in S(t) \tag{25}$$

where $\rho_t \equiv ln\pi_t$ for $t = 1, \ldots, T$ and $\beta_n \equiv ln\alpha_n$ for $n = 1, \ldots, N$. Note that Equation (25) form the basis for the *time dummy hedonic regression model* . This is Court's (1939, 109–111) hedonic suggestion number two. He chose to transform Equation (24) by the log transformation because the resulting regression model fit his data on automobiles better. Diewert (2003) also recommended the log transformation on the grounds that multiplicative errors were more plausible than additive errors.

Estimates for the unknown parameters ρ_t and β_n that appear in Equation (25) can be found by solving the following least squares minimization problem:

$$min_{\rho,\beta}\{\Sigma_{t=1}^{T} \Sigma_{n \in S(t)} \; [lnp_{tn} - \rho_t - \beta_n]^2\}. \tag{26}$$

In order to obtain a unique solution to Equation (26), we need to impose a full rank condition on the X matrix generated by the linear regression model defined by Equation (25) and $\rho_1 = 0$ (Diewert 2021c), and impose a normalization on the parameters. Choose the normalization $\rho_1 = 0$ (which corresponds to $\pi_1 = 1$). Denote the resulting solution by $\rho^* \equiv [1, \rho_2^*, \ldots, \rho_T^*]$ and $\beta^* \equiv [\beta_1^*, \ldots, \beta_N^*]$. Use these estimates to form estimates for $\pi_t^* \equiv exp[\rho_t^*]$ for $t = 1, \ldots, T$ and $\alpha_n^* \equiv exp[\beta_n^*]$ for $n = 1, \ldots, N$. It turns out that these estimates satisfy the following equations:

$$\pi_t^* = \Pi_{n \in S(t)}[p_{tn}/\alpha_n^*]^{1/N(t)}; \qquad t = 1, \ldots, T; \tag{27}$$

$$\alpha_n^* = \Pi_{t \in S^*(n)}[p_{tn}/\pi_t^*]^{1/T(n)}; \qquad n = 1, \ldots, N. \tag{28}$$

Note that p_{tn}/α_n^* is a *quality adjusted price* for product n in period t and p_{tn}/π_t^* is the corresponding *inflation adjusted price* for product n in period t. Thus the period t estimated price level, π_t^*, is the geometric mean of the quality adjusted prices for products that are available in period t and the estimated quality adjustment factor for product n, α_n^*, is the geometric mean of all of the inflation adjusted prices for product n over all periods. Note that if the set of available products in periods r and t is the same, then $\pi_t^*/\pi_r^* = [\Pi_{n \in S(t)}(p_{tn}/p_{rn})]^{1/N(t)}$ which is the Jevons index defined over the products that are present in both periods. These price levels generated by this method have satisfactory axiomatic properties; see Diewert (2021c). It turns out that the price levels satisfy an *identity test* so if prices are equal in periods r and t, then $\pi_r{*} = \pi_t{*}$. There are some additional choices that the statistical agency will have to make if it uses this method; that is, necessary to decide on the length of the window of observations T and it is necessary to decide on how to link the results of the latest window of estimates with the previous window of estimates for the price levels. The agency should be able to resolve these issues by experimenting with the different choices for the window length and for linking the price level estimates for a new window to the estimates of the previous window.

5.2. *Case 2: Price and Quantity Data Are Available*

5.2.1. Method 4: Apply the Section 4 Carry Forward Methodology

Little additional explanation is required here; just apply the methodology explained in Section 3 to the elementary index context. Diewert et al. (2018) have more details on how to apply the carry forward methodology for Paasche, Laspeyres, Fisher and Törnqvist indexes in the case of two observations.

5.2.2. Method 5: Apply the Weighted Time Product Dummy Multilateral Method

The basic economic model is still the price proportionality model defined by Equation (19) above but now we assume that we have expenditure or quantity information on household purchases in addition to price information. With this extra information, it is preferable to take the economic importance of each commodity into account and replace the least squares minimization problem defined by Equation (26) with the following weighted least squares minimization problem.

$$min_{\rho,\beta}\{\Sigma_{t=1}^{T}\Sigma_{n\in S(t)}s_{tn}[lnp_{tn}-\rho_t-\beta_n]^2\} \tag{29}$$

where the period t expenditure share on commodity n is $s_{tn} \equiv p_{tn}q_{tn}/p^t \cdot q^t$ for $t = 1,\ldots,T$ and $n \in S(t)$; see Diewert (2021c) for a discussion on the merits of different choices for the weights. (Rao (1995, 2004, 2005, 574) was the first to consider this model using expenditure share weights. However, Balk (1980, 70) suggested this class of models much earlier using different weights. See also De Haan and Krsinich (2012, 2014) and Diewert and Fox (2018). Again, we need to make the normalization $\rho_1 = 0$ to obtain a unique solution ρ^* and β^* to Equation (29). It turns out the solution will satisfy the following equations, which are the weighted counterparts to Equations (27) and (28) (Diewert 2021c):

$$\pi_t^* = exp[\Sigma_{n\in S(t)}s_{tn}ln(p_{tn}/\alpha_n^*)]; \qquad t = 1,\ldots,T; \tag{30}$$

$$\alpha_n^* = exp[\Sigma_{t\in S^*(n)}s_{tn}ln(p_{tn}/\pi_t^*)/\Sigma_{t\in S^*(n)}s_{tn}]; \qquad n = 1,\ldots,N. \tag{31}$$

From Equations (30) and (31), it can be seen that the period t estimated price level, π_t^*, is now a weighted geometric mean of the quality adjusted prices for products that are available in period t and the estimated quality adjustment factor for product n, α_n^*, is now a weighed geometric mean of all of the inflation adjusted prices for product n over all periods. Note that if the set of available products in periods r and t is the same, $\pi_t{*}/\pi_r{*}$ will not collapse to a weighted Jevons index unless the expenditure shares in the two periods under consideration are equal.

Once the estimates for the π_t^* and α_n^* have been computed, we have two methods for constructing period by period aggregate price and quantity (or volume) levels, P^t and Q^t for $t = 1,...,T$. The way to see this is to consider the underlying Equation (24) which were the equations $p_{tn} = \pi_t\alpha_n$ for $t = 1,...,T$ and $n \in S(t)$. Take this equation for some n and t and multiply both sides of it by the observed quantity, q_{tn}, and sum the resulting equations. We obtain the following equations using the fact that $q_{tn} = p_{tn} \equiv 0$ for $n \notin S(t)$:

$$\begin{aligned} p^t \cdot q^t &= \Sigma_{n\in S(t)}\ p_{tn}q_{tn} && t = 1,...,T \\ &= \pi_t \Sigma_{n\in S(t)}\ \alpha_n q_{tn} \\ &= \pi_t \Sigma_{n=1}^{N}\ \alpha_n q_{tn} && \text{since } q_{tn} = 0 \text{ if n does not belong to } S(t) \\ &= \pi_t \alpha \cdot q^t. \end{aligned} \tag{32}$$

Because Equation (24) will not hold exactly, with nonzero errors e_{tn}, Equation (32) will only hold approximately. However, the approximate version of Equation (32) allow us to form period t price and quantity aggregate levels, say P^t and Q^t, in two separate ways: the π_t^* estimates that are part of the solution to Equation (29) can be used to form P^{t*} and Q^{t*} via Equation (33) and the α_n^* estimates that are part of the solution to Equation (29) can be used to form the aggregates P^{t**} and Q^{t**} via Equation (34):

$$P^{t*} \equiv \pi_t^*; Q^{t*} \equiv p^t \cdot q^t / \pi_t^*; \qquad t = 1,...,T; \tag{33}$$

$$Q^{t**} \equiv \alpha^* \cdot q^t; P^{t**} \equiv p^t \cdot q^t / \alpha^* \cdot q^t; \qquad t = 1,...,T. \tag{34}$$

Define the error terms $e_{tn} \equiv lnp_{tn} - ln\pi^* - ln\alpha_n^*$ for $t = 1,...,T$ and $n = 1,...,N$. If all $e_{tn} = 0$, then P^{t*} will equal P^{t**} and Q^{t*} will equal Q^{t**} for $t = 1,...,T$. However, if the error terms are not all equal to zero, then the statistical agency will have to decide on pragmatic grounds on which option to choose to form the aggregate price and quantity levels. De Haan and Krsinich (2018) were the first to realize that the results of a hedonic regression would lead to two separate ways to define the resulting aggregate price and quantity levels. See also Diewert (2020c, 2020d). If the accurate measurement of price levels is the target, then it is probably best to use P^{t*}; if the target is to measure aggregate quantity levels (and hence welfare), then it is probably best to use P^{t**}.

It should be noted that $P^{t**} \equiv p^t \cdot q^t / \alpha^* q^t$ is a *quality adjusted unit value price level*. The term "quality adjusted unit value price index" was introduced by Dalén (2001). Its properties were further studied by De Haan (2004, 2010), Silver (2010, 2011), De Haan and Krsinich (2018), Von Auer (2014) and Diewert (2020c, 2020d). There is also an inequality between P^{t*} and P^{t**} that is due to De Haan and Krsinich (2018, 763).

From Equations (30) and (33), we have $P^{t*} = exp[\Sigma_{n\in S(t)} s_{tn} ln(p_{tn}/\alpha_n^*)]$, which is a share weighted geometric mean of the period t quality adjusted prices, p_{tn}/α_n^*, for products that are actually present in period t. From Equation (34), we have P^{t**} equal to the following expression:

$$\begin{aligned} P^{t**} &\equiv p^t \cdot q^t / \alpha^* \cdot q^t \qquad t = 1, \ldots, T \\ &= \Sigma_{n\in S(t)} p_{tn} q_{tn} / \Sigma_{n\in S(t)}\ \alpha_n^* q_{tn} \\ &= \Sigma_{n\in S(t)} p_{tn} q_{tn} / \Sigma_{n\in S(t)}\ \alpha_n^* (p_{tn})^{-1} p_{tn} q_{tn} \\ &= [\Sigma_{n\in S(t)} s_{tn} (p_{tn}/\alpha_n^*)^{-1}]^{-1} \\ &\leq P^{t*} \end{aligned} \tag{35}$$

since a share weighted harmonic mean of the quality adjusted prices present in period t is always equal to or less than the corresponding share weighted geometric mean using Schlömilch's inequality (see Hardy et al. 1934, 26). Note that $P^{t**} \leq P^{t*}$ implies that $Q^{t**} \geq Q^{t*}$ for $t = 1, \ldots, T$.

The axiomatic properties of the price levels π_t^* are studied in Diewert (2021c). They are reasonably good.

The issues of choosing a window length T for this multilateral method remain unresolved; statistical agencies can experiment with different choices for T. There is also the issue of linking the present window with the previous window.

From the viewpoint of the economic approach to index number theory, the use of this method should be confined to situations where the products in scope are close substitutes since the underlying economic assumption is that the products are perfect substitutes, except for random errors. Quality adjusted unit value price levels are appropriate in this situation but if the products are not close substitutes, it would be preferable to use the inflation adjusted carry forward prices methodology suggested by Diewert et al. (2018) if the target index is a superlative index. Finally, Method 5 should not be used at higher levels of aggregation where substitution between elementary index categories may be low. At the second stage of aggregation it would be preferable to use Fisher, Walsh or Törnqvist indexes if actual price and quantity data are available or use pseudo Fisher indexes if the quantity data can only be approximated.

5.2.3. Method 6: The Use of Quality Adjusted Unit Value Price Levels

From the discussion of Method 5, it is clear that quality adjusted unit values can be used as price levels, provided that the commodities in scope for the elementary aggregate are close substitutes. However, it is not necessary to use the Weighted Time Product Dummy multilateral index number method in order to obtain estimates for the quality adjustment parameters, the components of the vector α. If the group of products under consideration consists of highly substitutable products and all of the products were purchased in the prelockdown period 0, then simply set α equal to p^0, the (unit value) price vector for the products in the pre-lockdown period. If all of the products were purchased for a number of pre-lockdown periods, say periods 0, -1, -2 and -3, and the price vectors for these periods were p^0, p^{-1}, p^{-2} and p^{-3}, then define α as follows:

$$\alpha \equiv (1/4)[(p_{01})^{-1}p^0 + (p_{-1,1})^{-1}p^{-1} + (p{-}2,1)^{-1}p^{-2} + (p{-}3,1)^{-1}p^{-3}]. \tag{36}$$

Thus a is set equal to the average of past pre-lockdown price vectors for the commodities in the group of commodities under consideration but these vectors of past prices are deflated by the price of the first commodity in order to eliminate the effects of general inflation between past periods for the group of commodities. The first commodity should be chosen to be the commodity with the largest average expenditure share in the group of commodities. If there are missing prices in the pre-lockdown periods, then instead of using the a defined by Equation (36), the α defined by the Time Product Dummy multilateral method (Method 3 above) could be used to estimate the quality adjustment parameters.

From the viewpoint of the economic approach to index number theory, the use of quality adjusted unit values as estimates for price levels should only be applied if the commodities in the elementary group of commodities are close substitutes. It is possible to cluster N highly substitutable commodities in scope into quality groups based on their price per unit of a dominant characteristic. Group the N products into low quality, medium quality and high quality products based on their relative prices in the pre-lockdown period. Then aggregate price levels for each of the three groups of products could be constructed by simply taking unit values (without quality adjustmet) for each group of products. We would end up with three elementary indexes in place of the single elementary index. Then these three separate indexes could be aggregated up into a single index using a superlative index number formula. This is feasible because we are assuming the availability of price and quantity data for Method 6. The advantage of this method is that it avoids the need for imputation.

5.2.4. Method 7: Linking Based on Relative Price and Quantity Similarity

A desirable property of the Fisher price index between two periods is the fact that the Fisher index will equal unity if prices in the two periods are equal even if the quantities demanded in the two periods are not equal. Most multilateral methods do not satisfy this strong identity test; they tend to satisfy a weaker identity test that says that the relative aggregate price levels between any two periods in the window of observations will equal unity provided that both prices *and* quantities are identical in the two periods being compared.

There is a recently developed multilateral method that satisfies the above strong identity test and can deal with missing observations. The method is based on building a set of Fisher index bilateral comparisons where each comparison is based on linking the periods that have the most *similar relative price structures.* Hill (2001, 2004) was an early pioneer in using this similarity of relative prices approach to multilateral index number theory in the time series context. The real time linking method described here is due to Diewert (2021c).

Initially, periods 1 and 2 are linked by the usual bilateral Fisher price index. When the data of period 3 become available, the price and quantity data of period 3 are linked to the corresponding data of either period 1 or 2, depending on which of these two periods has the most similar structure of relative prices. The bilateral Fisher index is used to link period with period 1 if the measure of relative price similarity between periods 1 and 3 is higher than the measure of relative price similarity between periods 1 and 2. If the measure of relative price similarity between periods 2 and 3 is higher than the corresponding measure for comparing periods 1 and 3, then the bilateral Fisher index is used to link period 3 with period 2. When the data of period 4 become available, the data for period 4 are linked to the data of periods 1,2 or 3, depending on which of these 3 prior periods gives rise to the highest measure of price similarity. And so on. In practice, measures of *relative price dissimilarity* are used to link the data of two periods, using the lowest measure of dissimilarity to do the linking. At the first stage of the network of comparisons, the two periods that have the most similar structure of relative prices is chosen. At the next stage of the comparison, look for a third period that had the most similar relative price structure to the first two periods and link in this third country to the comparisons of volume between the first two countries and so on.

A key aspect of this linking methodology is the choice of the measure of similarity (or dissimilarity) of the relative price structures of two periods. Various measures of the similarity or dissimilarity of relative price structures have been proposed by Allen and Diewert (1981), Kravis et al. (1982, 104–106), Hill (1997, 2001, 2004, 2009), Aten and Heston (2009) and Diewert (2009). The predicted share dissimilarity measure recently proposed by Diewert (2021c) seems to be the most promising but the method needs to be more thoroughly tested before it can be suggested to statistical agencies for general use. A major advantage of this new method of linking periods is that the strong identity test will always be satisfied; that is, if prices in the current period are the same as the prices in a past period, the estimated price levels pertaining to these two periods will always be identical even if quantities or expenditures are not identical. If the prices in the current period are proportional to the prices in a prior period, then the ratio of the current period price level to the prior period price level will be equal to the factor of proportionality. Another advantage of Diewert's method is that it is not necessary to choose a window length. There can never be a chain drift problem using this new multilateral method.

6. Other Measurement Problems

6.1. No Agency Employee Price Collection

Most statistical agencies stopped sending employees to retail outlets to collect prices during pandemic periods. Some agencies have switched to web scraping; that is, they collect online prices over the internet. The collected prices will not be perfectly comparable with the previously collected in store prices. Cavallo (2017) did a large scale comparison of in store prices versus online prices (excluding transport costs) across ten countries and found little difference between in store and online prices; online prices over the comparable in-store prices were on average 4% lower. The average markup ranged from −13% for Japan to +5% for Australia. See also Cavallo (2013) and Cavallo and Rigobon (2016). These results provide some justification for comparing a web scraped price for a specific product with a collected price for the same product in a prior period. Under lockdown conditions, home delivery of products purchased online increased dramatically. On the other hand, household travel expenses decreased due to fewer in store shopping trips. As these travel expenses are in scope for household expenditures, it may make sense to collect online prices that include delivery since the delivered price is the price that the consumer actually faces for the product. The higher price for the delivered product will be offset by lower household transportation costs. In general, we endorse the collection of web scraped data to replace previous data that were collected by agency employees. However, some care should be taken to not collect online prices for goods or services which were never actually consumed by any household. Examples of such services are listed airline fares or listed prepaid holiday packages that are eventually cancelled. (How exactly should cancellation fees be treated in a CPI?)

6.2. Lack of Information on Current Household Expenditure Weights

It will be very difficult for statistical agencies to find current period expenditure share or quantity weights for their elementary index categories. The problem is that the "representative" basket for each month is changing rapidly as the virus spreads and lockdown rules change to react to current conditions. Here are some possible ways for NSOs to obtain current information on household expenditures:

- Some countries (such as the United States and the UK) have continuous household expenditure surveys. Usually, the sample size for such surveys is small so, for example, the U.S. Bureau of Labor Statistics Consumer Expenditure Survey does not have a big enough sample size to allow monthly publication of the implied monthly weights. It publishes semi-annual estimates. The way forward here is to increase the sample size. For countries that currently do not have a continuous consumer expenditure survey, it is recommended that they start one. National governments will have to allocate extra resources to fund a continuous survey.
- Some private companies collect consumer expenditure data (along with prices and quantities) on a continuous basis for a sample of households using scanner data. NSOs can purchase these data (at a fair price) or set up their own competing company if they are unable to establish a satisfactory consumer expenditure survey.
- National governments can appeal to their business communities to persuade large firms producing consumer products to donate their electronic data to the NSO. Many countries, including Canada, have a Statistics Act which can be used to compel firms and households to provide information to NSOs. However, in general, NSOs are reluctant to use compulsion in order to obtain data. Many large retailers around the world are already donating their data and it should be possible for more firms to be persuaded to do this. This information will help to produce a better CPI and it will also allow much better production accounts to be produced.
- Credit card companies collect information on household purchases of consumer goods and services. If the expenditure information could also be combined with product codes, this information would enable the construction of consumer price indexes by location and demographic group. For some countries, it may be possible to access this information source. For other countries, it may not be possible for the statistical agency to access this information due to privacy concerns. See Carvalho et al. (2020) and Dunn et al. (2020) for examples of how such information can be used to analyze changes in expenditure patterns.

6.3. Should the CPI be Revised?

From Subsection 6.2, it can be seen that NSOs will not be able to produce very accurate period t basket updates q^{bt} that approximate actual period t consumption q^t in a timely fashion (if they can produce them at all). However, in time, better estimates for actual consumption in past periods may become available. Smoothing a sample of collected monthly household expenditures (by taking a moving average for example) will probably lead to more accurate trend estimates for monthly household expenditures, but the trend can only be calculated after some months have passed. The question then arises: should the CPI be revised in the light of improved information that becomes available after the release date? From a statistical point of view, the answer to this question is yes. However, for many countries, a monthly CPI must be provided to the public and no revisions are allowed.

Scanner data along with the usual information on retail sales can be massaged to produce some rough and ready weights in real time (Scanner data from retail outlets is not perfectly well suited for a CPI: retail outlets sell to tourists, foreign firms, governments and to domestic firms as well as to households. Thus scanner data collected directly from households is preferable). NSOs will simply have to announce that their new estimates for inflation and economic growth are only very approximate estimates. A country's national accounts are allowed to be revised and this revision process is generally accepted by the public, hence estimates of economic growth can be revised. This is not the case for the CPI. A country could produce at least two CPIs: one that is not revised and is based on available information at the month of production of the index and another that is allowed to be revised in the light of information that becomes available at a later date.

This strategy has been successfully used by the U.S. Bureau of Labor Statistics in the United States where two indexes are released at the same time; the first one ("CPI-U") is not revisable and the second one ("C-CPI-U") is allowed to be revised (and approximates a superlative index after the last revision). The second CPI can be labeled as an analytic CPI and can be used by economic analysts who require more accurate historical information on inflation. The first type of traditional CPI produced under lockdown conditions will necessarily be much more inaccurate; it will be very difficult to obtain adequate approximations to actual consumption during the start of the lockdown period due to the absence of accurate survey information on consumer expenditures. Users need to be alerted to this problem.

In Subsection 6.2 we attempted to anticipate the problems that many statistical agencies will face in trying to update their baskets to reflect the lockdown realities. We realize that new lockdown baskets will not be available to many, if not most, NSOs. Our conclusion boils down to this: if later information shows that the early lockdown indexes are very inaccurate, then set the current CPI price level to the best estimate possible even if it is necessary to use a different methodology than was used in the pre-lockdown periods. For the revisable CPI, new information should be used to revise previous indexes.

6.4. The Stockpiling Problem

Lockdowns have led governments to limit trips to retail outlets for purchases of food and other essential goods such as pharmaceutical products. These regulations plus the reactions of households to cut down on their shopping trips to limit the risk of infection have led households to accumulate large *stockpiles* of essential storable goods. Thus at the initial stages of a lockdown, there were large increase in purchases of storable goods but actual consumption of these goods was less. In other words, it becomes necessary to distinguish actual household *consumption* of storable goods from the *acquisition* of the goods. In principle, the national statistical agency will have to decide between these two approaches to the production of a CPI. The acquisitions approach is of course much more practical. In order to implement an actual consumption approach, the NSO would require a household inventory survey which would be costly.

From a welfare point of view, it is monthly consumption of goods and services which is most relevant but it will generally be more convenient to stick to an acquisitions approach to the measurement of consumption. If the actual consumption approach to the scope of the CPI is chosen, then in principle, the stocks of storable items need to be measured at the beginning and end of each period. Real actual consumption of a storable good is equal to beginning of the period inventory stock plus new purchases of the good less end of period stock of the good. In principle, if actual consumption is the target concept, then household stocks of storables should be capitalized and added to household wealth. In normal times, the services provided by these storable stocks probably should not be added to the current flow of consumption unless one argues that these stocks are desirable in their own right as a form of insurance against future supply shocks. In times of a pandemic, such an argument seems reasonable. Note that not recognizing a flow of services from the storables stock is a different treatment from the treatment of the services that consumer durables provide over their useful lifetime. Stocks of consumer durables should also be capitalized and added to household wealth but the services that durables render during a month need to be recognized as part of actual consumption. If the acquisitions approach to storable goods is taken, then household purchases of essential storable goods at the beginning of the lockdown period will be very much larger than pre-lockdown purchases of the same goods. Once the lockdown has been in place for a month or two, then purchases of storables should fall back to pre-lockdown levels. But the problem here is that the assumption of a constant basket equal to a pre-lockdown basket for all post lockdown periods may be a rather poor assumption.

6.5. How Should Scanner Data be Combined with Web Scraped Data?

Many statistical agencies now have access to scanner data from some retailers. How exactly should the indexes which are generated by the use of these data be combined with traditional price data collected by statistical agency employees or by the use of web scraped data?

In general, it is preferable if the contribution of these two sources of price data be combined in an index which weights the prices according to their economic importance; that is, to their shares of expenditure in the elementary category under consideration. It is not a problem to calculate expenditure shares for the scanner data but the web scraped data will not come with the associated expenditure data and so weighting the two sources of data by their relative quantities or expenditure shares will not be possible. In the end, some rough explicit or implicit estimate of the relative economic importance of the two sources of data will have to be made. Area specialists in NSOs will have to provide approximate weights for each elementary category that uses the two sources of information.

7. Conclusion

We suggest that statistical offices concentrate on getting more up to date expenditure weights for the post-lockdown period so that inflation during the lockdown period can be more accurately measured. When the lockdown ends, we suggest the use of a Fisher index, linking the first post-lockdown period to the last pre-lockdown period.

We have shown how the use of inflation adjusted carry forward prices for missing products, as recommended by international agencies, will typically lead to an understatement of inflation using our concept of market clearing imputed prices for missing products. However, these imputed prices require econometric estimation and so implementing this approach can only be done on a retrospective basis using econometric techniques.

Three steps that NSOs can take to provide as much information as possible on price indexes during a time of lockdown are:

1. Collect whatever prices are available, including from non-traditional sources. For missing prices, use inflation adjusted carry forward prices. While we favour using reservation prices, we acknowledge that currently it is unlikely that NSOs will be able to estimate these in a timely fashion,
2. Start a program to obtain current expenditure weights for the consumption basket, and
3. Produce a revisable CPI as an analytical series that can be updated as new methodology is developed and new data sources are exploited. Statistics Canada has followed this advice.

The BLS, U.S. Bureau of Labor Statistics' (2020) approach is very much in line with the approach advocated in this article; that is, the BLS produces a headline non-revisable standard fixed basket Lowe index while at the same time, it produces an approximation to a Tornqvist index which is improved over a two year revision period. Thus this supplementary index eventually measures inflation using weights that reflect current consumer expenditure patterns. Given that lockdown conditions have applied in varying degrees in many countries for many months, it is important to have information on current period household expenditure patterns so that meaningful estimates of consumer price inflation can be produced during the lockdown periods. See Appendix A in Diewert and Fox (2020) and the references there for more materials on the BLS approach to dealing with the pandemic.

