Multilateral Hedonic House Price Indexes

Jan de Haan[a]

30 April 2022

Abstract: The Time Dummy Hedonic method has frequently been applied to construct quality-adjusted price indexes for residential property. This method belongs to the class of multilateral methods where price indexes for all periods are estimated simultaneously from the data pertaining to the entire window. In this short paper, I discuss an alternate multilateral method: hedonic imputation GEKS. It turns out that the imputation GEKS index is equal to the geometric mean of two other multilateral hedonic price indexes: a modified version of the "average characteristics price index" and an index which is very similar to the Time Dummy Hedonic index.

Keywords: hedonic imputation, house prices, multilateral index methods.

JEL Classification: C43, E31.

---

[a] Statistics Netherlands and Delft University of Technology; jandehaan1@gmail.com.

The views expressed in this paper are those of the author and do not necessarily reflect the views of Statistics Netherlands.

1. Introduction

"Eurostat recommends that the HPI [House Price Index] should be computed using a hedonic approach, but has not provided guidance to NSIs [National Statistical Institutes] as to which hedonic method should be used." (Hill et al., 2018, p. 222). The Handbook on Residential Property Price Indices (Eurostat, 2013) provides an overview of most of the available methods, both hedonic and non-hedonic. Some methods are multilateral, meaning that price index numbers for all periods are estimated simultaneously from the data pertaining to the entire estimation window. The only multilateral hedonic method discussed in the Handbook is the well-known Time Dummy Hedonic method.[1] Several NSIs in Europe are currently using this method for the construction of their house price index.

This short paper presents an alternate multilateral hedonic method to compute a house price index which, to the best of my knowledge, has not been discussed before: (unweighted) hedonic imputation GEKS. I am not proposing to implement this method in official statistics at this stage as there are some issues that need further investigation, but the approach is quite interesting and provides insight into the various multilateral hedonic house price methods. The paper shows that the imputation GEKS index equals the geometric mean of two (unfamiliar) multilateral hedonic price indexes: a modified version of the so-called average characteristics price index and an index which is similar to the Time Dummy Hedonic index.

Section 2 briefly describes the Time Dummy Hedonic method, which relies on a pooled regression. Section 3 presents the imputation GEKS method, which is based on regressions for all time periods separately. Section 4 addresses the above result, i.e., the decomposition of the imputation GEKS index into two other multilateral hedonic price indexes. Section 5 concludes.

---

[1] Hill (2011) presents a comprehensive overview of hedonic house price indexes. The most famous non-hedonic multilateral method to construct a house price index is the repeat sales method. For a comparison of various non-hedonic multilateral methods for housing and an application to Dutch data, see De Haan and Van de Laar (2021). Multilateral methods are increasingly being used by national statistical institutes to compile (non-housing) sub-components of the Consumer Price Index (CPI) from scanner data or from other types of transaction data. For an overview of methods and an application to Australian scanner data, see Van Kints, De Haan and Webster (2019). The use of multilateral index methods for the CPI was first proposed by Ivancic, Diewert and Fox (2011) to deal with chain drift in superlative price indexes due to stocking up goods that are on sale. Most countries using a multilateral method for scanner data selected the GEKS-Törnqvist method.

2. The Time Dummy Hedonic index

The estimation window consists of time periods $t = 0,...,T$; $T+1$ periods in total. Let S^t denote the set of houses sold in period *t* and N^t the corresponding number of houses (the size of S^t). The Time Dummy Hedonic (TDH) method estimates the following log-linear regression model on the pooled data with $N = \sum_{t=0}^{T} N^t$ observations:

$$\ln p_i^t = \alpha + \sum_{t=1}^{T} \delta^t D_i^t + \sum_{k=1}^{K} \beta_k z_{ik} + \varepsilon_i^t, \tag{1}$$

