APPLICATION OF ARIMA MODELS IN REAL ESTATE MARKET FORECASTING

Dorota Dejniak, Ph.D.
The Bronisław Markiewicz State Higher School of Technology and Economics in Jarosław
Jarosław, Poland
e-mail: dorota.dejniak@pwste.edu.pl

Janusz Dąbrowski, Ph.D.
The Bronisław Markiewicz State Higher School of Technology and Economics in Jarosław
Jarosław, Poland
e-mail: geo-staszic@wp.pl

Abstract

Forecasts for the real estate market, a key factor to reliably determine the most probable market value of property in the estimation process. The article presents the methodology of ARIMA models' construction and their use for forecasting univariate time series. There have been described the steps to create a model for the example data on price per square meter of office space.

The data acquired as part of the EU project, which studied in detail the real estate market and its broad environment in Jarosław in the years 2000-2013.

Key words: *real estate market analysis, time series, ARIMA models*

Introduction

The real estate market is subject to constant changes associated with the analysis of macro-economic indicators and local factors (BIEDA 2013 and 2016). This market is characterized by dynamics (BYDŁOSZ 2010) associated with the activities of the banking sector, customer expectations as well as risks (DYDENKO 2015) and uncertainties (CZAJA 2001), (DĄBROWSKI 2013). It is a consequence of the fact that property transactions are financed in large part not from own resources but from external sources. The forecasts related to real estate sales premises are extremely important for people who plan to sell or buy a flat (PARZYCH 2013). What is important, it is extraordinary sensitivity of the real estate market to macro and micro-economic indicators.

In the forecast presented in the article there was used the data from the years 2000 -2013 regarding real estate transactions premises for the city of Jarosław. The article presents the methodology of ARIMA models and their use for univariate time series forecasting series (KOT et al. 2007), (LUSZNIEWICZ 2001).. They was presented a forecast price per square meter of residential premises. One of the generally applicable approaches was used to build ARIMA models proposed by Box and Jenkins. IT tool deployed by the authors in the process of model development and determining the forecast was the *Time series module and forecasting* of STATISTICA software.

Analysis of time series

By examining the time series it must be assumed that the process of generating the data contains elements shaping the cyclical nature of the data, the trend and seasonality. The trend can be defined as a monotonic variable component depending on time. Cyclical component is understood as the fluctuations in the long term. Seasonal component determines the periodic fluctuations of less than a year. Random component includes all previously unrecognized factors generating the data. There are usually a lot and often are independent, so it can be assumed that the random component is a modelled normal distribution. Prognostic model based on time series involves identifying features:

$$y_t = f(T_t, S_t, E_t)$$

where:

T_t –component of the trend and cyclical.
S_t - component responsible for seasonal variations,
E_t – random component.

Figure 1 there was presented on an annualized basis, "the number of transactions" carried out in the city of Jarosław in the years 2000-2013, along with the function of a linear trend.

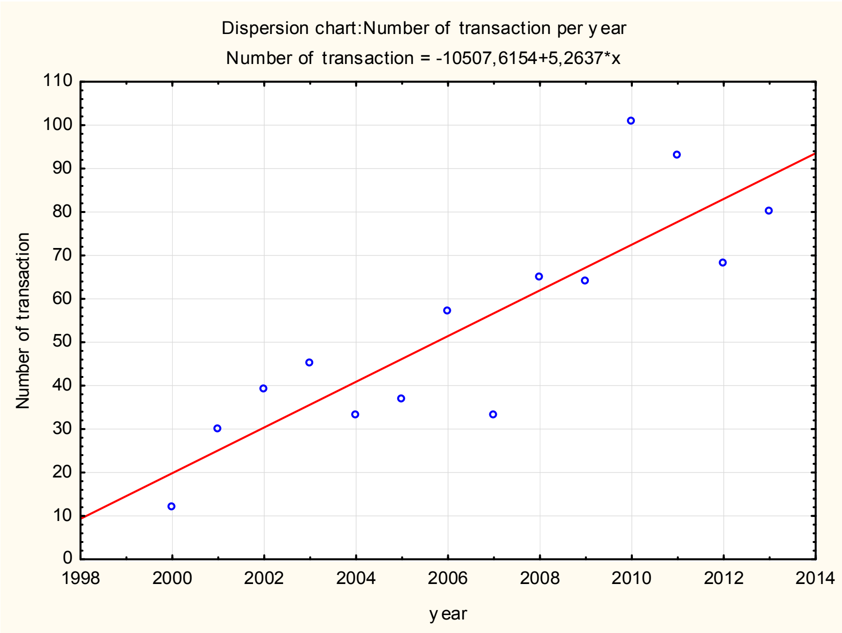

Fig. 1. A scatterplot of "the number of transactions" with respect to the "year". *Source: own work*

The software *Statistica* allows us to determine the basic characteristics of numerical example. The average value of the flat, median. In the analysed period, the average apartment price was close to 100 000 PLN.

