MEASURING HETEROGENEITY OF HOUSE PRICE DEVELOPMENTS IN HUNGARY, 1990–2016

Ádám BANAI – Nikolett VÁGÓ – Sándor WINKLER

This study presents the detailed methodology of generating house price indices for the Hungarian market. The index family is an expansion of the Hungarian housing market statistics in several regards. The nationwide index is derived from a database starting from 1990, and thus the national index is regarded as the longest in comparison to the house price indices available so far. The long time series allow us to observe and compare the real levels of house prices across economic cycles. Another important innovation of this index family is its ability to capture house developments by regions and settlement types, which sheds light on the strong regional heterogeneity underlying the Hungarian housing market.

Keywords: housing market, house price index, hedonic regression, Hungary

JEL classification indices: C430, R210, R310

Ádám *Banai*, Director at Magyar Nemzeti Bank (MNB, the central bank of Hungary).
E-mail: banaia@mnb.hu

Nikolett *Vágó*, corresponding author. Senior economist at Magyar Nemzeti Bank (MNB, the central bank of Hungary). E-mail: vagon@mnb.hu

Sándor *Winkler*, Senior economist at Magyar Nemzeti Bank (MNB, the central bank of Hungary).
E-mail: winklers@mnb.hu

1. INTRODUCTION AND MOTIVATION

Housing market is organically connected to every part of the economy and has a potentially large-scale impact on each area. As residential property is one of the most important assets of households, changes in house prices are likely to affect households' consumption and saving decisions, while having a social aspect through affecting housing affordability. Similarly, the real estate market developments directly affect the business sector. Price developments and the number of transactions affect demand for the new investment projects and ultimately, the construction industry. Finally, developments in the real estate market also exert a direct impact on the banking sector. Changes in the prices of the real estate collaterals behind mortgage loans may not only determine the performance of the loans, but the recovery through the sale of collaterals in the event of default. Beside the outstanding portfolio, an active real estate market boosts the demand for housing loans and affects the banking sector through a larger volume of new disbursements. Compared to the corporate loans, the banking sector can earn larger spreads on the mortgage loans, and the intensifying activity has a positive impact on profitability.

In addition to the factors described above, understanding the trends and identifying the risks arising in the real estate market are of key importance for the central banks. The house price index is designed to serve this particular purpose, providing insight into the processes in the real estate market and its individual segments.

In Hungarian practice, two indices have been in use so far: those published by FHB Bank and the Hungarian Central Statistical Office (HCSO). Although both indices provide a fair view of the domestic housing market developments, the construction of additional house price indices was necessary for Magyar Nemzeti Bank (MNB, the central bank of Hungary) from a financial stability perspective. Having been available since 1998, the FHB index is typically published with a considerable lag (5–8 months) and it is constructed on the basis of a sample that covers only around 50 per cent of all transactions. The HCSO launched its own house price indices in 2007, with a primary focus on presenting the differences in the dynamics of new and used property prices.

Compared to these two indices, the MNB house price index family[1] presented in this study represents a step forward in several regards. (1) The index developed by the authors has been constructed on the longest and most comprehensive time

---

[1]   Hereinafter, we refer to the indices developed by the authors as "MNB house price indices", because these indices are used for exploring the house price trends and identifying the financial stability risks, while these are also the official statistical house price indices of the MNB.

series: it captures the developments of Hungarian house prices on a national scale starting from 1990. This is important because both the modelling of house prices and the assessment of housing market developments require the longest possible time series. (2) Starting from 2001, the index sheds light on the heterogeneity of house prices across regions and municipality types. By contrast, the previous indices provided an overall view of the country as a whole, only obscuring the different behaviour patterns of individual regions. From a business, banking and central banking perpektive, however, it is essential to be aware of housing market heterogeneitis between regions and municipality types. Even when the national price movements do not indicate any problems, individual regions can be exposed to potentially harmful developments. The extreme spikes in house prices or housing market bubbles typically arise in the capital cities or larger municipalities, which can be largely attributed to the central role of major cities in the economy and their more advanced infrastructural and institutional coverage.

In Section 2, we provide an overview of the traditionally applied methods in the construction of house price indices. Section 3 describes the transaction data on which the price indices are based. Section 4 is dedicated to the methodology used for the construction of our indices. In Section 5, we outline our regression results and present the derived house price indices. In Section 6, we conduct a robustness analysis for the methodology of the indices. Finally, Section 7 provides conclusions.

2. LITERATURE REVIEW

The calculation of changes in house prices aims to determine the average percentage change of residential property in a specific area during the review period. The actual market price of residential properties evolves during the sale and purchase of the properties; consequently, the housing market turnover provides a suitable basis for measuring the changes in house prices. Although the residential property prices can also be determined by appraisal, the comprehensive and regular appraisals of the stock of dwellings are scarce even by the international standards, primarily because of the high costs of data collection.[2] Consequently, the changes in house prices are typically computed from transaction prices.

Ideally, for the calculation of average price changes, each dwelling should change owners in each review period (e.g. year); only in that case would the market price data be available on each individual property. In reality, however,

---

[2]   House price indices in New Zealand, Denmark, the Netherlands and Sweden are constructed from stock appraisals.

only a fraction of the dwelling stock is sold in a given location in each period, and the diverging composition of the transactions executed in the given period poses problems. Since the dwellings sold may represent different types and quality in each period, changes in the average or median transaction prices do not reveal meaningful information about the average or typical change in the value of the total housing stock. In order to receive information explicitly about the price changes, we should observe the trading of the houses that have the same quality and attributes. The house price indices are essentially designed to achieve this goal: controlling for the quality traits of the dwellings, they capture the average "pure" price change in the residential properties.

While computing the price index from transaction data, it should be borne in mind that despite controlling for the effect of diverging transaction composition of individual periods, the index remains susceptible to the types of dwelling traded in the given period and to the circumstances that certain locations are over-represented in the market turnover in a given country, relative to the distribution of the total housing stock. The availability of detailed regular statistics on the composition of the housing stock may help eliminate this problem. Based on such statistics, new weights can be assigned to transaction data that are not representative of the overall housing stock. In the absence of such information, the evaluation of the indices derived from the transaction data should factor in the constraints mentioned above. Finally, it should be noted that statistical offices typically rely on transaction data in the compilation of house price indices; consequently, the above mentioned limitations are also present in the international house price statistics.

The main difference between the construction methods of house price indices arises from their treatments of the bias, stemming from the compositional shifts observed in the consecutive periods. In the following section, we provide an overview of the possible house price index calculation methods based on the classification presented in the handbook issued by Eurostat (2013), turning more attention to the hedonic regression methods used also for the house price indices of this study.

Hedonic regression models

The estimation of hedonic regression models is the most widely used method for calculating house price indices. The basic assumption of this approach is that the residential property prices can be determined as a function of their individual characteristics, and therefore the use of explanatory variables expressing the characteristics of the house under a linear regression model can control for the

bias arising from the composition effect. The technique of hedonic regression estimates dates back to Court (1939) and Griliches (1961), while the conceptual bases of the method were laid down by Lancaster (1966) and Rosen (1974). In Hungary several authors have alredy analyzed hedonic regression models on domestic data, Horváth (2007) compared the hedonic regression models with other methods on a sample of advertisement data for a homogenous residental area of the capital city of Hungary, Budapest. Horváth–Székely (2009) presented a house price index for used homes calculated by a hedonic regression modell on transaction data obtained from the tax authority. Kutasi–Badics (2016) compared a hedonic regression and an artifical neural network modell in terms of market price prediction for the Budapest market using a large set of advertisement data.

From the observations of price changes, the hedonic regression method attempts to strip out the composition effect arising from the trading of real properties with different characteristics across periods by including the following variables: floor area of the structure, size of the land (for single detached dwelling units), the characteristics of its environment, age and type of the dwelling (e.g. detached house, row house, condominium, etc.), the materials used in the construction of the house, and the internal characteristics of the dwelling (e.g. number of bedrooms and bathrooms, energy efficiency, etc.). The hedonic regression models are traditionally estimated by the method of ordinary least squares (OLS). Based on the time horizon of the estimate, three main types of hedonic modelling can be distinguished: the *time dummy variable method,* the *adjacent-period approach* and the *multiperiod time dummy method.*

In the case of the *time dummy variable approach,*[3] a pooled OLS estimate is prepared based on the data of all periods. Except for the initial base period, the model uses separate dummy variables to capture price changes between each period as the price index is produced by exponentiating the time dummy coefficients. The regression equation can be written as:

$$log\ y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \sum_{t=2}^{T} \delta_t d_{ti} + \varepsilon_i,$$ (1)

where y is the price of the house, x denotes the characteristics of the house, d_t is the time dummy for period t, β expresses the coefficients of the control variables, δ_t means the coefficients of the time dummies and ε is the residual value. One potential drawback of this approach is that the coefficients of the model's explanatory variables are constant over time, and if the characteristics of the property exert a different impact on the property price over time, the index will be subject to bias. Moreover, it might be problematic from a practical point of view that a

---

[3]   The model was originally developed by Court (1939).

new estimate must be prepared for the entire time horizon in each period; consequently, the entire time series of the price index will be subject to revision in each period.

Under the *adjacent-period model,* estimates are produced for the observations of two consecutive periods. In practice, this technique is a restricted form of the previously described time dummy variable method. If T means the number of all review periods, then a total of $(T-1)$ estimates must be run in order to receive the price index. Since the estimation samples include the observations of two periods, each regression equation includes a single time dummy. The period-to-period price index can be computed by exponentiating the time dummy coefficients. The estimated regression for period t can be written as:

$$log \; y_i^t = \beta_0^t + \beta_1^t \, x_{1i}^t + \beta_2^t \, x_{2i}^t + \ldots + \beta_k^t \, x_{ki}^t + \delta^t \, d_i^t + \varepsilon_i^t \; , \qquad (t = 2, 3, \ldots, T) \; (2)$$

where y is the price of the house, x denotes the characteristics of the house, d is the time dummy, β expresses the coefficients of the control variables, δ represents the coefficients of the time dummy and ε is the residual value. Since a separate estimate is produced for each adjacent-period pair, the greatest advantage of this method is that the estimated coefficients of the explanatory variables can change over time. Therefore, this method eliminates the underlying assumption of the time dummy variable model, i.e. the parameters are constant over time. This is consistent with the assumption that the demand and supply conditions can change over time with respect to the specific characteristics of the properties. Compared to the time dummy variable model, the downside of this approach is the far smaller sample size on which the estimate can be run. Consequently, this approach is only recommended if a sufficient number of observations are available for each adjacent-period pair. That notwithstanding, the method is considered to be advantageous from a statistical standpoint, as the previous elements of the time series are not subject to revisions again and again during the estimation of the additional index values.

