Takarék House Price Index

Methodological guide

Table of contents

Published: 1st of October, 2009

Updated: 7th of August, 2018

The development of the methodology was a cooperation of Takarék Mortgage Bank Co. Plc. (FHB Mortgage Bank Co Plc. at the time and the National Bank of Hungary, under the guidance of Áron Horváth. The work of the housing expert of the Central Statistical Office, Mrs. Gáborné Székely was indispensable whom we hereby thank for her expertise and experience. Bertalan Papp, Zoltán Sápi and Bálint Szalai have also taken part in doing statistical calculations.

1. The Takarék House Price Index

The Takarék House Price Index measures the development of Hungarian residential real estate prices. It shows how the value of Hungarian residential properties has changed since 1998.

The index is published quarterly, and starts from first quarter 1998. At the time of quarterly disclosures, the time series will end with data of the previous quarter. I.e. at the first publication, in October 2009 it contains 46 pieces of data, from which the last one applies to 2nd quarter 2009. The value of the index is normalised with the average of year 2000, i.e. the average index value in 2000 is 100.

During the development of the index, we regarded such long-standing, internationally well-known indices as examples as the house price index of Halifax1 or Hypoport AG2. We made it our aim to prepare the Takarék House Price Index on similar standards, and to make it an unavoidable reference in the profession. This methodological guide was written for those who are interested in details of the construction of the index. We are going to present the features of the used database, then demonstrate the hedonic method applied when computing the index, with the help of which we developed an indicator of actual price movements from changeable data. Finally, we are going to share some conclusions which we experienced during the process. The description is closed with the planned direction of our upcoming work, as certain index parts can be formed from the compiled database as well that could be subject to the interest of the profession.

[1] http://www.lloydsbankinggroup.com/media1/research/halifax_hpi.asp

[2] http://www.hypoport.com/hpx_mean_en.html

2. Data

a) Data source, data

The index is based on purchase-and-sell transaction data of residential real estate, and was prepared processing purchase-and-sell data of more than 800 thousand residential properties from some 3000 settlements from 1998 onwards. Both the own observations of Takarék Group and a database purchased from the National Tax and Customs Administration served as sources of data. The following table shows the yearly distribution of used data[3]:

	Number of used data records
1998	14 810
1999	66 176
2000	71 581
2001	73 118
2002	100 468
2003	119 842
2004	68 327
2005	87 249
2006	109 567
2007	65 942
2008	59 064
2009	37 872
2010	42 479
2011	45 181
2012	44 849
2013	30 789
2014	53 794
2015	61 983
2016	66 455
2017	63 304

[3] We must note that the number of used data unfortunately does not show the number of Hungarian housing transactions from the given year obviously, because the data supply from the National Tax and Customs Administration, in our opinion, was not full in more quarters.

When providing loans for the purchase of real estate, Takarék Group devotes special attention to the registration of features of the properties. During valuation for the assessment of loan-to-value, several attributes of a property are investigated.

- Address, location.
- Area (gross, net, useful).
- Area of accompanying parcel.
- Year of construction.
- Distance from public transportation (train, local- and distance bus).
- Condition.
- Technical features of the built construction.
- Access to public utilities.

These data allow for the application of the hedonic method (to be discussed later).

Another part of the database is data collected by stamp duty offices coming from the National Tax and Customs Adminstration since 2008. Every single Hungarian residential property transaction must find its way to this database, which comprise the following.

- Date of purchase.
- Address (settlement, district if in Budapest).
- Purchase price documented through transaction.
- Sum of valuation by a stamp duty office.
- Type of building (detached house, terraced house, condominium or block flat).
- Area.

However the incoming data are rather incomplete. The area, which is indispensable for the calculations, is missing in more than a quarter of the database. An even more serious problem is that the quality of available data is weak, errors can be identified in numerous cases: often an impossible sum is stated to be the purchase price. That's why this dataset is first analysed by experts of Takarék Group one by one, and obvious errors are corrected and complemented based on formerly accumulated data (for instance, earlier information is often

available about the area of the property). Afterwards, data that are useless from a statistical point of view are filtered out from the still deficient database using automated filtering conditions, and remaining data are drawn into the calculations.

