Prediction of Real Estate Market Development Using Statistical Methods

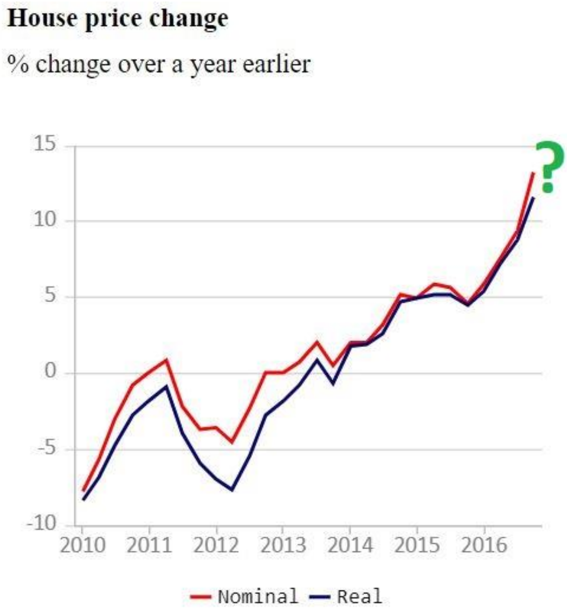
*Darina Tauberová**

1. Introduction

It is necessary to be well versed in the development of real estate prices when evaluating real property. This concerns not only cases of purchase for the purpose of real estate investment, construction, or market analysis, but also any usual pricing should include an analysis of the price development of the evaluated real property. These requirements call for the development of real estate valuation, as the prediction of the development of the real estate market cannot be determined solely on the basis of an analysis of the currently sold (offered) immovable property. It is necessary to approach other methodologies where it is appropriate to use mathematical and statistical methods. The aim of this article is to present a possible methodological approach and a detailed description of this approach, incl. a clear explanation of the model used and the tests carried out. The article assumes the knowledge of general real estate valuation issue and the basics of mathematical statistics.

The article focuses on the market in the Czech Republic, but the results are globally usable. Research is focused on residential real estate because of the fact that different rules apply to corporate real estate such as larger commercial buildings, industrial buildings or halls and agricultural land than for residential immovable property. The author of the research is trying to develop the field of real estate valuation, because this branch is one of the very important branches necessary for the state economy as a whole, its stability and influence on the global economy. Quality opinions can minimize market risks, such as the so-called “real estate bubble”.

Fig. 1 Development of real estate prices, Czech Republic



Source: www.globalpropertyguide.com

The author analysed possible approaches that appear to be suitable for creating a system approach for predicting real estate prices and the creation of a regression model for the prediction of the development of the real estate market was chosen as most appropriate,

* Mgr. Darina Tauberová, an expert and certified valuer, postgraduate student at ÚSI VÚT in Brno. Email: darina.tauberova@centrum.cz. The article is made as one of the results of the dissertation of the author.

namely the creation of a delayed multiple regression model, provided the regression model assumptions are fulfilled. The article is divided to theoretical and practical parts. The theoretical part explains regression models, assumptions of regression models and hypothesis testing. The practical part demonstrates the prediction of the real estate market development.

