

VALUATION METHODS FOR THE HOUSING MARKET: EVIDENCE FROM BUDAPEST

Dávid KUTASI – Milán Csaba BADICS

*(Received: 31 March 2015; Revision received: 7 August 2015,
accepted: 3 October 2015)*

Different valuation methods and determinants of housing prices in Budapest, Hungary are examined in this paper in order to describe price drivers by using an asking price dataset. The hedonic regression analysis and the valuation method of the artificial neural network are utilised and compared using both technical and spatial variables. In our analyses, we conclude that according to our sample from the Budapest real estate market, the Multi-Layer Preceptron (MLP) neural network is a better alternative for market price prediction than hedonic regression in all observed cases. To our knowledge, the estimation of housing price drivers based on a large-scale sample has never been explored before in Budapest or any other city in Hungary in detail; moreover, it is one of the first papers in this topic in the CEE region. The results of this paper lead to promising directions for the development of Hungarian real estate price statistics.

Keywords: housing prices, hedonic method, neural networks, Budapest residential market

JEL classification indices: G120, R310

Dávid Kutasi, corresponding author. PhD student at the Faculty of Architecture, Department of Construction Management, Budapest University of Technology and Economics.
E-mail: davidkutasi@ekt.bme.hu

Milán Csaba Badics, PhD student at the Faculty of Business Administration, Department of Finance, Corvinus University Budapest. E-mail: milancsaba.badics@uni-corvinus.hu

1. INTRODUCTION

Research on the housing price is important for property owners, developers, investors, banks, and policymakers. For the owners it is a valuable asset in their portfolio, and the appraisal of the value is important, especially if they are considering selling the property. For the developers and investors it is also vital to know current and potential future prices. It is also essential for banks giving mortgage and having real estate as a collateral to be aware of the housing prices to minimise the risks they encounter. The immense number of real estate as a collateral motivates banks to create a well-functioning appraisal system. Furthermore, it is an EU-requirement for Hungarian financial institutions to re-evaluate their real estate portfolio secured by mortgage at least once every three years.¹ Regarding macroeconomic and socio-economic considerations, it is vital for policymakers to be aware of housing prices as well. For instance, the recurring ideas of property tax or housing benefits also call for further analyses of the housing prices.

Research on this topic in the Central and Eastern European region is currently scarce (Pavlin 2006; Widlak – Tomczyk 2010).

Many valuation techniques exist that provide appraisals for the value of real estates. These methods were classified by Pagourtzi into traditional and advanced methods (*Table 1*). The simple traditional methods are either based on comparing the observed property to other similar properties in the same neighbourhood (Comparable method), or on creating a Discounted Cash Flow (DCF) model of the assets' incomes and expenses (Income method, Profit method). In this classification, the multiple regression method and the stepwise regression method are also considered to be traditional. The advanced techniques require statistical

Table 1. Classification of real estate valuation methods

Traditional valuation methods	Advanced valuation methods
Comparable method	Hedonic pricing method
Investment/Income method	Spatial analysis method
Profit method	Artificial neural networks (ANN)
Development/Residual method	Fuzzy logic
Constructor's/cost method	Autoregressive integrated moving average (ARIMA)
Multiple regression method	
Stepwise regression method	

Source: Pagourtzi et al. 2003.

¹ Regulation No. 575/2013 of the European Parliament and of the Council of June 26, 2013, on prudential requirements for credit institutions and investment firms and amending Regulation (EU) No. 648/2012 Article 208.

and data mining knowledge to estimate real estate prices. Hedonic and spatial methods determine and quantify the technical and geographical components of the value (hedonic pricing method, spatial analysis method), while other techniques use data mining methods for the description of the value (artificial neural network, fuzzy logic).

The volume edited by Kauko – d’Amato (2009) offers a comprehensive overview on the possible mass appraisal methods with practical examples. The description of the methods range from the traditional comparable sales method, mentioning the geographically weighted regression and the hedonic modelling, to artificial neural networks and self-organising maps (Kohonen Map).

In this paper, we shall focus on two advanced methods and compare the hedonic regression and the artificial neural networks.

2. RELATED LITERATURE

In the wake of Rosen’s paper (1974), the *hedonic method* became widespread in scientific mass appraisal. The hedonic methodology is generally used for the valuation of goods in terms of their characteristics. Hedonic pricing models are very useful for determining how the price of a commodity varies with the set of attributes it possesses. If the marginal shadow prices of these attributes can be estimated, this methodology will provide a method for value appraisal.

By using the regression method, the importance of various components of a housing real estate can be measured. By adding weighted components, the value of the observed real estate can be determined. This method is undoubtedly applicable for mass appraisal through the repeated use of the regression function. The main purpose of the hedonic analysis is to determine most of the components that influence the market value significantly. The more significant the variable at one’s disposal, the more precise the attainable appraisal.

The variables affecting residential values can be classified into three groups: technical variables, geographic/spatial variables, and macroeconomic variables. The technical variables consist of the quality, material, design, construction technique, and architectural aspects. The geographic- or location-specific variables derive from the absolute and relative locations. The expression “absolute location” refers to the city, district, zone, or altitude, while “relative location” denotes the distance from a specific place or urban function such as a market, city centre, or shopping mall. The macroeconomic factors also have an impact on the value. Inflation, GDP growth, average real wage, and unemployment rate are all factors that could influence residential prices; however, for a given moment or for a short period of time they can be assumed to be constant. These variables are not consid-

ered in our cross-sectional model; further panel database analyses are dedicated to these issues. In this paper, we used 39 technical and geographical variables, which conforms to similar international research methodologies such as the study of Kain – Quigley (1970), who also determined 39 variables in their study.

