USING KERNEL DENSITY ESTIMATION FOR MODELLING AND SIMULATING TRANSACTION LOCATION

Katarzyna KOBYLIŃSKA [a], Radosław CELLMER [b], Sabina ŹRÓBEK [b,*], Natalija LEPKOVA [c]

[a] *Department of Real Estate Resources, University of Warmia and Mazury in Olsztyn, Olsztyn, Poland*
[b] *Department of Real Estate Management and Regional Development, University of Warmia and Mazury in Olsztyn, Prawocheńskiego 15/117, 10–720 Olsztyn, Poland*
[c] *Department of Construction Economics and Property Management, Vilnius Gediminas Technical University, Saulėtekio al. 11, LT-10223 Vilnius, Lithuania*

ABSTRACT. Simulation modelling performs a prognostic function through model research and the shaping of the future. Thorough insight into the analysed system and exploring its characteristics for the selection of optimal tools of analysis is an extremely significant process that precedes the stage of the simulation itself. For modelling and transaction simulation, the problem concerning the optimal range of the kernel function used for exploring the spatial activity of a property market should be addressed first. A probability function is the basis for the subsequent phase of research, which allows one to answer the question of whether the transaction density in a given year can be reflected in the transactions of the following year and subsequent years, and whether transaction distribution is correlated, in any way, with the transaction density in the previous year. The final results of the work are maps of the dynamics of transactions on the market and of the simulated transaction density.

KEYWORDS: Real estate market; Simulation modelling; Kernel density estimation; Transaction density; Map of transaction

1. INTRODUCTION

The information stored within spatial information systems may concern not only the physical, but also the economic aspects related to the property market. Properties are key goods for the national economy, but they have also become a fundamental pillar of financial stabilisation. Open, transparent and reliable information on a property market is a high value good. Access to information and its skilful use determine the proper functioning and economic growth of a region, as well as the efficiency or effectiveness of managing commune or district property. Reliable information on the current situation and property market trends existing in social awareness is essential at the decision-making level, *inter alia*, for potential investors.

The property market is one of the most dynamically growing sectors of the economy. Thousands of transactions made every year on the property market bring in millions in capital. Scientific research on tendencies forming on the property market proves that one of the key determinants of price is location (Kiel, Zabel 2008). A relatively large amount of data regarding the property market is of a geospatial nature, relating to buildings (properties) and factors that determine demand and supply. Examples of spatial variables are the transactional prices of properties and attributes which are a function of their location. As one of its features, the location of a property within an area is a constant and invariable factor determined by spatial conditions that only partly explain the level of prices.

Information about the future regarding the place where a transaction on the property market will take place is a desirable good for many stakeholders participating in spatial management in a broad sense. This information is also required by entities making transactions on the property market, market service domains, public institutions, local government institutions, natural persons, organizations connected with the real estate market, and others. Simulation modelling performs a prognostic function through model research and

* Corresponding author. E-mail: zrobek@uwm.edu.pl

the shaping of the future. Thorough insight into an analysed system and exploring its characteristics in order to select the optimal tools of analysis is an extremely important process that precedes the stage of the actual simulation.

In the present work, an assumption was made that the probability of a transaction taking place in a given location in the future depends on the activity of the market in the surrounding areas of the analysed location in previous years. This distribution may be identified with the intensity of the analysed phenomenon in space. Kernel estimation is a useful tool for exploring the above.

For modelling and transaction simulation, the problem concerning the optimal range of the kernel function used for exploring the spatial activity of a property market must be addressed first. The effects of the events (transactions) that took place on future market developments will then be investigated further in the work. After this phase, trends showing changes in the number of transactions on the property market over time will be considered, with maps of their transaction dynamics and simulated density being the final results of the research. The above problems will be the subject of this article.

Research which uses modelling and simulation has been successfully conducted in numerous fields of studies and has inspired real estate researchers to carry out prediction research with the use of such methods on the real estate market. The motivation for the study is an attempt at explaining the possible directions of future processes that will take place on the real estate market. There have only been a few studies that relate to prediction methods and real estate simulation (applying mainly to commercial real estate: Miller 2014; Haß *et al.* 2012).

