STATISTICS DENMARK

Hedonic House Price Index

Prices and Consumption, Statistics Denmark

May 2016

-

Jakob Holmgaard[1]

Abstract

This paper describes different methods for constructing hedonic house price index

[1] Corresponding e-mail address: jho@dst.dk

1. Introduction

With a hedonic price index you can avoid the effect from changes in identified qualitative characteristics of the product. For example, a quarterly nationwide hedonic house price index is not affected by quarterly changes in the distribution of house sales over regions, if the regional sales distribution is included in the characteristics used in the hedonic calculation. The resulting hedonic price will not reflect a shift in sales from provincial regions to the capital; it will only reflect price developments inside the different regions. There are several methods for constructing hedonic price index. Triplett (2006) divides the methods overall into two main groups: Direct methods and indirect methods[2].

1. Characteristic method
2. Time-dummy method
} direct method

3. Imputation method
4. Re-pricing method[3]
} indirect method

The two direct methods are directly based on the products characteristics and used a hedonic regression to determine the price development of the products quantifiable characteristics. That is, the direct methods are fully based on the products characteristics.

The characteristics method is based on a regression equation that makes the price, often the logarithm of the price, a function of the products characteristics. By estimating the hedonic equation each period, we obtain coefficients of the characteristics for each period. These estimated coefficients represent the prices of the characteristics. The last step is to calculate a price index that describes the overall quality adjusted price from period to period. This calculation is done using index formula, like the formula for a Laspeyres or Paasche price index.

The time-dummy method, which is the other direct method, corresponds to the characteristics method with a restriction on the coefficients. It is assumed that the coefficients of the characteristics are unchanged

[2] See illustrative examples in annex 1.

[3] Triplett(2006) refers to this method as “the quality adjustment method”.

between two consecutive periods. When the price of the characteristics is restricted to be unchanged, the quality adjusted price development between two periods can be intercepted by a time-dummy which is 0 in period 1 and 1 in period 2.

The indirect methods are also referred to as composite methods. The starting point is that a product is available in various models. Some models change quality during their life cycles (no price match), while other models do not change quality (price match). For the models with price match, the price can directly be compared and the indirect methods use the unchanged price models directly in a price index. For the models with no match, a hedonic regression is used to calculate the pure price development and inserting these prices into the traditional price index with price match. Thus the resulting price index includes both observable prices and hedonic estimated prices.

The imputation method used the models in one period as representative products and imputes the price for these products in the second period. There is a distinction between single and double imputation – see annex 1 and 2.

The re-pricing method constructs the price development of models without price match by adjusting the pure price with the change in the volume in the models characteristics.

The most general hedonic regression method is the characteristics method. In principle, the other hedonic approaches (time-dummy method, imputation method and re-pricing method) can be seen as special cases of the characteristics method.

It is also easy to show that the hedonic regression is the most general method to construct quality adjusted house price indices. Many other methods (SPAR-method, Repeat sales method, Stratification method) can be seen as special cases of hedonic regression. As an example, the

SPAR-method is formulated as a restricted hedonic regression in annex 2.

Eurostat recommend the indirect hedonic methods. For example, is the re-pricing method recommended to the Harmonised Index of Consumer Prices (HICP), while the imputation method (hedonic Fisher price index) is recommended for Residential Property Price Index (RPPI). The indirect methods are recommended since they resemble the traditional price index theory. However, for house price index, there is no or very little conflict between the direct and indirect method – see later on in this paper.

In annex 1 and 2, all four methods mentioned above are illustrated using two different datasets. The following section presents the hedonic approach and the two direct methods – the characteristic method and the time-dummy method.

About the hedonic approach

The purpose of the hedonic method is to correct for quality differences between products across time. If we, for example, want to find the development in the house prices from quarter one to quarter two, we could start by using the average price of the houses sold. That is, calculate the average price of say 5.000 houses sold in quarter one and the average price of the 6.000 houses sold in quarter two. We may use simple arithmetic averages or perhaps geometric averages considering the wide dispersion in house quality and prices. Then we calculate the percentage change between the average price for quarter 1 and 2 and use it to represent the development in the house prices from quarter 1 to 2.

That is a rather simple method but it is not good enough. The simple price average does not explicitly adjust for differences in the quality of houses sold in the two quarters. For example, if sales increase faster in the big cities than in small towns and villages, the simple average price will increase because the location of houses sold has improved between the two quarters. Location is an example of a quality characteristic, and the average price of the sold houses should be corrected for changes in the geographical locations.

If the value of quality improvements is ignored or underestimated, it will cause the price index to overestimate the price increase over time. And if, for instance, a time series at current prices is deflated by an index that overestimates the price increase we will underestimate the real growth rate.

In order to correct the house price index for quality changes, a regression model is set up to describe some selected characteristics effects on the sales price of the house. More specifically, we formulate an equation that makes the price P_i of house i a function of three characteristics: usable area, location and year of construction:

$$lnP_i = \beta_0 + \beta_1 \cdot X_{i1} + \beta_2 \cdot X_{i2} + \beta_3 \cdot X_{i3} + \varepsilon_i \quad (2)$$

We use the logarithm of the price on the left side, so that a price change of 1 per cent corresponds to 0.01. This is a standard choice, because the distribution of lnP_i is more nicely bell shaped that the distribution of P_i. Usable area is also log-transformed but the transformation is not relevant for the other characteristics. This makes it a semi logarithmic equation, and the contribution of the characteristics must amount to 0.01 to move the house price by one per cent. The sales price and the three characteristics are all observable variables, and there is one price and one set of characteristics for every house sold in a given quarter. The noise term ε_i is an unknown variable that should be uncorrelated with the characteristics and follow a normal distribution, so that we can use OLS (Ordinary Least Square). The four β-parameters are unknown, but OLS gives us an estimate.

Equation (2) can be estimated for each quarter. In our example, there are 5.000 observations in quarter one, and the estimated equation looks like this:

$$lnP_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot X_{i1} + \hat{\beta}_2 \cdot X_{i2} + \hat{\beta}_3 \cdot X_{i3} + u_i \quad (3)$$

The hat on the four β parameters indicates that they are our estimates, and u_i is the calculated residual for each of the 5.000 sold houses. The properties of OLS imply that the sum and therefore the average of the residual is zero for the 5.000 houses, so that equation (3) holds without an error term, if we insert the average of the logarithmic prices for the 5.000 houses sold together with the average of the characteristics:

$$\overline{lnP} = \hat{\beta}_0 + \hat{\beta}_1 \cdot \bar{X}_1 + \hat{\beta}_2 \cdot \bar{X}_2 + \hat{\beta}_3 \cdot \bar{X}_3 \quad (4)$$

In this equation, the variables with a bar on top represent the average of the 5.000 individual observations in quarter one. We note that the subscript i running from 1 to 5.000 has disappeared in (4). Until now we have referred to a particular quarter when presenting the equations, but equation (2) can obviously be estimated for all quarters, so we should introduce a subscript for the period in equation (4):

$$\overline{lnP_t} = \hat{\beta}_{0t} + \hat{\beta}_{1t} \cdot \bar{X}_{1t} + \hat{\beta}_{2t} \cdot \bar{X}_{2t} + \hat{\beta}_{3t} \cdot \bar{X}_{3t} \quad (4a)$$

As the notation implies, we have a set of estimated β-parameters for each quarter, and as already mentioned, there is no error term, so equation (4a) can be seen as an exact or definitional breakdown of the average sales price on selected characteristics ($\bar{X}$'s) and their parameters ($\hat{\beta}$'s). The development over time in the $\bar{X}$'s represents a qualitative and hence a non-price component in the average price. The development over time in the $\hat{\beta}$'s, including the constant, represents the pure price component.

Even though equation (4a) always holds pr. construction, it does not mean that you can always trust the implied decomposition into quality and price. The estimated $\hat{\beta}$' may subject to bias whenever the regression model in (2) is misspecified, for instance due to omitted variables. All estimated parameters may be misleading if we are lacking an important explanatory variable in the regression model.

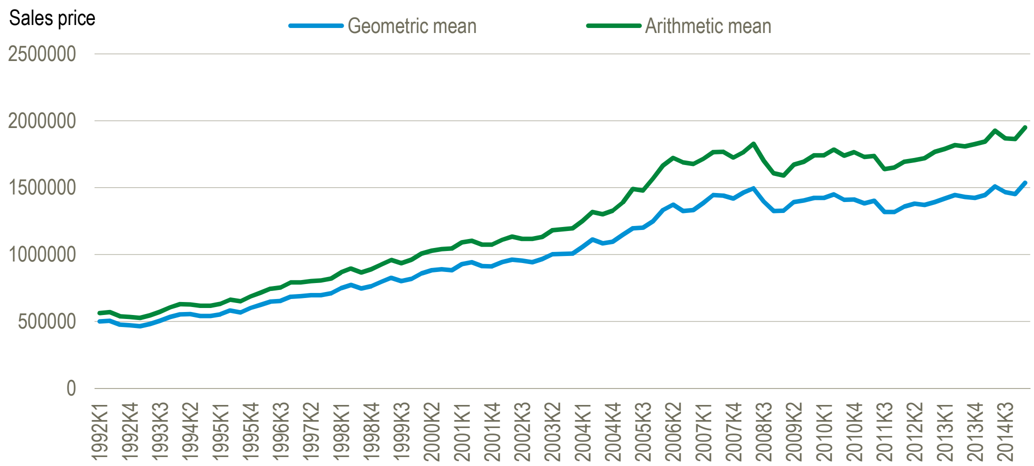
STATISTICS DENMARK

In general, the quality of equation (4a) reflects the quality of the data and the ability of the regression model to explain the observed house prices in each quarter.

Equation (4a) is the starting point for calculating a quality-adjusted hedonic house price index, and for this purpose you can use the decomposition of the average logarithmic price ($\overline{lnP_t}$) in (4a) in different ways. Taking the exponential value on both sides changes (4a) to:

$$exp\left(\overline{lnP_t}\right) = exp\left(\hat{\beta}_{0t} + \hat{\beta}_{1t} \cdot \bar{X}_{1t} + \hat{\beta}_{2t} \cdot \bar{X}_{2t} + \hat{\beta}_{3t} \cdot \bar{X}_{3t}\right) \qquad (4b)$$

The left-hand side of equation (4b) corresponds to the geometric mean of the house prices in period t[4], which is not the same as the arithmetic mean. In the figure below, the geometric mean price is compared to the arithmetic mean price for one-family houses:



As illustrated by the figure above, the arithmetic mean is more affected by the high prices than the geometric mean. For example, the arithmetic mean of 500.000 DKR and 2.000.000 DKR is 1.250.000 DKR, while the geometric mean is only 1.000.000 DKR. Using the geometric average, a price doubling and a half price is canceled out, while the price doubling dominates the arithmetic mean.

[4] In case of two houses sold , we have that:

$$exp\left(\frac{log(P_1)+log(P_2)}{2}\right) = exp\left(0{,}5 \cdot log(P_1) + 0{,}5 \cdot log(P_2)\right) = exp\left(log\left(P_1^{0,5} \cdot P_2^{0,5}\right)\right) = P_1^{0,5} \cdot P_2^{0,5}$$

If we used the simple house price as explained variable (2), we would end up with the arithmetic average of house prices in equation (4b). However, the regression model is explaining the logarithm of house prices, because it brings the distribution of the residuals closer to the normal distribution.