2. Regression Model

Based on research and testing, a multiple regression model was created under the conditions of fulfilling assumptions of regression models and hypothesis testing. Regression models determine the relationship of each quantity. Regression models comprise of simple linear, nonlinear, and multiple linear and nonlinear models. For simple models, there is only one explanatory variable x, for multiple regression models, there is a set of explanatory variables $x_1, \ldots, x_n$. Other types of models – logistic, multinomial, Poisson, etc. will not be described in this paper because they will not be used [1].The basis of regression modelling is the least squares method. Regression analysis, as a statistical method, makes it possible to estimate the value of a dependent variable on the basis of knowledge of independent variables. In the case of linear regression, it is a finding of a dependency formula, where x is an independent variable and y a dependent variable. The relationship between two variables is summarized by a straight line from which its parameters and its equation can be calculated. The data are plotted on a dot chart, the so-called correlation dot plot, and it is necessary to verify whether dependence exists between the quantities, the so-called regression. The purpose of the regression analysis is to determine the coefficients "*a*" and "*b*" since the values x (x coordinates are exact) and y (y coordinates can be subject to error) are known. Coefficient a determines the slope of the line, coefficient b determines the shift of the line and determines the intersection of the line with axis y, it is a regression coefficient which determines the properties of the line with its value. The correlation coefficient determines the relationship between two variables, whether or not it exists, and whether any prediction can be expressed. Moreover, the regression analysis is able to describe the magnitude of the influence of the independent variable x on the dependent variable y. It is true that the dependent variable y is measured at an interval level, and the independent variable x is interval or dichotomous (i.e. a variable having only two values).The distribution of both variables should be normal, this is not a condition for a larger set, because according to the probability theory and the central limit theorem, abnormal distribution does not have much effect on the results. Linear regression represents the approximation of values by a straight-line least squares method [4].If a set of independent variables and one dependent variable are considered in the calculation, then this is a multiple linear regression. The formula is as follows:

$$y = a + b1 * x1 + b2 * x2 + \ldots + bn * xn, \tag{1}$$

where *a, b* – determine the properties of the line with their value,
y – is the value of the dependent variable,
x1…2 – is the value of the independent variable,

In general, calculations can be made in professional statistical programmes or in programmes such as OpenOffice or Microsoft Excel, but it is preferable to use professional statistical software (GRETL, Statgraphics, etc.) for more accurate and simple calculations.

Other types of models, non-linear models, are commonly used (quadratic, cubic, logarithmic); however, if the linear regression model is only slightly inferior to the non-linear regression model, the linear regression model is preferred because of a clearer interpretation of regression coefficients, which is not possible in case of non-linear models [2].

2.1. Variables and Their Delay

Due to the fact that the economy responds to the stimulus with a certain delay, a possible delay of the variables must be considered when creating the regression model. If this delay is not detected, the exclusion of the delay may result in erroneous results and failure to identify significant correlations. Determination of the optimal delay is possible using statistical software which offer the possibility to automatically analyse the delay of the variables, alternatively MS Excel can be used, where the delay is determined by moving one period of time gradually, and then always calculating the correlation coefficient (path in MS Excel: Data Analysis – Correlation).Both in statistical software and in MS Excel, it is necessary to know the circumstances, the development and relationships of the variables and always analyse whether the delays are real or not in practice. At the same time, it is necessary in the final formation of the delayed multiple regression model to adjust some delays to meet some of the necessary assumptions of the regression model, in particular the assumption of autocorrelation of residues (explained below) [2].

2.2. Assumptions of Regression Models

Regression models must meet several assumptions, such as:

a) **Coefficient of determination, Pearson correlation coefficient.** Variables must be in a linear relationship. Multiple linear regression is based on the Pearson correlation coefficient, so the lack of linearity can cause important relationships to remain unclear. The coefficient of determination is monitored when creating regression models. The mutual relationship between the quantities is expressed by the correlation. If there is a correlation in the relation of quantities, then the quantities are likely dependant on each other. However, correlation does not address causality – the relationship between cause and effect. In the case of regression models, the coefficients R and R^2 *are used,* R is the correlation coefficient and R^2 is the coefficient of determination. The coefficient of determination R^2 tells us how much variance of the explained variable is explained, but it does not take into account the number of explanatory variables in the model. Therefore, in the case of multiple regression models, the adjusted coefficient of determination R^2_{ADJ}, which also takes into account the addition of insignificant variables, is used. The adjusted coefficient of determination is actually a purification of the coefficient of determination R^2 and better shows the suitability of adding or removing statistically significant or non-significant explanatory variables to the model.

b) **Normal data distribution.** Variables must be normally distributed, otherwise there is a possibility of inaccurate results. It is therefore necessary to examine the distribution of each variable entering the analysis within the validity of the central limit theorem.