The works undertaken aim primarily to expand knowledge and verify hitherto obtained information on certain spatial phenomena on the property market. The conducted research will serve as the basis for formulating new theories regarding property market processes and spatial phenomena occurring on this market. Simulation modelling may also serve as a foundation of knowledge on the possible processes that can occur in its structure.

2. RELATED WORK

Taking into account the set of elements entangled in the interrelations which comprise the property market, we should look at its structure in a systemic manner, including internal relations and structural features. The property market system comprises a set of subjects interacting with each other, i.e. a set of subjects, interrelations and relations between their features (Ho, Addae-Dapaah 2014). The subjects of a property market system are the entities participating in the processes occurring in its structure (e.g. a person selling a property) and the objects (e.g. a property – the subject of a sale) connected with them. Each subject has specific features that are attributed to it and characteristic of it (e.g. the size of a plot). Relations are the connections between the subjects of the system and their features. Each relation has its own characteristic features. Due to these relations, the entirety has features that are not displayed by its constituents (Kaklauskas *et al.* 2005; Xie *et al.* 2014). The individual subjects, features and relations have a dynamic character and are affected by specific factors that define their trajectory of movements in the space of conditions. A market system has a heterogeneous structure, considering that, e.g. demand is a combination of needs, purchasing power (income, credit capacity) and price, supply is a combination of a sale offer, good and price, and the market is a combination of demand and supply (Wyman *et al.* 2011; Cellmer 2010). In research regarding the property market as a system, an indirect method of research is used, i.e. analysing substitutes of the real world in the form of models. Direct research on the entirety of phenomena occurring on the property market may be a long-lasting and costly process, or even unachievable due to the complexity of the system. The components of a property market system and its interrelations may, thus, be presented by means of models mapping the real word and facilitating the identification, diagnosis and assessment of the state of reality. In addition, the identification of cause and effect connections enables one to assess and predict changes related to, *inter alia*, prices and property value.

Modelling property value was also explored, *inter alia*, by: (Renaud 1997; Kouwenberg, Zwinkels 2014; Kaklauskas *et al.* 2011; Zoppi *et al.* 2015). As the property market grows, scientists have been building increasingly more complex types of models, including rating models of the property market that facilitate the transparent evaluation of its condition and growth potential (Hurst 2012; Renigier-Biłozor, Wiśniewski 2012; Feige *et al.* 2013).

The modelling of structures of the property market for cognitive purposes has been attempted by many researchers, however only a few of them have attempted to carry out market research with

the use of simulation modelling (Diappi, Bolchi 2008; Mc Breen *et al.* 2010; Bao *et al.* 2012; Filatova 2015; Özbaş *et al.* 2014; Vorel 2014). The great advantage of this specific research approach is the lack of limitations connected with the structure and degree of complexity characterizing the analysed system, as well as the possibility of including stochastic processes, hence allowing one to carry out the modelling of real systems with a high degree of complexity and high percentage of random factors (Von Malottki 2008). This is of key importance for researchers focusing on the behaviours of the property market, where processes and relations may have a random nature. Until now, simulation tests in the field of the property market have been focused on streamlining the economic decisions made (Bao *et al.* 2012), the effectiveness of investing, *inter alia*, the maximization of profits from the rental sector (Mc Breen *et al.* 2010), or the influences of environmental or social and economic features on demand (Vorel 2014). The simulations also aimed to analyse and explore price fluctuations on the property market, as well as analyse the dynamics of a local property market (Özbaş *et al.* 2014; Diappi, Bolchi 2008; Filatova 2015), while the spatial aspects of the modelled system were less addressed. The area of study that needs investigation is the issue of a transaction occurring on the property market in the future (transaction location), which currently constitutes an information gap. Simulation modelling allows one to acquire knowledge on future processes forming on the property market to bridge this gap. In addition, this method has been successfully used to simulate the probabilistic nature of real phenomena, as well as enabling the mathematical modelling of real processes whose results could not be predicted with the use of analytical solutions due to their complexity.