where p_i^t denotes the price of property *i* in period *t*, z_{ik} is the *k*-th characteristics of *i* $(k = 1,...,K)$,[2] D_i^t is a dummy variable that has the value 1 if the observation relates to period *t* and 0 otherwise, and ε_i^t is an error term with zero mean. Because an intercept α is included in the model, the dummy variable for period 0 is excluded. I assume that (1) is estimated using Ordinary Least Squares (OLS) regression; the coefficients are $\hat{\alpha}$, $\hat{\delta}^t$ $(t = 1,...,T)$, and $\hat{\beta}_k$ $(k = 1,...,K)$. The predicted prices in period 0 and period *t* are given by $\hat{p}_i^0 = \exp(\hat{\alpha})\exp(\sum_{k=1}^{K} \hat{\beta}_k z_{ik})$ and $\hat{p}_i^t = \exp(\hat{\alpha})\exp(\hat{\delta}^t)\exp(\sum_{k=1}^{K} \hat{\beta}_k z_{ik})$, and the TPD index between 0 to *t* is equal to $P_{TPD}^{0t} = \exp(\hat{\delta}^t) = \hat{p}_i^t / \hat{p}_i^0$.

Due to the time dummy specification of model (1), the OLS regression residuals sum to zero in each period so that $\sum_{i \in S^t} \ln \hat{p}_i^t / N^t = \sum_{i \in S^t} \ln p_i^t / N^t$. Taking exponents yields $\prod_{i \in S^t} (\hat{p}_i^t)^{1/N^t} = \prod_{i \in S^t} (p_i^t)^{1/N^t}$ $(t = 0,...,T)$; in every period, the geometric mean of the estimated prices is equal to the geometric mean of the observable prices. Using this result, it can be shown that the TDH index can be written as (see e.g., De Haan and Krsinich, 2018)

$$P_{TDH}^{0t} = \frac{\overline{p}^t}{\overline{p}^0} \exp\left[\sum_{k=1}^{K} \hat{\beta}_k \left(\overline{z}_k^0 - z_k^t\right)\right] = \frac{\overline{p}^t / \overline{p}^0}{\exp\left[\sum_{k=1}^{K} \hat{\beta}_k \left(\overline{z}_k^t - \overline{z}_k^0\right)\right]}, \tag{2}$$

where $\overline{z}_k^0 = \sum_{i \in S^0} z_{ik} / N^0$ and $\overline{z}_k^t = \sum_{i \in S^t} z_{ik} / N^t$ denote the (arithmetic) means of the characteristics in the respective periods; $\overline{p}^0 = \prod_{i \in S^0} (p_i^0)^{1/N^0}$ and $\overline{p}^t = \prod_{i \in S^t} (p_i^t)^{1/N^t}$ are the geometric means of the prices. As can be seen, the TDH method adjusts the ratio of geometric average prices for changes in the average characteristics. The denominator of the second expression of equation (2), $\exp\left[\sum_{k=1}^{K} \hat{\beta}_k (\overline{z}_k^t - \overline{z}_k^0)\right]$, is sometimes referred to as a quality index.

---

[2] The housing characteristics are assumed fixed. In reality some are likely to change (due to depreciation and renovations) but the analysis will still hold.

3. The hedonic imputation GEKS index

Instead of running the “big” pooled regression, we can estimate the log-linear hedonic model

$$\ln p_i^t = \alpha^t + \sum_{k=1}^{K} \beta_k^t z_{ik} + \varepsilon_i^t \tag{3}$$

on the data of S^t, i.e., separately for each period $t = 0,...,T$. The (time-dependent) OLS coefficients are $\hat{\alpha}^t$ and $\hat{\beta}_k^t$, and the predicted prices are $\hat{p}_i^t = \exp(\hat{\alpha}^t)\exp(\sum_{k=1}^{K} \hat{\beta}_k^t z_{ik})$. Again, we have $\prod_{i\in S^t} (\hat{p}_i^t)^{1/N^t} = \exp(\hat{\alpha}^t)\exp(\sum_{k=1}^{K} \hat{\beta}_k^t \bar{z}_k^t) = \prod_{i\in S^t} (p_i^t)^{1/N^t}$.

