A Note on Concept and Measurement of Property Price Indices

Bert M. Balk[1]
Rotterdam School of Management
Erasmus University
bbalk@rsm.nl

Draft, 2 April 2013

0. *Introduction and summary*

Property price indices, whether residential or commercial, attempt to measure aggregate development of property value, broadly formulated. A number of specific measurement tools are available and, for the case of residential property, extensively documented in the *Handbook on Residential Property Price Indices*, its latest version being available at the Eurostat website.

This paper describes an attempt to integrate the various methods from a common perspective. Section 1 describes what could be called the bottom-up approach; that is, the approach from given transaction data. Section 2 turns to the top-down approach, where in a precise way the various measurement targets are conceptualized. Section 3 then follows with a discussion of a number of sample-based estimation methods. Section 4 looks at stochastic models – hedonic models being a special case – to see how they fit into the overall framework. This, however, is unfinished business.

1. *A transactions based index*

The first thing that comes into mind when contemplating the construction of a Residential or Commercial Property Price Index (RPPI or CPPI, respectively) is collecting all or a sample of prices belonging to residential or commercial property transactions during a certain period, which must not be too long, and in a certain region, such as a country, a province, or a town. The source of course determines the quality, representativity, and timeliness of the data. Some data cleaning might be necessary. For example one does not want to include properties which are sold more than once in a very short time period. Generally speaking, prices should reflect economic values.[2]

---

[1] With thanks to Erwin Diewert for his comments on the very first version. Earlier versions of this paper have been presented at the Statistische Woche (Nürnberg, 14 September 2010) and the Joint BIS ECB Eurostat IMF OECD Conference on Commercial Property Price Statistics (Frankfurt am Main, 10-11 May 2012).

[2] To facilitate comparisons, especially in the case of commercial properties, transaction prices of entire properties might be replaced by transaction prices per square meter.

The next step is to calculate a mean price and study its development through time. This can be done by comparing the mean price of a certain period to the mean price of the immediately preceding period, or by comparing the mean price of a certain period to the mean price of some other, earlier period. A presentation of the mean price development as index numbers relative to some base period makes both viewpoints possible.

A little bit of notation is needed to assist our further reflections. We consider two periods (say, years): a base period 0 and a (later) comparison period 1. These periods may or may not be adjacent. When the periods 0 and 1 are not adjacent and there are intermediate periods for which data are available then chained indices might be employed; this situation is not considered here.

Let S^t be the sample of properties of a certain type (residential or commercial[3]) for which prices are collected in period t (t = 0,1). Each set S^t can just be depicted as a list of addresses. Let P_h^t denote the transaction price of property h in period t ($h \in S^t$; t = 0,1). The (arithmetic) mean price in period t is then given by

$$P^t(S^t) \equiv \sum_{h \in S^t} P_h^t / n(S^t), \tag{1}$$

where $n(S^t)$ denotes the number of elements of the set S^t. The simplest R(C)PPI, reflecting the price development between periods 0 and 1, is then calculated as the ratio of mean prices,

$$P^{UV}(1,0) \equiv \frac{P^1(S^1)}{P^0(S^0)}. \tag{2}$$

This price index bears the superscript "UV" because, technically, it is a unit value index. If mean prices are calculated for various regions then a similar index can be defined for comparing price levels across regions.

Even this simplest form of a R(C)PPI, however, already invokes discussion. We have, more or less tacitly, assumed that mean = arithmetic mean. But why not use the geometric mean, the harmonic mean, or a location measure such as the median? Further, should the mean be based on the entire sample or on some trimmed version (where a percentage of the lowest and highest prices are removed)? These are questions which are not so simple to answer without specifying the measurement target from a more general perspective. But let us for a moment put aside such questions.

A more serious problem with the index defined by (2) is that the mix of properties transacted during period 1 may be vastly different from the mix of those transacted during period 0. Thus one seriously runs the risk of comparing the mean price of apples

[3] Of course, there must be fairly precise, operational definitions of the concepts 'residential' and 'commercial'. Commercial property is usually subdivided into office, retail (hotels, shops, etcetera), and industrial buildings.

with the mean price of oranges, which may be interesting for some purposes, but is not very helpful in general. The simplest cure is *standardization*.[4] What do we mean by this?