Characteristics method:

Given that equation (4b) decomposes the average house price (exp ($\overline{lnP_t}$)) in price and quantity, with parameters ($\hat{\beta}$'s) as prices and the three quality variables ($\bar{X}$'s) as quantities, it is straightforward to formulate a hedonic Laspeyres price index starting in quarter zero:

$$P_{o:t}^{LA} = \frac{exp(\hat{\beta}_{0t} + \hat{\beta}_{1t} \cdot \bar{X}_{10} + \hat{\beta}_{2t} \cdot \bar{X}_{20} + \hat{\beta}_{3t} \cdot \bar{X}_{30})}{exp(\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \bar{X}_{10} + \hat{\beta}_{20} \cdot \bar{X}_{20} + \hat{\beta}_{30} \cdot \bar{X}_{30})} \tag{5}$$

We note that the hedonic Laspeyres index in (5) resembles the normal Laspeyres index. The difference concerns the use of an exponential function in (5), the constants $\hat{\beta}_{00}$ and $\hat{\beta}_{0t}$ that are prices not multiplied by quantities (we may imagine that they are multiplied by 1). The hedonic Laspeyres price index in (5) describes the quality-adjusted price development, and it is also an example of the characteristics method.

Based on equation (4b), it is equally straightforward to make a hedonic Paasche price index:

$$P_{o:t}^{PA} = \frac{exp(\hat{\beta}_{0t} + \hat{\beta}_{1t} \cdot \bar{X}_{1t} + \hat{\beta}_{2t} \cdot \bar{X}_{2t} + \hat{\beta}_{3t} \cdot \bar{X}_{3t})}{exp(\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \bar{X}_{1t} + \hat{\beta}_{20} \cdot \bar{X}_{2t} + \hat{\beta}_{30} \cdot \bar{X}_{3t})} \tag{5a}$$

We can also use the characteristics method to calculate the price index *implicitly*. For example, if we want to calculate a hedonic Paasche price index, as in equation (5a), we can do this by dividing the value index, which consists solely of the geometric mean of the sales prices with the hedonic Laspeyres quantity index. We get that:

$$V_{0:t} = \frac{exp(\overline{lnP_t})}{exp(\overline{lnP_0})} \quad (5b)$$

$$Q^{LA}_{o:t} = \frac{exp(\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \bar{X}_{1t} + \hat{\beta}_{20} \cdot \bar{X}_{2t} + \hat{\beta}_{30} \cdot \bar{X}_{3t})}{exp(\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \bar{X}_{10} + \hat{\beta}_{20} \cdot \bar{X}_{20} + \hat{\beta}_{30} \cdot \bar{X}_{30})} \quad (5c)$$

By dividing the value index in (5b) with the Laspeyres quantity index in (5c) we get the Paasche price index:

$$P^{PA}_{o:t} = \frac{V_{0:t}}{Q^{LA}_{o:t}} = \frac{\frac{exp(\overline{lnP_t})}{exp(\overline{lnP_0})}}{\frac{exp(\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \bar{X}_{1t} + \hat{\beta}_{20} \cdot \bar{X}_{2t} + \hat{\beta}_{30} \cdot \bar{X}_{3t})}{exp(\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \bar{X}_{10} + \hat{\beta}_{20} \cdot \bar{X}_{20} + \hat{\beta}_{30} \cdot \bar{X}_{30})}}$$

$$= \frac{\frac{exp(\hat{\beta}_{0t} + \hat{\beta}_{1t} \cdot \bar{X}_{1t} + \hat{\beta}_{2t} \cdot \bar{X}_{2t} + \hat{\beta}_{3t} \cdot \bar{X}_{3t})}{exp(\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \bar{X}_{10} + \hat{\beta}_{20} \cdot \bar{X}_{20} + \hat{\beta}_{30} \cdot \bar{X}_{30})}}{\frac{exp(\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \bar{X}_{1t} + \hat{\beta}_{20} \cdot \bar{X}_{2t} + \hat{\beta}_{30} \cdot \bar{X}_{3t})}{exp(\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \bar{X}_{10} + \hat{\beta}_{20} \cdot \bar{X}_{20} + \hat{\beta}_{30} \cdot \bar{X}_{30})}}$$

$$= \frac{exp(\hat{\beta}_{0t} + \hat{\beta}_{1t} \cdot \bar{X}_{1t} + \hat{\beta}_{2t} \cdot \bar{X}_{2t} + \hat{\beta}_{3t} \cdot \bar{X}_{3t})}{exp(\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \bar{X}_{1t} + \hat{\beta}_{20} \cdot \bar{X}_{2t} + \hat{\beta}_{30} \cdot \bar{X}_{3t})} \quad (5d)$$

Notice that equation (5b) can be calculated without using hedonic regression as it is just the geometric mean of the sales prices. Equation (5c) is calculated using hedonic regression, but only for the reference period. Thus calculating the hedonic Paasche price index *implicitly*, it allows you to re-estimate the regression model with a lower frequency, perhaps once a year, while you are producing quarterly hedonic price indices. The coefficients should only be estimated whenever you change the reference period, which is an attractive feature for Statistical agencies. In addition, the method is suitable for estimating deflators to the National accounts, since National accounts volumes are measured as Laspeyres volume indices[5]. This method is used in Statistics Norway.

[5] The value index is a product of a price index and a volume index: $V_{0:t} = P^{LA}_{0:t} \cdot Q^{PA}_{0:t} = P^{PA}_{0:t} \cdot Q^{LA}_{0:t} = P^{FI}_{0:t} \cdot Q^{FI}_{0:t}$

Time-dummy method:

So far we have used the characteristics approach, where the regression model in equation (2) is estimated every period, but there are other approaches. For example, if we concentrate on two successive quarters at a time and choose to ignore the variation in the parameters of the X's, that is, we assume that only the constant in (2) varies between the two successive quarters. The change in the constant β_0 can be estimated by supplementing (2) by a time-dummy D_t that is zero in quarter one and one in quarter two:

$$lnP_i = \beta_0 + \beta_1 \cdot X_{i1} + \beta_2 \cdot X_{i2} + \beta_3 \cdot X_{i3} + \beta_4 \cdot D_t + \varepsilon_i \qquad (2x)$$

Equation (2x) is estimated for a sample comprising the house sales in both quarters, and the estimated equation is:

$$lnP_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot X_{i1} + \hat{\beta}_2 \cdot X_{i2} + \hat{\beta}_3 \cdot X_{i3} + \hat{\beta}_4 \cdot D_t + u_i \qquad (3x)$$

With this equation, the constant in period one is $(\hat{\beta}_0 = \hat{\beta}_0 + \hat{\beta}_4 \cdot 0)$ and the constant in period two is $(\hat{\beta}_0 + \hat{\beta}_4 = \hat{\beta}_0 + \hat{\beta}_4 \cdot 1)$. The other coefficients are assumed to be equal in the two periods. The latter assumption represents a restriction on the characteristics method, which estimates all coefficients in both periods. As the constant is estimated freely in the two quarters, the average of u_i is zero in both quarters. This means that the mean price in both periods can be decomposed exact:

$$exp(\overline{lnP_1}) = exp(\hat{\beta}_0 + \hat{\beta}_1 \cdot \bar{X}_{11} + \hat{\beta}_2 \cdot \bar{X}_{21} + \hat{\beta}_3 \cdot \bar{X}_{31})$$

$$exp(\overline{lnP_2}) = exp(\hat{\beta}_0 + \hat{\beta}_4 + \hat{\beta}_1 \cdot \bar{X}_{12} + \hat{\beta}_2 \cdot \bar{X}_{22} + \hat{\beta}_3 \cdot \bar{X}_{32}) \qquad (4x)$$

It is obvious from equation (4x) that the three characteristics only contribute by their quantity to the development in the average price $exp(\overline{lnP_t})$ between quarter 1 and 2, because their prices $\hat{\beta}_1, \hat{\beta}_2 and\ \hat{\beta}_3$ are unchanged. And it turns out that the X's can be eliminated from the hedonic price indices, here illustrated by the hedonic Laspeyres price index[6]:

$$P_{1:2}^{LA} = \frac{exp(\hat{\beta}_0 + \hat{\beta}_4 + \hat{\beta}_1 \cdot \bar{X}_{11} + \hat{\beta}_2 \cdot \bar{X}_{21} + \hat{\beta}_3 \cdot \bar{X}_{31})}{exp(\hat{\beta}_0 + \hat{\beta}_1 \cdot \bar{X}_{11} + \hat{\beta}_2 \cdot \bar{X}_{21} + \hat{\beta}_3 \cdot \bar{X}_{31})}$$

$$= \frac{exp(\hat{\beta}_0) \cdot exp(\hat{\beta}_4) \cdot exp(\hat{\beta}_1 \cdot \bar{X}_{11}) \cdot exp(\hat{\beta}_2 \cdot \bar{X}_{21}) \cdot exp(\hat{\beta}_3 \cdot \bar{X}_{31})}{exp(\hat{\beta}_0) \cdot exp(\hat{\beta}_1 \cdot \bar{X}_{11}) \cdot exp(\hat{\beta}_2 \cdot \bar{X}_{21}) \cdot exp(\hat{\beta}_3 \cdot \bar{X}_{31})}$$

$$= exp(\hat{\beta}_4) \tag{6}$$

Thus, the quality-adjusted hedonic price index is determined by the exponential value of $\hat{\beta}_4$, which is the estimated coefficient for the time-dummy. Since the characteristics variables cancel out in equation (6), there is no difference between a hedonic Laspeyres-, Paasche- or Fisher price index using the time dummy-method. Basically, the time dummy method is a restricted version of the characteristics method, and if the restrictions are not really binding, the two methods are very close.

The coefficient of the time dummy $\hat{\beta}_4$, is used whether it is significant or not. The other estimated coefficients are not irrelevant. They may be used to assess the quality and credibility of the hedonic regression.

So far, we have assumed that the regression model is estimated for two subsequent quarters. The model may also be estimated for say four subsequent quarters at a time using three time-dummies instead of one. Including the two quarters $t-3$ and $t-2$ in the rolling sample would increase the number of observations, but it would also imply three estimates for the time-dummy coefficient implying that the resulting price index would be revised twice, see Triplett (2006).

Below are described some practical issues concerning implementing hedonic price index.