In the *multiperiod time dummy approach,*[4] the hedonic regression model must be estimated separately for each individual period. The calculation of the price index requires the definition of a "benchmark property", and the house price index is defined as the price change of the benchmark property. For the computation of the pure price change, therefore, we need to determine the values of the property

---

[4]   This methodology was applied, among others, by Crone – Voith (1992), Knight et al. (1995), and Gatzlaff – Ling (1994); however, the authors used different terms to refer to this particular type of hedonic modelling.

characteristics included in the model. The regression equation for period t is the following:

$$\log y_i^t = \beta_0^t + \beta_1^t x_{1i}^t + \beta_2^t x_{2i}^t + \ldots + \beta_k^t x_{ki}^t + \varepsilon_i^t \,, \qquad (t = 2, 3, \ldots, T) \qquad (3)$$

where y is the price of the house, x denotes the characteristics of the house, β expresses the coefficients of the control variables, and ε is the residual value. This method should be preferred to the previously described approaches as the assumption is that the property characteristics captured by the explanatory variables of the model can change not only from half-year to half-year but also from quarter to quarter. It is problematic, however, that the definition of the "benchmark property" can be ambiguous (e.g. the typical property of the initial or the previous period), and the price index received largely depends on the benchmark property. In many cases, the Fisher price index is used, which is defined as the geometric average of the Laspeyres price index (which is computed from the average property characteristics of the base period) and the Paasche price index (which is weighted with the average property characteristics of the current period). Another disadvantage of this model is that its estimates may render even more uncertainty by the lack of sufficient observations for each period.

It should be noted that the risk of multicollinearity may arise in all three types of the hedonic models. A high correlation between the explanatory variables increases the standard errors of the coefficients, and some variables may become insignificant. Since the estimated coefficients, which are unbiased even in the event of multicollinearity, bear the most relevance for the house price index using as many variables as possible should be considered, hence reducing the risk of bias arising from the missing variables.

One possible extension of the methods drawing on hedonic regressions is the calculation of stratified indices. Sub-samples are separated according to some criteria relevant to the analysis (e.g. regions, municipality types). The hedonic models may be estimated for each individual sub-sample even according to different specifications, resulting in a separate sub-index for each sub-sample. With proper weighting, the sub-indices may aggregate up to a consistent house price index. One of the benefits of the approach is that the compilation of sub-indices supports the analysis of house prices separately for each sub-sample. In addition, since a separate estimate is prepared for each sub-sample and the coefficients of the explanatory variables may vary, we can factor in the fact that the characteristics of the reviewed properties included in the model may exert a different impact on the prices of the individual sub-samples. For the sake of reliability, each sub-sample should include a sufficient number of observations.

Other methods

The *stratified sample mean method* compiles the index based on average price changes computed within homogeneous groups that are created on the basis of various price determinant attributes. An aggregate house price index is built by taking the weighted average of the values computed for different groups/strata. The advantages of the method are that it is easy to apply and explain to users, and that the sub-indices can be interpreted independently. As a significant drawback, however, it is difficult in practice to truly control for the composition effects mentioned above.

The *repeat sales method* was first proposed by Bailey et al. (1963). Besides the Federal Housing Finance Agency (FHFA), Standard & Poor (2009) compiles a house price index based on this method for 20 cities in the United States. This method only considers price changes in those properties that have been sold more than once over a specific time horizon. The main advantage of the model is the irrelevance of control variables in the estimate; the only bias that may arise is due to the depreciation of the dwelling units or their renovation induced appreciation. The method can be applied efficiently if the sample of dwellings sold more than one time in the reviewed real estate market and period is high enough. Since the estimate covers the entire review period, the entire model must be re-estimated in each case, and therefore the price index is subject to revisions in each individual period. Beyond the models presented above, there are *combined methods* that use the hedonic regressions and the repeat sales method in conjunction with one another in order to maximise their benefits. In practice, however, such techniques are hardly used due to the complexity of the model and to the relatively minor observed improvement in efficiency.

The data requirements of the methodologies described above are different. The data available to us so far (presented in Section 3) are primarily suitable for the purposes of the hedonic regression models.

<div align="center">

3. DATA

</div>

The changes in house prices can be statistically measured by using two types of data: (1) the value data of the stock of dwellings, or (2) the transaction data evolved during the sales and purchases of the properties. The former are typically derived from the real estate property appraisals. Internationally, there are only a handful of examples of data being regularly released on the value of residential properties pertaining to the total stock. The latter are traditionally collected by tax authorities in relation to property transactions. Therefore, statisticians tend to rely

on the transaction data for the compilation of house price indices. We compiled the house price indices based on the property acquisition duty data collected by the Hungarian National Tax and Customs Administration (NTCA).

Aside from the current demand and supply conditions, house prices are determined by attributes falling into two main groups: (1) the characteristics of the residential property itself, i.e. the quality of the property, and (2) the location of the property, i.e. the characteristics of its environment and location. The data collected by the NTCA include the most basic information related to the sales of the properties. In addition to the sale price, information is available on the property's net internal area (NIA), the exact location and the type of the property (e.g. detached house or flat) and, starting from 2008, on whether the home is newly built or not. Information on the condition and qualitative characteristics of the residential properties is of insufficient quality and quantity. Overall, the data collected by the NTCA are mainly suitable for explaining the price of the properties with their NIA and location. With respect to the variables stored in the database, there are few differences[5] in the data collected before and after 2008, which we addressed by defining different model specifications for before and after 2008.

For our purposes, we also retain those transactions in the database where the party acquiring the property is a business organisation. These observations make up only about 7 per cent of the post-2008 data. In addition, since the types of the residential properties changing hands in these cases are similar to those traded in the transactions of private individuals – i.e. they also constitute an integral part of the housing market turnover and the market itself – their inclusion in the estimate appears to be warranted.

With all unusable observations stripped out, the NTCA duty database currently contains information on around 3.1 million property transfers between 1990 and 2016 Q2. The variables of the database and the list of municipality-level variables linked to the database and included in the estimations are illustrated in *Table 1*. Information on the NIA of the real estate is essential for determining the price of the properties. One of the greatest deficiencies of the database is the fact that in some cases, the value of the useful NIA is either incomplete or zero. In order to prevent the loss of an inordinate amount of data, this incomplete NIA data must be back-cast. The precise methodology of this exercise is explained in Subsection 4.1. We need to stress that the Budapest sales data have only been available in the NTCA duty database from 2001, which should be borne in mind during the assessment of the pre-2001 house price index values.

As the NTCA database contains limited information on the characteristics of the dwellings, numerous location-dependent variables have been included in the

---

[5]   For a detailed description of the deficiencies in the data see Banai et al. (2017).

Table 1. Definition of the dataset and the individual variables used in the estimation

Source	Variable	Description
NTCA duty database	Price (ln) (HUF)	The price of the real estate is the dependent variable of the regressions. It is the larger among National Tax Authority's valuation and the price in the transaction contract. The variable is in a logarithmic form.
	Quarter	Quarter of property acquisition duty.
	Type of property * Size of property (ln)	The net internal area (NIA) is in the regressions by categories of the type of dwelling. The category variable "type of dwelling" can have the following values: family house in inner and outer districts (in the case of Bp.), family house in county seat and other cities (in the case of cities), condominium, panel block of flats and homestead. In case of data before 2008 there are only two categories: family house and flat.
	New flat	Category variable: new or used property.
Variables based on the HCSO id of the settlement	Districts	Category variable: districts of Budapest.
	Agglomeration	Category variable: 8 districts distinguished: agglomeration of Szeged, Pécs, Debrecen, Miskolc, Székesfehérvár, Budapest, Győr and Sopron.
	Recreational area	Category variable: 7 seasonal property areas distinguished: Lake Balaton - near shore, Lake Balaton - other, Dunakanyar, Mátra-Bükk, Sopron-Kőszeghegyalja, Lake Tisza, Lake Velence - Vértes.
	County	Category variable: county of the settlement.
TSTAR database	Population (ln)	Population at the end of the year. The variable is in a logarithmic form.
	Size (ln)	Size of the municipality. The variable is in a logarithmic form.
	Local home support (ln) (HUF)	Amount of local housing subsidies. The variable is in a logarithmic form.
Geox database	Distance from Budapest (ln)	The shortest distance from Budapest expressed in minutes. The variable is in a logarithmic form.
	Distance from county seats (ln)	The shortest distance from the county seat expressed in minutes. The variable is in a logarithmic form.
NTCA PIT database	Total income per capita (ln) (HUF)	Net labour income per capita. The variable is in a logarithmic form.

Note: The Geox database contains the location of Hungarian municipalities relative to specific nodes and centres (e.g. distance from Budapest or from the nearest highway node). The distances are expressed both in time and kilometres. The TSAR database is maintained by the HCSO and contains comprehensive information on Hungarian municipalities (e.g. demography, institutional coverage, tourism, etc.). Inner districts in the Budapest model: I, II, III, V, VI, VII, VIII, IX, XI, XII, XIII, and XIV.

data used for the compilation of our index. We selected these explanatory variables based on two main criteria: on the one hand, we tried to include the variables that did not correlate strongly; on the other hand, in selecting the variables we focused on indicators that may carry significant additional house price information compared to the rest of the control variables. In addition, we obviously wanted to be certain that the sign of the explanatory variables included in the models is economically intuitive.

