

SIMULATION MODELING CONCEPT OF PROPERTY MARKET TRANSACTIONS

Radosław Cellmer, PhD

Faculty of Geodesy and Land Management
University of Warmia and Mazury in Olsztyn
e-mail: rcellmer@uwm.edu.pl

Katarzyna Szczepankowska, M.Sc

Faculty of Geodesy and Land Management
University of Warmia and Mazury in Olsztyn
e-mail: katarzyna.szczepankowska@uwm.edu.pl

Abstract

The real estate market is a specific and imperfect field of research, and its complex structure and the presence of information gaps necessitate the use of advanced analytical tools. One such tool is simulation modeling, which has a variety of practical applications and can be used to model real-life systems characterized by a high degree of complexity and a high share of random components. In this paper, virtual data was used to simulate transactions on the local real estate market. Simulation tools were applied to generate additional information about transactions, their spatial distribution and transaction prices. The applicability of the iterative Monte Carlo approach with a standard regression model and a spatial regression model was evaluated.

Key words: property market, simulation modeling, spatial regression.

JEL Classification: C53, R32.

Citation: Cellmer R., Szczepankowska K., 2014, *Simulation Modeling concept of property market transactions*, Real Estate Management and Valuation, vol. 22, no. 1, pp. 94-102.

DOI: 10.2478/remav-2014-0011.

1. Introduction

The real estate market is considered to be highly imperfect due to insufficient transaction data, which often prevents the use of quantitative methods in market analysis. In weakly developed markets, transactions can be simulated to supply the required data for the development of price and value models and to evaluate various development scenarios on the local real estate market.

Simulation modeling has a wide range of applications in production (service system models), transport systems (freeways, airports, railways), supply chains (production and system inventories, distribution networks), etc. Simulation models are also deployed in optimization research to determine the best or allowable parameter values (ATLIOK, MELAMED, 2007). Simulation tools have been used in major engineering projects, including the construction of the trans-Alaskan natural gas pipelines project and planning operations for the Suez Canal, and to simulate cosmic explosions in laser compression experiments (REMINGTON et al.).

The real estate market is a complex system that is relatively difficult to analyze, which justifies the use of simulation tools. Simulation modeling provides analysts with information about the way a system operates within a given set of input parameters. Such parameters include the specific characteristics of the market environment and transaction data.

2. The nature of simulation

The word "simulation" originates from the Latin term *similis* which means "similar" or "similarity". The

initial meaning of "simulation" can be derived from the Latin words *simulatio* – "pretend" or "pretense", *simulans* and *odsimulare* – "pretend", "make similar" (KOPALIŃSKI 2006). The contemporary meaning of the word can be traced back to the Greek term *mimeisthai* – "imitation" or the Latin term *imitatio* – "role playing". In current use, "simulation" is synonymous with the words "imitation" or "mimesis" (imitation, mimicking). In literature, simulation is defined as a problem-solving technique that involves observations of changes taking place in a dynamic model of a system over time (GORDON, 1974) or a numerical technique for conducting experiments on selected mathematical models that describe the behavior of a complex system over long periods of time (FISHMAN, 1981). A similar definition was proposed by LAW and KELTON (1991) who regard simulation as a numerical technique for developing experimental models that supply information about the influence of input parameters on selected output parameters. Simulation can also be defined as an experimental mathematical model of a real-life system written in the form of a computer program, where the experiment is conducted directly on the examined site (MIELCZAREK, 2009). According to PIETROŃ (2013), simulation is the artificial recreation of parameters describing an object, a phenomenon or space that occur in nature but are difficult to analyze. MACIĄG et al. (2013) argue that simulations can be performed in various contexts, including systemic, genetic, philosophical and cultural. All in all, it can be said that simulation is a unique analytical approach that combines various methods and techniques, including computer hardware and software, to describe the functioning of a real-life system.

3. System model and simulation modeling

A system can incorporate a set of interrelated objects characterized by attributes that can also be mutually connected (FISHMAN, 1981). According to JAJUGA et al. (1993), a system is an isolated set of interrelated elements (material or abstract) that are analyzed as a whole, from a given point of view, where the set is characterized by specific parameters that are not attributed to its individual elements. In view of the above, a system can be defined as an internally coordinated set of elements with a unique structure. On the outside, a system constitutes a whole, whereas on the inside, a system constitutes a set that is determined by mutual dependencies between its elements.