Since houses are unique and usually not sold more than once during a quarter, all quantities are equal to 1, and so the use of an unweighted price index seems appropriate. For a bilateral comparison, there are two geometric options: the imputation Jevons price index that is based on the set of houses S^0, $\prod_{i\in S^0} (\hat{p}_i^t / p_i^0)^{1/N^0} = \prod_{i\in S^0} (\hat{p}_i^t / \hat{p}_i^0)^{1/N^0}$, and the imputation Jevons price index based on S^t, $\prod_{i\in S^t} (p_i^t / \hat{p}_i^0)^{1/N^0} = \prod_{i\in S^0} (\hat{p}_i^t / \hat{p}_i^0)^{1/N^0}$. Taking the geometric mean of the two options is useful due to the symmetric treatment of the two periods. This leads to the following bilateral Hedonic Imputation price index (see also Diewert, Heravi and Silver, 2009):

$$P_{HI}^{0t} = \exp\left(\hat{\alpha}^t - \hat{\alpha}^0\right)\exp\left[\sum_{k=1}^{K}\left(\hat{\beta}_k^t - \hat{\beta}_k^0\right)\left(\frac{\bar{z}_k^0 + \bar{z}_k^t}{2}\right)\right]. \tag{4}$$

If transitivity is asked for, the bilateral hedonic imputation price indexes can be used as elements in the GEKS procedure (see e.g., De Haan and Daalmans, 2019). With l denoting the link period $(0 \le l \le T)$, this leads to the multilateral hedonic imputation GEKS price index

$$\begin{aligned} P_{HIGEKS}^{0t} &= \prod_{l=0}^{T}\left[P_{HI}^{0l} P_{HI}^{lt}\right]^{\frac{1}{T+1}} \\ &= \exp\left(\hat{\alpha}^t - \hat{\alpha}^0\right) \\ &\times \prod_{l=0}^{T}\left[\exp\left[\sum_{k=1}^{K}\left(\hat{\beta}_k^l - \hat{\beta}_k^0\right)\left(\frac{\bar{z}_k^0 + \bar{z}_k^l}{2}\right)\right]\right]^{\frac{1}{T+1}} \\ &\times \prod_{l=0}^{T}\left[\exp\left[\sum_{k=1}^{K}\left(\hat{\beta}_k^t - \hat{\beta}_k^l\right)\left(\frac{\bar{z}_k^l + \bar{z}_k^t}{2}\right)\right]\right]^{\frac{1}{T+1}}. \end{aligned} \tag{5}$$

4. A decomposition

In the Appendix, I show that the hedonic imputation GEKS price index can be written as the geometric mean of two other (unfamiliar) multilateral hedonic price indexes, P_A^{0t} and P_B^{0t}:

$$P_{HIGEKS}^{0t} = \left[P_A^{0t} P_B^{0t} \right]^{\frac{1}{2}}, \tag{6}$$

with

$$P_A^{0t} = \exp\left(\hat{\alpha}^t - \hat{\alpha}^0\right) \exp\left[\sum_{k=1}^{K} \left(\hat{\beta}_k^t - \hat{\beta}_k^0\right) \bar{z}_k^* \right]; \tag{7}$$

$$P_B^{0t} = \frac{\bar{p}^t / \bar{p}^0}{\exp\left[\sum_{k=1}^{K} \hat{\beta}_k^* \left(\bar{z}_k^t - \bar{z}_k^0\right) \right]}, \tag{8}$$

where $\bar{z}_k^* = \sum_{t=0}^{T} \bar{z}_k^t / (T+1)$ en $\hat{\beta}_k^* = \sum_{t=0}^{T} \hat{\beta}_k^t / (T+1)$.

