An evaluation of the methods used by European countries to compute their official house price indices

Robert J. Hill*, Michael Scholz*, Chihiro Shimizu, and Miriam Steurer***

Abstract – Since 2012, Eurostat requires the national statistical institutes (NSIs) in all European Union (EU) countries to compute official House Price Indices (HPIs) at a quarterly frequency. Eurostat recommends computing the HPI using a hedonic method. Most NSIs have followed this advice, although they differ in their choice of method. Some NSIs use stratified medians instead of hedonic methods. We evaluate the theoretical and empirical properties of both hedonic and stratified median methods. Of particular concern is the comparability of the HPIs across countries when computed using different methods. Our empirical comparisons use detailed micro-level data sets for Sydney and Tokyo, containing about 867,000 actual housing transactions. All the hedonic methods perform better than stratified medians. The hedonic methods generate quite similar results, except when applied to new dwellings in Tokyo. This finding shows that the choice of hedonic method can be important for smaller countries with less data. Also, the widely used hedonic repricing method becomes unreliable when the reference shadow prices are not updated frequently.

JEL Classification: C43, E31, R31

Keywords: housing market, price index, repricing, average characteristics, hedonic imputation, rolling time dummy

Reminder:

The opinions and analyses in this article are those of the author(s) and do not necessarily reflect their institution's or Insee's views.

** Department of Economics, University of Graz, Austria (robert.hill@uni-graz.at, michael.scholz@uni-graz.at, miriam.steurer@uni-graz.at)*
*** Nihon University, Tokyo (shimizu.chihiro@nihon-u.ac.jp)*

A preliminary draft was presented at 15th Ottawa Group Meeting, 10-12 May 2017, Eltville am Rhein, Germany. We acknowledge financial support for this project from the Austrian Research Promotion Agency (FFG), grant #10991131. Hill and Steurer were members of the Expert Team advising Eurostat on the treatment of Owner Occupied Housing (OOH) in the Harmonized Index of Consumer Prices (HICP). The idea for this paper developed out of this project. The views expressed here are of the authors, and are not necessarily representative of those of Eurostat.

Received on 9 June 2017, accepted after revisions on 19 February 2018

To cite this article: Hill, R. J., Scholz, M., Shimizu, C. & Steurer, M. (2018). An evaluation of the methods used by European countries to compute their official house price indices. *Economie et Statistique / Economics and Statistics*, 500-501-502, 221–238. https://doi.org/10.24187/ecostat.2018.500t.1953

The fundamental role played by housing in the broader economy has been demonstrated by the global financial crisis of 2007-2011, which began in the US housing market. It is essential therefore that governments, central banks and market participants are kept well informed of trends and fluctuations in house prices. In Europe, Eurostat – the statistical institute of the European Union (EU) – has required since 2012 (see Eurostat, 2017) that the national statistical institutes (NSIs) in all EU member countries compute official house price indices (HPIs). HPIs, however, can be highly sensitive to the method of construction, and this sensitivity can be a source of confusion amongst users (see Silver, 2015). In a European context it is also important that the HPIs of different countries are reasonably comparable, especially in the Eurozone where the HPIs are needed by the European Central Bank for its decisions on monetary policy, financial regulation, and the monitoring of financial stability.

The difficulty in measuring house price developments arises from every house being different both in terms of its physical characteristics and its location. HPIs need to take account of these quality differences. Otherwise, the price index will confound price changes and quality differences. The importance of these measurement problems has been recently recognized by the international community and the European Commission, Eurostat, the UN, ILO, OECD, World Bank and IMF together commissioned a Handbook on Residential Property Price Indices (RPPIs) that was completed in 2013 (see Eurostat, 2013).

Hedonic methods – which express house prices as a function of a vector of characteristics – are ideally suited for constructing quality-adjusted HPIs (see Diewert, 2010; Hill, 2013). Eurostat recommends that the HPI should be computed using a hedonic approach, but has not provided guidance to NSIs as to which hedonic method should be used. As a result, different countries have adopted different methods. In total six different methods are being used:

(i) Repricing: used by Austria, Belgium, Finland, Hungary, Italy, Latvia, Luxembourg, Norway, Slovenia;

(ii) Average characteristics: used by Romania, Spain;

(iii) Hedonic imputation: used by Germany, UK;

(iv) Rolling time dummy (RTD): used by Croatia, Cyprus, France, Ireland, Portugal;

(v) Stratified median (or mix-adjusted median): used by Bulgaria, Czech Republic, Estonia, Iceland, Lithuania, Poland, Slovakia;

(vi) Sales Price Appraisal Ratio (SPAR): used by Denmark, the Netherlands, and Sweden.

The sources for the methods used by each country are listed in Online complement C1. The first four methods are hedonic. Method (v) by averaging medians across strata provides some partial quality adjustment, although not to the same extent as a hedonic method. Method (vi) combines actual prices with expert valuations (see Haan *et al.*, 2008).

For each method, the taxonomy can be further refined, in that two countries using the same basic method in some cases differ slightly in the way it is formulated. For example, with regard to the RTD method, some countries use a two quarter rolling window while others use a four or five quarter window, while with the repricing method countries differ in the frequency with which the reference characteristics shadow prices are updated.

Our objective here is to evaluate the theoretical and empirical properties of the methods (i), (ii), (iii), (iv) and (v) used by NSIs in Europe to compute their HPIs. We do not consider method (vi) – the SPAR method – since for our data sets we do not have access to any expert valuations. Of particular concern is the comparability of the HPIs across countries when computed using different methods.

We show that the underlying structures of the repricing, average characteristics, and hedonic imputation methods share some common features. The RTD method is somewhat different in its approach.

Empirically we compare the hedonic methods and stratified medians using detailed micro-level data for Sydney and Tokyo. These data sets were chosen since together they contain about 867,000 actual housing transactions, and cover quite long time spans. The Sydney data covers 11 years, while the Tokyo data covers 30 years. When comparing hedonic methods, it is important to have a sufficiently long time series, since problems of drift or bias may only emerge over these kinds of time horizons. To understand how these hedonic methods perform in practice it is important that they are compared using real housing data sets, rather than just simulated data. Also, by evaluating EU methods using non-EU data we provide an independent check on method selection.

The empirical comparisons have two main objectives. The first is to establish how sensitive the HPI is to the choice of hedonic method. The second is to see whether any of the hedonic methods (computed on a quarterly basis) behave in anomalous ways, particularly over longer time horizons (e.g., 10+ years). This is potentially a concern especially for the widely used repricing method, which extrapolates to later periods using the estimated characteristic shadow prices of the base period.

The repricing method, when updated at least every five years, performs quite well on our datasets. The biggest surprise is that the Paasche and Laspeyres versions of the hedonic double imputation method exhibit substantial drift in the Sydney apartment dataset. Some drift is also observed in the Tokyo apartment dataset. Fortunately, no NSI in Europe is using either of these methods. The Törnqvist version of the hedonic double imputation method, used by Germany, is not affected by drift.

Eurostat recommends that each NSI compute separate hedonic indices for houses and apartments. We are able to do this for Sydney, but not for Tokyo, since almost all the transactions in the latter are for apartments. Furthermore, indices specifically for new housing are needed for the owner occupied housing price index (OOHI) which is being used on an experimental basis in the Harmonized Index of Consumer Prices (HICP) (see Eurostat, 2017). The age of dwellings is included as a characteristic in the Tokyo data set but not in the Sydney data set. Hence we are able to compute an HPI for new dwellings for Tokyo, but not for Sydney. Our findings for Tokyo in this regard have important implications for HPIs, OOHIs, and the HICP in Europe.

The remainder of the paper is structured as follows. The next section explains the theoretical properties of the hedonic methods used by NSIs in Europe to compute their HPIs. After that the hedonic methods are compared empirically using data for Sydney and Tokyo. Our main findings are then summarized in the conclusion.

Some Alternative Methods for Constructing Hedonic House Price Indices (HPIs)

All the methods considered here are formulated to be compatible with Eurostat guidelines. In other contexts, these methods could be structured in slightly different ways.

Repricing method

The repricing method is currently the most widely used hedonic method for computing the HPI in Europe. It is used by the NSIs of Austria, Belgium, Finland, Hungary, Italy, Latvia, Luxembourg, Norway, and Slovenia.

The repricing method begins by estimating a semilog hedonic model using only the data of year 1. The hedonic model can be written as follows:

$$\ln p_{(1,q),h} = \sum_{c=1}^{C} \beta_{1,c} z_{(1,q),h,c} + \varepsilon_{(1,q),h} \tag{1}$$

where $z_{(1,q),h,c}$ is the level of characteristic c in dwelling h sold in year 1, quarter q. Examples of characteristics include property type (e.g., house or apartment), number of bedrooms, and land area. Also, $\beta_{1,c}$ denotes the shadow price on characteristic c in year 1, and ε is a random error term.

The objective in (1) is to estimate the characteristic shadow prices $\beta_{1,c}$. These shadow prices are computed using the whole year's data.

As it is typically applied in the HPI, the repricing method compares one quarter $(t,q-1)$ with the next quarter (t,q) using the base year's shadow price vector $\hat{\beta}_1$.

The repricing price index formula consists of two components: a quality unadjusted price index (QUPI) and a quality adjustment factor (QAF). The QUPI is the ratio of the geometric mean prices in both periods $(t,q-1)$ and (t,q), computed as follows:

$$QUPI_{(t,q),(t,q-1)} = \frac{\tilde{p}_{(t,q)}}{\tilde{p}_{(t,q-1)}} \tag{2}$$

where $\tilde{p}_{(t,q-1)}$ and $\tilde{p}_{(t,q)}$ denote, respectively, the geometric mean price of dwellings sold in year-quarter $(t,q-1)$ and year-quarter (t,q).

$$\tilde{p}_{(t,q-1)} = \prod_{h}^{H_{(t,q-1)}} \left(p_{(t,q-1),h} \right)^{1/H_{(t,q-1)}},$$
$$\tilde{p}_{(t,q)} = \prod_{h}^{H_{(t,q)}} \left(p_{(t,q),h} \right)^{1/H_{(t,q)}} \tag{3}$$

where $H_{(t,q-1)}$ and $H_{(t,q)}$ denote the number of properties sold in $(t,q-1)$ and (t,q) respectively. Arithmetic means could be used instead. However, geometric means have the advantage of being more compatible with a semi-log regression model.