The present property market paradigm is based on the behavioural approach to explaining the behaviours of its participants, which entails both risk and uncertainty. A retrospective analysis of the property market shows that one of its main features is variability, defined, *inter alia*, by price or value variability. The price established on the property market is an effect of the influence of many factors determining its level. In addition, uncertainty also occurs within the property market, which testifies to the stochastic nature of information processes that determine the price of its space. According to (Hanappi 2011; Xiong 2011), uncertainty is a phenomenon that relates to situations of no information on or unfamiliarity with the probability distribution, or the impossibility of determining it. Uncertainty manifests itself in the impact of particular random factors on market actors; it is a characteristic and inseparable element that shapes the property market as an open system (Wiśniewski 2007; Bełej 2011). Uncertainty is defined as all factors and phenomena that are the effect of difficulties connected with identification, technical aspects (data collection), calculation, interpretation and others deemed to be important. It is a long-lasting condition that accompanies the functioning of the system and results from the limited predictability and ambiguity of the behaviours of system subjects, and the unreliability of real and informative processes. Uncertainty naturally exists on the property market. It should be treated as an inherent feature of the market, since it results from its dynamic and process characteristics, the subjective behaviours of market participants, information asymmetry, the specificity of property as a good, etc. (Heilperen 1998). Simulation modelling allows one to establish, in a creative and cognitive way, a new set of concepts and theories that relate to the contemporary understanding of a market as an economic category, as well as one that incorporates spatial aspects.

Both variability and uncertainty greatly hinder the building of a full analytical model of the functioning of the property market, while modelling and simulation approaches enable one to include the elements of uncertainty, randomness, complexity and risk in its modelling process.

The existing models of the property market concern short-term issues on the local market, which is limited by the number of property attributes assuming a constant environment of the modelled market. Furthermore, questions arise that cannot be answered by the existing model of analysis. Scientists search for knowledge regarding tendencies revealing possible price changes on the market in the future. The answer to these insightful questions can be found in the evaluation of market functioning, which often incorporates many concurrent and feedback factors. In such cases, traditional analysis and econometric modelling might be extremely labour intensive, or even unachievable. The simulation modelling of systems provides promising solutions as regards the answers to these questions. It is recommended that the knowledge gap on the property market in regards to simulation modelling of the spatial aspects of the market be bridged by an analysis of transactions on the property market in the future, including the impact of random factors.

3. RESEARCH METHODOLOGY

The dynamic nature of the processes that shape space and the complexity of procedures in spatial management necessitate the application of information on the future nature of the processes taking place in the system, which makes the decision-making process reliable. Data generation processes reveal mechanisms that determine the evolution of a modelled property market system on the basis of features contained in econometric data. Econometric models that quantify cause and effect dependences are built under uncertain conditions, as a complete set of causes determining the state of a modelled variable cannot usually be established (Kouwenberg, Zwinkels 2014). Prediction and simulation are processes implemented with the use of a property market model in order to improve the reliability of the decisions made. Prediction shall be construed as a forecasting process making use of knowledge regarding regularities or tendencies observed in the past (Wiśniewski 2007), thus rational reasoning on unknown events on the basis of known events. Its aim is to provide objective information necessary for taking decisions (Armstrong 2001). Making predictions regarding the property market system is difficult, complex, and has a certain degree of uncertainty. It is difficult, because it concerns properties, i.e. heterogeneous objects. It is complex, because there are a lot of market participants who have an impact on the value of a property, each of whom has a subjective approach to the property. It has a certain degree of uncertainty, as it concerns predicting future behaviours that have some degree of uncertainty (Wiśniewski 2007). The prognostic models that have been used so far in property market analyses allow one to predict some phenomena, while simulation models allow one not only to predict, but also to create phenomena in a virtual dimension (Giulioni, Bucciarelli 2008).

Literature presents a lot of definitions of simulation, which may be defined as, *inter alia*, a problem-solving technique that is based on tracking changes occurring over time in a dynamic system model (Gordon 1974), or a numerical technique used for carrying out experiments on certain types of mathematical models that describe the behaviour of a complex system in the long-term perspective (Fishman 1981). A similar definition has been formulated by (Law, Kelton 1991), defining simulation as a numerical technique that is based on carrying out experiments on a model to gain knowledge on the influence of input parameters on particular output indicators. In its essence, simulation is a specific research approach, which is not considered to be a single approach, but rather a set of methods and techniques for describing and analysing complex systems by referring to the natural human abilities to explore and shape the future (prognostic function), present (rational function) and past (retrospective function) by means of modelling (Von Malottki 2008). Simulation increases the possibilities of quantitative research with the qualitative research of spatial phenomena, enabling the solving of problems with "unclear" definitions, as well as open system or variable structure system research.