	N	Mean	Median	Minimum	Maximum	Variance	Dept. Stand.	Variation coefficient
Mean value of a flat	14	99215,71	84562	52548	151297	1,65E+09	40561,36	40,88

Time series analysis uses the occurrence of autocorrelation (linear relationship) between the series values at time t, and t-j. To estimate the autocorrelation of the series (row j) between the values at times t and t-j it is necessary to determine the implementation of the estimator:

$$\rho_j = \frac{\hat{\gamma}_j}{\hat{\gamma}_0}$$

where $\hat{\gamma}_j = cov\left(y_t, y_{t-j}\right) = \frac{1}{n}\sum_{t=j+1}^{n}(y_t - \bar{y})(y_{t-j} - \bar{y})$,
and $\bar{y}$ is the mean value calculated for the total number of observations.

Interpretation of the autocorrelation coefficients is the same as of the linear correlation coefficient. Determining the value of autocorrelation for further delays $\rho_1, \rho_2,$ ... we obtain a function $\rho(k)$, which is called the autocorrelation function. A special type of time series model is white noise. In this model, the following embodiments do not depend on the preceding and have the same distributions. The average white noise process is zero variance for all t are the same. Positive autocorrelation for the delay is zero. Autocorrelation function of white noise process should be zero for all the delays, therefore, is not significantly different from zero for any delay. Determining whether for the delay autocorrelation is different from zero requires building a significance test and providing a critical area. In the presented analysis there was used a test by Q Box - Piers, a critical region defined by the 95% confidence level has the form (-7.030, 7.030). When there is at least one delay for which the autocorrelation is important, the value of Q statistics belongs to the critical region.

Identifying the components of a series, for the variable under consideration there should also be provided a partial autocorrelation function. This is due to the fact and significant correlation can also be maintained between the more distant in time delays.

The following graphs show the correlation and autocorrelation function for the variable "The number of transactions".

At a significance level of 0.05 there proved to be a particularly significant delay first for the variable "number of transactions" (Fig. 3). In this case, the autocorrelation is 0.581. This demonstrates the influence of the number of transactions from the previous period on the number of transactions in the current period. This confirms the common opinions on shaping the real estate market.

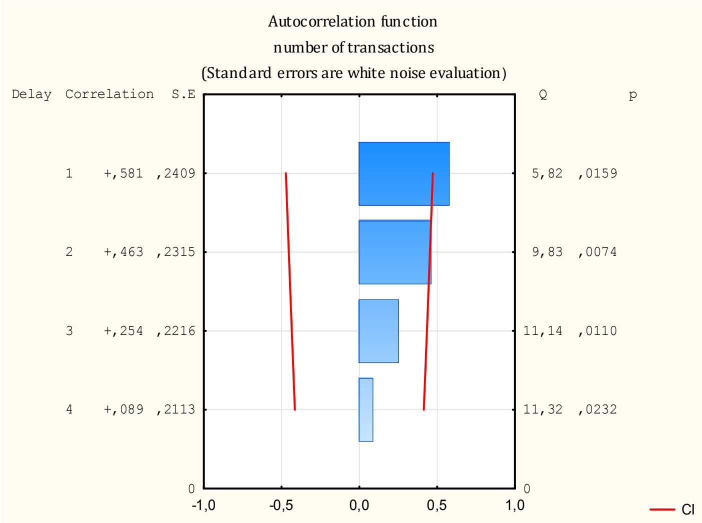

Fig. 2. Graph of the autocorrelation function for "the number of transactions". *Source: own work*

