The 9th International Conference "ENVIRONMENTAL ENGINEERING"
22–23 May 2014, Vilnius, Lithuania
SELECTED PAPERS
eISSN 2029-7092 / eISBN 978-609-457-640-9
Available online at *http://enviro.vgtu.lt*

Section: Sustainable Urban Development

Simulation modeling in a real estate market

Radosław Cellmer, Katarzyna Szczepankowska

Department of Real Estate Management and Regional Development, Faculty of Geodesy and Land Management, University of Warmia and Mazury in Olsztyn, Olsztyn, 10-719, Poland

Abstract

One of the main features of the real estate market is its imperfection, to which belongs, among others, the lack of sufficient data on transactions, often preventing the use of quantitative methods to the analysis. Some solution of the problem of insufficient amount of data can be an expansion of a research area. However, it also means that, the analysis is made on a much diversified data set, so it is difficult to meet the condition of similarity between objects. In the face of the absence of sufficient number of market data, especially in less developed markets, an attempt was made to simulate a transaction that provided at one hand, a necessary data to build models of prices and value, and at the other hand, to evaluate various opportunities and options shaping the local real estate market. In the presented paper a theoretical basis for simulation modeling was done. Also such simulation was presented. It was based on the classical model of linear, multiple regression in an analytical form, and also with the use of Monte Caro simulation. In the study an assumption was made that, the probability of the next transaction will depend on the density of current transactions. In order to determine the density a nuclear estimation was used. It allows to explicitly taking into consideration a spatial resolution.
As a result of the research, simulated transaction data, in a several input assumptions, were obtained, which fill the information gap on the researched fragment of the real estate market.

Keywords: simulation modeling; real estate market; Monte Carlo.

Nomenclature

$\hat{f}(x)$	kernel density estimator
m	size of random sample
h	positive real number (smoothing parameter)
k	symmetrical function relative to zero with a global maximum atthis point
$x-x_i$	randomsamples
w_i, w_j	kriging weights
$Cov\ (s_i,s_j)$	covariance between the value of variables in locations s_i, s_j
s_0	interpolated point
$Z(x_i)$	data values
x_i	locations at which measurements were performed
$N(h)$	number of point pairs $(xi, xi + h)$ separated by distance h
$\gamma(s_i - s_j)$	semivariogram value for sections connecting measurement points s_i and s_j, calculated from the adopted theoretical model
$\gamma(s_0 - s_i)$	semivariogram value for all sections connecting points s_i and s_0.

1. Introduction

The presence of information gaps obstructs the development of direct models illustrating the correlations and relationships on the real estate market. The real estate market is an imperfect system where processes and correlations can be predicted with a certain degree of probability and human factors contribute to the random nature of relationships in that system. The

Corresponding author: Katarzyna Szczepankowska. E-mail address: katarzyna.szczepankowska@uwm.edu.pl
http://dx.doi.org/10.3846/enviro.2014.113

real estate market is imperfect due to difficult access to information and insufficient data, and those problems are frequently encountered by property appraisers in daily practice.

The random nature of the real estate market largely obstructs the development of comprehensive analytical models illustrating market functions. Simulation modeling is a tool that enables decision-makers to analyze situations that are burdened with uncertainty. It is an optimal tool for recreating market processes in an experimental setting, it accounts for irregularities caused by random factors, and it supports the generation of additional information about the real estate market.

The advantages of simulation models and the resulting analytical data have been recognized by many researchers in various fields of science. Simulation models have several unquestioned benefits, including the possibility of analyzing the system through observations of the modeled behaviors and the option of conducting experiments that do not account for human factors and do not intervene in real-life systems. Simulation models can be developed for non-existing, hypothetical systems that are expected in the future. Such models provide researchers with access to a broad experimental knowledge base, iterative calculations can be rapidly repeated, and the structure of the model can be tested and improved numerous times. The main advantage of simulation models is that they can be applied in studies of random phenomena, which is why they play a key role in analyses of real markets that are characterized by random processes and relationships.

