"RESIDUALS ANALYSIS FOR CONSTRUCTING 'MORE REAL' PROPERTY VALUE"

Abstract

The price of a real estate is based on its value. The real estate market, where the real estate price is established, is dynamic and undergoes continuous change.

Considering the fact that nothing is fully deterministic or fully stochastic in nature, the author has put forward a compromise based on the conducted analyses in order to better diagnose the “live spatial structures”. The considerations presented in this paper provide grounds for the claim that integration of geo-deterministic inference (represented by the geostatistical model) and geo-stochastic inference (represented by maps of residuals), linked to space valuation, makes it possible to dynamically diagnose and characterise spatial phenomena and to make rational forecasts (and, consequently, planning) of changing in space and, therein, on the real estate market.

Key words: residuals from a model, stochastic and deterministic factors, geostatistical model, spatial model of residuals.

1. Introduction

The market economy is an evolving phenomenon and, according to its classic principles, the state of equilibrium in it means the absence of any emotion and total stagnation. Simplification of the market system in terms of equilibrium, where supply and demand are equal and which practically manifests itself by investors’ linear reactions to information, is easy to interpret and study. However, this approach means the absence of any development and is contrary to the real conditions in which the system works, which is confirmed by Peter’s [2] claim that if a system is to survive, it has to evolve and to be far from a state of equilibrium.

The above indicates that diagnoses of spatial value based on statistical modelling involving pure cause-and-effect principles, is frequently incomplete and insufficient to arrive at correct conclusions and decisions based on them. It is commonly known that the real estate market, with its environment made up of the legal, economic, technical and social conditions, is a dynamic and constantly evolving system, which makes it difficult to study and forecast. Consequently, the author of this study recognised that the supplementing of knowledge by a factor unexplained by

the model (residuals) provides an important complement to statistical (deterministic) formulas, devoid of elements of uncertainty and randomness.

In presenting the practical aspects of applying residuals in real estate appraisal, a set of 800 sold land real estates was used. The results of the analyses will be useful in establishing and forecasting the values used by various entities for various purposes, e.g. analysis of socio-economic spatial development, competitiveness and innovativeness of regions from the investors' perspective, forecasting the economic results of spatial changes on the real estate market, etc.

2. Interpreting residuals – stochastic relations on the real estate market

It was understood in the mid-20th century that seeking the deterministic principles of nature meets logical and practical obstacles [RAO 1994]. Understanding and accepting elements of uncertainty in all areas of human life and recognising randomness as an unavoidable part of numerous classes of objects and their manners of existence has brought about the development of theories based on stochastic principles of nature. According to PRIGOGINE [3] and RAO [4], nothing is fully indeterministic or fully deterministic; each phenomenon is a peculiar mixture of the two extremes.

The real estate market is a place where facts are created. It is undoubtedly a natural phenomenon, which according to PRIGOGINE'S [3] idea, always has an integral element of randomness and irreversibility, since only artificial creations can be fully deterministic and reversible. Therefore, an analysis of the effect of the market on the real estate price allows for selection of the factors which influence the facts according to the following diagram (Fig. 1).

Fig. 1. Place of residuals in causative sources of a real estate price.
Source: own study

The basic factors which differentiate the real estate price include time, place, technical and legal conditions and people, who make decisions based on their preferences and the knowledge

about the real estate market relations. The human factor which influences the real estate price is an outcome of certain reasons, motives and external information. It is usually predictable. Uncertainty and dynamism are linked to the information getting to a participant in the market. This process is usually non-linear and it significantly affects the final price of a real estate.

Between the basic (static) elements which influence the price of a real estate and the most important part of the diagram – human – there is a set of factors which are described as undefined. They are linked to random (stochastic) phenomena which apparently occur by chance and are thus difficult to predict. If they occur separately, they are hardly noticeable, but in a larger group they make a characteristic factor which makes the valuation of space more dynamic and influences the development of market structures.

Giving an unambiguous definition of the factor which represents chance phenomena and clear determination of their reasons of existence is difficult and can cause many technical and hypothetical problems. It is therefore difficult to prove the stochastic aspects of real estate appraisal, all the more so that they are complex (frequently exogenous), dynamic and random. Analysing the phenomenon can be important for forecasting the value of a real estate, which is closely correlated with the final price of a real estate on the market.

It was assumed that the stochastic factors which affect the real estate value are represented in a simplified manner by residuals from a model, according to the claim that **residuals** provide a formal measure of differences between specific (stochastic) phenomena and (deterministic) generalisations, provided by the created models.