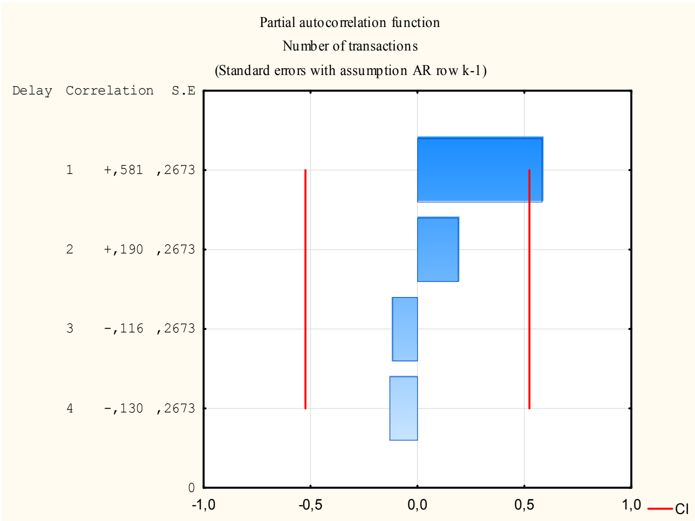

Fig. 3. Graph of partial autocorrelation function for "the number of transactions". *Source: own work*

Arima Models

ARIMA models were introduced to the study of time series in the second half of the twentieth century by Box and Jenkins. The process of building the model consists of three phases: identification, estimation and application. The first step is to identify the characteristics of the analysed time series. A decision is made about the need for data transformation to stabilize the variance or a series of differentiation in order to stabilize average. In the second step there are the estimated parameters' selected models. The final choice of model follows the analysis of several criteria: significance of the model parameters, mean square error. Then the model is subjected to diagnostic screening. The basis of diagnosis is to analyse the properties of model residuals. If the residuals of the process model are white noise, the model can be used for forecasting. In the third stage we should use the model to forecast preparation.

Abbreviation ARIMA consists of three parts AR, I, MA. The first is autoregress. The second is the integration row (indicates the minimum number of first differentiation in the model). The third member is a part of the moving average included in the model. Time series analysis using these models gives excellent forecasting capabilities.

To build the ARIMA model there was used weekly data base developed for the variable "unit price [m2] housing" from 2000 to 2013 (and received 691 observations). A scatterplot of this variable shows the resulting lack of stationarity of the series (Figure 4).

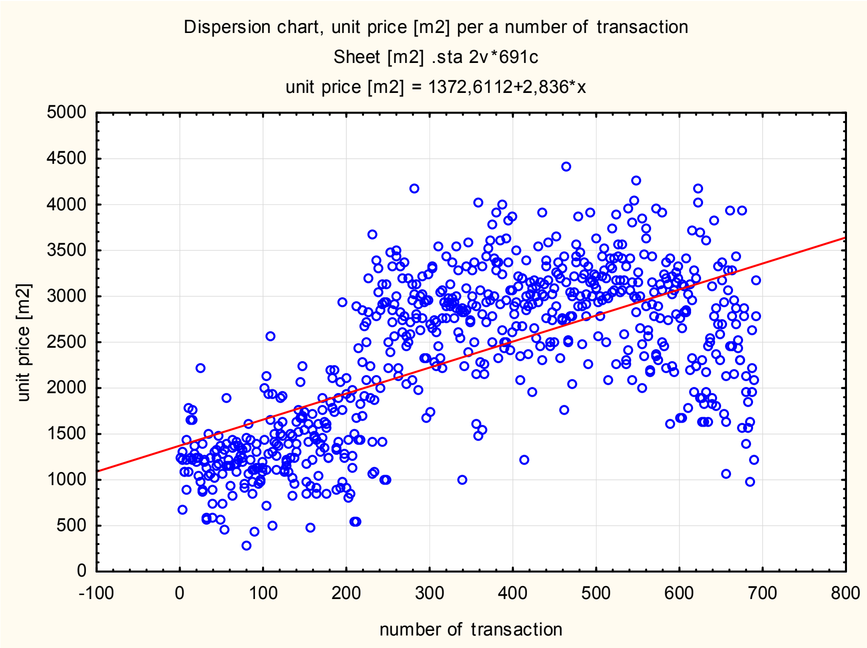

Fig. 4. Figure of variable "unit price [m2]" since 2002 until 2013. *Source: own work*

Basic characteristics of the figures for this variable are shown in the chart below.