P_A^{0t} is a modified version of the so-called average characteristics index that uses the full-sample average characteristics $\bar{z}_k = \sum_{t=0}^{T} \sum_{i \in S^t} z_{ki} / N = \sum_{t=0}^{T} (N^t / N) \bar{z}_k^t$ (Hill et al., 2018). In general, $\bar{z}_k^*$ will not be equal to $\bar{z}_k$. P_A^{0t} can be written as

$$P_A^{0t} = \prod_{r=0}^{T} \left(P_{(r)}^{0t}\right)^{\frac{1}{T+1}}, \tag{9}$$

where $P_{(r)}^{0t} = \prod_{i \in S^r} (\hat{p}_i^t / \hat{p}_i^0)^{1/N^r}$, with imputed prices $\hat{p}_i^0$ and $\hat{p}_i^t$ based on the separate regressions. Thus, P_A^{0t} is the geometric mean of the (non-symmetric) imputation Jevons price indexes using all possible "reference periods" r $(r = 0,...,T)$. This is a simple way to impose transitivity.

P_B^{0t} is an implicit price index, calculated as the ratio of geometric average prices divided by a quality index. A comparison of equation (8) with equation (2) shows that P_B^{0t} is similar to the TDH index. The only difference is that the average coefficients $\hat{\beta}_k^*$ from the $T+1$ separate regressions are used rather than the coefficients $\hat{\beta}_k$ from the pooled regression.

If P_A^{0t} and P_B^{0t} are deemed equally good, taking their geometric mean seems like a good idea, and this is exactly what P_{HIGEKS}^{0t} does. However, it is unclear whether P_A^{0t} and P_B^{0t} are equally good. More importantly perhaps, it is not clear either whether P_A^{0t} and P_B^{0t} are better choices than the usual average characteristics price index and the TDH index.

5. Conclusions

The multilateral TDH index is transitive. Transitivity is a nice property for a price index as the results will be independent of the choice of base period, but it is not necessarily a requirement. Because chain drift is unlikely to be a problem with unweighted indexes, chaining adjacent-period (i.e., bilateral) TDH indexes is also an option. This approach is followed in France and Portugal (Hill et al., 2018). The reason why some other NSIs in Europe prefer using the multilateral version is that pooling data from multiple periods stabilizes the coefficients so that the index will be less volatile.

The multilateral TDH method constrains the hedonic coefficients to be the same across all time periods of the window. But the "true" parameters may change over time. Price index P_A^{0t}, defined by (7), accounts for parameter changes and fixes the average characteristics at $\overline{z}_k^*$. Just like the TDH index, price index P_B^{0t}, defined by (8), fixes the coefficients, in this case at average values $\hat{\beta}_k^*$. By taking the mean of P_A^{0t} and P_B^{0t}, the imputation GEKS index has "the best of both worlds" and seems like a good alternative to the TDH index (though the latter is easier to compute). When the "true" parameters happen to be constant, the two methods are likely to produce similar results.

Yet, as mentioned earlier, it is unclear whether P_A^{0t} and P_B^{0t} are better choices than the usual average characteristics index, denoted P_{AC}^{0t} in (10) below, and the TDH index. It might be interesting to examine empirically how the geometric mean of the last two indexes differs from the hedonic imputation GEKS index. In any case, if $\overline{z}_k^* \simeq \overline{z}_k$ and $\hat{\beta}_k^* \simeq \hat{\beta}_k$, we find

$$P_{HIGEKS}^{0t} \simeq \left[P_{AC}^{0t} P_{TDH}^{0t} \right]^{\frac{1}{2}}. \tag{10}$$

While multilateral methods reduce volatility, which is useful, they do have some disadvantages including: revisions of previously estimated indexes when data is added; and a "loss of characteristicity" due to the use of data from the entire window rather than just from the periods compared. To deal with revisions, NSIs using TDH employ a rolling window approach with a form of linking to update the time series.[3] To mitigate the loss of characteristicity, a window length of less than 6 quarters is typically chosen – a window length of 2 quarters produces the chained adjacent-period TDH method. Both solutions can also be applied to GEKS (and to other multilateral indexes).