Finally, it is unlikely that expenditure patterns will revert to the pattern that prevailed in periods just before the first lockdown period. This reinforces the case for obtaining more current estimates for household expenditures by category.

8. References

Aizcorbe, A., C. Corrado, and M. Doms. 2000. "Constructing Price and Quantity Indexes for High Technology Goods", Industrial Output Section, Division of Research and

Statistics, Board of Governors of the Federal Reserve System, Washington DC. Available at: https://users.nber.org/~confer/2000/si2000/doms.pdf.

Allen, R.C., and W.E. Diewert. 1981. "Direct versus Implicit Superlative Index Number Formulae", *Review of Economics and Statistics* 63: 430–435. DOI: https://doi.org/10.2307/1924361.

Aten, B., and A. Heston. 2009. "Chaining Methods for International Real Product and Purchasing Power Comparisons: Issues and Alternatives": 245–273 in *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, edited by D.S. Prasada Rao. Cheltenham UK: Edward Elgar.

Balk, B.M. 1980. "A Method for Constructing Price Indices for Seasonal Commodities", *Journal of the Royal Statistical Society, Series A* 143: 68–75. DOI: https://doi.org/10.2307/2981770.

Carvalho, V.M., J.R. Garcia, S. Hansen, Á. Ortiz, T. Rodrigo, J.V. Rodriguez Mora, and J. Ruiz. 2020. "Tracking the COVID-19 Crisis with High-Resolution Transaction Data", Cambridge-INET Working Paper Series No: 2020/16, University of Cambridge. Available at: https://www.econ.cam.ac.uk/research-files/repec/cam/pdf/cwpe2030.pdf.

Cavallo, A. 2013. "Online and Official Price Indexes: Measuring Argentina's Inflation", *Journal of Monetary Economics* 60(2): 152–165. DOI: https://doi.org/10.1016/i.imoneco.2012.10.002.

Cavallo, A. 2017. "Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers", *American Economic Review* 107(1): 283–303. DOI: https://doi.org/10.1257/aer.2016Q542.

Cavallo, A. 2020. "Inflation with Covid Consumption Baskets," NBER Working Paper 27352, Cambridge, MA. Available at: https://www.nber.org/papers/w27352.

Cavallo, A., and R. Rigobon. 2016. "The Billion Prices Project: Using Online Prices for Inflation Measurement and Research", *Journal of Economic Perspectives* 30(2): 151–178. DOI: https://doi.org/10.1257/iep. 30.2.15.

Court, A.T. 1939. "Hedonic Price Indexes with Automotive Examples": 99–117 in The *Dynamics of Automobile Demand*, New York: General Motors Corporation.

Dalen, J. 2001. "Statistical Targets for Price Indexes in Dynamic Universes," Paper presented at the Sixth meeting of the Ottawa Group, April 2–6, Canberra, Canada.

Deaton, A., and W.E. Diewert. 2002. "Conceptual Foundations for Price and Cost of Living Indexes", Chapter 2: 38–93 in *At What Price? Conceptualizing and Measuring Cost of Living and Price Indexes*, edited by C.L. Schultzc, and C. Mackie, Washington D.C., National Academy Press.

De Haan, J. 2004. "Estimating Quality-Adjusted Unit Value Indices: Evidence from Scanner Data," Paper presented at the Seventh EMG Workshop, December 12–14, Sydney, Australia. Available at: http://www.ipeer.ca/papers/De%20Haan,%20Oct.%2018,2004%20(3).pdf.

De Haan, J. 2010. "Hedonic Price Indexes: A Comparison of Imputation, Time Dummy and Repricing Methods", *Jahrbücher für Nationökonomie und Statistik* 230: 772–791. DOI: https://doi.org/10.1515/ibnst-2010-Q611.

De Haan, J., and F. Krsinich. 2012. "The Treatment of Unmatched Items in Rolling Year GEKS Price Indexes: Evidence from New Zealand Scanner Data", paper presented at the Economic Measurement Workshop 2012, November 23, University of New South

Wales. Available at: https://statsnz.contentdm.oclc.org/digital/collection/p20045coll4/id/340/

De Haan, J., and F. Krsinich. 2014. "Scanner Data and the Treatment of Quality Change in Nonrevisable Price Indexes", *Journal of Business and Economic Statistics* 32(3): 341–358. DOI: https://doi.org/10.1080/07350015.2014.880Q59.

De Haan, J., and F. Krsinich. 2018. "Time Dummy Hedonic and Quality Adjusted Unit Value Indexes: Do They Really Differ?", *Review of Income and Wealth* 64(4): 757–776. DOI: https://doi.org/10.1111/roiw.12304.

De Haan, J., and H.A. van der Grient. 2011. "Eliminating Chain drift in Price Indexes Based on Scanner Data", *Journal of Econometrics* 161: 36–46. DOI: https://doi.Org/10.1016/i.ieconom.2010.09.004.

Diewert, W.E. 1976. "Exact and Superlative Index Numbers", *Journal of Econometrics* 4: 114–145. DOI: https://doi.org/10.1016/0304-4076(76)90009-9.

Diewert, W.E. 1978. "Superlative Index Numbers and Consistency in Aggregation", *Econometrica* 46: 883–900. DOI: https://doi.org/10.2307/1909755.

Diewert, W.E. 1992. "Fisher Ideal Output, Input and Productivity Indexes Revisited", *Journal of Productivity Analysis* 3: 211–248. DOI: https://doi.org/10.1007/BF00158354.

Diewert, W.E. 2003. "Hedonic Regressions: A Review of Some Unresolved Issues", Paper presented at the Seventh Meeting of the Ottawa Group, 27–29 May, Paris, France.

Diewert, W.E. 2009. "Similarity Indexes and Criteria for Spatial Linking": 183–216 in *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, edited by D.S. Prasada Rao, Cheltenham, UK: Edward Elgar.

Diewert, W.E. 2021a. "The Economic Approach to Index Number Theory", Draft Chapter 5 in Consumer Price Index Theory, Washington D.C.: International Monetary Fund. Available at: https://www.imf.ora/en/Data/Statistics/cpi-manual. (accessed April 2021).

Diewert, W.E. 2021b. "Elementary Indexes", Draft Chapter 6 in Consumer Price Index Theory, Washington D.C.: International Monetary Fund. Available at: https://www.imf.org/en/Data/Statistics/cpi-manual. (accessed April 2021).

Diewert, W.E. 2021c. "The Chain Drift Problem and Multilateral Indexes", Draft Chapter 7 in Consumer Price Index Theory, Washington D.C.: International Monetary Fund. https://www.imf.org/en/Data/Statistics/cpi-manual. (accessed April 2021).

Diewert, W.E. 2021d. "Quality Adjustment Methods", Draft Chapter 8 in Consumer Price Index Theory, Washington D.C.: International Monetary Fund. Available at: https://www.imf.org/en/Data/Statistics/cpi-manual. (accessed April 2021).

Diewert, W.E., and R. Feenstra. 2019. "Estimating the Benefits of New Products: Some Approximations", Discussion Paper 19–02, Vancouver School of Economics, University of British Columbia, Vancouver, B.C., Canada. Available at: https://econ2017.sites.olt.ubc.ca/files/2019/03/pdf_paper_Diewert_Feenstra-DP19-02-Benefits-of-New-Products-Some-Approximations-Mar2019.pdf.

Diewert, W.E., and K.J. Fox. 2018. "Substitution Bias in Multilateral Methods for CPI Construction Using Scanner Data," UNSW Business School Research Paper No. 2018-13. Available at: https://ssrn.com/abstract = 3276457.

Diewert, W.E., and K.J. Fox. 2020. "Measuring Real Consumption and CPI Bias under Lockdown Conditions", NBER Working Paper 27144, Cambridge, MA. Available at: https://www.nber.org/papers/w27144.

Diewert, W.E., K.J. Fox and P. Schreyer. 2018. "The Digital Economy, New Products and Consumer Welfare", Economic Statistics Centre of Excellence (ESCoE) Discussion Paper 2018-16, London, UK. Available at: https://www.escoe.ac.uk/publications/the-digital-economy-new-products-and-consumer-welfare/.

Dunn, A., K. Hood, and A. Driessen. 2020. "Measuring the Effects of the COVID-19 Pandemic on Consumer Spending using Card Transaction Data," U.S. Bureau of Economic Analysis Working Paper WP2020-5. Available at: https://www.bea.gov/research/papers/2020/measuring-effects-covid-19-pandemic-consumer-spending-using-card-transaction

Dutot, Charles. 1738. *Réflexionspolitiques sur les finances et le commerce*, Volume 1, La Haye: Les frères Vaillant et N. Prevost.

Eurostat 2020. "Guidance on the Compilation of the HICP in the Context of the COVID-19 Crisis", Methodological Note, Directorate C, Unit C4, April 3. Available at: https://ec.europa.eu/eurostat/documents/10186/10693286/HICPguidance.pdf (accessed May 2020).

Feenstra, R.C. 1994. "New Product Varieties and the Measurement of International Prices", *American Economic Review* 84(1): 157–177. DOI: https://doi.org/10.1257/0002828041301524.

Fisher, I. 1922. *The Making of Index Numbers*, Boston: Houghton-Mifflin.

Hardy, G.H., J.E. Littlewood, and G. Pólya. 1934. *Inequalities*, Cambridge: Cambridge University Press.

Hausman, J.A. 1996. "Valuation of New Goods under Perfect and Imperfect Competition": 20–236 in *The Economics of New Goods*, edited by T.F. Bresnahan and R.J. Gordon, Chicago: University of Chicago Press.

Hausman, J.A. 1999. "Cellular Telephone, New Products and the CPI", *Journal of Business and Economic Statistics* 17(2): 188–194. DOI: https://doi.org/10.2307/1392474.

Hausman, J. 2003. "Sources of Bias and Solutions to Bias in the Consumer Price Index", *Journal of Economic Perspectives* 17(1): 23–44. DOI: https://doi.org/10.1257/089533003321164930.

Hicks, J.R. 1940. "The Valuation of the Social Income", *Economica* 7: 105–124. DOI: https://doi.org/10.2307/2548691.

Hill, R.J. 1997. "A Taxonomy of Multilateral Methods for Making International Comparisons of Prices and Quantities", *Review of Income and Wealth* 43(1): 49–69. DOI: https://doi.org/10.1111/i.1475-4991.1997.tb°0200.x.

Hill, R.J. 2001. "Measuring Inflation and Growth Using Spanning Trees", *International Economic Review* 42: 167–185. DOI: https://doi.org/10.1111/1468-2354.00105.

Hill, R.J. 2004. "Constructing Price Indexes Across Space and Time: The Case of the European Union", *American Economic Review* 94: 1379–1410. DOI: https://doi.org/10.1257/0002828043052178.

Hill, R.J. 2009. "Comparing Per Capita Income Levels Across Countries Using Spanning Trees: Robustness, Prior Restrictions, Hybrids and Hierarchies": 217–244 in

Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications, edited by D.S. Prasada Rao, Cheltenham UK: Edward Elgar.

ILO, IMF, OECD, UNECE, Eurostat, and The World Bank 2004. *Consumer Price Index Manual: Theory and Practice*, edites by Peter Hill, Geneva: International Labour Office.

IMF 2020. "Consumer Price Index Continuity", Unit C-4, Price Statistics. Available at: https://stats\viki.unece.org/displav/CCD2/Compilation + of + CPI + in + times + of + COVID-19?preview = /278037166/279773977/STA%20CPI%20Business%20 Continuity.pdf (accessed April 2021).

Ivancic, L., W.E. Diewert, and K.J. Fox. 2011. "Scanner Data, Time Aggregation and the Construction of Price Indexes", *Journal of Econometrics* 161: 24–35. DOI: https://doi.org/10.1016/i.ieconom.2010.09.003.

IWGPS, Intersecretariat Working Group on Price Statistics 2020. "Consumer Price Index: Business Continuity Guidance". Available at: https://statswiki.unece.org/display/CCD2/Compilation + of + CPI + in + times + of + COVID-19?preview = /278037166/279776928/IWGPS%20CPI%20Continuity%20Note.pdf (accessed April 2021).

Jevons, W.S. 1865. "The Variation of Prices and the Value of the Currency since 1782", *Journal of the Statistical Society of London* 28: 294–320; reprinted in *Investigations in Currency and Finance* (1884), London: Macmillan and Co.: 119–150.

Konüs, A.A. 1924. "The Problem of the True Index of the Cost of Living", translated in *Econometrica* 7, (1939): 10–29. DOI: https://doi.org/10.2307/1906997.

Kravis, I.B., A. Heston, and R. Summers. 1982. *World Product and Income: International Comparisons of Real Gross Product*, Statistical Office of the United Nations and the World Bank, Baltimore: The Johns Hopkins University Press.

Pollak, R.A. 1983. "The Theory of the Cost-of-Living Index": 87–161 in *Price Level Measurement*, edited by W.E. Diewert and C. Montmarquette, Ottawa: Statistics Canada.

Rao, D.S.P. 1995. "On the Equivalence of the Generalized Country-Product-Dummy (CPD) Method and the Rao-System for Multilateral Comparisons", Working Paper No. 5, Centre for International Comparisons, University of Pennsylvania, Philadelphia.

Rao, D.S.P. 2004. "The Country-Product-Dummy Method: A Stochastic Approach to the Computation of Purchasing Power parities in the ICP", paper presented at the SSHRC Conference on Index Numbers and Productivity Measurement, June 30–July 3, 2004, Vancouver, Canada. Available at: https://ideas.repec.org/p/qld/uqcepa/08.html.

Rao, D.S.P. 2005. "On the Equivalence of the Weighted Country Product Dummy (CPD) Method and the Rao System for Multilateral Price Comparisons", *Review of Income and Wealth* 51(4): 571–580. DOI: https://doi.org/10.1111/i.1475-4991.2005. 00169.x.

Silver, M. 2010. "The Wrong and Rights of Unit Value Indices", *Review of Income and Wealth* 56:S1: 206–206. DOI: https://doi.org/10.1111/j.1475-4991.2010.00391.x.

Silver, M. 2011. "An Index Number Formula Problem: the Aggregation of Broadly Comparable Items", *Journal of Official Statistics* 27(4): 1–17. Available at: https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/an-index-number-formula-problem-the-aggregation-of-broadly-comparable-items.pdf (accessed January 2022).

Summers, R. 1973. "International Comparisons Based Upon Incomplete Data", *Review of Income and Wealth* 29(1): 1–16. DOI: https://doi.org/10.1111/i.1475-4991.1973.tb00870.x.

Theil, H. 1967. *Economics and Information Theory*, Amsterdam: North-Holland.

UNECE 2020. (A. Bredt and C. Boldsen), "Compilation of CPI in Times of COVID-19". Available at: https://statswiki.unece.org/display/CCD2/Compilation + of + CPI + in + times + of + COVID-19 (accessed April 2021).

U.S. Bureau of Labor Statistics 2020. "Effects of COVID-19 Pandemic on BLS Price Indexes". Available at: https://www.bls.gov/bls/effects-of-covid-19-pandemic-on-bls-price-indexes.htm (accessed April 2021).

Von Auer, L. 2014. "The Generalized Unit Value Index Family", *Review of Income and Wealth* 60: 843–861. DOI: https://doi.org/10.1111/roiw.12042.

Von Hofsten, E. 1952. *Price Indexes and Quality Change*, London: George Allen and Unwin.

Walsh, C.M. 1901. *The Measurement of General Exchange Value*, New York: Macmillan and Co.

Received July 2020
Revised January 2021
Accepted November 2021

Journal of Official Statistics, Vol. 38, No. 1, 2022, pp. 287–289, http://dx.doi.org/10.2478/JOS-2022-0013

A Comment on the Article by W. Erwin Diewert and Kevin J. Fox

Carsten Boldsen[1]

Comments

The lockdown that followed the outbreak of the COVID-19 pandemic in 2020 posed unprecedented challenges to compile the consumer price index (CPI) in many countries, including difficulties with collecting prices from markets that remained open and the lack of price observations from markets that were temporarily closed. The closing of markets meant that many goods and services in the CPI basket were not available for purchase in which cases statistical offices had to implement methods to adjust for the missing observations.

To assist countries several international initiatives were launched. The Intersecretariat Working Group on Price Statistics (2020) issued the *Consumer Price Index: Business Continuity Guidance* in May 2020; Eurostat in 2020 published guidelines and methodological notes for the compilation of the HICP, Harmonized consumer price index (Eurostat 2020); four webinars in October-November 2020 organised by the Ottawa Group on Price Indices and United Nations Economic Commission for Europe (UNECE) discussed the production of the CPI under lockdown (UNECE 2020); and in August 2021, UNECE published the *Guide to producing CPI under lockdown* (UNECE 2021).

Based on economic theory, Diewert and Fox argue that the price of a product that becomes unavailable should be estimated by its 'reservation price' (the theoretical price that would drive demand to zero) to capture the effect on wellbeing. The authors argue reservation prices are higher than the corresponding 'inflation adjusted carry forward prices' which are commonly used by statistical offices to estimate missing prices. Therefore, the use of inflation adjusted carry forward prices will result in a lower CPI and underestimate the negative impact of disappearing products on households' wellbeing. Furthermore, CPIs based on pre-pandemic annual expenditure weights will not reflect the effect of changing consumption pattern during a lockdown. Therefore, the authors suggest compiling an index based on monthly weights and present different strategies to approximate a monthly cost-of-living index (COLI) during lockdown depending on data availability.

Economic theory brings important insights into the construction of CPIs. However, the applied static equilibrium analysis and the conclusions that can be drawn from it rest on rather strict assumptions. Consumers are utility maximizing agents with perfect information and constant preferences. There is no real, historical, time in the analysis.

[1] UNECE, Statistics Division, Palais des Nations, 8–14 avenue de la Paix, Geneva, 1211, Switzerland. Email: carsten.boldsen@un.org

© Statistics Sweden

In reality, preferences and income changes over time and there is a continuous emergence of new products and disappearance of old ones, which impact demand and relative prices.

Assume the reservation price is used to price a disappearing product. To be consistent, when the product was first included in the index, it should have been introduced with the same reservation price for the pre-introduction period. Hence, products would enter and exit with the same reservation price and leave the COLI unaffected. When a product first appears on the market there may be consumers that would be ready to pay the reservation price in the pre-introduction period. However, it is difficult to see why consumers should be ready to pay the same price when the product leaves the market after, say, two three or five years, in particularly for products which over their life cycle experience a fall in their relative price, reflecting a decrease in their relative marginal utility. The suggested approach may work for some markets, but most likely not for all; the assumptions may be too far from actual consumer behaviour and market conditions to give useful results.

There are also conceptual and practical reasons why reservation prices would not be suitable for the CPI. Most countries publish one headline CPI, which is used both for compensation purposes and as a measure of inflation. In the former case, the target of the CPI would be a COLI, in the latter it would be a fixed basket index, that is, a cost-of-goods index (COGI). The choice of estimation method, therefore, cannot be based only on the COLI approach. In many countries there is a preference for estimating missing prices based on the development of observed prices of comparable products. This is an operational and transparent practice, which ensures that the CPI is based on market signals rather than estimates based on questionable assumptions. As recognized in the article, reservation prices cannot be calculated in real time. They can only be estimated retrospectively and when extensive data sets of prices and quantities are available, for example, in terms of scanner data. For products where reservation prices can be estimated, market imperfections may lead to unreliable estimates and bring additional uncertainty into the CPI.

The use of monthly expenditure weights in most countries is not an option because of lack of data and resource constraints. It may be possible to derive such weights for specific product groups, but it will be difficult to construct monthly weights for the full basket. A chained index based on monthly weights may change because of both price changes and changes in the weights; in theory it may change even if all prices remain constant only because of changes in the weights. This is not a suitable property if the index is used for measuring inflation. It could also be noted that a COLI ideally should be based on 'democratic' weights while most CPIs are using 'plutocratic' weighting.

While the article focuses on the situation under the pandemic, it feeds into the broader discussion about the treatment of disappearing and new products in the CPI and how to measure the effect on wellbeing caused by changes in the available basket of goods and services. In their article *Measuring Consumer Inflation in a Digital Economy*, Reinsdorf and Schreyer (2019) discuss the problems of measuring welfare gains of new products. They conclude that such welfare effects for both conceptual and practical reasons should be addressed in research work on welfare measurement "beyond GDP" rather than included in existing headline CPIs. The article by Diewert and Fox provides useful inputs to this line of research work.

References

Eurostat. 2020. *Guidelines and methodological notes in the context of the COVID-19 crisis.* Available at: https://ec.europa.eu/eurostat/data/metadata/covid-19-support-for-statisticians.

Intersecretariat Working Group on Price Statistics. 2020. *Consumer Price Index: Business Continuity Guidance*. Available at: https://statswiki.unece.org/x/roKSE.

Reinsdorf, M. and P. Schreyer. 2019. *Measuring Consumer Inflation in a Digital Economy*. OECD Statistics Working papers 2019/01. Available at: https://www.oecd-ilibrary.org/economics/measuring-consumer-inflation-in-a-digital-economy_1d002364-en.

UNECE. 2020, The proceedings of the webinars are available on https://unece.org/-statistics/events/webinars-producing-cpiunder-lockdown.

UNECE. 2021. *Guide on producing CPI under lockdown*. Available at: https://unece.org/info/Statistics/pub/359134.

Received March 2021
Revised October 2021
Accepted November 2021

sciendo

Journal of Official Statistics, Vol. 38, No. 1, 2022, pp. 291–293, http://dx.doi.org/10.2478/JOS-2022-0014

Creative and Exhaustive, but Less Practical – a Comment on the Article by Diewert and Fox

Bernhard Goldhammer[1]

The article by Diewert and Fox provides a comprehensive overview of challenges that NSOs face in producing the CPI in pandemic times by touching on many different fields. A focus is on the treatment of missing prices, where they propose different methods depending on the resources available to the NSO. However, some of the procedures proposed can be seen as being less practical like the use of reservation prices (which is also debatable from a theoretical point of view) and of alternative data sources for weights whose implementation supposedly takes longer than the pandemic itself. Overall, the article provides an important contribution for making CPI production more robust for similar crises in the future.

Key words: Consumer price index compilation; pandemic conditions.

Comments

Let us start the discussion about the article by Erwin Diewert and Kevin Fox with the guidance on price imputation by IMF, Eurostat and UNECE and their preferred term "carry forward price adjusted for inflation". The methodology of imputing prices by using a previous price and adjusting them with the price development of different products, different product categories or even the overall index boils down to reweighting the index in one way or another. For imputing with the all-items index, the imputed sub-index gets an implicit weight of 0. This property can be seen from Equation (9) for the Laspeyres index, Equation (12) for the Lowe index and Equation (18) for the Young index of the article by Diewert and Fox.

Similarly, using a particular sub-index for imputation attaches a higher weight to this sub-index, which might be acceptable under pandemic conditions. This implicit reweighting may help to accommodate the lockdown basket, while an explicit re-weighting would break the index. Another reason for using this method, not discussed in the article, is the continuity of established CPI production systems.

It allows returning to a normal index production after the lockdown, as time series of product prices will not be discontinued: normal price collection can be restarted, enabling a price comparison to the previous year. This basically relates to the price collection; a weight adjustment will be necessary, as overall consumption patterns may have changed. So, given a lockdown in the months March to June 2020, the price comparison would be possible also for the months March to June 2021 (to the imputed price, of course).

[1] European Central Bank, Directorate General Statistics, Taunustor 2, 60311 Frankfurt am Main, Germany. Email: Bernhard.Goldhammer@ecb.europa.eu

© Statistics Sweden

Regarding reservation prices, the authors acknowledge the practical problems that NSOs would face. But also from a theoretical point of view, the concept is not very intriguing: The authors assume that the supply curve "has shifted to become identical to the vertical price axis" and that the "(unobserved) market price" would be identical to the reservation price. A different interpretation would be that the supply curve has fully disappeared and no market – and therefore no market price – exists anymore. In such a situation, the calculation of a reservation price would be misleading. Furthermore, depending on the chosen preference concept, the value of the reservation price can be substantially different (Diewert and Feenstra 2019). Reservation prices themselves are mainly a construct derived from the "economic approach" of index theory; but most consumer price indices rather follow the basket approach, which does, in principle, rule out the use of non-observable prices like reservation prices. It may be argued that quality adjustment with hedonics in a basket index is already some kind of "economic index theory element". However, the fixed basket approach in its very nature needs constant quality and therefore quality adjustment. The statistician has a wide variety of methods at hand, not only hedonics. They are chosen because they can be applied to serve the basket concept, not because of their economic nature. Similarly, reservation prices could be applied for imputation, if they could be calculated reliably and lead to understandable, transparent and unambiguous results. Hausman's (1996) cereal reservation price of USD 7.14 instead of the initially observed price of USD 3.78 (Groshen et al. 2017) would put an enormous communication burden on the shoulders of an NSO.

It is a strength of the article that the authors try to accommodate the needs of NSOs facing different challenges. It is, however, surprising that the argumentation in Section 4 "the way forward" gives the impression that which case applies to which NSO depends solely on resources. Data availability plays, at least, an equally important role: Even with large resources, monthly consumption expenditure data with the necessary granularity may be not available. Both the data and the resources to exploit them are needed.

Leaving aside the concept of reservation prices, Section 5 provides advice on how to deal with missing prices for the construction of the index. While I generally like the idea of using different formulas to somehow cope with missing products within an elementary aggregate, it is not clear how to derive meaningful annual change rates from an index which, for a short period, uses a different calculation formula, which is then reverted. This may itself lead to a probably distorted index – over the period of one year's time.

An innovative method in this respect is method 7, a new multilateral index, which would deserve an own article. Such an article could be used to elaborate more in detail on the following aspects: first, "similar price structures" and their identification should be discussed. Second, the properties of the resulting index need more elaboration, especially the interpretation of the inflation rate between two periods in which the structures are quite different. Compared to a chained index, where the influence of the structural change compared to the price change can be calculated, this is also not obvious for this new method.

Turning to the question of up-to-date weighting information, one problem highlighted by the authors is the lack of information on current household expenditure. While their ideas for mitigation sound convincing, they all point to medium-term strategies and not to the short-term solutions necessary in a pandemic. For example, setting up a continuous consumer expenditure survey will take several years, will be very costly, and needs to take into consideration survey problems like underreporting, panel mortality or panel effect. Credit card data normally refer to the entire shop where the transaction took place; and scanner data will tell an incomplete story, as it can only be processed for few retailers due to heavy processing expenses. So, nobody should expect a quick solution of the pandemic weighting problems by referring to the potential data sources discussed by the authors.