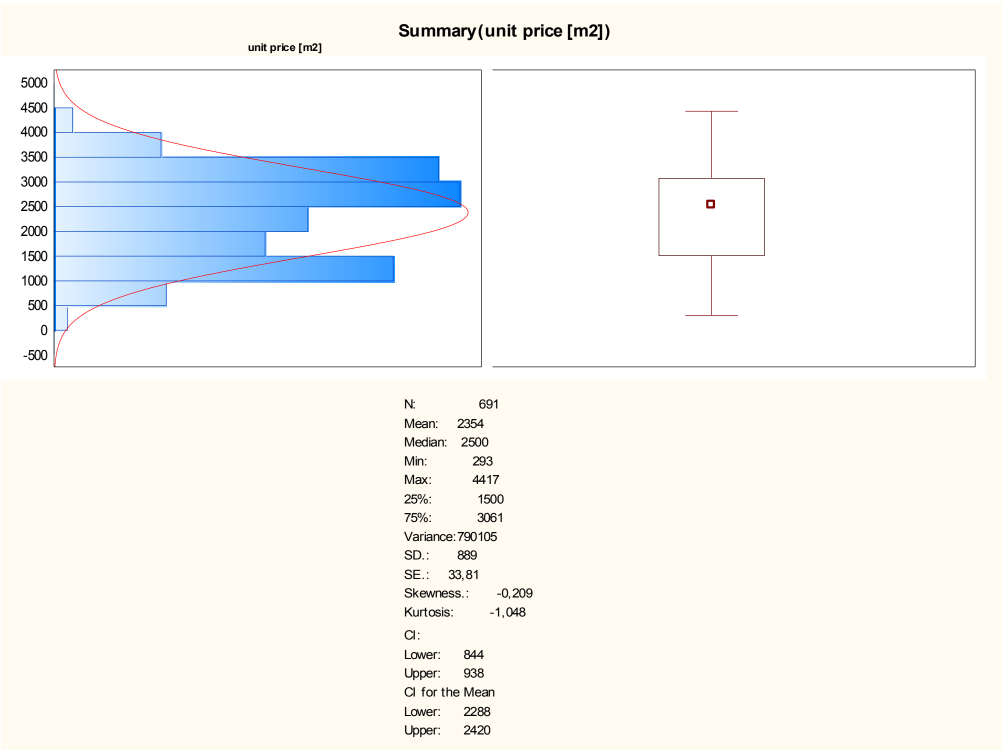

Fig. 5. Basic characteristics of figures for the variable "unit price [m2]" since 2002 until 2013. *Source: own work*

The supposition of a series instability has been verified using correlation analysis. For the adopted level of significance 0.05 autocorrelation coefficients are decreasing very slowly, they are statistically significantly different from zero. They reaffirm the importance of the value of transactions from previous periods.

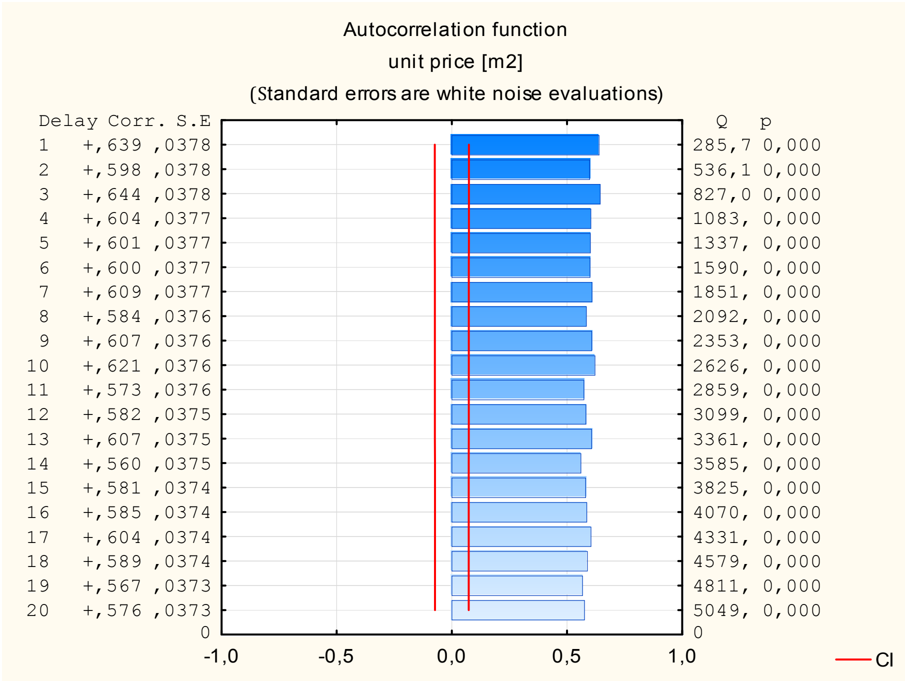

Fig. 6. Graph of autocorrelation function for the "unit price [m2] ". *Source: own work*