A model is a simplified object that is similar to the modeled object and supplies information that is useful for research. It is a representation of the analyzed object, and its form differs from that of the real-life object (FISHMAN 1981, GREGOR et al. 1998). A theoretical model is a hypothetical construction that emulates a real-life situation in a simplified manner. It is a formal representation of a theory, a formal description of an empirical observation, or a combination of the two (FISHMAN 1981). Models are often developed due to the complexity of real-life phenomena. They are used to analyze the possible outcomes of spatial management solutions, to test selected assumptions and choose optimal solutions.

Simulation modeling is a widely used descriptive tool for the analysis of stochastic systems that cannot be modeled easily by other methods. When the goal is to optimize a system or compare several proposed alternative systems for which the response can only be measured using simulation, an optimization algorithm must be used in conjunction with simulation. Such an algorithm needs to explore the solution space for new, improved solutions while exploiting the already visited solutions (HOLLAND, 1975). According to KONDRATOWICZ (1978), simulation modeling involves the identification and formalization of system attributes and the determination of the interactions that take place between those attributes over time with the aim of finding their equivalents in other systems. The end product is a simulation model that cohesively describes the significant attributes of the emulated system and algorithms describing changes in attributes over time.

ATLIOK et al. (2007) define simulation modeling as a common paradigm for analyzing complex systems. In a nutshell, this paradigm creates a simplified representation of the system under study. The paradigm then proceeds to experiment with the system, guided by a prescribed set of goals, such as improved system design, cost-benefit analysis, sensitivity to design parameters, and so on. Experimentation consists of generating system histories and observing system behavior over time, as well as its statistics. Thus, the created representation describes the system structure, while the generated histories describe system behavior. Three classical types of simulation models can be identified based on the description of system variables: continuous, discrete and mixed. In a continuous model, system variables and changes in system state have a continuous character.

Continuous variables can take on the value of any real number. In a discrete model, discrete variables assume the values from a given subset of real numbers, and changes in time occur at strictly defined moments. In mixed models, also known as hybrid models, changes in state are registered as continuous-discrete changes, and time is a continuous, discrete or pseudo-discrete function. The main advantage of simulation models is that there are no limitations regarding the structure and complexity of the analyzed system and that stochastic processes can be taken into account. The above supports the modeling of highly complex systems with a high share of random components (MIELCZAREK, 2009).

The real estate market is a complex system, and selected market elements can be described by cause-and-effect relationships expressed in mathematical form. Various market elements can be simulated, including demand, supply, transactions and transaction prices. This paper focuses on simulations of property transactions, their spatial distribution and transaction prices.

4. Materials and Methods

In this study, simulations were performed on the assumption that data relating to completed transactions, including the spatial distribution (location) and prices of transacted property, can be reflected in future transactions. Low market effectiveness significantly inhibits price simulations but, in many cases, statistical models can be effectively used to explain the changes in transaction prices on local markets. Price models can be applied to simulate transactions, diagnose and evaluate real estate markets and predict the future distribution of traded property, property prices and, consequently, property value.

In the simplest case, a transaction can be simulated with the use of a classical multiple linear regression model.

$$Y = \beta_0 + \sum_{i=1}^k \beta_i X_i + \varepsilon \quad (1)$$

where:

- Y - dependent variable
- β_0 - constant
- β_i - model parameters
- X_i - independent variable
- ε - random component

To predict the price of property, the model should incorporate values of independent variables corresponding to the property attributes for which the theoretical value of the dependent variable is calculated. If such a model is used in simulation, traded property should hypothetically be selected, and its price should be generated in line with probabilistic rules. The price can be simulated based on the modeled theoretical value and the distribution of random components. A higher number of transactions can be simulated with the use of the Monte Carlo approach based on the iteration process presented in Fig. 1.

The probability that a given property will be traded in a given location is determined by numerous factors, including demand, supply and the availability of market information. An analysis of potential supply supports the determination of the property's qualitative and quantitative attributes. The number of objects characterized by a given combination of attributes can be determined based on multidimensional contingency tables with cross-classified property variables (WALESIK 1996). In this case, the probability that a property characterized by a given set of attributes will be traded is directly correlated with the number of such objects. The same approach can be used to determine the frequency of transactions involving property with a given set of attributes. Such analyses can be used in the process of developing simulation models for generating random numbers according to the estimated distribution of frequency or probability.