The location of the property can influence the sale price significantly. Larger municipalities, regional centres or the centres of smaller geographical units typically have better infrastructural and institutional coverage, and such factors may increase the appeal of these municipalities and, hence, the housing market of the area drives up the local house prices. Another important factor in the assessment of a residential property is its temporal distance from geographically key locations and nodes or whether it is located in municipal agglomerations or seasonal property areas. Smaller municipalities located closer to motorways with easy access to larger municipalities are more attractive than hard-to-access locations. Additionally, we use the size (km^2) and population of the municipalities, the total amount of local housing subsidies and the net per capita income of the municipalities for the modelling of house prices. The descriptive statistics of the data used for the estimation of the house price indices are included in the *Annex* (*Table 6*).

4. THE METHODOLOGY OF THE MNB'S HOUSE PRICE INDEX

4.1. The backcasting of useful NIA

One of the deficiencies of the database is the insufficient information available on the NIA of the residential properties. Depending on the period, data are unavailable for around 30–40 per cent of transactions. The database also includes a variable referred to as the "property area" with a far higher, nearly 100 per cent availability; however, as to whether this information pertains to the area of the structure or to the plot of land varies for each observation and cannot be explicitly determined. We supply the missing values of the "useful NIA" variable from the appropriate values of the "property area" variable, and backcast the missing values with a regression method. In brief, we use the following method for the definition of the NIA:[6]

---

[6]   This procedure is consistent with the HCSO's treatment of incomplete or incorrect NIA values; the two methods differ from one another only in respect of the details of the regression estimate.

1. We considered all "useful NIA" data under 15m^2 and over 500m^2 to be missing parameters.
2. Wherever the "useful NIA" is specified as "missing" or zero and the "property area" is under 150m^2, we consider the "property area" to be the "NIA".
3. For observations where the "NIA" remains "missing" or zero even after the first two steps, we use a regression method to backcast the size of the "NIA".

Table 2 indicates the percentages of the observations by municipality type that must have been backcasted with the assistance of the linear regression method described in Section 3. Evidently, the missing NIA information mainly affects villages.

Table 2. Percentages of missing useful NIA information by year and by municipality type, %

	Budapest	Cities	Villages	Total
1990–2007	4.2	20.3	53.4	25.0
2008–2015	7.4	25.6	63.1	28.7

We used the ordinary least squares method to estimate the linear regression model fit to the logarithm of the NIA variable. We prepared an estimate for each municipality type (Budapest, cities and villages) both for the pre-2008 and post-2008 samples. We divided the database into sub-samples representing the old and the new structures in order to use the broadest possible information base for both of these periods, while breaking down the backcasting by municipality types was warranted by the significantly diverging distributions of property size in the individual municipality types.

4.2. Methodology of outlier filtering

The linear regression models can be sensitive to the outliers[7] in the database. Firstly, the database might contain observations that are most likely incorrect, for example, owing to measurement or recording errors. Secondly, although some outliers may reflect existing processes, we should consider discarding them from the sample anyway, as they may significantly distort the estimate and the resulting price change. This underpins the significance of outlier filtering; indeed, discarding the extreme and influential data points may improve the accuracy of the regression estimate and the reliability of the conclusions drawn from the results.

---

[7]   Zrínyi et al. (2012) cite numerous outlier definitions from the academic literature.

With that in mind, we opted for a multi-step outlier filtering technique. In the first step, we tried to strip out the incorrect data points by defining absolute bounds for the main variables of the duty database (adjusted for the consumer price index for each year in the case of price-type variables). Observations were removed from the estimation sample if:[8]

- the sale price was lower than HUF 100,000 or higher than HUF 1 billion;
- the NIA of the dwelling was smaller than 15m^2 or larger than 500m^2;
- the unit sale price per square metre was below HUF 2,000 or above HUF 10 million.

In the second step, we also performed filtering for statistical purposes. We estimated a regression equation (1) for the house price on the data points that were deemed correct based on the first step, and calculated the following four indicators to identify outliers and influential values:

1. Externally studentized residuals[9]:

$$r_i = \frac{y_i - \widehat{y_{(i)}}}{\sqrt{MSE_{(i)}(1 - h_{ii})}},$$

where y_i is the observed price for the *i*th observation, $\widehat{y_i}$ is the estimated value for the *i*th observation, $MSE_{(i)} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i}),^2$ is the mean squared error excluding the *i*th observation and h_{ii} is the leverage.[10]

For each observation, the indicator examines the standard error adjusted value of the estimated residuals (deviation between observed and estimated values). The computed residuals are not homoscedastic (their variance is different), and the observations with the greatest leverage have residuals with the smallest variance, which is addressed by the $\sqrt{1 - h_{ii}}$ term in the formula. The mean squared error included in the indicator is derived from a regression that does not include the reviewed *i*th observation. This is useful because the price estimate will be defined on the basis of the coefficients that are not skewed by the *i*th observation

---

[8]   The values provided for the sale and the unit price refer to 2015 Q4; thresholds for all other periods are received after adjustment for the consumer price index.

[9]   The indicator of externally studentized residuals is explained in detail by Belsley et al. (1980), Vellemen – Welsch (1981), Chatterjee – Hadi (1986), and Bollen – Jackman (1990).

[10]   Leverage is a measure of how far a data point deviates from the mean of the explanatory variables. In other words, it is a diagonal element of the hat matrix that shows the leverage exerted by the *i*th observed price (y_i) on the *i*th estimated value ($\widehat{y_i}$).

even if it is deemed to be an outlier; consequently, the indicator will not mistaken-ly yield price observations and price estimates that are too close to each other.[11]

2. Cook's distance[12]:

$$CD_i = \left(\frac{\widehat{y_i} - \widehat{y_{(i)}}}{\sqrt{MSE_{(i)}(1 - h_{ii})}} \right)^2 \frac{h_{ii}}{1 - h_{ii}} \frac{MSE_{(i)}}{\sqrt{MSE}} \frac{1}{p}$$

where $\widehat{y_i}$ is the estimated price for the ith observation, $\widehat{y_{(i)}}$ is the estimated value for the ith observation calculated from the coefficients of a regression obtained after the removal of the ith observation, p is the number of explanatory variables in the regression, MSE is the mean squared error, $MSE_{(i)}$ is the mean squared er-ror obtained after the removal of the ith observation and h_{ii} is the leverage of the ith observation.

As opposed to the studentized residual indicator that concentrates on the re-sidual values, this indicator focuses on the dependent variable estimates obtained with and without the ith observation. In addition to the dependent variable, the effect of explanatory variables appears indirectly in the calculation through the leverage. A high value of the indicator suggests that the observation has a signifi-cant effect on the size of the estimated regression coefficients.

3. Welsch distance[13]:

$$WD_i = \frac{\widehat{y_i} - \widehat{y_{(i)}}}{\sqrt{MSE_{(i)}(1 - h_{ii})}} \frac{\sqrt{h_{ii}\ (n - 1)}}{1 - h_{ii}}$$

where $\widehat{y_i}$ is the estimated price for the ith observation, $\widehat{y_{(i)}}$ is the estimated value for the ith observation calculated from the coefficients of a regression obtained after the removal of the ith observation, n is the number of elements in the sam-ple, $MSE_{(i)}$ is the mean squared error obtained after the removal of the ith observa-tion and h_{ii} is the leverage of the ith observation.

Similar to the previous indicator, this indicator measures the effect of the given observation on the estimated values; however, it uses a different normalisation and is more sensitive to the observations with high leverage. With similar outliers in the database, this indicator could be more efficient in identifying observations that are to be discarded.

---

[11]   It should be borne in mind that with several similar outliers in the database, the regression received after the removal of individual observations may be very similar to the regression that includes all of the observations.
[12]   For more detail, see Cook (1977), Hair et al. (1995) or Bollen – Jackman (1990).
[13]   For more detail, see Welsch (1982) or Chatterjee – Hadi (1986).

4. DFBETA[14]:

$$DFBETA_{ij} = \frac{b_j - b_{j\,(i)}}{\sqrt{MSE_{(i)}\,(\mathbf{X'X})^{-1}_{jj}}}$$

where b_j is the *j*th element of the vector of the coefficients, $b_{j(i)}$ is the *j*th element of the regression that results when the *i*th observation is removed, $MSE_{(i)}$ is the mean squared error computed after the removal of the *i*th observation and $\mathbf{X}$ is the matrix of the explanatory variables included in the linear regression model.

The indicator measures the sensitivity of the coefficients (deviation adjusted for variance) estimated with and without the *i*th observation for a randomly selected explanatory variable. Since the period's dummy coefficient bears most relevance to the price index in the case of adjacent-period estimates, one of the benefits of the indicator is its ability to specifically measure the effect of the given observation on this particular coefficient.

In selecting the indicators, we tried to limit the overlaps between them to the minimum. According to the literature, based on the sample size and the number of the explanatory variables, each indicator can be used to identify the values that can be considered as outliers or influential based on the specific indicator. An observation will be included in the final estimation sample if it is deemed valid by at least 3 of the 4 above mentioned indicators.[15] The results confirm that the selected indicators offer significant additional information relative to each other.

We identified the percentage of the observations that should be filtered out in the two steps, both for regions and for municipality types. As a result, we initially removed about 1 per cent of the observations from the estimation sample and in the second step 4–5 per cent of the observations were dismissed.

The effect of the filtering performed in the individual steps on the aggregate index is illustrated by *Figure 1*. We found that despite the far smaller number of observations excluded in the first step, these observations exert a far greater impact on the price index than those removed from the estimation sample in the second step. This is because in the first step we strip out obviously incorrect data points at the outer edge of the distributions.