From the mathematical point of view, the residuals can be expressed by the formula:

$$e_i = y_i - f(x_i, \beta_i) \quad (1)$$

where: e_i – residual from a model

y_i - empirical (observed) variable

x_i - model variables

β_i - model parameters

The residuals interpretation possibilities vary depending on the type of relations used and those whose presence is recognised. According to the stochastic theory, residuals are a manifestation of random processes and uncertainty. In statistical modelling of phenomena, residuals occur as a random factor, but they do not express random processes, but a measurement error or an error of selecting a functional form. Whereas in the deterministic theory, where events occur according to cause-and-effect principle, residuals should not occur at all as, to put it simply, everything can be forecast and formalised with formulae (expressions and equations).

Therefore, it can be claimed that **residuals** make up the area of phenomena which fail to be explained by the model because of a lack of the relevant statistical data, incomplete theoretical and practical knowledge about the modelled area and/or because of the dynamic, frequently unforeseeable, tempo of change of the study area.

3. Analysis of residuals as applied in real estate appraisal

The differences between experiment and theory, and precisely – between specific events and generalisations – have been used for a long time to obtain knowledge about the “behaviour” of certain events. John Stuart Mill was the first to have recognised the “residuals” as a source of better understanding of phenomena. This is what he wrote in his work “A system of logic”, published in 1874 [Thomas 1973]: *“Of all the methods used to study nature, this one is the most productive and gives the most unexpected results: it frequently informs us about a sequence for which neither the cause nor the result are clear enough to attract the observer’s attention”*. The residual analysis has been classified as one of extremely useful research methods; according to Horts [1] it is *“one of the most interesting and useful parts of predictive research, to be used where prediction failed”*.

Residuals have been used in numerous fields: statistics, psychology, sociology, geography, politics and ecology. The analysts who have employed residual analysis for various purposes have subscribed to the notion that studying and diagnosing a phenomenon based on purely deterministic rules and relations provides a certain picture of the analysed structures, yet it is usually narrowed, generalised and incomplete. An analysis and presentation of the real relations with the use of even the most complicated mathematical formulas allows only for the interpretation of what can be formalised with equations, showing a specific result for a given cause and vice versa. In the creation of a certain model, the most accurate reflection of a phenomenon should be strived for. This is usually done with ‘adulteration’ data. However, a part is usually left which is called a “residual” or an ordinary “error” caused by an inaccurate interpretation of the phenomenon.

It seems that the residual component cannot be eliminated (in phenomenon analysis), even if all the statistical rules and principles are observed. This is caused by the occurrence of factors and phenomena which cannot be simply formalised and expressed in the form of an equation. They are only felt intuitively and can be interpreted as a result of experience; they are revealed only after becoming visible.

The level of abstraction of the presented way of understanding phenomena, including the real estate market, is relatively high. Therefore, in order to make the application of residuals more versatile, two forms of the application of residuals to the analysis of spatial phenomena have been proposed. These can be the residuals presented in a numerical form (in orthodox analysis) and in a graphic form (in heretic analysis) – Fig. 2.

Fig. 2. Form of residuals presentation.
Source: author's study.

3.1. Numerical form – the structure of the geostatistical model

The model is called geostatistical in order to emphasise the role of the spatial factor in real estate analysis and in the field of research on spatial management and to differentiate the econometric models which stem from the theory of economics.

According to the assumption adopted in the work, the geostatistical model represents a general variability which results from deterministic premises. Including residuals in the process of the model construction allows for minimising the irregularities associated with the model formal structure. The aim of the assumption is to supplement the statistical (deterministic) formulas, devoid of the elements of uncertainty and randomness, by including the components unexplained by the model (i.e. the residuals) into the model itself.

Spatial systems are usually complex and heterogenic in structure and action. The diagram of the model creation, presented below, takes into account various phenomena association within the diagnosed spatial structure, allowing for non-linear relationships, as these frequently occur in the relations of complex phenomena.

Fig. 3 shows the consecutive steps of creating a parametric geostatistical model, using a residual component.

Fig. 3. Diagram of execution of subsequent stages of construction of a geostatistical model with the use of residuals.
Source: the author's study.

The model construction was based on the real estate market in Olsztyn. The study object is characterised in Table 1.