It can be assumed that the higher the number of transactions in a given neighborhood, the greater the possibility that a subsequent transaction will be completed in the same area. In the real estate market, traded property is not evenly distributed, and its location is largely determined by supply. In view of the observed spatial variation on local markets, it could be assumed that the probability of a subsequent transaction will be determined by the density of clusters representing completed transactions. Cluster density can be evaluated by kernel density estimation that accounts for spatial variation. This approach involves the modeling of a smoothed surface representing density that

is determined by the concentration of points in the surrounding area (Porta et al. 2009). The experiment was performed with the use of a kernel density estimator based on a quartic (biweight) kernel with the following form (Silverman 1986, Kulczycki 2005):

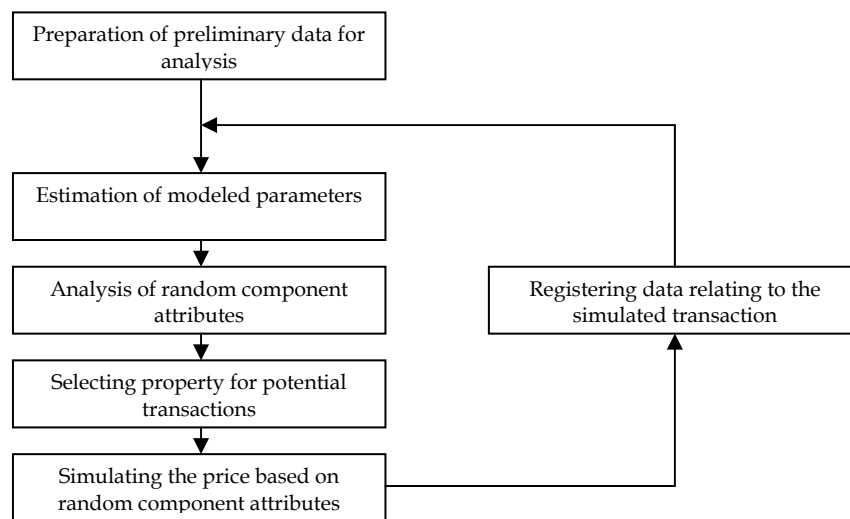


Fig. 1 General diagram of transaction simulations with the use of a multiple regression model. *Source:* own elaboration.

$$\hat{f}(x) = \frac{1}{mh^2} \sum_{i=1}^m 3\pi^{-1} \left(1 - \left(\frac{x}{h} \right)^2 \right)^2 \quad \text{dla} \quad x \leq h \quad (2)$$

and

$$\hat{f}(x) = 0 \quad \text{dla} \quad x > h \quad (3)$$

where:

- m - size of random sample,
- h - positive real number (smoothing parameter),
- x - distance from kernel, i.e., the point at which density is estimated.

The probability that a transaction is completed in a given location will be directly determined by the value of the kernel density estimator.

Combined probability may be difficult to estimate when the above factors, including potential supply, realized demand and location of traded property, are taken into account. The significance of these factors in the process of generating random numbers should be determined. This can only be achieved by conducting many experiments and comparing the results with the actual market behavior. The experiments can be performed on an *a priori* assumption that the analyzed factors have an identical influence on the estimated probability.

Prices can be simulated with the use of a statistical model that, unlike classical regression models, accounts not only for the physical attributes and legal status of property, but also its location. In line with the substitution principle for determining the market value of property, the most probable price is influenced by the prices of similar property. In this case, the compared property should be similar not only in its physical attributes or legal status, but also in terms of its location. In the developed model, simulation effects should be determined by the prices of property situated in direct vicinity. The above conditions are met by spatial autoregressive (SAR) models that rely on autocorrelation or spatial autoregression. In this study, the following spatial lag model (SLM) was used (Anselin 1988):

$$y = \rho W y + X \beta + \varepsilon \quad (4)$$

which can also be expressed in reduced form:

$$y = (I - \rho W)^{-1} X \beta + \xi \quad (5)$$

where X is the matrix of independent variables, β is the vector of coordinates (modeled parameters), ξ is the vector of model errors, Wy is a spatially lagged dependent variable, and ρ is the coefficient of spatial autocorrelation. If a spatial autocorrelation does not exist ($\rho = 0$), the result is a classical linear multiple regression model.

A simulation diagram based on the above assumptions has been presented in Fig. 2.