4.3. Methodology of regression estimation

In consideration of the characteristics of the available database and the advantages and disadvantages of the specific methodologies as described in Subsection 3.1,

---

[14]   For more detail, see Belsley et al. (1980) or Bollen – Jackman (1990).
[15]   We performed a robustness analysis for this criterion, which is discussed in Section 6.
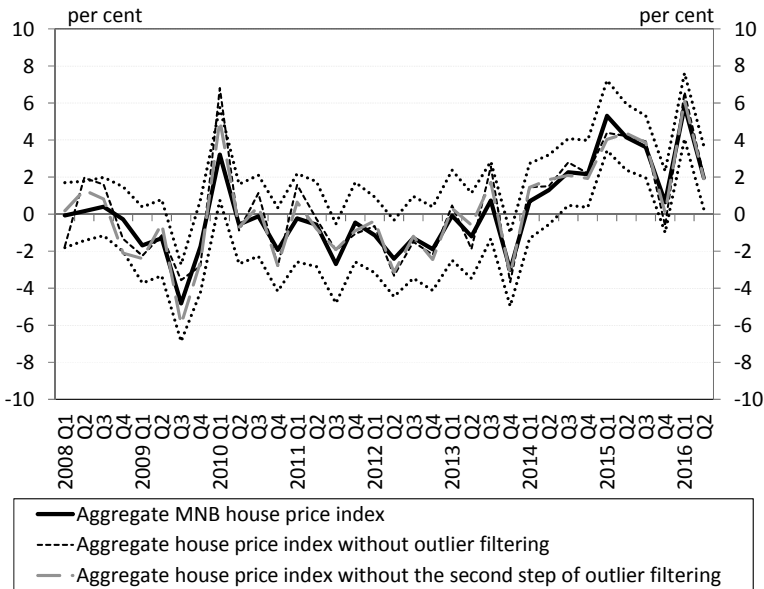
Figure 1. Nationwide MNB house price index with various outlier filtering procedures
(quarterly changes)

Note: The black dotted line indicates the 95 per cent confidence interval* of the MNB's aggregate house price index.

* The confidence interval of the aggregate national index was produced by weighting the 95 per cent confidence intervals of the quarter dummy parameters included in the models for individual municipality types with transactions.

the estimation of a hedonic regression model for adjacent periods appears to be the most appropriate method. We prefer the adjacent-period estimate to the multiperiod time dummy method because one of our key objectives was to examine the diverging characteristics of the individual municipalities and the regional processes separately. For this exercise, we need to divide the database into subsamples, but the number of observations in the sub-samples are insufficient to run a reliable estimate for each individual period.

The adjacent pair estimation procedure has numerous advantages over the time dummy variable approach. On the one hand, as a result of estimating a separate regression equation for each adjacent-period pair, the partial effects exerted on the house price by the model's control variables capturing the characteristics of the dwelling can be different over time, which is a more flexible approach and a better fit to economic intuition compared to the assumption of fixed effects irrespective of the periods. On the other hand, with the adjacent-period estimate we can compile a consistent house price index for the longest possible time horizon,

while using the broadest possible information base for all periods. An estimation for the entire period would entail a loss of information: it is the specificity of our database that there is less information available on each transactions until 2008, which would restrict the range of explanatory variables in the models from 2008. It is another important factor that the national index estimated for the pre-2001 period (for which no sub-indices can be produced due to the insufficient number of transactions) can be consistently added to the national index compiled with the adjacent-period method for the period between 2001 and 2016 from the sub-indices created for each region and municipality type. This would not have been possible using the time dummy variable approach. Finally, the time series resulting from the adjacent-period estimate is not subject to revisions for methodological reasons[16] with the release of new data.

4.4. Disaggregated indices

After cleaning the database, we divided it into sub-samples based on the legal status and region of the municipalities, and conducted the adjacent-period hedonic regressions separately for each sub-sample. According to the settlement types, we produced indices for Budapest, cities and villages. We defined the regional decomposition based on the number of observations available. As a result, we compiled 7 region-level city indices and produced the national house price indices for cities by weighting the city indices with the number of transactions. In the case of villages, the low number of observations prevented us from preparing reliable estimates for each region, and consequently, we do not compile the regional price indices for villages. Similar weights based upon the number of transactions are applied to receive the national quarterly price change from the quarterly price changes of homes in Budapest, cities and villages. Finally, we construct the aggregate house price index by chaining the previously received quarterly indices.

Since the number of real estate market transactions was significantly lower in the 1990s in Hungary, it was only from 2001 that the 9 disaggregated house price indices described above could be produced reliably. However, the adjacent-period estimate allows us to link the national-level quarterly price changes derived from the adjacent-period models for the period of 1990–2001 to the national aggregate index resulting from the weighting of the previously described disaggregated indices.

---

[16]   Irrespective of the estimation methodology, the values of the house price index are routinely subject to revisions because a certain part of the property transaction data used for the calculations are made available with a significant lag.

5. PRESENTATION OF THE RESULTS OF THE MNB'S HOUSE PRICE INDEX

5.1. Time series of the MNB's house price index

In this section, we present the house price indices estimated for the individual sub-samples and the MNB's aggregate house price index. Because of the limitations of our database described above, for the period before 2001 we constructed only a national index, while indices decomposed by the settlement types and regions (for cities) start from 2001. *Figure 2* illustrates the MNB's aggregate nominal and real house price indices on a long time series. Since the database does not contain observations on transactions executed in Budapest before 2001, an aggregate house price index for the period after 2001 was estimated without taking into account the observations on residential properties located in Budapest. There is a significant difference between the two nominal time series only after 2014, which suggests that the house price dynamics before 2001 is properly described by the aggregated MNB house price index without Budapest.
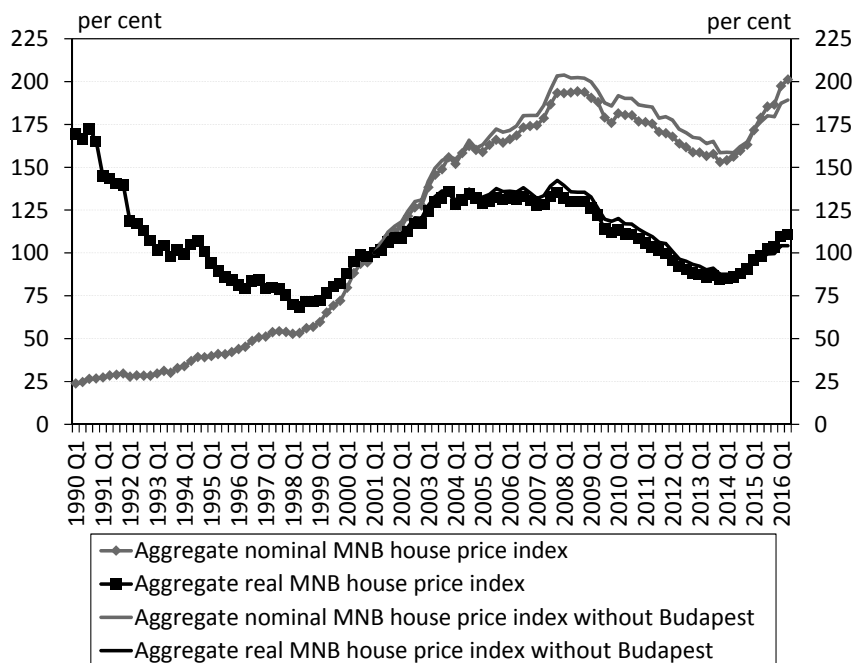

Figure 2. Nominal and real MNB house price index (2001 Q1 = 100%)

Note: The real index is deflated with the consumer price index. Based on a national estimation until 2001 and aggregate from the sub-indices from 2001. The current MNB house price index values are available for download at the following link: http://www.mnb.hu/letoltes/mnb-lakasarindex.xlsx

According to our calculations, between 1990 and 2007 the house price indices grew continuously in a nominal sense, but at varying rates in different phases. Prices increased at a relatively slower rate between 1990 and 1999; during these years the house price level roughly doubled. The increase between 1999 and mid-2003, however, is even more robust, in this four and a half years the prices rose by nearly 157 per cent. Although prices continued to increase until the 2008 crisis, the growth rate was far less pronounced. Based on the results of the MNB's indices and consistent with the HCSO's calculations, Hungarian house prices embarked on a continuous decline that lasted until 2014. From the upswing in early 2014 house prices started to grow dynamically once again.

By 2016, the nominal level of house prices rose to a historical peak. After the sharp rise of the past two and a half years, on average, prices have already exceeded the previous "peak" of 2007–2008. In real terms, however, house prices still fall significantly behind the levels recorded in 2003–2008.

Broken down by regions and settlement types, the house price indices indicate a considerable heterogeneity in the Hungarian housing market. Budapest has witnessed more dynamic price increases in recent years than those seen in municipalities outside of the capital (*Figure 3*), while differences are also evident
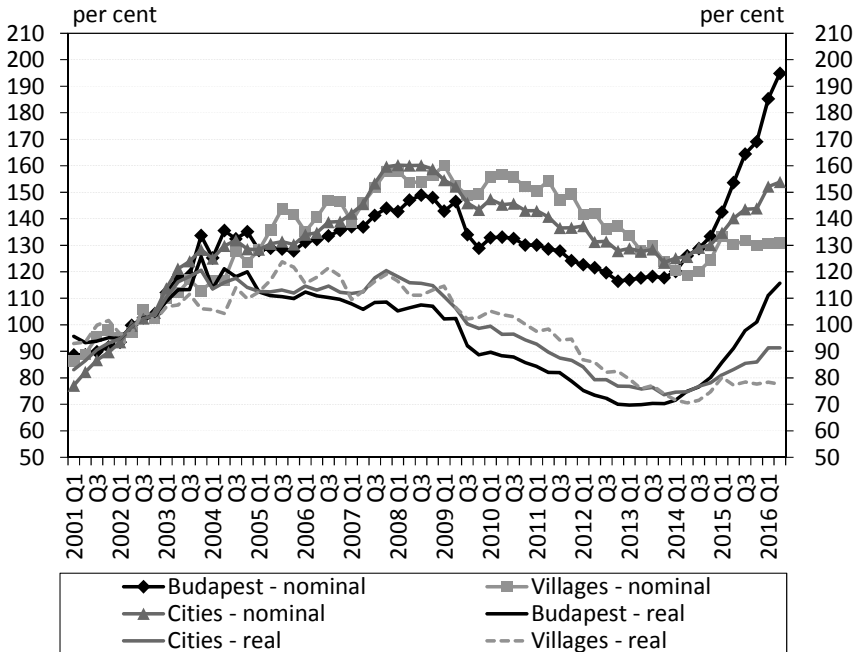
Figure 3. The MNB's nominal and real house price index by municipality typs
(2002 = 100%)
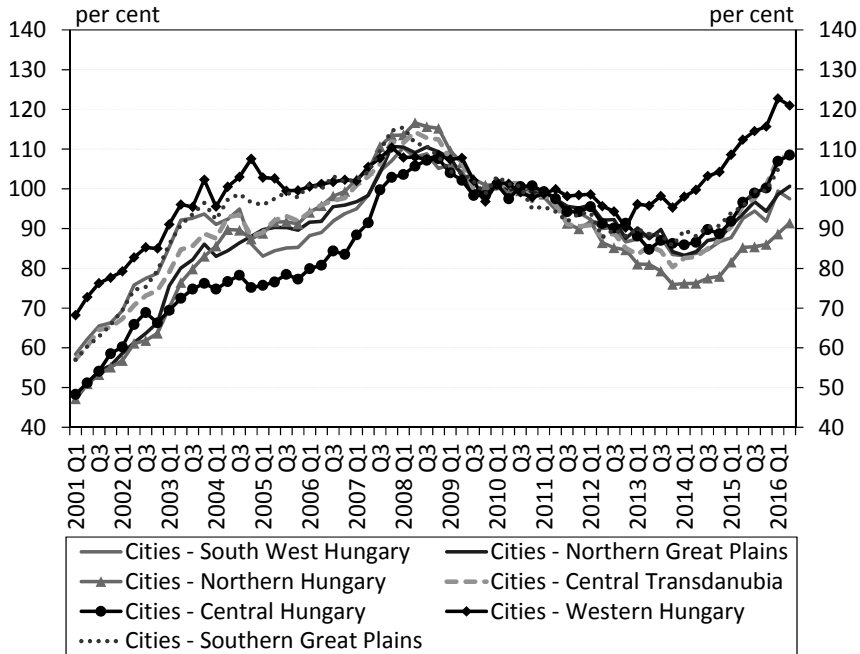
Figure 4. The MNB's nominal house price index for cities by regions (2010 = 100%)