Market growth poses new challenges related to the processes of analysing the complexity of the property market system. Progressing complexity brings about specific repercussions in the processes of market analysis, in particular, modelling processes. When building simulation models, the increasing complexity of a system structure as well as of the description of its elements should be taken into account.

A property market is a conceptual system existing in the mental space of its participants, while price is its emanation. In addition to being market facts, property prices are also information that participates in the processes occurring on the property market. System and comprehensive diagnosis of a simulated phenomenon should lead to an understanding of the structure of processes determining property prices as a phenomenon that reflects (represents) the activities of a property market.

Simulations allow one to define different states, which are the fruition of random functions with known distribution patterns. Extending the property market theory to include the main role of information, uncertainty and non-linearity results in the application of proper methods of simulating the essentials.

In practice, theoretical aspects provided by the property market theory and the role of spatial activity should be taken into account in the modelling and simulation of transaction location. Spatial activity of a property market may be measured by, *inter alia*, the number of transactions recorded in the analysed area in a given period of time. One of the main problems when estimating the intensity (or density) of a given phenomenon in space is the fact that these phenomena occur at points, or that their identification is only possible at certain measurement points. The density of a particular phenomenon should thus be construed as the number of points referring to the applied unit of area, e.g. the area within given administrative borders.

A convenient way of estimating density can be the application of kernel density estimation, which accounts for spatial resolution. When using this estimation, the surface reflecting the density resulting from the concentration of points in a given area is modelled (Porta *et al.* 2009; Battaglia *et al.* 2010; Thurstain-Goodwin, Unwin 2000). Kernel Density Estimation has been used to model the density and distribution of a population and central business district activities, to define a 'central area' and network distribution, as well as an estimate of the delineation of a central area. The kernel is a 3-D function that estimates the density of events within its searching radius or bandwidth (*h*), weighting them based on their distance from the estimation point (Gatrell *et al.* 1996).

A kernel probability density estimator is defined by the basic formula of (Silverman 1986; Kulczycki 2005):

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

where: *m* − size of a random sample; *h* − positive real number (smoothing parameter); *K* − function that satisfies the following requirements:

$$\int_{IR^n} K(x)dx = 1; \tag{2}$$

$$K(x) = K(-x) \quad \forall \quad x \in IR^n; \tag{3}$$

$$K(0) \geq K(x). \tag{4}$$

This function should reach global maximum at the zero point and, at the same time, must fulfil the requirement of symmetry with respect to zero. An example of the kernel function for a single-dimensional variable has been presented in Figure 1.

The value of the density function is largely influenced by the parameter connected with dis-

tance. Estimating density is based on reading the values of the density function for each object and adding these values in order to obtain a continuous area representing the density field (Longley *et al.* 2005).
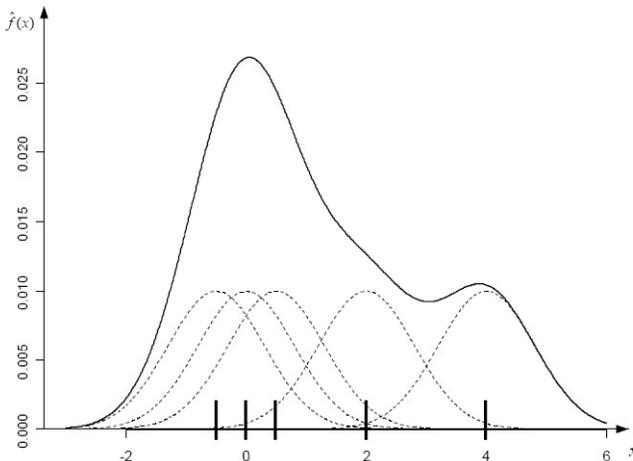
Depending on the aim of the studies, the type of data and the expected results, various forms of the kernel function, e.g. Epanechnikov (also referred to as 'optimal'), uniform, double-weight or normal kernel function are applied. In many applications (e.g. ArcGIS, SAGA GIS), an approximation of normal distribution is assumed. If normal distribution is assumed as the kernel, the function will be expressed as follows:

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right] \tag{5}$$

for *x* > 3, the values of the function cause its curve to be more flattened. Silverman (1986) proposed an approximation of the density function which, to a large degree, reflects the properties of normal distribution, as a double-weight kernel, also referred to as a quartic kernel, in the following form:

$$K(x) = \begin{cases} 3\pi^{-1}\left(1 - x^2\right)^2 & \text{for } x \leq 1 \\ 0 & \text{for } x > 1 \end{cases}, \tag{6}$$

where: *x* − is the distance from the 'kernel', i.e. the point at which density is estimated.