The next step is to compute a quality adjustment factor (QAF). This is done by using shadow prices of year 1 as a point of reference to compare quality of the average dwelling sold in periods $(t,q-1)$, and (t,q). The formula of the quality adjustment factor is as follows:

$$QAF_{(t,q-1),(t,q)} = \frac{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{1,c} \bar{z}_{(t,q),c}\right)}{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{1,c} \bar{z}_{(t,q-1),c}\right)}, \tag{4}$$

where

$$\bar{z}_{(t,q-1),c} = \frac{1}{H_{(t,q-1)}} \sum_{h=1}^{H_{(t,q-1)}} z_{(t,q-1),h,c},$$

$$\bar{z}_{(t,q),c} = \frac{1}{H_{(t,q)}} \sum_{h=1}^{H_{(t,q)}} z_{(t,q),h,c},$$

denote the average basket of the characteristic c of periods $(t,q-1)$ and (t,q), respectively, computed using the arithmetic mean formula. In the case of dummy variables, such as postcodes, the average measures the proportion of transactions that feature that postcode. For example, if 1 percent of the transactions occur in postcode 1, then the average basket for postcode 1 equals 0.01.

The repricing price index is now obtained by dividing the quality-unadjusted index (QUPI) in (2) by the quality adjustment factor (QAF) in (4) as follows:

$$\frac{P_{(t,q)}}{P_{(t,q-1)}} = \frac{QUPI_{(t,q),(t,q-1)}}{QAF_{(t,q-1),(t,q)}} = \frac{\tilde{p}_{(t,q)}}{\tilde{p}_{(t,q-1)}} \Bigg/ \frac{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{1,c} \bar{z}_{(t,q),c}\right)}{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{1,c} \bar{z}_{(t,q-1),c}\right)}$$
$$= \frac{\tilde{p}_{(t,q)}}{\tilde{p}_{(t,q-1)}} \Bigg/ \frac{Q^{L}_{1,(t,q)}}{Q^{L}_{1,(t,q-1)}}, \tag{5}$$

where $Q^{L}_{1,(t,q)}$ denotes a Laspeyres quantity index between year 1 and quarter (t,q). It can be seen that the QAF can be rewritten as a ratio of Laspeyres indices as follows:

$$QAF_{(t,q-1),(t,q)} = \frac{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{1,c} \bar{z}_{(t,q),c}\right)}{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{1,c} \bar{z}_{1,c}\right)} \Bigg/ \frac{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{1,c} \bar{z}_{(t,q-1),c}\right)}{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{1,c} \bar{z}_{1,c}\right)} = \frac{Q^{L}_{1,(t,q)}}{Q^{L}_{1,(t,q-1)}}.$$

More generally, relative to the first quarter in the data set (1,1), the price index for period (t,q) is calculated as follows:

$$\frac{P_{(t,q)}}{P_{(1,1)}} = \frac{\tilde{p}_{(t,q)}}{\tilde{p}_{(1,1)}} \Bigg/ \frac{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{1,c} \bar{z}_{(t,q),c}\right)}{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{1,c} \bar{z}_{(1,1),c}\right)}$$
$$= \frac{\tilde{p}_{(t,q)}}{\tilde{p}_{(1,1)}} \Bigg/ \frac{Q^{L}_{1,(t,q)}}{Q^{L}_{1,(1,1)}}, \tag{6}$$

where $\tilde{p}$ again denotes a geometric mean price as defined in (3). An interesting feature of the repricing method is that it only requires the hedonic model to be estimated once (in the base year). This is perhaps one reason why it has proved popular with NSIs.

The base year under the repricing method 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.

Average characteristics method

The average characteristics method and the hedonic imputation method both begin by estimating the following semilog hedonic model separately for each period. For example, for periods $(t,q-1)$ and (t,q), the regression model takes the following forms:

$$\ln p_{(t,q-1),h} = \sum_{c=1}^{C} \beta_{(t,q-1),c} z_{(t,q-1),h,c} + \varepsilon_{(t,q-1),h} \tag{7}$$

$$\ln p_{(t,q),h} = \sum_{c=1}^{C} \beta_{(t,q),c} z_{(t,q),h,c} + \varepsilon_{(t,q),h} \tag{8}$$

where h indexes the dwelling transactions in period (t,q), $p_{(t,q),h}$ the transaction price, and $z_{(t,q),h,c}$ is the level of characteristic c in dwelling h. Unlike under the repricing method, the estimated shadow prices on the characteristics,

$\beta_{(t,q),c}$, are specific to period (t,q) and are updated every period.

The next step is to construct an average basket of characteristics. The hedonic method then measures the change in the imputed price of the average dwelling over time. The version used by European NSIs computes an average basket $\overline{z}_{t,c}$ based on a whole year's data calculated using the arithmetic mean formula. The price index between two adjacent quarters in the same year therefore is now calculated as follows:

$$\begin{aligned}\frac{P_{(t,q)}}{P_{(t,q-1)}} &= \frac{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{(t,q),c}\overline{z}_{t-1,c}\right)}{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{(t,q-1),c}\overline{z}_{t-1,c}\right)} \\ &= \frac{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{(t,q),c}\overline{z}_{t-1,c}\right)}{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{t-1,c}\overline{z}_{t-1,c}\right)} \Bigg/ \frac{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{(t,q-1),c}\overline{z}_{t-1,c}\right)}{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{t-1,c}\overline{z}_{t-1,c}\right)} \\ &= \frac{P^{L}_{t-1,(t,q)}}{P^{L}_{t-1,(t,q-1)}}, \end{aligned} \tag{9}$$

where $P^{L}_{t-1,(t,q)}$ denotes a Laspeyres price index between periods t − 1 and (t,q). From the first line of (9) we can see that the overall price index can be written as a Lowe index (i.e., it is a fixed basket index where the time period of the basket is not the same as that of the two time periods being compared). The second line of (9) shows that the overall price can also be expressed as the ratio of two Laspeyres price indices.

Once a year, the average basket of characteristics is updated. This can be done at the end of the year, once all the data for that year are available. The price index between the fourth quarter in one year and the first quarter in the next year therefore is calculated as follows:

$$\begin{aligned}\frac{P_{(t+1,1)}}{P_{(t+4)}} &= \frac{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{(t+1,1),c}\overline{z}_{t,c}\right)}{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{(t,4),c}\overline{z}_{t,c}\right)} \\ &= \frac{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{(t+1,1),c}\overline{z}_{t,c}\right)}{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{t,c}\overline{z}_{t,c}\right)} \Bigg/ \frac{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{(t,4),c}\overline{z}_{t,c}\right)}{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{t,c}\overline{z}_{t,c}\right)} \\ &= \frac{P^{L}_{t,(t+1,1)}}{P^{L}_{t,(t,4)}}, \end{aligned} \tag{10}$$

Again the overall price index can be expressed as the ratio of two Laspeyres price indices.

Relative to the first quarter in the data set (1,1), the price index for period $(t+1,1)$ is calculated as follows:

$$\frac{P_{(t+1,1)}}{P_{(1,1)}} = \frac{P^{L}_{0,(1,2)}}{P^{L}_{0,(1,1)}} \times \frac{P^{L}_{0,(1,3)}}{P^{L}_{0,(1,2)}} \times \frac{P^{L}_{0,(1,4)}}{P^{L}_{0,(1,3)}} \times \frac{P^{L}_{1,(2,1)}}{P^{L}_{1,(1,4)}} \ldots \times \frac{P^{L}_{t-1,(t,4)}}{P^{L}_{t-1,(t,3)}} \times \frac{P^{L}_{t,(t+1,1)}}{P^{L}_{t,(t,4)}}.$$

It turns out that the repricing method can be represented as a fixed base average characteristics method. Suppose, as with the average characteristics and hedonic imputation methods, the hedonic model is estimated for a single quarter. The imputed errors from the semilog hedonic model for quarter s can then be written as follows:

$$\hat{\varepsilon}_{sh} = \ln p_{sh} - \sum_{c=1}^{C}\hat{\beta}_{s,c} z_{shc}.$$

By construction under OLS, $\sum_{h=1}^{H_1}\hat{\varepsilon}_{sh} = 0$. Hence

$$\sum_{h=1}^{H_s}\left[\ln p_{sh} - \sum_{c=1}^{C}\hat{\beta}_{s,c} z_{shc}\right] = 0$$

which in turn implies that the geometric mean price takes the following form:

$$\tilde{p}_s = \exp\left(\sum_{c=1}^{C}\hat{\beta}_{s,c}\overline{z}_{sc}\right).$$

Substituting this expression into the repricing formula (with shadow prices estimated using only the first quarter not the first year) yields the following:

$$\begin{aligned}\frac{P_{(t,q)}}{P_{(1,1)}} &= \frac{\tilde{p}_{(t,q)}}{\tilde{p}_{(t,q-1)}} \Bigg/ \frac{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{(1,1),c}\overline{z}_{(t,q),c}\right)}{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{(1,1),c}\overline{z}_{(t,q-1),c}\right)} \\ &= \frac{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{(t,q),c}\overline{z}_{(t,q),c}\right)}{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{(t,q-1),c}\overline{z}_{(t,q-1),c}\right)} \Bigg/ \frac{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{(1,1),c}\overline{z}_{(t,q),c}\right)}{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{(1,1),c}\overline{z}_{(t,q-1),c}\right)} \\ &= \frac{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{(t,q),c}\overline{z}_{(t,q),c}\right)}{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{(1,1),c}\overline{z}_{(t,q),c}\right)} \Bigg/ \frac{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{(t,q-1),c}\overline{z}_{(t,q-1),c}\right)}{\exp\left(\sum_{c=1}^{C}\hat{\beta}_{(1,1),c}\overline{z}_{(t,q-1),c}\right)} \\ &= \frac{P^{P}_{(1,1),(t,q)}}{P^{P}_{(1,1),(t,q-1)}}, \end{aligned} \tag{11}$$

where $P^{P}_{(1,1),(t,q)}$ denotes a Paasche price index between periods (1,1) and (t,q). Hence the repricing method can also be interpreted as an average characteristics method that uses the Paasche price index formula.