[6] We have used calculation rule: $\exp(a+b) = \exp(a) \cdot \exp(b)$

Handling of long reference periods:

The parameter estimates (coefficients) from the reference period are used as weights in the price index formula. Hence, in order to optimize these weights (coefficients), there should be no price development within the reference period.
However, using the characteristics method and the implicit method, the reference period typically consists of several periods (for example four quarters) and the mean sales price should therefore be revised, so it is not influenced by the underlying price development within the reference period[7]. The parameter estimates for the reference period can be adjusted in two ways. (1) You can comprise time dummy variables for n-1 of the quarters in the reference period with n quarters (i.e. three time dummy variables if the reference period is one year)[8] in order to estimate a constant for each quarter in the reference period. Adding n-1 time dummy variables to equation (2) gives us:

$$lnP_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot X_{i1} + \hat{\beta}_2 \cdot X_{i2} + \hat{\beta}_3 \cdot X_{i3} + \sum_{i=1}^{n-1} \hat{\beta}_i \cdot D_i + u_i \qquad (8)$$

The separate quarterly constants will reflect the price movement within the reference year. More specifically, we are not using the full right-hand side of the estimated relation in (8). Moving the time dummy term to the left-hand side and inserting the average values gives us a revised average price:

$$\overline{lnP} - \sum_{i=1}^{n-1} \hat{\beta}_i \cdot \bar{D}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot \bar{X}_1 + \hat{\beta}_2 \cdot \bar{X}_2 + \hat{\beta}_3 \cdot \bar{X}_3 \qquad (8x)$$

It is the right-hand side of (8x) that is used in the denominator in the price index formula, if the reference period consists of more than one quarter. This implies that we have to revise the simple relation between

[7] For the time dummy method, the reference period is the previous quarter and not the previous year, meaning that long reference periods is not an issue for the time dummy method.
[8] It is common to use the first quarter in the reference period as reference, meaning that there should be no time dummy variable for the first quarter if a constant is included in the regression.

the hedonic Laspeyres price and Paasche quantity indices and the average house price:

$$P_{o:t}^{LA} \cdot Q_{o:t}^{PA} = \frac{exp(\overline{lnP_t})}{exp(\overline{lnP_0})}$$

This relation was explained in (5c), but if using long reference periods it is replaced by:

$$P_{o:t}^{LA} \cdot Q_{o:t}^{PA} = \frac{exp(\hat{\beta}_{0t} + \hat{\beta}_{1t} \cdot \bar{X}_{10} + \hat{\beta}_{2t} \cdot \bar{X}_{20} + \hat{\beta}_{3t} \cdot \bar{X}_{30})}{exp(\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \bar{X}_{10} + \hat{\beta}_{20} \cdot \bar{X}_{20} + \hat{\beta}_{30} \cdot \bar{X}_{30})} \cdot \frac{exp(\hat{\beta}_{0t} + \hat{\beta}_{1t} \cdot \bar{X}_{1t} + \hat{\beta}_{2t} \cdot \bar{X}_{2t} + \hat{\beta}_{3t} \cdot \bar{X}_{3t})}{exp(\hat{\beta}_{0t} + \hat{\beta}_{1t} \cdot \bar{X}_{10} + \hat{\beta}_{2t} \cdot \bar{X}_{20} + \hat{\beta}_{3t} \cdot \bar{X}_{30})}$$

$$= \frac{exp(\hat{\beta}_{0t} + \hat{\beta}_{1t} \cdot \bar{X}_{1t} + \hat{\beta}_{2t} \cdot \bar{X}_{2t} + \hat{\beta}_{3t} \cdot \bar{X}_{3t})}{exp(\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \bar{X}_{10} + \hat{\beta}_{20} \cdot \bar{X}_{20} + \hat{\beta}_{30} \cdot \bar{X}_{30})} = \frac{exp(\overline{lnP_t})}{exp(\overline{lnP_0} - \sum_{i=1}^{n-1} \hat{\beta}_i \cdot \bar{D}_i)}$$

$$= \frac{exp(\overline{lnP_t})}{exp(\overline{lnP_0})/exp(\sum_{i=1}^{n-1} \hat{\beta}_i \cdot \bar{D}_i)} \qquad (5d)$$

where (8x) is applied to establish the last equals sign. The revision of (5c) amounts to dividing the average price $exp(\overline{lnP_0})$ in the reference year by $exp(\sum_{i=1}^{n-1} \hat{\beta}_i \cdot \bar{D}_i)$.

(2) Another approach is to estimate the reference equation for each of the quarters in the reference year[9]. In that case, there is no time dummy in equation and the reference price is calculated as an average of the regression results:

$$\hat{\hat{P}}_r = (\hat{P}_{1r} + \hat{P}_{2r} + \hat{P}_{3r} + \hat{P}_{4r})/4 \qquad (9)$$

, where

- $\hat{\hat{P}}_r$: Reference price, i.e. the average estimated sales price for the reference period
- $\hat{P}_{Qr}$: The estimated average sales price for quarter Q in the reference period.

[9] Statistics Netherlands (2013)

STATISTICS DENMARK

The difference between the two methods for calculating the reference price concerns the weights in the price index formula. If the quarters of the reference year should have equal weight, estimate the equation without time dummy for each quarter, i.e. four times, and use equation (9). If the quarters can have different weight, estimate and use an equation compromising time dummy variables.

Changing the reference period:

When calculating the hedonic price index in practice - using the characteristics method - you have to take changes in the reference year into consideration. For example replacing year 0 by year 1 as reference year in quarter s does not imply that we shift from index value $P_{o:s-1}^{LA}$ in quarter $s-1$ to index value $P_{1:s}^{LA}$ in quarter s and continue with index value $P_{1:s+1}^{LA}$ in quarter $s+1$, all calculated according to equation (5). The difference between $P_{o:s-1}^{LA}$ and $P_{1:s}^{LA}$ comprises the effect of replacing the $X's$ of year 0 by the $X's$ of year 1, and we do not want to include that effect in the resulting price index. That is, if all prices remain constant, the price index should stay constant as well. Consequently, we produce a chained index $I_t^{chained}$. We have year 0 as reference period for the first quarter in the calculation, and the chain index may be set equal to the hedonic index in equation (5) in a period going from quarter number 1 until quarter number s where the reference period changes:

$$I_t^{chained} = P_{o:t}^{LA} \quad for\ t = 1\ to\ s \qquad (10)$$

Starting in quarter number s+1, the formula for the chained index will reflect that the reference period is changed from year 0 to 1 in quarter number s. More specifically, the chained index will cumulate the changes in the hedonic Laspeyres formula $P_{1:t-1}^{LA}$:

$$I_t^{chained} = (P_{1:t}^{LA}/P_{1:t-1}^{LA}) \cdot I_{t-1}^{chained} \quad for\ t = s+1\ to\ next\ change\ of\ ref. \qquad (11)$$

STATISTICS DENMARK

Notice that for the time dummy method, the published hedonic price index is calculated as a chained index $I_t^{chained}$ cumulating the index that describe the quality-adjusted price change from period $t - 1$ to t:

$$I_t^{chained} = exp(\hat{\beta}_{4t}) \cdot I_{t-1}^{chained} \qquad \text{(time dummy method)} \qquad (12)$$

With this equation, the resulting price index $I_t^{chained}$ can be calculated for the entire period from quarter 1 to quarter t. Only the coefficient of the time dummy variable is used to calculate the resulting price index.

Introducing of new explanatory variables:

Concerning the practical implementation of $P_{o:t}^{LA}$ from equation (5), we note that new characteristics introduced in period t, for example a new geographical breakdown, can only be introduced in the price index when the new data covers a reference period.

Growth contribution from individual observations to the overall price index:

As in a traditional price index, it is possible to calculate the growth contribution from individual observations to the overall price index. For example if you have isolated some outliers in the regression analysis, you can run the regression with and without these outliers and then calculate the price index for both scenarios. The difference between the two price indices is the growth contribution from the outliers, which you can consider to exclude from the regression.

Growth contribution from each explanatory variable to the overall price index:

It is also possible to decompose the price index to get the formal contribution from the price development on each explanatory characteristic. To decompose we re-write the index formula and to simplify, there is only one characteristic:

$$P^{LA}_{0:t} = \frac{\exp(\hat{\beta}_{0t} + \hat{\beta}_{1t} \cdot \bar{X}_{10})}{\exp(\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \bar{X}_{10})} \tag{13}$$

Taking logs gives us:

$$\ln P^{LA}_{0:t} = \hat{\beta}_{0t} + \hat{\beta}_{1t} \cdot \bar{X}_{10} - (\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \bar{X}_{10}) = (\hat{\beta}_{0t} - \hat{\beta}_{00}) + (\hat{\beta}_{1t} - \hat{\beta}_{10}) \cdot \bar{X}_{10} \tag{14}$$

The left hand side of equation (14) is the log of the hedonic Laspeyres price index and the right hand side shows its decomposition into the change of the constant between period 0 and t and into the price change of the characteristic.

However, there is a caveat. The formal decomposition does not necessarily show the contributions to the hedonic price index, because β_0 and β_1 are correlated. In other words, equation (14) may just as well be normalized on β_0, which would indicate the contribution from the price of the characteristic to the "unspecified" price development captured by the constant. The caveat is, for example, particularly evident if the characteristic in (14) is a 1/0 dummy variable indicating by 1 that the house is outside the Copenhagen area.

A constant coefficient of say -0.2 for houses outside the Copenhagen region reflects that houses are 18,1 per cent (=exp(-0.2)) lower outside Copenhagen. And a change from -0.2 to -0.1 over a number of years reflects that the price differential between Copenhagen and the rest of the country has fallen, so that houses outside Copenhagen are only 9.5 per cent cheaper. This change in relative prices is of interest but it does not per se imply anything about the national house price. The price contribution of houses in Copenhagen is an unidentified part of the constant.

Confidence intervals for hedonic price indices:

In a regression analysis you estimate some parameters. Each of these parameter estimates is equal to the true (unknown) parameter estimate plus a noise term:

$$\hat{\beta} = \beta + \varepsilon \quad (15)$$

Since the estimated parameters are subject to noise, it may be appropriate to calculate a confidence interval for the hedonic price index. This can be done using the following formula[10]:

$$Confidence\ interval_t = exp\left(lnI_t \pm 1{,}96 \cdot \sqrt{Var(lnI_{0:t})}\right) \quad (16)$$

, where

- Exp: Natural exponential function
- Ln: Natural logarithm of the chain price index in period t
- $\mathrm{Var}(\mathrm{lnI}_{0:t})$: The variance of the logarithm of the price development between the reference period and the current period.

Confidence intervals for the characteristics method[11]:

Let us assume that we have a hedonic Laspeyres price index, which is calculated using the characteristics method (here shown with two explanatory variables for simplicity):

$$P_{0:t}^{LA} = \frac{exp\left(\hat{\beta}_{0t} + \hat{\beta}_{1t} \cdot \bar{X}_{10} + \hat{\beta}_{2t} \cdot \bar{X}_{20}\right)}{exp\left(\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \bar{X}_{10} + \hat{\beta}_{20} \cdot \bar{X}_{20}\right)} \quad (17)$$

We consider the numerator and denominator as two stochastic independent variables, where the parameter estimates are stochastic

[10] Another approach is bootstrapping, where the price index for a given period is produced for a given number of subsamples, given a lower- and upper bound.