As a result, the kernel density estimator takes on the following form:

$$\hat{f}(x) = \frac{1}{mh^2} \sum_{i=1}^{m} 3\pi^{-1}\left(1 - \left(\frac{x}{h}\right)^2\right)^2 \quad \text{for} \quad x \leq h \tag{7}$$

and

$$\hat{f}(x) = 0 \quad \text{for} \quad x > h. \tag{8}$$

It is easy to see that the smoothing parameter *h* is also the radius indicating the range of the elementary function. The length of this radius as the smoothing parameter has decisive influence on the quality of the kernel estimator. An example of estimating the density distribution of transactions on the real estate market while accounting for various values of the smoothing parameter has been presented in Figure 3.

The value of the smoothing parameter should not be too small in order to avoid an excessive amount of local extremes, which could be in conflict with the nature of the studied phenomenon. On the other hand, too high of a value of the *h* parameter may result in excessive smoothing of the estimator and, at the same time, hide certain specific features of the analysed distribution (Stępień 2010).

Fig. 1. Density function of single-dimensional variable − example

Source: own elaboration based on Stępień (2010).

One important field in which kernel estimation has been applied for many years is system analysis (Kulczycki 2005). However, there are not many works pertaining to its application for the spatial analyses of the real estate market. Kernel estimation was used, among others, to assess the accessibility of retail shopping centres (Porta *et al.* 2009), as well as to analyse accessibility to public health care (Spencer, Angeles 2007). Models representing the density of phenomena can also be the basis for studying correlative relationships. Serving as an example are the results of the analyses of relationships between features of the transportation system and selected socioeconomic indicators pertaining to the local society (Penn, Turner 2003).

Kernel density estimation method is not a perfect approach. The kernel function assumes, for a single point, an even spatial distribution in all directions. In actual spatial systems, the difference in the value of a feature depends not only on the distance but also on the direction. In the case of studies regarding the real estate market, the spatial distribution may be focused, among others, along transportation routes, which may indicate spatial anisotropy. Moreover, a certain difficulty over the course of carrying out analyses may be the presence of spatial barriers in a natural (e.g. the occurrence of rivers, forests or lakes) or anthropogenic form (e.g. transportation routes). The selected methodology therefore assumes a homogenous nature of the space in each place and in all directions. The assumption of spatial heterogeneity would thus require such modification of the kernel function that would result in its shape being adapted to the spatial conditions.

The application of kernel estimation is not limited to assessing density. It can also be used for studying the intensity of a given phenomenon. It is thus possible to apply when developing maps of market activity, the intensity of road traffic or the intensity of building development.

Specific aim

A transaction is a certain event that may influence space within a specific distance. The occurrence of a transaction in a given area does not have an influence at a distance of, e.g. ten kilometres; however, in the case of shorter distances, e.g. several hundred metres, the interactions may be relatively strong. The research problem that needs to be resolved is the selection of a kernel function (its range) for the needs of market data analysis.

In the case of the property market, if the goods are undeveloped land properties, we are dealing with surface objects. In this case, for purposes of the research, we can use simplified elements that represent the analysed properties – plot centroids.

For the selection of a kernel function, the modelling of transaction density with the use of a range of functions with different smoothing parameters was carried out (*h*). Particular density distributions were then compared to the location of a real transaction in the subsequent year of analysis, and an "adjustment gauge" was developed on the basis of the sum of values read from a raster regarding the kernel function with an assigned range. The optimal range of the kernel function was determined on the basis of the greatest sum of values. The process of transaction modelling was performed for the entire year of the analysis. As a result of density analysis, a certain space was received that characterizes the intensity of the analysed phenomenon, while testifying to the activity of the property market. As far as the volume under each space that represents the density function of a single event is the same and equal to unity, the total volume of all density functions is equal to the number of transactions. In order to use this space as a probability density function, transformation was carried out in such a way that the integral covering the entire space under analysis was equal to one. The conducted research provided an answer to the optimal selection of the range of function used for the spatial activity analysis of the property market. The thus determined probability function is the basis for the subsequent phase of the research, which allows one to answer the question of whether the transaction density in a given year can be reflected in transactions in the following year and subsequent years, and whether transaction distribution is correlated, in any way, with the transaction density in the previous year.