As far as we are aware this is a new result in the literature. It is also somewhat counterintuitive that this version of the repricing method can be written as a ratio of Paasche price indices, since these price indices require the estimated characteristic shadow prices of the periods $(t,q-1)$ and (t,q). By contrast, as can be seen from the first line of (11), in practice all that is needed are the characteristic shadow prices of period (1,1).

Hedonic imputation method

Once a hedonic model has been estimated, it allows one to ask counterfactual questions such as what a particular dwelling actually sold in say period t would have sold for instead in period $t + 1$. Using this approach, the hedonic imputation method constructs price relatives measuring how the price has changed from period t to $t + 1$ for every dwelling sold in period t, and likewise for very dwelling sold in period $t + 1$. These price relatives can then be averaged across dwellings to obtained the overall price index. Here we will present two slightly different variants of the hedonic imputation method. The first is used by the UK NSI and the second by the German NSI. Both versions use the same estimated hedonic model as the average characteristics method in (8) to impute prices for each dwelling. For example, let $\hat{p}_{(t,q),h}(z_{t-1,h})$ denote an imputed price in period (t,q) for dwelling h which was actually sold one year earlier in period $(t-1,q)$. The UK version is a chained Lowe index where the reference basket is all the dwellings sold in the previous year. When comparing two quarters in the same year (here t), the formula is as follows:

$$\frac{P_{(t,q)}}{P_{(t,q-1)}} = \prod_{h=1}^{H_{t-1}} \left[\frac{\hat{p}_{(t,q),h}\left(z_{t-1,h}\right)}{\hat{p}_{(t,q-1),h}\left(z_{t-1,h}\right)} \right]^{1/H_{t-1}}, \tag{12}$$

where H_{t-1} denotes the number of properties sold in year $t-1$. When the 4th quarter is compared with the 1st quarter of the next year the references basket is updated as follows:

$$\frac{P_{(t+1,1)}}{P_{(t,4)}} = \prod_{h=1}^{H_t} \left[\frac{\hat{p}_{(t+1,1),h}\left(z_{t,h}\right)}{\hat{p}_{(t,4),h}\left(z_{t,h}\right)} \right]^{1/H_t}. \tag{13}$$

When the underlying hedonic model has a semilog functional form, the UK method is in fact identical to the average characteristics method described above. This duality between the average characteristics method and the hedonic imputation method is explored in more detail in Hill and Melser (2008). In the case of the UK method, the duality can be demonstrated as follows:

$$\begin{aligned}\frac{P_{(t,q)}}{P_{(t,q-1)}} &= \prod_{h=1}^{H_{t-1}} \left[\frac{\hat{p}_{(t,q),h}\left(z_{t-1,h}\right)}{\hat{p}_{(t,q-1),h}\left(z_{t-1,h}\right)} \right]^{1/H_{t-1}} \\ &= \prod_{h=1}^{H_{t-1}} \left[\frac{\sum_{c=1}^{C} \exp\left(\hat{\beta}_{(t,q)} z_{t-1,h}\right)}{\sum_{c=1}^{C} \exp\left(\hat{\beta}_{(t,q-1)} z_{t-1,h}\right)} \right]^{1/H_{t-1}} \\ &= \frac{\frac{1}{H_{t-1}} \sum_{c=1}^{C} \sum_{h=1}^{H_{t-1}} \exp\left(\hat{\beta}_{(t,q)} z_{t-1,h}\right)}{\frac{1}{H_{t-1}} \sum_{c=1}^{C} \sum_{h=1}^{H_{t-1}} \exp\left(\hat{\beta}_{(t,q-1)} z_{t-1,h}\right)} \\ &= \frac{\exp \sum_{c=1}^{C} \hat{\beta}_{(t,q),c} \bar{z}_{t-1,c}}{\exp \sum_{c=1}^{C} \hat{\beta}_{(t,q-1),c} \bar{z}_{t-1,c}} = \frac{P^L_{t-1,(t,q)}}{P^L_{t-1,(t,q-1)}}.\end{aligned} \tag{14}$$

In an analogous way it can be shown that:

$$\begin{aligned}\frac{P_{(t+1,1)}}{P_{(t,4)}} &= \prod_{h=1}^{H_t} \left[\frac{\hat{p}_{(t+1,1),h}\left(z_{t,h}\right)}{\hat{p}_{(t,4),h}\left(z_{t,h}\right)} \right]^{1/H_t} \\ &= \frac{\exp \sum_{c=1}^{C} \hat{\beta}_{(t+1,1),c} \bar{z}_{t,c}}{\exp \sum_{c=1}^{C} \hat{\beta}_{(t,4),c} \bar{z}_{t,c}} = \frac{P^L_{t,(t+1,1)}}{P^L_{t,(t,4)}}.\end{aligned} \tag{15}$$

The German version by contrast uses a Törnqvist-type formula (i.e., the geometric mean of geometric-Laspeyres and geometric-Paasche-type formulas) defined as follows[1]:

Geometric Laspeyres (GL):

$$\frac{P_{(t,q)}}{P_{(t,q-1)}} = \left[\prod_{h=1}^{H_{(t,q-1)}} \frac{\hat{p}_{(t,q),h}\left(z_{(t,q-1),h}\right)}{\hat{p}_{(t,q-1),h}\left(z_{(t,q-1),h}\right)} \right]^{1/H_{(t,q-1)}} \tag{16}$$

Geometric Paasche (GP):

$$\frac{P_{(t,q)}}{P_{(t,q-1)}} = \left[\prod_{h=1}^{H_{(t,q)}} \frac{\hat{p}_{(t,q),h}\left(z_{(t,q),h}\right)}{\hat{p}_{(t,q-1),h}\left(z_{(t,q),h}\right)} \right]^{1/H_{(t,q)}} \tag{17}$$

Törnqvist:

$$\frac{P_{(t,q)}}{P_{(t,q-1)}} = \left\{ \left[\prod_{h=1}^{H_{(t,q-1)}} \frac{\hat{p}_{(t,q),h}\left(z_{(t,q-1),h}\right)}{\hat{p}_{(t,q-1),h}\left(z_{(t,q-1),h}\right)} \right]^{1/H_{(t,q-1)}} \left[\prod_{h=1}^{H_{(t,q)}} \frac{\hat{p}_{(t,q),h}\left(z_{(t,q),h}\right)}{\hat{p}_{(t,q-1),h}\left(z_{(t,q),h}\right)} \right]^{1/H_{(t,q)}} \right\}^{1/2} \tag{18}$$

Here it makes no difference whether we are comparing two quarters in the same year or the last quarter in one year with the first quarter in the next.

When the underlying hedonic model is semilog, the geometric-Laspeyres (GL), geometric-Paasche (GP), and Törnqvist hedonic imputation indices can be represented as average characteristic methods as follows:

$$\begin{aligned}\text{GL:} &\left[\prod_{h=1}^{H_{(t,q-1)}} \frac{\hat{p}_{(t,q),h}\left(z_{(t,q-1),h}\right)}{\hat{p}_{(t,q-1),h}\left(z_{(t,q-1),h}\right)} \right]^{1/H_{(t,q-1)}} \\ &= \frac{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{(t,q),c} \bar{z}_{(t,q-1),c}\right)}{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{(t,q-1),c} \bar{z}_{(t,q-1),c}\right)} = P^L_{(t,q-1),(t,q)}\end{aligned} \tag{19}$$

1. *Silver (2016, pp. 54–57) refers to the Törnqvist-type indices in (18) and (21) as hybrid Fisher-type indices.*

GP: $$\frac{P_{(t,q)}}{P_{(t,q-1)}} = \left[\prod_{h=1}^{H_{(t,q)}} \frac{\hat{p}_{(t,q),h}\left(z_{(t,q),h}\right)}{\hat{p}_{(t,q-1),h}\left(z_{(t,q),h}\right)} \right]^{1/H_{(t,q)}} \quad (20)$$

$$= \frac{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{(t,q),c} \bar{z}_{(t,q),c}\right)}{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{(t,q-1),c} \bar{z}_{(t,q),c}\right)} = P^{P}_{(t,q-1),(t,q)}$$

Törnqvist:

$$\left\{ \left[\prod_{h=1}^{H_{(t,q-1)}} \frac{\hat{p}_{(t,q),h}\left(z_{(t,q-1),h}\right)}{\hat{p}_{(t,q-1),h}\left(z_{(t,q-1),h}\right)} \right]^{1/H_{(t,q-1)}} \left[\prod_{h=1}^{H_{(t,q)}} \frac{\hat{p}_{(t,q),h}\left(z_{(t,q),h}\right)}{\hat{p}_{(t,q-1),h}\left(z_{(t,q),h}\right)} \right]^{1/H_{(t,q)}} \right\}^{1/2} \quad (21)$$

$$= \left\{ \frac{\exp\left[\sum_{c=1}^{C} \hat{\beta}_{(t,q),c}\left(\bar{z}_{(t,q-1),c} - \bar{z}_{(t,q),c}\right)\right]}{\exp\left[\sum_{c=1}^{C} \hat{\beta}_{(t,q-1),c}\left(\bar{z}_{(t,q-1),c} - \bar{z}_{(t,q),c}\right)\right]} \right\}^{1/2}$$

$$= \left(P^{L}_{(t,q-1),(t,q)} \times P^{P}_{(t,q-1),(t,q)}\right)^{1/2} = P^{F}_{(t,q-1),(t,q)}$$

where $P^{F}_{(t,q-1)(t,q)}$ denotes a Fisher price index comparison between periods $(t,q-1)$ and (t,q).