The filtering conditions concern that

- The transaction should be between private persons.
- The whole ownership should be transferred.
- An existing date should be connected to the purchase price.
- An area should be belong to the purchase price.
- The price should be reasonable.
- The price of one area unit should be reasonable.

In the 4th quarter of 2008 and in the beginning of 2009, we faced extreme deficiencies in the database, which may well have the reason that the stamp duty offices passed the supplying of data to the National Tax and Customs Administration (Tax and Financial Control Administration back than) at this time. The price movements of this period were finalised relying on the expertise of the Hungarian Central Statistical Office.

b) Change of the information set

The two sources of the database are available with a significant shift in time. Takarék Group registers its data with special care, so virtually up-to-date information is disposable based on them. However, a large part of data coming from APEH finds its way into the database at least half a year after the purchase transaction, a not negligible number of data have even arrived years after the transactions. On the other hand, we especially put emphasis on the index reflecting even the newest state of the housing market, thus we do not wait for all the purchase-and-sell data of the latest period when preparing the index. Based on our research, we concluded that fewer observations carry substantial information as well, so we decided for the soonest publication possible.

Using the data arriving continuously from the National Tax and Customs Administration, we recalculate the values of the index that are for the latest period, and as such, these may be subject to modification. Our insecurity because of the incomplete information set is indicated by dashed lines at the end of the graph.

3. The hedonic calculation method

For computing the index, we applied the process based on the hedonic method, which is usual in the profession, and the classic reference of which is the study of Kain and Quigley4. The method serves to filter out the composition effect appearing in observations of simple indicators (mean price, median price). The composition effect itself arises, because not every single property is sold in every quarter, thus our observations – even if the transaction data are complete - are accordingly a restricted sample of the real estate stock, and the sample is not representative. The hedonic method is the most widespread statistical way to measure aggregate property price change, and is based on the theoretical consideration that the value of houses/flats can be divided into the values of their attributes. So if we start from the assumption that the price of real estate depends on its attributes, we get the general change of the price level by filtering out the effect of the change of features in the sample from the change of average price. A simple example demonstrates the point of the method. Let's suppose that in the country of Residency, flats of unique design are more expensive than flats of the same type. Altogether six flats were built in Residency. The table below shows the distribution of the two types and the value of the flats.

of same type	of unique design
	20 thousand dollars
10 thousand dollars	20 thousand dollars
	20 thousand dollars
10 thousand dollars	20 thousand dollars

The value of the flats is stable, but it is random, which ones are purchased in the different periods. In the first period, two flats alike and one unique flat changes owners:

of same type	of unique design
	20 thousand dollars
10 thousand dollars	20 thousand dollars
	20 thousand dollars
10 thousand dollars	20 thousand dollars

In this case, the average price of the observed transactions is: (10 + 10 + 20) / 3 = **13.3** thousand dollars.

[4] Kain, John F. – John M. Quigley [1970]: Measuring the Value of Housing Quality. *Journal of the American Statistical Association* 65 (330), Jun 1970, pp. 532-548.

In the second period, the market gets momentum, and all the flats change owners.

of same type	of unique design
	20 thousand dollars
10 thousand dollars	20 thousand dollars
	20 thousand dollars
10 thousand dollars	20 thousand dollars

Now the mean price of observed transactions is (2 * 10 + 4 * 20) / 6 = **16.6** thousand dollars.

So the observed mean price has risen, because more flats of unique design were in circulation in the second time period. I.e. the composition of traded flats changed, and thus the composition effect increased the average price despite the nationwide valuation of the flats remaining unchanged. The hedonic method deals with this problem by trying to identify the value of attributes that define the value of the flats. In the above example, the flats have one feature, namely their type. According to the hedonic method, we observe that two flats of the same type cost 10 thousand on the average in the first period, while the only one of unique design cost 20 thousand. In the second period, again the quality of traded flats may be observed: the average price of flats of the same type is again 10 thousand, and that of unique flats is 20 thousand. The conclusion of the method now is that the price of the flats did not change, because neither the price of alike nor the price of unique design “attribute” changed from the first period to the second.