It could be assumed that the internal "memory" of a property market is made up of existing phenomena and the behaviours occurring there, while elements of this "memory" are reproduced in the future. On inclusion of market conditions, the following assumptions were accepted and categorized:

- the places where further transactions appear depend on factors, such as, supply, property prices, local preferences of purchasers, subjective, natural and anthropogenic factors and the global situation (tensions, conflicts);
- transaction density indicates the occurrence of particular factors within the analysed area;
- transaction density is a gauge of a bundle of factors that occurred in the past, and may

subsequently occur in the future;

– there is a higher probability of transactions occurring in a place where transactions have already occurred (assuming that these factors are not dynamic);

– external factors change over time; however, for short-term modelling and simulation, it can be assumed that these elements are fixed;

– future transaction density should correspond to existing transaction density while maintaining the principle of *ceteris paribus* (assuming the static nature of market conditions).

After receiving an answer to problems regarding the optimal range of the kernel function for transactional data and the influence of the occurrence of a transaction on future market events, trends regarding the number of transactions over time should be considered. For this purpose, a simple linear regression model was built in each control area. The result of the works is a map of the dynamics of transactions on the property market designed with the use of geostatistical simulation – ordinary kriging.

Kriging is a geostatistical method that provides the best linear unbiased estimators of the value of the analysed regionalized variable. It is often described with the use of the BLUE (best linear unbiased estimator) acronym. Kriging supports the identification of regularities in the spatial distribution of data, and those regularities are used in spatial interpolation. Kriging methods are applied to quantitative data expressed on a quotient or interval scale (Longley *et al.* 2005). The proposed methods were further developed and described in detail by Matheron (1967). The value estimated by means of the kriging method is an important linear combination of regionalised random variables. The estimator of kriging as a random function Z (si) is the following value:

$$Z^*\left(s_0\right) = \sum_{i=1}^{n} w_i Z\left(s_i\right). \qquad (9)$$

An estimation error is thus the difference between an estimated value and a random variable value modelling a real value. The kriging method differs from other point estimation methods mainly due to the minimization of error variance. Estimation variance by the kriging method equals:

$$\sigma^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j Cov(s_i, s_j) + 2\sum_{i=1}^{n} w_i Cov(s_i, s_0) + \\ Cov(s_0, s_0), \qquad (10)$$

where: 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 – point where interpolation is made.

Kriging requires the knowledge of covariance or semivariance values determined from a semivariogram. An empirical semivariogram can be computed with the use of the below formula (Matheron 1967; Journel, Huijbregts 1978; Cressie 1993; Sarma 2009):

$$\tilde{a}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[Z\left(x_i + h\right) - Z\left(x_i\right) \right]^2, \qquad (11)$$

where: $Z(x_i)$ – data values; x_i – locations at which measurements were performed; $N(h)$ – number of point pairs $(x_i, x_i + h)$ separated by the distance h.

An empirical semivariogram is approximated with the use of theoretical functions (e.g. a spherical model). Kriging is not a simple method to be applied in practice, as the assumption that the analysed variables are stationary is rarely met.

The final phase of the works of this article is preparing a map of simulated transaction density, which should be created on the basis of a map of the dynamics of transactions and the spatial activity of the market in the last year of analysis.