Table 1. Characteristics of the study object

Study object	Explained (dependent) variable	Period of analysis	Number of observations ¹	Number of attributes ²
Olsztyn – land real estate without buildings	real estate price	1997 - 2005	821	25

¹transactions have been collected and analysed by the author with the assumption that, in this type of analyses, the more the better

²the real estates have been described with 25 attributes in order that the loss of data in further statistical analyses should be as low as possible; these were the following.: area, date, form of possession, dimension of plot, location of plot, equipment into – water-supply, gas, phone, sewage, electric energy; form of area, approach and access to plot, difficulty of traffic in the neighbourhood, rail lines and other neighbourhood difficulties; allocation in local plan; form of transaction; attractiveness of position to forest areas and park and areas of water reservoirs; location connected with the distance from the main commerce and city centre and means of transport.

Source: the author's study.

The main assumption adopted in creating a geostatistical model was to reflect with the highest possible fidelity the factors which determine the picture of the spatial processes. Below is a brief presentation of the stages of the construction of such a model (detailed information can be found in [5] and [8]).

Stage 1 – *Preliminary specification of attributes*

Before the beginning of the model creation, the phenomenon has to be analysed thoroughly. In practical terms, it manifests itself in selecting a set of characteristics to describe a phenomenon which maximises the information capacity of the selected variables connected to the study subject. Such a set of variables should be appropriately quantified, should have a specific level of credibility and should maintain the appropriate accuracy. Twenty-five independent variables have been adopted in the study which describe the values of the real estate market in Olsztyn.

Stage 2 - *Selection of the a priori variables*

Beginning the process of selection of the *a priori* variables, a preliminary analysis should be conducted of the selected variables according to the assumptions adopted at the model construction. Not negating, but undermining the fact of categorical assumption linear relationships between the variables in a model, the optimal set of variables can be identified with the use of **single-factor analysis of variance**. The single-factor analysis of variance has been employed in order to determine the occurrence of a significant effect of the independent variable on the explained variable. The Pearson correlation coefficient, which is resistant to curvilinear relationships, has been applied in order to determine the strength of relations between the variables.

Stage 3 - *Determination of the model structure*

After the selection of the preliminary (*a priori*) variables, the relationships between the

attributes are determined and the model form is established.

In this study, the model form is selected by analysing the individual relations between a pair of variables (i.e. an independent and dependent variable), in effect, creating a multi-functional model.

The study employs 7 functions most frequently applied in this type of analyses: linear, hyperbolic, logarithmic, exponential, square, third and fourth degree polynomials. The results of the measures based on the model residuals – a) sum of squares of the residuals (*SSe*), b) the model adequacy (*A*), c) mean square error (*SSq*) - have been adopted as the criteria of the interrelations quality.

Stage 4 – *Selection of variables a posteriori*

The selection of the optimal set of characteristics is one of the most important stages of the model creation. It is not always that the set of characteristics which is optimal before creating the model retains its optimality after its structure has been determined. Considering the above, an additional procedure of selecting variables *a posteriori* has been proposed with the use of the model residuals (as the measures of fit), according to the principle of maximising the information capacity and minimising the size of the optimal set. The variables which after being removed in a simulation did not change or improved the measures of fit were eliminated from the model. The process aimed at eliminating the redundancy of variables.

Stage 5 – *Determination of the functional form of the interrelation of the model components*

The creation of this stage was inspired by an analysis of tendencies in the development of theory of prediction. It can be assumed that the interrelations between the explaining variables are not only additive. There can also be multiplicative (indicative) relations.

Therefore, the optimal form of components of the analysed model can be selected after the analysis of sensitivity of the results of fit in reaction to a multiplicative form of the occurrence of variables in the model. In practical terms, it means that those variables are present in the model in the multiplicative form for which the simulation indicated the improvement of the values of the measures of fit of the model. As a result of the analyses, part of the model variables have been noted in the additive form (model trend), i.e. those which coincide with the “trend” have been noted in the multiplicative form, e.g..

$$\hat{y} = a + \sum f(x_{pow}, x_E, x_W, x_K, x_{DD}, x_{atrW}, x_{atrL}) \times \prod (1 + f(x_T, x_{data}, x_{PZP}, x_{odlH}, x_{odlC})) \quad (2)$$

Stage 6 – *Elimination of inadequate cases (components)*

It is not always that the cases which stand out are “inadequate”. Frequently these are the phenomena which are connected with the property of the analysed phenomenon and the randomness of the market. It is assumed that the residuals are the best objects to determine the

diverging cases. Therefore, the diverging cases can be identified with the use of residual analysis in which the so called “three sigmas principle” was employed (plus/minus three standard deviations) together with taxonomic techniques (including the agglomerative technique and cluster analysis), which were used to classify the diverging spatial units. The agglomeration technique was used to determine the number of clusters. The cluster analysis itself was conducted by employing the method of grouping k-averages, which are included in the group of so-called "optimisation" processes.