The applicability of simulation models for real estate market analysis has been recognized by the authors in earlier studies where virtual data was used to simulate transactions on the local real estate market. The results of experiments performed with the use of virtual data were highly promising, and they laid the groundwork for studies of real-life systems. The present study address the uniqueness of the real estate market, which is characterized by limited availability of reliable information about the prices and terms of property transactions. Simulation models are developed to supply additional data, and they are characterized by an unlimited number of random, dynamic correlations. Price trends and, consequently, the value of real estate are simulated to predict various market scenarios for the nearest future.

2. Random nature of the real estate market

The real estate market is a specific and imperfect field of research with interdisciplinary and systemic character. It is shaped by identifiable processes and correlations, which can often be predicted with given probability, as well as by random processes and relationships (randomness – absence of order or anticipated behavior). Probability is an identifiable measure of a random event, and random elements (components) are associated with:

- absence of homogeneous data,
- non-homogeneous access to data,
- unavailability of comprehensive information for real estate market actors,
- uncertainty of market structures and functions,
- instability of property attributes,
- market actors' emotional approach to transactions, etc. [1–2].

Real estate market data is burdened with uncertainty, which implies that the consequences of decisions will vary in different market scenarios, and the probability that a given scenario will take place is unknown [3]. The above significantly obstructs the development of a comprehensive analytical model of the real estate system. Human factors are largely responsible for the random nature of relations in the real estate market. Simulation models are applied to recreate market functions in an experimental setting, where input data can be repeatedly manipulated and the irregularities resulting from random factors are taken into consideration.

The presence of information gaps obstructs the development of direct models illustrating the correlations and dependencies on the real estate market. The above problem can be solved with the use of simulation tools that generate additional market data. The available data and information gaps were identified, and a simulation model was built to supply additional transaction data. Both deterministic factors and random components were taken into consideration in the developed stochastic model. The proposed method expands the availability of market data and accounts for random factors.

3. Simulation modeling – Monte Carlo approach

The contemporaryapproach to simulation modeling dates back to the World War II when the Monte Carlo method was developed. The method was successfully used to simulate the probabilistic nature of real-life phenomena, and it supported mathematical modeling of real-life processes whose outcomes could not be predicted by analytical methods due to the complexity of those processes. The Monte Carlo approach belongs to the group of stochastic simulation methods where time does not play a key role. The discussed method is also applied in dynamic simulations where changes taking place over time in a system are observed, and the information collected in previous periods is used. In this approach, the Monte Carlo method is a variant of discrete stochastic simulation, and one of its greatest advantages is that it uses a spreadsheet for building simulation models.

According to Powell and Baker [4] in the Monte Carlo method, successive steps are performed after the formulation of a deterministic mathematical model, which is written as a logical formula in a spreadsheet. The constituent elements of simulation models are identified separately as the main stages and detailed stages. The first element (I) involves the determination of the random distribution of selected input variables during which: (a) randomly described variables are selected, and (b) random distribution is characterized. In stage (a), random model parameters that exert the greatest influence on the result variable are identified. The incorporation of successive random variables should be carefully monitored during model development because additional variables increase the variance of output values, and they are not always indispensable in the modeling process. Random distribution is described by identifyingdiscrete and continuous random variables [5]. Input data isanalyzed to formulate a cumulative distribution function for a discrete random variable; alternatively, a theoretical functional formula is developed for every continuous random variable.

Output variables are selected in stage (II). Stage (III) involves a single simulation experiment that can be sub-divided into the following stages: (a) series of random numbers are generated separately for every variable, (b) the resulting random numbers are used to calculate the value of random variables, (c) random variables are inserted into the analytical model, and (d) the process isrepeated n number of times [6–7]. The generation of a separate series of random numbers for every variable (stage a) involves two steps: random numbers with uniform distribution in the (0, 1) interval are generated, and successive random numbers are converted into any random variable distribution or any stochastic distribution. In stage (b), the values of random variables are automatically generated in the spreadsheet, whereas in stage (c), logical formulas are applied to combine input data (deterministic data and data from random distributions) with output data. Stage (d) involves an n number of repetitions, and it determines the accuracy of the final analysis. The number of repetitions is positively correlated with the accuracy of results. At the beginning of stage (d), an n number of random values is selected from the input distribution, they are successively inserted into the model to determine output values.