The hedonic method has been employed in various fields, not only for housing valuations, but also for the valuation of agricultural fields (Brosen et al. 1984) and for the malting barley market (Wilson 1984). One objective of the hedonic model is to estimate the dependent variable and to quantify the effect of the explanatory variables. In addition, the model enables the observation of such attributes and features as the effect of noise, air pollution, airport noise, or neighbourhood amenities (Uyeno et al. 1993; Harrison – Rubinfeld 2013).

Most of the studies use the hedonic method to analyse the above-mentioned effects, the relationship between the price or market value and the housing characteristics. Furthermore, with the usage of regression, housing price indices were created. However, certain doubts regarding the main assumptions and estimations of the hedonic model surface. The previously used linear model might not clearly explain the link between the dependent and explanatory variables, as this relation might be non-linear.

For the description of the co-functioning of the dependent and explanatory variables as well as to dispel the doubts, the *artificial neural network model* (ANN) is utilised. ANN models provide powerful tools for statistical data analysis. Their best feature is their ability to “learn” dependencies based on observations and to discover relationships among the variables that are less obvious. In recent years, neural network applications in economics and finance (pattern recognition, classification, and time series forecasting) have increased rapidly.

Neural networks appear in various forms in both technical and economic applications. In the field of real estate valuation, the most frequently used model is the Multi-Layer Perceptron (MLP) that belongs to the feedforward neural networks. Networks usually have 3 or 4 layers: an input, a hidden, and an output layer. Neurons can be found in each layer. The number of the first layer’s neurons corresponds to the number of the model’s explanatory variables, while the last layer’s equals the number of the target variables (usually 2 neurons: binary target variable, or 1 neuron: continuous target variable). The model’s complexity and forecast ability is determined by the number of neurons in the hidden layer. Apart from the previously mentioned neurons, the input layer and the output layer consist of 1 neuron each (distortion) that have a value of 1 and have the same role as the constant in the regression model. Generally, every single neuron is connected to all the other neurons in the next layer, and these edges represent weights. Every neuron receives inputs from the previous layer and, by the use of a non-linear function, it transforms to the next layer’s input. Since a neuron network with a

hidden layer can model a complex problem, it will be used for the modelling (Dai et al. 2012).

The learning of the network commences with the use of randomly chosen weights. Due to the learning algorithm (usually back-propagation), the weights change during the iteration process. The aim of the algorithm is to find those weights that minimise the error function (usually MSE, RMSE or MAPE) between the target variable and the current variable. The learning process is generally conducted by the gradient method. Since the network with a certain number of neurons in the hidden layer can learn any relationship on the learning data (even the outliers and noise), the prevention of the over-learning can be achieved by halting the learning algorithm early. As a conclusion, it is essential to distribute the teaching, the test, and the validation segment in the beginning. The learning process of the network stops when the test segment has reached its minimum. Then, with the given parameter, the network has to be run on the validation segment.

Let the network have n input neurons, m hidden neurons, and 1 output neuron. The learning process is given in the following two steps (Wang et al. 2011):

First step (hidden layer): the output of the hidden layer is given with the following equation:

$$net_j = \sum_{i=0}^n v_{ij} x_i, \quad j = 1, 2, \dots, m,$$

where

$$y_j = f_H(net_j), \quad j = 1, 2, \dots, m,$$

Here net_j is the activation value of the j th neuron, y_j the output of the hidden layer, and f_H is the transfer function, which is generally a sigmoid function:

$$f_H(x) = \frac{1}{1 + \exp(-x)}$$

Second step: The value of output layer is given by the following function:

$$O = f_o \left(\sum_{j=0}^m w_{jk} y_j \right)$$

where f_o is the transfer function and is mostly linear.

Since the weights are randomly determined, the network gives different results in the case of multiple runs. In most cases, it is eliminated by running every network multiple times, and the results are averaged. Thus, on the one hand, the

forecasts are more robust, while on the other hand, the outlier network is eliminated.

One parameter must be optimised in the neural network: the number of neurons in the hidden layer. The cross-validation or the grid search is usually used for choosing the optimal parameters. The former solution is chosen in this paper.

Comparisons of Multiple regression approach (MRA) and ANN methods are widely used in the real estate valuation literature. Brost (1995) tested different neural network models relying on US real estate data. The mean absolute errors ranged from 8.7% to 12.4%. He concludes that neural networks deserve more attention in the field of real estate valuation, as they perform better than the multiple regression model.

The comparison of the two methods on a huge Singaporean residential property dataset indicated that ANN has a mean absolute error of 3.9%, while the MRA mean absolute error is 7.5%. The authors mentioned that the result may be significantly improved if the outliers are filtered (Tay – Ho 1992).

Do – Grudnitski (1992) observed data based on a sample of 105 assets from the US residential property market. The model included eight explanatory variables, with the selling price being the dependent variable. Using three nodes in the hidden layer, the neural network model is more accurate than MRA. While regression had 11.3% of mean absolute error, ANN had just 6.9%.

Worzala et al. (1995) are not fully convinced either that ANN always leads to a better result. Observing a sample of 288 sales in Fort Collins, Colorado, they conclude that the result of the neural network depends on the software, and every attempt provides different results.

Rossini (1997) used a South Australian sample to compare the methods. According to him, MRA is ahead of ANN, but the results are not conclusive. It should also be mentioned that the usage of the neural network and the software for the calculations in the 1990s were relatively new and rudimental.

Nguyen – Cripps (2001) observed family housing sales in Tennessee and concluded that ANN performs better than MRA if one provides sufficient data training size and appropriate ANN parameters. Otherwise, the results vary.