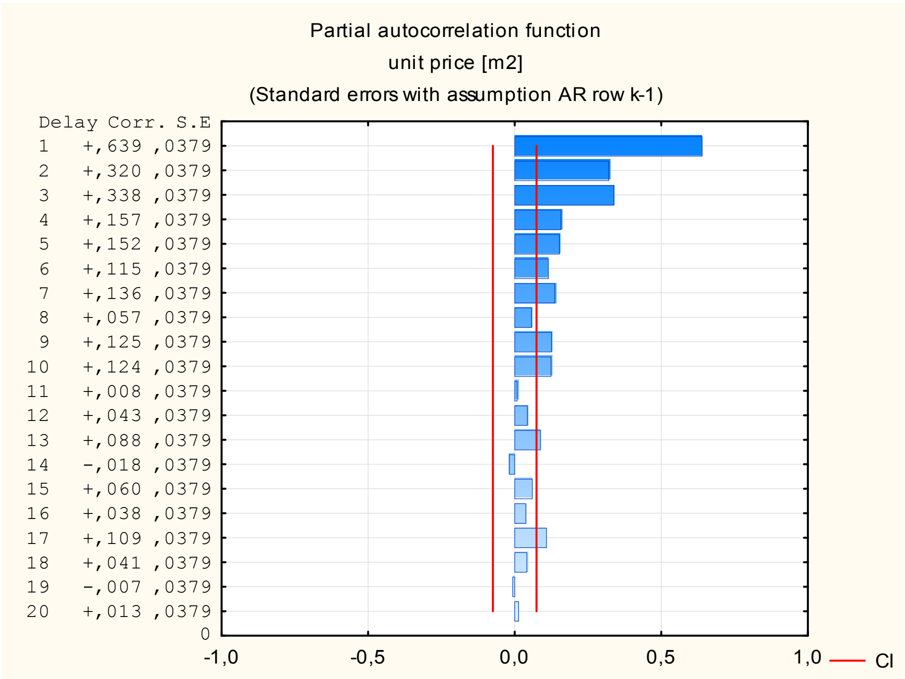

Fig. 7. Graph of partial autocorrelation function for the "unit price [m2] ". *Source: own work*

Due to the lack of stationarity of the model, its pre-differentiation and logarithmization have been made.

After analysing the charts there was initially carried out the identification process, which involved estimation of hypothetical parameters of the model. The best results were obtained using the test method parameters down. Adhered to the recommendation that the maximum government delays did not exceed 3. Estimation ARIMA models is carried out by high credibility. With the evaluation model, we can determine the values of theoretical and residuals.

The correctly obtained model residuals should be normally distributed and they are not characterized by the autocorrelation. As a result the simulation was obtained as the most optimal ARIMA model (1, 1, 1). MS symbol indicates the mean square of residuals.

Table 1. Estimation ARIMA model (1,1,1) to variable "unit price [m2] "

Specifications: the price of unit [m2] Transformation: ln (x) D (1) Model: (1,1,1) Residual MS =, 07412						
	Parametr	Assumpt. - Error std	Assumpt. - t(688)	p	Bottom - 95%p.ufn	Top - 95%p.ufn
p(1)	0,117633	0,041806	2,81378	0,005036	0,035550	0,199716
q(1)	0,923101	0,015687	58,84317	0,000000	0,892300	0,953902

Source: own work

It has also been carried out the analysis of normality residues and their autocorrelation.