Relative to the first quarter in the data set (1,1), the price index for period $(t+1,1)$ is calculated as follows:

$$\frac{P_{(t+1,1)}}{P_{(1,1)}} = P^{F}_{(1,1),(1,2)} \times P^{F}_{(1,2),(1,3)} \times \ldots \times P^{F}_{(t+1,1),(t,4)}.$$

In practice this means that the UK method is essentially equivalent to the average characteristic method used by Romania and Spain. While Germany's method can also be represented as an average characteristics method, it is the only country in the EU that uses the Törnqvist formula to construct its HPI.

Rolling time dummy method

The Rolling Time Dummy (RTD) method, as proposed by Shimizu *et al.* (2010) (see also O'Hanlon, 2011), is used by a number of NSIs in Europe. RTD is a variant on the widely used time-dummy hedonic method. The relationship between time-dummy and hedonic imputations methods is explored by Diewert *et al.* (2009) and Haan (2010). When discussing the RTD method we use a slightly different notation than we have used thus far in this paper. We refer simply to periods denoted by s and t, without distinguishing which year and quarter they are in. The RTD method begins by estimating the following hedonic model over a time window of $k+1$ periods starting with period s:

$$\ln p_{uh} = \sum_{c=1}^{C} \beta_{(s,s+k),c} z_{uhc} + \sum_{c=1}^{C} \delta_i D_{ih} + \varepsilon_{uh} \quad (22)$$

where h now indices the dwelling transactions in periods $s,...,s+k$, and D_{ih} is a dummy variable that equals 1 when $u = i$ is the period in which the dwelling sold, and zero otherwise. Now the characteristic shadow prices for each period in the window are assumed to be equal (i.e., $\beta_{s,c} = \beta_{s+1,c} = \cdots = \beta_{s+k,c} = \beta_{(s,s+k),c}$). The RTD method then moves the window forward one period, and re-estimates the model.

The RTD method derives the price index comparing period $t+k-1$ to period $t+k$ as follows:

$$\frac{P_{t+k}}{P_{t+k-1}} = \frac{\exp(\hat{\delta}^{t}_{t+k})}{\exp(\hat{\delta}^{t}_{t+k-1})} \quad (23)$$

A superscript t is included on the estimated δ coefficients to indicate that they obtained from the hedonic model with period t as the base. The hedonic model with period t as the base is only used to compute the change in dwelling 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 dwelling prices from period $t+k$ to period $t+k+1$ is now computed as follows:

$$\frac{P_{t+k+1}}{P_{t+k}} = \frac{\exp(\hat{\delta}^{t+1}_{t+k+1})}{\exp(\hat{\delta}^{t+1}_{t+k})} \quad (24)$$

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:

$$\frac{P_{t+k+1}}{P_{t}} = \frac{\exp(\hat{\delta}^{t-k}_{t+1})}{\exp(\hat{\delta}^{t-k}_{t})} \frac{\exp(\hat{\delta}^{t-k+1}_{t+2})}{\exp(\hat{\delta}^{t-k+1}_{t+1})} \cdots \frac{\exp(\hat{\delta}^{t+1}_{t+k+1})}{\exp(\hat{\delta}^{t+1}_{t+k})}. \quad (25)$$

A trade-off exists when choosing the window length. A longer window length increases the sample size and robustness of the price index. On the other hand, a longer window acts to smooth the price signal, providing a less timely and market relevant indicator. The optimal window length will differ depending on the datset. When data points are scarce, RTD4Q and RTD5Q (i.e., 4 or 5 quarter windows) are recommended over RTD2Q (i.e., a 2 quarter window). NSIs in Europe using the RTD method have selected the following window lengths: France = 2, Cyprus = 4, Ireland = 5, Portugal = 2, Croatia = 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 indices P_t, P_{t+1}, ..., P_{t+k} are already fixed. The sole objective when estimating the hedonic model inclusive of data from period $t+k+1$ is to compute P_{t+k+1}, irrespective of how many periods are included in the hedonic model. More generally, this property of never being revised is recommended by Eurostat (2017) and is shared by all the hedonic price indices considered here. To be clear, by non-revisability we mean that simply adding a new period of data does not change the results for earlier periods. If new data become available for earlier periods, this is another matter. In this case, it may be desirable to revise the existing index.

Stratification and stratified medians

The RPPI Manual published by Eurostat (2013) recommends that the data should be divided into broad strata by region and building type, and then hedonic methods should be applied separately to each stratum. The results are then averaged across strata typically using the arithmetic mean formula. One issue that arises is whether the arithmetic mean formula should be weighted by the number of transactions or the housing stock in each stratum. Weighting by the housing stock in each stratum might be preferable for macroeconomics analysis, when such stock weights are available. Failing that, weighting by number of transactions is probably preferable to equal weighting.

Sometimes however insufficient data or resources are available to compute hedonic indices. In such situations stratified medians are often used as a simpler and less reliable alternative to hedonic methods. The first step in computing a stratified (or mixadjusted) median index is to split the data set into strata. As with hedonic methods, the first split should be between houses and apartments. Each stratum should be further subdivided based on location, for example by province, county, district or postcode. When information on the physical characteristics of dwelling are available, splits can also be done based say on size (for example floor area less than 80 square meters and greater than 80 square meters), or age (e.g., new and existing). In the empirical applications, after splitting houses and apartments, we focus on locational stratification based on postcodes and Residex regions for Sydney, and wards in Tokyo.

Once the strata have been constructed, the median price for each stratum is computed. These medians are then averaged separately for houses and apartments, typically using the arithmetic mean formula. Again the issue arises as to whether the average should be weighted by the number of transactions or the housing stock in each stratum.

With regard to computational complexity, a stratified median method lies somewhere in between a simple median method and a quality-adjusted hedonic method[2]. Averaging medians across strata reduces the noise in the index resulting from compositional changes in the median dwelling over time. While in principle more strata should imply better quality adjustment, this approach soon runs into the problem when the classification becomes finer that some of the strata may be empty in some periods (i.e., there are no transactions with that particular mix of characteristics). This imposes limits on how far stratified median methods can take the quality-adjustment process.

Evaluations of the different methods for Sydney (2003-2014)

The Sydney data set

We use a data set obtained from Australian Property Monitors that consists of prices and characteristics of houses and apartments sold in Sydney (Australia) for the years 2002-2014. Results are presented for the years 2003-2014. For some methods, data for 2002 are needed to compute the reference baskets used in 2003.

The functional form for our hedonic models is semilog. The explanatory characteristics for houses are as follows:

- the actual sale price;
- time of sale;
- property type (i.e., detached or semi);
- number of bedrooms;
- number of bathrooms;
- land area;
- postcode (there are 202 postcodes in the data set).

For apartments we have the same set of characteristics. However, we drop the land area characteristic for apartments in our hedonic

2. *The median price per square meter could be viewed as a highly restrictive version of a hedonic method.*

analysis since it refers to the whole strata, and we do not have any information on the number of apartments in the building. 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, caused for example by the inclusion of erroneous extra zeroes. These extreme observations can distort the results. The exclusion criteria we applied are shown in Table 1. Complete data on all our hedonic characteristics are available for 380,414 house transactions. For apartments the corresponding figure is 250,005.

Summary of methods to be considered

The methods that will be compared (of which the first ten are hedonic) are listed below:

1. Repricing (no updating of base year);
2. Repricing (base year updated every five years);
3. Repricing (base year updated every year);
4. Average characteristics;
5. Double imputation Geometric-Laspeyres;
6. Double imputation Geometric-Paasche;
7. Double imputation Törnqvist;
8. RTD (2 quarters);
9. RTD (4 quarters);
10. RTD (5 quarters);
11. Stratified median.

In the case of Sydney, price indices will be computed separately for houses and apartments. An overall HPI for Sydney could then be computed using the standard method for aggregating strata briefly discussed above and recommended in chapter 5 of the RPPI Manual (see Eurostat, 2013). For Tokyo only data for apartments are available. The age of dwellings is available for Tokyo but not for Sydney. So, for Tokyo we compute price indices for all apartments and for new apartments.

It is particularly important to determine how well the methods used by NSIs perform on a data set for new dwellings, since a price index for new dwellings is a key input into the experimental owner-occupied housing price index (OOHI) in Europe. The OOHI is in turn being considered for inclusion in the harmonized index of consumer prices (HICP).

House and apartment price indices for Sydney

The house price indices (HPIs) for Sydney generated by the various methods discussed above are shown in Table C3-1 (in Online complement C3). Five of the series are graphed in Figure I. As is clear from Table C3-1 and Figure I, the HPI is quite robust to the choice of method. Over the whole sample period, depending on the choice of hedonic method, house prices rose by between 73.7 and 78.1 percent. The three repricing methods – RP1 which uses shadow prices from 2003, RP2 which updates the shadow prices every five years, and RP3 which updates the shadow prices every year – generate the lowest increase in house prices[3]. Also shown in Table C3-1 are stratified median results computed in two different ways. MIX-PC stratifies houses by postcodes of which there are 202. MIX-RX stratifies by Residex region of which there are 16[4].

The MIX-PC stratification is hence much finer than its MIX-RX counterpart. It is not surprising therefore that the MIX-PC index is less erratic and closer to the hedonic indices.

3. Examples of the estimated characteristic shadow prices from the hedonic models are provided for Sydney in 2003 and for Tokyo in 2002 in Online complement C2. It can be seen that most of the shadow prices are significantly different from zero at the 5 percent significance level, and the adjusted R-squared for are about 0.85.