Suppose that we know a bit more about the properties transacted than only their address; say, there is information about type, age, number of rooms, etcetera. Then one can split each sample S^t into a number of non-overlapping (or, disjunct) subsamples:

$$S^t = S_1^t \cup S_2^t \cup \cdots \cup S_K^t . \quad (3)$$

It is then immediately clear that the mean price, as calculated over the entire sample, can be written as a weighted mean of mean prices for the subsamples:

$$P^t(S^t) = \sum_{k=1}^{K} a(S_k^t) P^t(S_k^t), \quad (4)$$

where $a(S_k^t) \equiv n(S_k^t) / n(S^t)$ is the relative number of elements of subsample S_k^t and $P^t(S_k^t)$ is the mean price of subsample S_k^t (k=1,…,K). Notice that $\sum_{k=1}^{K} a(S_k^t) = 1$.

There are a number of options for executing standardization on the simple R(C)PPI index (2). The first is to consider the denominator of (2), which can be written as

$$P^0(S^0) = \sum_{k=1}^{K} a(S_k^0) P^0(S_k^0) . \quad (5)$$

Standardization with respect to the *base period* means that (2) is replaced by the Laspeyres index

$$P^L(1,0) \equiv \frac{\sum_{k=1}^{K} a(S_k^0) P^1(S_k^1)}{\sum_{k=1}^{K} a(S_k^0) P^0(S_k^0)} . \quad (6)$$

Thus, as it were, the base period mix is assumed to exist also in the comparison period, and then overall means are computed and compared. A problem arises whenever one or more of the subsamples S_k^1 are empty. The finer the classification in (5) is, the more likely this is to happen. Expression (6) can be rewritten as follows,

$$P^L(1,0) \equiv \sum_{k=1}^{K} \frac{a(S_k^0) P^0(S_k^0)}{\sum_{k=1}^{K} a(S_k^0) P^0(S_k^0)} \frac{P^1(S_k^1)}{P^0(S_k^0)} ; \quad (7)$$

[4] Also called mix adjustment, or (post-) stratification. The most sophisticated standardization method is to use an hedonic model.

that is, as a weighted arithmetic mean of simple R(C)PPIs for subsamples (or cells). If the prices concern entire properties, these weights are the relative transaction values of the base period.

The second standardization departs from the numerator of (2), which can be written as

$$P^1(S^1) = \sum_{k=1}^{K} a(S_k^1) P^1(S_k^1). \tag{8}$$

Then (2) is replaced by the Paasche index

$$P^P(1,0) \equiv \frac{\sum_{k=1}^{K} a(S_k^1) P^1(S_k^1)}{\sum_{k=1}^{K} a(S_k^1) P^0(S_k^0)}. \tag{9}$$

Here the *comparison period* mix is assumed to have existed also in the base period. And one runs into problems when one or more of the subsamples S_k^0 are empty. Expression (9) can be rewritten as

$$P^P(1,0) \equiv \left(\sum_{k=1}^{K} \frac{a(S_k^1) P^1(S_k^1)}{\sum_{k=1}^{K} a(S_k^1) P^1(S_k^1)} \left(\frac{P^1(S_k^1)}{P^0(S_k^0)} \right)^{-1} \right)^{-1}. \tag{10}$$

This expression resembles expression (7), the difference being that (10) is a weighted harmonic mean, whereby, if the prices concern entire properties, the weights are the relative transaction values of the comparison period.

The third variant standardizes (2) on the mix of some *third* period *b*, which delivers the Lowe index

$$P^{Lo}(1,0;b) \equiv \frac{\sum_{k=1}^{K} a(S_k^b) P^1(S_k^1)}{\sum_{k=1}^{K} a(S_k^b) P^0(S_k^0)}. \tag{11}$$

Here one runs into problems when S_k^0 is non-empty but S_k^1 is empty or *vice versa*. Usually the period *b* will be a period prior to period 0.

A fourth alternative would be to take the unweighted geometric mean of expressions (6) and (9); that is,

$$P^F(1,0) \equiv \left(P^L(1,0) P^P(1,0) \right)^{1/2}. \tag{11a}$$

This is the Fisher index.