between certain regions (*Figure 4*). After 2008, the house prices did not exhibit such a steep downward shift in Western Hungary as in the rest of the country, while the house price levels in Northern Hungary, for example, fall far behind. One important result of the regional breakdown of the house price index, overall, is the separate presentation of the house price changes in Budapest. In the current cycle, the pick-up in the housing market is strongly Budapest-oriented, which is well reflected in the 65 per cent nominal increase in the Budapest house prices between 2013 Q4 and 2016 Q2, compared to the national average of 31 per cent. The sharp increase in the Budapest house prices, however, appears to be less re-markable once we consider that prices did not reach the 2008 level until 2016 in real terms (*Figure 3*).

5.2. Regression results

Regressions for Budapest, cities and villages are described separately. Due to the limited scope of this study, we only present the Southern Transdanubian region for the city indices constructed for the individual regions, because this region

aptly illustrates the effect of the Balaton, the most important tourism catchment area apart from the Budapest agglomeration. Since the regression model is estimated over and over again for each quarter pair, of all the results we only present the regression results of the estimates required for the construction of the 2015 Q4 index.

Table 3 shows the regression outputs of the Budapest model. Estimated on a sample covering 2015 Q3 and 2015 Q4, the model has 69 per cent explanatory power.[17] The dummy variable denoting 2015 Q4 shows that a Budapest transac-

Table 3. Regression results of the Budapest house price index model for 2015 Q4

Variables		Variables	
Quarter 2015 Q4 (ref.: 2015 Q3)	0.0278***	Districts (ref.: 1)	
	(0.0042)	2	0.0361*
New flat (ref.: used flat)	0.3507***		(0.0197)
	(0.0162)	3	–0.3261***
Type of property * size of property (ln)			(0.0187)
Condominium	1.2556***	4	–0.4888***
	(0.0767)		(0.0194)
Panel block of flats	1.4491***	5	0.2110***
	(0.0819)		(0.0214)
Family house (inner city)	1.0233***	6	–0.1000***
	(0.0718)		(0.0198)
Family house (outer city)	1.1667***	7	–0.2811***
	(0.0692)		(0.0190)
Type of property * size of property(ln)^2		8	–0.4511***
Condominium	–0.0416***		(0.0188)
	(0.0097)	9	–0.2544***
Panel block of flats	–0.0954***		(0.0196)
	(0.0117)	10	–0.6225***
Family house (inner city)	0.0124		(0.0197)
	(0.0095)	11	–0.1284***
Family house (outer city)	–0.0265***		(0.0185)
	(0.0083)	12	–0.0038
Constant	12.5399***		(0.0205)
	(0.1524)	Further categories	
Number of observations	23386		
Adj. R-squared	0.6867		

Note: "ref." indicates reference category. "Further categories" indicate that the regression table does not include all categories of the variable because of the lack of space, for the complete version see Banai et al. (2017). Standard errors in brakets. *, **, *** indicates a coefficient significant at the 0.01, 0.05 and 1% level.

[17]   It should be noted that the backcasting of the data improves the explanatory power of the regressions artificially.

tion took place in 2015 Q4 as opposed to 2015 Q3, increased the price by $e^{0.0278} = 1.028$; in other words, in 2015 Q4 the pure price change in the Budapest housing market was 2.8 per cent.

The NIA variable is included in the regression in its interaction with the type of the property; in other words, a 1 per cent increase in NIA may generate different price increasing effects for different property types. In addition to the linear term, the squared term of the NIA variable was also included in the models because in our view, a 1 per cent increase in useful NIA may have a different price increasing effect in the case of larger dwellings. Because of the inclusion of the squared term, the partial effect exerted by a 1 per cent difference in NIA depends on the size of the NIA; consequently, the examination of the estimated coefficients alone

Table 4. Combined partial effect of the linear and squared terms of the NIA variable by average NIA

	Mean (sq metre) of sub-samples by property type	Average partial effect by property type	Mean (sq metre) of total sample by settlement type	Average partial effect by settlement type
Budapest				
Condominium	55.6	0.9209	61.2	0.9129
Panel block of flats	52.0	0.6953	61.2	0.6639
Family house (inner city)	119.5	1.1421	61.2	1.1255
Family house (outer city)	102.1	0.9217	61.2	0.9488
Cities in South West Hungary				
Condominium	58.0	0.8869	70.6	0.8683
Panel block of flats	52.7	0.9079	70.6	0.8856
Family house (county seat)	92.4	2.1536	70.6	2.0193
Family house (other)	84.5	2.6631	70.6	2.5259
Villages				
Condominium	67.3	1.1013	80.1	0.5848
Homestead	75.0	2.7992	80.1	2.6731
Family house	80.8	3.0359	80.1	3.0512

is not appropriate. For the model specified for Budapest, *Figure 8* in the *Annex* illustrates the combined partial effect of the linear and squared terms on the price by property type. *Table 4,* in turn, indicates the average partial effects.

Evidently, a 1 per cent increase in NIA has the greatest positive impact on the value of detached houses. Moreover, the regression results show that the partial effect of NIA is significantly higher in the case of detached homes in the inner

districts[18] of Budapest compared to those located in the outer districts. In addition, with respect to the price effect of the NIA, there is a significant difference between brick homes and panel buildings: the partial price increasing effect of the NIA is greater for brick houses, which are considered to be of better quality. Another interesting result from the regressions for Budapest is the fact that compared to the I. district, only the V. district has a price increasing effect; and in the rest of the districts, the residential properties tend to be cheaper on average. It is another intuitive result that the new residential properties are, *ceteris paribus,* more expensive on average.

Table 5 depicts the regression results of the model constructed for Southern Transdanubian cities, also run on a sample of 2015 Q3 and 2015 Q4 transactions.

Table 5. Regression results of the villages and the Southern Transdanubian cities' house price index model for 2015 Q4

Variables	Cities	Variables	Villages
Quarter 2015 Q4 (ref.: 2015 Q3)	−0.0270**	Quarter 2015 Q4 (ref.: 2015 Q3)	−0.0140
	(0.0124)		(0.0093)
New flat (ref.: used flat)	0.3202***	New flat (ref.: used flat)	0.3895***
	(0.0417)		(0.0827)
Type of property * size of property (ln)		Type of property * size of property (ln)	
Condominium	1.2706***	Condominium	13.6006***
	(0.4071)		(0.3561)
Panel block of flats	1.2105***	Municipality	11.0535***
	(0.4295)		(0.3412)
Family house (county seat)	−0.1041	Family House	10.8612***
	(0.3688)		(0.3288)
Family house (not county seat)	−0.7235*	Type of property * size of property $(\ln)^2$	
	(0.3734)	Condominium	−1.4847***
Type of property * size of property $(\ln)^2$			(0.0442)
Condominium	−0.0472	Municipality	−0.9559***
	(0.0503)		(0.0453)
Panel block of flats	−0.0382	Family House	−0.8908***
	(0.0570)		(0.0370)
Family house (county seat)	0.2494***	County (ref.: Pest)	
	(0.0428)	Györ-Moson-Sopron	0.0581
Family house (not county seat)	0.3816***		(0.0416)
	(0.0436)	Vas	0.0515
County (ref.: Baranya)			(0.0475)
Somogy	0.0635**	Further categories	
	(0.0308)	Agglomeration (ref.: not agglomeration)	

---

[18] Inner districts featured in the model: I, II, III, V, VI, VII, VIII, IX, XI, XII, XIII, and XIV.

Table 5. cont.

Variables	Cities	Variables	Villages
Tolna	0.0123	Pécs Agglomeration	0.4383***
	(0.0429)		(0.0585)
Agglomeration (ref.: not agglomeration)		Szeged Agglomeration	0.4817***
Pécs Agglomeration	0.2244***		(0.0542)
	(0.0409)	Further categories	
Recreational area (ref.: not recreational area)		Recreational area (ref.: not recreational area)	
Lake Balaton - near shore	0.8919***	Lake Balaton - near shore	0.8194***
	(0.0645)		(0.0325)
Lake Balaton - rest	–0.0044	Lake Balaton - rest	0.3928***
	(0.0528)		(0.0314)
		Dunakanyar	0.1991***
Local home support (ln)	–0.0105***		(0.0275)
	(0.0034)	Further categories	
Total income per capita (ln)	1.0131***	Total income per capita (ln)	0.5281***
	(0.0605)		(0.0234)
Distance from Budapest (ln)	–0.0978	Distance from Budapest (ln)	–0.2668***
	0.0771		(0.0273)
Distance from county seats (ln)	–0.0801***	Distance from county seats (ln)	–0.1884***
	(0.0090)		(0.0123)
Population (ln)	–0.1172***	Population (ln)	0.1156***
	(0.0200)		(0.0089)
Size of municipalities (ln)	0.1098***	Size of municipalities (ln)	–0.0815***
	(0.0251)		(0.0101)
Constant	–1.0344	Constant	–20.8978***
	(1.2943)		(0.8095)
Number of observations	4300	Number of observations	16006
Adj. R-squared	0.6253	Adj. R-squared	0.7366

Note: "ref." indicates reference category. "Further categories" indicate that the regression table does not include all categories of the variable because of lack of space, for the complete version see Banai et al. (2017). Standard errors in brakets. *, **, *** indicates coefficients significant at the 0.01, 0.05 and 1% level.