Stage (IV) is a comprehensive simulation process that involves a k number of experiments. In many cases, a single experiment has to be repeated many times. The higher number of repetitions and experiments, the greater the accuracy of the estimatedarithmeticmean [8].

The last stage (V) involves a detailed analysis of the results. Simulation results are presented with the use of mean values and in the form of histograms illustrating the random distribution of result variables. The accuracy of results is generally evaluated with the use of variance analysis, confidence interval analysis and means quared error analysis [9].

4. Materials and Methods

The location of real estate market transactions was simulated based on a probability distribution defined with kernel density estimators, whereas the simulated price was determined by geostatisticalsimulation and ordinary kriging.

The intensity (density) of a given phenomenon in space is difficult to estimate because such phenomena often constitute points or can be identified only in selected measurement points. Methods similar to the interpolation technique can be applied on the assumption that the density of the analyzed phenomenon (in this case, the number of transactions per unit of area) does not have discrete character. Density can be determined by kernel density estimation that openly accounts for spatial resolution. This approach involves modeling of a smoothed surface that represents density determined by the concentration of points in the surroundingarea [10].

In its basic form, the kernel density estimator is defined by the belowformula [11–12]:

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^{m} K\left(\frac{x - x_i}{h}\right) \quad (1)$$

The probability density function is determined by the distance parameter, and it is smoothed with an increase in distance. In density estimations, every object is replaced with a value calculated according to the probability density function, and function values are added to produce an aggregate areaor a continuous density field [13]. Although the kerneldensityfunctionis often associated with the normal distribution function, it merely acts as its approximation in many applications (ArcGIS, SAGA GIS). A kernel density function that represents the parameters of a normal distribution can be approximated with the use of biweight (quartic) kernels proposed by Silverman [12].

It should be noted that the value of the smoothing parameter considerably affects the quality of the kernel density estimator. Excessively low values produce a high number of local extrema, which could be inconsistent with the characteristic attributes of real-life populations. Excessively high values of parameter h lead to excessive smoothing of the estimator, which masks the specific attributes of the analyzed distribution [14].

Transaction prices were simulated with the use of geostatistical methods that support spatial interpolation and the determination of simulation errors. The values estimated by kriging constitute a weighed, linear combination of regionalized random variables. The following value is the kriging estimator of random function $Z(s_i)$:

$$Z^*(s_0) = \sum_{i=1}^{n} w_i Z(s_i) \quad (2)$$

The estimation error is the difference between the estimated value and the value of the random variable that models the real value:

The kriging method differs from other point estimation methods in that it minimizes error variance. Variance estimated by kriging is equal to:

$$\sigma^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j Cov(s_i, s_j) - 2\sum_{i=1}^{n} w_i Cov(s_i, s_0) + Cov(s_0, s_0) \quad (3)$$

Kriging requires a knowledge of covariance or semivariance values that are determined from a semivariogram. An empirical semivariogram can be computed with the use of the belowformula [15–16–17–18]:

$$\gamma(h) = \frac{1}{2N(h)}\sum_{i=1}^{N(h)}[Z(x_i + h) - Z(x_i)]^2 \quad (4)$$

An empirical semivariogram is approximated with the use of theoretical functions (e.g. a spherical model). Equivalent variance canrely on semivariance values [19–20–21–22], therefore, error variance will take on the following form:

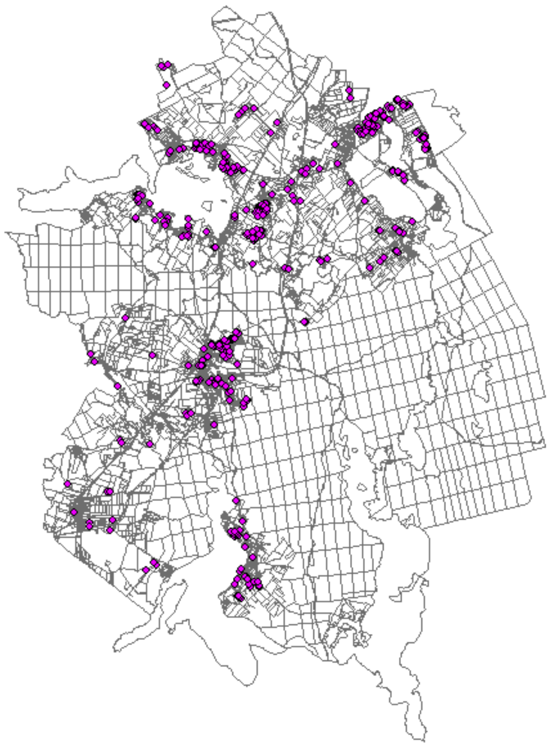
$$\sigma^2 = -\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \gamma(s_i - s_j) + 2\sum_{i=1}^{n} w_i \gamma(s_i - s_0) \quad (5)$$

