Ádám Banai–Nikolett Vágó–Sándor Winkler

The MNB's house price index methodology

MNB Occasional Papers 127

2017

Ádám Banai–Nikolett Vágó–Sándor Winkler

The MNB's house price index methodology

MNB Occasional Papers 127

2017

MNB Occasional Papers 127

The MNB's house price index methodology

(Az MNB lakásárindex módszertana)

Written by Ádám Banai, Nikolett Vágó, Sándor Winkler

Budapest, May 2017

Published by the Magyar Nemzeti Bank

Publisher in charge: Eszter Hergár

H-1054 Budapest, Szabadság tér 9.

www.mnb.hu

ISSN 1585-5678 (online)

Contents

Abstract

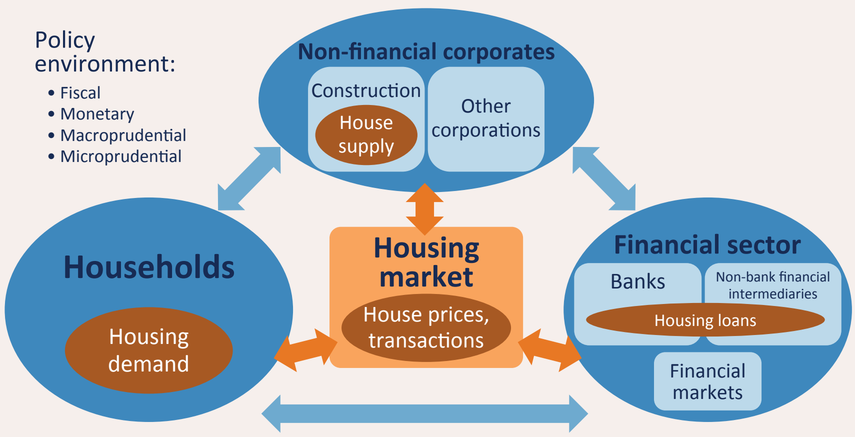
This study presents the detailed method of the MNB's house price index and the results of the new price indices. The index family is considered to be a novelty among Hungarian housing market statistics in several regards. Firstly, the national index was derived from a database starting in 1990, and thus the national index is regarded as the longest in comparison to the house price indices available so far. The long time series allows us to observe and compare the real levels of house prices across several cycles. Another important innovation of this index family is its ability to capture house price developments by region and settlement type, which sheds light on the strong regional heterogeneity underlying Hungarian housing market developments.

JEL codes: C430, R210, R310
Keywords: housing market, house price index, hedonic regression

1 Introduction, 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. Housing market trends are of key significance from a social perspective as well, especially in a country where the majority of citizens reside in their own property. As the availability of affordable housing may decline in line with rising real estate prices, an increasing part of the population may face housing problems. Similarly, real estate market developments directly affect the business sector. Price developments and the number of transactions affect demand for new projects and ultimately, the construction industry. Finally, real estate market developments also exert a direct impact on the banking sector. Changes in the prices of the real estate collateral behind mortgage loans may not only determine the performance of the loans, but the recovery through the sale of collateral in the event of default. Besides the outstanding portfolio, the real estate market also affects the banking sector through new disbursements. Compared to corporate loans, the banking sector can earn larger spreads on mortgage loans, and intensifying activity has a positive impact on profitability. In addition, a buoyant real estate market boosts demand for housing loans.

Figure 1
Interaction between market participants and the housing market



Source: ESRB, MNB.

In addition to the factors described above, understanding real estate market trends and identifying the risks arising in the real estate market are of key importance for the central bank. 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 domestic housing market developments, the construction of additional house price indices may be instrumental from several aspects. 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. Apart from this feature, however, both indices treat national housing market processes in a completely uniform fashion, which might conceal important information.

Compared to these two indices, the house price index family presented in this study represents a step forward in several regards. (1) The index developed by the authors of this study has been constructed on the longest and most comprehensive time series: it captures Hungarian house price developments 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, obscuring the different behaviour patterns of individual regions. From a business, banking and central banking perspective, however, it is essential to be aware of housing market heterogeneities between regions and municipality types. Even when national price movements do not indicate any problems, individual regions may be exposed to potentially harmful developments. Extreme spikes in house prices or housing market bubbles typically arise in 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.

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

2 A brief review of the relevant literature

The calculation of house price changes 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, housing market turnover provides a suitable basis for measuring the changes in house prices. Although residential property prices can also be determined by appraisal, comprehensive and regular appraisals of the stock of dwellings are scarce even by international standards, primarily because of the relatively high costs of data collection.[1] Consequently, house price changes are typically computed from transaction prices.

Ideally, for the calculation of average price changes each dwelling should change owners in each review period; in that case, market price data would be available on each individual property. In reality, however, only a fraction of the dwelling stock changes hands 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 different quality in each period, changes in the average or median transaction price 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 price changes, we should observe the trading of houses that have the same quality and attributes. 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 residential properties.

In 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 circumstance that certain locations are overrepresented in 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 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 abovementioned limitations are also present in international house price statistics.

The main difference between the construction methods of house price indices arises from their treatment of the bias stemming from the compositional shifts observed in consecutive periods. In the following, we provide an overview of possible house price index calculation methods based on the classification presented in the handbook issued by Eurostat (2013).

Stratified sample mean method

This approach 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 received by taking the weighted average of the values computed for the 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. This is partly because of the scarcity of information on the housing characteristics that

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

are most suitable for stratification. On the other hand, an adequately detailed stratification (classification of sub-groups by numerous qualitative variables) may significantly reduce the sample sizes used for the calculation of individual sub-indices, which reduces the reliability of the method.

Indices from stratified sample means have been constructed, among others, by the UK Department of the Environment (1982) and by the Australian Bureau of Statistics (ABS, 2006), whereas the approach is typically used as a point of reference in the academic literature (e.g. Mark and Goldberg (1984), Crone and Voith (1992), Gatzlaff and Ling (1994), and Wang and Zorn (1997)).

Hedonic regression models

The estimation of **hedonic regression models** is the most widely used method of calculating house price indices. The basic assumption of this approach is that 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).

From price change observations, the 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: the floor area of the structure, the size of the land (for single detached dwelling units) and the characteristics of its environment, age and type of the dwelling (e.g. detached house, row house, condominium), 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). 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**,[2] a pooled OLS estimate is prepared based on the data of all periods. Except for the initial base period, the model uses a separate dummy variable 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:

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

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 the 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 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 estimate 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:

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

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

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 the model is that the estimated coefficients of the explanatory variables can change over time; in other words, the model eliminates the underlying assumption of the time dummy variable model, i.e. that the parameters are constant over time. This is consistent with the assumption that 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 the 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 additional index values.

In the **multiperiod time dummy approach**,[3] a 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 characteristics included in the model. The regression equation for period t is the following:

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

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 approach when 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 defined. 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 be rendered even more uncertain by the lack of sufficiently numerous observations for each period.

It should be noted that the risk of multicollinearity may arise in all three types of the hedonic model. A high correlation between the explanatory variables increases the standard errors of the coefficients, and some variable 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 for the estimate as possible should be considered, hence reducing the risk of bias arising from 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. region, municipality type). 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 price in the individual sub-samples. For the sake of reliability, each sub-sample should include a sufficient number of observations.

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

Repeated sales method

Another frequently used approach is the **repeated sales method**, which 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. However, the downside is the potential selection bias that may arise due to the diverging market velocity of the properties. This problem can be mitigated by increasing the time horizon considered in between transactions. This, however, will also increase the bias stemming from the change in the condition of the units (unless a control variable is available in this regard). The method can be applied efficiently if the number of transactions completed in the reviewed real estate market is high enough, ensuring a sufficient number of observations for the estimation. 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. The repeated sales method was first proposed by Bailey, Muth and Nourse (1963). Besides the Federal Housing Finance Agency (FHFA), Standard & Poor's (2009) compiles a house price index based on this method for 20 cities in the United States. Residex and the UK Land Registry also compile repeated sale indices for Australia and the United Kingdom, respectively.

Beyond the models presented above, there are **combined methods** that use hedonic regressions and the repeated sales method in conjunction with one another in order to maximise their benefits. Since hedonic regressions entail a smaller risk of selection bias and the repeated sale method is less susceptible to model specification, using the method of ordinary least squares (OLS) in conjunction with the joint estimation of the equations may in theory lead to a more efficient outcome. 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 requirement of the methodologies described above is extremely different. The data available to us so far (presented in Chapter 3) are primarily suitable for the purposes of hedonic regression models.

3 Data

House price changes can be statistically measured by using two data types: (1) value data of the stock of dwellings, or (2) transaction data evolved during the sales and purchases of the properties. The former are typically derived from real estate property appraisals and, as shown at the beginning of the previous chapter, at the global level 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 transaction data for the compilation of house price indices. We compiled the MNB's house price index based on the property acquisition duty data collected by the National Tax and Customs Administration (hereinafter: NTCA) in relation to the transfer of residential properties. We describe the detailed contents of the duty database and review the additional data included in the modelling of house prices in the following.

Aside from 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 sale of 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.

Although duty information has been collected by the NTCA since January 1990, since then the structure and quality of the data have been subject to significant changes. From 2008, the structure of the database changed in the wake of the integration of regional duties offices and the introduction of the uniform duty system. Before the integration, regional duties offices used diverging IT systems and as a result, data were collected in different formats. After the duty integration in 2008, however, data on the transfer of real estate properties were collected and stored in a uniform structure. With respect to the variables stored in the database, there are four main differences in the data collected before and after 2008:

1. before 2008, it was unknown whether the person liable for duty payment in the transfer of property is a private individual or a business organisation;
2. before 2008, it was also unknown whether the property sold was new or used;
3. before 2008, the database did not include the Budapest district variable; and
4. before 2008, the only known information on the type of the property was whether it was a detached house or a flat, while a more detailed decomposition is available from 2008 in this regard.

For our purposes, we also retained those transactions in the database where the party acquiring the property is a business organisation – this allowed us to disregard the first difference mentioned above.[4] These observations make up only about 7 per cent of post-2008 data; in addition, since the types of 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 housing market turnover and the housing market – their inclusion in the estimate appears to be warranted. We addressed the other two main differences by defining different model specifications for the data collected before and after 2008.

[4] The database does not include the properties purchased by the National Asset Management Agency (NAMA).

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 estimates 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 Sub-Chapter 4.1. We need to stress that 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.

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

Source	Variable	Description
NTCA duty database	ar_ln	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.
	adnev	Quarter of property acquisition duty.
	jelleg * lat_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), condomnium, panel block of flats and homestead. In case of data before 2008 there are only two categories: family house and flat.
	uj	Category variable: new or used property.
Variables made based on the HCSO id of the settlement	bp_ker	Category variable: districts of Budapest.
	agglomerácio	Category variable: 8 districts distinguished: agglomeration of Szeged, Pécs, Debrecen, Miskolc, Székesfehérvár, Budapest, Győr and Sopron.
	udulokorzet	Category variable: 7 seasonal property areas distinguished: Lake Balaton - near shore, Lake Balaton - other, Dunakanyar, Mátra-Bükk, Sopron-Kőszeghegyalja, Lake Tisza, Lake Velence - Vértes.
	megye	Category variable: county of the settlement.
TSTAR database	de02_ln	Population at the end of the year. The variable is in a logarithmic form.
	de66_ln	Size of the municipality. The variable is in a logarithmic form.
	on23_ln	Amount of local housing subsidies. The variable is in a logarithmic form.
Geox database	ido_p_bp_ln	The shortest distance from Budapest expressed in minutes. The variable is in a logarithmic form.
	ido_p_msz_ln	The shortest distance from the county seat expressed in minutes. The variable is in a logarithmic form.
NTCA PIT database	teljovperfo_ln	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, XIV.

As mentioned before, the NTCA database contains limited information on the characteristics of the dwellings. Apart from the NIA of the dwelling, the only known variables are type (detached house, semi-detached or row house, condominium, block of flats or homestead) and age, i.e. whether the dwelling is new or used. Because of that, numerous location-dependent variables have been included in the data used for the compilation of our index. The legal status of the municipality 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 the municipalities concerned and hence, the housing market of the area, driving up 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 and presumably, the value of the properties located in such areas is also higher. In addition, we used the size and population of the municipalities, the total amount of local housing subsidies and the net per capita income of the municipality 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.