Peterson – Flanagan (2009) used a sample of 46,467 residential properties. They concluded that on a large sample, ANN generates significantly lower dollar pricing errors, has greater pricing precision out-of-sample, and extrapolates better from a more volatile pricing environment. The article mentions that multiple layered ANNs are capable of modelling complex non-linearities and have better results than the regression model.

Selim (2009) used a sample of 5,741 houses from the urban and the rural areas of Turkey to compare hedonic regression and ANN. The model included 46 explanatory variables, and the logarithm of selling price was the dependent vari-

able. The study demonstrates that ANN is a better alternative for predicting house prices.

From the previous literature review it is not unequivocal which method is ultimately better. The comparison and application of both methods are justified by the data of the Budapest housing market.

3. DATA AND VARIABLES

The data utilised in this study refer to asking house prices. It is not a common method, but in specific cases, it has already been employed (Falzon – Lanzon 2013). Because of the underdeveloped property registration system in Hungary, there is no official dataset providing sufficient information to perform an analysis on the Budapest real estate market. The official dataset made by the tax authorities and published by the Hungarian Central Statistical Office (CSO) is ineffective for a research of this type since the CSO publish only an average square meter price for those streets where at least three transactions happened in the given year. Previously, there was a study by Horváth – Székely (2009) that used 650,000 data from the CSO with a selling price for the period of 2001–2008 for Hungary. This paper had limited variables due to the lack of data: only the size, the city (or district), the type of house and the year of sale were involved. That was a reasonable way to have a country level view, but deeper analyses could not be performed due to data deficiency. Another paper from CSO (2005) made a regression on 6,624 houses with more variables, but they used the estimated value of the flat by the owner as a dependent variable.

Kauko (2007) published a study on the Budapest housing market based on Kohonen maps with the aim of comparing the results with Helsinki and Amsterdam. Data was used from CSO (2,087 observations) and Ecorys (215 observation), but the main goal was not to define the market value, but to shed light on how socio-demographic characteristics, price, and regulation are inter-related.

In order to attain the most relevant data for the Budapest housing market, a dataset based on the asking price from real estate agencies was used, which is modified according to the assumed difference between the supply and the transaction price.² Therefore, data from the largest on-line real estate website (www.ingatlan.com) were used, a website that had over 3,500 advertised apartments in April 2012. Both agencies and individuals may upload ads to this site and provide relevant technical and geographical information on the apartments. To avoid outlier

² The modification was based on Hungarian studies and real estate experts, who claim that the real transaction price is ~8% lower than the average supply price.

- figures and misleading numbers, the dataset was filtered through the following parameters:
1. Properties between the range of 35 to 200 square meters;
 2. Properties between HUF 5.5 to 100 million (~€19,000 to ~350,000);
 3. Properties without precise data on technical parameters or location were omitted.
 4. Properties with compromising information such as a huge mortgage burden or a perpetual lease were omitted.
 5. Properties that were mentioned in more than one ad were omitted. In these cases, the latest ad was kept, while the others were filtered.

After filtering, the dataset was supplemented with additional explanatory variables. New technical and geographical variables were added. From the ad descriptions, an additional garage and a balcony variable were created. Furthermore, geographical analyses were made with the Microsoft Mappoint and the MPMileCharter software products. The distances between the residential properties and the specific important places were calculated (such as shopping malls, cinemas, metro stations, etc.). Derived from the calculations involving more than 300 specific and 1,806 residential properties, new geographical variables were generated.

Besides the dependent variable, 23 technical and 16 geographical variables were used. Note, that the dependent variable was the logarithm of the modified supply price of the real estate.

The description and the descriptive statistics of the dependent, the non-location-specific, and the location-specific variables are presented in *Tables 2–5*.

Table 2. Description of variables³

Variables	Description
Price	The logarithm of price is used as dependent variable given in million HUF. The asking price is corrected with 8% bargaining possibility.
Area	Logarithm of total area in square meter ³ .
Balcony	Presence of a balcony – dummy variable.
Floor	The floor number of the apartment in 4 categories (0th–1st; 2nd–4th; 5–7th; 7–10th floor).
Construction method	Pre-fab panel building or other (brick, concrete) incorporated with the area.
Condition	Current condition of the flat (bad; average; good; very good; new). As the data was given by the advertiser, the consistency of the category is not assured. Area size also incorporated.

³ Logarithm is used, since additional square metres add more value if the property is small.

Heating	Heating system (electric; Hera system; convector; district heating; house central heating with or without individual meter; circulation heating; fan-coil heating).
# of rooms	Number of rooms (weighted sum of the semi- and normal rooms).
Parking	Presence of garage for the property included in the price.
Panorama	Panorama (street view, courtyard view, garden view, panorama). As the data was given by the advertiser, the consistency of the category is not assured.
Elevator	Elevator (bad – more than 3 floors without elevator; good – more than 3 floors with elevator, or 1 to 2 floors without elevator, very good – 1 to 2 floors with elevator).
Zone	Three different zones were set up with remarkable differences in Budapest: 1) Luxurious Buda – Districts I, II and XII; 2) Buda and Pest city – Districts III, XI, XXII, V, VI and VII; 3) Pest – Districts XIII, XIV, VIII, IX, X, IV, XV, XVI, XVII, XVIII, XIX, XX, XXI and XXIII.
Subway	Logarithm of the distance from the closest metro or HEV (local train) stations and the square of it ⁴ .
Train	Logarithm of the distance from the closest coach station or train station.
Airport ⁵	Logarithm of the bee line distance from the Liszt Ferenc International Airport ³ .
University	Logarithm of the distance from the closest university campus and the square of it ³ .
Market	Logarithm of the distance from the closest market and the square of it ³ .
Shopping mall	Logarithm of the distance from the closest shopping mall and the square of it ³ .
Park	Logarithm of the distance from the closest park.
Bath, pool	Logarithm of the distance from the closest swimming pool, beach or bath.
Cinema	Logarithm of the distance from the closest cinema.
Theatre	Number of theatres in the 1.5 kilometre radius of the property.
Stadium ⁴	Logarithm of the bee line distance from the closest professional football stadiums.
Cemetery	Logarithm of the distance from the closest cemetery.
Historic monuments	Number of monuments and touristic attraction in the 1.5 kilometre radius of the flat.