Thus, if one wants to standardize the simple R(C)PPI as defined by expression (2), there are (at least) four variants. Unfortunately, on the same basic data these variants will in general deliver different outcomes. It is difficult to judge these differences without having recourse to a more general framework, which reiterates a point made earlier. However, a comparison of standardized to unstandardized index numbers delivers important information about mix changes through time.

All the indices (6), (9), (11), and (11a) use mean transaction prices per subsample (cell). These means are weighted by relative transaction numbers from periods 0, 1, or *b*. An interesting innovation was developed by Prasad and Richards (2008). They considered strata consisting of small geographical regions grouped according to the long-term average price level of the residential property units in those regions, measured over a certain time span preceding the base period 0. The overall price index was calculated as an equally-weighted geometric mean of relative median prices[5]

$$P^J(1,0) \equiv \prod_{k=1}^{K} \left(\frac{P^1(S_k^1)}{P^0(S_k^0)} \right)^{1/K} . \tag{11b}$$

Though the behaviour of this index in Australia's six largest cities was shown to correspond closely to that of more sophisticated techniques, also here the question can be posed as to what precisely constitutes the measurement target.

We now turn to a more encompassing framework.

2. *Various targets of measurement*

Suppose that for each period all the existing residential or commercial properties can be listed, and let H^t ($t = 0,1$) be the set of those properties.[6] This set could be subdivided into subsets according to geographical area, type of structure, or age of structure. To keep the exposition simple this complication is for the time being disregarded.

Suppose that every property comes with a price (= value), and let P_h^t denote the price of property *h* in period *t* ($h \in H^t$; $t = 0,1$). This price is the value at which ownership should change on a free market (which is an idealized concept), and is not necessarily equal to some recorded price.[7] The value of the entire stock of residential or commercial properties in period *t* is then given by the sum $\sum_{h \in H^t} P_h^t$. The value change, going from period 0 to period 1, expressed as a ratio, is

[5] This is my interpretation because the paper is vague on the overall index formula. Technically this is called a Jevons price index formula.

[6] Some countries actually have such listings, in some cases operated by semi-governmental agencies, in other cases operated by tax authorities.

[7] Such as book values or values submitted to (or estimated by) tax authorities.

$$V(H^1,H^0) \equiv \frac{\sum_{h\in H^1} P_h^1}{\sum_{h\in H^0} P_h^0}. \tag{12}$$

This could be considered as a *primary target of measurement*. Note that (12) is not a price index, except when $H^0 = H^1$. In that case the expression can be written as a ratio of arithmetic means (and not as a ratio of medians or so). Note also that the value change defined by (12) is transitive; that is, for any three periods 0, 1 and 2 one easily checks that $V(H^2,H^0) = V(H^2,H^1)V(H^1,H^0)$.

In the course of time the composition of the set H^t changes, because (new) structures are added to the stock of properties and (old) structures are removed. Usually, however, the sets H^0 and H^1 have a lot of elements in common. Thus, let $H^{01} \equiv H^0 \cap H^1$ be the set of common or continuing properties. The set $X^0 \equiv H^0 - H^{01}$ contains the discarded properties, whereas the set $N^1 \equiv H^1 - H^{01}$ contains the new properties. Using these definitions, expression (12) can be written as

$$V(H^1,H^0) = \frac{\sum_{h\in H^{01}} P_h^1 + \sum_{h\in N^1} P_h^1}{\sum_{h\in H^{01}} P_h^0 + \sum_{h\in X^0} P_h^0}. \tag{13}$$

We want to split the value change into three parts: the contributions of the continuing, new, and discarded properties. A symmetric method of doing this is by using the logarithmic mean.[8] Then one obtains

$$\ln V(H^1,H^0) = \frac{L(\sum_{h\in H^{01}} P_h^1, \sum_{h\in H^{01}} P_h^0)}{L(\sum_{h\in H^1} P_h^1, \sum_{h\in H^0} P_h^0)} \ln\left(\frac{\sum_{h\in H^{01}} P_h^1}{\sum_{h\in H^{01}} P_h^0}\right) + \frac{\sum_{h\in N^1} P_h^1 - \sum_{h\in X^0} P_h^0}{L(\sum_{h\in H^1} P_h^1, \sum_{h\in H^0} P_h^0)}. \tag{14}$$