c) The independent variable must be **interval or dichotomous.**

d) **Multicollinearity.** Independent variables must not be highly correlated with each other, as this would violate the requirement for multicollinearity. Multicollinearity, mutual statistical dependence, must be eliminated in the case of multidimensional regression models. Multicollinearity itself is perceived as an unwanted correlation (dependence) between explanatory variables. If the dependence between the absolute value of explanatory variables is 0.8, it is high multicollinearity [2].Multicollinearity can be found in several ways, for example by determining a pair correlation coefficient between the explanatory variables, when it is true that if $|r| \geq 0.8$ there is multicollinearity, or it is possible to gradually change the individual explanatory variables to the explained variables and, thanks to the other explanatory variables, to model the linear regression and to determine the coefficient of determination (Meloun, Militký, 2002).The backward regression (ridge regression) is used for removing multicollinearity, where such explanatory variables that have the largest *p-values* are discarded.

e) There must not be **remote values** in the data. Differences between the real value y_i and the theoretical value y_{iTeor} are referred to as regression model residues and represent essentially the sides of squares we try to minimize. Residues should be distributed randomly around the horizontal axis and should not be outside the 95% confidence interval. Residues should roughly follow the line, they should have a normal division with a zero mean value [2]. $\chi 2$ test (Chi square, test of goodness of fit) tests the difference of file frequency and the residual normality. In the test, the critical values (quantiles) of the respective distributions according to the tests are specified – t-distribution, f-distribution, χ2-distribution [1].

f) **Homoskedasticity**. Relationships between variables should exhibit homoskedasticity, i.e. homogeneity of variance, which means that variance in the data of one variable will roughly correspond to variance of the other variables. Heteroskedasticity, on the contrary, results in the model being inappropriate and impartial [3].

g) **Autocorrelation of residues**. Absence of autocorrelation of residues in the model is another assumption of the linear regression model. Autocorrelation is defined as a violation of the assumption of mutual independence of random components from different observations. Autocorrelation can be caused by an inappropriate way of collecting data, excluding variables delays, environmental influences, selecting an incorrect model. The result is the deflection of model variance estimate, and the statistical tests are losing their strength. Autocorrelation can be removed, for example, by using more appropriate delays and using more appropriate trend. Autocorrelation in relation to delays between residues is distinguished as autocorrelation of the first, second and up to m-th order [2].Values are in the range of <0.4; if the value *D* is less than 2, it is a positive autocorrelation, values greater than 2 denote a negative autocorrelation [3].

Within the scope of the examination, the **statistical significance *p, the p-value,*** is monitored. The hypothesis testing uses the *p-value*. This is the lowest level of significance of the test where the zero hypothesis is rejected. This zero hypothesis *H0* may be the following: random errors are homoskedastic. In case of low values, there is a high probability that a zero hypothesis applies. The standard limit is 5%.The result of the test is compared to the chosen level of significance α. So if the test result is less than 5%, the zero hypothesis may be rejected and it is then a statistically significant result [3].

When selecting variables, it is advisable to prioritize quality over quantity, there is no direct relationship that the more variables included in the calculation, the more accurate the result will begin terms of efficiency, it is advisable to achieve the maximum possible effect with minimal inputs, i.e. to minimize the number of variables as much as possible. For regression models, it is therefore necessary to assure the following assumptions: proper function form of the given dependence, homoskedasticity of residues, absence of autocorrelation of residues, absence of unrelated values and residue normality [1].

3. Creating the Model and Calculating Prediction Values

The resultant choice of problem solution method is to create a delayed multiple regression model provided that the assumptions of the regression models listed above are met. Quantities (factors) are standardized and parameterized to base indices. On the basis of correlation and delayed correlations, it is possible to identify significant explanatory variables from these quantities. The House Price Index (HPI) was selected as the explained variable. This synthetic price index, published by the Czech Statistical Office, measures the development of the price level of residential real estates and is calculated on the basis of a harmonized European Union standard. Its advantage is international comparability. The data are published quarterly in base indices relative to 2015 (2015=100).This is an index that clearly shows the growth/fall in the price level of residential real estates.