⁴ Previous expectation (the higher the number, the better).

⁵ Previous expectation (the higher the number, the better).

Table 3. Descriptive statistics of dependent variable

Attribute	Variables	Min.	Max.	Mean	Std.dev.
Price	*price	5.152	91.08	20.184	13.77
	LNPRICE	1.639	4.512	2.832	0.563

Table 4. Descriptive statistics of non-location-specific variables

Attribute	Variables	Min.	Max	Mean	Std. dev.	PE ⁶
Area	*area	35	200	68.858	26.516	
	LNAREA	3.555	5.298	4.168	0.348	+
Balcony	BALCONYYES	0	1	0.516	0.500	+
	BALCONYNO	0	1	0.484	0.500	–
Floor	*floor number	0	10	2.589	2.155	
	FLOOR01	0	1	0.368	0.482	2
	FLOOR24	0	1	0.479	0.500	4
	FLOOR57	0	1	0.109	0.311	3
	FLOOR810	0	1	0.045	0.207	1
Construction method	INTNOPANELLNAREA	0	1	0.842	0.365	+
	INTPANELLNAREA	0	1	0.158	0.365	–
Condition	INTBADLNAREA	0	1	0.071	0.258	1
	INTAVERAGELNAREA	0	1	0.110	0.313	2
	INTGOODLNAREA	0	1	0.326	0.469	3
	INTVERYGOODLNAREA	0	1	0.437	0.496	4
	INTNEWLNAREA	0	1	0.056	0.230	5
Heating	HEATINGELECTRIC	0	1	0.008	0.088	1
	HEATINGHERA	0	1	0.028	0.164	1
	HEATINGDISTRICT	0	1	0.139	0.346	4
	HEATINGCONVECTOR	0	1	0.192	0.394	1
	HEATINGCHSUM	0	1	0.279	0.449	5
	HEATINGCIRCULATION	0	1	0.344	0.475	5
	HEATINGFANCOIL	0	1	0.011	0.102	7
# of rooms	ROOMNUMBER	1	6	2.286	0.810	+
Parking	PARKINGGARAGE	0	1	0.047	0.212	+
	PARKINGNO	0	1	0.953	0.212	–
Panorama	PANORAMAPANORAMA	0	1	0.205	0.404	4
	PANORAMAGARDEN	0	1	0.287	0.452	3
	PANORAMACOURTYARD	0	1	0.140	0.347	2
	PANORAMASTREET	0	1	0.368	0.482	1
Elevator	ELEVATORVERYGOOD	0	1	0.230	0.421	3
	ELEVATORGOOD	0	1	0.656	0.475	2
	ELEVATORBAD	0	1	0.114	0.318	1

⁶ Previous expectation (the higher the number, the better).

Table 5. Descriptive statistics of location-specific variables

Attribute	Variables	Min.	Max	Mean	Std. dev.	PE ⁷
Zone	ZONEBUDALUX	0	1	0.177	0.381	3
	ZONEBUDA+PESTCITY	0	1	0.337	0.473	2
	ZONEPEST	0	1	0.486	0.500	1
Subway	LNSUBWAY	0	9.366	7.219	0.9115	–
	LNSUBWAYSQ	0	87.731	52.95	12.706	–
Train	LNTRAINSTATION	5.011	9.443	7.86	0.602	+
Airport	LNAIRPORT	7.634	10.249	9.731	0.2573	+
University	LNUNIVERSITY	0	9.48	8.192	0.476	–
	LNUNIVERSITYSQ	0	89.877	60.651	14.217	–
Market	LNMARKET	3.689	9.461	7.513	0.795	–
	LNMARKETSQ	13.608	89.512	57.07	11.778	–
Shopping mall	LNSHOPPINGMALL	0	9.589	7.488	0.678	–
	LNSHOPPINGMALLSQ	0	91.945	56.523	9.78	–
Park	LNPAK	4.942	9.721	7.832	0.709	–
Bath, pool	LNBATH	4.382	9.441	7.645	0.607	–
Cinema	LNCINEMA	0	9.875	7.738	0.774	–
	LNCINEMASQ	0	97.507	60.474	11.649	–
Theatre	THEATER15KM	0	11	0.859	2.030	–
Stadium	LNSTADIUM	4.501	9.391	7.988	0.639	+
Cemetery	LNCEMETRY	6.292	9.525	8.192	0.477	+
Historic monuments	LNHISTORIC15KM	0	5	0.468	0.964	–

4. METHODS

4.1. Hedonic regression model

Firstly, the regression method was implemented for the database. In order to evaluate and compare the two methods, 80% of the data was used for the creation of the regression function, while 20% was left for the evaluation. Most of the given variables indicated a 10% significance level. The outcome (Table 6) corresponded to the previous expectations.

The semi-logarithmic function that was used in our regression method is described by the following equation:

$$\ln(y) = \beta_0 + \beta_1 * \ln(x_1) + \beta_2 * \ln(x_2) + \dots + \beta_m * \ln(x_m) + \beta_{m+1} * x_{m+1} + \dots + \beta_s * x_s$$

where y is the dependent variable and x_1, x_2, \dots, x_s are the independent variables.