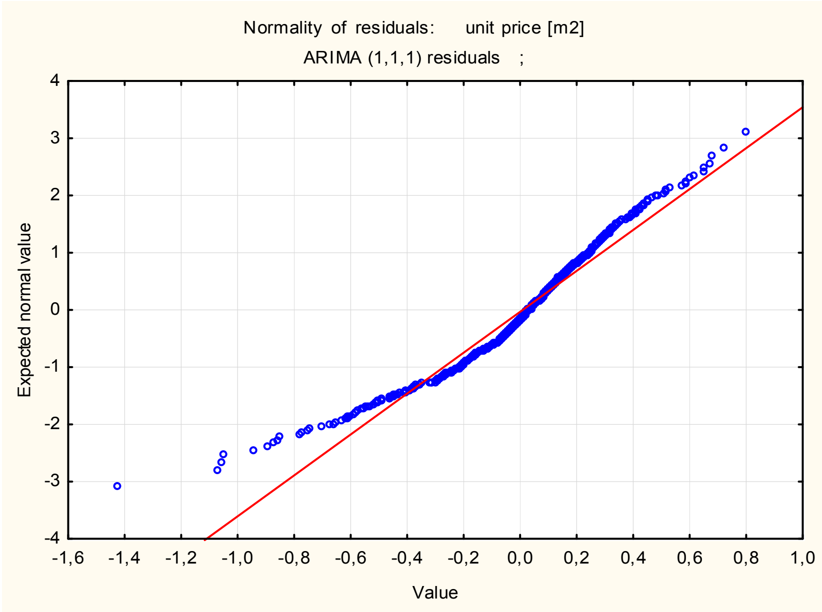

Fig. 8. Chart of normality residues of the variable "unit price [m2]" for the ARIMA model (1,1,1). *Source: own work*

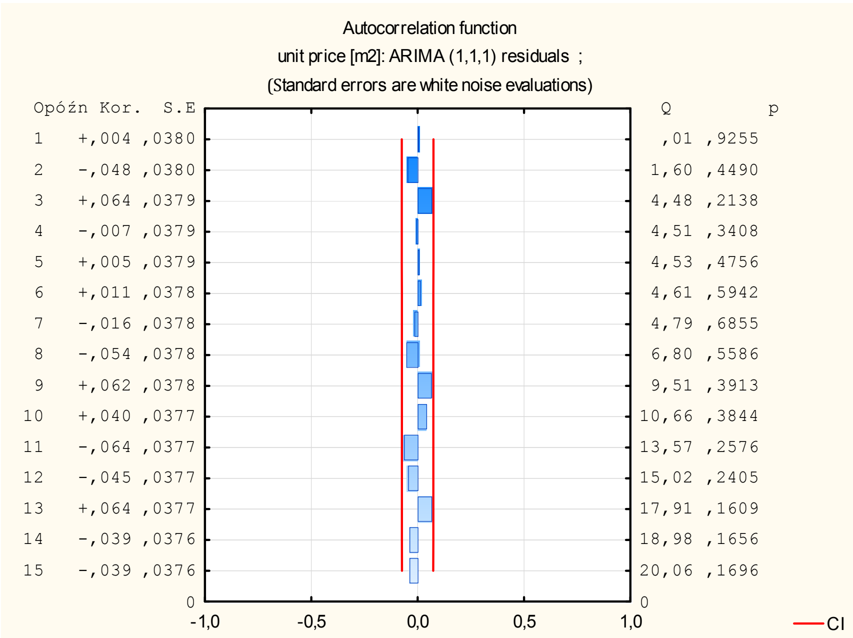

Fig. 9. Chart of autocorrelation function for the residuals of ARIMA model (1,1,1) "unit price [m2] ". Source: own work