Fig. 2 HPI, Czech Republic

Source: https://tradingeconomics.com

Relevant variables must be selected in the regression model. These are the variables that enter into the prediction of real estate market development. These are macroeconomic, microeconomic, socio-economic or political factors. Factor data have been thoroughly statistically investigated, and it has been found that some factors cannot be included in the statistical investigation as no relevant data are available, or there are data only for a short period of time. The basic data sources are the official webpages of the Czech Statistical Office and the Czech National Bank. It is important to point out that it is not possible to name and describe all factors influencing the analysed issue as human activity is influenced not only objectively but also subjectively. Furthermore, creation of any models in this paper is based on the so-called no-event scenario, which assumes that there will be no extraordinary event, such as the escalation of the Eurozone's debt and banking crisis, migration, state of war, or other geopolitical events with significant impact on the Czech economy. At the same time, a major breakthrough in solving problems in the Czech Republic, or the Eurozone, is not expected. Relevant explanatory variables are the factors listed in Tab. 1.These data are further investigated. Twenty-four explanatory variables were selected for further statistical research from the point of view of relevance for the modelling of the market prediction model. Based on scientific research, expert searches, and literature, these variables have been identified as the most relevant. A problem with some factors that could influence prediction on the theoretical level is data that cannot be parameterized and quantified, or inadequate data sources. The individual explanatory variables will not be described in this article. Explanatory variables are:

Tab. 1: Selection of explanatory variables

Marking	Description of the explanatory variable
y	*House price index*
x_1	*Gross domestic product*
x_2	*Inflation*
x_3	*Housing construction in the Czech Republic*
x_4	*Number of building permits*
x_5	*New floor areas according to building permits*
$x6$	*Index of housing availability*
x_7	*General rate of unemployment*
x_8	*Construction Production Index, Building construction*
x_9	*Average interest rate*
x_{10}	*Number of mortgages*
x_{11}	*Availability of mortgages*
x_{12}	*General rate of unemployment*
x_{13}	*Economic activity rate 15-64 years*
x_{14}	*Median gross monthly wages*
x_{15}	*Indices of offer prices of flats*
x_{16}	*Indices of realized prices of new flats*
x_{17}	*Indices of realized prices of older flats*
$x18$	*CZK/Euro*
x_{19}	*CZK/USD*
x_{20}	*Population*
x_{21}	*Domestic realized demand*
x_{22}	*Index of housing availability*

Source: author's own work

In order to further investigate the data of all factors, it is necessary to adjust them to comparable values. Base indices are appropriate in this case. Indices are a key tool for any comparison in time and place. Base index explains the difference between the values now and in a selected period (for example, the Statistical Office uses 2015 as a selected period, and other years are compared to that period).Base indices of all periods are therefore always related to one given date, to the same selected value, and best depict the development in the time series. The base index relative to 2015 is calculated by averaging 2015, which is 2015=100.It is also possible that even in 2015 itself in individual months or quarters the value is different from 100.This is precisely because the base is the average 2015=100, based on the average of the whole year 2015.On the grounds that the Czech Statistical Office most frequently uses the base index 2015=100, the data in this paper are also related to the average of 2015.

3.1. Creating Correlation Matrix without Delay

In the first step, the variables were examined without delay, the correlation coefficient was determined by the least squares method. The correlation matrix contains pair correlation coefficients of the individual explanatory variables that provide information on the occurrence of the so-called multicollinearity (i.e. dependence) between two or more variables. Multicollinearity occurs in the model when the pair coefficient value exceeds the absolute value of 0.8.For the above 24 explanatory variables $x_1, \ldots, x_{24}$, a correlation matrix of Pearson correlation coefficients is created. The correlation matrix can also be created in MS Excel.

Procedure in MS Excel: Data – Data Analysis – Correlation – Highlight selected data (all input data, $y, x_1, \ldots, x_{24}$), check boxes "*Labels in first row*" and "*New Worksheet Ply*" – a correlation matrix is created on a separate sheet. The correlation matrix for all variables will appear in this new worksheet.