---

[3] Shimizu, Nishimura and Watanabe (2010) introduced rolling-window hedonic indexes into the housing literature. Hill et al. (2022) discuss various linking options for dealing with low transaction volume.

Appendix: Derivation of equation (6)

The second and third components of equation (5) can be written as follows:

$$\prod_{l=0}^{T}\left[\exp\left[\sum_{k=1}^{K}\left(\hat{\beta}_k^l-\hat{\beta}_k^0\right)\left(\frac{\overline{z}_k^0+\overline{z}_k^l}{2}\right)\right]\right]^{\frac{1}{T+1}}$$

$$=\left[\prod_{l=0}^{T}\left[\exp\left(\sum_{k=1}^{K}\hat{\beta}_k^l\overline{z}_k^0\right)\right]^{\frac{1}{T+1}}\right]^{\frac{1}{2}}\times\left[\prod_{l=0}^{T}\left[\exp\left(\sum_{k=1}^{K}\hat{\beta}_k^l\overline{z}_k^l\right)\right]^{\frac{1}{T+1}}\right]^{\frac{1}{2}}$$

$$\times\left[\prod_{l=0}^{T}\left[\exp\left(\sum_{k=1}^{K}\hat{\beta}_k^0\overline{z}_k^0\right)\right]^{\frac{1}{T+1}}\right]^{-\frac{1}{2}}\times\left[\prod_{l=0}^{T}\left[\exp\left(\sum_{k=1}^{K}\hat{\beta}_k^0\overline{z}_k^l\right)\right]^{\frac{1}{T+1}}\right]^{-\frac{1}{2}}; \qquad \text{(A.1)}$$

and

$$\prod_{l=0}^{T}\left[\exp\left[\sum_{k=1}^{K}\left(\hat{\beta}_k^t-\hat{\beta}_k^l\right)\left(\frac{\overline{z}_k^l+\overline{z}_k^t}{2}\right)\right]\right]^{\frac{1}{T+1}}$$

$$=\left[\prod_{l=0}^{T}\left[\exp\left(\sum_{k=1}^{K}\hat{\beta}_k^t\overline{z}_k^l\right)\right]^{\frac{1}{T+1}}\right]^{\frac{1}{2}}\times\left[\prod_{l=0}^{T}\left[\exp\left(\sum_{k=1}^{K}\hat{\beta}_k^t\overline{z}_k^t\right)\right]^{\frac{1}{T+1}}\right]^{\frac{1}{2}}$$

$$\times\left[\prod_{l=0}^{T}\left[\exp\left(\sum_{k=1}^{K}\hat{\beta}_k^l\overline{z}_k^l\right)\right]^{\frac{1}{T+1}}\right]^{-\frac{1}{2}}\times\left[\prod_{l=0}^{T}\left[\exp\left(\sum_{k=1}^{K}\hat{\beta}_k^l\overline{z}_k^t\right)\right]^{\frac{1}{T+1}}\right]^{-\frac{1}{2}}. \qquad \text{(A.2)}$$

When multiplying (A.1) and (A.2), two terms cancel out. Using $\hat{\beta}_k^* = \sum_{l=1}^{T}\hat{\beta}_k^t/(T+1)$, $\overline{z}_k^* = \sum_{l=1}^{T}\overline{z}_k^t/(T+1)$, $\exp(\hat{\alpha}^0)\exp\left(\sum_{k=1}^{K}\hat{\beta}_k^0\overline{z}_k^0\right)=\overline{p}^0$, and $\exp(\hat{\alpha}^t)\exp\left(\sum_{k=1}^{K}\hat{\beta}_k^t\overline{z}_k^t\right)=\overline{p}^t$, it follows from (A.1) and (A.2) that equation (5) can be written as