If there is heteroscedasticity among the error terms, the F-test and t-statistics cannot be considered efficient. Hence, one might suspect incorrect conclusions.

⁷ Previous expectation (the higher the number, the better).

Table 6. Hedonic regression estimates

	TOTAL		BUDALUX		BUDA + PESTCITY		PEST	
Sample size	1806		319		609		878	
Number of observation	1446		254		493		705	
Out-of-sample	360		65		116		173	
Variable	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
LNAREA	0.807	34.16***	1.055	26.35***	0.738	17.06***	0.773	25.07***
INTNOPANELLNAREA	0.103	16.13***	–	–	0.145	12.30***	0.081	11.64***
INTPANELLNAREA		base		–		base		base
BALCONYYES	0.075	6.12***	0.051	1.68*	0.097	4.11***	0.059	3.55***
BALCONYN0		base		base		base		base
FLOOR01	0.090	2.37**	0.491	2.36**	0.129	2.48**		base
FLOOR24	0.101	2.72***	0.493	2.37**	0.159	3.08***	0.042	2.27**
FLOOR57	0.084	2.31**	0.423	1.95*	0.104	2.1**	0.054	2.06**
FLOOR810		base		base		base		base
INTBADLNAREA	–0.094	–10.27***	–0.116	–6.11***	–0.085	–5.31***	–0.096	–9.07***
INTAVERAGELNAREA	–0.089	–10.94***	–0.111	–6.11***	–0.090	–6.71***	–0.082	–8.35***
INTGOODLNAREA	–0.062	–8.37***	–0.077	–5.18***	–0.063	–5.26***	–0.054	–6.35***
INTVERYGOODLNAREA	–0.027	–3.94***	–0.038	–2.67***	–0.024	–2.10**	–0.018	–2.53**
INTNEVLNAREA		base		base		base		base
ZONEPEST	–0.083	–4.42***	–	–	–	–	–	–
ZONEBUDA+PESTCITY		base		–		–		–
ZONEBUDALUX	0.125	5.68***	–	–	–	–	–	–
HEATINGELECTRIC	–0.432	–5.54***	–0.581	–3.62***	–0.357	–2.93***	–0.357	–2.21**
HEATINGHERA	–0.531	–7.81***	–0.675	–4.55***	–0.526	–5.31***	–0.500	–3.05***
HEATINGDISTRICT	–0.482	–7.74***	–0.535	–2.97***	–0.303	–3.29***	–0.469	–2.96***
HEATINGCONVECTOR	–0.522	–8.53***	–0.663	–4.82***	–0.449	–5.16***	–0.471	–2.98***
HEATINGCHSUM	–0.405	–6.85***	–0.638	–4.81***	–0.285	–3.29***	–0.346	–2.24**
HEATINGCIRCULATION	–0.424	–7.39***	–0.623	–4.68***	–0.362	–4.66***	–0.359	–2.33**
HEATINGFAN–COIL		base		base		base		base
ROOMNUMBER		–		–		–		–
PARKINGGARAGE	0.165	5.57***	0.094	1.84*	0.169	3.10***	0.161	3.93***
PARKINGNO		base		base		base		base
ELEVATORVERYGOOD	0.071	3.00***	–	–		base	0.055	1.83*
ELEVATORGOOD	0.065	3.48***	–	–		base	0.043	1.75*
ELEVATORBAD		base		–		–0.073		–2.24**
PANORAMASTREET	–0.160	–8.05***	–0.206	–5.69***	–0.110	–3.44***	–0.124	–3.56***
PANORAMACOUTYARD	–0.157	–6.36***	–0.218	–3.31***	–0.103	–2.62***	–0.110	–2.75***
PANORAMAGARDEN	–0.097	–4.95***		base		–0.108	–3.13***	–0.070
PANORAMAPANORAMA		base		base		base		base
THEATER15KM	0.011	2.54**	–	–	–	–	–	–
LNHISTORIC15KM		–		–		–		–
LNAIRPORT	0.180	4.83***	–	–	–	–	0.201	5.15***
LNCEMETERY	0.093	6.67***	–	–	–	–	0.175	9.68***
LNCINEMA	0.624	3.08***	–	–	–	–	–0.049	–2.76***

LNMARKET	0.645	4.96***	-0.071	-2.51**	-	0.383	2.29**
LNSHOPPINGMALL	-0.586	-2.88***	-	-	-	-	-
LNSTADIUM	0.047	3.71***	-	-	0.253	7.36***	-
LNSUBWAY	0.074	2.00**	-	-	0.168	3.75***	0.041
LNTRAINSTATION	0.065	4.06***	0.123	3.42***	0.070	3.30***	0.081
LNUNIVERSITY	0.134	3.02***	-	-	0.481	3.20***	0.090
LNBATH	-	-	-	-	-	-	-
LNPARK	-	-	-	-	-	-	-
LNCINEMASQ	-0.043	-3.17***	-	-	-	-	-
LNMARKETSQ	-0.045	-5.09***	-	-	-	-0.029	-2.54**
LNSHOPPINGMALLSQ	0.040	2.95***	-	-	-	-	-
LNSUBWAYSQ	-0.005	-1.85*	-	-	-0.017	-4.83***	-
LNUNIVERSITYSQ	-0.009	-2.93***	-	-	-0.034	-3.24***	-0.008
C	-6.850	-9.53***	-1.224	-3.28***	-4.846	-6.76***	-5.602
R-squared	0.85596		0.8486		0.85551		0.83174
Adjusted R-squared	0.85197		0.8356		0.84712		0.82425
F-statistic	241.239		65.255		101.970		111.053

Note: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Since this problem arises in many economic fields, this phenomenon received special attention. The White-test showed a p-value of 0.000, thus the null hypothesis was rejected (H_0 : error terms are homoscedastics), whereas heteroscedasticity is present. This error can be eliminated by the usage of the estimation using White heteroscedasticity-consistent standard errors and covariance. All of the appraisals were calculated according to this principle.