[11] Bootstrapping is another approach which we find give similar results.

variables and the explanatory variables are constant. To avoid dividing by stochastic variables, we take the logarithm to the price index:

$$lnP_{o:t}^{LA} = \left(\hat{\beta}_{0t} + \hat{\beta}_{1t} \cdot \bar{X}_{10} + \hat{\beta}_{2t} \cdot \bar{X}_{20}\right) - \left(\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \bar{X}_{10} + \hat{\beta}_{20} \cdot \bar{X}_{20}\right) \qquad (18)$$

The variance for the difference between two stochastic independent variables is given by the following rule:

$$Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y) \qquad (19)$$

We will assume that the covariance is zero between the parameter estimates in different periods, which seems a plausible assumption. Thus, equation (19) can be reduced to:

$$Var(X - Y) = Var(X) + Var(Y) \qquad (20)$$

Keeping equation (20) in mind, the variance of the first term in equation (18) is[12]:

$$Var\left(\hat{\beta}_{0t} + \hat{\beta}_{1t} \cdot \bar{X}_{10} + \hat{\beta}_{2t} \cdot \bar{X}_{20}\right) =$$

$$[1 \quad \bar{X}_{10} \quad \bar{X}_{20}] \cdot \begin{bmatrix} var(\hat{\beta}_{0t}) & cov(\hat{\beta}_{0t}, \hat{\beta}_{1t}) & cov(\hat{\beta}_{0t}, \hat{\beta}_{2t}) \\ cov(\hat{\beta}_{1t}, \hat{\beta}_{0t}) & var(\hat{\beta}_{1t}) & cov(\hat{\beta}_{1t}, \hat{\beta}_{2t}) \\ cov(\hat{\beta}_{2t}, \hat{\beta}_{0t}) & cov(\hat{\beta}_{2t}, \hat{\beta}_{1t}) & var(\hat{\beta}_{2t}) \end{bmatrix} \cdot \begin{bmatrix} 1 \\ \bar{X}_{10} \\ \bar{X}_{20} \end{bmatrix} \qquad (21)$$

The variance of the second term of equation (18) is:

$$Var\left(\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \bar{X}_{10} + \hat{\beta}_{20} \cdot \bar{X}_{20}\right) =$$

[12] $Var(a \cdot X) = a^2 \cdot Var(X)$ and $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$

$$[1 \quad \bar{X}_{10} \quad \bar{X}_{20}] \cdot \begin{bmatrix} var(\hat{\beta}_{00}) & cov(\hat{\beta}_{00}, \hat{\beta}_{10}) & cov(\hat{\beta}_{00}, \hat{\beta}_{20}) \\ cov(\hat{\beta}_{10}, \hat{\beta}_{00}) & var(\hat{\beta}_{10}) & cov(\hat{\beta}_{10}, \hat{\beta}_{20}) \\ cov(\hat{\beta}_{20}, \hat{\beta}_{00}) & cov(\hat{\beta}_{20}, \hat{\beta}_{10}) & var(\hat{\beta}_{20}) \end{bmatrix} \cdot \begin{bmatrix} 1 \\ \bar{X}_{10} \\ \bar{X}_{20} \end{bmatrix} \quad (22)$$

Thus, the variance of the logarithm to the price index is:

$$var(lnP_{o:t}^{LA}) = var(\hat{\beta}_{0t} + \hat{\beta}_{1t} \cdot \bar{X}_{10} + \hat{\beta}_{2t} \cdot \bar{X}_{20})$$
$$+var(\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \bar{X}_{10} + \hat{\beta}_{0t} \cdot \bar{X}_{20}) =$$

$$[1 \quad \bar{X}_{10} \quad \bar{X}_{20}] \cdot \begin{bmatrix} var(\hat{\beta}_{0t}) & cov(\hat{\beta}_{0t}, \hat{\beta}_{1t}) & cov(\hat{\beta}_{0t}, \hat{\beta}_{2t}) \\ cov(\hat{\beta}_{1t}, \hat{\beta}_{0t}) & var(\hat{\beta}_{1t}) & cov(\hat{\beta}_{1t}, \hat{\beta}_{2t}) \\ cov(\hat{\beta}_{2t}, \hat{\beta}_{0t}) & cov(\hat{\beta}_{2t}, \hat{\beta}_{1t}) & var(\hat{\beta}_{2t}) \end{bmatrix} \cdot \begin{bmatrix} 1 \\ \bar{X}_{10} \\ \bar{X}_{20} \end{bmatrix} +$$

$$[1 \quad \bar{X}_{10} \quad \bar{X}_{20}] \cdot \begin{bmatrix} var(\hat{\beta}_{00}) & cov(\hat{\beta}_{00}, \hat{\beta}_{10}) & cov(\hat{\beta}_{00}, \hat{\beta}_{20}) \\ cov(\hat{\beta}_{10}, \hat{\beta}_{00}) & var(\hat{\beta}_{10}) & cov(\hat{\beta}_{10}, \hat{\beta}_{20}) \\ cov(\hat{\beta}_{20}, \hat{\beta}_{00}) & cov(\hat{\beta}_{20}, \hat{\beta}_{10}) & var(\hat{\beta}_{20}) \end{bmatrix} \cdot \begin{bmatrix} 1 \\ \bar{X}_{10} \\ \bar{X}_{20} \end{bmatrix} \quad (23)$$

This method can also be used for hedonic Paasche price index, which is equivalent to the implicit price index method.

Confidence intervals for the time dummy method:

For time dummy method it is much simpler to calculate the variance of the price index, as the only estimated parameter is the coefficient for the time dummy variable. Thus the confidence interval becomes:

$$Confidence\ interval_t = I_t \pm 1{,}96 \cdot \sqrt{Var(\hat{\beta}_4)} \quad (24)$$

Sales of real property as case study:

In order to compare the three types of direct hedonic price indices with more traditional price index methods, we will use sales of real property (one family houses) as a case study, which is based on the value weighted SPAR (Sales Price Appraisal Ratio) method. The value weighted SPAR method calculate the average sales price divided by the average appraisal value og simply the sum of the sales prices divided by the sum of the appraisals. In this way the appraisals are used to quality adjustment the prices and thus the price index. In addition, we have also calculated a volume index based on the sales prices without any adjustments. Then we can see how much the hedonic price indices and the SPAR index adjust.

The following regression has been used for the different hedonic price indices:

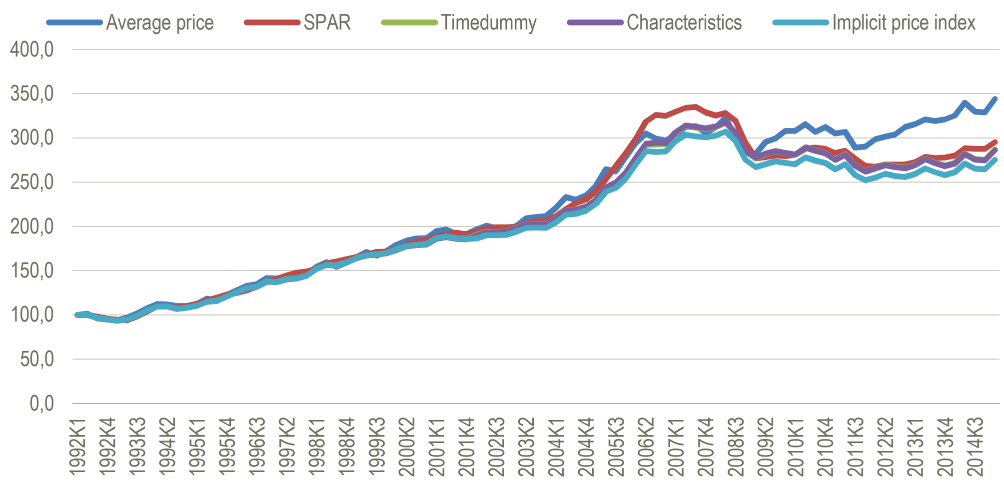
$$\ln\_price = \ln\_ua \text{ zone2 zone3 zone4 zone5 zone6 zone7 zone8 zone9 zone10 zone11 cy2 cy3 cy4 timedummy2 timedummy3 timedummy4} \quad (25)$$

, where

- ln_price: The natural logarithm to the sales price of the house
- ln_ua: The natural logarithm to the usable area
- zone1: Province Byen København (reference variable)
- zone2: Province Københavns omegn
- zone3: Province Nordsjælland
- zone4: Province Bornholm
- zone5: Province Østsjælland
- zone6: Province Vest- and Sydsjælland
- zone7: Province Fyn
- zone8: Province Sydjylland
- zone9: Province Østjylland
- zone10: Province Vestjylland
- zone11: Province Nordjylland

- cy1: Construction year < 10 (reference variable)
- cy2: Construction year between 10-19
- cy3: Construction year between 20-34
- cy4: Construction year > 35
- timedummy1: Quarter 1 in the reference period (ref. variable)
- timedummy2: Quarter 2 in the reference period
- timedummy3: Quarter 3 in the reference period
- timedummy4: Quarter 4 in the reference period

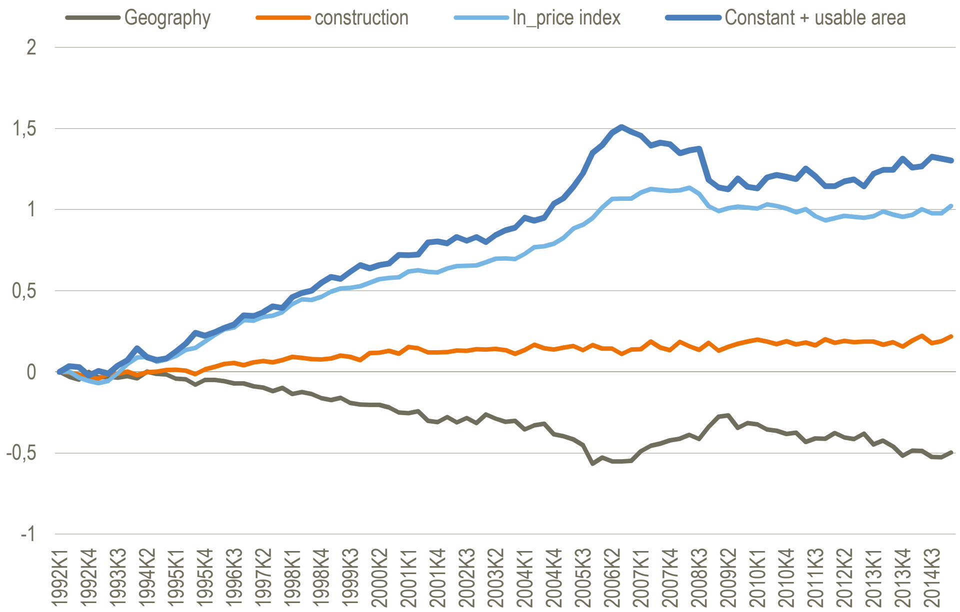
The results are shown in the figure below:



The figure shows that the hedonic Laspeyres price index (characteristics method) and the time dummy method are very similar. The price indices are also shown in the annex. The Laspeyres price index has a tendency to overestimate the price development as the weights are derived from the reference period. The Paasche price index (implicit price index) has a tendency to underestimate the price development as the weights are derived from the actual period.

The gap between the value index (average price) and the quality adjusted hedonic price indices reflects two things: (1) The value index are described on an arithmetic average, while the hedonic price indices are based on a geometric means, as we use the logarithm of the sales price as response variable. (2) The hedonic price indices quality adjusts the sales price, which is ignored in the value index.

As illustrated by equation (14) we can find the price contribution from ach characteristic. For each of the three main groups of characteristics, we show the contribution to the hedonic Laspeyres price index in the figure below:



There is a lot of volatility in the price contribution of usable area. The increasing trend in the contribution shows that the price or market value of space is increasing. More specifically, the beta for (the log of) usable area in square meters has increased, from around 0.5 in the beginning of the nineties to around 0.8. Notice that a beta of 1 implies that the price tends to be proportional with the size of the usable area.

Conclusion:

Hedonic regression is the most general method to construct quality adjusted house price indices. Many other methods (SPAR-method, Repeat sales method, Stratification method) can be seen as special cases of hedonic regression. As an example, the SPAR-method is formulated as a restricted hedonic regression.

The most general hedonic regression method is the characteristics method. In principle, the other hedonic approaches (time-dummy method, imputation method and re-pricing method) can be seen as special cases of the characteristics method. For instance, the re-pricing method is quite similar to the characteristics method since the two methods will produce the same volume change and consequently the same price change when applied to the same data set.