Kriging is not a simple method in practical applications because the assumption that the analyzed variables are stationary is rarely met. The simulated price was determined with a generator of random numbers with normal distribution, where the expected value was the result of spatial interpolation, and standard deviation was the root of kriging variance.

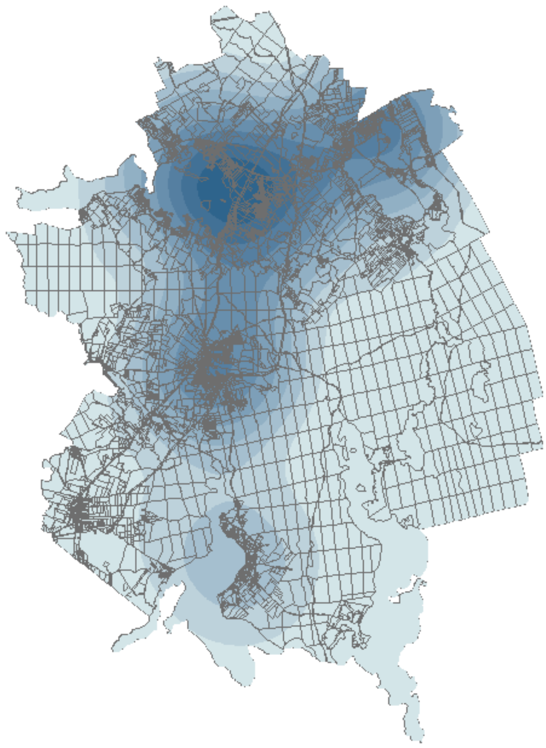
5. Results

The study analyzed transaction data relating to undeveloped property traded on the local real estate market in the municipality of Stawiguda (Olsztyn county, Region of Warmia and Mazury, north-eastern Poland). Transaction data was supplied by the Olsztyn County Office. The analyzed transactions were conducted in 2009–2013.

Areas where property transactions were unlikely or impossible (waterbodies, denseforests) were eliminated from the analyzed site. The resulting set of input data consisted of 560 items. The spatial distribution of input data ispresented in Figure 1. The geometric centers of traded land plots were marked with points, and the density of distribution of the analyzed transactions was determined with the use of the kernel density estimator.



a)



b)

Fig. 1. Spatialdistribution of input data (a) intput data for simulating transactions on the real estate market (the geometriccenters of traded land plotsare market with points), (b)spatialdistribution of density
Source: own elaboration

In the experiment, a single iteration was conducted in line with the followingprocedure:

1. The spatial distribution of transaction density was determined with the use of the kerneldensityfunction.
2. The location of a subsequent transaction was simulated based on the densitydistribution.
3. The price of property was spatiallyinterpolated by kriging.
4. The price of property was simulated based on interpolation and the distribution of interpolationerrors.

A total of 100 transactions were simulated. The location of simulated transactions is shown in Figure 2.