In the region, a 1 per cent increase in NIA has a higher partial effect among detached homes than among flats, while the price increasing effect is nearly identical for the brick or panel buildings within the flat category (*Table 4*).

Importantly, the recreation areas located closer to the shore of Lake Balaton have a significant price increasing effect, whereas the locations close to but not directly on the shore of the lake do not have a significant price effect in the case of cities. In the Southern Transdanubian region, the "Distance from Budapest" is not a significant variable. This seems to be plausible if we take into consideration that the cities on the southern shore of Lake Balaton can be accessed more

easily from Budapest than other cities in the Southern Transdanubian region that is, the effect of the "Distance from Budapest" can be somewhat absorbed by the categories of the Balaton "Recreation Area" variable referring to the closeness of Lake Balaton. The accessibility of the given county seat, however, proved to be an important factor. Moreover, an increase in the net per capita income for the settlement is combined with a higher transaction price on average, while the amount of the local subsidies granted for housing purposes may have a negative effect on the sale price. Presumably, the latter may be attributed to the potential positive correlation between the subsidy amounts and the ratio of the socially disadvantaged households and hence, the ratio of the lower-quality properties.

We present the results of the model estimated for villages of a sample covering 2015 Q3 and 2015 Q4 in *Table* 5. Similar to Budapest and to the cities located in rural Hungary, it is also true for villages that a 1 per cent difference in useful NIA exerts the greatest partial impact on the price for detached homes (*Table 4*). In addition, consistent with our intuition, a 1 per cent increase in NIA has a somewhat smaller price increasing effect in the case of rural farms, which are often poorer in quality. It is only in Győr-Moson-Sopron and Vas counties, located close to the Western border of Hungary that prices do not significantly differ from those observed in Pest county villages; and in all other counties the residential homes are, *ceteris paribus,* cheaper. Among recreational areas, the locations close to the shore of Lake Balaton exhibit the strongest price increasing effect and, as opposed to the cities, even the vacation spots not directly on the shore of Lake Balaton have a significant, albeit lesser, price increasing effect. For the most part, belonging to the agglomeration of regional centres raises house prices significantly. We did not receive a significant coefficient for the Budapest agglomeration, which might be because the model includes the distance from Budapest variable, with a significant negative coefficient. In addition, in the model constructed for villages, all other municipality-level variables have a significant effect. All else being equal, the house prices increase with the income and population size of the municipality, while the size of the municipality has a negative effect on prices. The significance of the municipality-level variables is fairly high; the model's 74 per cent explanatory power suggests that the inclusion of the municipality-level variables contributes the largest share of value added in the smaller, rather heterogeneous municipalities.

6. ROBUSTNESS ANALYSIS

We examined the robustness of the house price index models from four key perspectives: the backcasting of the useful NIA variable, the filtering of outliers

and influential values (hereinafter: outlier filtering), the range of the explanatory variables used for the estimate, and the estimation methodology of the model. The robustness analysis was intended to identify the extent to which a change in the methodology, *ceteris paribus,* affects the final values of the index, i.e. influences the result of the calculations. Moreover, we also examined the temporal stability of the parameters included in the models, as adjacent-period estimates allow for the temporal deviations of the nexus between house prices and its determinants.[19]

First and foremost, we wanted to ascertain that the residential properties purchased by business organisations do not excessively alter our results. According to our calculations, even with the exclusion of the latter transactions the aggregate house price index does not change perceivably overall. Such transactions, as mentioned before, form an integral part of the housing market turnover; consequently, we retained them in the final models.

Since the useful NIA variable is incomplete for numerous observations, we checked the extent to which the removal of these observations alter the aggregate price index (*Figure 5*). Although a considerable difference can be observed across the entire time series, the MNB's aggregate price index remains within the 95 per cent confidence interval in nearly all review periods.

In connection with the entry of the useful NIA, we also performed another type of robustness analysis. In the first step of the entry of missing NIA information, the "property size" variable was only selected in cases where it was smaller than 150m^2. Even adjusting the 150m^2 limit within the range of 100m^2 and 200m^2 did not generate a material difference in the final outcome of the index.

In the next step, we examined the deviation caused by the backcasting of missing NIA information in the quarterly change of the house price index. As mentioned in Section 4, we estimated the missing NIA information by regression models, with separate models constructed for each municipality type. The main reason for this exercise was the fact that we observed a perceivable divergence in the distribution of property size in individual settlement types. The backcasting of NIA information on the total sample, which is also performed by the HCSO, yielded considerably different results in certain periods.

We calculated the effect of tightening the limit applicable to per-square-metre prices – raising the lower limit of outlier filtering from HUF 2,000 per square metre to HUF 4,000 per square metre – on the quarterly change of the aggregate price index. This modification increases the volatility of the estimated price change in certain parts of the reference horizon.

---

[19]   For more information about the robustness analysis and for more illustrations see Banai et al. (2017).

Figure 5. Robustness analysis for the backcasting of the useful NIA (quarterly price changes)

Note: The black dotted line indicates the 95 per cent confidence interval of the MNB's aggregate house index.

Finally, we examined the effect generated by the use of a different method both in the first and second steps relative to the outlier filtering technique applied by the HCSO. We found that the latter may partly explain the differences between the house price indices of the HCSO and the MNB.

In the second step of the outlier filtering, we retained the observations that were considered valid by at least three of the four indicators discussed in Sub-section 4.2. We found that it barely changes the quarterly price change values of the index if the observations must be deemed valid by four or two indicators; therefore, our model is robust in this regard.

In the case of the explanatory variables, we explored two important questions. Firstly, we examined the value added to the final outcome of the house price indices by the municipality-level variables linked to the NAV duty database from the GEOX, TSTAR and NAV PIT databases. Secondly, we tested whether the temporal fixation of the municipality-level variables derived from the latter databases had any bearing on the results. We found that the municipality's income, distance from specific centres, size and population had significant impacts on the house price developments. Although the inclusion of the latter variables in the regressions only moderately influenced the outcome of the final aggregate house
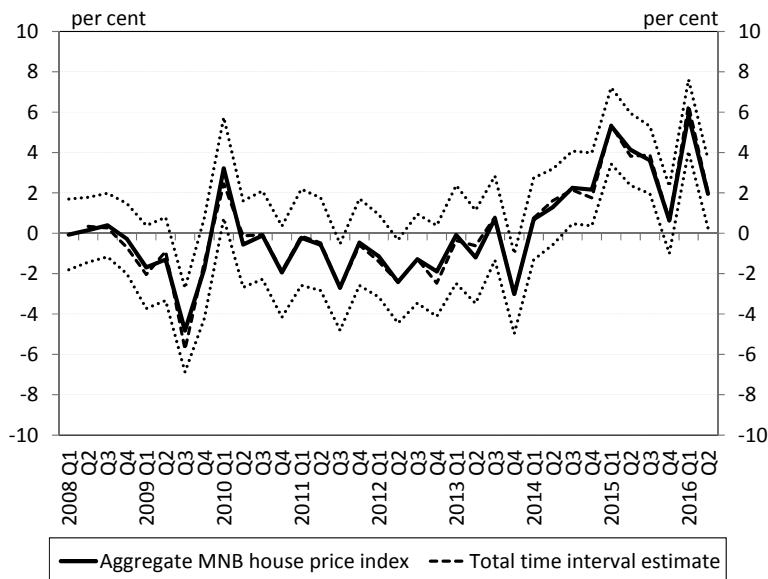
Figure 6. Robustness analysis for the estimation methodology (quarterly price changes)

Note: The black dotted line indicates the 95 per cent confidence interval of the MNB's aggregate house index.

price index, they had a considerable impact on the indices constructed for cities and villages. Moreover, according to our calculations, the temporal fixation[20] of municipality-level variables would only marginally influence the results.

With respect to the estimation methodology, we conducted two robustness analyses. On the one hand, instead of the sub-samples created for individual municipality types, we prepared estimates for the sample representing the whole of Hungary, and on the other hand, we made a disaggregate estimation on the sub-samples covering the entire time horizon instead of adjacent-period pairs. We found that the results did not deviate significantly from the final house price indices (*Figure 6*).

Beside the robustness of the results, we finally examined the stability of the parameters of the explanatory variables over time.[21] Although one of the main advantages of the use of adjacent-period estimates is specifically the fact that the

---

[20]   For the purposes of the robustness analysis, we included the 2013 values of the municipality-level variables in the regression equations.