In general, Eurostat recommends the imputation approach for house price indices. However, houses are heterogeneous goods and it is difficult to find a lot of price match between houses sold in two consecutive quarters. Without price match, the imputation method and the characteristics method will be equivalent in practice. The time dummy method is the simplest method, but it does not allow decomposing the price development. In addition, the time-dummy method does not distinguish between Laspeyres, Paasche or Fisher indices while all three hedonic price indices can be calculated with the characteristics method.

The hedonic Fisher price index is preferred since it is a superlative index. However, using the characteristics method to calculate hedonic Paasche price index, it is not required to run the regression for the current period, which is a main advantage from a Statistical agency point of view. Thus, it is relatively easy to produce the hedonic Paasche price index, and it can be sufficient to calculate the Paasche price index if it is close to the Laspeyres and Fisher index in the historical period.

Annex 1: Illustrative examples of hedonic price index methods

We will here use the same data source as presented in the CENEX handbook for the Harmonised Consumer Price Index (HICP) and compute hedonic price index using all of the four hedonic methods mentioned in this paper[13].

Consider a product group with a sample of 6 models. For times t=1 and t=2, prices for each of the 6 models have been collected in the following table:

t=1				t=2			
P	**X_1**	**X_2**	**Model**	**P**	**X_1**	**X_2**	**Model**
390	23	0	1	290	23	0	1
480	39	0	2	519	39	0	2
700	51	1	3	700	51	1	3
550	39	0	4	550	39	0	4
520	35	1	5	520	35	1	5
490	43	0	6	698	53	1	6
Average:				**Average:**			
513,6*	38,33	0,33		525,4*	40	0,5	

*) Geometric average

There are two characteristics (X_1 and X_2) for each of the models. The variable P denotes the price of each model. Notice that models 1-5 does not change characteristics during the two periods. Only model 6 change characteristics.

It should be stressed that in practice one should never use such small data source as the precision of the regression coefficients then becomes unacceptably poor. The example presented here is only meant as an "illustrative example" for demonstration of the computation technique. However, you can imagine that the estimated coefficients are computed from some larger data source.

[13] Handbook of the application of quality adjustment methods in the Harmonised Index of Consumer Prices. Developed within the project "CENEX HICP Quality Adjustment". Volume 13. Federal Statistical Office of Germany, page 147.

Characteristics method:

In the characteristic method a regression is made for each of the two periods. These yields:

$\overline{\ln P_1} = 5{,}604 + 0{,}0155 \cdot \overline{X}_{11} + 0{,}1331 \cdot \overline{X}_{21}$ (period 1) (a.1.1)

$\overline{\ln P_2} = 5{,}168 + 0{,}0270 \cdot \overline{X}_{12} + 0{,}0317 \cdot \overline{X}_{22}$ (period 2) (a.1.2)

Where the bar over a variable indicates the average value for the six products and the subscript indicates the period. Since the residuals sums to zero in an OLS-regression with a constant the two equations holds pr. construction.

By inserting these coefficients and the average values of the characteristics we can compute a hedonic Laspeyres price index and a hedonic Paasche price index. The Fisher price index is just the geometric mean of these two price indices. The results are shown below:

$$P_{1:2}^{LA} = \frac{exp(5{,}168 + 0{,}0270 \cdot 38{,}33 + 0{,}0317 \cdot 0{,}33)}{exp(5{,}604 + 0{,}0155 \cdot 38{,}33 + 0{,}1331 \cdot 0{,}33)} \cdot 100 = 97{,}27 \qquad (a.1.3)$$

$$P_{1:2}^{PA} = \frac{exp(5{,}168 + 0{,}0270 \cdot 40 + 0{,}0317 \cdot 0{,}5)}{exp(5{,}604 + 0{,}0155 \cdot 40 + 0{,}1331 \cdot 0{,}5)} \cdot 100 = 97{,}50 \qquad (a.1.4)$$

$$P_{1:2}^{FI} = (97{,}27 \cdot 97{,}50)^{0{,}5} = 97{,}38 \qquad (a.1.5)$$

For Residential Property Price Index Eurostat recommend the Fisher price index to eliminate the bias in the Laspeyres (overstimating) and Paasche (underestimating) price index. The Fisher price index is a superlative index.

We can also use the characteristic method to calculate the price index *implicitly*. For example, if we want to calculate a hedonice Paasche price index, as in equation (a.1.4), we can do this by dividing the value index, which consists solely of the geometric mean of the sales prices with the hedonic Laspeyres quantity (volume) index. We get that:

$$V_{1:2} = \left(\frac{390 \cdot 480 \cdot 700 \cdot 550 \cdot 520 \cdot 490}{290 \cdot 519 \cdot 700 \cdot 550 \cdot 520 \cdot 698}\right)^{1/6} \cdot 100 = 102{,}29 \qquad (a.1.6)$$

$$Q_{1:2}^{LA} = \frac{exp(5{,}604 + 0{,}0155 \cdot 40 + 0{,}1331 \cdot 0{,}5)}{exp(5{,}604 + 0{,}0155 \cdot 38{,}33 + 0{,}1331 \cdot 0{,}33)} \cdot 100 = 104{,}91 \qquad (a.1.7)$$

Notice that the coefficients are fixed in the quantity index in (a.1.7).

We can now compute the hedonic Paasche price index as:

$$P_{1:2}^{PA} = \frac{102,29}{104,91} \cdot 100 = 97,50 \qquad \text{(a.1.8)}$$

The result is identical with the explicitly Paasche price index in equation (a.1.4). The idea of calculating quantity index and thus only re-estimating the coefficients with a lower frequency, for example once a year, even though producing quarterly price index is also used in the re-pricing method. See example later on.

The time-dummy method:

In the time-dummy method the regression is based on the pooled data set consisting of both period 1 and 2, i.e. 12 observations. The result is shown below:

$$\overline{\text{lnP}} = 5,3918 + 0,0213 \cdot \bar{X}_1 + 0,0984 \cdot \bar{X}_2 - 0,0293 \cdot \bar{D}_1 \qquad \text{(a.1.9)}$$

D_1 is the time-dummy variable, which is 0 in period 1 and 1 in period 2. Thus, the time-dummy method is equivalent to calculating a constant for each period:

Constant in period 1: $5,3918 + (-0,0293 \cdot 0) = 5,3918$ (a.1.10)

Constant in period 2: $5,3918 + (-0,0293 \cdot 1) = 5,3624$ (a.1.11)

The price index from period 1 to 2 then yields:

$$P_{1:2} = \frac{exp(5,3624)}{\exp(5,3918)} \cdot 100 = exp(-0,0293) \cdot 100 = 97,11 \qquad \text{(a.1.12)}$$

Since the time dummy model assumes that the two characteristic variables have the same price (estimate) in period 1 and 2, there is per. construction coincidence between the hedonic Laspeyres and Paasche price indices:

$$P_{1:2}^{LA} = \frac{exp(5,3624 + 0,0213 \cdot 38,33 + 0,0984 \cdot 0,33)}{exp(5,3918 + 0,0213 \cdot 38,33 + 0,0984 \cdot 0,33)} \cdot 100 = 97,11 \qquad (a.1.13)$$

$$P_{1:2}^{PA} = \frac{exp(5,3624 + 0,0213 \cdot 40 + 0,0984 \cdot 0,5)}{exp(5,3918 + 0,0213 \cdot 40 + 0,0984 \cdot 0,5)} \cdot 100 = 97,11 \qquad (a.1.14)$$

The Fisher price index is pr. definition also 97,11. There is also coincidence between the Laspeyres and Paasche quantity index.

Imputation method:

In the table above, the characteristics for model 1-5 is unchanged, but you should use a hedonic method to calculate the price development for model 6.

For model 6 the prices for the two periods does not reflect the pure price change, since the characteristics have also changed. For the imputation method the starting point is model 6 in period 1, for instance and then impute the price for this model in period 2.

For the imputation one can for example use the estimated prices from the characteristics method. By inserting model 6's characteristic from period 1 in the estimated equation for period 2, the imputed price for period 2 is 560,77:

$$exp(5{,}168 + 0{,}0270 \cdot 43 + 0{,}0317 \cdot 0) = 560{,}77 \quad \text{(period 2)} \qquad \text{(a.1.15)}$$

If we calculate an equally weighted geometric basis index, we get:

$$P_{1:2}^{LA} = \left(\frac{290 \cdot 519 \cdot 700 \cdot 550 \cdot 520 \cdot 560{,}77}{390 \cdot 480 \cdot 700 \cdot 550 \cdot 520 \cdot 490}\right)^{\frac{1}{6}} \cdot 100 = 98{,}62 \qquad (a.1.16)$$

This index is higher than the price index based on the characteristics method. The difference reflects that the observable price of 490 for model 6 in period 1 is lower than the estimated price resulting from using the characteristics methods equation (a.1.1) for the characteristics in period 1:

$$exp(5{,}604 + 0{,}0155 \cdot 43 + 0{,}1331 \cdot 0) = 528.12 \quad \text{(period 1)} \qquad \text{(a.1.17)}$$

If we also use the imputed price of model 6 in period 1, corresponding to a double imputation instead of single imputation, the basic index become slightly lower:

$$P_{1:2}^{LA} = \left(\frac{290 \cdot 519 \cdot 700 \cdot 550 \cdot 520 \cdot 560{,}77}{390 \cdot 480 \cdot 700 \cdot 550 \cdot 520 \cdot 528.12}\right)^{\frac{1}{6}} \cdot 100 = 97{,}40 \qquad (a.1.18)$$

Since the characteristics of the two imputed prices of model 6 are derived from period 1, this is a hedonic Laspeyres price index. The 97,40 is also similar to the Laspeyres price index using the characteristic method, however it is not exactly the same. This reflects that the pure price development of the 5 models with price match can differ from the price development using the characteristic method of these 5 models.

If the characteristics of the two imputed prices of model 6 instead are derived from period 2, we get the following prices:

$$exp(5{,}168 + 0{,}0270 \cdot 53 + 0{,}0317 \cdot 1) = 758{,}32 \quad \text{(period 2)} \qquad \text{(a.1.19)}$$

$$exp(5.604 + 0.0155 \cdot 53 + 0.1331 \cdot 1) = 704{,}25 \quad \text{(period 1)} \qquad \text{(a.1.20)}$$

The basis index then becomes:

$$P_{1:2}^{PA} = \left(\frac{290 \cdot 519 \cdot 700 \cdot 550 \cdot 520 \cdot 758{,}32}{390 \cdot 480 \cdot 700 \cdot 550 \cdot 520 \cdot 704{,}25}\right)^{\frac{1}{6}} \cdot 100 = 97{,}63 \qquad (a.1.21)$$

This is a hedonic Paasche price index since the characteristics are derived from the current period.

Notice that if the characteristics had changed for all 6 models in the sample (i.e. no price match at all), the imputation method is equivalent to the characteristics method. That is, the imputation method becomes a direct method. Eurostat recommend the imputation approach for house price index. However since houses (dwellings) sold in different periods are heterogeneous (no price match from quarter to quarter), the imputation method (single and double imputation) and characteristics method will be equivalent in practice. See examples in annex 2.