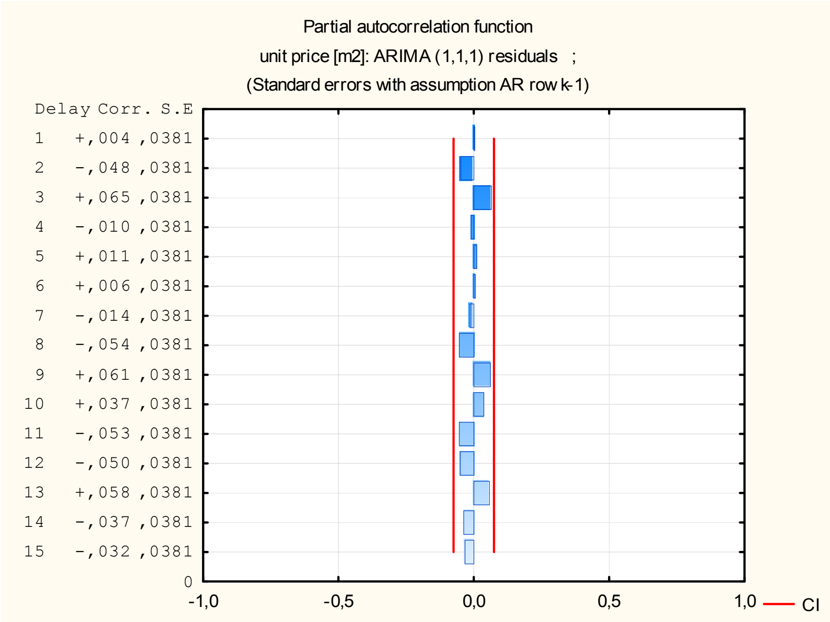

Fig. 10. Chart of partial autocorrelation function for the residuals of ARIMA model (1,1,1) "unit price [m2] ". *Source: own work*

The estimated model was subjected to diagnostic check. Figure 9 and 10 indicate a lack of autocorrelation of residues, which suggests that the residuals are white noise process.

Statistica program for ARIMA models enables too forecast the tested variable. The table below shows the forecasted values for 25 weeks.

Table 2. Forecast value for the variable "unit price [m2]" for the next 25 weeks

Prediction; Model: (1,1,1) Season delay: 12 Specifications: the price of a meas. unit [m2] Home Base: 1 End Base: 691			
	Prediction	Bottom - 90%	Top - 90%
692	2262,476	1444,901	3542,664
693	2173,345	1376,360	3431,825
694	2163,093	1366,876	3423,113
695	2161,891	1363,772	3427,091
696	2161,749	1361,419	3432,564
697	2161,733	1359,165	3438,204
698	2161,731	1356,933	3443,853
699	2161,730	1354,714	3449,495
700	2161,730	1352,506	3455,126
701	2161,730	1350,309	3460,747
702	2161,730	1348,123	3466,358
703	2161,730	1345,948	3471,959
704	2161,730	1343,785	3477,550
705	2161,730	1341,632	3483,130
706	2161,730	1339,489	3488,701
707	2161,730	1337,357	3494,263
708	2161,730	1335,236	3499,814
709	2161,730	1333,125	3505,357
710	2161,730	1331,024	3510,890
711	2161,730	1328,933	3516,413
712	2161,730	1326,852	3521,928
713	2161,730	1324,781	3527,433
714	2161,730	1322,720	3532,930
715	2161,730	1320,669	3538,418
716	2161,730	1318,627	3543,897

Source: own work

There is also a possibility of graphical visualization of the resulting predictions.

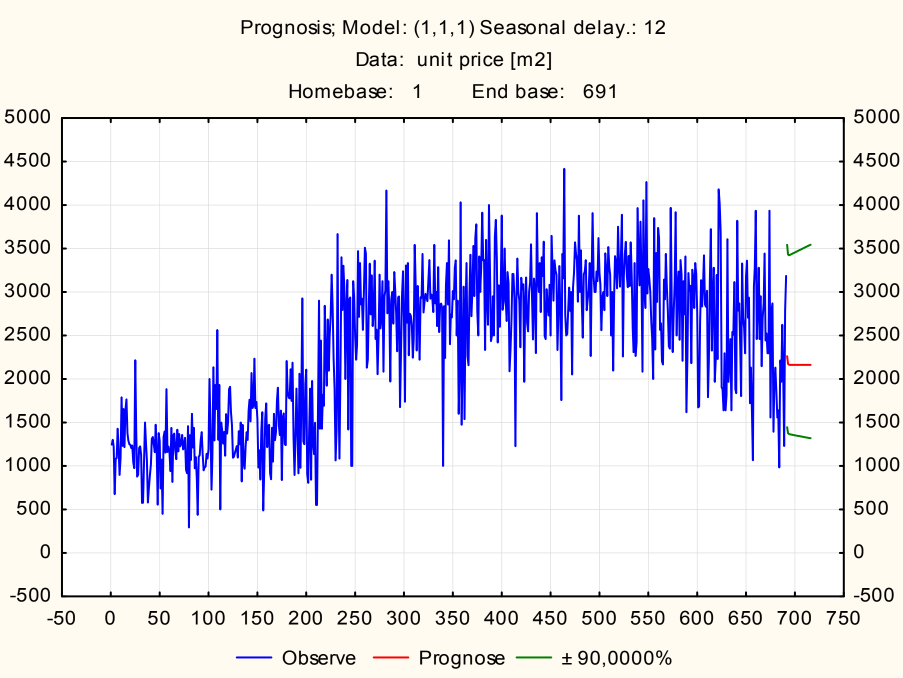

Fig. 11. Graphical forecast for the variable "unit price [m2]" for the next 25 weeks. *Source: own work*