[21]   It hinders the testing of the temporal stability of the parameters that even the constant is different over time, the effect of which on the parameters of individual variables cannot be factored in properly.
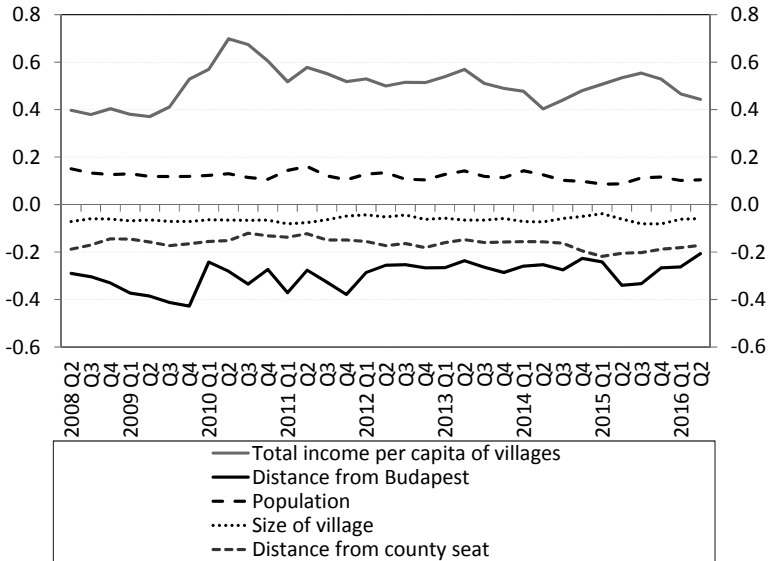
Figure 7. Dynamic analysis of the parameters of selected municipality-level variables
in the models constructed for villages

effect of certain variables on house prices may change over time, we should still examine whether a quarter-to-quarter change in the parameters gives rise to extreme volatility and whether the sign of the parameters remains stable over time.

In the case of villages, the entry of the municipality-level characteristics proved to be especially relevant; it is therefore important to examine whether the parameters of these characteristics are stable in size and magnitude over time. *Figure 7* depicts the coefficients of the municipality-level variables included in the models for villages. The figure indicates that both the per capita income and population of the municipality had positive parameters throughout the entire review period, while the estimated coefficients of the distance from the capital city/county seats and municipality size remained negative in all of the models. In the case of NIA variable, the time series of its partial effect, taking into account both the linear and quadratic terms were calculated for all property types for every disaggregate index[22] (by regions and settlement types). It holds for every category that the sign of the partial effect does not change over time. The relationship among the partial effects of different property types is also the same in almost every quarter. Beside these, there are differences in the magnitude of the partial effects that is consistent with our chosen adjacent-period model specification that allows the

---

[22]   Figure 8 in the Annex shows the results for Budapest.

nexus between NIA and the change in prices of the property over time. All in all, neither the municipality-level variables, nor the interaction term between NIA and property type had a coefficient that changed its sign over time that indicate the stability of the relation between these variables and house prices.

7. CONCLUSION

The Hungarian housing market developments are of key importance both for the banking sector and the real economy and accordingly, it is also the central bank's interest to gain insight into these processes. With that in mind, we have constructed the most detailed index family heretofore, taking a significant step toward a deeper understanding of the housing market processes. For constructing the index, we used the property acquisition duty data collected by the National Tax and Customs Administration (NTCA) in relation to the transfer of the residential properties. The MNB has specifically adapted the total duty database from the NTCA, which is suitable for individual identification for the reference period 1990–2015. This resulted in the longest and broadest Hungarian housing market database available so far. In our study, we presented the house price indices estimated using this unique database as well as their methodological background.

Compared to the information available in the past, we succeeded in taking a step forward in two main aspects. (1) First and foremost, the newly constructed index family is capable of providing segmented information on house price developments. We have constructed separate indices for different municipality types and individual regions. This was particularly important because the separate indices shed light on the significant heterogeneity behind nationwide developments. While prices continue to soar in Budapest, by early 2016 the market had already begun to stagnate in the cities of certain regions such as Northern Hungary, and even municipalities in rural areas recorded negligible growth. Indeed, the rise in national indices can be largely attributed to Budapest prices. (2) The newly constructed national index is available from 1990. The long time series allows us to make an assessment of the current level. We found that the prices still fall short of the pre-crisis levels in real terms, and even Budapest prices were unable to approach their 2008 levels until 2016. In addition, the national values indicate that the current price level continues to lag behind the levels recorded in 1990 in real terms.

The developments described above provide valuable supports to decision-makers by offering a more accurate view of the areas that are in need of intervention. Moreover, a more precise view may also help the market participants in their investment decisions or portfolio evaluations.

REFERENCES

Bailey, M. J. – Muth, R. F. – Nourse, H. O. (1963): A Regression Method for Real Estate Price Construction. *Journal of the American Statistical Association,* 58: 933–942.

Banai, Á. – Vágó, N. – Winkler, S. (2017): The MNB's House Price Index Methodology. *MNB Occasional Papers,* No. 127.

Belsley, D. A. – Kuh, E. – Welsch, R. E. (1980): *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* New York: Wiley.

Bollen, K. A. – Jackman, R. W. (1990): Regression Diagnostics: An Expository Treatment of Outliers and Influential Cases. In: Fox, J. – Long, J. S. (eds): *Modern Methods of Data Analysis.* Newbury Park, CA: Sage, pp. 257–291.

Chatterjee, S. – Hadi, A. S. (1986): Influential Observations, High Leverage Points, and Outliers in Linear Regression. *Statistical Science,* 1(3): 379–393.

Cook, R. D. (1977): Detection of Influential Observation in Linear Regression. *Technometrics,* 19(1): 15–18.

Court, A. T. (1939): Hedonic Price Indexes with Automotive Examples. In: *The Dynamics of Automobile Demand.* New York: General Motors Corporation, pp. 99–117.

Crone, T. M. – Voith, R. P. (1992): Estimating House Price Appreciation: A Comparison of Methods. *Journal of Housing Economics,* 2(4): 324–338.

European Central Bank (2011): *Financial Stability Review,* June 2011.

European Central Bank (2015): *Financial Stability Review,* November 2015.

Eurostat (2013): *Handbook on Residential Property Price Indices (RPPIs).*

Gatzlaff, D. H. – Ling, D. (1994): Measuring Changes in Local House Prices: An Empirical Investigation of Alternative Methodologies. *Journal of Urban Economics,* 35(2): 221–244.

Griliches, Z. (1961): Hedonic Price Indexes for Automobiles: An Econometric Analysis for Quality Change. In: Price Statistics Review Committee (ed.): *The Price Statistics of the Federal Government.* National Bureau of Economic Research, Inc., pp. 173–196.

Hair, J. F. – Anderson, R. E. – Tatham, R. L. – Black, W. C. (1995): *Multivariate Data Analysis.* Prentice-Hall.

Horváth, Á. (2007): Ingatlanárindexek számításának módszertana (On the Methodology of Real Estate Price Indexes). *Statisztikai Szemle,* 85(3): 213–239.

Horváth, Á. – Székely, G. (2009): Hedonikus módszer alkalmazása a használt lakások áralakulásának megfigyelésében (Applying Hedonic Methods in Observing Price Trends of Second-Hand Properties), *Statisztikai Szemle*, 87(6): 594–607.

Knight, J. R. – Dombrow, J. – Sirmans, C. F. (1995): A Varying Parameters Approach to Constructing House Price Indexes. *Real Estate Economics,* 23(2): 187–205.

Kutasi, D. – Badics, M. Cs. (2016): Comparison of the Hedonic Regression and the Neural Network Valuation Methods: Evidence from the Budapest Housing Market. *Acta Oeconomica*, 66(3): 527–546.

Lancaster, K. J. (1966): A New Approach to Consumer Theory. *Journal of Political Economy,* 74(2): 132–157.

O'Brien, D. – Westermann, T. – Vidalis, N. (2014): *Review of the ECB's House Price Valuations Framework.* ECB, Directorate General Economics, Prices and Costs Division.

Oikarinen, E. (2012): Empirical Evidence on the Reaction Speeds of Housing Prices and Sales to Demand Shocks. *Journal of Housing Economics,* 21(1): 41–54.

Rosen, S. (1974): Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy,* 82(1): 34–55.

Standard & Poor's (2009): *S&P Case-Shiller Home Price Indices Index Methodology.* New York.

Vellemen, P. F. – Welsch, R. E. (1981): Efficient Computing of Regression Diagnostics. *The American Statistician,* 35: 234–242.

Welsch, R. E. (1982): Influence Functions and Regression Diagnostics. In: Launer, R. L. – Siegel, A. F. (eds): *Modern Data Analysis.* New York: Academic Press, pp. 149–169.

Zrínyi, M. – Katona, É. – Szántó, I. – Páll, D. (2012): A lineáris regressziót befolyásoló esetek diagnosztikája (Diagnostics of Cases Influencing Linear Regressions). *Statisztikai Szemle,* 90(7-8): 719–732.

Table 6. Descriptive statistics of continuous variables

			1990–2000				
Variable	Category	Subcategory	Number of obs.	Mean	Minimum	Median	Maximum
Price (HUF)	Budapest		362 037	3 253 945	10 000	2 350 000	400 000 000
	Cities		133 425	1 959 699	11 000	1 100 000	185 000 000
	Villages						
Population (capita)	Budapest and cities		362 803	62 429	0	33 203	2 018 035
	Villages		133 425	2 204	0	1 906	8 615
Size (sqkm)	Budapest and cities		362 803	15 658	0	10 477	52 516
	Villages		133 425	3 468	0	2 871	28 458
Subsidies (thousand HUF/year)	Budapest and cities		362 803	0	0	0	0
	Villages		133 425	0	0	0	0
Distance from Bp. (minutes)	Budapest and cities		362 803	105	0	106	234
	Villages		133 425	106	0	100	255
Distance from county seat (minutes)	Budapest and cities		362 803	27	0	26	109
	Villages		133 425	43	0	42	138
Income of settlement (HUF/capita/year)	Budapest and cities		362 788	296 550	105 255	299 903	542 522
	Villages		133 186	212 510	17 033	209 903	1 109 502
Net internal area of residential property (sqm)	Budapest	house					
		flat					
	Cities	house county seat	51 585	69	15	64	499
		house other	153 279	70	15	66	499
		flat	157 172	55	15	53	498
	Villages	house	128 114	70	15	69	498
		flat	5 311	55	15	51	472

Note: Values received after first-step outlier filtering.

Table 6. cont.