On the whole, to construct the MNB's house price index we compiled a housing market transaction database that relies on the broadest range of information available. The broad range of the data available since 1990 and the municipality-level variables included in the model create a unique opportunity for the modelling of house prices.

The MNB's house price indices are published quarterly after the fourth month following the reference quarter. At the time of the calculation of the latest house price indices, around 50–70 per cent of the transactions have been processed, and nearly a year passes before the database containing the transactions of the given quarter becomes complete. As a result, index values for the last three quarters are subject to revision upon publication on every occasion.

4 The MNB's house price index methodology

4.1 THE BACKCASTING OF USEFUL NIA

One of the deficiencies of the database used 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 the "NIA" variable that indicates the NIA of the structures. 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 summary, we use the following method for the definition of the NIA:[5]

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 be 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	Municipalities	Total
1990–2007	4.2	20.3	53.4	25
2008–2015	7.4	25.6	63.1	28.7

We use the ordinary least squares method to estimate the linear regression model fit to the logarithm of the NIA variable. We prepare 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 type was warranted by the significantly diverging property size distributions in the individual municipality types (Table 3). The models which were run on data between 1990 and 2007 included the following explanatory variables: property price, quarter dummy, property type (house or flat), legal status of the municipality, category variables for county, agglomeration and seasonal property area and per capita income for the municipality. In the estimates prepared on the basis of 2008–2016 data, we were also able to control for a variable that indicates whether the property is newly built, for the district (in the case of Budapest) and – within the property-type variable – we distinguished between panel buildings and brick dwellings.

[5] 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.

Table 3
Percentiles of NIA by municipality type and by sub-sample

Percentile	1990–2007			2008–2016 Q1		
	Budapest	Cities	Municipalities	Budapest	Cities	Municipalities
5%	27	33	39	27	34	40
10%	31	37	45	31	38	45
25%	39	49	56	40	50	60
50%	52	56	70	53	57	75
75%	67	70	90	68	75	96
90%	86	95	120	90	103	120
95%	105	116	137	111	126	143

4.2 METHODOLOGY OF OUTLIER FILTERING

Linear regression models may be fairly sensitive to the outliers[6] 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 extreme and influential data points may improve the accuracy of the regression estimate and the reliability of the conclusions drawn from the results.

With that in mind, we opted for a multi-step outlier filtering technique. In the first step, we tried to strip out 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:[7]

- 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:[8]

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

where y_i is the observed price for the ith observation, $\widehat{y}_i$ is the estimated value for the ith observation, $MSE_{(i)} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2$ is the mean squared error excluding the ith observation and h_{ii} is the leverage.[9]

For each observation, the indicator examines the standard error adjusted value of the estimate's residual (deviation between observed and estimated values). The computed residuals are not homoscedastic (their

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

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

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

[9] 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 ith observed price (y_i) on the ith estimated value ($\widehat{y}_i$).

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 coefficients that are not skewed by the *i*th observation even if it is deemed to be an outlier; consequently, the indicator will not mistakenly yield price observations and price estimates that are too close to each other.[10]

2. Cook's distance:[11]

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

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

As opposed to the studentized residual indicator that concentrates on residual values, this indicator focuses on dependent variable estimates obtained with and without the *i*th 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 significant effect on the size of the estimated regression coefficients.

3. Welsch distance:[12]

$$WD_i = \frac{\widehat{y}_i - \widehat{y}_{(i)}}{\sqrt{MSE_{(i)}\left(1-h_{ii}\right)}} \frac{\sqrt{h_{ii}\left(n-1\right)}}{1-h_{ii}}$$

where $\widehat{y}_i$ is the estimated price for the *i*th observation, $\widehat{y_{(i)}}$ is the estimated value for the *i*th observation calculated from the coefficients of a regression obtained after the removal of the *i*th observation, n is the number of elements in the sample, $MSE_{(i)}$ is the mean squared error obtained after the removal of the *i*th observation and h_{ii} is the leverage of the *i*th 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 observations with high leverage. With similar outliers in the database, this indicator may be more efficient in identifying observations that are to be discarded.

4. DFBETA:[13]

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

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 **X** is the matrix of the explanatory variables included in the linear regression model.

[10] 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.

[11] For more detail, see Cook (1977), Hair et al. (1995) or Bollen and Jackman (1990).

[12] For more detail, see the publications of Welsch (1982) or Chatterjee and Hadi (1986).

[13] For more detail, see Belsley et al. (1980) or Bollen and Jackman (1990).

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 sample size and the number of the explanatory variables, each indicator can be used to identify the values that can be considered 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 above 4 indicators.[14] The results confirm that the selected indicators offer significant additional information content 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 (Table 4).

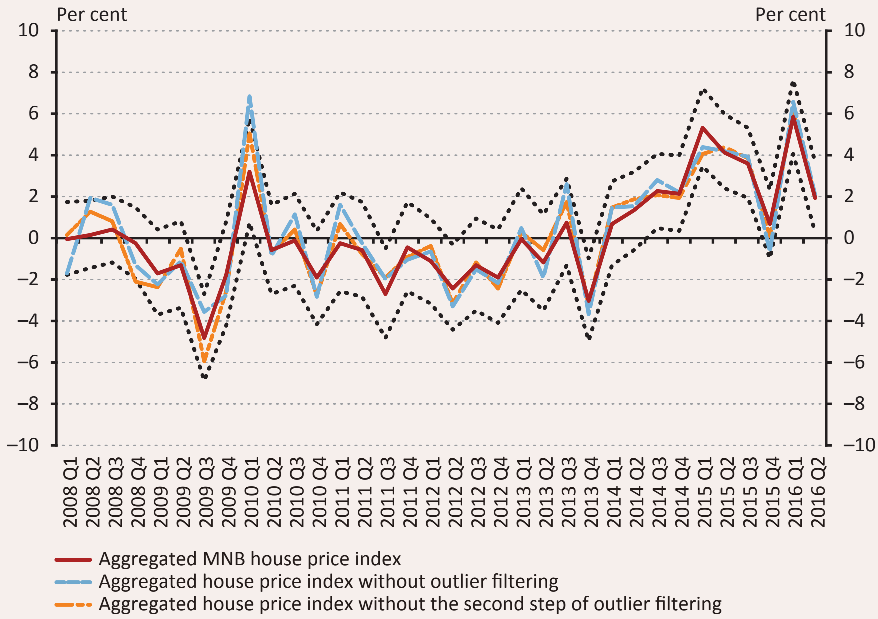
Table 4
Percentage of discarded observations in a multi-step filtering procedure

		2001–2007				2008–2015			
		Total	1st step of filtering	2nd step of filtering	Total Outliers	Total	1st step of filtering	2nd step of filtering	Total Outliers
		Number of obs.	%	%	%	Number of obs.	%	%	%
Budapest		364269	0.2	4.3	4.5	308734	0.2	5.4	5.6
Municipalities		317219	2.2	4.1	6.3	241947	2.5	5.1	7.6
Cities	Southern Great Plains	138899	0.7	4.9	5.6	110286	1.4	6.0	7.4
	South West Hungary	88198	0.6	5.1	5.7	64286	0.5	5.8	6.3
	Northern Great Plains	148379	0.7	5.2	5.8	100227	0.5	5.8	6.3
	Northern Hungary	88907	0.7	4.5	5.2	62331	0.7	6.0	6.7
	Central Transdanubia	99829	0.6	4.9	5.4	77462	0.6	5.7	6.3
	Central Hungary	108341	1.4	4.2	5.6	88414	0.9	5.7	6.6
	Western Hungary	73726	0.4	4.9	5.3	71439	1.0	5.5	6.5
Total		**1427767**	**0.5**	**4.5**	**5.0**	**1125126**	**1.0**	**5.5**	**6.6**

The effect of the filtering performed in the individual steps on the aggregate index is illustrated by Figure 2. 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.

[14] We performed a robustness analysis for this criterion, which is discussed in Chapter 6.

Figure 2
National 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.

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 Sub-chapter 3.1, 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 individual municipalities and regional processes separately. For this exercise, we need to divide the database into sub-samples, 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 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 in all periods. An estimate run 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 transaction until 2008, which would restrict the range of explanatory variables in the models run 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, which would not have been possible based on the time dummy variable approach. Finally, the time series resulting from the adjacent-period estimate is not subject to revisions for methodological reasons[15] with the release of new data.

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

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 regression estimate separately for each sub-sample. According to settlement type, we produced indices for Budapest, Cities and Villages. We defined the regional decomposition based on the number of observations available. Table 5 presents the number of observed transactions for each region and settlement type for each year of the review period in a breakdown that is consistent with that of the final house price indices. Since in the adjacent-period approach an estimate is prepared on the data of two quarters simultaneously, one estimate sample includes around half of the observations contained in the table cells on average. In the case of cities, the high number of observations allowed us to estimate separate models for each region. As a result, we compiled 7 region-level city indices and produced national house price indices for cities by weighting the 7 city indices by 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 regional price indices for villages.

Table 5
Number of observations included in the estimate for each year of the review period

Year	Budapest	Municipa-lities	Cities							Total
			Southern Great Plains	South West Hungary	Northern Great Plains	Northern Hungary	Central Trans-danubia	Central Hungary	Western Hungary	
2001	27,348	34,703	18,340	10,224	18,648	5,202	9,698	11,620	9,190	**144,971**
2002	38,649	39,351	19,441	13,395	23,843	12,259	12,650	13,974	11,485	**185,044**
2003	62,771	45,160	21,242	14,301	24,969	16,121	15,709	17,121	12,205	**229,597**
2004	46,584	40,182	14,794	10,479	15,730	10,380	11,241	12,874	7,393	**169,654**
2005	45,247	44,447	15,615	10,135	16,990	12,066	12,401	12,792	7,763	**177,455**
2006	56,830	43,946	20,153	11,737	20,022	13,215	14,195	15,268	10,518	**205,883**
2007	54,117	45,586	19,967	11,790	18,574	12,951	16,185	16,831	10,297	**206,296**
2008	46,925	38,902	17,442	10,164	15,143	10,898	13,246	15,268	9,821	**177,806**
2009	30,928	28,229	11,930	7,261	11,157	6,695	7,855	9,903	7,215	**121,170**
2010	30,484	24,077	10,672	6,549	9,458	6,207	7,289	8,678	6,806	**110,218**
2011	29,389	23,422	10,533	6,262	9,422	5,708	7,587	8,280	7,053	**107,654**
2012	30,522	23,095	11,195	6,253	9,762	5,800	7,299	8,108	7,862	**109,894**
2013	29,693	23,189	10,716	6,335	9,838	5,979	7,484	7,966	7,720	**108,917**
2014	38,370	27,181	12,572	7,391	12,047	7,410	9,052	9,954	8,793	**132,769**
2015	46,272	28,326	13,406	8,074	13,134	7,709	10,265	12,073	8,977	**148,233**
2016	20,324	15,316	7,127	4,334	6,751	4,036	5,369	5,699	4,492	**73,446**

Note: values for 2016 refer to Q1 and Q2.

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 mentioned 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 national aggregate index resulting from the weighting of the previously described disaggregated indices.

From 2001, the aggregate, national-level house price index is constructed by weighting the disaggregated indices with transaction data. In the first step, we aggregate the quarterly price changes of houses in cities outside of Budapest by using the number of transactions as weights. Next, similar weights 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.

4.5 SELECTION OF EXPLANATORY VARIABLES

We selected the explanatory variables based on two main criteria: on the one hand, we tried to include in the models 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 make certain that the sign of the explanatory variables included in the models is economically intuitive.

Explanatory variables were selected only in the case of the TSAR, GEOX and NTCA PIT databases, as we have used all variables with sufficient data for the regressions from the NTCA duty database that serves as a basis of our calculations. As the variables included in the NTCA PIT database strongly correlate to each other, from this database we used a single variable derived from the tax base, tax payments and the population of the municipality: per capita net income.

The TSTAR database maintained by the HCSO contains nearly 1,800 municipality-level variables. We selected around 50 variables – linked to demography, the number and composition of enterprises, construction activity, labour market, education, the financial management of local governments and tourism – that may intuitively influence the average house prices observed in the municipality. Eventually, from these variables we only selected those that correlated relatively mildly with each other (number of population, municipality size and the amount of local subsidies granted for housing purposes).