Finally, another special point discussed by the authors is stockpiling. Interestingly, a traditional fixed-basket approach may be closer to the consumption approach than an acquisition approach. Another question is whether a CPI should really account fully and immediately for one-off effects like a singular rush to stockpiling or be a more robust measure of the development of purchase power over the medium-term.

In conclusion, the article provides a comprehensive overview of challenges which NSOs have been facing in their CPI compilation since the pandemic has started. Touching on many different fields may also help the reader who is not familiar with procedures in an NSO to understand the problems in these times. While not all recommendations may stand the test of practice or are feasible given time, data and resources constraints, the article provides an important contribution for setting the scene towards more robust CPI compilation systems that are better placed to cope with a crisis like the current pandemic.

References

Diewert, W.E. and R. Feenstra. 2019. *Estimating the Benefits of New Products: Some Approximations*. Discussion Paper 19–02, Vancouver School of Economics. Vancouver: University of British Columbia.

Groshen, E., B. Moyer, A. Aizcorbe, R. Bradley, and D. Friedman. 2017. “How Government Statistics Adjust for Potential Biases from Quality Change and New Goods in an Age of Digital Technologies: A View from the Trenches.” *Journal of Economic Perspectives* 31(2): 187–210. DOI: http://dx.doi.org/10.1257/jep.31.2.187.

Hausman, J. 1996. “Valuation of New Goods under Perfect and Imperfect Competition.” In *The Economics of New Goods*, edited by T. Bresnahan, and R. Gordon: 207–248. Chicago: University of Chicago Press.

Received October 2021
Accepted November 2021

Journal of Official Statistics, Vol. 38, No. 1, 2022, pp. 295–300, http://dx.doi.org/10.2478/JOS-2022-0015

"Measuring Inflation under Pandemic Conditions": A Comment

Naohito Abe[1]

Diewert and Fox (2022) examine various implications of the 2020 COVID-19 pandemic for constructing consumer price indexes. The authors state that the pandemic caused major changes in consumption expenditures and shares which makes fixed basket index number formulae inapplicable. They emphasize the need for more frequent surveys of consumer expenditure which will enable compilation of the Fisher index which is considered superior to the traditional Laspeyres or Young indexes. In addition, Diewert and Fox discuss the use of various "new" technologies such as web scraping, scanner data, and information from transactions through credit cards to estimate consumption expenditure.

Key words: Price index; COVID-19; demand shocks; supply shocks.

The Diewert and Fox article investigates a critical topic and offers useful discussion and valuable insights and advice, particularly for people who are constructing official price indices. In my view, the COVID-19 pandemic has highlighted the limitation of the traditional Laspeyres, Young, and Lowe price index number formulae. The changes in the consumption expenditure caused by the pandemic were so large that the price indexes that ignore those changes can be misleading and are inadequate as deflators for consumption expenditures.

My main comment on the Diewert and Fox paper concerns their treatment of missing or unobservable prices during the 2020 COVID-19 pandemic. During the pandmic, in many countries, statistical offices were not able to collect price data, which lead to missing or unobservable prices. Diewert and Fox (2022) introduce "the market clearing reservation prices" and name the price index that uses the reservation prices for missing prices as a "true" index. They point out that the inflation adjusted carry forward prices by national statistical offices are below the corresponding reservation prices, thus understate the "true" inflation during the pandemic. In this comment, I raise two issues. The first issue is how relevant the market clearing reservation prices are when inferring missing price data. The second is whether the inflation adjusted carry forward prices are likely to be below the "true" but unobserved prices or not.

Diewert and Fox (2022) state:

> "What has happened is that the supply curve for product n has become straight line that is parallel to the price axis and this line has shifted to become identical to the vertical

[1] The Institute of Economic Research, Hitotsubashi University, Institute of Economic Research 2-1, Naka, Kunitachi, Tokyo, 186-8603, Japan. Emails: nabe@ier.hit-u.acu.ac.jp

Acknowledgments: I have benefitted greatly from discussion with the Statistical Bureau in Japan, D.S. Prasada Rao, and Chihiro Shimizu. I also thank the editor and the reviewers for carefully reading the manuscript and comments. Financial supports from JSPS KAKENHI, 19H01467 and 20H00082 are gratefully acknowledged.

© Statistics Sweden

price axis. Thus the (unobserved) market price for the product n under consideration is the price where the demand curve intersects the vertical price axis; (pp.XX) . . ."

Using Figure 1 below, I demonstrate that the case discussed by Diewert and Fox is only one of several possible types of shocks that may have been observed during the Covid-19 pandemic. I discuss demand side and supply side shocks observed during the past year. In fact, the situation considered by Diewert and Fox (2022) can be illustrated as Case 1 as follows.

Case 1: Pure Supply Shock (Diewert and Fox)

Suppose the economy is at *E*, where the demand curve, *D*, intersects the supply curve, *S*. The pandemic shifts the supply curve from *S* to *S'* so that the curve overlaps the vertical axis, while the demand curve is unaffected. The economy moves to equilibrium point *E1* where the quantity demanded is zero. The price level at *E1*, identified by Diewert and Fox as the market clearing reservation price, is greater than the price level at *E* as long as the demand curve slopes downward.

To make use of the market clearing reservation price as defined in Diewert and Fox (2022), some assumptions on the demand functions are necessary. First, to obtain a finite value for the reservation price, the marginal substitution of the commodity at zero consumption must also be finite, which excludes a popular class of utility functions that use logarithms of quantities. The second and more controversial assumption is that the demand curve remains invariant through the pandemic. Generally, given prices, shifts in demand occur when people's marginal willingness to pay for commodities change, which can arise without changes in technologies, prices of other commodities, and due to legal restrictions. Because the 2020 COVID-19 pandemic affected our daily life to a great extent, the assumption that the pandemic did not affect demands appears particularly strong.

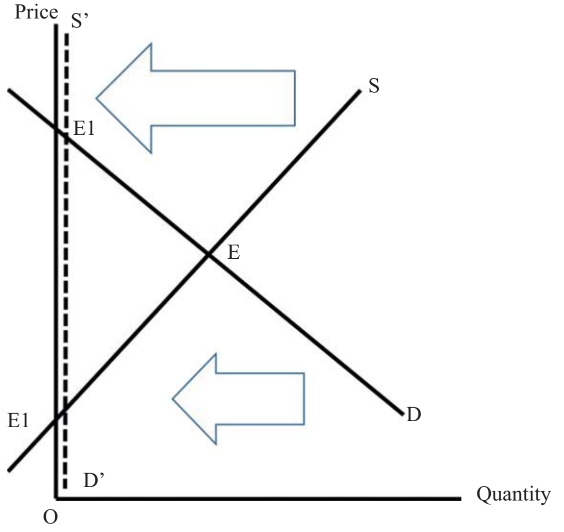


Fig. 1. Demand and supply during pandemic.

In Figure 1, we can easily consider demand shocks as the underlying cause for the missing price problem.

Case 2: Pure Demand Shock

Suppose the economy is at E before the pandemic. Then, consider the case where the 2020 COVID-19 pandemic shifts the demand curve left so that it overlaps the vertical axis, while the supply curve is unchanged. In this case, the economy moves to new equilibrium $E2$ and the "true" or equilibrium prices are lower than those before the pandemic.

Similar to Case 1, to obtain non-zero "true" price, the following assumptions are required: (1) supply curve is upward sloping and (2) the supply curve intersects the vertical axis at positive price level, which implies that the marginal cost at zero quantity should be positive. Similar to Case 1, the second assumption excludes a popular class of cost functions that use logarithms of quantities. However, in the short run, the assumption of positive marginal costs at zero seems plausible

Diewert and Fox (2022) derive a sufficient condition for the true price index to be higher than the index based on the inflation adjusted carry forward price. Because the treatment of supply and demand is symmetric between Cases1 and 2, it is possible to derive a corresponding sufficient condition under Case 2.

Following notations by Diewert and Fox (2022), define P^{11}, P_{LQ}, P_B^{CI} as the inflation adjusted carry forward price, the Laspeyres and Young indexes using P1I, respectively. Also define $P_n^{1*}, P_{LQ}^*, P_B^*$ as the "true" prices computed based on economic theory, the Laspyres and Young indexes using P_n^{1*}, respectively. Then, a sufficient condition for $P_{LQ}^* < P_{LQ}$ and $P_B^* < P_B^{CI}$ is given by

$$P_n^{1*} < P^{11}. \tag{1}$$

The above condition is very similar to Equations (10) and (13) in Diewert and Fox (2022). The only difference is the direction of the inequality sign. That is, if the "true" prices are lower than the inflation adjusted carry forward prices, the fixed basket indexes such as the Laspeyres and Young indexes with the carry forward price overstate the true price changes.

Thus, to infer the missing prices due to zero quantities caused by the pandemic, we have two options. In Case 1, the case considered by Diewert and Fox, we estimate the prices by fixing the demand function. In Case 2, the "true" prices are obtained by fixing the supply function. Which of the two cases is relatively more appropriate to infer the missing prices during the pandemic is an empirical question which is addressed in the remaining part of this paper. Using Japanese examples of the missing price problem due to the 2020 COVID-19 pandemic, I discuss which shocks, demand or supply, are more plausible to interpret the data.

The Statistical Bureau in Japan faced the missing price problem in October and November 2020 for some items in international tour packages. The bureau collects six specific packages, such as a round trip between Tokyo and Seoul for three days, two persons including various fees and surcharges. During the pandemic, even though the demand was limited, travel agencies in Japan were selling tours with their list prices, which enabled the bureau to collect the list prices. In October 2020, travel agencies stopped selling some tour packages, which forced the bureau to infer prices for these packages. The Statistical Bureau in Japan adopted the inflation adjusted carry forward prices for the missing items.

Figure 2 depicts movements of the expenditure and price index for international tour packages in Japan since 2015. We can observe that the expenditures are close to zero since March, 2020, which was seven months before the missing price problem occurred. That is, the occurrence of zero expenditure did not automatically cause a missing prices problem. Until October 2020, travel agencies kept selling international tour packages even if demand for the tours was small. In other words, the supply curve before October 2020 was not the vertical axis. Some travel agencies offered tours at similar prices before the pandemic, which suggests that Case 2 seems more appropriate than Case 1 to interpret the movement of prices and quantities in Figure 2. Because the demand curve became the vertical axis before the missing price problem occurred, the price level reached *E2* in Figure 2 before October 2020. The "true" price for the missing items do not seem to differ much from the list prices in previous months. In this case, the adoption of the inflation adjusted carry forward prices by the bureau seems a reasonable decision.

Another missing price problem the Statistical Bureau in Japan encountered was for admission tickets to professional baseball and soccer games. Figure 3 depicts the movements of expenditures and prices for the category. In April and May 2020, all the professional games were canceled so that price data became unavailable for the statistical bureau.

Because the prices of the admission tickets were stable until July, the Statistical Bureau adopted the prices in previous months for the missing prices. Professional games resumed in June 2020. Since July, the prices of admission tickets went up by few percentage points

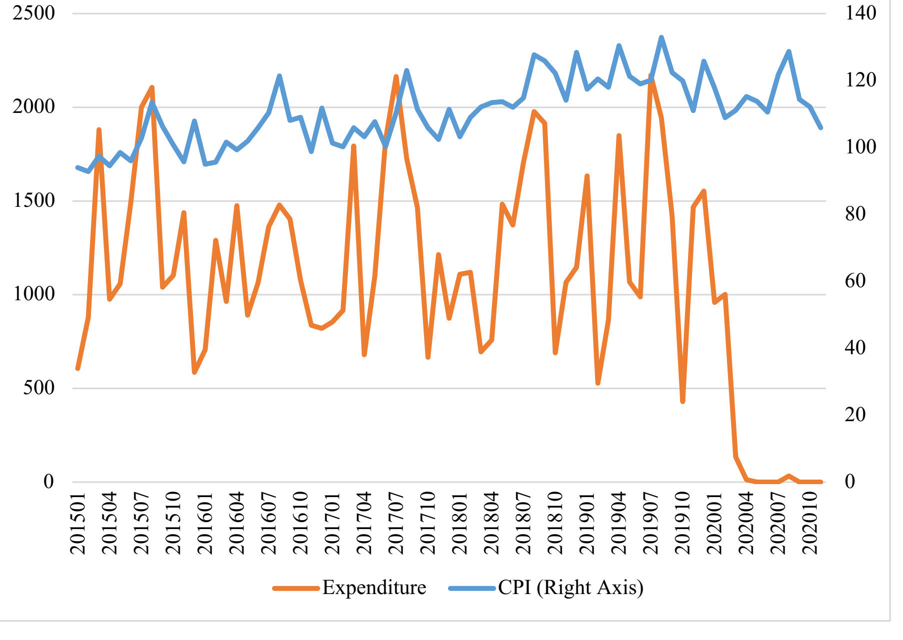


Fig. 2. Expenditures and prices of international tour packages in Japan.

Note: The data are from *the Family Income and Expenditure Survey* (two-or-more person households) and *the Consumer Price Index* by the Statistics Bureau of Japan. The unit is the Yen for the expenditure. The base year of the price index is 2015.

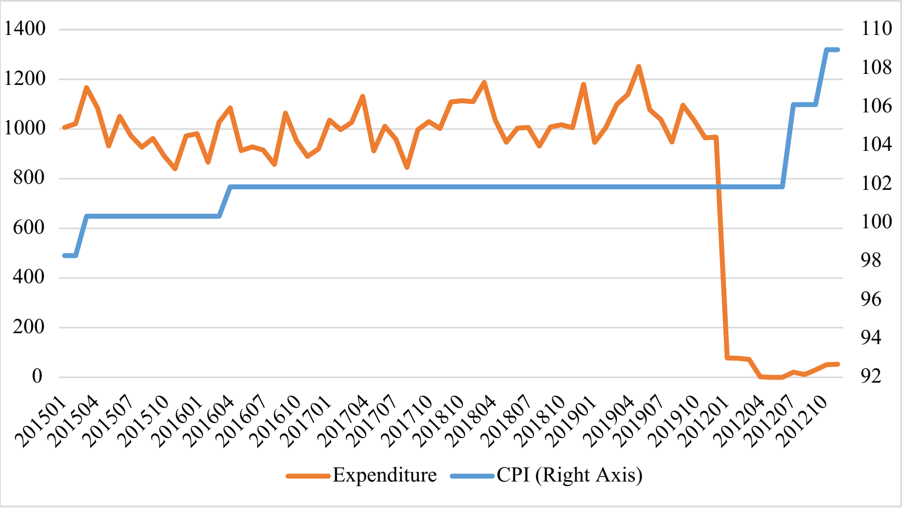


Fig. 3. Expenditures and prices of admissions to professional Baseball and soccer games.
Note: The data source is the same in Figure 2.

probably due to increases in costs for sanitization of seats. In September, the number of maximum admissions for each game was set to between 33% to 50% of the maximum number before the pandemic. However, the expenditure remained 1% to 5 % of the usual level.

If we assume Case 1 occurred for professional games in Japan, the market clearing reservation prices during April and May 2020 would have been much higher than the official price index. However, the statistical office would have faced a problem when considering the situation after the pandemic. In September 2020, many seats were sold at relatively higher prices than before the pandemic, while the expenditures remained at low level. If the demand curve is unchanged, Figure 3 suggests that the supply curve is close to the vertical axis in September-November 2020. The prices during the period that are higher than the pre-pandemic level by 5–8 percent points must be close to the market clearing reservation prices. The magnitude of the increase in prices seems too small because it implies an unlikely large price elasticity of demand. Therefore, it seems natural to think that the pandemic affected the demand curve even after the lockdown to a great extent. In other words, the scenario in Case 2 seems more appropriate to interpret Figure 3 than that in Case 1.

The two examples discussed above suggest occurrences of significant demand shocks during the pandemic, which renders Case 1 unsuitable as a description for the mechanism behind the missing price problem.

Before concluding this short comment, let me mention a possibility that the 2020 COVID-19 pandemic makes it difficult to evaluate changes in consumer welfare from the cost of living index. The combination of the positive prices and low expenditure depicted in Figures 2 and 3 suggests the occurrence of excess supplies. If prices are not determined at the intersection of demand and supply, there are some people who cannot purchase commodities at the list prices, or there exist some sellers who cannot find customers who purchase the products at the prices. In fact, during the pandemic, broadcasts reported many empty seats in flights and long-distance trains. Long queues for purchasing face masks and hand sanitizers were frequently observed. If excess supply or excess demand exists, the observed price becomes higher or lower than the marginal willingness to pay. The cost of living index based on the observed prices ignores the plight of people who cannot purchase at observed prices even if they wish to buy, thus, the cost of living index becomes inappropriate to evaluate their welfare.

Missing price problem caused by the 2020 COVID-19 pandemic is a serious challenge. The inference of the missing prices is difficult even if we have plenty of data. The market clearing reservation price proposed by Diewert and Fox (2022) is one of several possible methods to infer the prices, which can potentially result in huge price changes from the previous periods. However, if we allow for changes in demand functions, as opposed to Diewert and Fox (2022), it seems that the missing prices do not differ much from the carry forward prices from the previous periods. Based on the analysis of the two cases, supply versus demand shocks, illustrated using Figure 1 and the data illustrated in Figures 2 and 3, the decision by statistical agencies to use inflation adjusted carry forward prices seems to be a reasonable practice.

References

Diewert, E.W. and K.J. Fox. 2022. "Measuring Inflation under Pandemic Conditions." *Journal of Official Statistics*, this issue.

Statistical Bureau of Japan 2022. *Frequently Asked Questions: the Consumer Price Index*. Available at: http://www.stat.go.jp/data/cpi/4-1.html#Z8 (accessed February 2022).

Received February 2021
Revised October 2021
Accepted October 2021

Journal of Official Statistics, Vol. 38, No. 1, 2022, pp. 301–317, http://dx.doi.org/10.2478/JOS-2022-0016

Price Index Numbers under Large-Scale Demand Shocks– The Japanese Experience of the COVID-19 Pandemic

Naohito Abe[1], *Toshikatsu Inoue*[1], *and Hideyasu Sato*[2]

We investigate the prices and quantities of face masks when the 2020 COVID-19 pandemic was particularly serious to understand the impact of demand shocks on the cost of living index (COLI). Using a recently developed index number formula that is exact for the constant elasticity of substitution utility function with variable preferences, we quantified the degree of demand shock caused by the pandemic. Our empirical analysis revealed that shifts in preferences during the pandemic were so large that the COLI with variable tastes became very different from the standard superlative indexes. While the prices of face masks decreased in the Fisher index in May 2020 by 0.76% per week, the COLI increased by 1.92% per week.

Key words: COVID-19; pandemic; coronavirus; price index; demand shocks.

1. Introduction

The coronavirus 2019 (COVID-19) pandemic promoted massive stockpiling behavior among consumers worldwide, and Japan was no exception. The first case of COVID-19 in Japan was reported in mid-January, 2020. Immediately after the news was reported, the demand for face masks and sanitizers surged. On March 23, 2020, the Governor of Tokyo warned that a lockdown might be imposed, causing several people to flock to supermarkets and grocery stores to purchase food and other necessary items.

Due to the COVID-19 threat, people have changed their consumption behaviors to a large extent. Figure 1 shows the weekly rates of change of the chained Laspeyres and Paasche indexes of face masks in Japan based on scanner data. As Ivancic et al. (2011) show, weekly scanner data often exhibit large discrepancies between Laspeyres and Paasche indexes, which can be observed in the figure. In the middle of January 2020, both indexes increased to a great extent; then, the Paasche index overtook the Laspeyres index. According to the Bortkiewicz decomposition of the Laspeyres- Paasche (L-P) gap, the negative L-P gap implies that the correlation between quantities and prices is positive. Although, theoretically, a positive correlation between quantities and prices is not impossible, it is quite unlikely. A natural interpretation is that during this period, large-

[1] The Institute of Economic Research, Hitotsubashi University, Institute of Economic Research 2-1, Naka, Kunitachi, Tokyo, 186-8601, Japan. Emails: nabe@ier.hit-u.ac.jp and ed195001@g.hit-u.ac.jp
[2] Department of Food and Nutritional Sciences, Toyo University, 1-1-1 Izumino, Itakura-machi, Ora-gun, Gunma 374-0193, Japan. Email: sato.hideyasu@gmail.com
Acknowledgments: The authors gratefully acknowledge helpful suggestions and comments from D.S.P. Rao. Abe's work was supported by JSPS KAKENHI, 19H01467, 16H06322, 18H00864, and 20H00082. The authors also wish to thank anonymous referees for very constructive and helpful comments.

© Statistics Sweden

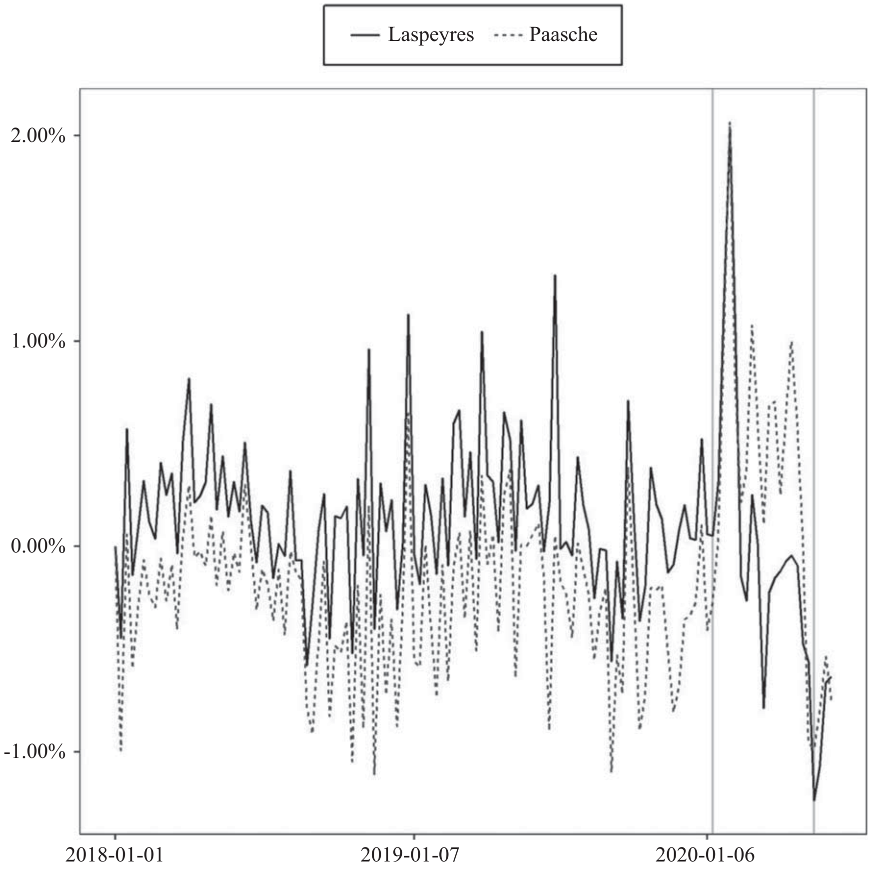


Fig. 1. Chained Laspeyres and Paasche Indexes of Face Masks.
Notes: Based on weekly Japanese scanner data. See Section 5 for the detail of the dataset.

scale demand shocks occurred, which shifted prices and quantities along an upward-sloping supply curve.

In this study, using Japanese scanner data for face masks, we investigated the impacts of demand shocks on the COLI caused by the COVID-19 in 2020. The traditional theory of COLI assumes that the preference is unchanged. Therefore, changes in demands or preferences are not captured in the COLI. One notable exception is Fisher and Shell (1972) that proposes calculating the difference between two COLIs, one using the old preferences, and one using the new preference. Although this carries information on the effects of having different preferences on the COLI, it does not provide us with information on how the cost of living changes when preferences vary. Philips (1974) criticizes Fisher and Shell (1972) and proposes a cardinal COLI that compares the minimum expenditures between two time periods assuming two different utility levels are comparable, that is, the utility function is cardinal. Balk (1989) proposes a COLI based on ordinal utility functions. He introduces the reference vector. The minimum expenditure is arrived at which the utility level at the reference vector are assured. Martin (2020) provides us with a brief survey on the cost of living index with variable preferences. In a recent path-breaking paper, Redding and Weinstein (2020) propose the constant elasticity of substitution (CES) unified price index (CUPI). The CUPI has several important characteristics. First, CUPI is exact for CES COLI without any restrictions on the relationship between quantities and prices. Second, CUPI can be decomposed into two effects; price and taste effects. Price effects can take various forms. In this study, following Redding and Weinstein (2020), we adopt the Sato-Vartia index as the price effect. Taste effects capture changes in demand or preferences. Redding and Weinstein (2020) call this effect "bias" in the traditional COLI such as the Fisher or Tornqvist indexes.

Our empirical analyses based on Japanese weekly scanner data of face masks revealed that the 2020 COVID-19 caused large-scale demand shocks. This increased the discrepancies between the traditional COLI such as the Fisher and Tornqvist indexes and CUPI to a great extent. That is, the demand shocks caused by the pandemic caused a significant change in the COLI. More specifically, while the prices of face masks decreased in the Jevons and Fisher indexes in May 2020 by 0.06% and 0.76% per week, respectively, the COLI increased by 1.92% per week. The magnitude of the changes caused by the demand shock is so substantial that traditional index numbers may carry incorrect information on the cost of living among consumers.

The article is organized as follows. Section 2 presents a brief history of the COVID-19 pandemic in Japan. Section 3 introduces the index number formula by Redding and Weinstein (2020) and discusses the measures of demand shocks. Section 4 describes the data set. Section 5 presents the empirical results. Section 6 concludes.

2. The 2020 COVID-19 Pandemic in Japan and Face Masks

The first COVID-19 infection was reported in Japan on January 16, 2020. From Figure 2, which depicts the change in the number of infected persons by reported date, it can be seen that the number of COVID-19 cases started to increase significantly from February 2020. In response to this pandemic, the Japanese government announced its first emergency plan on February 13. The government also requested manufacturers to increase the production of masks, which had already started to run short, and prefectures to allocate stockpiles to medical institutions. On March 10, the second emergency plan was formulated, and the resale of masks, still in short supply, was legally prohibited. In addition, the government decided to purchase 20 million reusable cloth masks in bulk to be distributed to nursing homes and nursery schools. They also decided to secure 15 million masks for medical institutions as mask imports augmented, and manufacturers were requested to increase their production.