The bigger the disposable sample, and the more features of the flats observed, the bigger possibility that the deviations deriving from the change in composition can be filtered out. Based on the method, the composition effect may be filtered out as well that during the introduction of the housing subsidy system bound to value limits, the number of transactions with cheaper flats may well increase, which ceteris paribus reduces the observed mean price.

When applying the hedonic method, we need to run a regression. In the regressions, the prices of properties (or because of the skewness of the price distribution, the logarithms of prices, log p_j) are explained by their observable features.

$$log\ p_j = b_0 + b_1 q_{1j} + b_2 q_{2j} + ... + b_p q_{pj} + \sum_{i=2}^{t} m_i x_{ij} + u_j$$

The variables indicated with *q* are the attributes of the flats (size, type or location), and coefficients *b* are coefficients of the attributes drawn in as explanatory variables. The coefficients show how one more unit from the attributes (e.g. one more bathroom) raises the price of the property on the average. That’s why these coefficients are referred to as shadow prices of the features. The above specification is a so-called restricted hedonic regression, because *b* coefficients have no time index, i.e. shadow prices are unchanged through

time. Finally, x_{ij} variables in the summa part are the dummy variables showing the period of observation, i.e. x_{ij} = 1 if the j^{th} property was sold in the i^{th} period. Accordingly, the m_i estimated coefficients explain the "value" of the given year, and the price index will be the result of raising the coefficients to the exponential power. I.e. the coefficients got by model calculus, connected to the year-dummies, express the price change compared to the reference year.

The hedonic method gives us a more accurate measurement than the more simple indicators (averages) if we succeed to involve variables into the analysis that help tackle the distortion of the composition effect. As it is difficult to describe several qualitative parameters with a quantitative indicator, or there is no indicator available, it is common to include proxy variables among the explanatory variables. (For instance, the address is not a hedonic qualitative feature, still it contains important information. It does make a difference whether a flat or house is in the city center or in the suburb of Budapest.) The reason these variables are important is, because the composition effect mostly arises from the change of the settlement distribution of purchase transactions. Thus with the help of variables drawn into the regression, it can be filtered out if in some periods there were more transactions from Budapest than at other times. The set of variables was formed so that it handles the geographical, settlement size-related, property type-related and flat size-related instability of the composition. Where data are available, besides these it is possible to correct the composition effect resulting from the instability of the year of construction, construction technology, and the location inside the settlement.