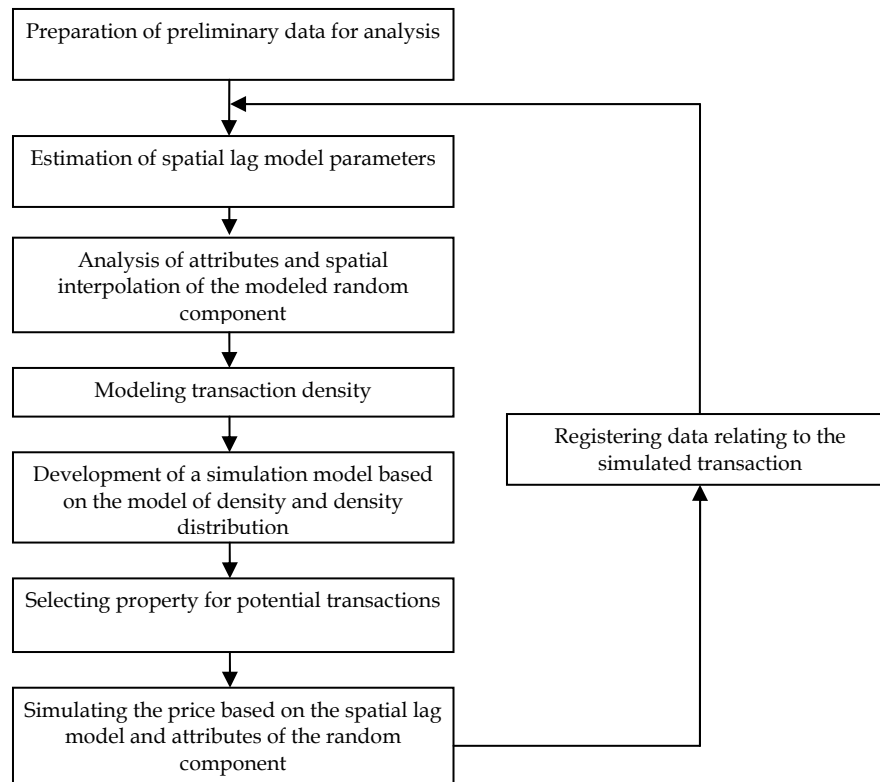


Fig. 2. General diagram of transactions simulated with the use of the spatial lag model and the density distribution model. *Source:* own elaboration.

The above assumptions support simulations of any number of transactions on the local market, which is a relatively laborious process. The adequacy of generated data and its applicability to the local real estate market has to be analyzed. In principle, such evaluations can only be performed for real-life transactions, and their results can be used to verify and adjust the tested assumptions.

4. Results

The experiment was performed with the use of virtual data to simulate transactions on the local real estate market. It was assumed that the analyzed area comprises 1,419 identically shaped and identically sized undeveloped land plots. Every property was described with the use of three attributes (A, B and C) expressed on a binary scale. The set of output data comprised 10 items. The spatial distribution of output data and price data has been presented in Fig. 3.

In the first stage of the experiment, 50 transactions were simulated according to the diagram shown in Fig. 1. It was assumed that the probability that a given property will be traded is identical for all land plots. The property for virtual transactions was selected with the use of a generator of random numbers with uniform distribution. The location of simulated transactions has been shown in Fig. 4.

The price of every property was simulated with the use of a generator of random numbers with normal distribution the average value of which was equal to the modeled theoretical value and standard deviation calculated based on the residuals (random components) from the model. The parameters of the multiple regression model for output data and all data generated by the simulation, including output data, have been presented in Table 1.