Re-pricing method[14]:

In this method the development in the volume (quantity) of model 6 is adjusted. For example, one can use the estimated prices from the characteristics method to calculate the Paasche quantity index from period 1 to 2 based on the parameter estimates from period 2:

$$Q_{model\,6}^{PA} = \frac{exp(5{,}168 + 0{,}0270 \cdot 53 + 0{,}0317 \cdot 1)}{exp(5{,}168 + 0{,}0270 \cdot 43 + 0{,}0317 \cdot 0)} \cdot 100 = 135{,}23 \qquad (a.1.22)$$

According to this estimate, a model 6 in period 2 corresponds to 1,3523 model 6 in period 1. If the price of model 6 in period 2, i.e. 698, is corrected for this difference, we get the following Laspeyres price index for the 6 models:

[14] This method is illustrated in the CENEX Handbook, page 147.

$$P_{1:2}^{LA} = \left(\frac{290 \cdot 519 \cdot 700 \cdot 550 \cdot 520 \cdot 698}{390 \cdot 480 \cdot 700 \cdot 550 \cdot 520 \cdot 490 \cdot 1,3523}\right)^{\frac{1}{6}} \cdot 100 = 97,27 \qquad (a.1.23)$$

The result of 97,27 is identical to the hedonic Laspeyres Price index derived from the characteristics method and the coincidence is not accidentally.

According to the characteristics method, it is only model 6 that contribute to the volume change in the characteristics from period 1 to 2. The characteristics method indicate a non-change in the volume for model 1-5 with price match, since both the parameter estimates and characteristics are unchanged. This can easily be seen by calculating a volume index for model 1, for instance:

$$Q_{model\ 1}^{PA} = \frac{exp(5,168 + 0,0270 \cdot 23 + 0,0317 \cdot 0)}{exp(5,168 + 0,0270 \cdot 23 + 0,0317 \cdot 0)} \cdot 100 = 100 \qquad (a.1.24)$$

We can also use the parameter estimates from period 1 to calculate the Laspeyres volume index for model 6 with no price match. We then get a slightly different volume index than in (a.1.22):

$$Q_{model\ 6}^{LA} = \frac{exp(5,604 + 0,0155 \cdot 53 + 0,1331 \cdot 1)}{exp(5,604 + 0,0155 \cdot 43 + 0,1331 \cdot 0)} \cdot 100 = 133,35 \qquad (a.1.25)$$

If we adjust the pure price of model 6 with 33,35 per cent, we obtain the following base index for the 6 models:

$$P_{1:2}^{PA} = \left(\frac{290 \cdot 519 \cdot 700 \cdot 550 \cdot 520 \cdot 698}{390 \cdot 480 \cdot 700 \cdot 550 \cdot 520 \cdot 490 \cdot 1,3335}\right)^{\frac{1}{6}} \cdot 100 = 97,50 \qquad (a.1.26)$$

The result of 97,50 is identical to the hedonic Paasche price index derived from the characteristics method.

Results:

The table below shows an overview over the results obtained above:

Method	Laspeyres	Paasche	Fisher
Characteristics	97,27	97,50	97,38
Time-dummy	97,11	97,11	97,11
Imputation (single)	98,62	96,29	97,45
Imputation (double)	97,40	97,63	97,51
Re-pricing	97,27	97,50	97,38

The characteristics method and the re-pricing method gets the same result, because they agree that the volume change for model 1-5 with price match is zero. Therefore it is sufficient to adjust the volume for model 6. The time-dummy method has a slightly different result than the characteristics method, since the parameter estimates are not quite constant in the two periods. The imputation method also gets a slightly different result than the characteristics method. Here the explanation is that the methods only agree on the volume change for model 6 with no price match. For the 5 models with price match the methods disagree on the volume change. However, had the entire product sample consisted solely of models without price match, the imputation method would be a direct method and be equivalent to the characteristics method. This will typically be the case for dwellings, since different dwellings are sold in different periods – see annex 2.

STATISTICS DENMARK

Annex 2: Illustrative examples of hedonic price index methods

The dataset applied in annex 1 will typically not be relevant for houses (dwellings), since different houses are sold in different periods and the number of transactions in period 1 and period 2 is not likely to be the same. In this annex we therefore use a more relevant dataset for houses with no price match in the two periods. The number of house sales in the two periods is also different. The imputation method now becomes a direct method imputing prices for all house sales in period 2, and the result of the imputation method is equivalent to the outcome of the characteristics method applied to the same dataset. To see this, consider the following table with 6 houses sold in period 1 (t=1) and 5 houses sold in period 2 (t=2):

t=1			t=2		
P	**X_1**	**X_2**	**P**	**X_1**	**X_2**
390	25	1	290	23	0
480	38	1	519	39	0
700	42	1	700	51	1
550	25	1	550	39	0
520	41	0	520	35	1
490	43	0	-	-	-
Average:			**Average:**		
513,6*	35,67	0,67	496,4*	37,4	0,4

*) Geometric average

There are two characteristics (X_1 and X_2) for each house model. The variable P denotes the price of each model. Notice that in contrast to the dataset applied in annex 1 there is no price match since the characteristics differ between the two periods. There is no correspondence between house models in period 1 and house models in period 2, and the number of observations is also different in the two periods.

As in annex 1, it should be stressed that the data set in the example is too small to be used in practice where more degrees of freedom are needed to determine the regression coefficients with adequate certainty. The example is only an "illustrative example" for demonstrating the computation technique.

Characteristics method:

With the characteristic method, a regression is made for each of the two periods:

$$\overline{\ln P_1} = 5{,}498 + 0{,}0173 \cdot \bar{X}_{11} + 0{,}1903 \cdot \bar{X}_{21} \quad \text{(period 1)} \qquad \text{(a.2.1)}$$

$$\overline{\ln P_2} = 5{,}074 + 0{,}0298 \cdot \bar{X}_{12} + 0{,}0472 \cdot \bar{X}_{22} \quad \text{(period 2)} \qquad \text{(a.2.2)}$$

Where a variable with a bar indicates the period average and the subscript indicates period and characteristic. There are two periods and two characteristics. Since the residuals sums to zero in an OLS-regression with a constant, the two equations hold pr. construction.

By inserting the estimated coefficients and the average values of the characteristics we can compute a hedonic Laspeyres price index and a hedonic Paasche price index and the relating Fisher price index, which is the geometric mean. The results are shown below:

$$P_{1:2}^{LA} = \frac{exp(5{,}074 + 0{,}0298 \cdot 35{,}67 + 0{,}0472 \cdot 0{,}67)}{exp(5{,}498 + 0{,}0173 \cdot 35{,}67 + 0{,}1903 \cdot 0{,}67)} \cdot 100 = 92{,}94 \qquad (a.2.3)$$

$$P_{1:2}^{PA} = \frac{exp(5{,}074 + 0{,}0298 \cdot 37{,}4 + 0{,}0472 \cdot 0{,}4)}{exp(5{,}498 + 0{,}0173 \cdot 37{,}4 + 0{,}1903 \cdot 0{,}4)} \cdot 100 = 98{,}67 \qquad (a.2.4)$$

$$P_{1:2}^{FI} = (92{,}94 \cdot 98{,}67)^{0{,}5} = 95{,}76 \qquad \text{(a.2.5)}$$

We can also use the characteristic method to calculate the price index *implicitly*. For example, if we want to calculate a hedonice Paasche price index, as in equation (a.2.4), we can do this by calculating the value index, which represents the geometric mean of the, respectively, five and six sales prices and divied by the hedonic Laspeyres quantity (volume) index. We get:

$$V_{1:2} = \frac{(290 \cdot 519 \cdot 700 \cdot 550 \cdot 520)^{1/5}}{(390 \cdot 480 \cdot 700 \cdot 550 \cdot 520 \cdot 490)^{1/6}} \cdot 100 = 96{,}64 \qquad (a.2.6)$$

$$Q_{1:2}^{LA} = \frac{exp(5{,}498 + 0{,}0173 \cdot 37{,}4 + 0{,}1903 \cdot 0{,}4)}{exp(5{,}498 + 0{,}0173 \cdot 35{,}67 + 0{,}1903 \cdot 0{,}67)} \cdot 100 = 97{,}94 \qquad (a.2.7)$$

Notice that the coefficients are fixed in the quantity index in (a.2.7).

We can now compute the hedonic Paasche price index as:

$$P_{1:2}^{PA} = \frac{96{,}64}{97{,}94} \cdot 100 = 98{,}67 \qquad \text{(a.2.8)}$$

The result is identical with the explicitly calculated Paasche price index in equation (a.2.4).

The time-dummy method:

In the time-dummy method the regression is based on a pooled data set consisting of both period 1 and 2, i.e. 11 observations. The result is shown below:

$$\overline{\ln P} = 5{,}3439 + 0{,}0218 \cdot \overline{X}_1 + 0{,}1804 \cdot \overline{X}_2 - 0{,}0239 \cdot \overline{D}_1 \qquad (a.2.9)$$

D_1 is the time-dummy variable, which is 0 in period 1 and 1 in period 2. The time-dummy method is equivalent to calculating a constant for each period:

Constant in period 1: $5{,}3439 + (-0{,}0239 \cdot 0) = 5{,}3439$ (a.2.10)

Constant in period 2: $5{,}3439 + (-0{,}0239 \cdot 1) = 5{,}3200$ (a.2.11)

The price index from period 1 to 2 then yields:

$$P_{1:2} = \frac{exp(5{,}3200)}{\exp(5{,}3439)} \cdot 100 = exp(-0{,}0239) \cdot 100 = 97{,}64 \qquad (a.2.12)$$

Since the time dummy model assumes that the two characteristic variables have the same price in period 1 and 2, there is per. construction coincidence between the hedonic Laspeyres and Paasche price indices:

$$P_{1:2}^{LA} = \frac{exp(5{,}3200 + 0{,}0218 \cdot 35{,}67 + 0{,}1804 \cdot 0{,}67)}{exp(5{,}3439 + 0{,}0218 \cdot 35{,}67 + 0{,}1804 \cdot 0{,}67)} \cdot 100 = 97{,}64 \qquad (a.2.13)$$

$$P_{1:2}^{PA} = \frac{exp(5{,}3200 + 0{,}0218 \cdot 37{,}4 + 0{,}1804 \cdot 0{,}4)}{exp(5{,}3439 + 0{,}0218 \cdot 37{,}4 + 0{,}1804 \cdot 0{,}4)} \cdot 100 = 97{,}64 \qquad (a.2.14)$$

The Fisher price index is pr. definition also 97,64, and the Laspeyres and Paasche quantity index coincide.

STATISTICS DENMARK

Imputation method:

For the imputation method one can use the estimated prices from the characteristics method. By inserting the characteristics from period 1 in the estimated equation for period 2, you get the imputed price for period 2. For example for the first observation in period 1, the imputed price for period 2 is:

$$exp(5{,}074 + 0{,}0298 \cdot 25 + 0{,}0472 \cdot 1) = 353{,}0 \quad \text{(period 2)} \qquad \text{(a.2.15)}$$

If we impute period 2 prices for all six observations in period 1 and calculate an equally weighted geometric basis index, we get:

$$P_{1:2}^{LA} = \left(\frac{353{,}0 \cdot 519{,}8 \cdot 585{,}6 \cdot 353{,}0 \cdot 542{,}2 \cdot 575{,}5}{390 \cdot 480 \cdot 700 \cdot 550 \cdot 520 \cdot 490}\right)^{\frac{1}{6}} \cdot 100 = 92{,}94 \qquad (a.2.16)$$

This result is identical with the Laspeyres price index calculated by the characteristic method (a.2.3) and that is not a coincidence since the normally indirect price imputation method becomes a direct method adjusting all observations when there is no price match in the applied dataset.