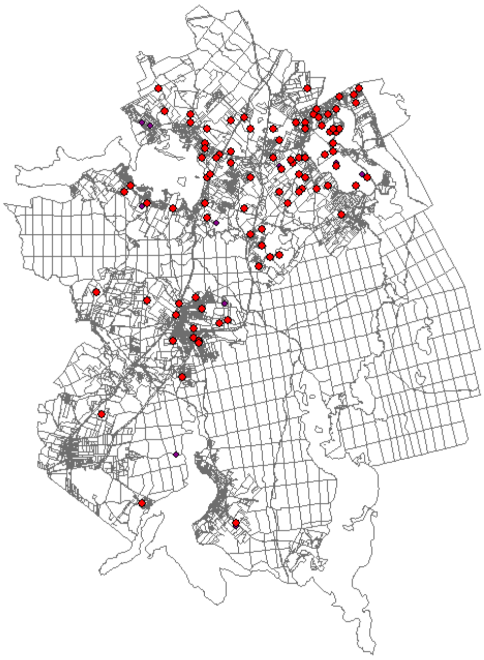


Fig. 2. Spatial distribution of simulated transactions determined with the use of a generator of random numbers with uniform distribution. Source: ownelaboration

In eachiteration, the simulated price was determined by geostatisticalsimulationbased on the previously described procedure. The presented process generated 100 additional data items relating to hypothetical transactions. The results of spatial interpolation based on output data and additionally generated data relating to hypothetical transactions are presented in Figure 3.

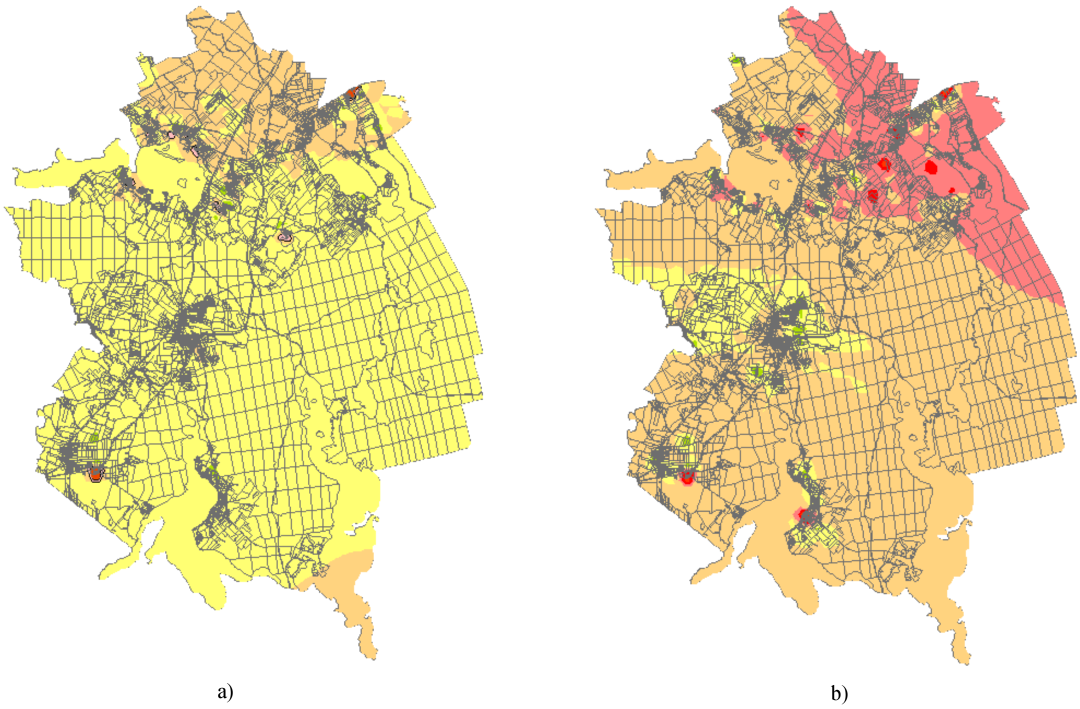


Fig. 3. The results of spatial interpolation (a) based on output data and (b) additionally generated data relating to hypothetical transactions. Source: ownelaboration

It should be notedthat simulation results are determined by the adopted assumption. It was assumed that the location of a simulated transaction would be determined by the density of the existing transactions, but location can also be influenced by other factors that determine its attractiveness. The selection of the kernel estimation function and spatial lag parameters are also important considerations. It was assumed that location is the critical parameter, but in some cases, non-spatial attributes, such as the shape of the land plot or the availability of utility services, can be of decisive importance. If this is the case, universal kriging or regression-kriging models should be applied.