Carried out in this way analysis of one of the variables tested, allows for justification of the use of ARIMA models in the analysis of real estate market. The problem with the unit price forecasts [m2] was a lack of stationarity of the test series, as a consequence of sensitivity to the "initial conditions". It was defeated by logorithmization and diversity of series.

Summary

Using the ARIMA models for forecasting time series, we should expect a number of problems concerning the order of differentiation or type of model. The development of such models is labour intensive and requires expertise and IT support, e.g. using STATISTICA. Their application does not guarantee a better performance in comparison to other, sometimes simpler methods. However, the advantage of ARIMA models is the presentation of the internal structure of the chain and its generating mechanism.

Because the real estate market is extremely sensitive to the "initial conditions", creating models for the variables describing this market is extremely difficult, and adjusting a model and forecasting is burdened with high risk and uncertainties. The proposed quantitative approach is the enrichment of decision information and broadens the forecasting methodology in the analysis of real estate market. It may be used in decision support systems.

References

BIEDA Agnieszka, PARZYCH Piotr. (2013). Development of spatial politics of monumental towns based on Krakow example. International Multidisciplinary Scientific GeoConference: SGEM: Surveying Geology & mining Ecology Management, 2, 143.

BIEDA Agnieszka, BYDŁOSZ Jarosław, PARZYCH Piotr (2016): Preliminary assessment of urban planning documents in the Cracow area published on the Internet. SGEM 2016. 16th International Multidisciplinary Scientific Geoconference, conference proceedings. Vol. 3, Cartography and GIS. Sofia : STEF92 Technology Ltd., cop. 2016. ISSN 1314-2704; ISBN: 978-619-7105-60-5. S. 367–374.

BYDŁOSZ Jarosław, PARZYCH Piotr, DĄBROWSKI Janusz (2010): Fiscal impact of selected legislative solutions in Poland on sustainable development in the context of the real estate market analysis. SGEM 2010 : 10th International Multidisciplinary Scientific Geoconference: conference proceedings, Vol. 2. Sofia : STEF92 Technology Ltd., cop. 2010. ISSN 1314-2704; ISBN 978-954-91818-1-4. S. 1165–1172.

CZAJA Józef, Metody szacowania wartości rynkowej i katastralnej nieruchomości, 2001, Wydawnictwo AGH.

DĄBROWSKI Janusz, FIREK Karol (2013) The assessment of real estate market as an economic tool affecting the protection of environment, W: SGEM2013 : GeoConference on Ecology, economics, education and legislation, Albena, Bulgaria S. 293–300. ISSN 1314-2704). — ISBN: 978-619-7105-05-6. — S. 293–300. — Bibliogr. S. 299–300.

DYDENKO Jerzy [red], 2015, Szacowanie nieruchomości, rzeczoznawstwo majątkowe, Wolters Kluwer SA.

KOT Stanisław, JAKUBOWSKI Jacek, SOKOŁOWSKI Andrzej, (2007), Statystyka, Difin.

LUSZNIEWICZ Andrzej, SŁABY Teresa, 2001, Statystyka z pakietem komputerowym STATISTICA PL, Teoria i zastosowania, Wydawnictwo C. H. Beck, Warszawa.

PARZYCH Piotr, BYDŁOSZ Jarosław, DĄBROWSKI Janusz (2013). Analysis of additional economic parameters necessary to grant a loan to Local Government Units based on real estate market analysis. 13 th International Multidisciplinar Scientyfic GeoConference SGEM 2013 Bułgaria. Environmental legislation, multilateral relations and funding opportunities. — Sofia : STEF92 Technology Ltd., cop. 2013. ISSN 1314-2704). — ISBN: 978-619-7105-05-6. — S. 41–48. — Bibliogr. S. 48