The above correlation matrix is applicable only to the first pre-selection of the relevant variables because it does not include knowledge of a particular market situation and does not address the shift of variables in time. If the model was compiled only on the basis of undelayed variables, the value of the adjusted coefficient of determination R^2_{ADJ} would be lower than in the case of use of delayed explanatory variables. The result of the model can be further improved by delayed explanatory variables, as will be demonstrated in the following chapters. By pre-selecting and using statistical software, 8 explanatory variables were shortlisted. However, evaluation using MS Excel is time consuming, so it is more appropriate to use professional statistical software. GRETL will be used to perform all the calculations in one step. An unavoidable reason for using delayed explanatory variables is also the fact that prediction based on delayed variables uses real data rather than predicted ones. The values of the coefficient of determination R^2 *on the basis of the calculation in GRETL are* 0.974011 and the values of the adjusted coefficient of determination R^2_{ADJ} are 0.966842.These results will be evaluated below. It is highly probable that the coefficient of determination R^2 and the adjusted coefficient of determination R^2_{ADJ} will increase if the optimal delay of the explanatory variables is detected and the statistically insignificant explanatory variables are selected out.

In the next step, the optimal delay of the explanatory variables was determined. The statistical software GRETL was used to determine the optimal delay according to the highest delayed correlation coefficient. If the optimal delay was detected using MS Excel, the procedure would be such that data rows would be shifted and correlation matrices created at the given delay. This is a very laborious and time-consuming process. GRETL suggests the optimal delay of the individual explanatory variables and at the same time selects out statistically insignificant explanatory variables. Variables that do not correlate are removed from the model by sequential ridge regression, as demonstrated in the following chapter. It is determined whether the model improves by delaying and removing selected explanatory variables (according to the adjusted coefficient R^2_{ADJ}).

After the optimal delay is detected, reverse ridge regression is performed to detect the relevant variables and to remove those explanatory variables that had the highest *p-value*. After the tests had been performed, it was found that there were 6 relevant variables and the coefficient was established, i.e. the values that make up the regression model, see Table 2.Thus, the reverse ridge regression determined that it is convenient to use the variables x_1, x_9, x_{15}, x_{16}, x_{21} *and* x_{22} as statistically significant explanatory variables. The value of the adjusted coefficient of determination R^2_{ADJ} was the highest (0.973977, i.e. 97.40%).The values of the Durbin-Watson statistic in this model are 1.742615, a value close to 2, i.e. a random component of the model is simple autocorrelation. In addition, the p-value of the autocorrelation test has a value of 0.534743, i.e. the value is greater than 0.05, and there is a presumption of the absence of autocorrelation of residues in the model. The White test for heteroskedasticity, i.e. the *p-value* of the test for heteroskedasticity, is 0.211673, and therefore greater than 0.05, and the zero hypothesis “H0: There is no heteroskedasticity” cannot be rejected at a significance level of 0.05.The random component of the model is homoskedastic, therefore the assumption of the absence of heteroskedasticity in the model is not violated. The residual normality test, the Chi square test, shows the normal error distribution, *p-value* at the significance level of 0.05 is 0.4475.

3.2. Delayed Multiple Regression Model

The dependent variable, i.e. the explained variable, is in this case the House Price Index (HPI), the values of which are to be predicted, as this index best illustrates the rise in the price level of residential real estates. The independent variables, i.e. the explanatory variables, were chosen to be x_1 *Gross Domestic Product (HDP),* x_9 *Average Interest Rate (US),* x_{15} *Indices of Offer Prices of Flats, whole Czech Republic (NC),* x_{16} *Indices of Realized Prices of New Flats, whole Czech Republic (RC),* x_{21} *Domestic Realized Demand* (RP) and x_{22} *Index of Housing Availability (DB)*.