The GEOX database features seven pieces of distance information for each municipality: the municipality's distance from Budapest, the county seat, the regional centre, the administrative district, the micro-regional centre and from the nearest highway and railway node. Each distance is optimised for the shortest time and shortest distance, expressed in time (minutes) and distance (kilometres). In our estimates we used the shortest time expressed in minutes. Among the distance data, we included two variables in our estimates: the distance from Budapest and from the county seat as the rest of the variables did not increase the model's explanatory power considerably.

Since it would not be right to remove temporal changes in municipality-level characteristics from the change in house prices, in the case of adjacent quarters where the sample includes transactions for Q4 and the adjacent Q1,[16] we run the regression in consideration of the latest municipality-level variable for all observations (irrespective of whether it is a Q4 or Q1 data item).

[16] The issue is only relevant in this particular case, as municipality-level variables are available at a yearly frequency.

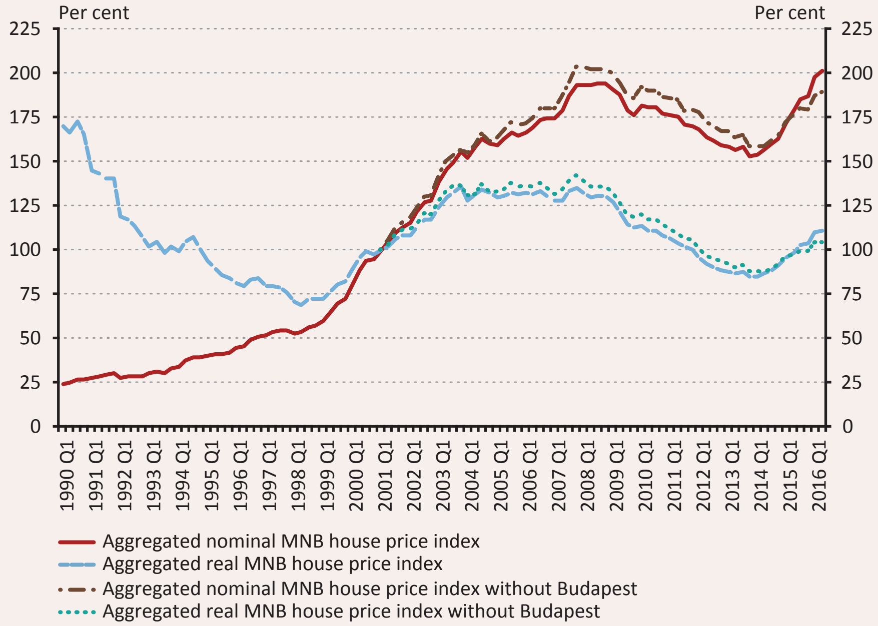
5 Presentation of the results of the MNB's house price index

In this chapter, we first present the time series of the MNB's house price indices and then provide an overview of the historic developments of Hungarian house prices. In the second half of the chapter, we discuss the regression results underlying our house price indices and in this context, we describe the partial effects of individual explanatory variables on the prices of residential properties.

5.1 RESULTS OF THE MNB'S HOUSE PRICE INDEX

In the following, we present the house price indices estimated for the individual sub-samples and the MNB's aggregate house price index. The NTCA database allowed us to produce the longest time series so far on Hungarian house price dynamics. We were able to estimate the MNB's house price index on a long time series from 1990 in an aggregated form, but in a disaggregated form we could only construct the indices from 2001 due mainly to the low number of observations of the initial years and the lack of Budapest observations before 2001. As a result, for the period before 2001 we constructed only a national index, while indices decomposed by settlement type and by region (for cities) start from 2001. From 2001, national indices are produced by weighting together the national index with the quarterly sub-indices, where the weights are the observation numbers used for estimating the index values. Figure 3 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 aggregated 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 3
Nominal and real MNB house price index
(2001 Q1 = 100%)



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

According to our calculations, between 1990 and 2007 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 4 and a half years prices rose by nearly 157 per cent. Although prices continued to increase up 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 region and settlement type, 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 4), while differences are also evident between certain regions (Figure 6). After 2008, house prices did not exhibit such a steep downward shift in Western Hungary as in the rest of the country, while 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 Budapest house price changes. At present, the pick-up in the housing market is strongly Budapest-oriented, which is well reflected in the 65 per cent nominal increase in Budapest house prices over the past two and a half years, compared to the national average of 31 per cent. The sharp increase in Budapest house prices, however, appears to be less remarkable once we consider that prices did not reach the 2008 level until 2016 in real terms (Figure 5).

Figure 4
The MNB's nominal house price index by municipality type
(2010 = 100%)

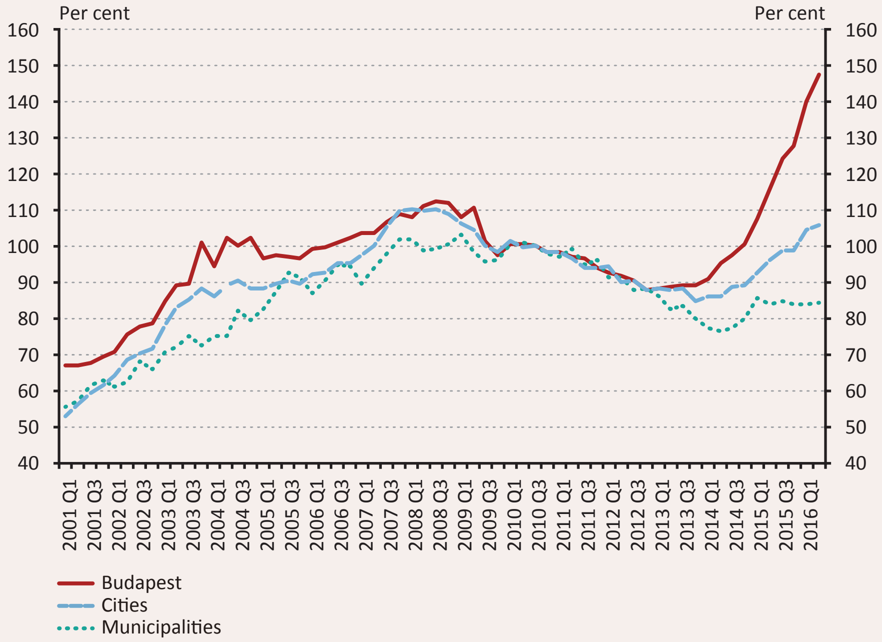
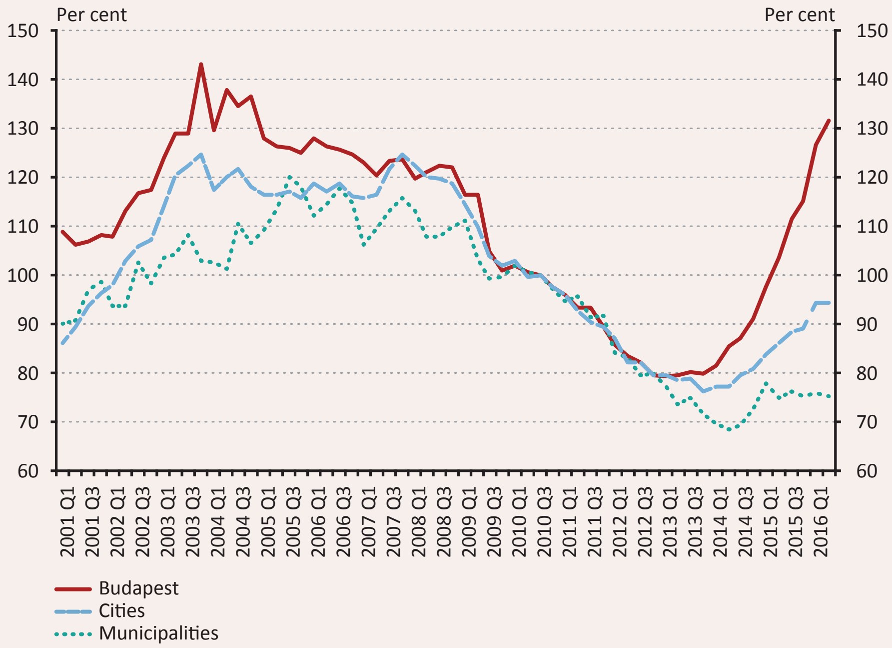


Figure 5
The MNB's real house price index by municipality type
(2010 = 100%)



Note: Deflated by the consumer price index.

Figure 6
The MNB's nominal house price index for cities by region
(2010 = 100%)

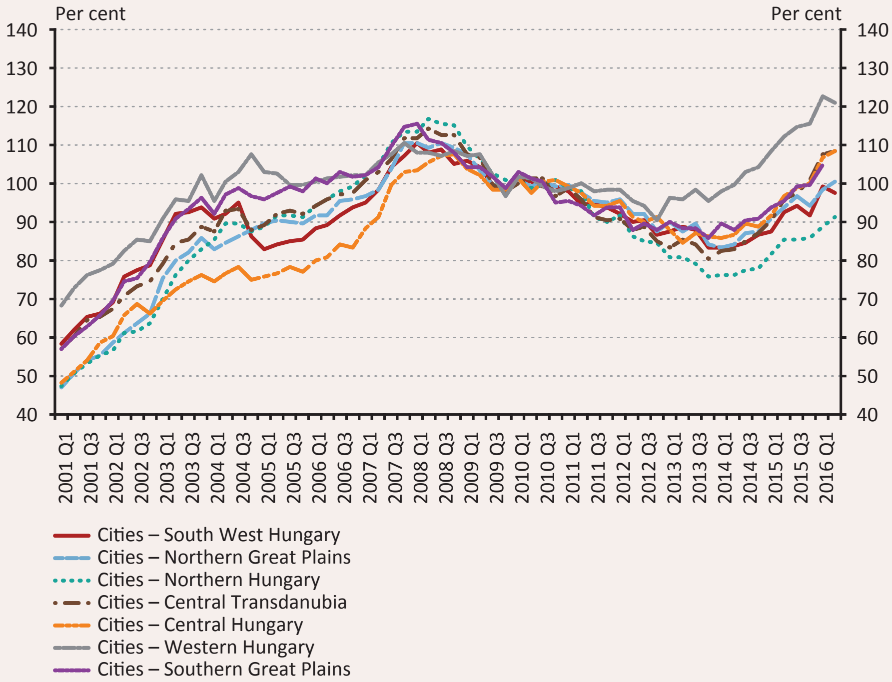
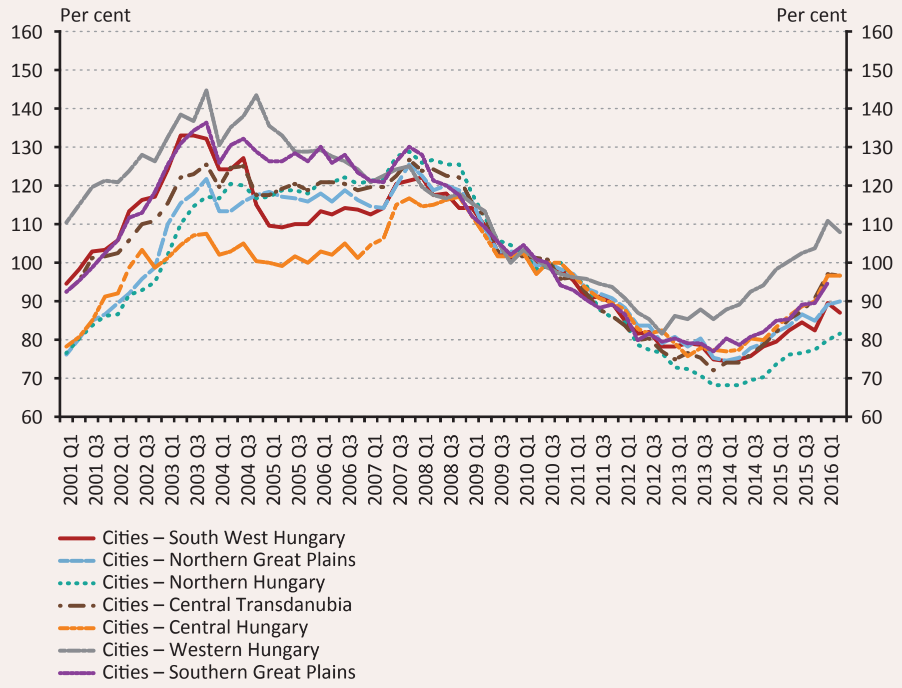
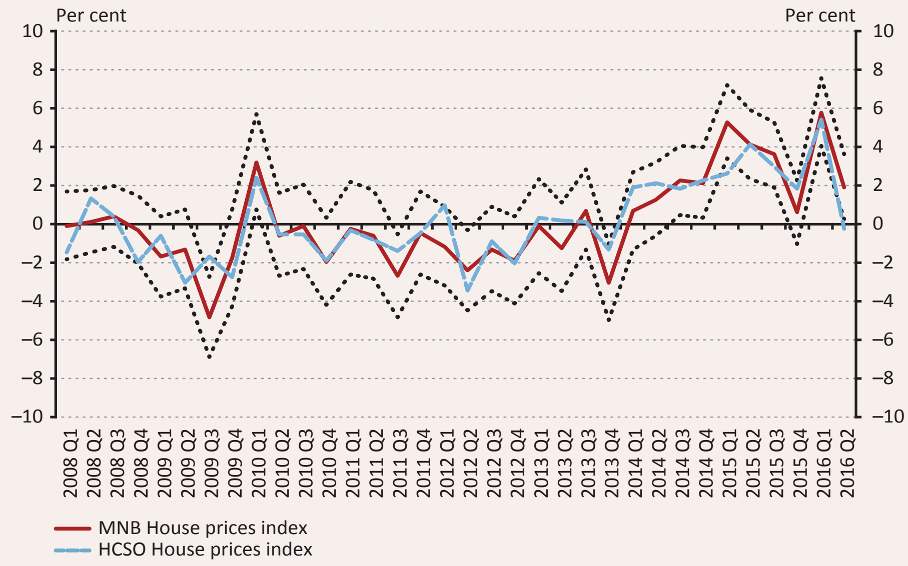