			2001–2007			
		Number of obs.	Mean	Minimum	Median	Maximum
Price (HUF)	Budapest	363 415	14 013 435	69 360	10 500 000	665 000 000
	Cities	741 363	8 109 460	57 000	6 500 000	763 440 000
	Villages	310 107	4 519 724	57 000	2 800 000	602 844 032
Population (capita)	Budapest and cities	1 104 778	601 809	1 085	81 818	1 739 569
	Villages	310 107	2 144	0	1 833	9 983
Size (sqkm)	Budapest and cities	1 104 778	27 496	575	21 673	52 516
	Villages	310 107	3 417	0	2 778	28 458
Subsidies (thousand HUF/year)	Budapest and cities	1 104 778	172 818	0	5 100	831 095
	Villages	310 107	300	0	0	100 000
Distance from Bp. (minutes)	Budapest and cities	1 104 778	77	0	75	234
	Villages	310 107	122	0	120	255
Distance from county seat (minutes)	Budapest and cities	1 104 778	18	0	0	109
	Villages	310 107	44	0	42	138
Income of settlement (HUF/capita/year)	Budapest and cities	1 104 778	523 941	124 979	516 128	965 277
	Villages	309 975	344 009	16 731	325 811	1 271 065
Net internal area of residential property (sqm)	Budapest — house	52 184	73	15	69	499
	Budapest — flat	310 364	55	15	51	498
	Cities — house county seat	102 739	69	15	65	495
	Cities — house other	317 197	70	15	68	499
	Cities — flat	319 631	56	15	54	498
	Villages — house	299 261	73	15	72	496
	Villages — flat	10 561	61	15	55	483

Table 6. cont.

				2008–2016		
		Number of obs.	Mean	Minimum	Median	Maximum
Price (HUF)	Budapest	340 191	16 664 362	90 000	12 500 000	850 000 000
	Cities	637 411	9 964 693	79 328	7 800 000	660 000 000
	Villages	266 151	6 074 002	79 400	3 500 000	441 249 984
Population (capita)	Budapest and cities	977 602	644 731	1 000	111 836	1 757 618
	Villages	266 151	2 081	0	1 745	10 282
Size (sqkm)	Budapest and cities	977 602	13 586	6	525	52 513
	Villages	266 151	1 718	0	527	28 458
Subsidies (thousand HUF/year)	Budapest and cities	977 602	50 947	0	4 190	378 960
	Villages	266 151	194	0	0	250 000
Distance from Bp. (minutes)	Budapest and cities	977 602	71	0	63	236
	Villages	266 151	118	0	117	259
Distance from county seat (minutes)	Budapest and cities	977 602	17	0	0	113
	Villages	266 151	44	0	42	149
Income of settlement (HUF/capita/year)	Budapest and cities	977 602	794 693	283 727	819 877	1 348 615
	Villages	266 150	583 976	-47 703	564 139	2 531 264
Net internal area of residential property (sqm) — Budapest	Condominium	276 948	56	15	52	499
	Panel block of flats	18 104	53	15	52	409
	Family house inner districts	12 066	112	15	100	499
	Family house outer districts	32 508	102	15	93	499
Net internal area of residential property (sqm) — Cities	Condominium	256 275	58	15	55	499
	Panel block of flats	77 686	53	15	53	471
	Family house county seat	69 584	89	15	83	498
	Family house other	233 686	83	15	80	499
Net internal area of residential property (sqm) — Villages	Condominium	11 380	66	15	60	485
	Homestead	7 023	70	20	67	496
	Family house	247 671	79	15	76	498

Table 7. Aggregate nominal and real MNB house price index (2001 Q1 = 100%)

Quarters	Aggregate nomi- nal MNB house price index	Aggregate real MNB house price index	Quarters	Aggregate nomi- nal MNB house price index	Aggregate real MNB house price index
1990 Q1	23.9	169.6	1995 Q3	41.0	86.0
Q2	24.7	166.6	Q4	42.2	84.1
Q3	26.6	172.3	1996 Q1	44.0	81.2
Q4	26.8	165.3	Q2	45.3	79.4
1991 Q1	27.4	144.7	Q3	48.7	83.3
Q2	28.5	143.2	Q4	50.8	84.2
Q3	29.0	140.4	1997 Q1	51.2	79.5
Q4	29.7	139.9	Q2	53.8	79.6
1992 Q1	27.8	118.6	Q3	54.3	78.8
Q2	28.5	117.3	Q4	53.9	75.6
Q3	28.5	113.2	1998 Q1	52.9	70.2
Q4	28.4	107.2	Q2	53.3	68.4
1993 Q1	29.7	101.5	Q3	56.1	71.8
Q2	31.2	104.5	Q4	56.9	71.8
Q3	30.2	98.0	1999 Q1	59.8	72.4
Q4	32.9	101.8	Q2	65.3	76.8
1994 Q1	33.9	99.4	Q3	69.3	80.2
Q2	37.0	104.6	Q4	72.1	82.1
Q3	39.3	107.0	2000 Q1	79.7	88.0
Q4	39.2	100.7	Q2	88.3	95.2
1995 Q1	39.9	94.0	Q3	93.8	98.8
Q2	41.1	89.2	Q4	94.6	97.6

Quarters	Aggregate nominal MNB house price index	Aggregate real MNB house price index	Aggregate nominal MNB house price in- dex without Budapest	Aggregate real MNB house price index without Budapest
2001 Q1	100.0	100.0	100.0	100.0
Q2	104.3	101.7	105.6	103.0
Q3	109.5	106.1	112.1	108.6
Q4	112.6	108.4	115.6	111.3
2002 Q1	115.0	108.3	118.1	111.2
Q2	121.6	112.4	124.5	115.1
Q3	126.7	117.4	130.1	120.5
Q4	127.7	117.3	130.9	120.3
2003 Q1	138.4	124.5	142.1	127.8
Q2	145.6	129.5	149.6	133.1
Q3	148.9	131.8	153.9	136.2
Q4	155.9	135.9	156.2	136.2
2004 Q1	152.0	128.0	154.8	130.4
Q2	158.3	131.1	158.9	131.7
Q3	162.5	134.4	165.8	137.2
Q4	160.1	131.8	161.2	132.7
2005 Q1	159.0	129.4	163.0	132.6
Q2	163.0	130.2	168.0	134.2

Quarters	Aggregate nominal MNB house price index	Aggregate real MNB house price index	Aggregate nominal MNB house price index without Budapest	Aggregate real MNB house price index without Budapest
2005 Q3	166.0	132.5	172.3	137.5
Q4	164.4	131.1	170.4	135.9
2006 Q1	166.4	132.0	171.5	136.0
Q2	168.6	131.2	174.2	135.6
Q3	173.2	132.8	180.1	138.1
Q4	174.0		180.2	135.0
2007 Q1	174.5	127.6	180.2	131.7
Q2	178.7	128.1	186.1	133.4
Q3	186.7	132.9	195.3	139.0
Q4	193.4	135.3	203.3	142.3
2008 Q1	193.2	132.1	203.8	139.4
Q2	193.6	130.0	202.1	135.7
Q3	194.3	130.1	202.3	135.4
Q4	193.8	130.0	201.9	135.5
2009 Q1	190.5	126.4	199.8	132.6
Q2	188.0	121.8	194.6	126.1
Q3	179.0	114.1	187.6	119.7
Q4	175.8	112.2	185.7	118.5
2010 Q1	181.5	113.6	191.7	120.0
Q2	180.5	111.0	190.1	117.0
Q3	180.3	110.7	190.1	116.8
Q4	176.8	108.1	186.4	114.0
2011 Q1	176.4	106.0	185.8	111.6
Q2	175.4	103.7	185.1	109.5
Q3	170.6	101.4	178.7	106.2
Q4	169.9	99.8	179.5	105.5
2012 Q1	167.9	95.5	177.6	101.0
Q2	163.9	91.9	172.2	96.5
Q3	161.8	90.5	170.1	95.2
Q4	158.7	88.5	167.4	93.3
2013 Q1	158.6	87.7	166.9	92.3
Q2	156.7	86.3	163.8	90.2
Q3	157.8	87.0	165.1	91.1
Q4	153.1	84.7	158.5	87.7
2014 Q1	154.1	85.2	158.7	87.7
Q2	156.1	86.1	158.5	87.4
Q3	159.7	88.1	162.2	89.5
Q4	163.1	90.9	164.7	91.8
2015 Q1	171.8	95.9	172.2	96.1
Q2	178.9	98.4	176.3	97.0
Q3	185.3	102.3	180.1	99.4
Q4	186.5	103.4	179.5	99.5
2016 Q1	197.4	109.9	187.3	104.3
Q2	201.2	110.8	189.2	104.2

Source: MNB.

Note: The real index is deflated with the consumer price index. Aggregate from the nation estimates until 2001 and from the sub-indices from 2001.
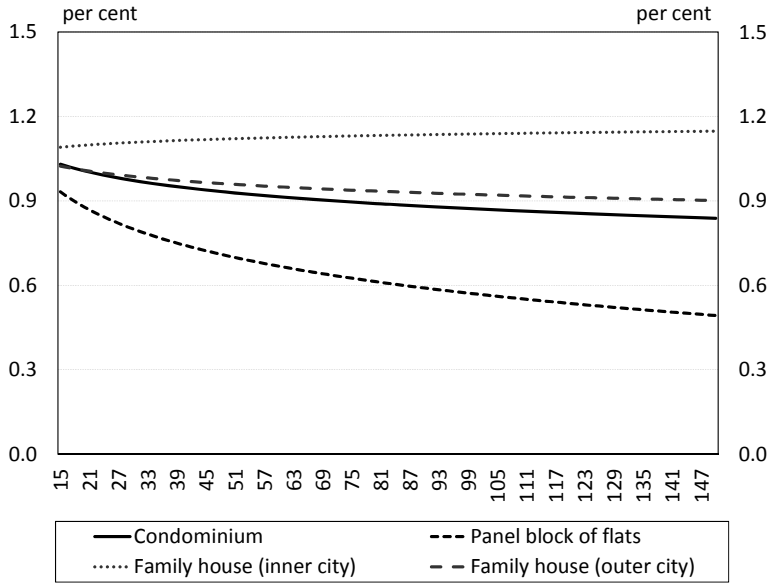
ANNEX

Figure 8. Partial effect of the interaction between NIA and property type in a model
specified for Budapest

Note: The horizontal axis indicates the size of the property expressed in square metres. The figure shows the
price increase generated, *ceteris paribus,* by a 1 per cent increase in the NIA of a given property based on an
estimate run on a sample covering 2015 Q3 and 2015 Q4. If the model only included linear terms, the figure
would present the constant functions.