Tab. 2: Selected explanatory variables, delays and coefficient

Marking	Description of the explanatory variable		Optimal delay (quarterly)	Coeficient
y	*House price index*	HPI	x	**-2,1721**
x_1	*Gross domestic product*	GDP	1	**0,183845**
x_9	*Average interest rate*	IR	5	**0,0142312**
x_{15}	*Indices of offer prices of flats with detected delay*	OF	1	**0,395358**
x_{16}	*Indexes of actual new apartments*	NF	1	**0,611055**
x_{21}	*Domestic realized demand*	RD	1	**-0,018163**
x_{22}	*Index of housing availability*	HA	1	**-0,152422**

Source: Calculations in SW GRETL, table is author's own work

The final delayed multiple linear regression model is as follows:

$$HPI = -2{,}17210 + 0{,}183845 \cdot GDP1 + 0{,}0142312 \cdot IR5 + 0{,}395358 \cdot OF1 + 0{,}611055 \cdot NF1 - 0{,}0181625 \cdot RD1 - 0{,}152422 \cdot HA1$$

where

HPI	*House price index*
GDP1	*Gross domestic product with detected delay*
IR5	*Average interest rate with detected delay*
OF1	*Indices of offer prices of flats with detected delay*
NF1	*Indices of realized prices of new flats with detected delay*
RP1	*Domestic realized demand with detected delay*
DB1	*Index of housing availability with detected delay*

Once specific data were entered into the model, it was possible to determine and thus predict the development of residential real estate prices. Of the originally selected 24 explanatory variables, the 6 most statistically significant explanatory variables were statistically selected, while observing the rules of the regression model. Optimal delay of these variables was also found.

3.3. Prediction of the Real Estate Market Development

Based on the created model, prediction can be approached. When looking for an HPI value for remote periods (2019, 2020), the predicted data already enters the calculation and therefore the result is very sensitive to the input data. If the data of explanatory variables cannot be traced, it is possible to find out these future data via a simple linear regression model. The calculation procedure is as follows: specific numbers were substituted into the model *"HPI = -2.17210 + 0.183845*GDP1 + 0.0142312*IR5 + 0.395358*OF1 + 0.611055*NF1 - 0.0181625*RD1 - 0.152422*HA1*", marked by colours in the table Tab. 3.

Tab. 3: **Calculation of prediction 4q/2017 to 3q/2018**

	y	x_1	x_9	x_{15}	x_{16}	x_{21}	x_{22}	
Quertet/year	House Price Index (HPI)	GDP	Average interest rate	Indices of offer prices of flats	Indices of realized prices of new flats	Domestic realized demand	Index dostup-nosti bydlení	Resulting Values
2q/2016	**105,10**	105,10	88,99	109,01	104,25	31,28	102,54	
3q/2016	**107,70**	105,70	86,65	111,16	104,65	31,28	101,64	
4q/2016	**112,80**	105,95	84,31	114,06	110,42	31,28	106,86	
1q/2017	**116,10**	108,28	88,06	116,21	115,31	41,55	112,85	
2q/2017	**117,20**	109,20	95,08	119,39	119,12	41,55	121,32	
3q/2017	**121,00**	112,01	94,75	125,10	127,14	41,55	122,37	
4q/2017	?	113,20	96,02	126,00	128,00	41,55	123,00	**128,31**
1q/2018	?	116,91	98,36	127,00	129,00	49,50	125,00	**128,37**
2q/2018	?	117,00	99,30	128,00	131,00	51,00	127,00	**129,66**
3q/2018	?							**131,06**
Coeficient	**-2,1721**	**0,183845**	**0,0142312**	**0,395358**	**0,611055**	**-0,0181625**	**-0,152422**	

Source: Calculations in SW GRETL, table is author's own work

When the values have been substituted, the calculations are as follows:

*HPI 4q/2017 = -2.17210 + 0.183845*112.01 + 0.0142312*86.65 + 0.395358*125.10 + + 0.611055*127.14 - 0.0181625*41.55 - 0.152422*122.37 = 128.31*

*HPI 1q/2018 = -2.17210 + 0.183845*113.20 + 0.0142312*84.31 + 0.395358*126.00 + + 0.611055*128.00 - 0.0181625*41.55 - 0.152422*123.00 = 128.37*

*HPI 2q/2018 = -2.17210 + 0.183845*116.91 + 0.0142312*88.06 + 0.395358*127.00 + + 0.611055*129.00 - 0.0181625*49.50 - 0.152422* 125.00 = 129.66*

*HPI 3q/2018 = -2.17210 + 0.183845*117.00 + 0.0142312*95.08 + 0.395358*128.00 + + 0.611055*131.00 - 0.0181625*51.00 - 0.152422*127.00 = 131.06*

By the above procedure, HPI values were also found for other quarters, and the same procedure can also be applied to prediction for later periods. The results of the prediction are presented in the table Tab. 3, including the above increments compared to the previous quarter of the same year and against the same quarter of the previous year.