Figure 7
The MNB's real house price index for cities by region
(2010 = 100%)



The MNB's aggregate house price index is consistent with the house price indices constructed by the HCSO. At present, the HCSO compiles a separate house price index for new and used houses, and also publishes a national house price index through the Eurostat's database. Figure 8 indicates that the house price changes reflected in the HCSO's national index show similar dynamics. The differences between the methodologies of the HCSO and MNB indices and their effect on the estimated quarterly price changes are presented in more detail in the chapter describing the robustness analysis.

Figure 8
Quarterly change of the MNB's aggregate house price index and the HCSO's national house price index



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

5.2 REGRESSION RESULTS

In this chapter, we present the regression results of the house price index. 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 catchment area apart from the Budapest agglomeration. On the other hand, since the regression model is run 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 6 shows the regression outputs of the Budapest model. Run on a sample covering 2015 Q3 and 2015 Q4, the model has a 69 per cent explanatory power.[17] The dummy variable denoting 2015 Q4 shows that the mere fact that a Budapest transaction took place in 2015 Q4 as opposed to 2015 Q3 increased the price by $\varepsilon^{0.0278}=1.028$; in other words, in 2015 Q4 the pure price change in the Budapest housing market was 2.8 per cent.

As in the other models, 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, to examine the estimated coefficients alone is not instrumental. For the model specified for Budapest, Figure 18 in the Annex illustrates the combined partial effect of the linear and squared terms on the price by property type. Table 7, in turn, indicates average partial effects.

Evidently, a 1 per cent increase in NIA has the greatest positive impact on the value of detached houses. Moreover, 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 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 of the Budapest regressions is the fact that compared to the first district, only the fifth district has a price increasing effect; in the rest of the districts, residential properties tend to be cheaper on average. It is another intuitive result that new residential properties are, *ceteris paribus,* more expensive on average.

Table 8 depicts the regression results of the model constructed for Southern Transdanubian cities, also run on a sample of 2015 Q3 and 2015 Q4 transactions. 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 brick or panel buildings within the flat category (Table 7 and Figure 19 in the Annex).

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

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

Table 6
Regression results of the Budapest house price index model for 2015 Q4

Number of obs	=	23386
F(32, 23353)	=	1602,81
Prob > F	=	0
R-squared	=	0,6871
Adj R-squared	=	0,6867
Root MSE	=	0,3176

Price (ln)	**Coef.**	**Std. Err.**	**t**	**P>|t|**	**[95% Conf.**	**Interval]**
Quarter (reference: 2015 Q3)						
2015 Q4	0.0278	0.0042	6.6700	0.0000	0.0196	0.0360
Type of property * size of property (ln)						
Condonimium	1.2556	0.0767	16.3600	0.0000	1.1053	1.4060
Panel block of flats	1.4491	0.0819	17.7000	0.0000	1.2886	1.6096
Detached house (inner city)	1.0233	0.0718	14.2500	0.0000	0.8825	1.1640
Detached house (outer city)	1.1667	0.0692	16.8700	0.0000	1.0311	1.3023
Type of property * (size of property(ln))2						
Condonimium	–0.0416	0.0097	–4.3000	0.0000	–0.0606	–0.0227
Panel block of flats	–0.0954	0.0117	–8.1500	0.0000	–0.1183	–0.0725
Family house (inner city)	0.0124	0.0095	1.3000	0.1920	–0.0062	0.0311
Family house (outer city)	–0.0265	0.0083	–3.2000	0.0010	–0.0427	–0.0102
Districts of Budapest (reference: 1)						
2	0.0361	0.0197	1.8300	0.0670	–0.0026	0.0747
3	–0.3261	0.0187	–17.4200	0.0000	–0.3628	–0.2894
4	–0.4888	0.0194	–25.2400	0.0000	–0.5267	–0.4508
5	0.2110	0.0214	9.8700	0.0000	0.1691	0.2528
6	–0.1000	0.0198	–5.0600	0.0000	–0.1387	–0.0613
7	–0.2811	0.0190	–14.8000	0.0000	–0.3183	–0.2439
8	–0.4511	0.0188	–23.9800	0.0000	–0.4880	–0.4142
9	–0.2544	0.0196	–12.9500	0.0000	–0.2929	–0.2159
10	–0.6225	0.0197	–31.5400	0.0000	–0.6612	–0.5838
11	–0.1284	0.0185	–6.9500	0.0000	–0.1646	–0.0921
12	–0.0038	0.0205	–0.1900	0.8530	–0.0439	0.0363
13	–0.1747	0.0186	–9.4000	0.0000	–0.2111	–0.1382
14	–0.3210	0.0185	–17.3300	0.0000	–0.3574	–0.2847
15	–0.5841	0.0199	–29.3500	0.0000	–0.6231	–0.5451
16	–0.3297	0.0219	–15.0700	0.0000	–0.3725	–0.2868
17	–0.5175	0.0211	–24.5200	0.0000	–0.5588	–0.4761
18	–0.5632	0.0199	–28.2700	0.0000	–0.6023	–0.5242
19	–0.5846	0.0212	–27.5800	0.0000	–0.6261	–0.5430
20	–0.6932	0.0208	–33.2700	0.0000	–0.7340	–0.6523
21	–0.7378	0.0199	–37.0600	0.0000	–0.7768	–0.6988
22	–0.4066	0.0222	–18.3100	0.0000	–0.4501	–0.3631
23	–1.0009	0.0373	–26.8300	0.0000	–1.0741	–0.9278
New flat (reference: used)	0.3507	0.0162	21.6400	0.0000	0.3189	0.3824
Constant	12.5399	0.1524	82.3000	0.0000	12.2413	12.8386

Table 7
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				
Condonimium	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				
Condonimium	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
Municipalities				
Condonimium	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

Table 8
Regression results of the Southern Transdanubian house price index model for 2015 Q4

Number of obs	=	4300					
F(21, 4278)	=	342.65					
Prob > F	=	0					
R-squared	=	0.6271					
Adj R-squared	=	0.6253					
Root MSE	=	0.4021					

Price (ln)	**Coef.**	**Std. Err.**	**t**	**P>|t|**	**[95% Conf.**	**Interval]**
Quarter (reference: 2015 Q3)						
2015 Q4	–0.0270	0.0124	–2.1800	0.0290	–0.0512	–0.0028
Type of property * size of property (ln)						
Condonimium	1.2706	0.4071	3.1200	0.0020	0.4725	2.0686
Panel block of flats	1.2105	0.4295	2.8200	0.0050	0.3685	2.0526
Family house (inner city)	–0.1041	0.3688	–0.2800	0.7780	–0.8272	0.6189
Family house (outer city)	–0.7235	0.3734	–1.9400	0.0530	–1.4556	0.0086
Type of property * size of property (ln))2						
Condonimium	–0.0472	0.0503	–0.9400	0.3480	–0.1458	0.0513
Panel block of flats	–0.0382	0.0570	–0.6700	0.5030	–0.1499	0.0736
Family house (inner city)	0.2494	0.0428	5.8200	0.0000	0.1654	0.3334
Family house (outer city)	0.3816	0.0436	8.7500	0.0000	0.2961	0.4671
County (reference: Baranya)						
Somogy	0.0635	0.0308	2.0600	0.0390	0.0031	0.1238
Tolna	0.0123	0.0429	0.2900	0.7740	–0.0717	0.0964
New flat (reference: used)	0.3202	0.0417	7.6800	0.0000	0.2385	0.4019
Agglomeration (reference: not agglomeration)						
Agglomeration of Pécs	0.2244	0.0409	5.4900	0.0000	0.1442	0.3045
Recreational area (reference: not recreational area)						
Lake Balaton - near shore	0.8919	0.0645	13.8300	0.0000	0.7655	1.0182
Lake Balaton - rest	–0.0044	0.0528	–0.0800	0.9330	–0.1079	0.0990
Total income per capita (ln)	1.0131	0.0605	16.7500	0.0000	0.8945	1.1317
Distance from Budapest(ln)	–0.0978	0.0771	–1.2700	0.2050	–0.2490	0.0534
Distance from county seats (ln)	–0.0801	0.0090	–8.8600	0.0000	–0.0979	–0.0624
Population(ln)	–0.1172	0.0200	–5.8500	0.0000	–0.1565	–0.0780
Size of municipalities(ln)	0.1098	0.0251	4.3700	0.0000	0.0605	0.1591
Local home support (ln)	–0.0105	0.0034	–3.0700	0.0020	–0.0172	–0.0038
Constant	–1.0344	1.2943	–0.8000	0.4240	–3.5719	1.5031

Importantly, recreation areas located closer to the shore of Lake Balaton have a significant price increasing effect, whereas locations close to – but not directly on the shore of – the lake do not have a significant price effect. In the Southern Transdanubian region, the distance from Budapest does not influence house prices considerably, which may be because the price increasing effect of municipalities on the southern shore of Lake Balaton – which can be accessed more easily from Budapest than other areas in the Southern Transdanubian region – may be somewhat absorbed by the Balaton recreation area. The accessibility of the given county seat, however, proved to be an important factor. Moreover, an increase in net per capita income for the municipality is combined with a higher transaction price on average, while the amount of 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 subsidy amounts and the ratio of socially disadvantaged households and hence, the ratio of lower-quality properties.

In Table 9, we present the results of the model estimated for villages on a sample covering 2015 Q3 and 2015 Q4. 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 7 and Figure 20 in the Annex). 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; in all other counties residential homes are, *ceteris paribus,* cheaper. Among recreational areas, locations close to the shore of Lake Balaton exhibit the strongest price increasing effect and, as opposed to cities, even 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, 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 municipality-level variables is fairly high; the model's 74 per cent explanatory power suggests that the inclusion of municipality-level variables contributes the largest share of value added in smaller, typically rather heterogeneous municipalities.