4. The Residex regions (with their constituent postcodes listed in brackets) are as follows: Inner Sydney (2000 to 2020), Eastern Suburbs (2021 to 2036), Inner West (2037 to 2059), Lower North Shore (2060 to 2069), Upper North Shore (2070 to 2087), Mosman-Cremorne (2088 to 2091), Manly-Warringah (2092 to 2109), North Western (2110 to 2126), Western Suburbs (2127 to 2145), Parramatta Hills (2146 to 2159), Fairfield-Liverpool (2160 to 2189), Canterbury-Bankstown (2190 to 2200), St George (2201 to 2223), Cronulla-Sutherland (2224 to 2249), Campbelltown (2552 to 2570), Penrith-Windsor (2740 to 2777).

Table 1
Criteria for removing outliers

	Price (in dollars)	Bed	Bath	Area (m^2)
Minimum Allowed	100,000	1	1	100
Maximum Allowed	4,000,000	6	6	10,000

The MIX-PC index rises by 82 percent while MIX-RX rises by 87 percent. The concern is not just that it rises faster than the hedonic indices, but also that it is more volatile, as can be seen from Figure I.

Volatility is an important issue. A higher level of volatility can indicate insufficient quality adjustment[5]. Two measures of volatility are the root mean squared error (RMSE) and mean absolute deviation (MAD) stated in (26) and (27) for the case of year-on-year comparisons for the same quarter. Here we define both RSME and MAD in terms of deviations of log ratios.

$$\mathrm{RMSE} = \sqrt{\frac{1}{T-1}\sum_{t-1}^{T-1}\left[\ln\left(\frac{p_{(t+1,q)}}{p_{(t,q)}}\right) - \frac{1}{T-1}\ln\left(\frac{p_{(T,q)}}{p_{(1,q)}}\right)\right]^2} \quad (26)$$

$$\mathrm{MAD} = \frac{1}{T-1}\sum_{t-1}^{T-1}\left|\ln\left(\frac{p_{(t+1,q)}}{p_{(t,q)}}\right) - \frac{1}{T-1}\ln\left(\frac{p_{(T,q)}}{p_{(1,q)}}\right)\right| \quad (27)$$

$$\mathrm{MIN} = \mathrm{Min}_{1,\ldots,T-1}\left\{100\left[\left(\frac{p_{(t+1,q)}}{p_{(t,q)}}\right)-1\right]\right\} \quad (28)$$

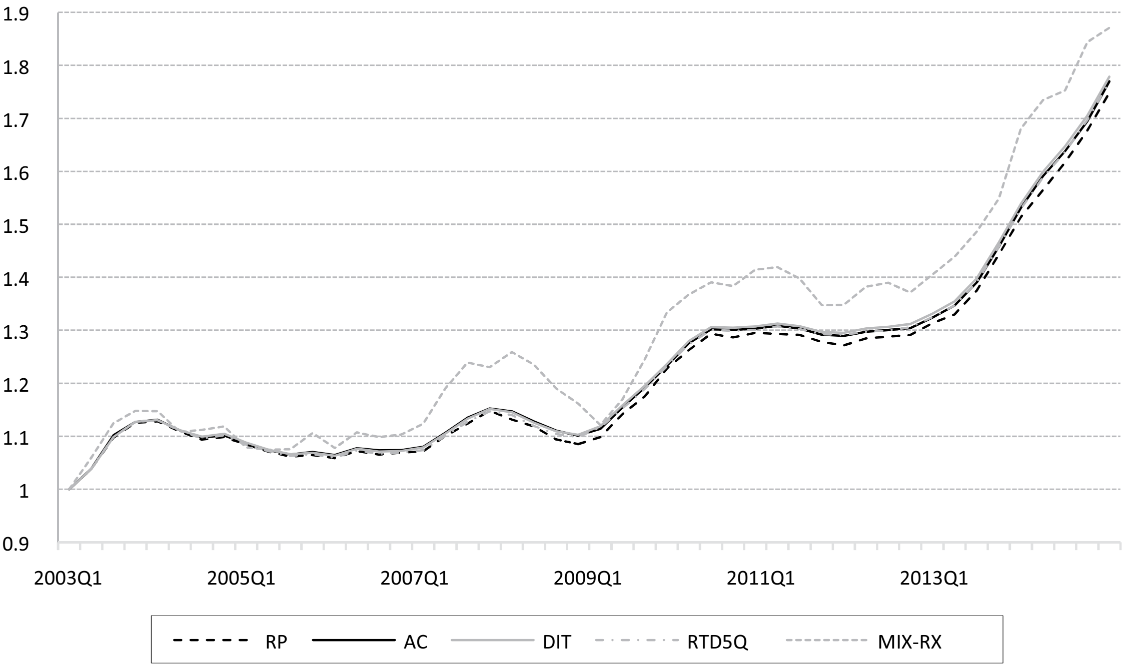
$$\mathrm{MAX} = \mathrm{Max}_{1,\ldots,T-1}\left\{100\left[\left(\frac{p_{(t+1,q)}}{p_{(t,q)}}\right)-1\right]\right\} \quad (29)$$

RSME, MAD, MAX and MIN statistics for Sydney houses are given in Table 2. These statistics are computed both on a year-by-year and quarter-by-quarter basis. It can be seen in Table 2 that the stratified median indices are more volatile than the hedonic indices (especially in the quarter-on-quarter comparisons). This is to be expected since the stratified medians fail to fully adjust for changes in the quality of the median over time. For the same reason, the volatility of the MIX.PC stratified median is lower than that of MIX.RX. This is because the finer stratification of MIX.PC allows it to do a better job of quality adjusting the price index.

The results for apartments in Sydney shown in Figure II are also reasonably robust to the choice of method, when we restrict the comparison to the hedonic methods actually used by NSIs to compute the HPI. The measured cumulative rise in apartment prices for the hedonic methods ranges between 68.1 and 72.6 percent. The stratified median index MIX-RX, by contrast, rises by 80 percent.

5. However, one must be careful in this regard since in a volatile market a good price index should capture this volatility.

Figure I
Estimates of Price Indices for Houses in Sydney (2003Q1 = 1)



Note: Hedonic methods: RP = Repricing; AC = Average characteristics; DIT = Double imputation Törnqvist; RTD5Q = Rolling time dummy with five quarter window; Stratified median method: MIX-RX = Mix adjusted stratified by Residex region. Period: 2002-2014.
Coverage: Houses in Sydney, Australia.
Sources: Australian Property Monitors; authors' calculations.

Table 2
Volatility of the House Price Indices in Sydney

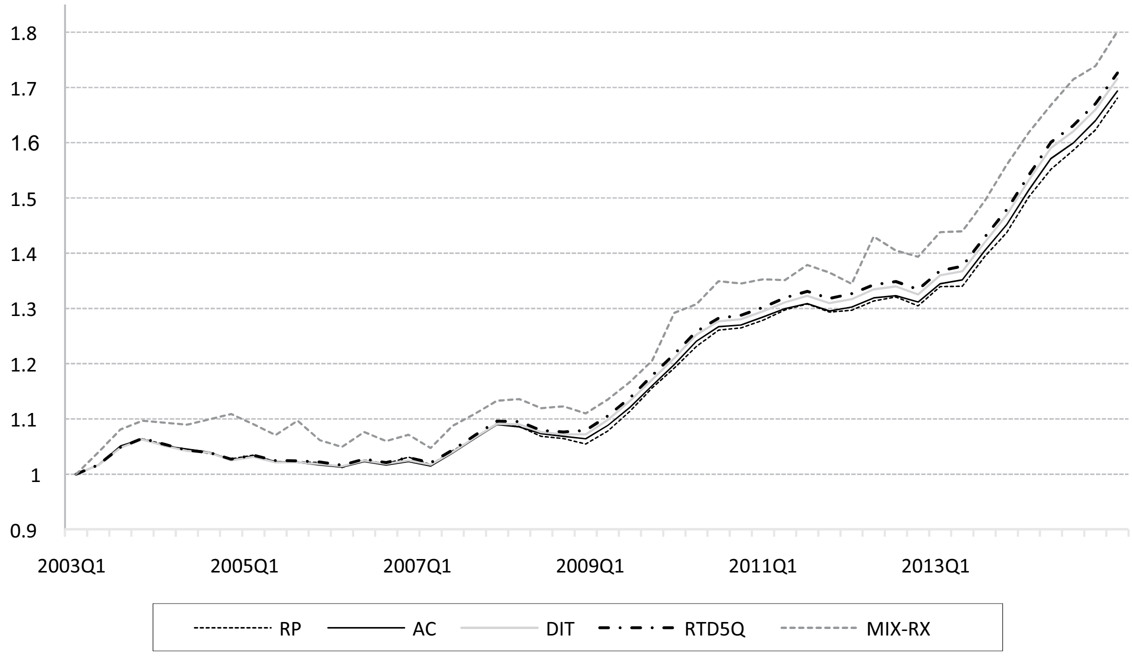
	RP1	RP2	RP3	AC	DIL	DIP	DIT	RTD2Q	RTD4Q	RTD5Q	MIX-PC	MIX-RX
Year-on-Year (Q1)												
RMSE	0.068	0.065	0.066	0.068	0.067	0.066	0.067	0.067	0.066	0.066	0.086	0.096
MAD	0.057	0.055	0.056	0.058	0.057	0.057	0.057	0.057	0.056	0.056	0.072	0.079
MIN	-3.90	3.90	-3.93	-3.96	-3.69	-3.79	-3.74	-3.76	-3.95	-4.03	-6.31	-10.95
MAX	17.69	17.93	17.93	18.14	18.13	18.01	18.07	18.06	18.00	17.99	20.14	21.97
Year-on-Year (Q2)												
RMSE	0.057	0.056	0.056	0.056	0.056	0.055	0.056	0.056	0.056	0.056	0.068	0.069
MAD	0.047	0.047	0.046	0.046	0.047	0.045	0.046	0.046	0.046	0.046	0.054	0.054
MIN	-3.47	-3.47	-3.56	-3.36	-3.36	-3.25	-3.30	-3.31	-3.40	-3.46	-4.28	-5.29
MAX	17.63	17.73	17.75	17.84	17.97	17.74	17.86	17.85	17.76	17.73	18.56	18.87
Year-on-Year (Q3)												
RMSE	0.058	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.057	0.066	0.071
MAD	0.052	0.051	0.051	0.051	0.051	0.052	0.051	0.051	0.051	0.051	0.059	0.062
MIN	-2.93	-2.93	-3.10	-2.95	-2.95	-3.07	-3.01	-2.99	-3.04	-3.04	-3.79	-3.95
MAX	16.20	16.46	16.33	16.25	16.38	16.28	16.33	16.37	16.41	16.40	14.53	19.07
Year-on-Year (Q4)												
RMSE	0.069	0.067	0.066	0.067	0.067	0.066	0.066	0.066	0.067	0.067	0.071	0.077
MAD	0.061	0.06	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.059	0.062	0.067
MIN	-5.46	-5.02	-4.21	-4.45	-4.43	-4.03	-4.23	-4.21	-4.17	-4.24	-5.64	-5.58
MAX	15.50	15.65	15.65	15.62	15.72	15.51	15.60	15.63	15.69	15.66	17.14	19.63
Quarter-on-Quarter												
RMSE	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.027	0.030
MAD	0.018	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.021	0.025
MIN	-2.22	-1.74	-1.74	-1.75	-1.76	-1.74	-1.74	-1.76	-1.81	-1.79	-4.82	-3.64
MAX	5.69	5.69	5.76	6.08	5.78	5.86	5.82	5.79	5.75	5.73	7.53	8.54