The first term at the right-hand side of equation (14) is the contribution of the continuing properties to the (logarithm of the) entire value change, whereas the second term is the net contribution of the new and discarded properties. Notice that (14) remains valid when there are no new or discarded properties. When the two sets X^0 and N^1 are non-empty then expression (14) can be rewritten as

$$\ln V(H^1,H^0) = \frac{L(\sum_{h\in H^{01}} P_h^1, \sum_{h\in H^{01}} P_h^0)}{L(\sum_{h\in H^1} P_h^1, \sum_{h\in H^0} P_h^0)} \ln\left(\frac{\sum_{h\in H^{01}} P_h^1}{\sum_{h\in H^{01}} P_h^0}\right)$$

[8] The logarithmic mean of two strictly positive real numbers *a* and *b* is defined by $L(a,b) = (a-b)/\ln(a/b)$ if $a \neq b$ and $L(a,a) = a$. See Balk (2008) for its properties.

$$+\frac{L(\sum_{h\in N^1} P_h^1, \sum_{h\in X^0} P_h^0)}{L(\sum_{h\in H^1} P_h^1, \sum_{h\in H^0} P_h^0)} \ln\left(\frac{\sum_{h\in N^1} P_h^1}{\sum_{h\in X^0} P_h^0}\right). \quad (15)$$

This expression admits a relatively simple interpretation: the rate of change of the entire stock value is a weighted mean of the rate of change of the value of the continuing properties and the relative value of the new over the discarded properties, the weights being mean value shares.

The value change of the continuing properties is obviously given by

$$V(H^{01}) \equiv \frac{\sum_{h\in H^{01}} P_h^1}{\sum_{h\in H^{01}} P_h^0}. \quad (16)$$

This is a value ratio but, if the quantity of any property is defined as always being equal to 1, then expression (16) is at the same time a price index. Specifically, it is a Dutot index. Expression (16) could be considered as a *secondary target of measurement*. Notice that the index defined by (16) is not transitive; that is, in general it will be the case that $V(H^{02}) \neq V(H^{01})V(H^{12})$.

Notice that (16) can be rewritten as

$$V(H^{01}) = \frac{\sum_{h\in H^{01}} \frac{P_h^1}{P_h^0} P_h^0}{\sum_{h\in H^{01}} P_h^0}; \quad (17)$$

that is, as a weighted arithmetic average of individual value changes; or as

$$V(H^{01}) = \left(\frac{\sum_{h\in H^{01}} \left(\frac{P_h^1}{P_h^0}\right)^{-1} P_h^1}{\sum_{h\in H^{01}} P_h^1}\right)^{-1}; \quad (18)$$

that is, as a weighted harmonic average of individual value changes; or as the simple geometric mean of (17) and (18),

$$V(H^{01}) = \left(\frac{\sum_{h\in H^{01}} \frac{P_h^1}{P_h^0} P_h^0}{\sum_{h\in H^{01}} P_h^0}\right)^{1/2} \left(\frac{\sum_{h\in H^{01}} \left(\frac{P_h^1}{P_h^0}\right)^{-1} P_h^1}{\sum_{h\in H^{01}} P_h^1}\right)^{-1/2}. \quad (18a)$$

We need, however, to go one step further. Properties change through time by depreciation, which expectedly has a negative effect on their value, and by renovation, which expectedly has a positive effect. A simple way to account for these two effects is to define the period 1 constant quality price (relative to period 0) of property *h* by

$$P_h^{1*} \equiv \frac{P_h^1}{(1-\delta_h^{01})(1+\varepsilon_h^{01})}, \tag{19}$$

where $\delta_h^{01} \geq 0$ and $\varepsilon_h^{01} \geq 0$. Notice that depreciation is given with the progress of time – thus the percentage δ_h^{01} depends on the length of the time span between the periods 0 and 1 –, whereas renovation, the percentage of which is given by ε_h^{01}, may or may not happen.