If we also use imputed prices for period 1, corresponding to double imputation instead of single imputation, the basic index still yields the same result:

$$P_{1:2}^{LA} = \left(\frac{353{,}0 \cdot 519{,}8 \cdot 585{,}6 \cdot 353{,}0 \cdot 542{,}2 \cdot 575{,}5}{455{,}1 \cdot 569{,}8 \cdot 610{,}6 \cdot 455{,}1 \cdot 496{,}1 \cdot 513{,}6}\right)^{\frac{1}{6}} \cdot 100 = 92{,}94 \qquad (a.2.17)$$

For example, for the first observation in period 1, the imputed price for period 1 is:

$$exp(5{,}498 + 0{,}0173 \cdot 25 + 0{,}1903 \cdot 1) = 455{,}1 \quad \text{(period 1)} \qquad \text{(a.2.18)}$$

Since the characteristics of the imputed prices are derived from period 1, this is a hedonic Laspeyres price index.

If the characteristics of the imputed prices are instead taken from period 2, we get the following imputed price for the first observation of period 2, in period 1, respectively, period 2:

$$exp(5{,}498 + 0{,}0173 \cdot 23 + 0{,}1903 \cdot 0) = 363{,}5 \quad \text{(period 1)} \qquad \text{(a.2.19)}$$

$$exp(5{,}074 + 0{,}0298 \cdot 23 + 0{,}0472 \cdot 0) = 317{,}2 \quad \text{(period 2)} \qquad \text{(a.2.20)}$$

The basis index then becomes:

$$P_{1:2}^{PA} = \left(\frac{317{,}2 \cdot 510{,}9 \cdot 765{,}7 \cdot 510{,}9 \cdot 475{,}4}{363{,}5 \cdot 479{,}3 \cdot 713{,}3 \cdot 479{,}3 \cdot 541{,}0}\right)^{\frac{1}{5}} \cdot 100 = 98{,}67 \qquad \text{(a.2.21)}$$

This is a hedonic Paasche price index since the characteristics are taken from period 2.

Re-pricing method[15]:

In this method the development in the volume (quantity) is adjusted for each house. Notice that since there are 6 observations in period 1, but only 5 observations in period 2, we introduce the average of the 5 original observations in period 2 as model 6 in period 2. The full dataset then becomes:

t=1			t=2		
P	**X_1**	**X_2**	**P**	**X_1**	**X_2**
390	25	1	290	23	0
480	38	1	519	39	0
700	42	1	700	51	1
550	25	1	550	39	0
520	41	0	520	35	1
490	43	0	496,4	37,4	0,4
Average:			**Average:**		
513,6*	35,67	0,67	496,4*	37,4	0,4

*) Geometric average

We now have a dataset with 6 observations in both periods and we can for example use the estimated prices from the characteristics method to calculate the Paasche quantity index based on the parameter estimates from period 2. For instance, the volume change from house 1 in the sample of period 1 to house 1 in the sample of period 2 can be written:

$$Q_{model\ 1}^{PA} = \frac{exp(5{,}074 + 0{,}0298 \cdot 23 + 0{,}0472 \cdot 0)}{exp(5{,}074 + 0{,}0298 \cdot 25 + 0{,}0472 \cdot 1)} \cdot 100 = 89{,}87 \qquad (a.2.22)$$

According to this estimate, house 1 in period 2 corresponds to 0,8987 of house 1 in period 1 implying that the price of house 1 in period 1 should be corrected by this amount to be comparable:

15 This method is illustrated in the CENEX Handbook, page 147, applied to the dataset in annex 1.

$$Price_{model1} = 390 \cdot 0{,}8987 = 350{,}5 \qquad (a.2.23)$$

The basic index for the 6 houses is:

$$P_{1:2}^{LA} = \left(\frac{290 \cdot 519 \cdot 700 \cdot 550 \cdot 520 \cdot 496{,}4}{350{,}5 \cdot 471{,}7 \cdot 915{,}2 \cdot 796{,}1 \cdot 455{,}9 \cdot 422{,}6}\right)^{\frac{1}{6}} \cdot 100 = 92{,}94 \qquad (a.2.24)$$

The result of 92,94 is identical to the hedonic Laspeyres Price index derived from the characteristics method.

We can also use the parameter estimates from period 1 to calculate the Laspeyres volume index for each unique house. For instance, for model 1 in period 1 we get:

$$Q_{model\ 1}^{LA} = \frac{exp(5{,}498 + 0{,}0173 \cdot 23 + 0{,}1903 \cdot 0)}{exp(5{,}498 + 0{,}0173 \cdot 25 + 0{,}1903 \cdot 1)} \cdot 100 = 79{,}86 \qquad (a.2.25)$$

Therefore the price of model 1 in period 1 should be corrected by this amount:

$$Price_{model1} = 390 \cdot 0{,}7986 = 311{,}5 \qquad (a.2.26)$$

The basic index for the 6 houses is:

$$P_{1:2}^{PA} = \left(\frac{290 \cdot 519 \cdot 700 \cdot 550 \cdot 520 \cdot 496{,}4}{311{,}5 \cdot 403{,}7 \cdot 817{,}8 \cdot 579{,}2 \cdot 567{,}0 \cdot 480{,}0}\right)^{\frac{1}{6}} \cdot 100 = 98{,}67 \qquad (a.2.29)$$

The result of 98,67 is identical to the hedonic Paasche price index derived from the characteristics method.

Summing up on the two indirect methods, imputation and re-pricing. With the imputation method, the Laspeyres price index is based on the 6 houses traded in period 1, and the Paasche price index is based on the 5 houses traded in period 2. The period 2 price of the 6 houses is imputed using the results of the characteristics method, and so is the period 1 price of the 5 houses traded in period 2.

With the re-pricing method the period 2 sample of 5 houses is extended with a house that represents an average of the characteristics of the 5 houses actually traded in period 2. With this extension, both samples contain 6 houses and we have 6 pair of houses for the basic price index.

STATISTICS DENMARK

Results:

The table below shows an overview over the results obtained above:

Method	Laspeyres	Paasche	Fisher
Characteristics	92,94	98,67	95,76
Time-dummy	97,64	97,64	97,64
Imputation (single)	92,94	98,67	95,76
Imputation (double)	92,94	98,67	95,76
Re-pricing	92,94	98,67	95,76

When applied to the same dataset with no price match, the imputation method becomes a direct method, which will be equivalent to the characteristics method and the re-pricing method, if the characteristics method is used to calculate the prices. Only the time-dummy method, which can be seen as a constrained characteristics approach, gives a slightly different result.

STATISTICS DENMARK

Annex 3: SPAR-method formulated as a hedonic regression (characteristic method)

1. SPAR-method with arithmetic average

Regression: $P_i = \beta_1 \cdot V_{1i} + \varepsilon_i$ ($\beta_0 = 0$, Poisson estimator[16]) (a.3.1)

Price index: $$P_{0:t}^{LA} = \frac{\hat{\beta}_{1t} \cdot \bar{V}_{10}}{\hat{\beta}_{10} \cdot \bar{V}_{10}} = \frac{\hat{\beta}_{1t}}{\hat{\beta}_{10}} \quad (a.3.2)$$

$$P_{0:t}^{PA} = \frac{\hat{\beta}_{1t} \cdot \bar{V}_{1t}}{\hat{\beta}_{10} \cdot \bar{V}_{1t}} = \frac{\hat{\beta}_{1t}}{\hat{\beta}_{10}} \quad (a.3.3)$$

$$P_{0:t}^{FI} = \sqrt{P_{0:1}^{LA} \cdot P_{0:1}^{PA}} = \frac{\hat{\beta}_{1t}}{\hat{\beta}_{10}} \quad (a.3.4)$$

2. SPAR-method with geometric average

Regression: $\log(P_i) = \beta_0 + \log(V_i) + \epsilon_i$ ($\beta_1 = 1$, OLS estimator) (a.3.5)

Price index: $$P_{0:t}^{LA} = \frac{\exp(\hat{\beta}_{0t} + \overline{\log(V_{10})})}{\exp(\hat{\beta}_{00} + \overline{\log(V_{10})})} = \frac{\exp(\hat{\beta}_{0t}) \cdot exp(\overline{\log(V_{10})})}{\exp(\hat{\beta}_{00}) \cdot exp(\overline{\log(V_{10})})} = \frac{\exp(\hat{\beta}_{0t})}{\exp(\hat{\beta}_{00})} \quad (a.3.6)$$

$$P_{0:t}^{PA} = \frac{\exp(\hat{\beta}_{0t} + \overline{\log(V_{1t})})}{\exp(\hat{\beta}_{00} + \log(V_{1t}))} = \frac{\exp(\hat{\beta}_{0t}) \cdot exp(\overline{\log(V_{1t})})}{\exp(\hat{\beta}_{00}) \cdot exp(\overline{\log(V_{1t})})} = \frac{\exp(\hat{\beta}_{0t})}{\exp(\hat{\beta}_{00})} \quad (a.3.7)$$

$$P_{0:t}^{FI} = \sqrt{P_{0:1}^{LA} \cdot P_{0:1}^{PA}} = \frac{\exp(\hat{\beta}_{0t})}{\exp(\hat{\beta}_{00})} \quad (a.3.8)$$

3. Extended SPAR-method with geometric average

Regression: $\log(P_i) = \beta_0 + \log(V_i) + \beta_2 \cdot X_{2i} + \epsilon_i$ ($\beta_1 = 1$, OLS estimator) (a.3.9)

Price index: $$P_{0:t}^{LA} = \frac{exp(\hat{\beta}_{0t} + \overline{\log(V_{10})} + \hat{\beta}_{2t} \cdot \bar{X}_{20})}{exp(\hat{\beta}_{00} + \overline{\log(V_{10})} + \hat{\beta}_{20} \cdot \bar{X}_{20})} = \frac{\exp(\hat{\beta}_{0t}) \cdot exp(\overline{\log(V_{10})}) \cdot exp(\hat{\beta}_{2t} \cdot \bar{X}_{20})}{\exp(\hat{\beta}_{00}) \cdot exp(\overline{\log(V_{10})}) \cdot exp(\hat{\beta}_{20} \cdot \bar{X}_{20})} =$$

$$\frac{\exp(\hat{\beta}_{0t}) \cdot \exp(\hat{\beta}_{2t} \cdot \bar{X}_{20})}{\exp(\hat{\beta}_{00}) \cdot \exp(\hat{\beta}_{20} \cdot \bar{X}_{20})} \quad (a.3.10)$$

$$P_{0:t}^{PA} = \frac{exp(\hat{\beta}_{0t} + \overline{\log(V_{1t})} + \hat{\beta}_{2t} \cdot \bar{X}_{2t})}{exp(\hat{\beta}_{00} + \overline{\log(V_{1t})} + \hat{\beta}_{20} \cdot \bar{X}_{2t})} = \frac{\exp(\hat{\beta}_{0t}) \cdot exp(\overline{\log(V_{1t})}) \cdot exp(\hat{\beta}_{2t} \cdot \bar{X}_{2t})}{\exp(\hat{\beta}_{00}) \cdot exp(\overline{\log(V_{1t})}) \cdot exp(\hat{\beta}_{20} \cdot \bar{X}_{2t})} =$$

$$\frac{\exp(\hat{\beta}_{0t}) \cdot \exp(\hat{\beta}_{2t} \cdot \bar{X}_{2t})}{\exp(\hat{\beta}_{00}) \cdot \exp(\hat{\beta}_{20} \cdot \bar{X}_{2t})} \quad (a.3.11)$$