Note: The RSME, MAD, MIN and MAX statistics are defined in (26), (27), (28) and (29). The hedonic methods are as follows: RP1 = Repricing without updating; RP2 = Repricing where the base period is updated every five years; RP3 = Repricing where the base period is updated every year; AC = Average characteristics; DIL = Double imputation Laspeyres; DIP = Double imputation Paasche; DIT = Double imputation Törnqvist; RTD2Q = Rolling time dummy with a 2 quarter rolling window; RTD4Q and RTD5Q have 4 and 5 quarter rolling windows; the stratified median methods are as follows: MIX-PC = Mix adjusted stratified by postcode; MIX-RX = Mix adjusted stratified by Residex region. Period: 2002-2014. Coverage: Houses in Sydney, Australia.
Sources: Australian Property Monitors; authors' calculations.

The double imputation Paasche (DIP) and Laspeyres (DIL) indices – shown in Table C3-2 (see Online complement C3) but excluded from Figure II – exhibit clear evidence of drift. According to DIP, prices rise by only 65.3 percent while according to DIL prices rise by 78.1 percent. It is fortunate therefore that none of the NSIs are using either DIP or DIL. The German NSI uses the double imputation Törnqvist (DIT) method, which is the geometric mean of DIP and DIL. The results indicate that the drift in DIP and DIL is offsetting, and hence DIT seems to be unaffected by any drift problems.

Given the duality between average characteristic and hedonic imputation methods, we should also consider the implications of this finding for the former. The average characteristics method, which uses a Laspeyres type formula, is potentially also at risk of drift. However, the drift arises here when the average dwelling is updated each quarter based on the previous quarter's data. The average characteristics method used by NSIs only updates the average dwelling annually and computes it based on a whole year's data. This seems to be enough to prevent drift.

Figure II
Estimates of Price Indices for Apartments in Sydney (2003Q1 = 1)



Note: Hedonic methods: RP = Repricing; AC = Average characteristics; DIT = Double imputation Törnqvist; RTD5Q = Rolling time dummy with five quarter window; Stratified median method: MIX-RX = Mix adjusted stratified by Residex region. Period: 2002-2014.
Coverage: Apartments in Sydney, Australia.
Sources: Australian Property Monitors; authors' calculations.

The results for the RSME, MAD, MAX and MIN statistics for Sydney apartments are given in Table 3. Again the stratified median indices are more volatile than the hedonic indices. Overall, the results in Tables C3-1 and C3-2 should be reassuring to Eurostat. They indicate that the HPIs of different countries should be broadly comparable even when computed using different hedonic methods. For those countries using stratified medians, it is important that the strata are sufficiently finely defined. Otherwise, like MIX-RX, the index will behave erratically.

Evaluations of the different methods for Tokyo (1986-2016)

The Tokyo data set

The Tokyo data set consists of 23 wards of the Tokyo metropolitan area (621 square kilometers), and the analysis period is approximately 30 years between January 1986 and June 2016. The data set covers previously-owned condominiums (apartments transactions) published in *Residential Information Weekly* (or *Shukan Jyutaku Joho* in Japanese) 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 are: floor area, age of building, travel time to nearest station, travel time to Tokyo central station, and the 23 wards (i.e., city codes). The hedonic model for Tokyo is estimated over 237,190 observations. A few observations were deleted since they were incomplete, or contained clear errors. The total number of deletions was less than 1 percent. The functional form for our hedonic models is again semilog. The explanatory variables used are:

- log of floor area;
- age (included as a quadratic);
- time to nearest station;
- time to Tokyo central station (included as a quadratic);
- ward dummy.

Table 3
Volatility of the Apartment Price Indices in Sydney

	RP1	RP2	RP3	AC	DIL	DIP	DIT	RTD2Q	RTD4Q	RTD5Q	MIX-PC	MIX-RX
Year-on-Year (Q1)												
RMSE	0.055	0.053	0.053	0.054	0.054	0.053	0.054	0.054	0.053	0.053	0.054	0.059
MAD	0.045	0.043	0.044	0.044	0.044	0.042	0.043	0.043	0.043	0.043	0.044	0.051
MIN	-2.17	-2.17	-1.75	-2.00	-2.26	-1.51	-1.89	-1.90	-1.88	-1.86	-3.52	-3.77
MAX	15.81	16.25	16.25	16.25	16.42	16.29	16.36	16.38	16.33	16.32	15.22	15.87
Year-on-Year (Q2)												
RMSE	0.045	0.044	0.044	0.045	0.045	0.044	0.045	0.045	0.044	0.044	0.049	0.050
MAD	0.034	0.035	0.034	0.035	0.036	0.034	0.035	0.035	0.035	0.035	0.039	0.038
MIN	-1.90	-1.90	-1.78	-2.20	-2.24	-1.75	-1.99	-2.01	-1.85	-1.80	-3.24	-1.74
MAX	13.82	14.16	14.16	13.92	14.18	14.17	14.17	14.18	14.18	14.15	14.40	15.63
Year-on-Year (Q3)												
RMSE	0.047	0.048	0.047	0.048	0.049	0.047	0.048	0.048	0.048	0.048	0.051	0.048
MAD	0.043	0.043	0.043	0.044	0.044	0.043	0.044	0.044	0.043	0.044	0.047	0.042
MIN	-1.50	-1.50	-1.47	-1.77	-1.84	-1.55	-1.69	-1.70	-1.49	-1.45	-2.39	-3.36
MAX	12.93	12.95	12.94	12.95	12.97	12.92	12.94	13.00	13.00	12.96	11.69	11.97
Year-on-Year (Q4)												
RMSE	0.055	0.053	0.053	0.055	0.055	0.053	0.054	0.054	0.054	0.054	0.059	0.059
MAD	0.049	0.048	0.047	0.049	0.049	0.047	0.048	0.048	0.048	0.048	0.053	0.050
MIN	-3.44	-3.40	-3.02	-3.56	-4.10	-3.05	-3.58	-3.48	-3.49	-3.52	-3.92	-4.27
MAX	13.05	12.82	12.75	12.62	12.65	13.52	12.89	12.83	12.90	12.82	14.11	16.39
Quarter-on-Quarter												
RMSE	0.017	0.016	0.016	0.016	0.017	0.016	0.016	0.016	0.016	0.016	0.021	0.023
MAD	0.015	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.018	0.019
MIN	-1.65	-1.48	-1.48	-1.34	-1.54	-1.22	-1.38	-1.34	-1.43	-1.44	-2.60	-3.20
MAX	4.49	4.17	4.17	4.28	4.34	4.15	4.25	4.24	4.22	4.21	5.91	7.17

Note: The RSME, MAD, MIN and MAX statistics are defined in (26), (27), (28) and (29). The hedonic methods are as follows: RP1 = Repricing without updating; RP2 = Repricing where the base period is updated every five years; RP3 = Repricing where the base period is updated every year; AC = Average characteristics; DIL = Double imputation Laspeyres; DIP = Double imputation Paasche; DIT = Double imputation Törnqvist; RTD2Q = Rolling time dummy with a 2 quarter rolling window; RTD4Q and RTD5Q have 4 and 5 quarter rolling windows; the stratified median methods are as follows: MIX-PC = Mix adjusted stratified by postcode; MIX-RX = Mix adjusted stratified by Residex region. Period: 2002-2014.
Coverage: Apartments in Sydney, Australia.
Sources: Australian Property Monitors; authors' calculations.

The hedonic method considered is essentially the same as for Sydney. The reason for including quadratics for *age* and *time to Tokyo central station* is that the impact of these variables on log(price) may be nonlinear and even possibly non monotonic. For example, there may be an optimal time to Tokyo central station (i.e., one may not want to live too near and not too far way either). This quadratic specification, however, can create problems with the repricing method, as is explained below.

Price indices for all apartments in Tokyo

The results for Tokyo for the years 1986 to 2016 for all apartments are shown in Table C3-3 (in Online complement C3) and Figure III. The general pattern that emerges is similar to that observed for Sydney, although there are some important differences.

Focusing first on the differences, two versions of the repricing method without rebasing – RP1(qd) and RP1 – are presented in Table C3-3. RP1 is much closer to the other methods than RP1(qd). RP1(qd) and RP1 differ in that the former uses the functional form discussed above that includes *age* and *time to Tokyo central station* as quadratics. RP1 includes these variables as linear functions. The problem with RP1(qd) is that while the quadratics by construction fit the data well in

1987 – the first full year of the data set – this specification does not perform so well when it is applied to data in other years. The squared term in the quadratics can distort the results for later years. The implication is that there is a trade-off between model fit in the base period and overall performance of the HPI. When the repricing method is used, quadratic terms in the hedonic model should be avoided. It is better to stick with a simpler more linear model. This problem was not observed for the Sydney data set because these variables were not included in the hedonic model.