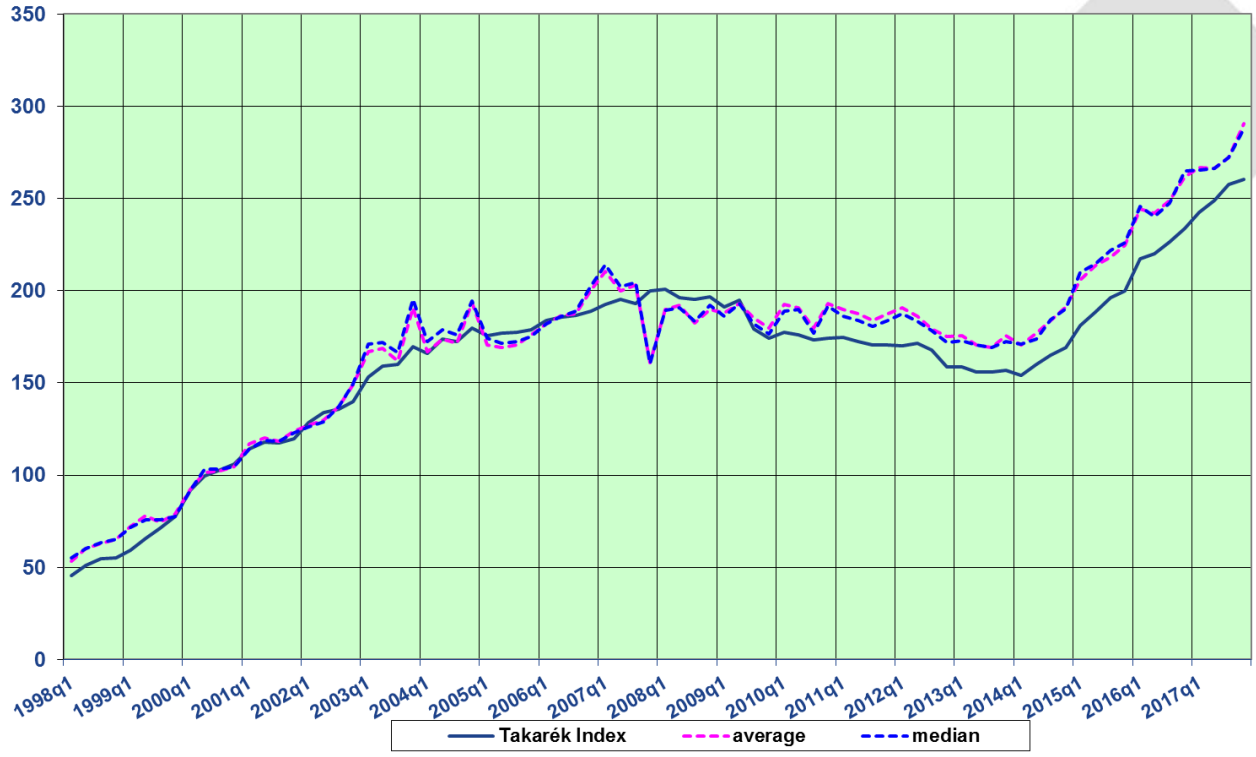
The values of Takarék House Price Index were computed by making sample parts that comprise 5 consecutive quarters from the complete database containing 46 quarters. That means consecutive sample parts overlap for 4 quarters. The above regression was estimated using these sample parts with 5 quarters, and the value of the index was always calculated based on the estimation, in the sample of which the given quarter is the last datum. This method ensures that after one and a half years, we can seal the past, as from now on the new data do not modify the previously estimated index values any more.

4. Conclusions, remarks

In the following, we are going to share some observations that we deem instructive, or may answer some occurrent questions.