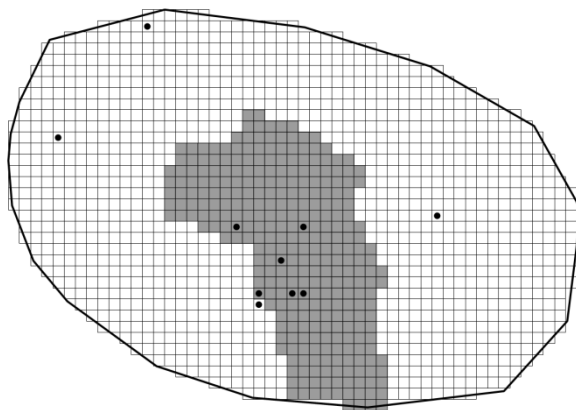
Table 1

Regression model before and after simulations of 50 transactions

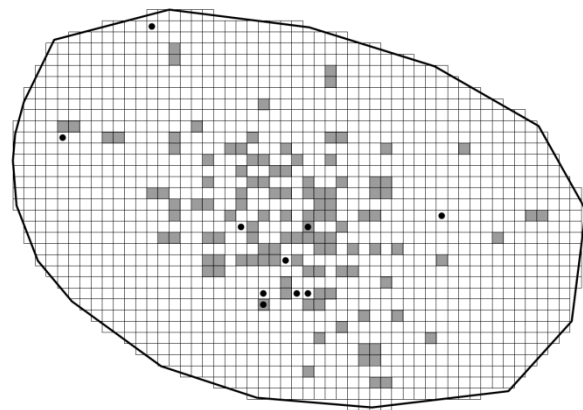
	preliminary data				data after simulations of 50 transactions			
	β	error β	t	P	β	error β	t	p
Constant	91.336	11.216	8.143	0.000	87.893	2.063	42.613	0.000
A	44.892	9.255	4.851	0.003	45.634	3.154	14.468	0.000
B	30.367	11.568	2.625	0.039	33.033	4.742	6.965	0.000
C	27.364	10.085	2.713	0.035	29.008	2.721	10.661	0.000
R ²	0.837				0.884			
Std. error.	14.168				10.297			

Source: own elaboration.

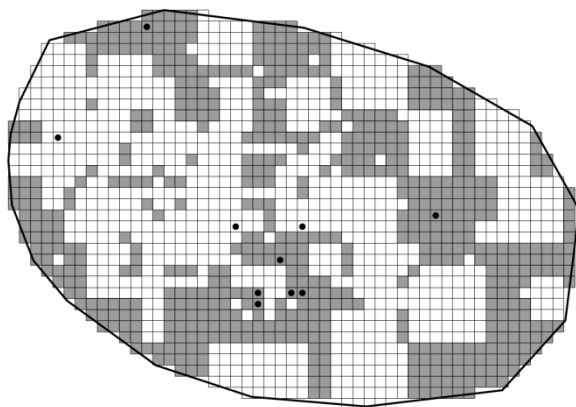
Distribution of attribute A



Distribution of attribute B



Distribution of attribute C



Set of output data

No.	A	B	C	Price 1m ²
1	0	0	1	103.13
2	0	0	0	96.23
3	0	0	1	130.46
4	1	0	0	130.25
5	1	1	0	167.68
6	1	0	1	165.20
7	1	0	1	183.90
8	1	0	1	165.12
9	1	0	1	145.04
10	0	1	1	147.98

Fig. 3. Output data for simulating transactions on the real estate market (the geometric centers of traded land plots are marked with points), Source: own elaboration.

The experiment generated an additional data set, which seemingly improved the precision parameters of the regression model. The ostensible improvement in model parameters results from a much larger data set and the assumed simulation and evaluation criteria. The fact that the resulting model better fits data does not automatically imply that it describes real-life events more reliably. The model merely illustrates one of the many possible scenarios for the development of selected real estate market components in the future.

In the second stage of the experiment, the simulation diagram shown in Fig. 2 was used. It was assumed that the probability that a given property will be traded is determined by the combined density of completed and simulated transactions. Density was determined with the use of the quartic kernel estimator, with the minimum value of the smoothing parameter set at a level ensuring that the probability for every property is greater than zero. The property for virtual transactions was selected

with the use of a generator of random numbers the distribution of which corresponded to the density of the completed transactions. The location of 50 simulated transactions based on the analysis of the density of geometric centers of virtually traded land plots has been presented in Fig. 5.

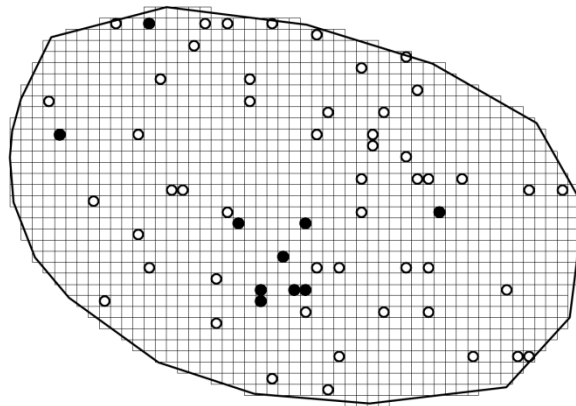


Fig. 4. Spatial distribution of simulated transactions determined with the use of a generator of random numbers with uniform distribution, *Source:* own elaboration.