Table 9
Regression results of the village house price index model for 2015 Q4

Number of obs	=	16006					
F(46, 15959)	=	974,21					
Prob > F	=	0					
R-squared	=	0,7374					
Adj R-squared	=	0,7366					
Root MSE	=	0,5868					
Price (ln)		**Coef.**	**Std. Err.**	**t**	**P>|t|**	**[95% Conf.**	**Interval]**
Quarter (reference: 2015 Q3)							
2015 Q4		–0.0140	0.0093	–1.5000	0.1340	–0.0323	0.0043
Type of property * size of property (ln)							
Condonimium		13.6006	0.3561	38.2000	0.0000	12.9027	14.2986
Municipality		11.0535	0.3412	32.4000	0.0000	10.3847	11.7223
Family House		10.8612	0.3288	33.0300	0.0000	10.2166	11.5057
Type of property * size of property (ln))2							
Condonimium		–1.4847	0.0442	–33.6300	0.0000	–1.5712	–1.3981
Municipality		–0.9559	0.0453	–21.1000	0.0000	–1.0447	–0.8671
Family House		–0.8908	0.0370	–24.0800	0.0000	–0.9634	–0.8183
County (reference: Pest)							
Györ-Moson-Sopron		0.0581	0.0416	1.3900	0.1630	–0.0235	0.1397
Vas		0.0515	0.0475	1.0900	0.2780	–0.0415	0.1446
…							
New flat (reference: used)		0.3895	0.0827	4.7100	0.0000	0.2274	0.5515
Agglomeration (reference: not agglomeration)							
Szeged Agglomeration		0.4817	0.0542	8.8900	0.0000	0.3755	0.5880
Pécs Agglomeration		0.4383	0.0586	7.4800	0.0000	0.3235	0.5531
Debrecen Agglomeration		0.5747	0.0606	9.4800	0.0000	0.4559	0.6935
Miskolc Agglomeration		0.3117	0.0509	6.1200	0.0000	0.2119	0.4115
Székesfehérvár Agglomeration		0.0239	0.0430	0.5600	0.5780	–0.0603	0.1081
Budapest Agglomeration		–0.0293	0.0303	–0.9700	0.3330	–0.0886	0.0300
Györ Agglomeration		–0.3023	0.0399	–7.5800	0.0000	–0.3804	–0.2241
Sopron Agglomeration		0.6258	0.0845	7.4000	0.0000	0.4601	0.7915
Recreational area (reference: not recreational area)							
Lake Balaton - near shore		0.8194	0.0325	25.2200	0.0000	0.7557	0.8830
Lake Balaton - rest		0.3928	0.0314	12.5100	0.0000	0.3312	0.4543
Dunakanyar		0.1991	0.0275	7.2400	0.0000	0.1452	0.2530
Mátra-Bükk		0.3469	0.0327	10.6200	0.0000	0.2829	0.4110
Sopron-Kőszeghegyalja		0.1432	0.0555	2.5800	0.0100	0.0344	0.2520
Lake Tisza		0.0100	0.0437	0.2300	0.8190	–0.0756	0.0956
Lake Velence –Vértes		0.0309	0.0424	0.7300	0.4660	–0.0523	0.1141
Total income per capita (ln)		0.5281	0.0234	22.6100	0.0000	0.4823	0.5739
Distance from Budapest(ln)		–0.2668	0.0273	–9.7900	0.0000	–0.3202	–0.2134
Distance from county seats (ln)		–0.1884	0.0123	–15.2800	0.0000	–0.2126	–0.1642
Population(ln)		0.1156	0.0089	12.9500	0.0000	0.0981	0.1330
Size of municipalities(ln)		–0.0815	0.0101	–8.0300	0.0000	–0.1014	–0.0616
Constant		–20.8978	0.8095	–25.8100	0.0000	–22.4845	–19.3110

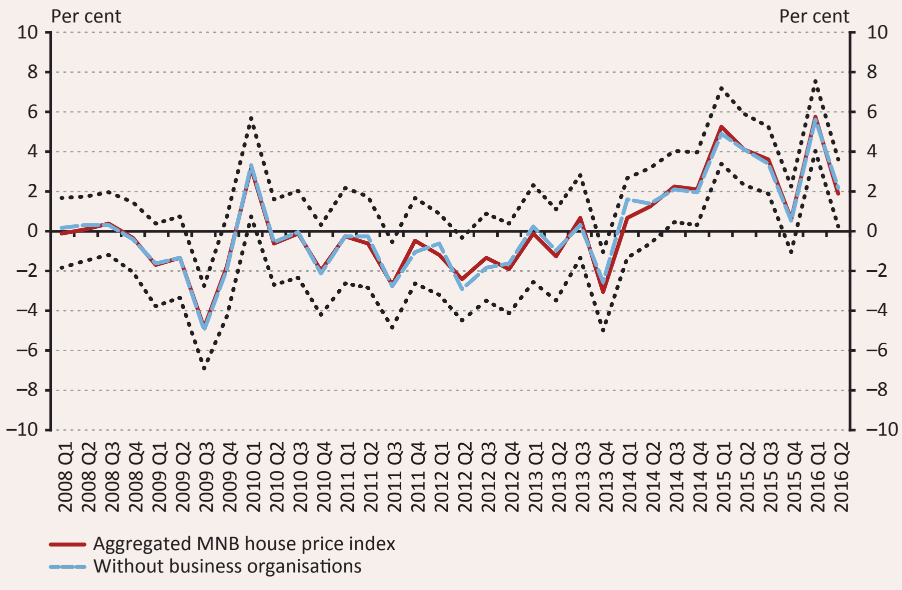
Note: While the model includes all counties as a dummy variable, due to space constraints, only the counties with a positive sign were presented in the table.

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 house price determinants.

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 (Figure 9). Such transactions, as mentioned before, form an integral part of housing market turnover; consequently, we retained them in the final models.

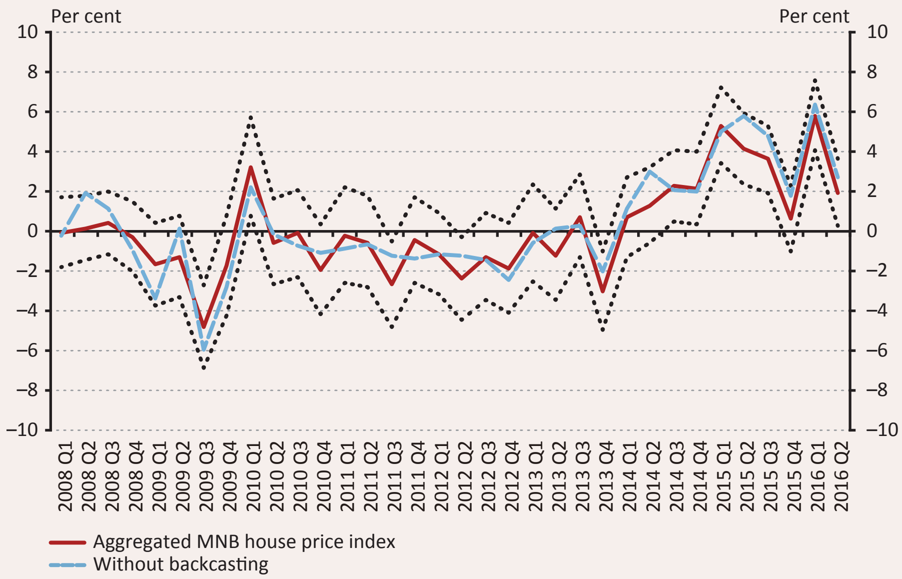
Figure 9
Robustness analysis for the filtering out of transactions by business organisations
(quarterly price changes)



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

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 10). 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.

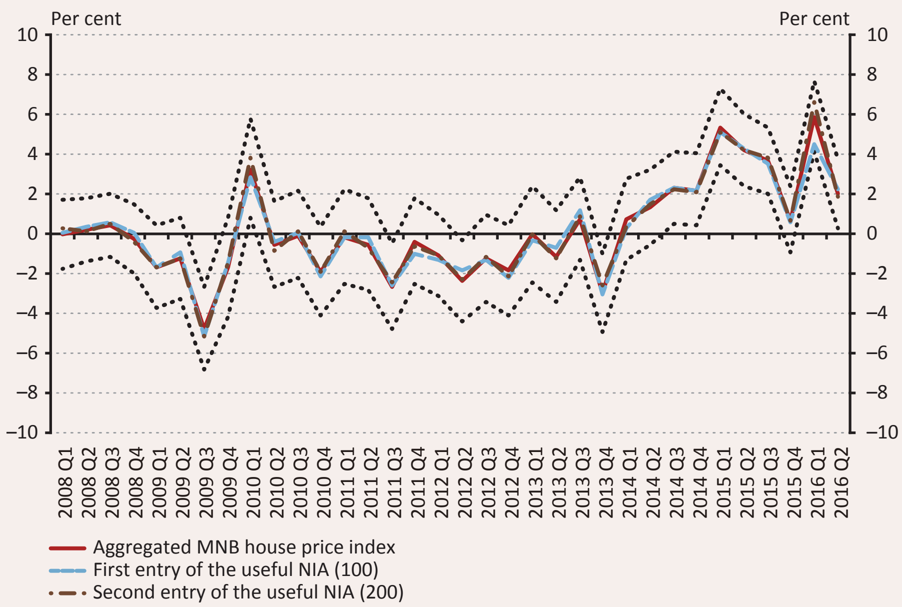
Figure 10
Robustness analysis for the backcasting of the useful NIA
(quarterly price changes)



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

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 (Figure 11).

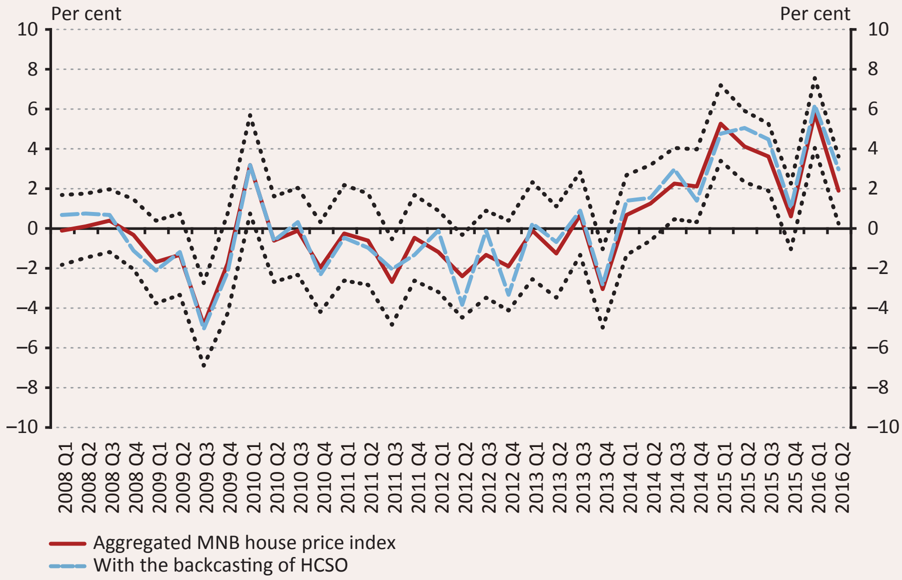
Figure 11
Robustness analysis for the first entry of the useful NIA
(quarterly price changes)



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

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 Chapter 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 (Figure 12).

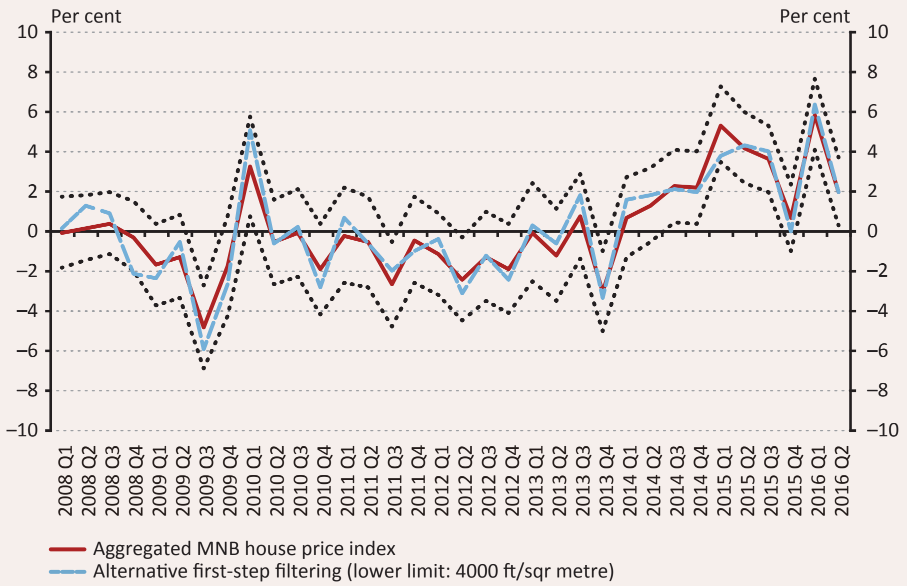
Figure 12
Robustness analysis for the backcasting of the missing values of the NIA variable
(quarterly price changes)



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

Upon testing the methodology of outlier filtering, in the first step we examined 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. According to Figure 13, this modification increases the volatility of the estimated price change in certain parts of the reference horizon.