Now expression (16) can be rewritten as

$$V(H^{01}) = \frac{\sum_{h\in H^{01}} P_h^{1*}(P_h^1 / P_h^{1*})}{\sum_{h\in H^{01}} P_h^0 1}. \tag{20}$$

The value change is here expressed in terms of prices and 'quantities'. Going from period 0 to period 1, the price of property *h* changes from P_h^0 to P_h^{1*}, while its 'quantity' changes from 1 to $P_h^1 / P_h^{1*} = (1-\delta_h^{01})(1+\varepsilon_h^{01})$.

Thus the next task is to decompose the value change, given by expression (20), into a price index and a 'quantity' index:

$$V(H^{01}) = P(H^{01})Q(H^{01}). \tag{21}$$

The price index $P(H^{01})$ could be considered as a *tertiary target of measurement.* Given this price index, from expression (15) the *residential (or commercial) property price index (R(C)PPI)* can be defined as

$$\ln R(C)PPI(H^1, H^0) = \frac{L(\sum_{h\in H^{01}} P_h^1, \sum_{h\in H^{01}} P_h^0)}{L(\sum_{h\in H^1} P_h^1, \sum_{h\in H^0} P_h^0)} \ln P(H^{01}) + \frac{\sum_{h\in N^1} P_h^1 - \sum_{h\in X^0} P_h^0}{L(\sum_{h\in H^1} P_h^1, \sum_{h\in H^0} P_h^0)}. \tag{22}$$

Put in words, the *R(C)PPI* measures the value change of residential (or commercial) property stock whereby the value change of the continuing elements is adjusted for the effects of depreciation and renovation.

Index number theory tells us that there are several ways of implementing the decomposition (21); see Balk (2008). For example, using the Laspeyres-Paasche decomposition one obtains

$$P^L(H^{01}) \equiv \frac{\sum_{h\in H^{01}} P_h^{1*}1}{\sum_{h\in H^{01}} P_h^0 1} = \sum_{h\in H^{01}} \frac{P_h^0}{\sum_{h\in H^{01}} P_h^0} \frac{P_h^{1*}}{P_h^0}. \quad (23)$$

Here the constant quality price relatives P_h^{1*} / P_h^0 are weighted by period 0 value shares. Alternatively, using the Paasche-Laspeyres decomposition one obtains

$$P^P(H^{01}) \equiv \frac{\sum_{h\in H^{01}} P_h^{1*}(P_h^1 / P_h^{1*})}{\sum_{h\in H^{01}} P_h^0 (P_h^1 / P_h^{1*})} = \left(\sum_{h\in H^{01}} \frac{P_h^1}{\sum_{h\in H^{01}} P_h^1} \left(\frac{P_h^{1*}}{P_h^0} \right)^{-1} \right)^{-1}. \quad (24)$$

Here the constant quality price relatives P_h^{1*} / P_h^0 are weighted by period 1 value shares. A symmetric price index is given by the Fisher-type formula

$$P^F(H^{01}) \equiv \left[P^L(H^{01}) P^P(H^{01}) \right]^{1/2}. \quad (25)$$

In practice, instead of (20) measurement is usually limited to the simpler expression (16).

3. *Estimation methods*

The problem with calculating (16) is that prices P_h^t are observed only for the subsets of properties transacted $S^t \subset H^t$ $(t = 0,1)$.

The *first* method of estimating (16) is based on the insight that a Dutot price index can be written as a ratio of average prices; that is,

$$V(H^{01}) = \frac{\sum_{h\in H^{01}} P_h^1 / n(H^{01})}{\sum_{h\in H^{01}} P_h^0 / n(H^{01})}, \quad (26)$$

where $n(A)$ denotes the number of elements of the set A.[9] The price index $V(H^{01})$ is then estimated by

$$\hat{V}(H^{01}; SM) \equiv \frac{\sum_{h\in S^1 \cap H^{01}} P_h^1 / n(S^1 \cap H^{01})}{\sum_{h\in S^0 \cap H^{01}} P_h^0 / n(S^0 \cap H^{01})}. \quad (27)$$

[9] Prices of entire properties can be replaced by prices per square meter and, correspondingly, $n(.)$ by total number of square meters.