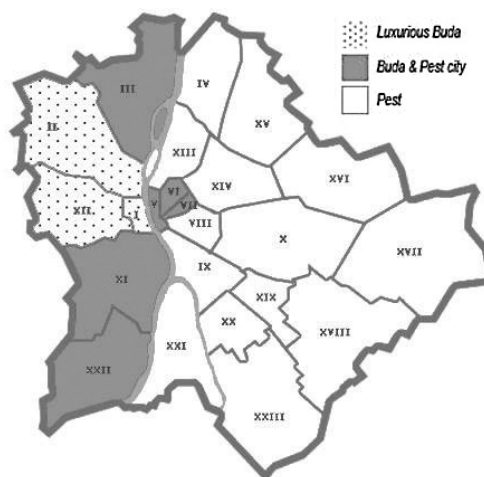


Figure 1. Classification of zones

The area of the residential property is significant with a positive coefficient. Brick or concrete buildings are more valuable than pre-fab panel houses, adding an additional 10.82% to the property value. Moreover, the presence of a balcony is also a positive characteristic, increasing the value by 7.81%.

The worst floors are the 8th to 10th floors. These are mainly pre-fab panel buildings, and owing to their age, the elevator can break down, causing inconveniences for the residents. Similar effects are perceived in the case of 5th to 7th floor properties, while ground and first floor flats are exposed to noise, air pollution, or burglary. Regarding the energy spent on heating and cooling, the top and ground floor flats are the worst. Thus, the outcome is reasonable, the best ones are the 2nd to 4th floors.

As for the condition of the properties, the better the condition, the higher the value. This finding is consistent with previous expectations. A property in good condition is worth as much as 93.9% of a new property with similar characteristics.

The rank of prestige is reflected in the results between the three areas. The best location is the luxurious and green Buda, including the 1st, 2nd, and 12th districts. It is followed by the less prestigious city centre in Pest and the rest of Buda. Although these areas do not all have the same characteristics, their prestige level is about the same. The areas with the less valuable residential properties are the suburban areas of Pest.

One minor issue in terms of the coefficient of the heating system was unexpected. The coefficient of the electric heating is somewhat better than previously expected. There are only 14 flats out of the 1,806 in the sample with electric heating and that could lead to the slightly higher coefficient. Hera⁸ and convector heating are proven to be the worst, while central heating and circulation systems are better. Changing the convector heating to central heating could increase the property value by 12.4%. The fan-coil system indicated the best results. It is assumed that properties with fan-coil heating are luxurious properties and that they are equipped with extras such as sauna, jacuzzi, and security guard that were not observed in our study. That gives an explanation for the exceptionally high added-value of the fan-coil system.

The room number variable turned out to be insignificant, which is reasonable in the light of the high correlation (0.81) between the area and the room number.

The presence of a garage improves the value, and so does an elevator. Regarding the panorama, a garden or a nice panorama give additional value to the property, while a courtyard or street view is less preferred.

Regarding spatial parameters, the square of the component was also added to the regression, in cases when the price and the distance is assumed to be not

⁸ Gas heating located in the place of the old fireplace.

simply logarithmic. For instance, the closure of a subway station is not good because of the high traffic and pollution. Long distances to be covered compose another deterring factor. Similar effects are detected in the case of cinemas, shopping centres, markets, and universities. The findings indicate a common desire to distance oneself from airports, stadiums and cemeteries, whereas the number of theatres in the neighbourhood increases the property value.⁹ Three of the variables turned out to be insignificant. The bath, the park, and the closure of historical monuments do not affect the value. The main reason behind this supposedly is that only the large parks were included in the study and it is the quality of the park that might be important rather than its size or presence; similarly, only the 18 best tourist spots were considered. The closure of a bath might not affect the property value.

Following the analyses of the whole dataset, a geographical segmentation was made to reach better estimation functions. The three zones have different characteristics and diverse levers affect the price.

The 1st, 2nd and 12th Buda districts completely lack pre-fab panel buildings, and these territories have the highest area coefficients. The importance of a balcony is less relevant since most of the houses provide a garden. The area is rich in nature, thus the difference between the panorama and the garden turned out to be minor, so the two variables were aggregated. Of the location-specific variables, only the market and the train or bus station turned out to be significant. The proportion of the elderly population is higher in these areas, thus the closure of market turned out to be a positive lever, while the noisy train and bus stations are not welcomed by the wealthy inhabitants of these zones.

The second group consists of the rest of Buda and the Pest city centre. The technical- and location-specific results are in accordance with the previous expectations. Since several train stations and stadiums are located near this area, these variables turned out to be significant and disturbing factors. Although the relationship between the subways and universities is not linear, we can nonetheless conclude that excessively long distances from these facilities are not preferred.

The third category encapsulating the rest of Pest has the lowest area added value. Those are the cheapest flats. Most of the location-specific variables turned out to be significant in this area, such as the distance from the national airport that is located in the side of this area.

⁹ The theatre variable is an indicator that the flat is located in the city centre. As the theatres are located in the central spots, probably it is not the presence of the theatre itself that is interesting, but the fact that the property is in a central location.