6. Conclusions

Real estate market phenomena are difficult to model due to the imperfect nature of the property market and the human influence. Human factors considerably affect the reliability of mathematical models. In order to effectively represent real-life markets, quantitative models should be developed based on very large amounts of data. Unfortunately, the scarcity of real estate transaction significantly obstructs the development of mathematical models. This problem can be effectively solved through simulation. Transactions on the real estate market can be simulated to formulate various predictions and market development scenarios. 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. The developed method can be used by analysts to address the problem of insufficient data on real estate markets.

References

[1] Radzewicz, A.; Wiśniewski, R. 2011. The uncertainty of real estate market, *Journal of the Polish Real Estate Scientific Society*19(1): 47–59.
[2] Kucharska-Stasiak, E. 2014.Uncertainty of property valuation as a subject of academic research, *Real Estate Management and Valuation* 21(4): 17–25. http://dx.doi.org/10.2478/remav-2013-0033
[3] French, N.; Mallinson, M. 2000. Uncertainty in property valuation. The Nature and Relevance of uncertainty and how it might be measured and reported, *Journal of Property and Finance* 18(1): 13–32. http://dx.doi.org/10.1108/14635780010316636
[4] Powell, S. G.; Baker, K. R. 2004. *The art of modeling with spreadsheets*. John Wiley &Sons, Ltd.
[5] Mielczarek, B. 2009. *Simulation modeling in management, discrete simulation.* Wrocław, Oficyna Wydawnicza Politechniki Wrocławskiej.
[6] Kondratowicz, L. 1978. *Simulation modeling in systems*, Warszawa, Wydawnictwo Naukowowo-techniczne.
[7] Law, A. M.; Kelton, W. D. 1991. *Simulation modeling and analysis.* McGraw-Hill, Inc.
[8] Atliok, T.; Melamed, B. 2007. *Simulation Modeling and Analysis with Arena*. Elsevier.
[9] Gregor, W.; Haluškova,M.; Hromada, J.; Košturiak, J.; Matuszek, J. 1998. *Symulation of Manufacturing System.* Bielsko-Biała, Wydawnictwo Politechniki Łódzkiej – Filii w Bielsku Białej.
[10] 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* 36: 450–465.
[11] Kulczycki, P. 2005. *Kernel density estimators for system analysis.* Warszawa, Wydawnictwa Naukowo-Techniczne.
[12] Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis*. New York, Chapman and Hall. http://dx.doi.org/10.1007/978-1-4899-3324-9
[13] Longley, P.; Goodchild, M.; Maguire, D.; Rhind, D. 2005. *Geographic Information Systems and Science*. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester.
[14] Stępień, B. 2010. *Estimate of uncertainty indicators of environmental noise risk.* doctoral dissertation, Faculty of Mechanical Engineering and Robotics, University of Science and Technology in Krakow.
[15] Matheron, G. 1967. Kriging or Polynomial Interpolation Procedures, *CIMM Transactions* 70: 240–244.
[16] Journel, A. G.; Huijbregts, C. J. 1978. *Mining Geostatistics*. London, UK,Academic Press Inc.
[17] Cressie, N. 1993. *Statistics for Spatial Data, revised edition.* New York,.John Willey & Sons.
[18] Sarma, D. D. 2009. *Geostatistics with Applications in Earth Sciences.* Springer. http://dx.doi.org/10.1007/978-1-4020-9380-7
[19] Bárdossy, A. 1997. *Introduction to Geostatistics.* Institute of Hydraulic Engineering, University of Stuttgart.
[20] Chiles, J.; Delfiner, B. 1999. *Geostatistics. Modelling Spatial Uncertainty*. New York ,John Wiley & Sons Inc.
[21] Waller, L. A.; Gotway, C. A. 2004. *Applied Spatial Statistics for Public Health Data.* New Jersey,Willey and Sons Inc. Pub., Heboken. http://dx.doi.org/10.1002/0471662682
[22] Gotway, C.; Stroup, W. 1997. A generalized linear modelapproach to spatial data analysis and prediction, *Journal of Agricultural, Biological, and Environmental Statistics* 2(2): 157–198. http://dx.doi.org/10.2307/1400401