$$P_{HIGEKS}^{0t}=\left[\exp\left(\hat{\alpha}^t-\hat{\alpha}^0\right)\right]^{\frac{1}{2}}\left[\frac{\overline{p}^t}{\overline{p}^0}\right]^{\frac{1}{2}}\left[\exp\left[\sum_{k=1}^{K}\hat{\beta}_k^*\left(\overline{z}_k^0-\overline{z}_k^t\right)\right]\right]^{\frac{1}{2}}\left[\exp\left[\sum_{k=1}^{K}\left(\hat{\beta}_k^t-\hat{\beta}_k^0\right)\overline{z}_k^*\right]\right]^{\frac{1}{2}}$$

$$=\left[\exp\left(\hat{\alpha}^t-\hat{\alpha}^0\right)\exp\left[\sum_{k=1}^{K}\left(\hat{\beta}_k^t-\hat{\beta}_k^0\right)\overline{z}_k^*\right]\right]^{\frac{1}{2}}\left[\frac{\overline{p}^t/\overline{p}^0}{\exp\left[\sum_{k=1}^{K}\hat{\beta}_k^*\left(\overline{z}_k^t-\overline{z}_k^0\right)\right]}\right]^{\frac{1}{2}}, \qquad \text{(A.3)}$$

which is equal to equation (6), using definitions (7) and (8) for P_A^{0t} and P_B^{0t}.

References

De Haan, J. en F. Krsinich (2018), "Time Dummy Hedonic and Quality-Adjusted Unit Value Indexes: Do They Really Differ?", *Review of Income and Wealth* 64, 757-776.

De Haan, J. and J. Daalmans (2019), "Scanner Data in the CPI: The Imputation CCDI Index Revisited", Paper presented at the 16th meeting of the Ottawa Group, 8-10 May 2019, Rio de Janeiro, Brazil.

De Haan, J. and R. van de Laar (2021), "House Price Indexes: A Comparison of Repeat Sales and Other Multilateral Methods", Paper presented at the 21st EMG Workshop (online), 26 November 2021, University of New South Wales.

Diewert, W.E., S. Heravi and M. Silver (2009), "Hedonic Imputation versus Time Dummy Hedonic Indexes", pp. 87-116 in W.E. Diewert, J. Greenlees and C. Hulten (eds.), *Price Index Concepts and Measurement*. Chicago: University of Chicago Press.

Eurostat (2013), Handbook on Residential Property Price Indices (RPPIs). Luxembourg: Publications Office of the European Union.

Hill, R.J., M. Scholz, C. Shimizu and M. Steurer (2018), "An Evaluation of the Methods Used by European National Statistical Institutes to Compute their Official House Price Indices", *Economie et Statistique/Economics and Statistics* 500-502, 221-238.

Hill, R.J., M. Scholz, C. Shimizu en M. Steurer (2022), "Rolling-Time-Dummy House Price Indexes: Window Length, Linking and Options for Dealing with Low Transaction Volume", *Journal of Official Statistics* 38, 127-151.

Hill, R.J. (2011), "Hedonic Price Indexes for Housing", OECD Statistics Working Paper 2011/01, OECD, Paris, France.

Ivancic, L., W.E. Diewert and K.J. Fox (2011), "Scanner Data, Time Aggregation and the Construction of Price Indexes", *Journal of Econometrics* 161, 24-35.

Shimizu, C., K.G. Nishimura and T. Watanabe (2010), "Housing Prices in Tokyo: A Comparison of Hedonic and Repeat Sales Measures", *Journal of Economics and Statistics* 230, 792-813.

Van Kints, M., J. de Haan and M. Webster (2019), "Utilizing Big Data and Multilateral Index Methods to Produce the Australian CPI: Theory, Implementation and Empirical Results", *Statistical Journal of the IAOS* 35, 387-405.