A second difference is that there is no clear evidence of drift in the DIL and DIP results in Table C3-3, as compared with what was observed for Sydney apartments. Over the whole sample period, the rise in apartment prices for all hedonic methods, excluding repricing without updating, ranges between 8.5 and 13.8 percent. The average masks a rollercoaster ride where prices first went way up and then way down before gradually returning to near their starting point.

Turning now to the similarities between the results for Sydney and Tokyo, the drift in the repricing results is again downward; although smaller for RP1 than RP1(qd). According to RP1, prices rose by about 7 percent, as opposed to a 6 percent fall when RP1(qd) is used. It

Table 4
Volatility of the Apartment Price Indices in Tokyo

	RP1(qd)	RP1	RP2	RP3	AC	DIL	DIP	DIT	RTD2Q	RTD4Q	RTD5Q	MIX
Year-on-Year (Q1)												
RMSE	0.106	0.102	0.096	0.091	0.092	0.092	0.091	0.091	0.092	0.092	0.092	0.103
MAD	0.087	0.082	0.074	0.07	0.073	0.072	0.072	0.072	0.072	0.072	0.072	0.085
MIN	-17.12	-16.58	-15.33	-15.55	-15.47	-15.61	-15.48	-15.54	-15.55	-15.52	-15.52	-15.46
MAX	33.88	34.42	34.42	30.79	30.14	30.09	30.41	30.25	30.27	30.67	30.84	32.38
Year-on-Year (Q2)												
RMSE	0.104	0.100	0.095	0.093	0.093	0.093	0.092	0.093	0.093	0.093	0.093	0.098
MAD	0.08	0.077	0.073	0.072	0.072	0.073	0.071	0.072	0.072	0.072	0.072	0.077
MIN	-20.31	-19.45	-17.93	-18.07	-18.34	-18.36	-18.20	-18.28	-18.28	-18.28	-18.28	-15.88
MAX	30.43	30.60	30.60	27.52	26.82	27.11	27.45	27.28	27.30	27.75	27.93	28.90
Year-on-Year (Q3)												
RMSE	0.100	0.098	0.095	0.094	0.092	0.093	0.093	0.093	0.093	0.093	0.093	0.106
MAD	0.077	0.075	0.074	0.073	0.072	0.073	0.072	0.072	0.073	0.073	0.073	0.082
MIN	-20.93	-20.32	-19.12	-19.20	-19.35	-19.43	-19.39	-19.41	-19.41	-19.38	-19.35	-19.07
MAX	21.36	21.70	21.70	21.13	20.17	19.94	20.03	19.95	19.99	20.32	20.42	30.99
Year-on-Year (Q4)												
RMSE	0.109	0.106	0.101	0.098	0.100	0.099	0.100	0.100	0.100	0.101	0.101	0.116
MAD	0.084	0.078	0.074	0.073	0.074	0.074	0.074	0.074	0.074	0.075	0.075	0.088
MIN	-19.25	-18.38	-16.98	-17.14	-17.12	-17.14	-17.38	-17.26	-17.27	-17.30	-17.29	-17.92
MAX	30.17	30.24	30.24	29.17	30.43	29.53	30.63	30.08	30.02	31.80	31.58	38.98
Quarter-on-Quarter												
RMSE	0.037	0.035	0.033	0.033	0.032	0.032	0.032	0.032	0.032	0.033	0.033	0.042
MAD	0.027	0.025	0.023	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.033
MIN	-6.41	-6.72	-6.72	-6.52	-6.30	-6.23	-6.29	-6.26	-6.27	-6.32	-6.35	-9.52
MAX	18.14	18.05	18.05	22.35	17.50	17.90	16.89	17.39	17.35	18.49	18.95	17.47

Note: The RSME, MAD, MIN and MAX statistics are defined in (26), (27), (28) and (29). The hedonic methods are as follows: RP1 (qd) = Repricing without updating where the impact of age and time to Tokyo central station are modelled using quadratics; RP1 = Repricing without updating; RP2 = Repricing where the base period is updated every five years; RP3 = Repricing where the base period is updated every year; AC = Average characteristics; Double imputation Laspeyres = DIL; Double imputation Paasche = DIP; Double imputation Törnqvist = DIT; RTD2Q = Rolling time dummy with a 2 quarter rolling window; RTD4Q and RTD5Q have 4 and 5 quarter rolling windows; the stratified median method is: MIX = Mix adjusted stratified by ward. Period: 1986-2016.
Coverage: Apartments in Tokyo, Japan.
Sources: Residential Information Weekly (RECRUIT, Co); authors' calculations.

is noticeable that, as with Sydney apartments, RP2 (repricing where the base period is updated every five years) generates very similar results to the average characteristics (AC) method. Given the duality that exists between the repricing and average characteristics methods this result is not so surprising. However, such similarity in the RP2 and AC results was not observed for houses in Sydney in Table C3-1.

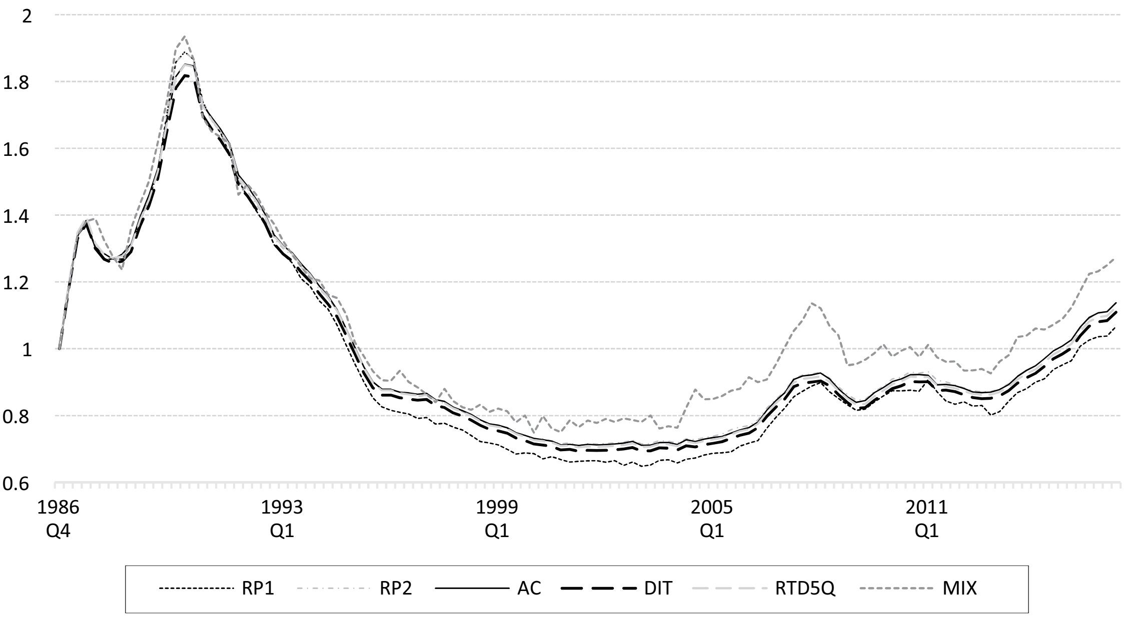
The stratified median index differs quite significantly from the hedonic indices. It rises by 27.4 percent, as compared with the hedonic range of 8.5 and 13.8 percent. It is worth noting that the stratified median indices in all three Figures rise faster than their hedonic counterparts. One possible explanation for this finding is that the average quality of dwellings sold has increased over time. The results for the RSME, MAD, MAX and MIN statistics for Tokyo apartments are given in Table 4. Again the stratified median index MIX is more volatile than the hedonic indices.

Price indices for new apartments in Tokyo

Estimating a price index for new apartments is difficult for the Tokyo data set due to the small sample size. We define a new build as any apartment that is less than three years old. We would have preferred less than two years old, but this is not really feasible. The main problem here is to estimate shadow prices for the locational ward dummy variables. For some wards, no transactions on new dwellings are observed in some quarters. This is a problem particularly for the average characteristics and hedonic imputation methods that begin by estimating a separate hedonic model for each quarter as stated in (8). One way of dealing with missing wards is to restrict a comparison between adjacent quarters to the apartments sold in wards that are observed in both quarters. This means that two different hedonic models need to be estimated for each quarter q. The first includes the apartments sold in wards that are observed in both $q-1$ and q, while the second includes the apartments sold in wards observed in both q and $q+1$. If the characteristics vectors z being priced include any wards not included in the estimated hedonic model, then these wards are dropped and the weights on the remaining wards are adjusted so that they still sum to one.

This problem is not as severe for the repricing method since it estimates the hedonic model

Figure III
Estimates of Price Indices for Apartments in Tokyo (1986Q1 = 1)



Note: Hedonic methods: RP1 = Repricing; RP2 = Repricing where the base period is updated every five years; AC = Average characteristics; DIT = Double imputation Törnqvist; RTD5Q = Rolling time dummy with five quarter window; Stratified median method: MIX = Mix adjusted stratified by ward. Period: 1986-2016.
Coverage: Apartments in Tokyo, Japan.
Sources: Residential Information Weekly (RECRUIT, Co); authors' calculations.

based on a whole year's data, as stated in (1). Again, though, if a ward is nor observed in the base year, then all apartments sold in that ward in future periods are excluded from the comparison. An alternative approach would be to substitute an adjacent ward for these apartments.

By contrast, the problem of missing wards does not arise for the RTD method. Any wards that are observed in any quarter can be included in the RTD hedonic model, as stated in (22). This example illustrates an important advantage of the RTD method, in that it performs well on smaller data sets.