The infection continued to spread, and the Governor of Tokyo mentioned the possibility of a lockdown at a press conference on March 23, resulting in a temporary increase in consumer demand for food and daily necessities. On April 7, a one-month state of emergency was declared in seven prefectures, including Tokyo and Osaka, and residents were requested to avoid leaving their prefectures as much as possible. The declaration was extended to all prefectures on April 16, and the period was extended to May 31. However, as the number of infected persons began to decrease in May, the declaration was lifted in 39 out of 43 prefectures on May 14, followed by three more prefectures on May 21. In response to chronic mask shortages, two reusable cloth masks were distributed during

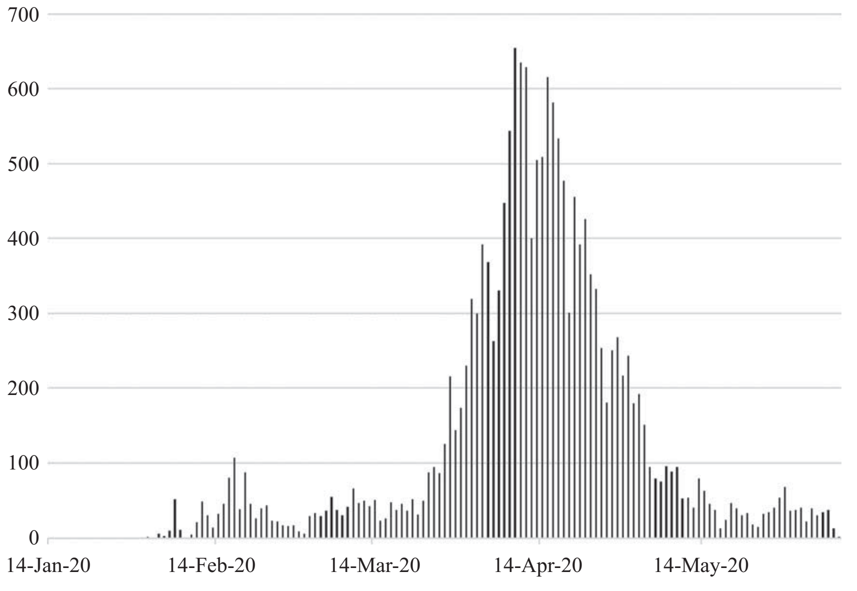


Fig. 2. Number of COVID-19 Infections in Japan.
Source: The National Institute of Infectious Diseases, Japan.

the state of emergency to each child, student, faculty, and staff member attending or working at schools, in addition to two masks to each household nationwide. These distributions were completed on June 20.

3. The Price and Cost of Living Index With Taste Shocks

The CUPI by Redding and Weinstein (2020) consists of the two price indexes. The first is the CES common variety (CCV) price index, and the second is the Redding-Weinstein (RW) index, which includes the effects of changing product variety. The CCV between time s and t is defined as

$$\ln CCV\ (p_s, q_s, p_t, q_t) = \sum_{i=1}^{N} \omega^*_{ist}\ (\ln p_{it} - \ln p_{is}) + \sum_{i=1}^{N} \omega^*_{ist}(\ln \varphi_{is} - \ln \varphi_{it}), \quad (1)$$

$$\omega^*_{ist} = \frac{w_{it} - w_{is}}{\ln (w_{it}) - \ln (w_{is})} \Big/ \sum_{i=1}^{N} \frac{w_{it} - w_{is}}{\ln (w_{it}) - \ln (w_{is})}, \quad (2)$$

$$w_{it} = p_{it} q_{it} \Big/ \sum_{i=1}^{N} p_{it} q_{it}. \quad (3)$$

where φ_{it} and q_{it} are the taste parameter and the quantity of a commodity i, at time t, respectively. We denote the vector of prices, quantities, and taste parameters at time t as follows.

$$p_t = (p_{1t}, p_{2t}, \ldots p_{Nt}),\ \ q_t = (q_{1t}, q_{2t}, \ldots, q_{Nt}),\ \ \varphi_t = (\varphi_{1t}, \varphi_{2t}, \ldots, \varphi_{Nt}),$$

The taste parameter φ_{it} is also a function of prices and quantities as follows.

$$\varphi_{it} = \left(\frac{p_{it}}{p_{1t}}\right)\left(\frac{w_{it}}{w_{1t}}\right)^{\frac{1}{\sigma-1}}\left[\prod_{k=2}^{N}\left\{\left(\frac{p_{kt}}{p_{1t}}\right)\left(\frac{w_{kt}}{w_{1t}}\right)^{\frac{1}{\sigma-1}}\right\}^{(-1/N)}\right]\varphi, \tag{4}$$

where $\varphi > 0$ is a positive constant. Please see Appendix (Subsection 7.1) for the derivation of Equation (4).

The first term on the right-hand side of the Equation (1) is the Sato-Vartia (SV) index. The second term. the taste term, captures the changes in the COLI caused by the changes in the preference parameters, φ_{it}. over time. RW shows that the CCV defined in Equations (1) and (4) is the COLI for the following utility function and the normalization condition.

$$U_t(q_t;\varphi_t,\sigma) = \left(\sum_{i=1}^{N}(\varphi_{it}q_{it})^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}, \tag{5}$$

$$\prod_{i=1}^{N}\varphi_{it}^{1/N} = \varphi, \tag{6}$$

where $\sigma > 1$ is the elasticity of substitution and $N > 1$ is the number of commodities. Since the above utility function is linearly homogeneous with respect to the quantities. the minimum expenditure function can be written as the product of the unit expenditure function. $C(p_t;\varphi_t)$ and utility level.

$$E(p_t, U_t;\varphi_t) = C(p_t;\varphi_t) \times U_t.$$

Here the unit expenditure function takes the following functional form.

$$C(p_t;\varphi_t) = \left(\sum_{i=1}^{N}\left(\frac{p_{it}}{\varphi_{it}}\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}}.$$

A notable feature of the utility function in Equation (5) is that the taste parameter, φ_{it}, can vary over time. The COLI corresponding to Equation (1) is given by,

$$COLI(s,t) = \frac{E(p_t, U_t = U;\varphi_t)}{E(p_s, U_s = U;\varphi_s)},$$

$$= \frac{C(p_t;\varphi_t) \times U}{C(p_s;\varphi_s) \times U},$$

$$= \frac{C(p_t;\varphi_t)}{C(p_s;\varphi_s)}$$

$$= \left(\frac{\sum_{i=1}^{N} \left(\frac{p_{it}}{\varphi_{it}} \right)^{1-\sigma}}{\sum_{i=1}^{N} \left(\frac{p_{is}}{\varphi_{is}} \right)^{1-\sigma}} \right)^{\frac{1}{1-\sigma}} \tag{7}$$

The important characteristics of the COLI in Equation (7) are: (A) its concavity with respect to the taste-adjusted prices, p_{it} / φ_{it}, and (B) its symmetric treatment of the taste-adjusted prices. The concavity of COLI comes from cost minimization. The symmetric treatment stems from our assumption that the differences among quantities in the utility function are represented by the taste parameter, φ_{it}. These two characteristics of the COLI imply that if the taste-adjusted prices become more heterogeneous, that is, if the taste-adjusted prices are more dispersed, the consumers benefit from the diversity, which enables them to attain the given utility level by smaller expenses. In other words, if the taste-adjusted prices become more dispersed, the COLI becomes smaller. Please note that this result comes not from the assumption of the CES preferences but from the equal treatments of the taste-adjusted prices. As Redding and Weinstein (2020) found, it is not difficult to generalize the CCV so that the utility functions can take the form of the translog function with variable taste parameters.

While the CCV by RW enables us to construct the COLI with variable tastes, there are several issues to be considered. First, we could not identify one of the taste parameters, φ_t, from data. For example, assume that we multiply all the taste parameters φ_t by a constant, say $\kappa > 0$, while φ_s is unchanged, since such a change is a monotonic transformation of the utility function at time t, we obtain identical demand functions at time t while the demand functions at time s are unchanged. However, the COLI will take a different value because it is a decreasing function of the taste parameters. This identification problem may seem severe because the choice of the normalization condition affects the index number. Comparing various different normalization condition, Kurtzon (2020) argues that an arbitrary choice of normalization can yield any desired CCV. Recently, Abe and Rao (2020) investigated the axiomatic properties of normalization conditions using RW. They found that the normalization condition in Equation (6) is the necessary and sufficient condition for the CCV; first to pass the commensurability test, the index number must be free from the measurement units of price and quantities, and second, to treat all the quantities equally in the normalization conditions. For example, instead of Equation (6), if we adopt the arithmetic mean, such as $(1/N)\sum_{i=1}^{N}\varphi_{it} = \varphi$, the CCV becomes sensitive to the choice of the measurement units of commodities such as pound or kilogram. Abe and Rao (2020) show that the CCV passes the transitivity test as well as the monotonicity test but fails the identity test. Therefore, in this study, we also use Equation (6) as the normalization condition.

The second issue is the interpretation of the taste term, the second term Equation (1). Martin (2020) argues that the magnitude of the taste term can be so large that the contribution of the price changes is swamped. Martin (2020) concludes that pure taste change effects are arguably out of the scope of a consumer price index and further clarifies the difference between the average price changes and the cost of living index. The CCV captures the effects of differences in taste, in addition to the effects of price changes. The taste term in the CCV reflects a strict concavity in the expenditure function. If the expenditure function is linear, that is, if σ is infinite, the taste term disappears from Equation (1). As RW shows, it is not difficult to generalize the assumption of CES preferences to a more general class of utility functions, such as Translog preferences. That is, the CCV can be regarded as the simplest case of the COLI with demand shocks, which provides us with information on the impact of demand shocks on economic welfare.

A demand shock for commodity i occurs at time t when the taste parameter changes, that is,

$$\varphi_{it} \neq \varphi_{it-1}.$$

Using Equations (4) and (6), we can estimate the taste parameter, φ_{it}, from the data of the expenditure shares and prices at time t. That is, we can observe changes in taste parameters over time. A natural measure of the degree of the demand shock at time t is the root mean square deviation (RMSD), such as,

$$RMSD_t = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\ln\,\varphi_{it} - \ln\,\varphi_{it-1})^2}. \tag{8}$$

If the RMSD increases, then the departure between the SV and COLI is expected to be greater. The actual effects of the demand shock on the price index can be captured by the taste term in Equation (1), $\sum_{i=1}^{N}\omega_{ist}^{*}(\ln\,\varphi_{is} - \ln\,\varphi_{it})$.

It is worth noting that to show the equivalence between the Sato-Vartia index and the COLI for the CES utility function, the taste parameters must be fixed over time. That is, the Sato-Vartia index or other superlative indexes become COLI only when the preferences are constant over time. In other words, the observed prices and quantities must always be on the time-invariant demand function, which is a strong assumption, particularly for the COVID-19 pandemic period in 2020. Using the CCV, we can construct the cost of living index without assuming constant demand curves.

4. Data

In this study, we used the scanner data of face masks provided by Intage Holdings Inc. The data set contains barcode level weekly sales and quantity information from nationwide retail stores in Japan. The scanner data provided by Intage is the largest point of sales data in Japan collected from more than 3,000 retail stores such as general merchandise stores, supermarkets, convenience stores, and drug stores all over Japan. Moreover, the retailers were chosen to get a nationally representative sample. We chose data for face masks between the week starting January 1, 2018, and the week starting June 8, 2020. Price information is obtained by dividing the weekly total sales at each store and the barcode by the quantities sold. When constructing such unit values, we must choose the data frequency. RW uses quarterly unit values while Diewert (2018) adopts a monthly frequency. Comparing price index numbers based on the unit values at various frequencies, Bradley (2005) found that a monthly unit value would lead to an upward bias in the cereal price index because of the aggregation of transactions at different prices. To mitigate the problems of using the unit values, we chose the weekly store-barcode level unit value, which were the finest data available. Although Japanese article number (JAN) code is supposed to be the unique identifier of products, sometimes, manufactures keep the identical JAN codes when they change the contents of the products. To deal with this problem, Intage creates an additional code, sequential code, to identify the difference of the commodities with the identical JAN codes if there are any differences. In this article, as the commodity identifier, we use the combination of both JAN and sequential codes. The total number of commodities x stores is about 47,000.

The general movements of the total sales of face masks are shown in Figure 3. In the week starting January 13, 2020, the demand for face masks surged. The period of the 2020 COVID-19 pandemic was set as the period between the week starting January 13, 2020, and the week starting May 18, 2020. This period is illustrated as the interval between the two vertical gray lines in Figure 3. The impact of the pandemic on the sales of face masks is clear in the figure. A surge in sales appeared in the week starting from January 13.

Table 1 reports the descriptive statistics for each weekly aggregated variable. As shown in the first row, the maximum value of total sales is enormous compared to the 95th percentile point. This distortion of sales distribution mainly appeared in the week when the demand for masks increased sharply in January 2020. In the second to the fifth row, we report the mean and standard deviation of the log change in prices and expenditure share in the common product calculated per week. The table shows that changes in expenditure shares are more volatile than changes in prices.

To see the changes in variables during the COVID-19 pandemic, we conducted the following regression.

$$y_t = \alpha + \beta D_t \sum_{i=2}^{12} \lambda_i M_t^i + \delta t.$$

The second term on the right-hand side, D_t, is a dummy variable that is set to unity during the 2020 COVID-19. The third term controls seasonality by the dummy variables, M_t^i, which represents the months from February to December. The fourth term represents the trend term. We use multiple variables as dependent variables to examine their changes during the COVID-19.

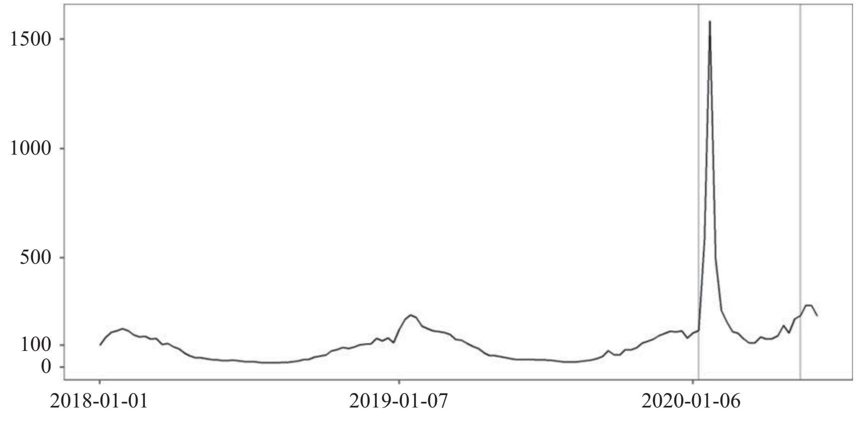


Fig. 3. Movements of the Total Sales of Face Masks.
Notes: Source: Scanner data provided by Image. The total sales in the first week of 2018 are normalized as 100.

Table 1. Descriptive statistics of face masks in Japan.

	Mean	Std. dev	Min	P5	P25	P50	P75	P95	Max
Total sales (million yen)	101	131	16.4	19.3	31.7	86.5	130	200	1330
Mean Δ (ln price) (%)	−0.03	0.26	−0.49	−0.43	−0.19	−0.05	0.12	0.39	1.00
Std. dev. Δ (ln price)	0.07	0.01	0.05	0.06	0.07	0.08	0.08	0.08	0.09
Mean Δ (ln share) (%)	0.77	6.48	−41.3	−6.09	−1.87	0.44	3.09	10.7	21.6
Std. dev Δ (ln share)	0.80	0.16	0.69	0.70	0.72	0.74	0.75	1.18	1.29

Notes: Scanner data of face masks between the week starting January 1, 2018, and the week starting June 8, 2020. Data is provided by Intage, covering approximately 3,000 retail stores all over Japan.

Table 2. Changes in variables during the COVID-19.

Dependent variable	Constant	Changes after the outbreak
ln(Sales)	4.90	0.615
	[4.69, 5.10]	[0.137, 1.090]
Mean ln Δp_{it} (%)	−0.0794	0.113
	[−0.183, 0.0245]	[−0.090, 0.315]
Std. dev. ln Δp_{it}	0.0708	−0.0164
	[0.0563, 0.0853]	[−0.0238, −0.0090]
Std. dev ln Δw_{it}	0.630	0.371
	[0.502, 0.758]	[0.302, 0.440]
L-P gap (%)	0.530	−0.772
	[0.290, 0.770]	[−1.110, −0.433]

Notes: Table 2 shows the results of regression analysis that measures the change in statistics during the COVID-19 outbreak. The seasonality and trend changes were controlled using monthly dummies and a trend term. In brackets, we report 95% confidence intervals estimated by heteroskedasticity and autocorrelation consistent standard errors. The first column shows the statistics used as dependent variables. L-P Gap is the difference between the logarithms of the chained Laspeyres and Paasche indexes. The second column shows the estimated value of the constant term in January. The third column shows the estimated coefficients of the dummy variables after January 13, 2020, that is after the outbreak.

The results of the regression analysis are shown in Table 2. The sales of face masks increased by 61.5% during the 2020 COVID-19. The arithmetic mean of price changes did not show a statistically significant change and the standard deviation of price changes fell. Contrary to price changes, the standard deviation of the change in expenditure share increased during the disaster. As shown in Figure 2, the L-P gap decreased significantly.

Figure 4 reports the RMSD of the changes in logged prices and logged expenditure shares. During the first week of the 2020 COVID-19 pandemic, prices become less volatile while the fluctuation of market shares surged. If the demand curve is stable, smaller volatility in prices should come with stable market shares. Thus, Figure 4 suggests that the demand curve changes during the first week of the COVID-19 pandemic.

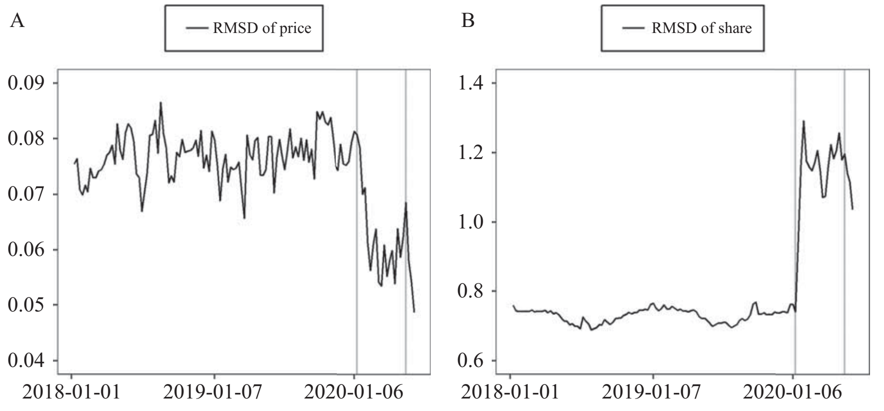


Fig. 4. Movements of the RMSD of prices and shares of face masks in Japan.
Notes: Source: Scanner data provided by Intage.

5. Empirical Results

Figure 5 shows the weekly change rates of several chained price indexes including the CCV. Panel A shows the movements of the simple geometric average price, the Jevons index, which is known to be free from chain drift. Panels B and C report the movements of the Fisher and SV indexes, respectively. The Fisher and SV indexes were close to each other. Although not depicted in the figure, the Tornqvist index is also very close to the Fisher index. The Jevons, Fisher, and SV indexes exhibited a sharp increase in the first week of the COVID-19 period. However, when we consider changes in preferences, the movements of prices become very different. The CCV shows a sharp drop in prices during the first week of the COVID-19 pandemic, which then increased substantially. The point estimate of the elasticity of substation is 5.87, which is between the 25th and 50th percentiles reported by Redding and Weinstein (2020). We adopt the methods developed by Feenstra (1994) to estimate the elasticity of substitution using balanced data during 2018–2019. We chose the periods because if we include observations during 2020, the estimates become unstable. The constant elasticity over time is surely a restrictive assumption. However, the estimation methods by Feenstra (1994) and Redding and Weinstein (2020) as well as the CCV critically depend on the assumption that elasticity of substitution is constant over time. The considerations of variable elasticity will be our future tasks. Table 3 reports the movements of the indexes including changes in sales during January-February, 2020. Appendix Table reports the index numbers of the entire sample periods. In the week starting from January 20, the sales of face masks surged, and an over 50% increase from the previous week was recorded. However, traditional price indexes, such as the Fisher price index changed little from the previous week. The CCV dropped by 8.37% in the week when the sales of face masks surged.

Figure 6 shows the movement of the RMSD of the taste parameters (Panel A) and the taste shock defined as $\sum_{i=1}^{N} \omega_{ist}^{*}(\ln \varphi_{is} - \ln \varphi_{it})$ (Panel B). First, from Panel A, we can observe that the RMSD of the taste parameters are always positive even before the COVID-19 period, which indicates that the assumption of the constant taste parameters is violated. From the figure, we can also observe that in January 2020, large negative taste

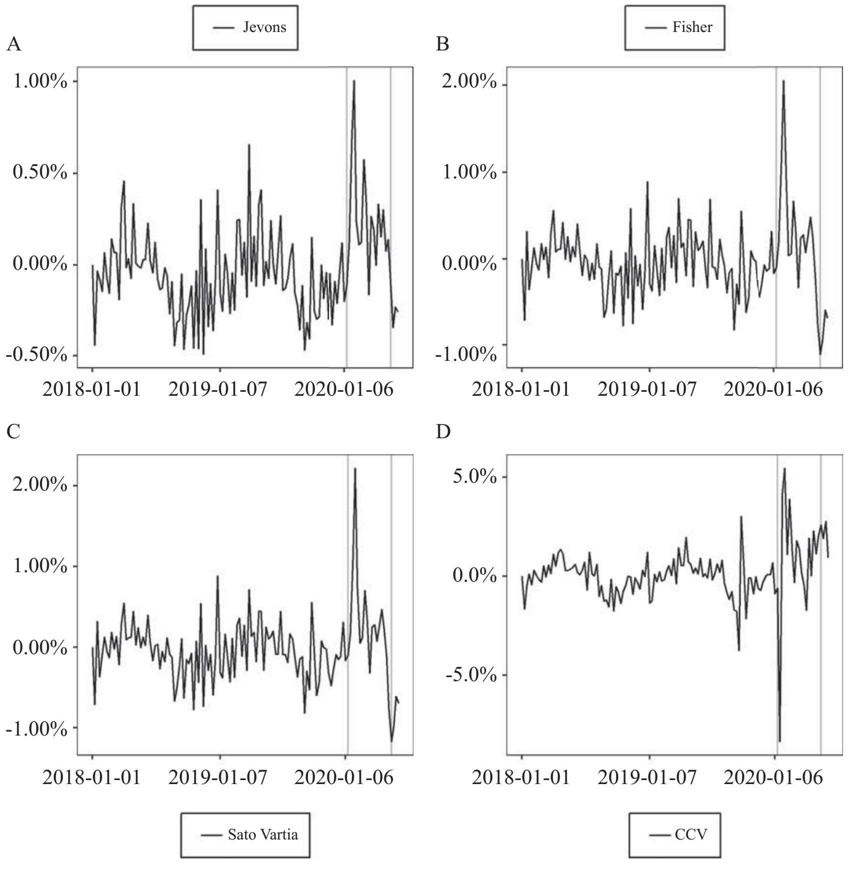


Fig. 5. Weekly change rates of several price indexes of face masks.
Notes: Weekly change rates of chained indexes. CCV stands for CES common variety price index defined in Equation (1).

Table 3. Comparisons of price indexes.

	Sales	Jevons	Fisher	Sato Vartia	CCV
2020/1/13	2.97	−0.11	−0.11	−0.10	−0.60
2020/1/20	53.92	0.12	0.18	0.17	−8.37
2020/1/27	43.46	0.68	1.09	1.08	4.21
2020/2/3	−50.59	1.00	2.05	2.22	5.45
2020/2/10	−27.88	0.23	0.91	0.73	1.10
2020/2/17	−9.90	0.11	0.03	0.05	3.87
2020/2/24	−10.38	0.12	0.06	0.12	1.27
Average. May, 2018	−2.93	0.00	0.09	0.10	0.33
May, 2019	−3.52	0.06	0.13	0.13	0.33
May, 2020	4.29	−0.06	−0.76	−0.76	1.92

Notes: The weekly rates of change (%) of the chained indexes. More comprehensive numbers are reported in the Appendix Table.

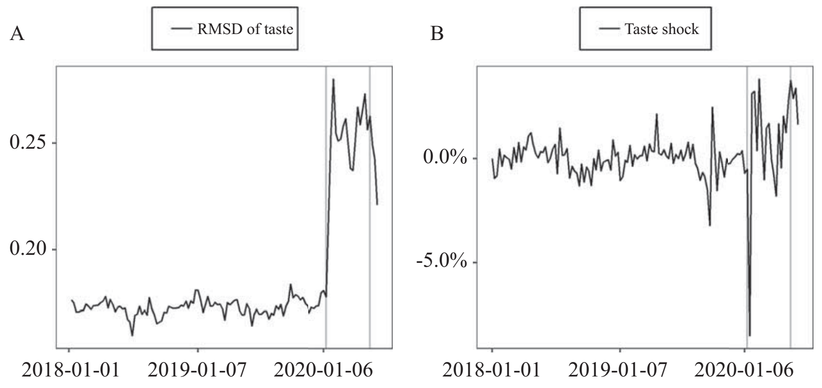


Fig. 6. The RMSD of Taste Parameters and the Taste Shock.
Notes: The RMSD of taste is defined in (8), while the taste shock is defined as the second term of the R.H.S. of (1)

shocks with a surge in the RMSD occurred, which makes the CCV drop to a great extent. The intuition behind the decline the CCV is as follows. As Equation (7) indicates, the CCV is a concave and symmetric function of the taste-adjusted prices, p_{it}/φ_{it}. This implies that people can obtain greater utility when taste-adjusted prices are more dispersed. Before the pandemic, some face masks were more popular than others, regardless of the price, which was represented as variations in the taste parameters among commodities. When people realized that face masks are effective in avoiding COVID-19 infections, their evaluation of each face mask changed to a great extent, which led to greater variations in the taste-adjusted prices than before the pandemic. The greater the dispersion, the more CCV dropped as discussed in Section 4. Opposite effects were observed in May 2020. As Table 3 shows, the CCV became positive while the standard price indexes were negative. At that time, the RMSD of tastes began to decrease, which led to smaller dispersions in the taste-adjusted prices, thus, resulting in positive taste effects.