4.2. Artificial neural networks

After the hedonic regression, the MLP neural network method was applied to the sample. Similarly to the previously discussed model, the dependent variable was the logarithm of the price (LNPRICE), while the explanatory variables were all the previously introduced variables (*Table 2*). The sample was divided into three segments: 60% was the teaching segment, 20% was the test segment, and 20% was the validation segment. The validation segment was composed of the out-of-sample part of the hedonic regression, thus the two methods became comparable after finding the best network structure.

During the sensibility analysis, different numbers of hidden layers and neurons were applied, and the ones with the best fitting results were chosen in order to reach the best network topology. The decision of the best fitting result was based on the mean absolute percentage error (MAPE) and 5-fold cross-validation.

For the whole sample, the results of fitting the MLP neural network is presented in *Table 7*. As our focus is on the best topology, the results are given for the validation segment. The figures in the headings represent the number of neurons in the hidden layers. Apparently, the best result is achieved when the MLP has 14 neurons.

The next step of modelling was to segment the sample similarly to the hedonic regression, using the three different zones, and to seek for the best network structure. The one layer with 14 neurons is the best everywhere, thus this network choice is considered to be robust (*Table 8*).

Table 7. Comparison of the appraisal accuracy for different MLP neural networks regarding the whole sample, %

# of neurons	7	8	9	10	11	12	13	14	15
MAPE	13.08	13.05	12.60	12.62	12.54	12.15	12.31	11.99	12.05

Table 8. Comparison of the appraisal accuracy for different MLP neural networks regarding the zone categories, %

Luxurious Buda									
# of neurons	7	8	9	10	11	12	13	14	15
MAPE	11.12	10.73	10.89	10.80	10.64	10.25	9.92	9.52	9.86
Buda and Pest city									
# of neurons	7	8	9	10	11	12	13	14	15
MAPE	11.97	11.63	11.33	11.43	10.78	10.58	10.45	9.98	10.62
Pest									
# of neurons	7	8	9	10	11	12	13	14	15
MAPE	11.79	11.69	11.47	11.08	11.14	11.30	10.59	10.08	10.74

5. RESULTS

After creating the hedonic regression model and identifying the best fitting neural networks, the outputs were compared in order to determine whether hedonic regression or the neural network is the better method for our sample. The basis of the comparison was the out-of-sample part for both the entire sample and the established three categories (number of observations: 360, 65, 120, 173). The evaluation of the two methods was based on three accuracy ratios: mean absolute percentage error (MAPE) and the percentage of the properties with less than 10% and more than 25% estimation error. All of the observed variables were used in all cases.

Table 9. Comparison of the appraisal accuracy for hedonic regression and MLP neural network (14 neurons in the hidden layers) for the total sample, %

	Hedonic	MLP
MAPE	17.90	10.94
<10%	36.67	56.67
>25%	25.56	7.78

For the total sample, the better outcome is achieved through neural networks based on *Table 9*. The discrepancy between the results is significant considering the 6.96% better MAPE ratio, and there are 20% additional observations giving less than 10% errors in the neural model as opposed to the hedonic model. The graphical representation also shows the superiority of the neural networks. The appraisal accuracy is visibly weaker in both cases if the value of the property is above HUF 35 million (~€110 000). Certain special characteristics of expensive flats that are not considered in this paper are responsible for this phenomenon. The observed variables, such as the distances from specific points in the city, do not considerably affect the value of expensive properties, but there are other features like sauna, jacuzzi, swimming pool, or security guard that were not observed and can cause huge differences in the luxurious segment. These elements are not present in the non-luxurious segment; the basic needs of living and location-specific variables are the ones that influence the property value. It is interesting to observe the sharp plummet of the appraisal accuracy ratios between these two segments (below and above HUF 35 million) (*Table 10*). These figures support the hypothesis concerning the omitted luxurious variables.

The appraisal accuracy was also observed in the three previously defined zones (*Table 11*). The table shows that, as opposed to suggestions of the previous cited literature, the size of the sample does not correlate with the preciseness of the neural network. Thus the hypothesis of the related literature that MLP shows

Table 10. Comparison of the appraisal accuracy for hedonic regression and MLP neural network for the segments below and above HUF 35 million, %

	Hedonic		MLP	
	below HUF 35 M	above HUF 35 M	below HUF 35 M	above HUF 35 M
MAPE	17.56	20.03	10.15	15.93
<10%	38.59	24.49	60.45	32.65
>25%	24.12	34.69	6.11	18.37

Table 11. Comparison of the appraisal accuracy for hedonic regressions and MLP neural networks for the zone segments, %

	Luxurious Buda		Buda and Pest city		Pest	
	Hedonic	MLP	Hedonic	MLP	Hedonic	MLP
MAPE	17.64	9.97	17.55	10.75	17.42	9.57
<10%	38.46	58.46	37.50	54.17	39.31	61.27
>25%	23.08	12.31	24.17	8.33	21.97	4.62

Table 12. Summary of the better methods in the observed segments

	Better Method	Sample size	Δ MAPE
Budalux	MLP	319	7.67%
Buda and Pest city	MLP	609	6.8%
Pest	MLP	878	7.85%
Total	MLP	1806	6.96%

better results for large sample does not stand in our case. The results indicate a similarly tendency as in the case of the total sample: in all of the three zones, the neural network gives more precise results than the hedonic regression, and it has almost the same amount in all the three distributions (6.8%–7.85%, *Table 12.*).

6. CONCLUSIONS

The study examines the determinants of residential property prices in Budapest, as one of the first analyses on housing valuation in Hungary with a relevant sample size. Two types of modelling approaches were employed: hedonic regression and the MLP neural network model. Most of the observed variables are shown to affect the residential property prices significantly. The outcome shows that the MLP neural network gives a better result than hedonic regression.