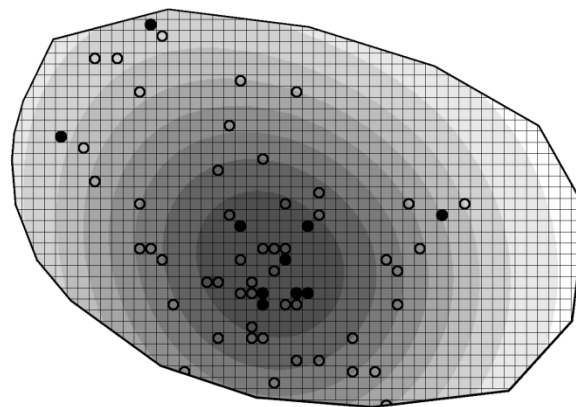


Fig. 5. Spatial distribution of transactions simulated based on density analysis, *Source:* own elaboration.

Prices were simulated with the use of the spatial lag model expressed by formulas (4) and (5), and the matrix representing the spatial structure of data was based on the inverse distance between the geometric centers of land plots. The first simulation relied only on output data, and in successive simulations, input data was used together with the results of previous simulations according to the diagram in Fig. 2. Predictions are difficult to make based on spatial models because the influence of the spatial factor can be estimated only when the spatial lag of the independent variable is known. The structure of spatial models prevents the direct estimation of spatial lags and the relevant matrices. The value of the dependent variable can be determined for a new location without a spatial structure matrix. In a spatial lag model, predictions comprise two components – the trend and the signal (KOPCZEWSKA 2011), where the former is a non-spatial factor expressed as $X\beta$, and the latter is a spatial factor expressed as $\rho W y$. The trend component can be determined directly from the model, whereas the signal component can be estimated by spatial interpolation methods such as kriging. An example of the spatial interpolation of the signal component for output data and data generated by 50 simulated transactions has been shown in Fig. 6.

The simulation of every transaction requires the preparation of output data, semivariogram models and the estimation of unknown value in a location selected based on density analysis. Transaction density has to be determined for every location with the use of the kernel function. The proposed solution is quite time-consuming, but it is likely to produce more satisfactory results than simple random sampling combined with a classical multiple regression model. In the proposed simulation process, individual results are largely dependent of previous results. Therefore, the final simulation result will be sensitive to the initial assumptions. The results generated by the spatial lag model for

output data and a combination of output and simulation data are presented in Table 2.

The model developed with the use of both preliminary data and data generated by simulations is somewhat more effective than the model containing only preliminary data. This does not imply that it simulates real-life events more reliably because the result is burdened with significant uncertainty that generally accompanies predictions. Similarly to simulations based on simple random sampling and classical regression models, it merely illustrates potential trends on the real estate market in the near future.

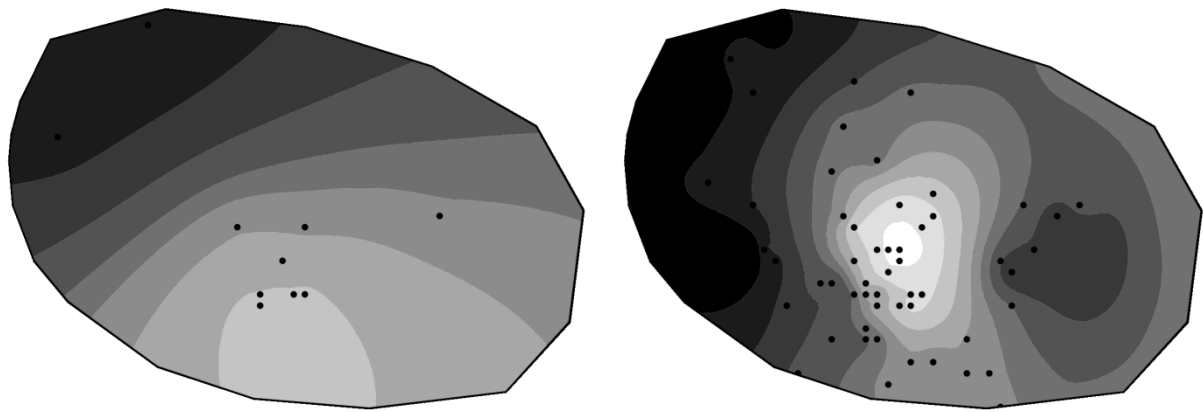


Fig. 6. Spatial interpolation of the signal component in a spatial lag model for output data and data generated by simulations, *Source:* Own elaboration.