Finally, Figure 7 shows the levels of several price indexes. In the figure, the chained Fisher and SV indexes do not depart from the Jevons index much, which suggests that chain drifts of the face mask are not serious. The deep trough of the CCV in January 2020 disappeared within a few weeks. The Jevons, Fisher, and SV indexes became smaller after the pandemic, while the CCV continued to increase. In early June 2020, the Jevons, Fisher, and SV indexes were around 97, while the CCV was more than 115. In other words, the magnitude of the cumulative effects of taste shocks was large.

6. Conclusion

By investigating the prices and quantities of face masks in Japan during the serious threat of the COVID-19 pandemic in 2020, we considered the impact of demand shocks on the COLI. We found that the demand shock that occurred during this period was large, which makes the Laspeyres index to be smaller than the Paasche index. The demand shocks measured by the changes in the taste parameters for the CES utility function created large taste effects that are not captured in the Sato-Vartia or superlative indexes such as the Fisher index. While the prices of face masks decreased in the Jevons and Fisher indexes in

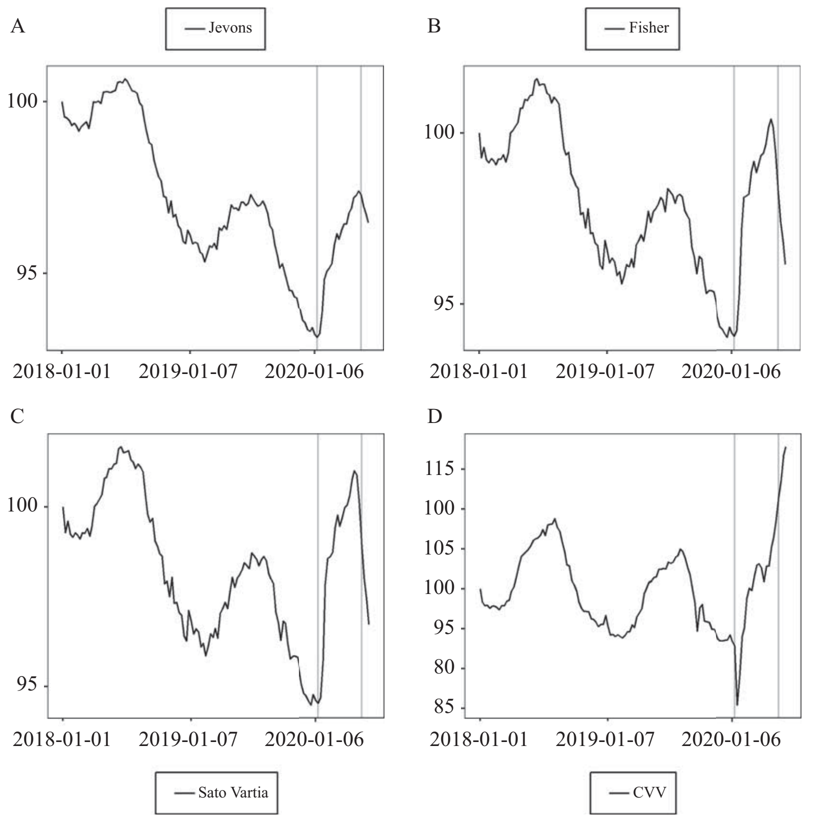


Fig. 7. Level of chained price indexes.
Notes: The levels were obtained by taking the cumulative logged weekly changes in the chained indexes. The indexes were normalized to 100 in the first week of 2018.

May 2020 by 0.06% and 0.76% per week, respectively, the COLI increased by 1.92% per week.

This study has several limitations. When the demand for face masks surged, face masks were rationed, which complicates the construction of the COLI. If we could identify a product that was not rationed during the sample period, it could be possible to adopt the method developed by Tobie and Houthakker (1950–1951) and Neary and Roberts (1980) to construct the cost of living under rationing. However, as long as we use scanner data, the existence of rationing cannot be identified. If rationing occurs, COLI tends to be greater than the index without rationing. Therefore, our estimates of the cost of living in this study should be regarded as a lower bound. Second, although the CCV allows for variable taste parameters, we need to assume that the elasticity of substitution is constant over time, which is a restrictive assumption when a strong demand shock occurred. Although we could assume some form of stochastic processes for the elasticity of substitution and conduct estimations, we have not been able to obtain stable and robust estimates. Finally, we did not discuss the variety of effects developed by Feenstra (1994) and Redding and Weinstein (2020) on COLI. Appendix (Subsection 7.2) reports some results of the various effects; however, we have obtained unreasonably large negative variety effects on the COLI. Investigations of the effects of rationing, variable elasticities over time, and the effects of changing variety will be our next task.

7. Appendix

7.1. Derivation of Equation (4)

The demand function generated by the Equation (5) can be written in terms of the expenditure share as follows,

$$\ln w_{it} = (\sigma - 1)(\ln P_t + \ln \varphi_{it} - \ln p_{it}), \tag{9}$$

where $\ln P_t = \ln C(p_t; \varphi_t)$.
(9) can be rewritten as

$$\ln \varphi_{it} = \frac{1}{\sigma - 1} \ln \left(\frac{w_{it}}{w_{1t}} \right) + \ln \left(\frac{p_{it}}{p_{1t}} \right) + \ln \varphi_{1t}.$$

Combined with the normalization condition in Equation (6), we can obtain Equation (4).

7.2. The Variety Effects

During the COVID-19 pandemic, due to the increasing demand for face masks, the variety of masks changed over time. One method of quantifying the effects of the changes in the product variety on the price index is provided by Feenstra (1994). Redding and Weinstein (2020) also consider a case wherein the variety of commodities changes over time. This is the second CUPI in their paper. The RW index, which is the COLI when the product variety changes, is defined as follows,

$$\ln RW(p_s, q_s, p_t, q_t) = \ln CCV(p_s, q_s, p_t, q_t) + \frac{1}{\sigma - 1}(\ln(\lambda_t^s) - \ln(\lambda_s^t)). \tag{10}$$

Here, λ_t^s is the ratio of the expenditure share of common products in the periods t and s to the total expenditure at time t,

$$\lambda_t^s = \frac{\sum_{i \in C_{t,s}} p_{i,r} x_{i,r}}{\sum_{i \in I_t} p_{i,r} x_{i,r}} \tag{11}$$

I_t : Set of all commodities at time t.

$C_{t,s}$: Set of common commodities at time t and s.

The second term on the right-hand side of Equation (10) is called the log λ ratio. Note that if we replace ln RW in Equation (10) with SV, then ln RW becomes the price index by Feenstra (1994).

The movements of λ, the log λ ratio, Feenstra's index, and RW index are reported in the Appendix. As is clear from the figure, the magnitudes of the various effects during the COVID-19 period are extremely large. We suspect that this occurs because of the product turnover of the identical products. Suppose face mask A was sold at store X. Then, the next week, mask A did not appear in the store because of the huge demand for the mask. Two weeks later, mask A returned to store X. Although, this turnover is not related to the introduction of the new product, the log λ ratio is interpreted as the introduction of new products, thus affecting the cost of living index.

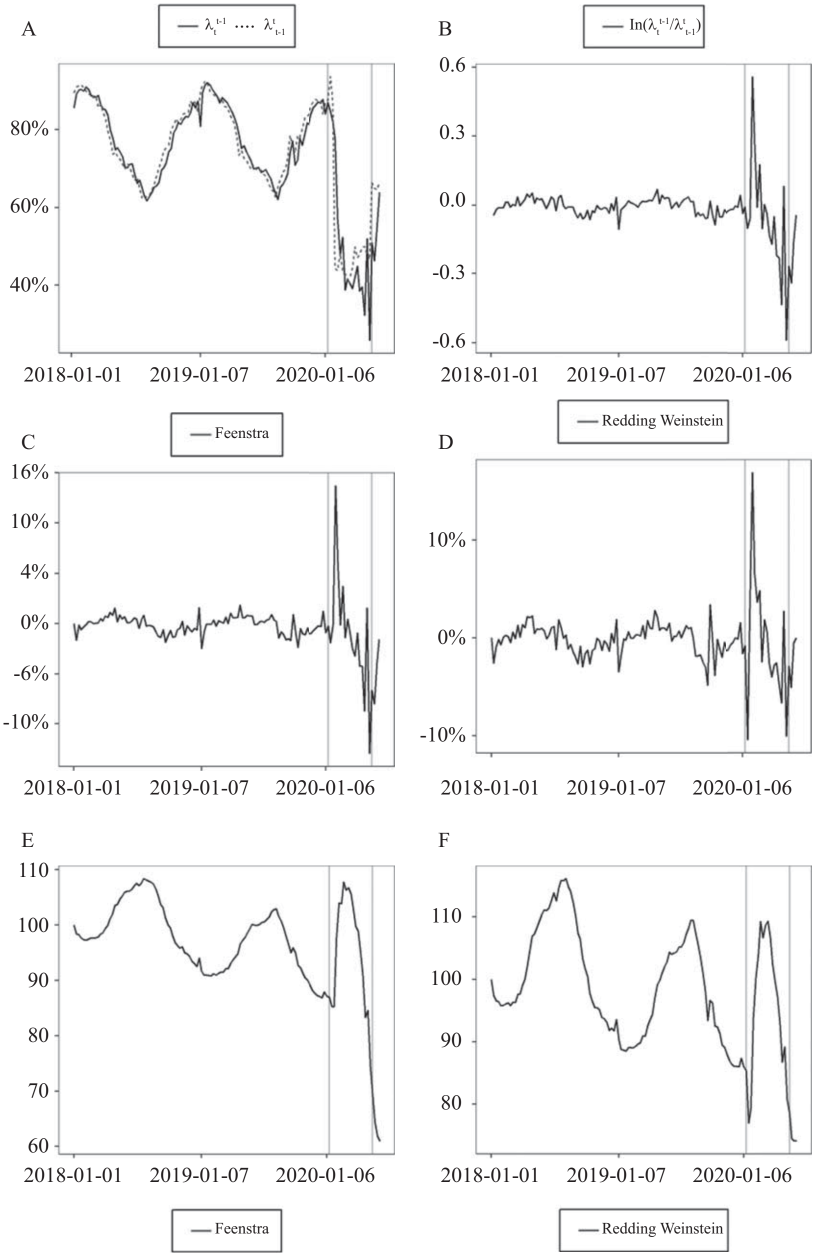


Fig. 8. Variety Effects

Table 4. Index numbers

	Sales	Jevons	Fisher	Sato-Vartia	CCV		Sales	Jevons	Fisher	Sato-Vartia	CCV
2019/1/7	0.190	-0.002	-0.003	-0.003	-0.014	2020/1/6	0.073	-0.002	-0.002	-0.002	-0.009
2019/1/14	0.100	-0.003	-0.004	-0.004	-0.012	2020/1/13	0.030	-0.001	-0.001	-0.001	-0.006
2019/1/21	0.039	0.001	0.002	0.002	0.001	2020/1/20	0.539	0.001	0.002	0.002	-0.084
2019/1/28	-0.021	0.000	-0.001	-0.001	-0.003	2020/1/27	0.435	0.007	0.011	0.011	0.042
2019/2/4	-0.084	-0.003	-0.004	-0.004	0.002	2020/2/3	-0.506	0.010	0.020	0.022	0.054
2019/2/11	-0.026	0.000	0.001	0.001	-0.002	2020/2/10	-0.279	0.002	0.009	0.007	0.011
2019/2/18	-0.026	-0.003	-0.004	-0.004	-0.002	2020/2/17	-0.099	0.001	0.000	0.001	0.039
2019/2/25	-0.007	0.002	0.002	0.003	0.003	2020/2/24	-0.104	0.001	0.001	0.001	0.013
2019/3/4	-0.015	0.003	0.004	0.004	0.005	2020/3/2	-0.023	0.006	0.007	0.007	-0.003
2019/3/11	-0.022	-0.001	-0.001	-0.001	0.000	2020/3/9	-0.082	0.003	0.003	0.003	0.018
2019/3/18	-0.072	0.001	0.003	0.003	0.009	2020/3/16	-0.066	-0.002	-0.003	-0.003	0.014
2019/3/25	-0.014	-0.002	-0.003	-0.003	-0.004	2020/3/23	0.009	0.003	0.002	0.002	0.002
2019/4/1	-0.055	0.007	0.007	0.007	0.014	2020/3/30	0.091	0.002	0.003	0.003	-0.005
2019/4/8	-0.058	-0.001	0.001	0.001	0.005	2020/4/6	-0.031	0.000	0.001	0.001	-0.017
2019/4/15	-0.041	0.002	0.002	0.002	0.005	2020/4/13	0.006	0.003	0.003	0.003	0.019
2019/4/22	-0.123	-0.001	-0.002	-0.002	0.020	2020/4/20	0.038	0.002	0.005	0.005	0.000
2019/4/29	-0.087	0.003	0.005	0.004	0.007	2020/4/27	0.126	0.003	0.002	0.002	0.023
2019/5/6	-0.003	0.004	0.004	0.004	0.006	2020/5/4	-0.087	0.001	-0.002	-0.001	0.011
2019/5/13	-0.052	-0.001	-0.003	-0.003	0.002	2020/5/11	0.151	0.001	-0.008	-0.008	0.020
2019/5/20	-0.048	0.000	0.003	0.003	0.004	2020/5/18	0.029	-0.001	-0.011	-0.012	0.026
2019/5/27	-0.038	-0.001	0.001	0.001	0.001	2020/5/25	0.079	-0.003	-0.009	-0.010	0.019
2019/6/3	-0.043	0.002	0.001	0.001	0.009	2020/6/1	-0.001	-0.002	-0.006	-0.006	0.028
2019/6/10	0.016	0.000	0.002	0.002	0.000	2020/6/8	-0.082	-0.003	-0.007	-0.007	0.009
2019/6/17	-0.011	-0.001	-0.001	-0.001	0.001						
2019/6/24	-0.017	0.001	-0.003	-0.001	-0.001						
2019/7/1	-0.011	0.003	0.007	0.004	0.009						
2019/7/8	-0.003	-0.001	-0.001	-0.001	-0.002						
2019/7/15	-0.025	-0.001	-0.001	-0.001	0.002						
2019/7/22	-0.055	-0.001	-0.002	-0.002	0.006						
2019/7/29	-0.037	0.000	0.002	0.002	0.002						
2019/8/5	-0.038	0.001	0.000	0.001	0.008						
2019/8/12	-0.005	-0.002	-0.001	-0.001	-0.003						
2019/8/19	0.059	-0.002	-0.004	-0.004	-0.008						
2019/8/26	0.055	-0.004	-0.002	-0.002	-0.012						
2019/9/2	0.054	-0.001	-0.001	-0.001	-0.008						
2019/9/9	0.062	-0.005	-0.008	-0.008	-0.017						
2019/9/16	0.108	-0.003	-0.003	-0.003	-0.018						
2019/9/23	0.184	-0.004	-0.005	-0.005	-0.037						
2019/9/30	-0.123	0.002	0.005	0.006	0.030						
2019/10/7	-0.008	-0.003	-0.001	-0.001	0.005						
2019/10/14	0.157	-0.003	-0.006	-0.006	-0.021						
2019/10/21	-0.009	-0.003	-0.005	-0.004	-0.001						
2019/10/28	0.055	0.000	0.001	0.001	-0.001						
2019/11/4	0.091	-0.002	0.000	0.000	-0.009						
2019/11/11	0.033	0.000	0.000	0.000	0.000						
2019/11/18	0.033	-0.003	-0.003	-0.003	-0.006						
2019/11/25	0.051	0.000	-0.004	-0.005	-0.007						
2019/12/2	0.028	-0.003	-0.003	-0.003	-0.003						
2019/12/9	0.028	-0.001	-0.001	-0.001	0.000						
2019/12/16	-0.009	-0.002	-0.001	-0.002	0.001						
2019/12/23	0.012	-0.001	-0.001	-0.001	0.001						
2019/12/30	-0.096	0.001	0.003	0.003	0.007						

Notes: The weekly rates of change (%) of the chained indexes.

8. References

Abe, N., and D.S.P. Rao. 2020. *Generalized Logarithmic Index Numbers with Demand Shocks-Bridging the Gap between Theory and Practice.* Tokyo: Hitotsubashi University. (RCESR Discussion Paper Series NO.DP20-1) Available at: https://risk.ier.hit-u.ac.jp/Japanese/pdf/dp20-1_rcesr.pdf (accessed February 2022).

Balk, B.M. 1989. "Changing Consumer Preferences and the Cost-of-Living Index: Theory and Nonparametric Expressions." *Journal of Economics* 50(2): 157–169. DOI: https://doi.org/10.1007/BF01239124.

Bradley, R. 2005. "Pitfalls of using unit values as a price measure or price index." *Journal of Economic and Social Measurement* 30: 39–61. DOI: https://dx.doi.org/10.3233/-JEM-2005-0243.

Diewert, W.E. 2018. *Scanner Data, Elementary Price Indexes and the Chain Drift Problem.* Vancouver: Vancouver school of Economics. (Microeconomics.ca working papers erwin_diewert-2018-10, revised 25 Oct 2018.) Available at: https://pubdocs.worldbank.org/en/849511558350534104/pdf/ICP-TAG03-S34-RD-Diewert-on-Scanner-Data-and-Elementary-Price-Indexes.pdf (accessed February 2022).

Feenstra, R.C. 1994. "New Product Varieties and the Measurement of International Prices." *American Economic Review* 84(1): 157–177. Available at: https://www.jstor.org/stable/2117976 (accessed February 2022).

Fisher, F.M., and K. Shell. 1972. "Taste and Quality Change in the Pure Theory of the True Cost of Living Index." In *Price Indexes and Quality Change*, edited by Zvi Griliches: 16–54. Cambridge: Harvard University Press.

Ivancic, L., W.E. Diewert, and K.J. Fox. 2011. "Scanner data, time aggregation andthe construction of price indexes." *Journal of Economerics* 161(1): 24–35. DOI: https://doi.or/10.1016/i.ieconom.2010.09.003.

Kurtzon, G. 2020. *The Problem with Normalizing Preferences that Change in a Cost-of-Living Index.* Washington DC: BLS (ECONOMIC WORKING PAPERS: 534) Available at: https://www.bls.gov/osmr/research-papers/2020/pdf/ec200160.pdf (accessed February 2022).

Martin, R.S. 2020. *Revisiting Taste Change in Cost-of-Living Measurement.* Washington DC: BLS, Economic Working Papers: 515. Availabe at: https://www.bls. gov/osmr/research-papers/2019/pdf/ec190050.pdf (accessed February 2022).

Neary, J.P., and K.W.S. Roberts. 1980. "The Theory of Household Behaviour under Rationing." *European Economic Review* 13(1): 25–42. DOI: https://doi.org/10.1016/0014–2921(80)90045-8.

Philips, L. 1974. *Applied Consumption Analysis*. Amsterdam: North-Holland.

Redding, S.J., and D.E. Weinstein. 2020. "Measuring Aggregate Price Indexes with Taste Shocks: Theory and Evidence for CES Preferences." *The Quarterly Journal of Economics* 135(1): 503–560. DOI: https://doi.org/10.1093/qie/qiz031.

Tobie, J., and H.S. Houthakker. 1950–1951. "The Effects of Rationing on Demand Elasticities." *The Review of Economic Studies* 18(3): 140–153. DOI: https://doi.org/10.2307/2295974.

Received July 2020
Revised March 2021
Accepted June 2021

Journal of Official Statistics, Vol. 38, No. 1, 2022, pp. 319–351, http://dx.doi.org/10.2478/JOS-2022-0017

Early Real Estate Indicators during the COVID-19 Crisis

Norbert Pfeifer[1] *and Miriam Steurer*[1]

In this article we construct a number of early housing-market indicators from daily-scraped list-price data and investigate how they inform on housing-market trends prior to the release of official transaction data. We minimize the impact of non-market conform list prices by our data selection- and cleaning routine and thereby improve the overlap between list and transaction data. In addition to some well-known real-estate indicators (price and rent indices, price-rent ratios, market volume, *time-on-market*, and market turnover) we develop a new market sentiment indicator which we construct from the direction and frequency of online price changes for individual listings. We then use this group of indicators to investigate how housing markets in London and Vienna react to the COVID-19 pandemic during 2020. For London these indicators show high volatility during much of 2020 and we find that the sales and rental markets drift apart: The sales market shows an overall positive price trend, while the rental market becomes significantly weaker after the onset of the COVID-19 crisis. For Vienna we find that all indicators signal positive market developments for the first eight months after the start of the COVID-19 pandemic.

Key words: List price index; COVID-19; real-estate indicators; online sentiment indicator; leading indicators.

1. Introduction

For government agencies and central banks, it is important to be informed about the housing market, as housing market crises can have serious consequences for the rest of the economy and threaten the banking system via mortgage defaults (see e.g., discussion in Mishkin 2011). For this reason, National Statistical Agencies, as well as National Banks, compute a number of indicators to track the development of the housing market. House price indices are the most important of them.

In official statistics, the preferred option to track the development of housing prices is via price indices that are constructed with transaction data (Eurostat 2013), but the time-lags involved in processing transaction data pose obstacles to the timely production of real estate indices and indicators. Several months can pass from a property being listed until an offer is agreed upon and several more until contracts are exchanged, completed, and entered into land registries. The typical time-lag between the availability of offer and transaction prices varies between housing markets; for example, Shimizu et al. (2016) find an average time-lag of seven months between list and transaction prices for apartment buildings in Tokyo, Brounen and Kok (2011) find an average lag of four, and a half months

[1] Department of Economics, University of Graz, Universitatsstrasse 15/F4, 8010 Graz, Austria. Emails: norbert.pfeifer@uni-graz.at and am.steurer@uni-graz.at

Acknowledgments: We acknowledge financial support for this project from the Austrian Research Promotion Agency (FFG), grant #10991131. We also thank ZT datenforum for giving us access to the Austrian transaction data set.

© Statistics Sweden

for the Netherlands, and a Bank of England estimate (presented in Eurostat 2013, 103) indicates a five, and a half months lag in the United Kingdom (UK). The implication of this time-lag is that governments and central banks often need to announce policies before official housing market indicators are available. This matters particularly during periods of crisis, such as the COVID-19 pandemic.

Real-estate data from online platforms are a potential alternative to transaction data for measuring developments in the housing market. Such list-price data have multiple benefits: They are immediately available, continuously updated, and often much cheaper to access than transaction prices. In addition to information on prices, they provide a range of other information that typically far exceeds the characteristics available in transaction data (Shimizu et al. 2016; Kolbe et al. 2021). Such listed housing data also allow for insights into market dynamics that are not available from transaction-based data sets. For example, frequent scraping enables the tracking of individual properties over time.

Some researchers have already demonstrated the usefulness of online data for spotting early signs of housing market changes. The work by Choi and Varian (2012), Chauvet et al (2014), Wu and Brynjolfsson (2015), or Lyons (2019) fall into this category. Similarly, Van Dijk and Francke (2018) illustrate with Dutch data that market tightness indicators built from online search data (online *list volume* and number of clicks per postcode) are able to predict changes in transaction prices and market liquidity. Others find list-price data less reliable for subsequent housing market development (see e.g., Knight et al. 1994; Kolbe et al. 2021).

The relationship between list and transaction prices is not straightforward. Sellers are more reluctant to lower prices during a market down-turn than increase them during a market rise (Genesove and Mayer 2001). As a result, the gap between list-prices and transaction prices varies anti-cyclically, and during housing booms, transaction prices can even exceed asking prices (for discussion and empirical examples, see e.g. Haurin et al. 2013). Additionally, the spread between list-prices and transaction prices also tends to be higher at the upper end of the market, which further complicates the relationship. This tendency can be explained by sellers' uncertainty over buyers' willingness to pay, which tends to be higher for atypical and highend properties (Haurin et al. 2010). The higher spread for high-end properties can also be explained by market liquidity. This inverse relationship between market liquidity and the price spread is wellknown in financial markets (see e.g., Bagehot 1971). And finally, each market is different, which implies that evidence from one market cannot easily be transferred to others. The regulatory environment and real-estate agent incentives also play a role in how list prices are pitched to the marketplace.

In this article we present a collection of leading indicators and apply them to two distinct housing markets: London and Vienna. Our sentiment indicator, constructed from the relative number of positive and negative price changes occurring in the market – is new to the literature. The other indicators are well known, but list-price data allow us to track them in a continuous fashion throughout the housing cycle.

We build our list-price data sets for both cities by scraping data from leading online housing platforms. Our list-price data run from the end of 2018 to the end of 2020. For 2019 we also have access to transaction data for both cities and can illustrate the relationship between the two data distributions (see Section 3). We show that our data preparation routine, which consists of excluding extreme values and limiting each listed property to enter the data set only once (the last entry before it leaves the market), moves the price distribution of the list-price data set towards that of the transaction price data set. We believe that this closeness in price distributions translates into more accurate leading indicators.

However, our main aim in this article is not to investigate the relationship between transaction prices and list prices per se, but to illustrate how to derive list price indicators that can be used to "feel the temperature" of the market before official transaction prices are available. We do not suggest that the levels of these indicators should be interpreted as substitutes to their transaction-based counterparts. Rather, we see these indicators as performing the role of lights on a dashboard that informs on the general market direction before transaction data become available: The more indicators align in their message, the stronger the message on market tendency.

The pictures these indicators paint for the London and Vienna housing markets in 2020 after the start of the COVID-19 crisis in February/March are as follows: Neither London nor Vienna experienced a negative price development between February and December 2020. This is surprising, since previous research has shown that – at least in the short run – pandemics tend to have a negative impact on house prices (Wong 2008; Francke and Korevaar 2021). There are however significant differences in the development of the London and Vienna real estate markets during 2020.

In Vienna, both the sales and rental market exhibited all signs of "hot" property markets throughout 2020 in the form of increasing (list) prices and increasing price-rent ratios. Liquidity and market sentiment levels (for the sales as well as the rental market) are stable throughout 2020 – but at higher levels than during 2019.