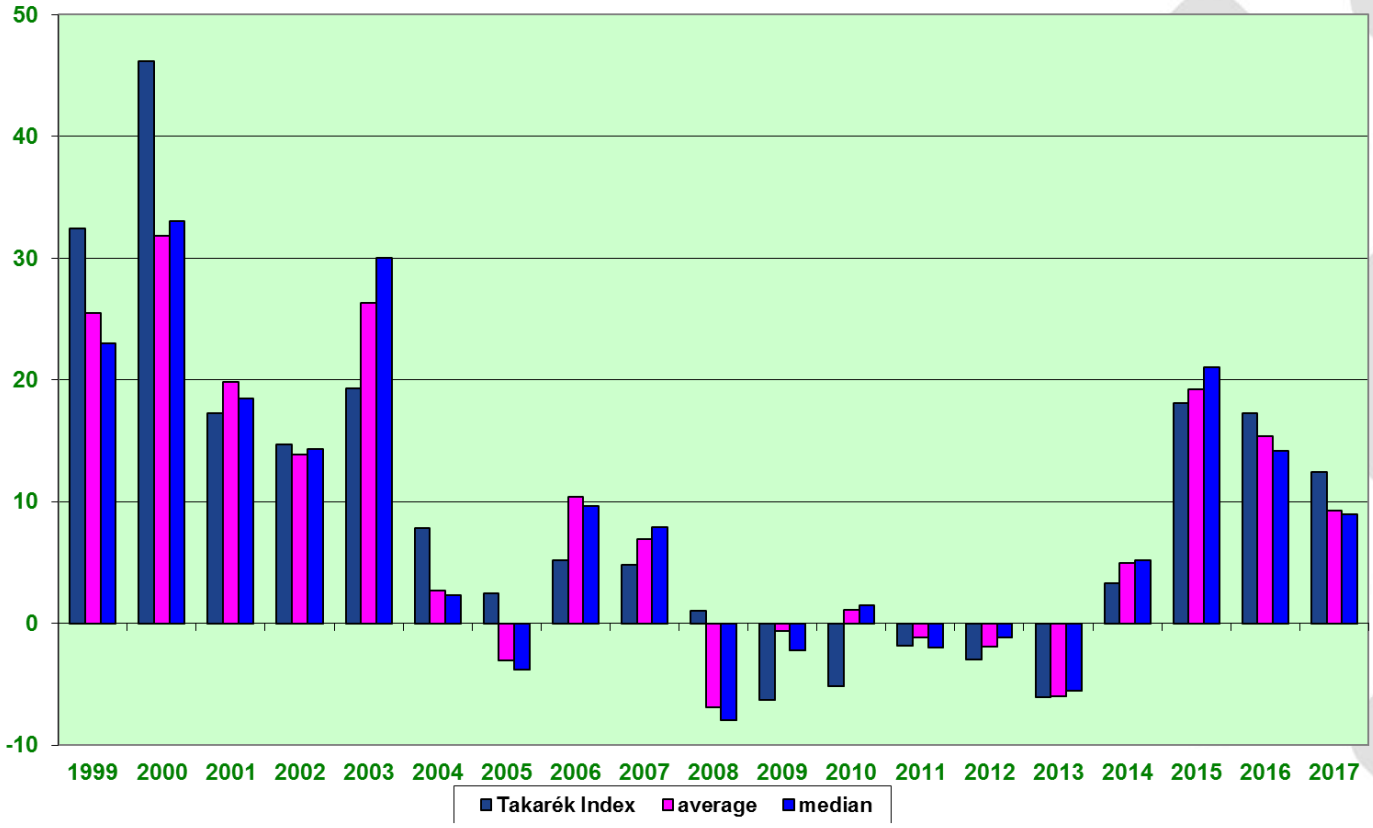
First, through two figures we illustrate why we chose the hedonic method for the index. *Figure 1* features the run of the hedonic index, the mean and median price of the sample. The trends of simpler methods and Takarék House Price Index calculated with the hedonic method are similar, they share the same conclusions on the long run. The figure also demonstrates that on the short run, simpler indicators have bigger volatility, while the index produced with the regression method has a smoother run, i.e. changes in the composition of sold properties, which is corrected by the hedonic method, may have a role in short-haul volatilities.

Figure 1: run of Takarék House Price Index and simple indicators (2000=100)



In *Figure 2*, yearly changes of the previous indicators can be seen.

Figure 2: change of Takarék House Price Index and simple indicators compared to previous year (%)



For us, the difference between 2000 and 2003 was especially instructive. During these years, the mean prices rose similarly, while the value of the hedonic index rose significantly more in 2000. Having analysed the data, it turned out that the number of flats in Budapest – which are relatively more expensive – was much lower in the database in 2000 than usual. As a result, although house prices increased significantly both in the countryside and in Budapest, there were relatively more properties from the countryside (cheaper properties) among the observations, and this change in the composition decreased the rise of the average price. The biggest differences are to be seen in data from 10 years ago, as in this time period the supplying of data was even more incomplete than today.

As a second conclusion, we would like to mention that having analysed the seasonality of the index, we found that the index contains minimal seasonality, so we decided to publish seasonally **un**adjusted data. *Figure 3* shows what a result the seasonal adjustment with the Tramo-Seats method brings.

Figure 3: run of Takarék House Price Index and its seasonally adjusted version (2000=100)