Notice that the average in the numerator is based on a set of properties that potentially differs from the set used in the denominator. Bias arises when the set $S^0 \cap H^{01}$ or the set $S^1 \cap H^{01}$ is not a representative sample from H^{01}. Moreover, it is clear that if one of these sets is empty, then we are facing a serious problem.

The method as outlined here is the one-stratum version of the so-called *stratification method (SM)* as described by Diewert (2006).[10] The index defined by (27) appears to be a special case of the simple R(C)PPI index as discussed in Section 1 and defined by expression (2). The distinctive feature of (27) is that the samples are restricted to properties which existed in both periods (though not necessarily sold in both periods).

The *second* method is based on those elements of H^{01} which are sold in both periods; that is, the set $S^0 \cap S^1 \cap H^{01}$. This is called the *repeat sales (RS)* method. The price index $V(H^{01})$ is now estimated by

$$\hat{V}(H^{01};RS) \equiv \frac{\sum_{h \in S^0 \cap S^1 \cap H^{01}} P_h^1 / n(S^0 \cap S^1 \cap H^{01})}{\sum_{h \in S^0 \cap S^1 \cap H^{01}} P_h^0 / n(S^0 \cap S^1 \cap H^{01})}. \tag{28}$$

Numerator and denominator of this ratio are based on the same set of properties. Bias arises when the set $S^0 \cap S^1 \cap H^{01}$ is not a representative sample from H^{01}. Moreover, it is clear that we are facing a serious problem when this set is empty.

The *third* method is based on the assumption that for some period b (usually prior to the base period) estimates (assessments) $\hat{P}_h^b$ of all the prices P_h^b ($h \in H^b$) are available. There are two steps involved. The first is that the price index $V(H^{01})$ is estimated by

$$V(H^b \cap H^{01}) \equiv \frac{\sum_{h \in H^b \cap H^{01}} P_h^1}{\sum_{h \in H^b \cap H^{01}} P_h^0} = \frac{\sum_{h \in H^b \cap H^{01}} P_h^1 / \sum_{h \in H^b \cap H^{01}} \hat{P}_h^b}{\sum_{h \in H^b \cap H^{01}} P_h^0 / \sum_{h \in H^b \cap H^{01}} \hat{P}_h^b}. \tag{29}$$

Usually it will be the case that $H^{01} \subset H^b$. However, bias arises when the set $H^b \cap H^{01}$ is not a representative sample from H^{01}. The second step consists in estimating the right-hand side of (29) by

$$\hat{V}(H^{01};DSPAR) \equiv \frac{\sum_{h \in S^1 \cap H^b \cap H^{01}} P_h^1 / \sum_{h \in S^1 \cap H^b \cap H^{01}} \hat{P}_h^b}{\sum_{h \in S^0 \cap H^b \cap H^{01}} P_h^0 / \sum_{h \in S^0 \cap H^b \cap H^{01}} \hat{P}_h^b}. \tag{30}$$

The sets of properties in numerator and denominator are now different. The numerator is based on the properties transacted in period 1 and relates their average price to the average price of these properties in the assessment period. The denominator does the same with the, potentially different, properties transacted in period 0. Additional bias

[10] Note that the stratified version of (26) requires stock-based weights.

arises when the set $S^0 \cap H^b \cap H^{01}$ or the set $S^1 \cap H^b \cap H^{01}$ is not a representative sample from $H^b \cap H^{01}$. This method is an instance of the so-called *sales price appraisal ratio (SPAR)* method, namely the D(utot)SPAR method.

Interestingly, the first method can be seen as a specific case of the third method. Notice that expression (27) can be rewritten as

$$\hat{V}(H^{01};SM) = \frac{\sum_{h\in S^1 \cap H^{01}} P_h^1 / \sum_{h\in S^1 \cap H^{01}} 1}{\sum_{h\in S^0 \cap H^{01}} P_h^0 / \sum_{h\in S^0 \cap H^{01}} 1}. \tag{31}$$

But then it is clear that (30) should be preferred to (31) as estimator of $V(H^{01})$. The ratio $\sum_{h\in S^1 \cap H^b \cap H^{01}} \hat{P}_h^b / \sum_{h\in S^0 \cap H^b \cap H^{01}} \hat{P}_h^b$ adjusts $\hat{V}(H^{01};SM)$ for the price effect of compositional change between the two periods.