The London property market indicators behaved more nervously after the start of the COVID-19 pandemic: The *list-price index* development for 2020 was stable in the sales market but negative in the rental market, and list-price volatility increased in the sales as well as the rental market after the start of the COVID-19 crisis. Volatility is also visible in the development of the liquidity indicators: *list volume*, *list time-on-market*, and *list turnover rate*. However, after the initial drop in liquidity during the first lock-down period in April 2020, liquidity increased for the second half of 2020. Finally, the *sentiment indicator* suggests a stable tendency in the sales market but a slightly negative (and quite volatile) trend in the London rental market for much of 2020.

2. List-data Indicators

We compute and discuss the following six list-data indicators:

1. List-price index as well as a list-rent index
2. List price-rent ratio
3. List volume
4. List time-on-market
5. List turnover rate
6. Market sentiment indicator based on list price changes

Transaction prices are the result of negotiations between buyers and sellers, and the *price indices* (or *rent indices*) constructed from them measure the development of these market prices over time. List-prices (list-rents), on the other hand, illustrate the market perception of sellers (landlords) only, and the resulting *list-price indices* (*list-rent indices*) measure how this perception changes over time. The direct impact of the demand side of the market is missing in list-data and as a result, the relationship between list and transaction prices is not straightforward.

As sellers are typically slower to adjust their price ideas in a downward market than in an upward one (Genesove and Mayer 2001), we do not expect the *list-price index* to show negative price movements as fast as positive price movements. This reluctance by sellers to reduce prices during a market down-turn also implies that list-prices are not always good indicators of market turning points (Knight et al. 1994). This asymmetry in list-price reaction to underlying market tendency probably explains why the literature provides a mixed verdict on whether list prices are good leading indicators for transaction prices: While some researchers find them to be good leading indicators (see e.g., Ahmed et al. 2016; Anenberg and Laufer 2017; Lyons 2019) others find the opposite (Kolbe et al. 2021).

The *price-rent ratio* is the ratio of what a property sells for relative to its rental price. When constructed with list-price data, the *list-price ratio* answers the question of how the list price of a property relates to the annual list rent. We again only use the last price entries (for sales as well as rental properties) to construct this ratio.

It is interesting to monitor the evolution of the price-rent ratio in housing markets, as it is an indicator of whether housing markets are valued fairly or in a bubble. The relative velocity and magnitude of price changes in the housing market also leads to particular patterns: In particular, prices tend to react faster and stronger to market pressures than rents, typically leading to *decreasing* price-rent ratios during down-turns (see e.g., Geneseve and Mayer, 2001; Engsted and Pedersen 2015; Hill and Syed 2016; Francke and Korevaar 2021).

We compute list *price-rent* ratios using the "double-prediction" (also called "double-imputation") approach a hedonic method that is well established in the hedonic price index literature (see e.g., Hill and Syed 2016). Its main benefit is that it partially addresses the omitted variable problem, which can be particularly important when comparing properties from two different data sets. We employ it for the price-rent ratio calculation because we are concerned that properties in the sales market systematically outperform those in the rental market on characteristics that are not contained in our data sets; for example, sales properties that are better maintained or have higher quality fittings. We do not expect such systematic biases to be present *within* each data set (e.g., when going from one year to the next), which is why we use the more straightforward Time Dummy hedonic method to establish hedonic indices when only one data set is concerned.

List volume, *list time-on-market*, and *list turnover rate*, are three measures of market liquidity. Liquidity in the housing market tends to be positively correlated with price, a phenomenon that is explained in the literature by "loss aversion" – the reluctance of sellers to sell at (nominal) losses (see Genesove and Mayer 2001) – and/or equity constraints – individuals being ''locked" into a mortgage unable to move (Stein 1995). Each indicator captures a separate aspect of housing market liquidity: *List volume* measures the number of unique properties on the market at a particular point in time or during a certain time period; it is a measure of supply and thus of the liquidity going *into* the market. *Time-on-market* measures how fast properties are selling and thus captures the liquidity *leaving* the market. *Time-on-market* tends to be negatively related to house prices, and has long been used as a measure of liquidity in the real-estate market (Miller 1978; Lippman and McCall 1986; Genesove and Mayer 2001, 1997). The market *turnover rate* is another popular measure of housing market liquidity (Genesove and Mayer 2001; De Wit et al. 2013; Van Dijk and Francke 2018). An alternative name for the turnover rate in the literature is rate-of-sales. It combines volume and *time-on-market* aspects of liquidity, but loses some detail while doing so.

When scraping data from online property platforms in regular intervals (e.g., daily or weekly), the history of properties on the market can be tracked. We use these individual histories to construct a new type of *sentiment indicator* based on the number and direction of price changes at the individual level. This indicator is new to the literature, and we believe it is a useful addition to monitor the "nervousness" of sellers and landlords in the housing market. Overall, the market sentiment analysis confirms that sellers are more likely to lower prices the longer properties remain on the listing platform. However, the rate at which this happens differs significantly between London and Vienna. London sellers and landlords frequently change prices while their counterparts in Vienna rarely do so.

We construct all list-market indicators on a monthly basis for the London and Vienna market throughout 2019 and 2020. While indicators mostly agree in their assessment of the market, sometimes they point in different directions. The findings of this manuscript therefore reiterate the importance of analysing multiple indicators before judging the behaviour of the housing market based on data from online platforms.

3. Data

For list data, we scrape from the leading platform in each market: for Vienna this is Willhaben (2021), for London it is Zoopla (2021). We started to collect sales as well as rental offers for both cities in the autumn of 2018 and include data-updates until December 14, 2020. We focus on apartment transactions rather than houses as the majority of transactions in both markets are for apartments. In 2019, 82% of residential transactions where for apartments in Vienna, in London the proportion was 59%.

When constructing the price indices and price-rent-ratios we want to get as "close" as possible to true market transaction prices. To do this, we track each property based on its url-address and include only the last price of a property before it leaves the platform. Thus, each property in the data set is included only once in the price index for the entire two-year time period. See Table 1 for the relationship between total entries and last entries. Other indicators need the first as well as the last entry on the platform (e.g., list *time-on-market*), or even the entire pricing history (e.g., *market sentiment indicator*).

In terms of data cleaning, we proceed as follows: For London we remove all listings with more than seven bedrooms or five bathrooms. In the Vienna data set – where we have information about size in square meters – we include all apartments between 20- and 180 square meters. For the list-price index calculations, we keep only the last entry before a property leaves the platform; then, we exclude the biggest price outliers by removing the top and bottom 3% of the price distribution in the London data set, and the top and bottom 3% of the price/square meter distribution in the Vienna data set. We follow the same steps for the transaction data sets, where the only data entry of a property is automatically the last entry.

Table 1 provides an overview of our list-price data set before and after the cleaning stage. More details on the data and cleaning stages can be found in Appendix (Subsection

Table 1. Descriptive statistics of data sets (2018–2020).

City		Apartments for sale			Apartments for rent		
		Raw data set		Cleaned data set	Raw data set		Cleaned data set
		Total	Unique		Total	Unique	
London	Observations	4,119,119	335,506	167,391	4,233,084	939,416	589,199
	Median price (Pounds)	590,000	525,000	495,000	1,820	1,733	1,772
	CV	2.39	2.29	0.56	1.58	1.39	0.43
Vienna	Observations	434,69	163,640	137,650	347,935	177,280	151,176
	Median price (EUR)	345,000	320,000	299,990	949.48	899	898
	CV	1.40	1.31	0.53	1.05	1.12	0.42

Note: We scraped the London market daily, while the Vienna market was scraped only once per week.
Note: To allow a comparison between the price distributions with different units of measure, we also include the Coefficient of Variation (CV).

6.2.). Note that, as apartments are typically advertised for weeks or months before being rented or sold, the number of total listings is much higher than the number of unique listings in both rental and sales markets.

In the hedonic regressions, to provide a true comparison between markets, we included only those characteristics that had relatively few missing entries. Also, for each city, we limited ourselves to those characteristics that were available for both the sales as well as the rental market. In particular, we included the following characteristics:

- inside space: the number of bedrooms (1, 2, 3, 4+) and bathrooms (1, 2, 3, 4+) in London and square meters in Vienna.
- outside space (garden or terrace) for Vienna
- historical building yes/no for Vienna
- distance to the city centre (which we constructed from individual GIS locations)
- type of property (flat/maisonette for London and flat/penthouse for Vienna)
- region dummies

We construct our own region dummies. To achieve this, we divide the city area into nine separate areas by forming two concentric circles around the city center (defined as Trafalgar's Square in London and Stephansplatz in Vienna) and dividing them by the North/South East/West axis. Figure 1a and Figure 1b illustrate the resulting regional dummies. Note, that the innermost circle is considered as one area (not divided into four) and provides the base area in the hedonic regressions (see Tables 5 and 6).

As London and Vienna are of similar size with respect to the area enclosed within their city limits, we use the same calculation for both cities, however London's property density in the outer regions is much higher than that of Vienna.

In addition to these characteristics, the frequent scraping of data allows us to keep track of each individual apartment throughout its time on the market. Individual identifiers like url-addresses allow us to establish a personal (price) history of each property with respect

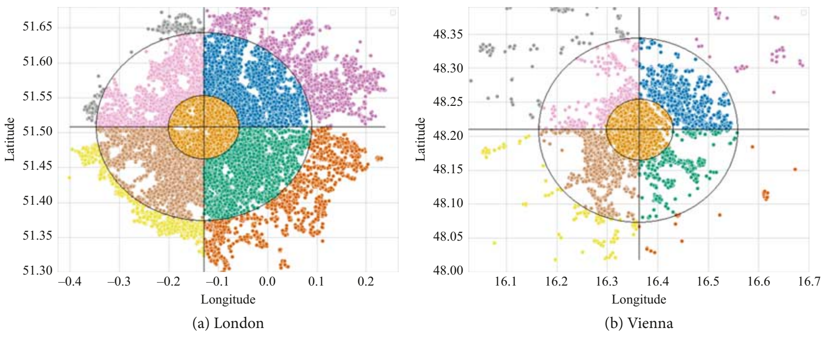


Fig. 1. Construction of regional dummies for London and Vienna.

to time on the market and/or price movements. This is definitely a benefit of (frequently scraped) list data, and we will use the pricing history to derive our sentiment indicator in Subsection 4.6.

We have access to a sample of transaction data for Vienna via the Austrian data provider ZT datenforum (2021). These data cover a 12 month period in 2019 during which all contracts are included that contain information on the interior space of the property (which is available for approximately 50% of transactions). The transaction data for London come from the UK Land Registry (2021a) and contain almost 100% of transacted properties. Table 2 shows a summary of the transaction data sets for the year 2019.

Figures 2 and 3 illustrate the relative geographic coverage of the list and transaction data sets. The spatial coverage of the list-price data overlaps with the transaction data, with the exception of a small part on the north-west side of London.

Figures 4 and 5 illustrate the distribution of prices before and after our cleaning process for list-price data compared with the corresponding transaction price data sets in 2019. Price negotiations between buyers and sellers, as well as inflated prices for properties that

Table 2. Summary statistics of transaction data sets, year 2019.

City	Observations	Median price	CV
London	56,601	415,000 Pounds	0.48
Vienna	5,755	258,864 EUR	0.59

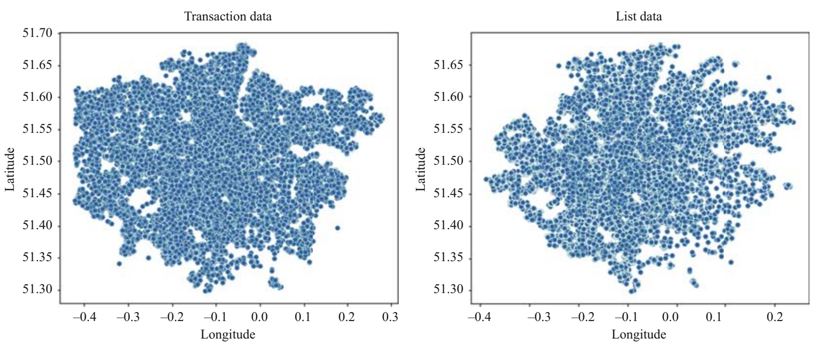


Fig. 2. Geographic coverage of list and transaction data London, 2019.

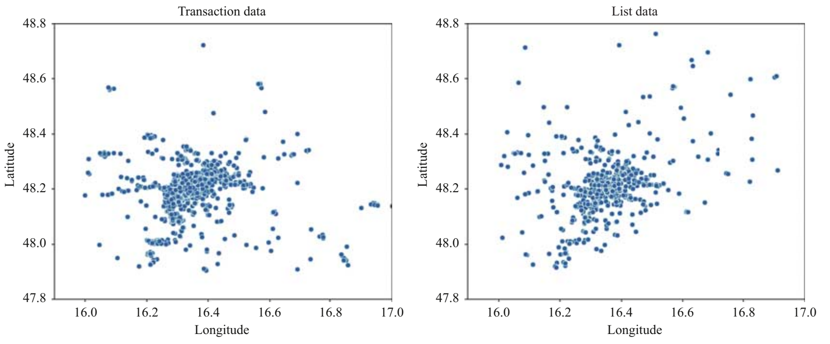


Fig. 3. Geographic coverage of list and transaction data Vienna, 2019.

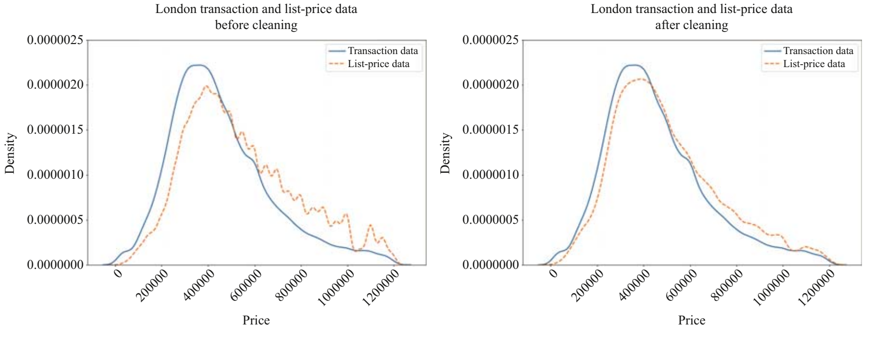


Fig. 4. London price data.

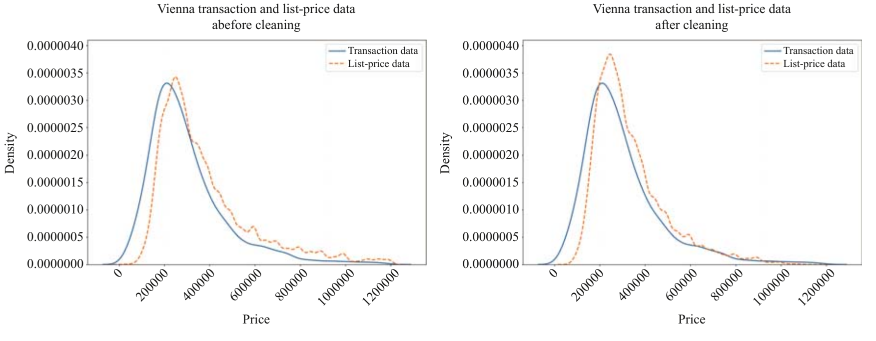


Fig. 5. Vienna price data.

are not being sold, will make list prices lie above transactions prices. Our cleaning process (especially selecting only the last price before a property leaves the platform) shifts the distribution of the list-price data towards that of transaction prices; this concerns in particular the highly-priced right tail of the listing data sets. With this process we find a closer correspondence between list and transaction data than others in the literature (see e.g., Shimizu 2016; Kolbe et al. 2021). Interestingly, on the left-hand side of the price distributions transaction prices lie above listing prices. This is probably due to non-market transactions between family members, friends, or neighbours. The properties associated with such non-market transactions will show up in the transaction data set but be missing from the listing platforms (as there is no reason to advertise them).

4. Methods and Results

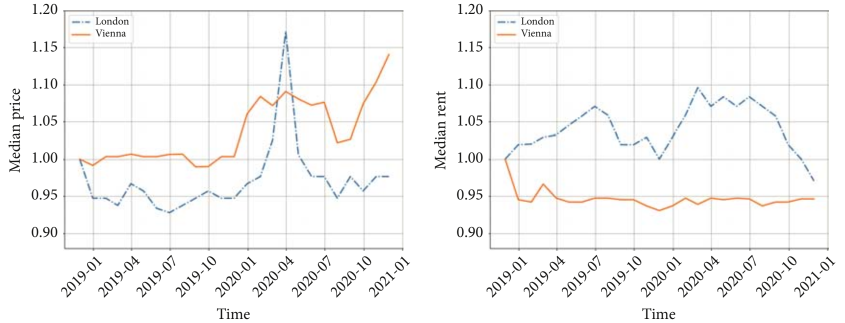
4.1. Indicator 1: List-price and List-rent Indicators

A house price index (HPI) measures changes in property prices across a designated market and hedonic indices are the standard tool for quality adjusting house price indices (see e.g., Eurostat 2013; IMF 2020). For some markets, sales or rent-determining characteristics are not available, and median price indices are used instead. We illustrate the difference between median and hedonic indices by looking at both types in this section.

Figure 6a depicts the percentage median price change per month for sales in London and Vienna. Figure 6b shows the median price index for rental properties in those cities. The most striking feature of Figure 6a is the dramatic increase in London median prices in April 2020 (by 14%) and the subsequent dramatic drop in the subsequent months. The period of the sudden increase in the median index coincides with the first physical COVID-19 shutdown period in London, which occurred from mid-March to mid-April 2020.

However, when inspecting the data closer, we find that this shutdown price spike is in a large part due to changes in market composition rather than actual price increases. During this first UK COVID-19 shutdown period properties still left the platform (presumably because they were sold) but hardly any new properties were added to the platform. Furthermore, the composition of properties on the London market changed during this period: Up-market properties were more likely to remain (and become) listed during the shutdown period than lower-priced properties. This resulted in the dramatic increase in the median apartment price in April, which is depicted in Figure 6a. Interestingly, the composition of listed apartments went almost back to "normal" in May and June 2020. This point illustrates how listings data could be utilised in parallel with transaction data as an early warning indicator of stock mix issues.

Hedonic indices are better able to control for compositional differences in the data than median price indices. We use the time-dummy method to construct hedonic list price index using data from November 2018 to December 2020 (for references on the time-dummy



(a) Sale-market (b) Rental-market

Fig. 6. Monthly median price development.

method, see e.g., Diewert et al. 2009; Eurostat 2013; Hill 2013). For each city we estimate separate semi-log hedonic sales as well as rental models as follows:

$$\ln(P^S) = X^S\beta^S + \gamma^S D^S + u^S \tag{1}$$

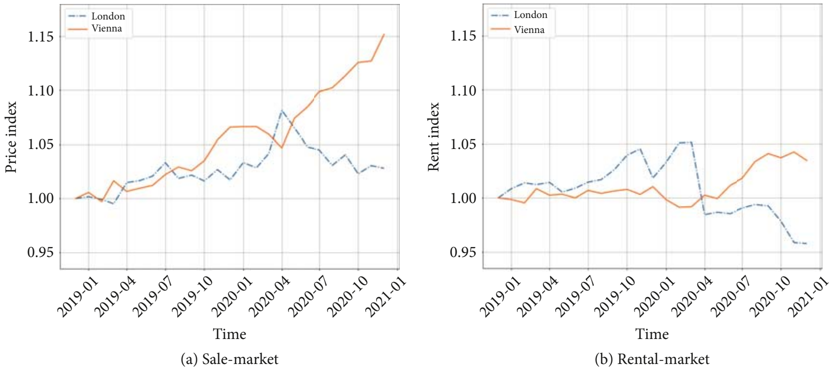
and

$$\ln(P^R) = X^R\beta^R + \gamma^R D^R + u^R. \tag{2}$$

The superscripts S and R in Equations 1 and 2 denote the sales and rental market, respectively. The matrices X^S and X^R contain all information on the characteristics for each property, the β terms indicate the corresponding estimated shadow prices, the γ are the estimated coefficients on the time dummy, and u^S and u^R are the error terms. $\ln(P^S)$ and $\ln(P^R)$ denote logged sales and rental list prices. The D matrices have the dimension H times T, where H is the total number of observations, and T corresponds to the number of time periods. Hence, each entry in our data set is linked to one row in the D matrices, where the elements for each row are filled with zeros, except for the column with the dates of the observed listings. The X matrices contain information on inside space, outside space (yes/no), new building (yes/no), and location (see Section 2 on the construction of region dummies).

The resulting list-price indices are derived by applying the exponent to the estimated time dummy coefficients γ and are illustrated in Figure 7. Both London and Vienna see rising prices in the sales market in the lead-up to the COVID-19 crisis, but price developments start to diverge in April 2020: the Vienna market continues its previous trend of strong price increases, while the London market softens.

The regression results for the London and Vienna markets are shown in Appendix (Section 6). The descriptive variables are highly significant for sales as well as rental markets and have the expected signs. Overall, the goodness of fit is better for the Vienna market (with an adjusted R^2s of 0.80 for the rental market and 0.77 for the sales market) than for London (with adjusted R^2 of 0.58 and 0.65 respectively). This is due to the better description of properties in Vienna, where we have information about the exact size of the property in square meters, whether it has an outside space, is located in a historic building, is in need of renovation, or can be considered a “penthouse”. In contrast, in London, we



(a) Sale-market (b) Rental-market

Fig. 7. Monthly list-price and list-rent indices for London and Vienna.

have much less detailed information on the individual properties, and size is measured in the form of bedroom and bathroom dummies rather than exact measurements. Distance to the city center has a negative impact in both cities, and the location dummies have the expected sign and are almost all highly significant.

Figure 7a illustrates the price movements for London and Vienna properties during our study period. Prices for apartments in Vienna increased during the entire period of observation (with the exception of a small dip in April during the first physical shut-down period). Compared to the median price index (see Figure 6a) the hedonic price index shows a much smoother price development. Comparing Figures 6a and 7a for London, the large price spike for April 2020 in the median price index is reduced to less than half of its previous size with the hedonic index. This example illustrates how important it is to control for quality in the construction of price indices, especially when working with online data that are not “tested by the market”.

In terms of market direction since the beginning of the COVID-19 crisis in March 2020 (when the first lock-down periods were imposed in both the UK and Austria), the picture is clear in Vienna, where we find a rapidly increasing price index from April 2020 to December 2020. The interpretation of the London sales market is more difficult. After an initial price rise followed by a price fall during the first part of 2020, prices stabilized in the second half of the year. This could be (at least partially) due to a policy intervention by the UK government: it announced a sudden change in property taxation in July 2020, under which properties costing up to 500,000 pounds became excluded from property taxation between July 2020 and March 2021. Roughly half of the London transactions fall within this limit. The rental development throughout 2020 is very different in the two cities: rental prices in London decreased from April 2020 to December 2020, while Vienna rents increased during the same period.

For the rental markets, the differences between median and hedonic price indices are less dramatic, but still present (compare Figures 7b and 6b). Figure 7b illustrates the quality-adjusted rent-price indices for London and Vienna. Similar to the sales markets, the price developments for London and Vienna also diverge in the rental markets. While Figure 7b shows falling rents for London since the beginning of the COVID-19 crisis in February 2020, it indicates rising rents in Vienna for the same time period. Interestingly, this rent increase depicted by the hedonic rent index for Vienna is not visible in the median rent index; The median index cannot pick up that over the two-year period rental properties in Vienna became increasingly smaller.

We now turn to the question of whether list price indices are good predictors for transaction-based price indices. Statistics Austria computes transaction price indices for apartments in Austria but does not publish a separate price index for Vienna. The Austrian National Bank provides a separate price index for Vienna, but it is based mostly on list-price data. Thus, for Vienna, we cannot compare list and transaction price indices directly but will illustrate the close co-movement of our list-price index with the list-price index of the Austrian National Bank (OeNB, 2020). Since the OeNB index is provided on a quarterly basis, we also include a quarterly index for Vienna (see Figure 8).

For London we can compare list-price and transaction-price indices: In Figure 9 we plot the official ONS transaction price index for London together with our lagged hedonic *list-price index*. The official HPI for London (https://landregistry.data.gov.uk/app/ukhpi) is

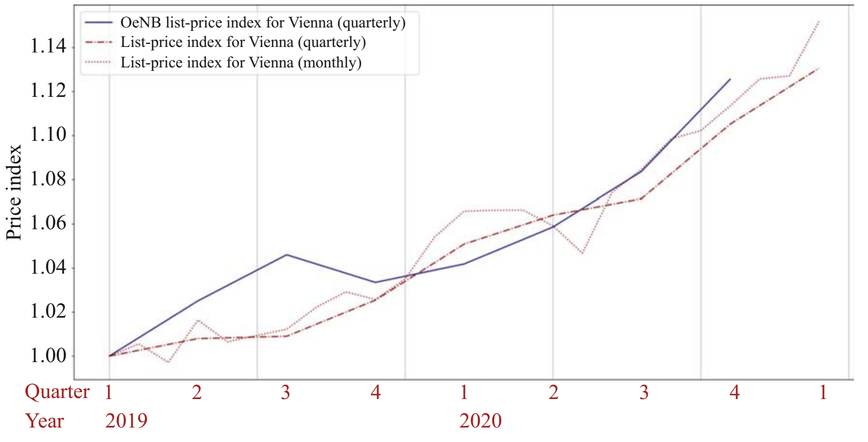


Fig. 8. Comparison of list price indices for Vienna.
Source: Own calculation; Austrian National Bank (OeNB).