Table 2

Spatial lag model for output data and data generated by 50 simulated transactions								
	preliminary data				data after simulations of 50 transactions			
	β	error β	z	$\Pr(> z)$	β	error β	z	$\Pr(> z)$
Constant	134.429	72.454	1.855	0.063	133.091	24.680	5.393	0.000
A	46.005	7.421	6.200	0.000	52.972	2.916	18.163	0.000
B	32.998	9.016	3.660	0.000	37.032	2.978	12.436	0.000
C	29.016	7.867	3.689	0.000	27.541	1.987	13.862	0.000
P	-0.297				-0.366			

Source: Own elaboration.

5. Conclusions

The processes observed on the real estate market are very difficult to describe with the formalized language of mathematical models. They explain only selected phenomena that determine the conclusion of property transactions, characterized by specific attributes and price, in a given location. Market phenomena can be described with the use of methods other than hedonic models that describe the correlations between the price and price-shaping factors, mainly property attributes. Selected phenomena can be described with simulation models that rely on the discussed price models. Simulations of real estate market transactions support the formulation of predictions and market development variants.

The proposed method and the presented results indicate that simulation tools can be effectively deployed in the real estate market, in particular in weakly developed markets where the number of transactions is low but sufficient to build a simple statistical model.

6. References

- ANSELIN L., 1988, *Spatial Econometrics: Methods and Models*, Kluwer Academic Publisher.
- ATLIOK T., MELAMED B., 2007, *Simulation Modeling and Analysis with Arena*, Elsevier.
- GORDON G., 1974, *Symulacja systemów*, WNT, Warszawa.
- GREGOR W., HALUŠKOVÁ M., HROMADA J., KOŠTURIK J., MATUSZEK J., 1998, *Symulation of Manufacturing System*, Wydawnictwo Politechniki Łódzkiej – Filii w Bielsku Białej, Bielsko-Biała.
- FISHMAN G. S., 1981, *Symulacja komputerowa*, Pojęcia i metody, PWE, Warszawa.

- HOLLAND JH., 1975, *Adaptation in Natural and Artificial Systems*, University of Michigan Press: Ann Arbor, MI.
- JAJUGA T, JAJUGA K. WRZOSEK K., WRZOSEK S., 1993, *Elementy teorii systemów i analizy systemowej*, Wyd. AE, Wrocław.
- KONDRATOWICZ L., 1978, *Modelowanie symulacyjne systemów*, Wydawnictwo Naukowowo-techniczne, Warszawa.
- KOPALIŃSKI W., 2006, *Podręczny słownik wyrazów obcych*, Oficyna Wydawnicza Rytm, Warszawa.
- KOPCZEWSKA K., 2011, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDEWU, Wydawnictwa Fachowe, Warszawa.
- KULCZYCKI P., 2005, *Estymatory jądrowe w analizie systemowej*, Wydawnictwa Naukowo-Techniczne, Warszawa.
- LAW A.M., KELTON W. D., 1991, *Simulation modeling and analysis*, McGraw-Hill, Inc.
- MACIĄG A., PIETROŃ R. KUKLA S., 2013, *Prognozowanie i symulacja w przedsiębiorstwie*, Polskie Wydawnictwo Ekonomiczne, Warszawa.
- MIELCZAREK B., 2009, *Modelowanie symulacyjne w zarządzaniu*, Symulacja dyskretna, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław.
- PIETROŃ R., 2013, *Modelowanie symulacyjne*, Wybrane zagadnienia, E-Materiał PWr.
- PORTA S., LATORA V., STRANO E., CARDILLO A., SCELLATO S., IACOVIELLO V., MESSORA R., 2009, *Street centrality vs. commerce and service locations in cities: a Kernel Density Correlation case study in Bologna, Italy*, Urban & Planning B 36, pp. 450-465.
- REMINGTON B.A., DRAKE R.P., RYUTOV D.D., 2006, *Experimental astrophysics with high power lasers and Z pinches*, Reviews of Modern Physics, pp. 755-807
- SILVERMAN B.W., 1986, *Density Estimation for Statistics and Data Analysis*, New York, Chapman and Hall.
- WALESIAK M., 1996, *Metody analizy danych marketingowych*, Wydawnictwo Naukowe PWN, Warszawa.