The fact of classifying the identified cases as inadequate (hyper-diverging) was ultimately confirmed by conducting a analysis of case occurrence in terms of time and space. The cases are classified as ultimately diverging if the previously identified objects do not occur in one time or place in space, i.e. they result from an action of random factors.

Eleven cases have been identified in the study (Fig. 4) which were classified as hyper-diverging and eliminated from further work.

Fig. 4. Space-time analysis of the occurrence of cases identified as diverging.
Source: the author's study

The research method used in the study and the statistical methods employed are aimed at minimising the possibility of making formal errors associated with the model structure. The proposed method is flexible enough to be used in diagnosing the results of various spatial processes; what is even more important, it is not necessary to employ the whole procedure to improve the model structure – only some of the above-mentioned stages have to be executed

3.2. Graphic form - maps

One of the main reasons for the complexity of the model structure, when the model reflects real relationships, is the complicated and dynamic nature of the phenomena – which are extremely difficult to present in simple mathematical terms. All the changes and transformations of the structures under study (including spatial ones) are caused by emerging innovations, both in terms of inventions or things and innovative ideas and behaviour. The innovations have their source in the divergence between reality and one's notion of it, in unexpected events, in changes in the perception of phenomena and problems, and, consequently, in the emergence of new awareness and knowledge.

Spatial innovations manifest themselves in the emergence of a certain factor (improvement) which effects the evolution of spatial-social structures. The factor disturbs the given structure and its order and creates new relationships which are revealed later. Spatial innovations are, therefore, spatial processes which are a manifestation of qualitative and quantitative changes, unreflected in long- and short-term tendencies.

The market phenomena, when related pointwise, provide incomplete and simplified information. As a result, the interpretation of results from the spatial point of view is a very important element of space valuation. In conducting the residuals analysis according to the assumption and study hypotheses, one can identify the areas which deviate from the model values. In this study, a map of residuals has been created with the use of SURFER software fig. 5.

Fig. 5. The spatial model of residuals, land real estate without buildings – a three-dimensional map.
Source: own study

The map of residuals shows the spatial distribution of the part of the explained variable which is connected with phenomena other than those taken into account in the statistical analysis of the model. According to the above presented assumptions, these can be spatial innovations. An analysis of the residuals distribution suggests that the extreme deviations of residuals present in a cluster can indicate areas of change in investors' preferences and, consequently, the value transformation in the area. The phenomenon in relation to the real estate market in Olsztyn could be caused by the following issues:

- availability of free areas – there is a downward tendency in the real estate market in Olsztyn in the availability of free areas, which affects the investors' behaviour and changes the value of real estates;
- availability of mortgages – increased supply and availability of mortgages by decreasing the interest and the required credit rating;
- the price of apartments – increased demand and a significant increase in prices of flats in blocks in the years 2000 - 2006;

- expansion of the University – increased demand for flats in its surroundings;
- change in social awareness – a tendency to buy houses associated with prestige and changes in social status after 1990;
- increased wealth of the society – changes in the economic situation in the country and in the region;
- the forecast of the economic situation – positive outlook for the future, faith in the improvement of the country's economic situation, accession to the EU;
- local government – better management of local resources, changes in policies on the local level, more favourable policy towards investors.

3.2.1 Application the residuals with a spatial connotation

The interpretation of residuals from the spatial perspective can be applied in spatial management in many ways and can play a variety of functions:

- a) identifying investment niches,
- b) defining sets of uniform real estates*,
- c) identifying the unbalanced areas based on spatial phenomena*,
- d) measuring the model quality *.

*these applications have been described in detail in [6] and [7].

The most important function in estimating the value of a set of real estates is the one of identifying investment niches. From the forecast point of view, accumulation of the “negative” residuals from the model in an area can be a sign of predicted spatial innovations, which indicate the emerging of investment niches (diagram 1 and Fig. 5 – the niches marked in green to blue). The negative residuals are a result of the predicted value, which is higher than that indicated by the empirical data. The areas where the “positive” residuals are clustered represent areas of intensive development.

Diagram 1. The process of determining an anticipated area of investment niches.