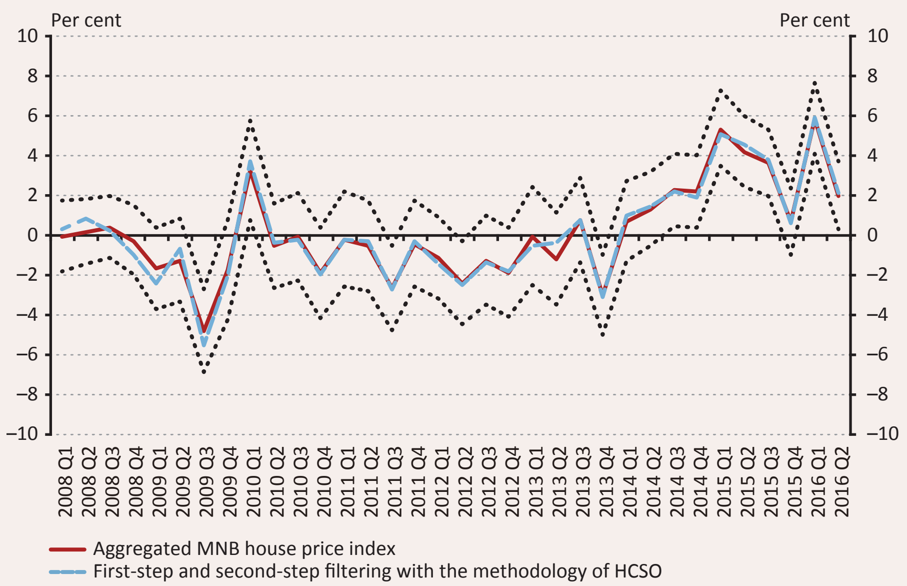
Figure 13
Robustness analysis for first-step filtering
(quarterly price changes)



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 step and in the second step relative to the outlier filtering technique applied by the HCSO (Figure 14). We found that the latter may partly explain the differences between the house price indices of the HCSO and the MNB.

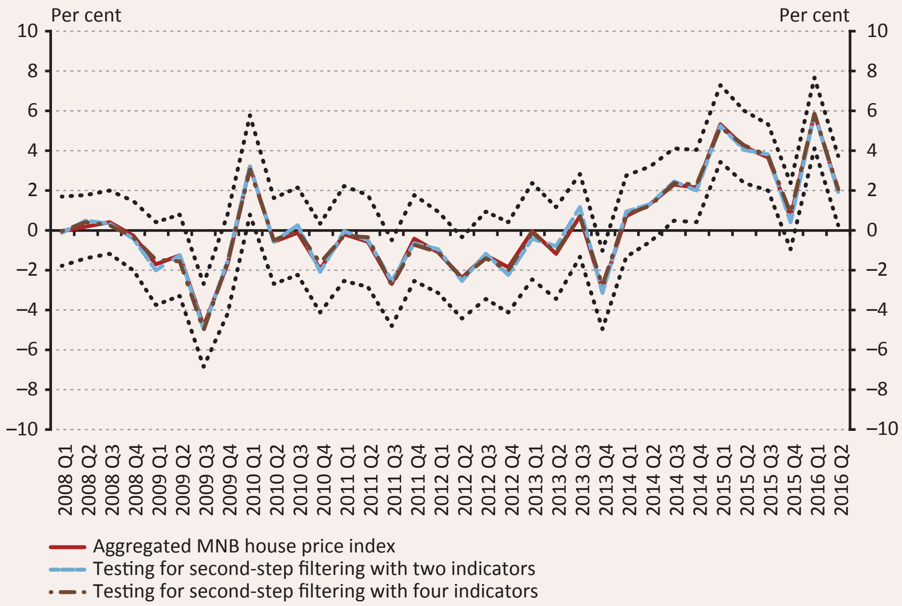
Figure 14
Robustness analysis for outlier filtering
(quarterly price changes)



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

In the second step of the outlier filtering used for the construction of the MNB's price indices, we retained the observations that were considered valid by at least three of the four indicators discussed in Sub-chapter 4.2. According to our calculations, it barely changes the quarterly price change values of the index (Figure 15) if the observations must be deemed valid by four or two indicators; therefore, our model is robust in this regard.

Figure 15
Robustness analysis for second-step filtering
(quarterly price changes)



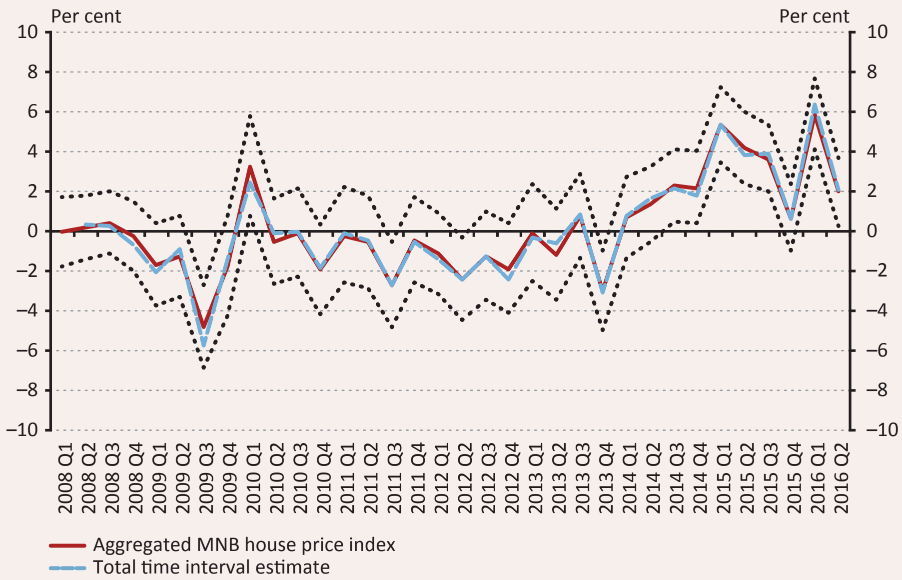
Note: The black dotted line indicates the 95 per cent confidence interval of the MNB's aggregate house index.

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 and 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 a significant impact on house price developments. Although the inclusion of the latter variables in the regressions only moderately influenced the outcome of the final aggregate house price index, they had a considerable impact on the indices constructed for cities and municipalities. Moreover, according to our calculations, the temporal fixation[19] 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 sub-samples created for individual municipality types, we also prepared estimates for the sample representing the whole of Hungary, and on the other hand, we made a disaggregated estimation on the sub-samples, but instead of adjacent-period pairs, the estimates covered the entire time horizon (Figure 16). We found that the results did not deviate significantly from the final house price indices.

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

Figure 16
Robustness analysis for the estimation methodology
(quarterly price changes)



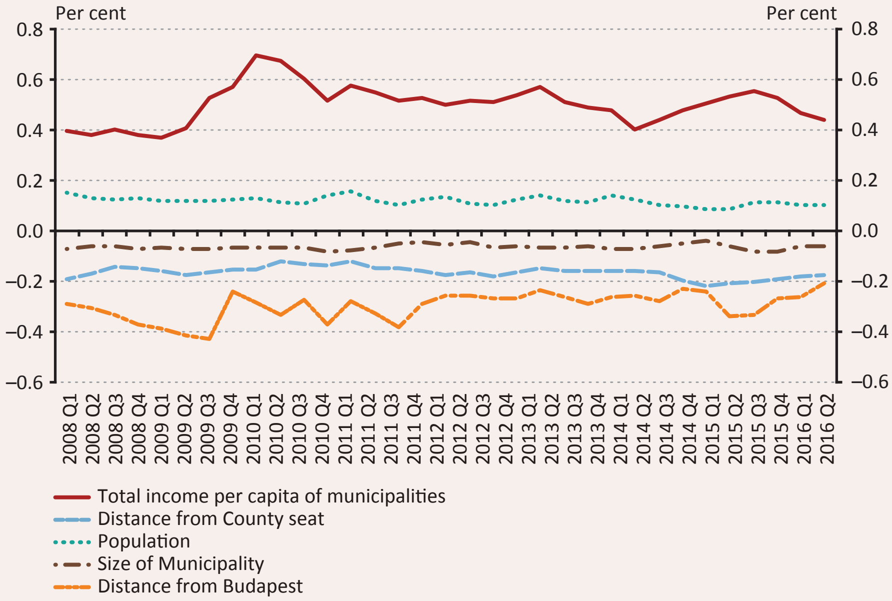
Note: The black dotted line indicates the 95 per cent confidence interval of the MNB's aggregate house index.

Besides the robustness of the results, we finally examined the stability of the explanatory variables' parameters over time.[20] Although one of the main advantages of the use of adjacent-period estimates is specifically the fact that the 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 municipality-level characteristics proved to be especially relevant; it is therefore important to examine whether the parameters of these municipality-level characteristics are stable in size and magnitude over time. Figure 17 depicts the coefficients of the municipality-level variables included in the models constructed for villages. The figure indicates that the per capita income and population of the municipality had a positive parameter 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. Neither municipality-level variable had a coefficient that changed its sign over time, which points to the stability of the correlation between these variables and house prices.

[20] 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 17
Dynamic analysis of the parameters of selected municipality-level variables in the models constructed for villages



7 Conclusions

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 housing market processes. For constructing the index, we used the property acquisition duty data collected by the National Tax and Customs Administration in relation to the transfer of residential properties. The MNB has specifically adapted – suitable for individual identification – the total duty data base from the NTCA for the reference period 1990–2015, which resulted in the longest and broadest Hungarian housing market database available so far. In the future, the database will be continuously updated with the HCSO's assistance. In our study, we presented the resulting indices and 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 index family constructed by the MNB 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 prices still fall short of pre-crisis levels in real terms, and even Budapest prices were unable to approach their 2008 levels until 2016. In addition, 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 support 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 market participants in their investment decisions or portfolio evaluations.

References

Australian Bureau of Statistics (ABS) (2006): A Guide to House Price Indexes. Information Paper, Cat. No. 6464

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

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

Bollen, K. A. – Jackman, R. W. (1990): Regression diagnostics: An expository treatment of outliers and influential cases. In: *Modern Methods of Data Analysis* (eds.: J. Fox and J. S. Long), pp. 257–291. Sage. Newbury Park, CA

Chatterjee, S. – Hadi, A. S. (1986): Influential observations, high leverage points, and outliers in linear regression. *Statistical Science,* Vol. 1, pp. 379–393

Cook, R. D. (1977): Detection of influential observation in linear regression. *Technometrics,* Vol. 19, pp. 15–18

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

Crone, T. M. – Voith, R. P. (1992): Estimating House Price Appreciation: A Comparison of Methods. *Journal of Housing Economics,* 2(4), pp. 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),* 2013 edition

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

Griliches, Z. (1961, 1971a): Hedonic Price Indexes for Automobiles: An Econometric Analysis for Quality Change. In: Z. Griliches (ed.): *Price Indexes and Quality Change: Studies in New Methods of Measurement.* Harvard University Press. Cambridge, MA

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). *Statistical Review,* Vol. 85, Issue 3

Horváth, Á. – Székely, J. (2009): Applying Hedonic Methods in Observing Price Trends of Second-Hand Properties. *Statistical Review,* Vol. 87, Issue 6

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

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

Mark, J. H. – Goldberg, M. A. (1984): Alternative Housing Price Indices: An Evaluation. *American Real Estate and Urban Economics Association Journal,* 12(1), pp. 30–49

O'Brien, D. – Westermann, T. – Vidalis, N. (2014): *Review of the ECB's house price valuations framework.* Directorate General Economics, Prices and Costs Division, January 2014

Oikarinen, E. (2012): Empirical evidence on the reaction speeds of housing prices and sales to demand shocks. *Journal of Housing Economics,* 21(1), pp. 41–54

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

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

UK Department of the Environment (1982): A New Index of Average House Prices. *Economic Trends,* 348, pp. 134–138