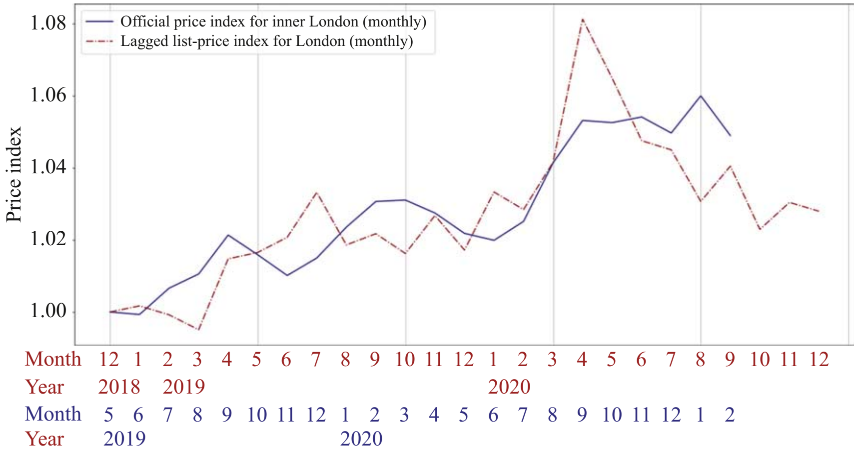


Fig. 9. Official house price index – London.
Note: There are two labels on the x-axis. The first row (red) corresponds to the list-price index that we lagged by five months. The blue labels correspond to the results for Inner London computed by the UK Land Registry. Own calculation; Source: UK Land Registry.

available at quarterly frequency for Inner London (this corresponds to our boundaries) as well as for Greater London, and the City of London. Last retrieval: 20.05.2020 together with our lagged hedonic list-price index (see UK Land Registry 2021b). As we do not have long enough time-series data to estimate the lag between prices appearing on the internet platform and transactions econometrically, we chose a five-month lag based on visual inspection. Our five-month lag fits well within the range of lags discussed in the literature. Longer lags of about seven months have been found for Tokyo (Shimizu et al. 2016), while Brounen and Kok (2011) find a shorter lag of four, and a half months for the Dutch housing market. An estimate for the UK market presented in Eurostat (2013, 103) indicates a five, and a half months lag.

The overall movements of the two indices coincide, even though the ONS index contains price movements of apartments and houses while our index concentrates on apartments only. The transaction price index is a bit more "jumpy" compared to the *list-price index*. This is probably due to the greater price stickiness of list prices compared to transaction prices – a phenomenon that has been discussed in the literature by Haurin et al. (2013). Note, that we cannot draw any conclusions about relative price levels from these figures as we normalized the listing and transaction price indices to 1 at the beginning of the period. Actual transaction data would be needed for a comparison of list- and transaction prices in levels.

4.2. Indicator 2: List Price-Rent Ratios

We compute *list price-rent ratios* using the "double-prediction" (also called "double-imputation") approach a hedonic method that is well established in the price index literature (see e.g., Hill and Syed 2016).

The method works as follows: First, we estimate two hedonic models, one for the sales market and one for the rental market as done in the previous section. Second, we use these models to impute a rental price for each sales object as well as a sales price for each rental object. Third, we compute "own-imputation": this involves imputing the price of each sales property via the hedonic sales model as well as the rental price for each rental property via the hedonic rental model. Fourth, we compute price-rent ratios for each property. Finally, these individual (imputed) price-rent ratios are aggregated to the market price-rent ratio.

Thus, for apartment j with characteristics c, which is listed for sale, we estimate:

$$\frac{\hat{P}^S(x_j^S)}{\hat{P}^R(x_j^S)} = \frac{exp\left(\left[\sum_{c=1}^{C}\hat{\beta}_c^S x_{j,c}^S\right] + \gamma^S d_j^S + \varepsilon_j^S\right)}{exp\left(\left[\sum_{c=1}^{C}\hat{\beta}_c^R x_{j,c}^S\right] + \gamma^R d_j^S + \varepsilon_j^R\right)}, \tag{3}$$

where $\ln \hat{P}^S(x_j^S)$ and $\ln \hat{P}^R(x_j^S)$ denote the predicted log offer price and log rental price for property j in the sales data set, and $\hat{\beta}^S$ and $\hat{\beta}^R$ are the estimated characteristic shadow prices obtained from Equations (1) and (2). γ^S and γ^R are the derived time coefficients, and ε^S and ε^R are the error terms for the sales and rental market respectively.

Similarly, for apartment h with characteristics c, which is offered for rent, we estimate the following:

$$\frac{\hat{P}^S(x_h^R)}{\hat{P}^R(x_h^R)} = \frac{exp\left(\left[\sum_{c=1}^{C}\hat{\beta}_c^S x_{h,c}^R\right] + \gamma^S d_h^R + \varepsilon_h^S\right)}{exp\left(\left[\sum_{c=1}^{C}\hat{\beta}_c^S x_{h,c}^R\right] + \gamma^R d_h^R + \varepsilon_h^R\right)}, \tag{4}$$

To derive at the overall price-rent ratio, we take the geometric mean as follows:

$$Med[\hat{P}^S/\hat{P}^R] = \sqrt{Med[\hat{P}^S/\hat{P}^R]_{sales} \times Med[\hat{P}^S/\hat{P}^R]_{rental}}, \tag{5}$$

where $Med[\hat{P}^S/\hat{P}^R]_{sales}$ and $Med[\hat{P}^S/\hat{P}^R]_{rental}$ represents the median of the price to rent ratio derived from Equations (3) and (4) respectively.

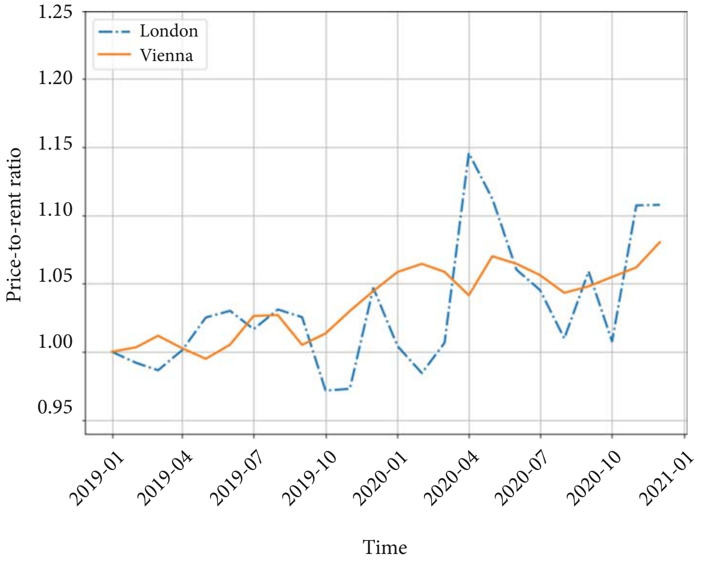


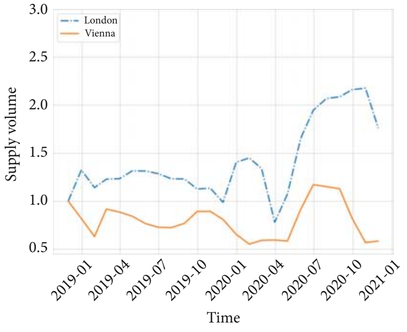
Fig. 10. Monthly list price-rent ratios.

The resulting *list price-rent* ratios for both cities are shown in Figure 10. In the Vienna market the *list price-rent ratio* increased in a smooth pattern over the entire two year period. Results for London again show an "April-peak" and a much more volatile up and down movement of the price-rent ratio – especially for 2020. As a result, it is unclear how the price-rent ratio should be interpreted for the London market.

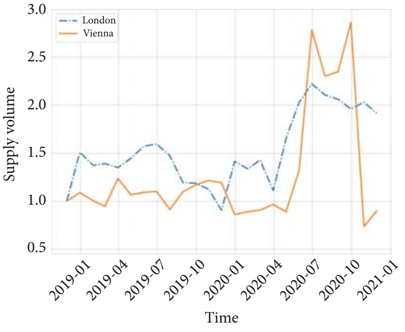
4.3. Indicator 3: List Volume

Market volume, the number of properties transacted during a particular time period, is an important measure of market tendency. This is the case since housing markets tend to be more liquid during market up-turns than during down-turns. The higher liquidity during market up-turns implies that more transactions are taking place and that houses are selling more quickly. During market down-turns there are typically fewer sales and houses tend to remain on the market for longer. Different theories for this stylized fact of the housing market exist. Stein (1995) links the phenomenon to credit constraints, while Genesove and Mayer (2001) illustrate how loss aversion could be the reason sellers refuse to sell during market down-turns (at prices below the original purchase price).

List volume is a related, but different, aspect of market liquidity and measures the number of unique properties on the market at a particular point in time. To construct this indicator we compute the number of unique listings per month for both the sales as well as the rental market. Listings that are available for several months will enter the volume count once each month (duplicates within one month are ignored). To compare "like with like" we include the same number of days for each month by shortening the longer months. Furthermore, we normalize the series to the number of unique properties listed in December 2018 in each market. It is important to note here that we can only measure short-term housing supply with this indicator.



(a) Sale-market



(b) Rental-market

Fig. 11. Total number of listings per month.

Figure 11a illustrates the *list volume* in the sales market for both cities. As already discussed above, there is a sharp decrease in listed properties in the London market during March/April 2020, followed by a strong increase in subsequent months. On the online platform the number of unique properties for sale in London dropped from over 10,000 per month in February 2020 to less than 6,000 in April, only to rise to over 14,000 by the end of July 2020. No such drop in *list volume* is visible for the Vienna market but *list volume* is more volatile in 2020 than in 2019.

List volume for rental properties is shown in Figure 11b. The number of listed rental apartments in London doubled from April to June 2020 and stayed at this high level for the rest of the study period. The sudden increase in available rental properties after April 2020 was even stronger in Vienna. Suggestions in the media speculate about an "Airbnb effect" that led to the sudden jump in available rental properties (Fonds Online 2020). This theory is plausible for Vienna as these new rentals are located to a large degree in the touristic inner districts of Vienna. The number of listed properties stayed high until a sudden drop in November 2020.

4.4. Indicator 4: List Time-on-market

A large benefit of daily list price data is that *time-on-market* can be measured at the level of individual properties. The aggregate *time-on-market* indicator here is defined as the mean number of days properties have been advertised on the platform. However, as not all properties that are taken off the market are also sold, the *list time-on-market* indicator will generate a lower bound for actual *time-on-market*. Figure 12 below illustrates this concept. The importance of relisted properties in defining and assessing *time-on-market* is investigated in Benefield and Hardin (2015) who also note that earlier estimates of *time-on-market* based on printed adverts in newspapers are probably even more biased downwards than calculations based on listing platforms.

Listing starts have to be approximated for the Vienna data set as first appearance on the website is not one of the available characteristics. This means that *time-on-market* will steadily increase in the initial months before settling into its true pattern. We highlighted this initial period with gray in Figure 13a and 13b. For the London market, we are able to derive the *time-on-market* measure based on the provided information of each first listing

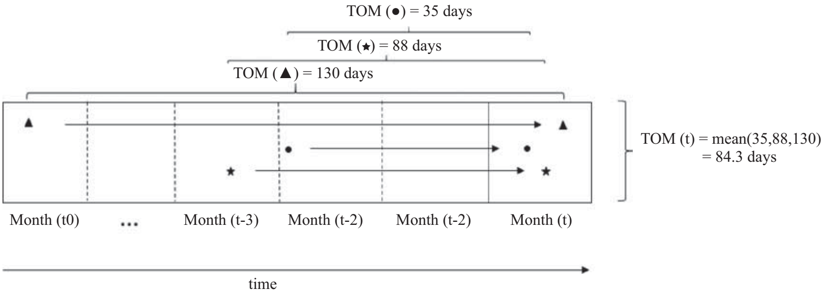


Fig. 12. Illustration of time-on-market calculation.
Note: TOM refers to *time-on-market*.

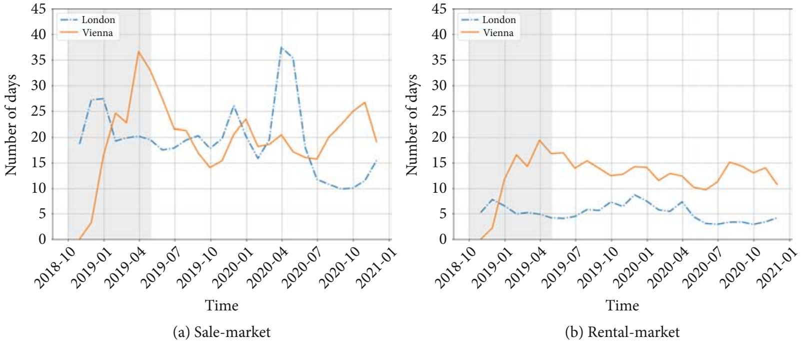


Fig. 13. Mean time-on-market.

entry from the website provider. Hence, the *time-on-market* can be calculated from the beginning (i.e., without lead time).

Time-on-market for rental apartments are also computed in the same way and are shown in Figure 13b. As expected, rental apartments have a shorter *time-on-market* than apartments for sale.

After the onset of the COVID-19 crisis, the *time-on-market* for apartments for sale in London exhibits a sharp up and then down-turn: while it sharply increased from February to May, there was a sharp trend-reversal in the summer of 2020. This drop in *time-on-market* coincided with the suddenly imposed stamp duty holiday for properties under 500,000 pounds for England and Northern Ireland, which started in July 2020 and lasted until March 2021. Roughly half of the apartments in London lie below this 500,000 pound threshold (see Table 1) and are thus exempt from stamp duty during this period (before July 2020 the cut-off level for stamp duty exemption was 125,000 pounds). In contrast to the decreasing *time-on-market* for sales properties in London, we see a smooth and stable *time-on-market* in the London rental market for the entire period of observation.

In Vienna, *time-on-market* for both the sales and rental market was relatively stable throughout 2020. The Vienna sales market even saw a decline in *time-on-market* for the first six months after the onset of the COVID-19 crisis. A similar slight downward trend in *time-on-market* is also noticeable for the Vienna rental market until mid-2020.

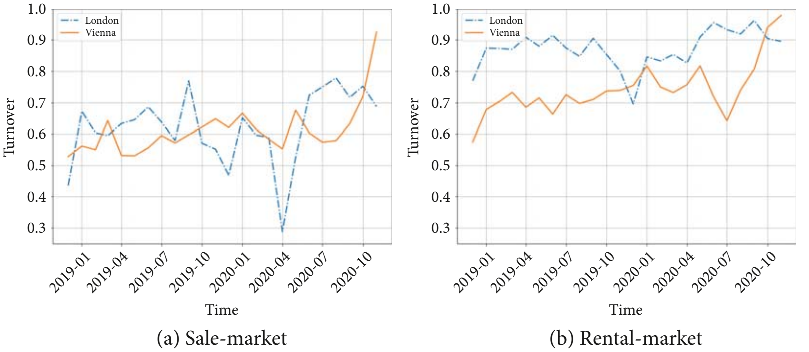
4.5. *Indicator 5: List Turnover Rate*

The *list turnover rate* is constructed using a similar methodology to the classic market turnover rate, which is generally defined as the number of sales (transactions) in a period divided by the number of properties available at the beginning of the period. The market turnover rate is a frequently discussed measure of market liquidity which (like market volume) tends to be positively related to price (see Genesove and Mayer 2001; De Wit et al. 2013; Van Dijk and Francke 2018). For theoretical explanations of this positive relationship see discussions on (nominal) loss aversion (e.g. Genesove and Mayer 2001), and equity constraints (e.g. Stein 1995; Genesove and Mayer 2001).

Here, the list *turnover rate* is constructed by dividing the number of properties that leave the platform (i.e., are available at some time during the month but disappear before the end of the month) by the total number of (unique) properties available on the platform during that month. The results for the *turnover rate* are shown in Figures 14a and 14b.

The *turnover rate* lies around 0.6 for both the London and Vienna sales markets until the start of the COVID-19 crisis in March 2020. A turn-over rate of 0.6 implies that 60% of properties that are available at some time during the month are no longer available at the end of the month. Interestingly, the turnover rates for the two cities then diverged dramatically in the spring of 2020. Turnover dropped sharply in London between February and April 2020 and only 27% of the properties on the platform in April 2020 also left the platform during that month. However, this turnover rate rebounded sharply and two months later almost 80% of properties left the platform within one month. The higher *list turnover rate* in the London sales market in the second half of 2020 is probably related to the stamp duty holiday, which was put in place in July (and runs until March 2021). The *list turnover rate* in the sales market in Vienna also becomes more volatile after the start of the COVID-19 crisis, but at a higher level than in London. Overall the *list turnover rate* in Vienna increased after the start of the COVID-19 crisis, another indication of a heating property market.

In line with the lower *time-on-market*, *list turnover rates* in rental markets are higher than in sales markets and for both markets there seems to be an upward trend in 2020. The fall in turnover in the Vienna rental market during the summer of 2020 could be influenced by Airbnb landlords shifting short-term rental properties to the longer-term rental market.



(a) Sale-market

(b) Rental-market

Fig. 14. Turnover Rate

4.6. *Indicator 6: Sentiment Indicator Changes*

A great advantage of daily scraped list price data is that it has a history: Each property can be identified via its unique url-address which allows us to observe not only its current offer price but also its entire pricing history. Here we use this pricing history (in particular the direction and frequency of price changes) to construct an indicator that informs about sellers' market sentiments. Properties tend to have more negative than positive price changes during their stay in the market as sellers often "try out" higher prices first before decreasing them later. However, disturbances to this "normal" pattern (i.e., more positive or negative price changes than usual) can be used as an indicator for shifting market sentiment. As markets differ with respect to what is considered a normal pricing strategy, an indicator based on such price changes will be specific to each market.

We derive our market sentiment indicator as follows:

1. At the end of each month, we establish the direction of the last price change for each apartment,
2. We construct a variable that reports "+1" if the last change was positive, "-1" if the last change was negative, and "0" if the price remained unchanged since the apartment entered the market. Figure 15 illustrates the process, and
3. We then take the sum of this variable over all properties and divide it by the total number of apartments for sale during the month.

Results for this variable are shown in Figures 16a and 16b. The most striking finding is that overall pricing strategies differ dramatically between London and Vienna. While London sellers tend to lower prices during a property's stay on the platform, sellers in the Vienna market hardly ever change prices – and if they do, they are almost as likely to increase as to decrease them. To further illustrate this difference we include Table 3, which

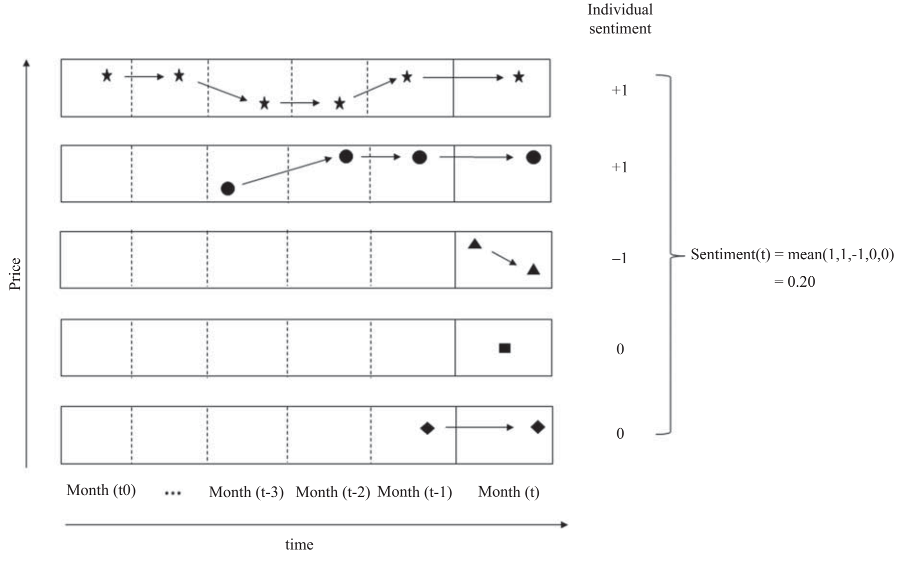
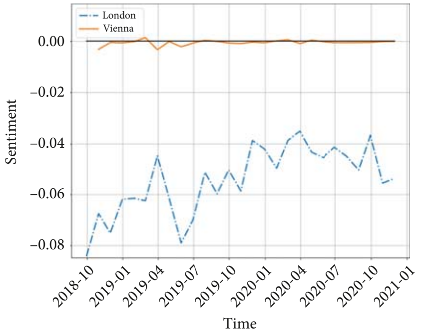
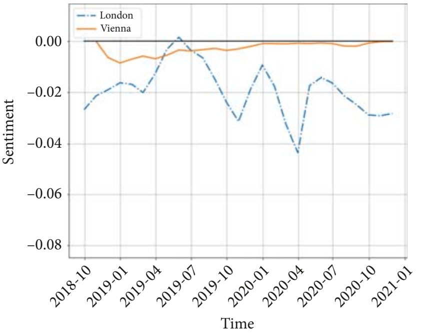


Fig. 15. Illustration of sentiment calculation.



(a) Sales-market



(b) Rental-market

Fig. 16. Market sentiment per month.

shows the total number as well as the composition of monthly price changes in all submarkets. It clearly illustrates that more changes are made by London property holders than by their Vienna counterparts.

We turn now to the history of the sentiment indicators over the two-year period. Even though the sentiment indicators for the sales markets in London and Vienna differ in levels, Figure 16a shows that sellers in both markets were less likely to lower prices in 2020 than 2019. This result fits well with the positive (even though, in the case of London, somewhat volatile) price developments shown by the hedonic price indices in Subsection 4.1.

Figure 16b illustrates the sentiment indicators for the rental markets. Similarly to the sales market, London landlords are more likely to lower prices than their counterparts in Vienna. It is notable that the London landlords' sentiment was lower in 2020 than in 2019. Thus, the sentiment of London landlords and London sellers in 2020 did not match according to this indicator. In contrast, Viennese landlords were less likely to lower rents for advertised apartments in 2020 than in 2019, and the sentiment in the rental market shows a prolonged upward trend since the beginning of 2019.

4.7. *Summary of Indicator Results Throughout 2020*

Table 4 provides a summary overview of the results discussed in Subsections 4.1 to 4.6. Here we focus on the broad direction of the indicators from the start of the COVID-19 crisis until December 2020. In Table 4, we take the average indicator values for January, February, and March 2020 as the pre-crisis reference values to compare the December 2020 values with. To test the robustness of these results we also investigated two alternative definitions of the pre-crisis period; the direction of the arrows in Table 4 remain unchanged (see C for more details).

The overall liquidity measure in Table 4 is based on the combined movement of the three liquidity measures – *volume*, *time-on-market*, and *list turnover rate*. Each month we measure the tendency for each of these three liquidity measures; the overall monthly liquidity measure is positive if two of the three liquidity measures indicate a positive liquidity tendency. Table 14 in the Appendix illustrates the movements for each of these liquidity measures as well as their combined direction each month.

Table 3. Number of positive (+) and negative (−) price changes per month.

Time	London						Vienna					
	Sale		Total change	Rent		Total change	Sale		Total change	Rent		Total change
	(+)	(−)		(+)	(−)		(+)	(−)		(+)	(−)	
10-2018	793	6,746	10.69%	1,214	2,984	6.38%	–	–		–	–	
11-2018	1,609	10,177	9.30%	2,440	5,584	5.46%	0	7	0.31%	0	0	0.00%
12-2018	865	6,766	9.73%	1,819	4,044	5.03%	18	23	0.31%	5	56	0.73%
01-2019	1,261	6,766	9.04%	2,002	3,923	5.03%	74	85	0.85%	79	199	1.97%
02-2019	1,215	5,062	10.04%	1,531	3,051	5.09%	152	161	1.38%	110	245	1.81%
03-2019	830	4,549	9.05%	1,342	2,827	5.65%	163	126	0.91%	46	196	0.95%
04-2019	953	3,889	7.43%	1,643	2,693	5.35%	82	205	0.76%	63	291	1.06%
05-2019	957	5,359	8.85%	1,968	2,230	4.86%	202	213	0.94%	49	263	0.80%
06-2019	757	5,639	10.37%	2,488	2,349	5.64%	72	176	0.52%	45	193	0.54%
07-2019	888	5,645	9.63%	1,852	2,157	4.50%	151	189	0.64%	55	245	0.59%
08-2019	941	4,960	7.54%	2,577	3,229	5.56%	130	106	0.40%	60	247	0.54%
09-2019	1,136	4,718	9.77%	2,197	3,349	7.13%	128	123	0.40%	53	228	0.45%
10-2019	1,152	5,412	7.77%	2,271	4,414	7.44%	110	157	0.37%	63	315	0.53%
11-2019	1,080	5,901	8.50%	1,995	4,620	7.96%	110	181	0.37%	57	291	0.44%
12-2019	1,424	4,908	7.05%	2,738	4,977	6.43%	135	160	0.32%	77	259	0.37%
01-2020	1,693	6,236	7.37%	3,526	4,806	6.03%	151	205	0.36%	91	191	0.24%
02-2020	1,979	8,202	8.14%	3,008	5,299	6.41%	120	90	0.16%	57	181	0.10%
03-2020	2,999	10,133	7.13%	3,300	8,637	7.26%	277	197	0.36%	93	279	0.22%
04-2020	703	2,191	6.87%	2,466	8,316	8.02%	94	218	0.25%	40	144	0.12%
05-2020	573	2.267	7.28%	1,349	2,692	5.23%	160	155	0.22%	69	238	0.19%
06-2020	794	3,684	7.06%	1,258	2,461	4.37%	55	99	0.11%	24	12	0.09%
07-2020	759	3,869	6.17%	1,194	2,700	4.22%	78	148	0.18%	65	178	0.18%
08-2020	688	3,785	6.50%	1,222	3,101	4.94%	110	179	0.22%	102	367	0.33%
09-2020	720	4,462	6.98%	1,057	1,158	5.00%	144	212	0.25%	64	364	0.27%
10-2020	851	3,901	5.75%	1,014	3,732	5.06%	33	94	0.09%	18	128	0.09%
11-2020	553	3,539	7.62%	815	2,738	5.41%	20	35	0.04%	1	19	0.01%
12-2020	168	1,136	7.25%	517	1,752	5.21%	5	6	0.01%	1	6	0.00%

Table 4. Overall market development for London and Vienna for 2020.

		Price development	Price-to-rent ratio	Liquidity measure	Market sentiment
London	Sales	→			→
	Rental	↘	↗	↗	
Vienna	Sales			→	→
	Rental	↗	↗	↗	↗

5. Conclusion

In this article we used data from online real-estate platforms to investigate the initial impact of the COVID-19 crisis on the London and Vienna real-estate markets in 2020. Based on our data set, which we scraped from leading housing platforms throughout 2019 and 2020, we constructed early market indicators for *list-price*, *list-rent*, *price-rent ratio*, *market volume*, *time-on-market*, *market turnover*, and *market sentiment*. The last of these, market sentiment, is new to the literature. Even though the other indicators have been discussed before, we add to the literature by presenting them all within one article and by illustrating how to construct them using online price data. The examination of trends in the real estate market is particularly important during times of economic upheaval, and these online indicators allow us to do so before transaction data become available.