Source: The authors' own elaboration

4. The use of residual analysis by entities on the real estate market

The residuals which supplement the information about the real estate market by stochastic factors in space valuation can have wide applications in real estate management, mainly in identifying the real market tendencies and in creating real estate value. The information and analysis results presented here can be used by various entities on the real estate market for various purposes, taking into account specific forms of the residual analysis (encoded as symbols *) e.g.:

- ❑ local governments in developing local development strategies (m);
- ❑ communal governments for developing studies of conditions of communal development (m), (mn);
- ❑ communal governments for forecasting the financial results of adopting and changing the local area development plan (m), (mn);
- ❑ local governments in identifying groups of real estates similar in terms of value for fiscal purposes (n), (mn);
- ❑ real estate administrators, real estate agents and appraisers in individual consulting in real estate investment (m), (mn), (n);
- ❑ highly-qualified real estate appraisers and consultants in creating individual value – speculative consulting (n), (mn);
- ❑ bank experts and real estate appraisers in establishing the value of real estate for bank mortgage purposes – particularly in high risk investments (n), (mn);
- ❑ real estate consultants in creating “real estate portfolios” for investors (m), (mn), (n);
- ❑ academic circles, researchers and analysts who study the real relationships on the market, taking into account the randomness and uncertainty, mainly in forecasting the real estate value and seeking non-deterministic variables (m), (mn), (n);
- ❑ individual investors – for determining the competitiveness of regions (m).

* map – complete (m); map-numerical – partial (mn); numerical – complete (n)

These methods of residuals analysis have been assigned to the specific problem groups while observing the principle of optimisation between the cost of studies and the value of results and their applications.

Summary

The phenomena which take place on the real estate market make the real estate value one of the factors which determines the development and transformation of spatial structures and systems. Its establishment in a credible manner requires not only the study and analysis of the spatial value based on the commonly known value-creating factors taken into account in the model, but also elements which can be provided by a spatial analysis of residuals from a

geostatistical model. Supplementing this knowledge by a component unexplained by the model, and its thorough analysis, allows for predicting untypical behaviour, and, consequently, for determining the real estate value with greater precision. Perceiving space as a purely deterministic structure does not result in complete and satisfying effects of its valuation.

The innovations (perturbations), which disturb the “life” of the real estate market, can be represented by the residuals. The market system and its structure changes when intensive accumulation of residuals is observed in certain areas of the market. Therefore, the residuals can indicate the creation of a new structure and a change – in this case, a change in the value of space.

Using the fact mentioned above and sorting the issues connected to real estate management, where it is necessary (or highly useful) to determine the “speculative” value of real estate, the application of residual factors can be presented as in the following diagram – Fig. 6. According to this diagram, geostatistical modelling and “spatial” residuals play specific functions in the dynamic system of spatial valuation. The diagram is based on the assumption of integrating residual factors as an element of deterministic concluding, supplemented by a stochastic factor (because the geostatistical model is based on residuals), and stochastic conclusions based on maps of residuals in real estate management procedures.

* REM – Real Estate Management; SM – Spatial Management

Fig. 6 Use of the residual and the geostatistical model in diagnosing the value of land in the context of real estate management issues.

Source: own study.

References

- [1]. Horst P and others 1941. *Prediction of Personal Adjustment*. Social Science Research Council, Bulletin 48;
- [2]. Peters E. *The theory of chaos and capital markets*. Wig-Press Warsaw 1997;
- [3]. Prigogine I., Stengers I. 1990. *From chaos towards order*. Państwowy Instytut Wydawniczy, Warsaw;
- [4]. RAO C. R. 1994. *Statistics and truth*. Wydawnictwo PWN, Warsaw;
- [5]. Renigier M. *The residual as the criterion of the econometric model matching*. Conference materials of the 12th National Scientific Conference of the TNN "Analysis and modelling of the real estate market for the needs of evaluation". Olsztyn 2004;
- [6]. Renigier M. *The use of methods for analysis of the residual from statistical models in real estate management*. Doctoral thesis, Olsztyn 2004;
- [7]. M. Renigier, S. Żróbek, R. Żróbek. *The residual from statistical models in valuation and management of real estate*. Annual conference of the European Real Estate Society – ERES. Dublin. 2005
- [8]. Renigier M. *Construction of the geostatistical model with the use of the residual*. ACTA SCIENTARIUM POLONORIUM, Administratio Locorum (Spatial Management)
- [9]. Thomas E.N. 1973 *Maps of Residuals from Reresson: Their Properties and Applications in Geographic Research*. A Review of Foreign Geographic Literature;