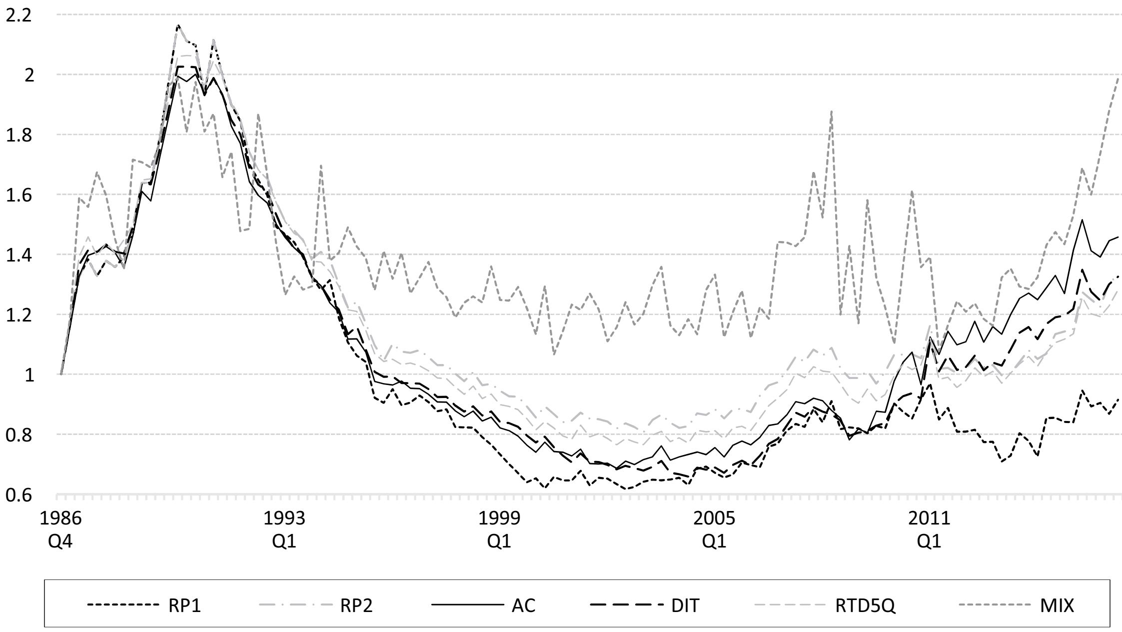
The new apartment price indices are shown in Figure IV. It can be seen that the index is much more sensitive to the choice of method than in Figures I, II, and III.

The stratified median index is particularly badly affected by the small sample size. Faced with a small sample problem, we have greatest confidence in the RTD method with a relatively long window (e.g., RTD5Q). Using RTD5Q as our benchmark, the downward drift in the repricing index (RP1) is much more pronounced than in the previous figures. Updating the reference shadow prices every five years (RP2) solves this problem. Indeed RP2 approximates RTD5Q quite closely. The average characteristics and hedonic imputation methods in this case are somewhat erratic. This is presumably because there are not enough data points to justify estimating a separate hedonic model each quarter.

The results for new builds in Tokyo have important implications for the HPI in smaller EU countries. In cases where there are less data, as Figure IV clearly illustrates, the choice of method for constructing the HPI becomes much more important. The number of new built apartments in Tokyo each quarter may well be higher than the total number of house or apartment transactions in countries such as Slovenia, Malta and Cyprus.

Figure IV also illustrates one of the problems with the acquisitions method for including owner-occupied housing (OOH) in the HICP. The acquisitions method as recommended by Eurostat requires, where possible, a price index specifically for new builds. It is much harder, however, to construct a reliable quality-adjusted HPI for new builds than it is to construct an index covering all housing transactions.

Figure IV
Estimates of Price Indices for New Apartments in Tokyo (1986Q1 = 1)



Note: Hedonic methods: RP1 = Repricing; RP2 = Repricing where the base period is updated every five years; AC = Average characteristics; DIT = Double imputation Törnqvist; RTD5Q = Rolling time dummy with five quarter window; Stratified median method: MIX = Mix adjusted stratified by ward. Period: 1986-2016.
Coverage: New Apartments in Tokyo, Japan.
Sources: Residential Information Weekly (RECRUIT, Co); authors' calculations.

While a price index for new builds may be needed in Europe for the HICP (when OOH is included using the acquisitions method), in the context of the HPI it does not make sense to compute separate HPIs for new builds and existing dwellings and then combine them. Rather, when available, age (or a dummy variable for new builds) should be included directly as a characteristic in a single hedonic model that encompasses both new and existing dwellings. Separate hedonic models, however, should be estimated for houses and apartments since the list of available characteristics for each may differ, and even when they do coincide, shadow prices that are representative for one may not be for the other. The methods outlined in chapter 5 of the RPPI Manual can then be used to construct an overall HPI that combines both houses and apartments (see Eurostat, 2013).

* *
*

Our main findings are as follows:

• The price indices seem to be quite robust over the range of hedonic methods used by NSIs in Europe to compute their HPIs.

• In smaller data sets (e.g., new builds in Tokyo), the HPI becomes more sensitive to the choice of method. Hence smaller countries in the EU need to be more careful when choosing a method. We recommend RTD4Q or RTD5Q for smaller countries with less housing transactions.

• The double imputation Paasche and Lasepyres (DIP and DIL) indices for apartments in Sydney are subject to drift. Evidence of a small amount of drift is also apparent in the Tokyo data. The results for Sydney apartments indicate that DIP and DIL should not be used. Fortunately, no NSIs are using either of these methods.

• The repricing method seems to have a downward bias relative to the other hedonic indices when the reference shadow prices are updated only every five years or not at all. However, this bias is no longer in evidence when the reference shadow prices are updated every year.

• With the repricing method, a hedonic model that performs well in the base period may not provide a good fit in later periods. In particular, for Tokyo the quadratic terms for *age* and *time to Tokyo central station* cause problems. In this sense, there is a greater risk of problems with the repricing method. Hence we recommend keeping the functional form of the hedonic model quite simple (e.g., with no quadratic terms) when the repricing method is used.

• We recommend that NSIs using the repricing method update the reference shadow prices frequently, preferably every year and at least every five years.

• Where possible, the use of stratified median indices should be avoided. This is because they fail to properly adjust for changes in quality over time. The upward bias of stratified medians over the whole sample period in both datasets can be attributed to an upward trend in the quality of transacted dwellings over time. The higher RSME and MAD statistics can be attributed to the stratified-medians not properly adjusting for changes in the quality of transacted dwellings on a period-to-period basis.

• It is more difficult to construct a quality-adjusted price index for new builds. Again, RTD5Q is recommended for computing an HPI for new dwellings when there is a shortage of data points.

• For the HPI we recommend not splitting new and existing dwellings. It is better to combine them in the same hedonic model, with age as one of the explanatory characteristics.

• Houses and apartments should be estimated using separate hedonic models, and then combined using the standard Eurostat method for combining strata (see Eurostat, 2013, chapter X).

• Finally, it should be noted that computing an HPI for a large city is easier than for a whole country, particularly if that country is small. Hence our empirical comparisons may err on side of underestimating the sensitivity of a national HPI to the choice of method used for constructing it. ☐

BIBLIOGRAPHY

Diewert, W. E. (2011). Alternative Approaches to Measuring House Price Inflation. *Vancouver School of Economics – Economics Working Paper* 2010-10.
https://econ.sites.olt.ubc.ca/files/2013/06/pdf_paper_erwin-diewert-10-10-alternative-approaches.pdf.

Diewert, W. E., Heravi, S. & Silver, M. (2009). Hedonic Imputation versus Time Dummy Hedonic Indexes. In W. E. Diewert, W. E., Greenlees, J. & Hulten, C. (Eds.), *Price Index Concepts and Measurement*. Chicago: University of Chicago Press, 161–196.
http://www.nber.org/books/diew08-1

Eurostat (2013). *Handbook on Residential Property Prices Indices (RPPIs)*. Luxembourg: European Union.
http://ec.europa.eu/eurostat/documents/3859598/5925925/KS-RA-12-022-EN.PDF

Eurostat (2017). *Detailed Technical Manual on Owner-Occupied Housing for Harmonised Index of Consumer Prices*. Luxembourg: Eurostat.
http://ec.europa.eu/eurostat/documents/7590317/0/Technical-Manual-OOH-HPI-2017/

Haan, J. de, Wal, E. van der, & Vries, P. de (2008). The Measurement of House Prices: A Review of the SPAR Method. *Statistics Netherlands Working Paper*.
https://www.cbs.nl/nr/rdonlyres/1392243b-5bf2-4c56-8a4b-6c0c1a1cc6ee/0/20080814sparmethodart.pdf

Haan J. de (2010). Hedonic Price Indexes: A Comparison of Imputation, Time Dummy and Re-Pricing Methods. *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)*, 230(6), 772–791.
https://doi.org/10.1515/jbnst-2010-0611

Hill, R. J. (2013). Hedonic Price Indexes for Housing: A Survey, Evaluation and Taxonomy. *Journal of Economic Surveys*, 27(5), 879–914.
https://doi.org/10.1111/j.1467-6419.2012.00731.x

Hill, R. J. & Melser, D. (2008). Hedonic Imputation and the Price Index Problem: An Application to Housing. *Economic Inquiry*, 46(4), 593–609.
https://doi.org/10.1111/j.1465-7295.2007.00110.x

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.
http://www.tara.tcd.ie/bitstream/handle/2262/62349/o%27hanlon%20pdf.pdf?sequence=1&isAllowed=y

Shimizu, C, Takatsuji, H., Ono, H. & Nishimura, K. G. (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.
https://doi.org/10.1108/17538271011080655

Silver, M. (2015). The Degree and Impact of Differences in House Price Index Measurement. *Journal of Economic and Social Measurement*, 39(4), 305–328.
https://doi.org/10.3233/JEM-150406

Silver, M. (2016). How to Better Measure Residential Property Price Indexes. *IMF Working Paper* WP/16/213.
https://www.imf.org/external/pubs/ft/wp/2016/wp16213.pdf