Notice that expression (30) can be written as

$$\hat{V}(H^{01};DSPAR) = \frac{\sum_{h\in S^1 \cap H^b \cap H^{01}} \hat{P}_h^b (P_h^1 / \hat{P}_h^b) / \sum_{h\in S^1 \cap H^b \cap H^{01}} \hat{P}_h^b}{\sum_{h\in S^0 \cap H^b \cap H^{01}} \hat{P}_h^b (P_h^0 / \hat{P}_h^b) / \sum_{h\in S^0 \cap H^b \cap H^{01}} \hat{P}_h^b}. \tag{32}$$

By setting $\hat{P}_h^b = c$ for all $h \in H^b$ the right-hand side of expression (32) reduces to

$$\hat{V}(H^{01};CSPAR) \equiv \frac{\sum_{h\in S^1 \cap H^b \cap H^{01}} (P_h^1 / \hat{P}_h^b) / n(S^1 \cap H^b \cap H^{01})}{\sum_{h\in S^0 \cap H^b \cap H^{01}} (P_h^0 / \hat{P}_h^b) / n(S^0 \cap H^b \cap H^{01})}, \tag{33}$$

which defines the C(arli)SPAR method. The J(evons)STAR method is defined by

$$\hat{V}(H^{01};JSPAR) \equiv \frac{\prod_{h\in S^1 \cap H^b \cap H^{01}} (P_h^1 / \hat{P}_h^b)^{1/n(S^1 \cap H^b \cap H^{01})}}{\prod_{h\in S^0 \cap H^b \cap H^{01}} (P_h^0 / \hat{P}_h^b)^{1/n(S^0 \cap H^b \cap H^{01})}}, \tag{34}$$

which differs from (33) in that geometric instead of arithmetic averages are used.

4. *Note on stochastic models*

In the literature the repeat sales method is usually based on a stochastic model of the form

$$P_h^\tau = \alpha_h \beta^\tau \varepsilon_h^\tau \qquad (h \in H^\tau; \tau = 0,1,...,T),$$

where α_h is a time-invariant property specific factor, β^τ a price level indicator, and ε_h^τ remaining noise. Then

$$\ln P_h^t - \ln P_h^s = \beta^t - \beta^s + \varepsilon_h^{ts} \quad (h \in H^{st}; s,t = 0,1,...,T),$$

or

$$\ln(P_h^t / P_h^s) = \sum_{\tau=0}^{T} \beta^\tau D_h^\tau + \varepsilon_h^{ts},$$

where D_h^τ is a dummy variable taking the value -1 if $\tau = s$, $+1$ if $\tau = t$, and 0 otherwise.

Francke's (2010) model can be seen as a special case. He specified

$$\varepsilon_h^{ts} = \gamma_0 + \gamma_1 (t-s)^{-1} + \delta_h^{ts} \quad (t > s)\ (h \in S^s \cap S^t \cap H^{st}),$$

where δ_h^{ts} is supposed to be white noise.

References

Balk, B. M., 2008, *Price and Quantity Index Numbers: Models for Measuring Aggregate Change and Difference* (Cambridge University Press, New York).

Diewert, W.E., 2009, "The Paris OECD-IMF Workshop on Real Estate Price Indexes: Conclusions and Future Directions", in *Price and Productivity Measurement: Volume 1—Housing* (Trafford Press, Victoria, Canada).

Francke, M. K., 2010, "Repeat Sales Index for Thin Markets; A Structural Time Series Approach", *The Journal of Real Estate Finance and Economics* 41, 24-52.

Haan, J. de, and W. E. Diewert (editors), 2011, *Handbook on Residential Property Price Indices*, November 2011 version downloadable from Eurostat website.

Prasad, N. and A. Richards, 2008, "Improving Median Housing Price Indexes through Stratification", *Journal of Real Estate Research* 30(1), 45-71.