Vellemen, P. F. – Welsch, R. E. (1981): Efficient computing of regression diagnostics. *The American Statistician,* Vol. 35, pp. 234–242

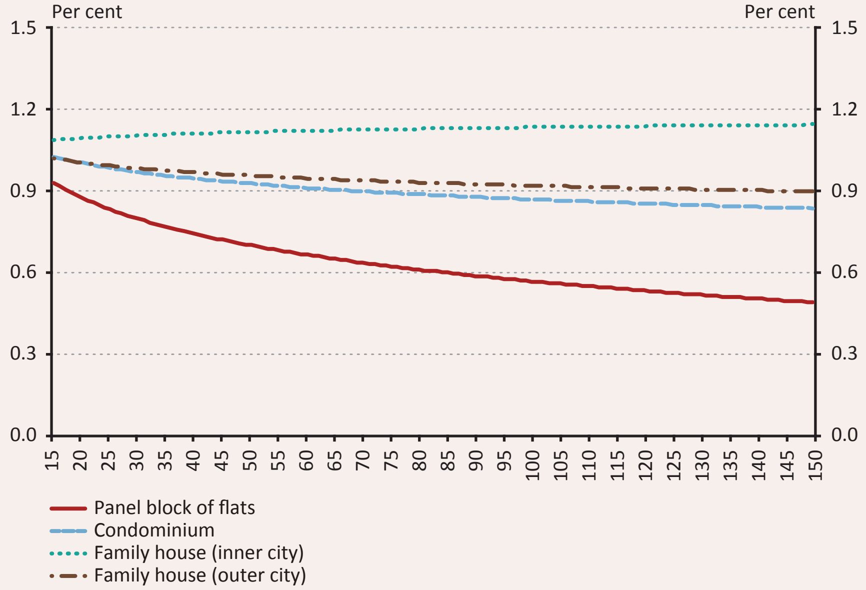
Wang, T. – Zorn, P. M. (1997): Estimating House Price Growth with Repeat Sales Data: What's the Aim of the Game? *Journal of Housing Economics,* 6(2), pp. 93–118

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

Zrínyi, M. – Katona, É. – Szántó, I. – Páll, D. (2012): A lineáris regressziót befolyásoló esetek diagnosztikája (Diagnostics of cases influencing linear regressions). *Statistical Review,* Vol. 90, Issue 7–8

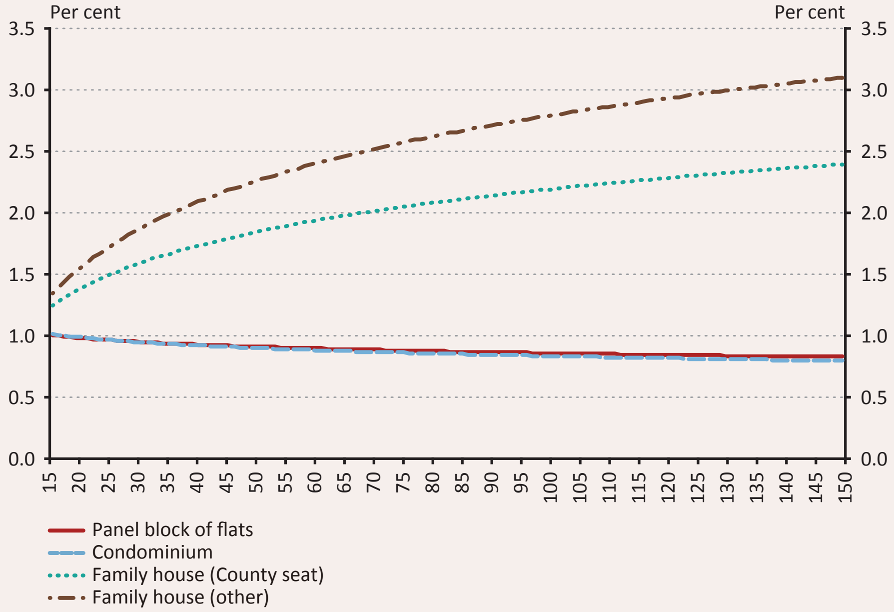
Annex

Figure 18
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 constant functions.

Figure 19
Partial effect of the interaction between NIA and property type in a model specified for Southern Transdanubian cities



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 constant functions.

Figure 20
Partial effect of the interaction between NIA and property type in a model specified for municipalities

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 constant functions.

Table 10
Distribution of category variables for the period of 1990–2000

		Budapest		Cities		Municipalities		Total	
		number of obs.	%	number of obs.	%	number of obs.	%	number of obs.	%
Property type	Family house			204,864	56.6	128,114	96.0	332,978	67.2
	County seat			51,585	14.2			51,585	10.4
	Other			153,279	42.3			153,279	30.9
	Flat			157,172	43.4	5,311	4.0	162,483	32.8
Agglomeration	Not			237,585	65.6	113,917	85.4	351,502	70.9
	Szeged			25,223	7.0	1,831	1.4	27,054	5.5
	Pécs			4,051	1.1	482	0.4	4,533	0.9
	Debrecen			25,786	7.1	1,649	1.2	27,435	5.5
	Miskolc			678	0.2	103	0.1	781	0.2
	Székesfehérvár			8,726	2.4	2,703	2.0	11,429	2.3
	Budapest			48,819	13.5	9,344	7.0	58,163	11.7
	Győr			7,880	2.2	2,937	2.2	10,817	2.2
	Sopron			3,289	0.9	459	0.3	3,748	0.8
Recreational area	Not			289,246	79.9	106,851	80.1	396,097	79.9
	Lake Balaton - near shore			8,963	2.5	2,903	2.2	11,866	2.4
	Lake Balaton - other			2,580	0.7	3,860	2.9	6,440	1.3
	Dunakanyar			20,604	5.7	6,265	4.7	26,869	5.4
	Mátra-Bükk			27,918	7.7	7,201	5.4	35,119	7.1
	Sopron-Kőszeghegyalja			6,079	1.7	1,270	1.0	7,349	1.5
	Lake Tisza			1,865	0.5	2,626	2.0	4,491	0.9
	Lake Velence–Vértes			4,782	1.3	2,449	1.8	7,231	1.5
County	Budapest								
	Baranya			6,475	1.8	2,315	1.7	8,790	1.8
	Bács-Kiskun			8,909	2.5	3,402	2.5	12,311	2.5
	Békés			18,487	5.1	3,476	2.6	21,963	4.4
	Borsod-Abaúj-Zemplén			1,221	0.3	700	0.5	1,921	0.4
	Csongrád			39,342	10.9	5,658	4.2	45,000	9.1
	Fejér			18,483	5.1	8,244	6.2	26,727	5.4
	Győr-Moson-Sopron			14,180	3.9	5,935	4.4	20,115	4.1
	Hajdú-Bihar			38,106	10.5	6,473	4.9	44,579	9.0
	Heves			35,964	9.9	24,230	18.2	60,194	12.1
	Komárom-Esztergom			31,794	8.8	9,453	7.1	41,247	8.3
	Nógrád			232	0.1	199	0.1	431	0.1
	Pest			63,103	17.4	22,050	16.5	85,153	17.2
	Somogy			20,525	5.7	13,288	10.0	33,813	6.8
	Szabolcs-Szatmár-Bereg			6,903	1.9	3,228	2.4	10,131	2.0
	Jász-Nagykun-Szolnok			21,571	6.0	6,729	5.0	28,300	5.7
	Tolna			8,111	2.2	4,375	3.3	12,486	2.5
	Vas			10,515	2.9	5,062	3.8	15,577	3.1
	Veszprém			6,579	1.8	2,609	2.0	9,188	1.9
	Zala			11,537	3.2	5,999	4.5	17,536	3.5

Note: Values received after first-step outlier filtering.

Table 11
Distribution of category variables for the period of 2001–2007

		Budapest		Cities		Municipalities		Total	
		number of obs.	%	number of obs.	%	number of obs.	%	number of obs.	%
Property type	Family house	52,184	14.4	419,936	56.8	299,261	96.6	771,452	54.6
	County seat			102,739	13.9			102,753	7.3
	Other			317,197	42.9			317,240	22.5
	Flat	310,364	85.6	319,631	43.2	10,561	3.4	640,685	45.4
Agglomeration	Not	284,580	100.0	484,678	65.4	267,583	86.3	1,037,006	77.6
	Szeged			30,617	4.1	2,705	0.9	33,326	2.5
	Pécs			28,627	3.9	3,407	1.1	32,038	2.4
	Debrecen			39,596	5.3	2,817	0.9	42,418	3.2
	Miskolc			29,698	4.0	3,857	1.2	33,559	2.5
	Székesfehérvár			18,461	2.5	5,845	1.9	24,308	1.8
	Budapest			82,900	11.2	15,652	5.0	98,563	7.4
	Györ			17,232	2.3	7,157	2.3	24,391	1.8
	Sopron			9,554	1.3	1,084	0.3	10,639	0.8
Recreational area	Not	363,415	100.0	615,302	83.0	266,038	85.8	1,244,938	88.0
	Lake Balaton - near shore			17,325	2.3	5,570	1.8	22,897	1.6
	Lake Balaton - other			4,357	0.6	6,556	2.1	10,914	0.8
	Dunakanyar			32,111	4.3	11,160	3.6	43,275	3.1
	Mátra-Bükk			46,747	6.3	8,902	2.9	55,655	3.9
	Sopron-Kőszeghegyalja			15,634	2.1	2,916	0.9	18,552	1.3
	Lake Tisza			3,226	0.4	4,507	1.5	7,733	0.5
	Lake Velence–Vértes			6,661	0.9	4,458	1.4	11,120	0.8
County	Budapest	284,580	100.0					284,680	21.3
	Baranya			42,728	5.8	15,933	5.1	58,667	4.4
	Bács-Kiskun			46,256	6.2	17,676	5.7	63,938	4.8
	Békés			41,502	5.6	12,348	4.0	53,856	4.0
	Borsod-Abaúj-Zemplén			52,973	7.1	27,196	8.8	80,176	6.0
	Csongrád			50,211	6.8	8,939	2.9	59,157	4.4
	Fejér			41,415	5.6	17,895	5.8	59,316	4.4
	Györ-Moson-Sopron			34,489	4.7	14,765	4.8	49,259	3.7
	Hajdú-Bihar			63,513	8.6	13,521	4.4	77,043	5.8
	Heves			23,344	3.1	20,773	6.7	44,120	3.3
	Komárom-Esztergom			32,409	4.4	10,442	3.4	42,855	3.2
	Nógrád			12,072	1.6	12,008	3.9	24,082	1.8
	Pest			107,057	14.4	37,470	12.1	144,541	10.8
	Somogy			27,249	3.7	20,255	6.5	47,508	3.6
	Szabolcs-Szatmár-Bereg			40,785	5.5	23,949	7.7	64,740	4.8
	Jász-Nagykun-Szolnok			43,153	5.8	14,470	4.7	57,629	4.3
	Tolna			17,757	2.4	12,240	3.9	29,999	2.2
	Vas			19,907	2.7	9,134	2.9	29,044	2.2
	Veszprém			25,451	3.4	10,410	3.4	35,864	2.7
	Zala			19,092	2.6	10,683	3.4	29,778	2.2

Note: Values received after first-step outlier filtering.

Table 12
Distribution of category variables for the period after 2008

		Budapest		Cities		Municipalities		Total	
		number of obs.	%	number of obs.	%	number of obs.	%	number of obs.	%
Property type	Family Houses	44,574	13.1	303,270	47.6	247,671	95.6	595,576	47.9
	Inner city of Budapest	12,066	3.6					12,070	1.0
	Outer city of Budapest	32,508	9.6					32,518	2.6
	County seat			69,584	10.9			69,595	5.6
	Other cities			233,686	36.7			233,723	18.8
	Condominium	276,948	81.5	256,275	40.2	11,380	4.4	544,725	43.8
	Panel block of flats	18,104	5.3	77,686	12.2			95,808	7.7
	Homestead					7,023	2.7	7,023	0.6
New/used	New	7,710	2.3	14,868	2.3	2,045	0.8	24,623	2.0
	Used	332,481	97.7	622,543	97.7	264,106	99.2	1,219,130	98.0
Districts of Budapest	I	6,051	1.8					6,051	1.8
	II	17,371	5.1					17,371	5.1
	III	22,711	6.7					22,711	6.7
	IV	16,716	4.9					16,716	4.9
	V	8,638	2.5					8,638	2.5
	VI	13,304	3.9					13,304	3.9
	VII	18,202	5.4					18,202	5.4
	VIII	21,258	6.3					21,258	6.3
	IX	15,915	4.7					15,915	4.7
	X	14,444	4.3					14,444	4.3
	XI	30,253	8.9					30,253	8.9
	XII	11,361	3.3					11,361	3.3
	XIII	29,072	8.6					29,072	8.6
	XIV	28,358	8.3					28,358	8.3
	XV	12,060	3.6					12,060	3.6
	XVI	9,463	2.8					9,463	2.8
	XVII	10,558	3.1					10,558	3.1
	XVIII	14,397	4.2					14,397	4.2
	XIX	9,378	2.8					9,378	2.8
	XX	10,390	3.1					10,390	3.1
	XXI	10,663	3.1					10,663	3.1
	XXII	6,906	2.0					6,906	2.0
	XXIII	2,199	0.6					2,199	0.6

The table is continued on the next page.