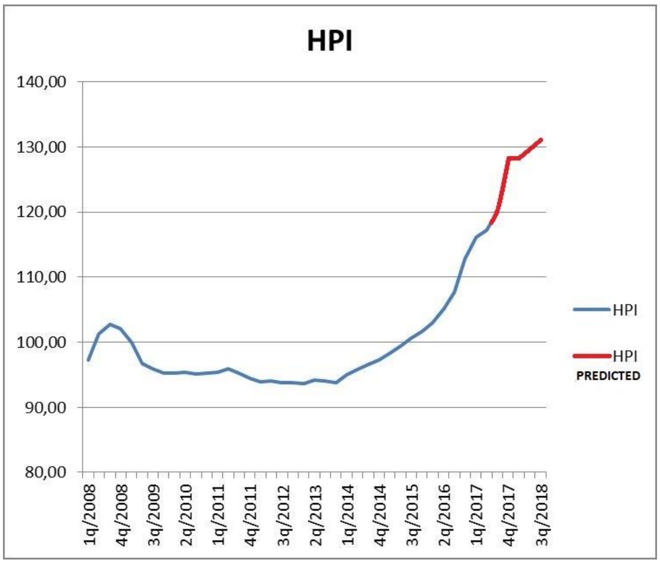
Tab. 4: **Prediction of HPI development**

Quertet/year	HPI, Predicted Value (base index, 2015=100)	Additions compared to the same querter of the same year (%)	Additions compared to the same quarter of the previous year (%)
4q/2017	***128,31***	***6,04***	***13,75***
1q/2018	***128,37***	***0,04***	***10,47***
2q/2018	***129,66***	***1,01***	***9,05***
3q/2018	***131,06***	***1,08***	***8,31***

Source: Calculations in SW GRETL, table is author's own work

A graphical representation of HPI development is presented in the graph of Fig. 3.

Fig. 3 HPI development prediction, Czech Republic



Source: author's own work

4. Conclusion

The article dealt with the presentation of a possible procedure in predicting the development of real estate market prices. A creation of a delayed multiple linear regression model appears to be appropriate. The model is also suitable for use in routine expert practice thanks to the simplicity of calculation without owning any computational programme. Thanks to the created model, experts are able to predict the development of the real estate market. The result is bound to the accuracy of the input data, which can be problematic, especially in cases of an attempt to predict price developments over a longer period of time. The novelty of the overall approach consists in the use of mathematical and statistical methods, which are not used in the ordinary expert practice. The model was set for the real estate market in the Czech Republic, however, based on the detailed procedure described in this article, this model can be used in any other country.

Bibliography:

[1] MELOUN, M., MILITKÝ, J.: *Kompendium statistického zpracování dat*. Prague. ACADEMIA, 2002. ISBN: 9788024621968

[2] RYBÁŘ, M.: *Regresní modely a jejich výuka.* Praha 2014. Rigorous work. Charles University in Prague, Mathematics and Physics Faculty.

[3] HEBÁK, P. *Regrese I.a II. část.* Praha 1998. University of Economics, Prague, 1998. ISBN: 80-7079-909-9

[4] ŘEZANKOVÁ, H., LÖSTER, T.: *Základy statistiky.* 1st edition. Prague: Oeconomica, 2013. ISBN 978-80-245-1957-9

[5] Global property guide: *House prices.* Available on-line: www.globalpropertyguide.com

[6] Trading economics: *House price index*. Available on-line: https://tradingeconomics.com