The findings of the study provide opportunities both for practical applications and further research. Regarding the practical applications, the results could be used in three main fields: automated valuation for mortgage-backed loans, taxation controlling purposes, and real estate development decisions. In the case of a huge mortgage portfolio, a quick check can be done using the results of this paper, controlling the traditional one-by-one valuations. The reported transaction prices are often lower than the real ones for tax evasion purposes. The results of this paper are applicable for the control of reported values and in cases when property tax is imposed. Regarding the price components and their coefficients, the study can back real estate development decisions, defining the valuable features of a residential property. Furthermore, through the change in housing prices, the valuation of public buildings or facilities may also be supported by this paper.

A better understanding and qualitative results of the real estate market processes could be beneficial for all the shareholders in the real estate market. Through the observation of the time changes in the housing values using panel data or the results of other data mining methods such as fuzzy logic, SVM (Support Vector Machine), or decision-trees, the introduction of new studies becomes possible. Further analyses could point out the reasons behind the outlier results in the samples, and this could make the obtained results more precise.

REFERENCES

- Borst, R. A. (1995): Artificial Neural Networks in Mass Appraisal. *Journal of Property and Tax Assessment Community*, 10(1): 69–94.
- Brosen, B. – Grant, W. R. – Rister, M. (1984): A Hedonic Price Model for Rough Rice Bid/Acceptance Markets. *American Journal of Agricultural Economy*, 66(2): 156–163.
- Central Statistical Office (CSO) (2005): Housing Conditions at the Turn of the Century. Budapest.
- Dai, W. – Wu, J-Y. – Lu, C-J. (2012): Combining Nonlinear Independent Component Analysis and Neural Network for the Prediction of Asian Stock Market Indexes. *Expert Systems with Applications*, 39(4): 4444–4452.
- Do, Q. – Grudnitski, G. A. (1992): Neural Network Approach to Residential Property Appraisal. *The Real Estate Appraiser*, 58: 38–45.
- Falzon, J. – Lanzon, D. (2013): Comparing Alternative House Price Indices: Evidence from Asking Prices in Malta. *International Journal of Housing Markets and Analysis*, 6(1): 98–135.
- Harrison, D. – Rubinfeld, D. L. (1978): Hedonic Housing Prices and the Demand for Clean Air. *Journal of Environmental Economics and Management*, 5(1): 81–102.
- Horváth, Á. – Székely, G. (2009): Usage of Hedonic Method for the Observation of Value Changes of Housing Real eEstate (in Hungarian: Hedonikus módszer alkalmazása a használt lakások áralakulásának megfigyelésében). *Statisztikai Szemle*, 87(6): 594–607.
- Kain, J. F. – Quigley, J. M. (1970): Measuring the Value of Housing Quality. *Journal of the American Statistical Association*, 65(330): 532–548.
- Kauko, T. (2007): A Heterodox Economic Analysis of the Housing Market Structure in Budapest Using Neural Network Classification. *Journal of Real Estate Literature*, 15(1): 85–124.

- Kauko, T. – d'Amato, M. (eds) (2009): *Mass Appraisal Methods: An International Perspective for Property Valuers*. Wiley-Blackwell, Oxford, United Kingdom.
- Nguyen, N. – Cripps, A. (2001): Predicting Housing Value: A Comparison of Multiple Regression Analysis and Artificial Neural Networks. *Journal of Real Estate Research*, 22(3): 313–336.
- Pagourtzi, E. – Assimakopoulos, V. – Hatzichristos, T. – French, N. (2003): Real Estate Appraisal: A Review of Valuation Methods. *Journal of Property Investment & Finance*, 21(4): 383–401.
- Pavlin, B. (2006): *Dwelling Price Index in Slovenia – Pilot Study of Hedonic Approaches*. Real Estate Price Indexes Conference, OECD–IMF Workshop, Paris, 6–7 November 2006.
- Peterson, S. – Flanagan, A. B. (2009): Neural Network Hedonic Pricing Models in Mass Real Estate Appraisal. *Journal of Real Estate Research*, 31(2): 147–164.
- Rosen, S. (1974): Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1): 34–55.
- Rossini, P. (1997): *Application of Artificial Neural Networks to the Valuation of Residential Property*. The Third Annual Pacific Rim Real Estate Society Conference, Palmerston North, New Zealand.
- Selim, H. (2009): Determinants of House Prices in Turkey: Hedonic Regression versus Artificial Neural Networks. *Expert Systems with Applications*, 36(2): 2843–2852.
- Tay, D. P. H. – Ho, D. K. H. (1992): Artificial Intelligence and the Mass Appraisal of Residential Apartments. *Journal of Property Valuation & Investment*, 10(2): 525–540.
- Uyeno, D. – Hamilton, S. W. – Biggs, A. J. G. (1993): Density of Residential Land Use and the Impact of Airport Noise. *Journal of Transport Economics and Policy*, 27(1): 3–18.
- Wang, J. Z. – Wang, J. J. – Zhang, Z. G. – Guo, S. P. (2011): Forecasting Stock Indices with Back Propagation Neural Network. *Expert Systems with Application*, 38(11): 14346–14355.
- Widlak, M. – Tomczyk, E. (2010): Measuring Price Dynamics: Evidence from the Warsaw Housing Market. *Journal of European Real Estate Research*, 3(3): 203–227.
- Wilson, W. (1984): Hedonic Prices in the Malting Barley Market. *Western Journal of Agricultural Economics*, 9(1): 29–40.
- Worzala, E. M. – Lenk, M. – Silva, A. (1995): An Exploration of Neural Networks and Its Application to Real Estate Valuation. *Journal of Real Estate Research*, 10(2): 185–202.