		Budapest		Cities		Municipalities		Total	
		number of obs.	%	number of obs.	%	number of obs.	%	number of obs.	%
Agglomeration	Not	340,191	100.0	396,399	62.2	225,291	84.6	962,043	77.3
	Szeged			28,978	4.5	3,199	1.2	32,182	2.6
	Pécs			25,333	4.0	2,511	0.9	27,848	2.2
	Debrecen			33,811	5.3	2,032	0.8	35,848	2.9
	Miskolc			24,972	3.9	3,108	1.2	28,084	2.3
	Székesfehérvár			15,790	2.5	5,129	1.9	20,921	1.7
	Budapest			81,590	12.8	15,659	5.9	97,262	7.8
	Györ			19,773	3.1	7,788	2.9	27,564	2.2
	Sopron			10,765	1.7	1,434	0.5	12,201	1.0
Recreational Area	not	340,191	100.0	517,811	81.2	223,535	84.0	1,081,718	87.0
	Lake Balaton - near shore			21,023	3.3	7,451	2.8	28,477	2.3
	Lake Balaton - other			3,781	0.6	6,850	2.6	10,632	0.9
	Dunakanyar			30,382	4.8	9,530	3.6	39,917	3.2
	Mátra-Bükk			39,197	6.1	7,601	2.9	46,804	3.8
	Sopron-Kőszeghegyalja			16,575	2.6	3,515	1.3	20,093	1.6
	Lake Tisza			2,302	0.4	3,360	1.3	5,662	0.5
	Lake Velence–Vértes			6,340	1.0	4,309	1.6	10,650	0.9
County	Budapest	340,191	100.0					340,291	27.4
	Baranya			35,100	5.5	11,551	4.3	46,657	3.8
	Bács-Kiskun			45,608	7.2	17,886	6.7	63,501	5.1
	Békés			29,733	4.7	8,715	3.3	38,453	3.1
	Borsod-Abaúj-Zemplén			43,747	6.9	21,705	8.2	65,459	5.3
	Csongrád			44,632	7.0	9,900	3.7	54,539	4.4
	Fejér			33,606	5.3	15,063	5.7	48,674	3.9
	Györ-Moson-Sopron			39,028	6.1	17,105	6.4	56,139	4.5
	Hajdú-Bihar			50,998	8.0	9,984	3.8	60,990	4.9
	Heves			18,433	2.9	15,531	5.8	33,967	2.7
	Komárom-Esztergom			25,922	4.1	8,619	3.2	34,545	2.8
	Nógrád			8,126	1.3	9,182	3.4	17,309	1.4
	Pest			100,800	15.8	33,258	12.5	134,074	10.8
	Somogy			22,696	3.6	16,033	6.0	38,733	3.1
	Szabolcs-Szatmár-Bereg			28,343	4.4	18,257	6.9	46,604	3.7
	Jász-Nagykun-Szolnok			29,313	4.6	10,422	3.9	39,740	3.2
	Tolna			14,635	2.3	9,242	3.5	23,879	1.9
	Vas			18,536	2.9	8,735	3.3	27,274	2.2
	Veszprém			28,525	4.5	12,823	4.8	41,352	3.3
	Zala			19,630	3.1	12,140	4.6	31,773	2.6

Note: Values received after first-step outlier filtering.

Table 13
Descriptive statistics of continuous variables until 1990–2000

			Number of obs.	Mean	Standard deviation	Minimum	5th percentile	25th percentile	Median	75th percentile	95th percentile	Maximum
Price (HUF)	Budapest											
	Cities		362,037	3,253,945	4,177,062	10,000	449,000	1,300,000	2,350,000	4,000,000	8,769,000	400,000,000
	Municipalities		133,425	1,959,699	3,050,159	11,000	150,000	500,000	1,100,000	2,400,000	6,000,000	185,000,000
Population (capita)	Budapest and cities		362,803	62,429	102,095	0	5,421	15,266	33,203	73,648	203,648	2,018,035
	Municipalities		133,425	2,204	1,489	0	375	1,090	1,906	3,004	5,104	8,615
Size (sqkm)	Budapest and cities		362,803	15,658	12,674	0	2,586	6,163	10,477	20,997	46,165	52,516
	Municipalities		133,425	3,468	2,398	0	779	1,811	2,871	4,491	7,918	28,458
Subsidies (thousand HUF/year)	Budapest and cities		362,803	0	0	0	0	0	0	0	0	0
	Municipalities		133,425	0	0	0	0	0	0	0	0	0
Distance from Bp. (minutes)	Budapest and cities		362,803	105	49	0	34	57	106	151	186	234
	Municipalities		133,425	106	48	0	40	66	100	140	191	255
Distance from county seat (minutes)	Budapest and cities		362,803	27	26	0	0	0	26	49	74	109
	Municipalities		133,425	43	20	0	14	27	42	55	77	138
Income of settlement (HUF/capita/year)	Budapest and cities		362,788	296,550	60,029	105,255	190,035	256,644	299,903	330,796	392,599	542,522
	Municipalities		133,186	212,510	64,621	17,033	115,114	164,742	209,903	254,585	314,919	1,109,502
Net internal area of residential property (sqm)	Budapest	house										
		flat										
	Cities	house county seat	51,585	69	36	15	36	53	64	76	118	499
		house other	153,279	70	32	15	38	55	66	79	110	499
		flat	157,172	55	16	15	35	47	53	60	78	498
	Municipalities	house	128,114	70	23	15	45	59	69	79	100	498
		flat	5,311	55	21	15	34	45	51	59	88	472

Note: Values received after first-step outlier filtering.

Table 14
Descriptive statistics of continuous variables until 2001–2007

			Number of obs.	Mean	Standard deviation	Minimum	5th percentile	25th percentile	Median	75th percentile	95th percentile	Maximum
Price (HUF)	Budapest		363,415	14,013,435	15,189,257	69,360	3,000,000	7,400,000	10,500,000	16,460,000	33,250,000	665,000,000
	Cities		741,363	8,109,460	8,674,177	57,000	800,000	3,900,000	6,500,000	10,000,000	20,300,000	763,440,000
	Municipalities		310,107	4,519,724	6,750,351	57,000	250,000	1,000,000	2,800,000	5,585,000	15,000,000	602,844,032
Population (capita)	Budapest and cities		1,104,778	601,809	775,145	1,085	6,074	21,291	81,818	1,697,343	1,719,342	1,739,569
	Municipalities		310,107	2,144	1,564	0	330	980	1,833	2,900	5,201	9,983
Size (sqkm)	Budapest and cities		1,104,778	27,496	19,889	575	3,106	9,145	21,673	52,512	52,516	52,516
	Municipalities		310,107	3,417	2,405	0	784	1,701	2,778	4,488	8,020	28,458
Subsidies (thousand HUF/year)	Budapest and cities		1,104,778	172,818	276,316	0	0	0	5,100	378,960	831,095	831,095
	Municipalities		310,107	300	1,258	0	0	0	0	150	1,575	100,000
Distance from Bp. (minutes)	Budapest and cities		1,104,778	77	68	0	0	0	75	140	180	234
	Municipalities		310,107	122	51	0	44	81	120	163	203	255
Distance from county seat (minutes)	Budapest and cities		1,104,778	18	25	0	0	0	0	37	66	109
	Municipalities		310,107	44	22	0	14	28	42	57	88	138
Income of settlement (HUF/capita/year)	Budapest and cities		1,104,778	523,941	143,657	124,979	293,701	417,290	516,128	621,751	803,533	965,277
	Municipalities		309,975	344,009	128,432	16,731	167,056	250,738	325,811	419,383	585,803	1,271,065
Net internal area of residential property (sqm)	Budapest	house	52,184	73	36	15	30	53	69	86	130	499
		flat	310,364	55	25	15	27	40	51	65	98	498
	Cities	house county seat	102,739	69	30	15	35	53	65	78	120	495
		house other	317,197	70	27	15	38	55	68	80	110	499
		flat	319,631	56	19	15	34	48	54	61	85	498
	Municipalities	house	299,261	73	22	15	45	60	72	83	103	496
		flat	10,561	61	26	15	34	47	55	68	108	483

Note: Values received after first-step outlier filtering.

Table 15
Descriptive statistics of continuous variables from 2008

			Number of obs.	Mean	Standard deviation	Minimum	5th percentile	25th percentile	Median	75th percentile	95th percentile	Maximum
Price (HUF)	Budapest		340,191	16,664,362	17,662,256	90,000	4,250,000	8,450,000	12,500,000	19,653,180	40,000,000	850,000,000
	Cities		637,411	9,964,693	9,301,686	79,328	1,250,000	5,000,000	7,800,000	12,500,000	25,000,000	660,000,000
	Municipalities		266,151	6,074,002	8,484,386	79,400	343,000	1,500,000	3,500,000	7,500,000	20,000,000	441,249,984
Population (capita)	Budapest and cities		977,602	644,731	800,110	1,000	6,007	23,573	111,836	1,721,556	1,757,618	1,757,618
	Municipalities		266,151	2,081	1,652	0	283	891	1,745	2,759	5,128	10,282
Size (sqkm)	Budapest and cities		977,602	13,586	19,582	6	46	228	525	19,393	52,513	52,513
	Municipalities		266,151	1,718	2,382	0	10	27	527	2,717	6,581	28,458
Subsidies (thousand HUF/year)	Budapest and cities		977,602	50,947	86,976	0	0	0	4,190	58,090	303,985	378,960
	Municipalities		266,151	194	3,438	0	0	0	0	0	800	250,000
Distance from Bp. (minutes)	Budapest and cities		977,602	71	66	0	0	0	63	128	178	236
	Municipalities		266,151	118	50	0	38	78	117	157	200	259
Distance from county seat (minutes)	Budapest and cities		977,602	17	25	0	0	0	0	34	67	113
	Municipalities		266,151	44	23	0	14	28	42	57	87	149
Income of settlement (HUF/capita/year)	Budapest and cities		977,602	794,693	142,988	283,727	543,267	692,342	819,877	885,874	971,593	1,348,615
	Municipalities		266,150	583,976	191,088	-47,703	318,639	449,726	564,139	691,716	926,838	2,531,264
Net internal area of residential property (sqm)	Budapest	Condominium	276,948	56	24	15	27	39	52	66	100	499
		Panel block of flats	18,104	53	17	15	30	43	52	61	77	409
		Family house inner districts	12,066	112	69	15	39	70	100	128	250	499
		Family house outer districts	32,508	102	63	15	44	74	93	109	198	499
	Cities	Condominium	256,275	58	21	15	33	47	55	65	95	499
		Panel block of flats	77,686	53	13	15	35	47	53	59	73	471
		Family house county seat	69,584	89	42	15	45	69	83	97	156	498
		Family house other	233,686	83	33	15	50	69	80	90	130	499
	Municipalities	Condominium	11,380	66	30	15	34	49	60	76	119	485
		Homestead	7,023	70	23	20	49	59	67	76	96	496
		Family house	247,671	79	26	15	50	65	76	87	114	498

Note: Values received after first-step outlier filtering.

MNB OCCASIONAL PAPERS 127

THE MNB'S HOUSE PRICE INDEX METHODOLOGY

May 2017

Print: Prospektus–SPL consortium

6 Tartu u., Veszprém H-8200