[16] Since there is no constant, the residuals does not sum to zero when using OLS. However, if using the poisson estimator, the equation holds – see Ramalho (sept. 2011): Hedonic functions, hedonic methods, estimation methods and Dutot and Jevons house price indexes: are there any links?
V = appraisal (=vurdering in Danish).

$$P_{0:t}^{FI} = \sqrt{P_{0:1}^{LA} \cdot P_{0:1}^{PA}} = \sqrt{\frac{\exp(\hat{\beta}_{0t}) \cdot \exp(\hat{\beta}_{2t} \cdot \bar{X}_{20})}{\exp(\hat{\beta}_{00}) \cdot \exp(\hat{\beta}_{20} \cdot \bar{X}_{20})} \cdot \frac{\exp(\hat{\beta}_{0t}) \cdot \exp(\hat{\beta}_{2t} \cdot \bar{X}_{2t})}{\exp(\hat{\beta}_{00}) \cdot \exp(\hat{\beta}_{20} \cdot \bar{X}_{2t})}} \quad \text{(a.3.12)}$$

4. Hedonic regression with geometric average (not SPAR-method)

Regression: $\log(P_i) = \beta_0 + \beta_1 \cdot \log(V_i) + \beta_2 \cdot X_{2i} + \epsilon_i$ (OLS estimator) (a.3.13)

Price index: $$P_{0:t}^{LA} = \frac{exp(\hat{\beta}_{0t} + \hat{\beta}_{1t} \cdot \overline{\log(V_{10})} + \hat{\beta}_{2t} \cdot \bar{X}_{20})}{exp(\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \overline{\log(V_{10})} + \hat{\beta}_{20} \cdot \bar{X}_{20})} = \frac{exp(\hat{\beta}_{0t}) \cdot exp(\hat{\beta}_{1t} \cdot \overline{\log(V_{10})}) \cdot exp(\hat{\beta}_{2t} \cdot \bar{X}_{20})}{exp(\hat{\beta}_{00}) \cdot exp(\hat{\beta}_{10} \cdot \overline{\log(V_{10})}) \cdot exp(\hat{\beta}_{20} \cdot \bar{X}_{20})} =$$

$$\frac{exp(\hat{\beta}_{0t}) \cdot exp\left((\hat{\beta}_{1t} - 1) \cdot \overline{\log(V_{10})}\right) \cdot exp(\hat{\beta}_{2t} \cdot \bar{X}_{20})}{exp(\hat{\beta}_{00}) \cdot exp\left((\hat{\beta}_{10} - 1) \cdot \overline{\log(V_{10})}\right) \cdot exp(\hat{\beta}_{20} \cdot \bar{X}_{20})} \quad \text{(a.3.14)}$$

$$P_{0:t}^{PA} = \frac{exp(\hat{\beta}_{0t} + \hat{\beta}_{1t} \cdot \overline{\log(V_{1t})} + \hat{\beta}_{2t} \cdot \bar{X}_{20})}{exp(\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \overline{\log(V_{1t})} + \hat{\beta}_{20} \cdot \bar{X}_{20})} = \frac{exp(\hat{\beta}_{0t}) \cdot exp(\hat{\beta}_{1t} \cdot \overline{\log(V_{1t})}) \cdot exp(\hat{\beta}_{2t} \cdot \bar{X}_{2t})}{exp(\hat{\beta}_{00}) \cdot exp(\hat{\beta}_{10} \cdot \overline{\log(V_{1t})}) \cdot exp(\hat{\beta}_{20} \cdot \bar{X}_{2t})} =$$

$$\frac{exp(\hat{\beta}_{0t}) \cdot exp\left((\hat{\beta}_{1t} - 1) \cdot \overline{\log(V_{1t})}\right) \cdot exp(\hat{\beta}_{2t} \cdot \bar{X}_{2t})}{exp(\hat{\beta}_{00}) \cdot exp\left((\hat{\beta}_{10} - 1) \cdot \overline{\log(V_{1t})}\right) \cdot exp(\hat{\beta}_{20} \cdot \bar{X}_{2t})} \quad \text{(a.3.15)}$$

$$P_{0:t}^{FI} = \sqrt{P_{0:1}^{LA} \cdot P_{0:1}^{PA}} = \sqrt{\frac{exp(\hat{\beta}_{0t} + \hat{\beta}_{1t} \cdot \overline{\log(V_{10})} + \hat{\beta}_{2t} \cdot \bar{X}_{20})}{exp(\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \overline{\log(V_{10})} + \hat{\beta}_{20} \cdot \bar{X}_{20})} \cdot \frac{exp(\hat{\beta}_{0t} + \hat{\beta}_{1t} \cdot \overline{\log(V_{10})} + \hat{\beta}_{2t} \cdot \bar{X}_{20})}{exp(\hat{\beta}_{00} + \hat{\beta}_{10} \cdot \overline{\log(V_{10})} + \hat{\beta}_{20} \cdot \bar{X}_{20})}} \quad \text{(a.3.16)}$$

Annex 4: Numbers to the figure on page 22

Period	Volume index	SPAR	Characteristics	Timedummy	Implicit price index
1992K1	100,0	100,0	100,0	100,0	100,0
1992K2	101,3	100,0	100,2	100,2	100,2
1992K3	95,7	98,1	96,5	96,5	96,5
1992K4	94,9	95,6	94,7	94,7	94,7
1993K1	93,9	94,3	93,3	93,3	93,3
1993K2	96,9	93,7	94,7	94,7	94,7
1993K3	101,8	98,1	99,2	99,3	99,2
1993K4	107,6	103,8	105,0	105,0	105,0
1994K1	112,1	109,8	109,2	109,3	109,3
1994K2	111,7	109,8	109,6	109,7	109,7
1994K3	109,8	108,2	106,6	106,7	106,6
1994K4	109,9	109,8	108,1	108,2	108,0
1995K1	112,4	112,3	110,3	110,4	110,2
1995K2	118,1	116,5	114,6	114,7	114,6
1995K3	115,9	119,3	116,0	116,1	115,8
1995K4	122,2	122,8	120,7	120,8	120,5
1996K1	127,7	124,7	125,6	125,7	125,5
1996K2	132,7	127,8	130,0	130,1	129,9
1996K3	134,5	131,6	131,5	131,7	131,6
1996K4	141,3	137,0	137,6	137,8	137,5
1997K1	141,0	140,5	137,0	137,2	137,0
1997K2	142,9	144,3	140,2	140,4	140,1
1997K3	143,7	147,5	141,3	141,5	141,3

1997K4	146,3	149,1	144,2	144,5	144,2
1998K1	154,3	151,9	151,9	152,2	151,9
1998K2	159,6	158,2	156,3	156,6	156,3
1998K3	154,2	160,4	155,7	155,9	155,7
1998K4	158,8	163,0	159,1	159,4	159,1
1999K1	165,3	165,5	164,0	164,3	164,0
1999K2	171,1	168,0	167,1	167,3	167,2
1999K3	166,9	171,2	168,0	168,2	168,1
1999K4	171,7	171,5	169,6	169,8	169,7
2000K1	179,1	174,1	173,3	173,6	173,5
2000K2	183,7	178,8	177,3	177,6	177,5
2000K3	186,1	182,9	178,7	178,9	179,2
2000K4	186,6	184,5	179,5	180,0	180,1
2001K1	194,6	188,0	186,1	186,7	186,7
2001K2	196,6	190,5	187,8	188,4	188,4
2001K3	191,8	192,7	185,9	186,5	186,7
2001K4	191,4	191,1	185,3	185,8	185,7
2002K1	196,4	194,0	189,7	190,2	186,2
2002K2	200,9	197,8	193,0	193,5	190,0
2002K3	197,9	199,1	193,2	193,6	189,9
2002K4	197,8	199,1	194,0	194,4	190,2
2003K1	200,1	199,4	197,7	198,2	194,0
2003K2	209,1	203,5	202,2	202,6	198,3
2003K3	210,4	206,0	202,5	202,9	198,6
2003K4	211,5	206,0	202,1	202,5	198,1
2004K1	221,1	211,7	208,4	208,9	204,4

2004K2	233,1	219,6	217,3	217,7	213,2
2004K3	230,1	225,9	218,5	218,9	214,3
2004K4	235,0	230,4	222,2	222,7	217,9
2005K1	246,0	239,6	229,6	230,0	225,2
2005K2	264,3	253,5	243,9	244,1	239,0
2005K3	262,4	268,0	249,5	249,1	243,6
2005K4	278,3	282,6	260,8	259,8	253,9
2006K1	294,6	299,7	277,7	276,6	270,3
2006K2	304,9	318,4	293,5	292,2	284,9
2006K3	299,2	325,9	294,3	292,4	284,0
2006K4	297,1	325,0	294,2	293,1	284,8
2007K1	303,1	329,4	306,3	305,3	296,5
2007K2	312,4	333,9	313,9	312,9	303,4
2007K3	313,2	334,8	312,5	311,4	301,8
2007K4	305,3	328,8	310,8	310,0	300,8
2008K1	311,9	325,3	312,8	311,9	302,3
2008K2	322,9	328,2	317,8	317,0	307,0
2008K3	300,5	319,3	306,2	305,4	296,6
2008K4	284,4	294,3	284,9	284,5	275,4
2009K1	281,9	276,9	277,6	276,7	267,0
2009K2	295,4	278,2	282,3	281,5	270,3
2009K3	299,3	280,4	285,0	284,6	273,5
2009K4	307,7	279,4	282,8	283,0	271,9
2010K1	307,8	281,0	281,0	281,2	270,2
2010K2	315,6	288,0	289,0	289,3	278,0
2010K3	306,9	288,9	285,4	285,8	274,1

2010K4	312,2	287,7	282,5	282,9	271,8
2011K1	304,9	282,9	275,0	275,4	264,6
2011K2	306,9	285,4	280,9	281,4	270,3
2011K3	289,2	276,9	268,3	268,8	258,1
2011K4	290,3	268,4	262,0	262,4	252,2
2012K1	298,6	267,4	265,4	265,7	255,2
2012K2	301,2	269,9	269,3	269,9	259,2
2012K3	303,9	269,9	267,1	267,6	256,9
2012K4	312,1	269,9	265,7	266,2	255,7
2013K1	315,4	272,5	269,1	269,6	258,9
2013K2	320,8	278,5	276,2	276,6	265,5
2013K3	319,1	277,2	271,4	272,0	261,2
2013K4	320,8	277,8	267,9	268,7	258,0
2014K1	325,2	280,1	271,1	271,8	261,1
2014K2	339,7	288,3	281,4	282,1	271,1
2014K3	329,5	287,7	275,2	276,0	265,1
2014K4	329,0	287,7	274,5	275,3	264,3
2015K1	343,9	295,3	286,1	286,9	275,3

Li

Literature:

- Ramalho, Esmeralda A. (sept. 2011): Hedonic functions, hedonic methods, estimation methods and Dutot and Jevons house price indexes: are there any links?
- Triplett, Jack (2006): Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes.
- Statistics Netherlands (2013): Method description, New dwellings, output price indices building costs

- Statistics Norway: http://www.ssb.no/priser-og-prisindekser/artikler-og-publikasjoner/boligprisindeksen--65148