We took great care in preparing our online data set and constructing the indicators. In particular, our method of selecting only the last price before a property leaves the platform shifts the distribution of the list-price data towards that of transaction prices. This is particularly noticeable at the right tail of the listing data set. We believe that this closeness in price distributions illustrates that the list price-indices developed here can help predict future market developments. For the period of (lagged) overlap between list and transaction data, the official transaction price indices follow our list-price indices with a five-month lag. For Vienna, the *list-price index* anticipated the strong positive price development and, in London, the stable price level throughout 2020.

For the construction of the *list price-rent ratios*, we used the hedonic “double-prediction” approach in order to limit the influence of unobserved differences between the online sales and online rental properties. The *list price-rent ratios* increase throughout 2020 in both markets, driven by the price increases in both markets, and – in London – also by falling rents.

To measure housing volume, we looked at three different measures: *list volume*, *list time-on-market*, and *list turnover rate*. Each of these measures captures a slightly different aspect of market volume. When all three agree in their market assessment, they send a clear signal about current market tendencies. For the 2020 COVID-months this signal indicated increased liquidity in both markets.

The *sentiment indicator*, which we constructed from information on the direction and frequency of online price changes, is new to the literature. It aims to capture current sellers’ market perception. According to this indicator, the sentiment of London sellers was relatively stable throughout 2020 and higher than during 2019. London landlords on the other hand were more positive in 2019 than in 2020. In addition to being more negative than the year before, the sentiment indicator for the 2020 London rental market is also very volatile, exhibiting a big drop in measured sentiment during the first COVID-19 lock-down a sharp recovery in the summer, and a subsequent slow decline until the end of the year. In Vienna, both sales and rental market sentiments were buoyant during 2020.

Even with six different indicators it is hard to judge the overall state of the London sales market during much 2020. We see a sudden rise and then fall in sales prices after the first COVID-19 lockdown in March/April 2020. Then prices stabilize in the summer. A similar volatile pattern is present in the *list price-rent ratio*, the sentiment indicator, as well as the three market liquidity indicators. In contrast to this behaviour in the sales market, the direction of the rental market is easy to interpret: it became negative as soon as the crisis hit.

In contrast to the somewhat contradictory signals that we find for the London market, the Vienna housing market shows a strong upward trend over the entire two year period we investigated, and this trend strengthened after the start of the COVID-19 crisis in March 2020. In Vienna, both price and rent indicators continued their previous upwards trends throughout 2020, and time-on-market fell dramatically after the start of the pandemic. Underlining this market tendency is a price-rent ratio that is also moving upwards. Liquidity indicators have been more volatile in Vienna after the start of the COVID-19 crisis, but at a higher level than in London. In terms of historical perspective, the Vienna property market is no stranger to rallying in the wake of a global crisis: It also did so during the Great Financial Crisis between 2007 and 2010 when Vienna's prices rallied while other real estate markets fell (OeNB 2019).

We believe we illustrated with this article that list-data indicators can provide useful early warning signals for the development of the housing market. While we advise against interpreting list-data indicators as perfect substitutes for their transaction-based counterparts – they primarily illustrate the behaviour of the supply side of the market they are a rich data source that can be used to get market insights. In particular, constructing a number of these list-data indices and interpreting their joint behaviour can provide useful and timely information on future market movements.

6. Appendix

6.1. Model Statistics

London

Table 5. Model summary for London list-price index.

	Goodness of fit: London	
	Rental model	Sale model
Constant	7.655***	13.449***
	(0.002)	(0.006)
2019_1	0.009555	0.002
	(0.003)	(0.006)
2019_2	0.014***	−0.001
	(0.003)	(0.007)
2019_3	0.012***	−0.005
	(0.003)	(0.007)

Table 5. Continued

	Goodness of fit: London	
	Rental model	Sale model
2019_4	0.014***	0.016**
	(0.003)	(0.007)
2019_5	0.005*	0.016**
	(0.003)	(0.006)
2019_6	0.009***	0.021***
	(0.003)	(0.006)
2019_7	0.014***	0.033***
	(0.003)	(0.006)
2019_8	0.017***	0.018***
	(0.003)	(0.007)
2019_9	0.026***	0.021***
	(0.003)	(0.006)
2019_10	0.039***	0.016**
	(0.003)	(0.007)
2019_11	0.044***	0.026***
	(0.003)	(0.007)
2019_12	0.018***	0.017**
	(0.003)	(0.007)
2029_1	0.032***	0.033***
	(0.003)	(0.006)
2020_2	0.050***	0.028***
	(0.003)	(0.006)
2020_3	0.050***	0.041***
	(0.003)	(0.007)
2020_4	− 0.016***	0.078***
	(0.003)	(0.009)
2020_5	− 0.013***	0.063***
	(0.003)	(0.007)
2020_6	− 0.015***	0.046***
	(0.003)	(0.006)
2020_7	− 0.010***	0.044***
	(0.003)	(0.006)
2020_8	− 0.006**	0.030***
	(0.003)	(0.006)
2020_9	− 0.007***	0.040***
	(0.003)	(0.006)
2020_10	− 0.022***	0.023***
	(0.003)	(0.006)
2020_11	− 0.042***	0.030***
	(0.003)	(0.006)
2020_12	− 0.043***	0.028***
	(0.003)	(0.006)
bathroom_2.0	0.251***	0.327***
	(0.001)	(0.002)
bathroom_3.0	0.453***	0.623***
	(0.003)	(0.006)
bathroom_4.0	0.414***	0.682***
	(0.010)	(0.029)
bedroom_2.0	0.180***	0.173***

Table 5. Continued

	Goodness of fit: London	
	Rental model	Sale model
	(0.001)	(0.002)
bedroom_3.0	0.327***	0.308***
	(0.001)	(0.003)
bedroom_4.0	0.371***	0.347***
	(0.002)	(0.006)
Distance to centre	−0.041***	−0.056***
	(0.000)	(0.000)
Maisonette	−0.022***	−0.040***
	(0.002)	(0.003)
$>5km$ from centre ne	−0.135***	−0.213***
	(0.001)	(0.003)
$>5km$ from centre nw	−0.048***	−0.017***
	(0.001)	(0.003)
$>5km$ from centre se	−0.120***	−0.173***
	(0.001)	(0.003)
$>5km$ from centre sw	−0.026***	−0.007**
	(0.001)	(0.003)
$>15km$ from centre ne	0.005	−0.045***
	(0.003)	(0.006)
$>15km$ from centre nw	0.021***	0.116***
	(0.005)	(0.009)
$>15km$ from centre se	−0.061***	−0.052***
	(0.004)	(0.006)
$>15km$ from centre sw	0.003	−0.047***
	(0.003)	(0.006)
Observations	551,413	159,570
adjusted R^2	0.575	0.650
Residual std. error	0.253 (df = 551372)	0.286 (df = 159529)
F statistic	18681.900***	7410.401***
	(df = 40.0; 551372.0)	(df = 40.0; 159529.0)

Note: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

Vienna

Table 6. Model summary for Vienna list-price index.

	Goodness of fit: Vienna	
	Rental model	Sale model
Constant	5.985***	11.719***
	(0.003)	(0.003)
2019_1	−0.002	0.006
	(0.003)	(0.004)
2019_2	−0.005	−0.003
	(0.003)	(0.004)
2019_3	0.008**	0.016***
	(0.003)	(0.004)
2019_4	0.003	0.007
	(0.003)	(0.004)
2019_5	0.003	0.009**
	(0.003)	(0.004)
2019_6	−0.000	0.012***
	(0.003)	(0.004)
2019_7	0.007**	0.022***
	(0.003)	(0.004)
2019_8	0.004	0.029***
	(0.003)	(0.004)
2019_9	0.006**	0.025***
	(0.003)	(0.004)
2019_10	0.008**	0.034***
	(0.003)	(0.004)
2019_11	0.003	0.053***
	(0.003)	(0.004)
2019_12	0.010***	0.064***
	(0.003)	(0.004)
2020_1	−0.002	0.064***
	(0.003)	(0.004)
2020_2	−0.009**	0.064***
	(0.004)	(0.005)
2020_3	−0.008**	0.058***
	(0.003)	(0.004)
2020_4	0.003	0.046***
	(0.003)	(0.005)
2020_5	−0.001	0.072***
	(0.003)	(0.004)
2020_6	0.012***	0.081***
	(0.003)	(0.004)
2020_7	0.018***	0.094***
	(0.003)	(0.004)
2020_8	0.033***	0.097***
	(0.003)	(0.004)
2020_9	0.040***	0.107***
	(0.003)	(0.004)
2020_10	0.036***	0.118***

Table 6. Continued

	Goodness of fit: Vienna	
	Rental model	Sale model
	(0.003)	(0.004)
2020_11	0.042***	0.120***
	(0.004)	(0.005)
2020_12	0.034***	0.141***
	(0.005)	(0.006)
Historic building	0.094***	0.068***
	(0.001)	(0.001)
Penthouse	0.107***	0.056***
	(0.012)	(0.008)
Need of renovation	−0.248***	−0.238***
	(0.017)	(0.005)
Outspace	0.073***	0.078***
	(0.001)	(0.002)
Square meter	0.012***	0.014***
	(0.000)	(0.000)
Distance to centre	−0.010***	−0.015***
	(0.000)	(0.000)
$> 5km$ from centre ne	0.028***	−0.053***
	(0.002)	(0.002)
$> 5km$ from centre nw	−0.006**	0.052***
	(0.003)	(0.004)
$> 5km$ from centre se	−0.019***	−0.179***
	(0.002)	(0.003)
$> 5km$ from centre sw	−0.011***	−0.026***
	(0.002)	(0.002)
$> 15km$ from centre ne	0.031***	0.039***
	(0.008)	(0.013)
$> 15km$ from centre nw	−0.006	−0.067***
	(0.005)	(0.008)
$> 15km$ from centre se	−0.059***	−0.259***
	(0.009)	(0.015)
$> 15km$ from centre sw	−0.001	0.013*
	(0.004)	(0.007)
Observations	151,176	137,650
adjusted R^2	0.802	0.770
Residual std. error	0.164	0.222
	(df=151137)	(df=137611)
F Statistic	16134.656***	12099.229***
	(df=38.0; 151137.0)	(df=38.0; 137611.0)

Note: *p < 0.1; **p < 0.05; ***p < 0.01

6.2. Descriptive Statistics of Missing Characteristics

Table 7. Data cleaning steps – London.

	Rental		Sale	
	Observations	Median price	Observations	Median price
Total unique	939,466	1,733	335,506	525,000
Property type not flat or maisonette	232,903	1,549	129,452	600,000
missing bedrooms and/or missing bathrooms	78,441	1,842	27,702	475,000
bedrooms >7 and/or bathrooms >5	250	15,167	83	7,600,000
price $>Q_{97}$	19,020	7,800	5,523	3,200,000
price $<Q_3$	19,653	693	5,355	137,500
Final	589,199	1,772	167,391	495,000

Table 8. Data cleaning steps – Vienna.

	Rental		Sale	
	Observations	Median price	Observations	Median price
Total unique	177,280	899	163,640	320,000
Property type not flat, maisonette, penthouse	9,022	850	10,934	559,000
Square meter $>Q_{97}$	4,958	3,345	4,494	1,620,000
Square meter $<Q_3$	4,368	497	4,090	145,100
price per square meter $> Q_{97}$	4,821	1,498	4,388	999,000
price per square meter $<Q_3$	4,815	550	4,384	179,000
Final	151,176	898	137,650	299,990

6.3. *Appendix – Alternative Pre-Crisis Definitions*

6.3.1. List-Price Change

Table 9. Price change since december 2018.

	List-price		London Sale	London Rent	Vienna Sale	Vienna Rent
Development since Dec. 2018	To pre-crisis definition	1.	3.10%	4.19%	6.60%	−0.54%
		2.	3.46%	4.50%	6.39%	−0.63%
		3.	2.72%	3.97%	5.78%	0.03%
	Until end of 2020		2.74%	−3.50%	13.47%	3.80%
Development during COVID-19 crisis			→	↘	↗	↗

Note: We include three possible "pre-COVID-19 crisis" definitions here: In definition 1 the pre-crisis period covers January and February 2020, in definition 2 the period is expanded to include January, February, and March 2020, and in definition 3 we include the period between October 2019 and March 2020. Mean values over these months are displayed. Similarly, we define the three months of our data sets (October to December 2020) as the "end of 2020".

6.3.2. List-Price-Rent Ratio

Table 10. Change of list-price-rent ratio since january 2019.

	List-price to rent ratio		London	Vienna
Development since Jan. 2019	To pre-crisis definition	1.	−0.58%	6.16%
		2.	−0.17%	6.06%
		3.	−0.23%	4.49%
	Until end of 2020		7.41%	6.58%
Development during COVID-19 crisis			↗	↗

Note: We include three possible "pre-COVID-19 crisis" definitions here: In definition 1 the pre-crisis period covers January and February 2020, in definition 2 the period is expanded to include January, February, and March 2020, and in definition 3 we include the period between October 2019 and March 2020. Mean values over these months are displayed. Similarly, we define the three months of our data sets (October to December 2020) as the "end of 2020".

6.3.3. List Volume Development

Table 11. Percentage list volume change compared to december 2018.

	List volume		London		Vienna	
			Sale	Rent	Sale	Rent
Development since Dec. 2018	To pre-crisis definition	1.	42.78%	37.59%	−39.71%	−12.46%
		2.	39.82%	39.33%	−40.09%	−11.42%
		3.	24.07%	23.25%	−26.80%	3.91%
	Until end of 2020		103.00%	96.49%	−34.39%	49.74%
Development during COVID-19 crisis			↗	↗	→	↗

Note: We include three possible “pre-COVID-19 crisis” definitions here: In definition 1 the pre-crisis period covers January and February 2020, in definition 2 the period is expanded to include January, February, and March 2020, and in definition 3 we include the period between October 2019 and March 2020. Mean values over these months are displayed. Similarly, we define the three months of our data sets (October to December 2020) as the “end of 2020”.

6.3.4. List Time-on-Market

Table 12. Average time-on-market in days.

	Monthly time-on-market		London		Vienna	
			Sale	Rent	Sale	Rent
Development since Dec. 2018	To pre-crisis definition	1.	17.99	6.63	20.76	12.79
		2.	18.48	6.24	20.02	12.82
		3.	19.84	6.86	18.33	12.97
	Until end of 2020		12.30	3.53	23.53	12.57
Development during COVID-19 crisis			↘	↘	↗	→

Note: We include three possible “pre-COVID-19 crisis” definitions here: In definition 1 the pre-crisis period covers January and February 2020, in definition 2 the period is expanded to include January, February, and March 2020, and in definition 3 we include the period between October 2019 and March 2020. Mean values over these months are displayed. Similarly, we define the three months of our data sets (October to December 2020) as the “end of 2020”.

6.3.5. List Turnover Rate

Table 13. Percentage average turnover.

	List turnover		London		Vienna	
			Sale	Rent	Sale	Rent
Development since Dec. 2018	To pre-crisis definition	1.	62.33%	83.92%	63.98%	78.28%
		2.	61.19%	84.39%	62.09%	76.36%
		3.	57.09%	81.35%	62.53%	75.42%
	Until end of 2020		71.81%	92.07%	75.89%	75.42%
Development during COVID-19 crisis			↗	↗	↗	↗

Note: We include three possible "pre-COVID-19 crisis" definitions here: In definition 1 the pre-crisis period covers January and February 2020, in definition 2 the period is expanded to include January, February, and March 2020, and in definition 3 we include the period between October 2019 and March 2020. Mean values over these months are displayed. Similarly, we define the three months of our data sets (October to December 2020) as the "end of 2020".

Table 14 presents the monthly development of the three liquidity measurements – *volume*, *time-on-market*, and *list turnover rate* – in one table. Each monthly indicator value is compared to the average value of the same indicator over the previous four months. Changing this moving average to include three or five months does not change the result. We assign an upward liquidity trend for a month if two out of the three liquidity indicators indicate a positive market trend (i.e., increasing *volume*, decreasing *time-on-market*, or increasing *list turnover rate*).

Table 14. Monthly liquidity indicators for London and Vienna from march to december 2020 – sales market.

Month	London				Vienna			
(2020)	Volume	TOM	Turnover	Liquidity	Volume	TOM	Turnover	Liquidity
Mar.	↗	↘	↗	(+)	↘	↘	↘	(−)
Apr.	↘	↗	↘	(−)	↘	↗	↘	(−)
May	↘	↗	↗	(−)	↗	↘	↗	(+)
Jun.	↗	↘	↗	(+)	↗	↘	↘	(+)
Jul.	↗	↘	↗	(+)	↗	↘	↘	(+)
Aug.	↗	↘	↗	(+)	↗	↗	↘	(−)
Sep.	↗	↘	↘	(+)	↗	↗	↗	(+)
Oct.	↗	↘	↗	(+)	↘	↗	↗	(−)
Nov.	↘	↘	↘	(−)	↘	↗	↗	(−)
Dec.	↘	↗	–	(−)	↘	↘	–	(+)

7. References

Ahmed, A., D. Ardila, D. Sanadgol, and D. Sornette. 2016. "Comparing Ask and Transaction Prices in the Swiss Housing Market", *Swiss Finance Institute Research Paper* No. 16–80. DOI: http://dx.doi.org/10.2139/ssrn.2894404.

Anenberg, E., and S. Laufer. 2017. "A More Timely House Price Index", *The Review of Economics and Statistics* 99 (4): 722–734. DOI: https://doi.org/10.1162/REST_a_00634.

Bagehot, W. 1971. "The Only Game in Town", *Financial Analysts Journal* 27(2): 12–14, DOI: https://doi.org/10.2469/faj.v27.n2.12

Benefield, J.D., and W.G. Hardin. 2015. "Does Time-on-Market Measurement Matter?', *The Journal of Real Estate Finance and Econonmics* 50: 52–73, DOI: https://doi.org/10.1007/s11146-013-9450-z.

Brounen, D., and N. Kok. 2011: "On the economics of energy labels in the housing market", *Journal of Environmental Economics and Management*, 62(2): 166–179. DOI: https://doi.org/10.1016/j.jeem.2010.11.006.

Chauvet, M., S. Gabriel, and C. Lutz. 2014. "Mortgage default risk: New evidence from internet search queries", *Journal of Urban Economics* 96: 91–111, DOI: https://doi.org/10.1016/j.jue.2016.08.004.

Choi, H., and H. Varian. 2012. "Predicting the Present with Google Trends", *Economic Record*, 88(1): 2–9. DOI: https://doi.org/10.1111/j.1475-4932.2012.00809.x.

De Haan, J. 2004. "Direct and Indirect Time Dummy Approaches to Hedonic Price Measurement", *Journal of Economic and Social Measurement*, 29: 427–443. DOI: https://doi.org/10.3233/JEM-2004-0235.

De Wit, E.R., P. Englund, and M.K. Francke. 2013. "Price and transaction volume in the Dutch housing market", *Regional Science and Urban Economics*, 43(2): 220–241. DOI: https://doi.org/10.2139/ssrn.1589296

Diewert, W.E., S. Heravi, and M. Silver. 2009. "Hedonic Imputation versus Time Dummy Hedonic Indexes", In *Price Index Concepts and Measurement*, edited by W.E. Diewert, J.S. Greenlees, and C.R. Hulten: 161–196. National Bureau of Economic Research, University of Chicago Press.

Engsted, T., and T.Q. Pedersen. 2015. "Predicting returns and rent growth in the housing market using the rent-price ratio: Evidence from the OECD countries", *Journal ofInternational Money and Finance* 53: 257–275. DOI: https://doi.org/10.1016/j.jimonfin.2015.02.001.

Eurostat. 2013. "Handbook on Residential Property Price Indices (RPPIs)", *Eurostat Methodologies and Working Papers*, Luxembourg, 2013. Available at: https://ec.europa.eu/eurostat/documents/3859598/5925925/KS-RA-12-022-EN.PDF (accessed February 2022).

Fonds Online. 2020. "Airbnb-Effekt: Pandemie schwappt auf Wiener Mietmarkt über'". Available at: https://www.fondsprofessionell.at/sachwerte/news/headline/airbnb-effekt-pandemie-schwappt-auf-wiener-mietmarkt-ueber-198790/ (accessed February 2022).

Francke, M., and M. Korevaar. 2021. "Housing markets in a pandemic: Evidence from historical outbreaks", *Journal of Urban Economics*, 123. DOI: https://doi.org/10.1016/j.jue.2021.103333.

Genesove, D., and C.J. Mayer. 1997. "Equity and Time to Sale in the Real Estate Market", *The American Economic Review*, 87(3): 255–269. Available at: http://www.jstor.org/stable/2951345.

Genesove, D., and C.J. Mayer. 2001. "Loss aversion and seller behavior: Evidence from the housing market", *The Quarterly Journal of Economics*, 116(4): 1233–1260. DOI: https://doi.org/10.1162/003355301753265561.

Haurin, D.R., J.L. Haurin, T. Nadauld, and A. Sanders. 2010. "List Prices, Sale Prices and Marketing Time: An Application to U.S. Housing Markets", *Real Estate Economics*, 38(4): 659–685. DOI: https://doi.org/10.1111/j.1540-6229.2010.00279.x.

Haurin, D.R., S. McGreal, A. Adair, L. Brown, and J.R. Webb. 2013. "List price and sales prices of residential properties during booms and busts", *Journal of Housing Economics*, 22(1): 1–10. DOI: https://doi.org/10.1016/j.jhe.2013.01.003.

Hill, R.J. 2013. "Hedonic price indexes for residential housing: a survey, evaluation and taxonomy", *Journal of Economic Surveys* 27(5): 879–914. DOI: https://doi.org/10.1111/j.1467-6419.2012.00731.x.

Hill, R.J., and D. Melser. 2008. "Hedonic imputation and the price index problem: an application to housing", *Economic Inquiry*, 46(4): 593–609. DOI: https://doi.org/10.1111/j.1465-7295.2007.00110.x.

Hill, R.J., and I.A. Syed. 2016. "Hedonic price-rent ratios, user cost, and departures from equilibrium in the housing market", *Regional Science and Urban Economics*, 56: 60–72. DOI: https://doi.org/10.1016/j.regsciurbeco.2015.11.001.

IMF. 2020. World Economic Outlook, October 2020: A Long and Difficult Ascent. IMF. Available at: https://www.imf.org/en/Publications/WEO/Issues/2020/09/30/world-economic-outlook-october-2020.

Kolbe, J.R., Schulz, M. Wersing, and A. Werwatz. 2021. "Real estate listings and their usefulness for hedonic regressions", *Empirical Economics*, 1–31. DOI: https://doi.org/10.1007/s00181-020-01992-3.

Knight, J., C.F. Sirmans, and G.K. Turnbull. 1994. "List Price Signaling and Buyer Behavior in the Housing Market", *The Journal of Real Estate Finance and Economics*, 9(3): 177–192. DOI: https://doi.org/10.1007/BF01099271.

Lippmann, S.A., and J.J. McCall. 1986. "An Operational Measure of Liquidity", *American Economic Review* 76(1): 43–55. Available at: https://www.jstor.org/stable/1804126.

Lyons, R.C. 2019. "Can list prices accurately capture housing price trends? Insights from extreme markets conditions", *Finance Research Letters*, 30: 228–232. DOI: https://doi.org/10.1016Zj.frl.2018.10.004.

Miller, N. 1978. "Time on the Market and Selling Price", *Real Estate Economics*. DOI: https://doi.org/10.1111/1540-6229.00174.

Mishkin, F.S. 2011. "Monetary Policy Strategy: Lessons from the Crisis", *NBER Working Papers*, 16755. Available at: https://www.nber.org/system/files/working_papers/w16755/w16755.pdf (accessed February 2022).

OeNB. 2019. *Immobilien Aktuell – International, die Immobilienmarktanalyse der OeNB.* Available at: https://www.oenb.at/Publikationen/Volkswirtschaft/immobilien-aktuell.html (accessed February 2022).
OeNB. 2020. *Wohnimmobilienpreisindex*, ÖsterreichischeNationalbank, Available at: https://www.oenb.at/isaweb/report.do?report = 6.6 (accessed February 2022).
Shimizu, C., K.G. Nishimura, and T. Watanabe. 2016. "House prices at different stages of the buying/selling process", *Regional Science and Urban Economics* 59: 37–53. DOI: https://doi.org/10.1016/j.regsciurbeco.2016.04.001.
Silver, M., and S. Heravi. 2001. "Scanner Data and the Measurement of Inflation", *The Economic Journal*, 111(472): 383–404. DOI: https://doi.org/10.1111/1468-0297.00636.
Stein, J. 1995. "Prices and Trading Volume in the Housing Market: A Model with Down-Payment Effects", *The Quarterly Journal of Economics* 110: 379–406. DOI: https://doi.org/10.2307/2118444.
UK Land Registry. 2021a. Available at: https://www.gov.uk/government/organisations/-land-registry (accessed February 2022).
UK Land Registry. 2021b. Available at: https://landregistry.data.gov.uk/app/ukhpi (accessed February 2022).
Van Dijk, D.W., and M.K. Francke. 2018. "Internet search behavior, liquidity and prices in the housing market", *Real Estate Economics* 46(2): 368–403. DOI: https://doi.org/10.1111/1540-6229.12187.
Willhaben. 2021. Available at: https://www.willhaben.at/ (accessed February 2022).
Wu, L., and E. Brynjolfsson. 2015. "The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales", A chapter in Economic Analysis of the Digital Economy, *University of Chicago Press*, 89–118. DOI: 10.7208/chicago/9780226206981.003.0003.
Wong, G. 2008. "Has SARS Infected the Property Market? Evidence From Hong Kong", *Journal of Urban Economics* 63(1): 74–95. DOI: https://doi.org/10.1016Zj.jue.2006.12.007.
ZT datenforum. 2021. Available at: http://www.zt.co.at/ (accessed February 2022).
Zoopla. 2021. Available at: https://www.zoopla.co.uk/ (accessed February 2022).

Received July 2020
Revised January 2021
Accepted August 2021