4. RESULTS

Research works were carried out on the basis of data regarding purchase and sales transactions of undeveloped land properties that were made on the property market in the Olsztyn Commune (Olsztyn District, Warmia-Masuria Province, north-eastern part of Poland). Olsztyn is the capital of the Warmia-Masuria Province, with a total surface area of 88.33 km^2 and a population count that fluctuates at around 200,000. The real estate market in Olsztyn is balanced, with no particular conditions that influence its functioning. The data regarding property market transactions were provided by the Municipal Council of Olsztyn, where the data are collected as a register of real estate prices and values, which is an integral part of the cadastre. In the register, every real estate is characterized by features such as: a transaction identifier, the date of a transaction, the price of the transaction, the type of real estate, etc. The analysed data pertained to the time period 2004–2012. The number of transactions in the respective years varied from 115 in 2009 to 327 in 2011, with an average of 210 properties marketed per year. The properties that were traded as non-marketable items, e.g. land used for the extension of existing roads, were excluded from the analysis. Instead, properties that provided an objective picture of the market activity were selected. The research area has been marked in Figure 2.
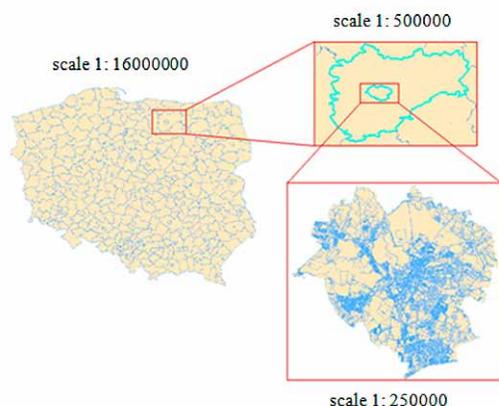
Fig. 2. Research area – Poland, Warmia-Masuria Province, Olsztyn Municipal Commune

Source: own study.

In order to select the range of the kernel function for market data analysis, transaction density modelling was carried out with the use of different ranges of the function. Exemplary analysis for 2004 (range of the kernel function is accordingly: 500 m, 1000 m, 1500 m) has been presented in Figure 3.
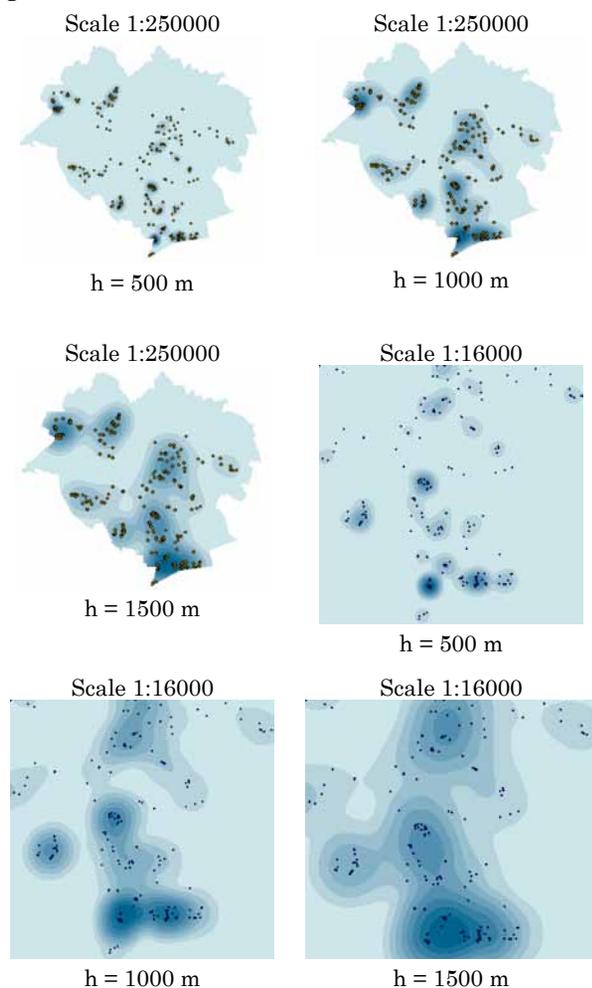
Fig. 3. Analysis of kernel function density distribution for 2004

Source: own study.

The above procedure was repeated for the respective years of the analysis. The received density distributions were then compared with the location of actual transactions in the consecutive years. For this purpose, the values of respective rasters from the previous year were read at points representing transactions in the following year; a fragment of this is illustrated by Figure 4.

Fig. 4. A fragment of the table with values of respective rasters from 2004 as a result of data extraction and placed over the points representing transactions in 2005 (range of kernel function – 500 metres)

Source: own study.

The sum of the raster values may be a quality gauge of reflecting density by the kernel function. In line with the assumptions, the highest value of the sum also informs us about the optimal range of the kernel function. The results of the analysis have been presented in Table 1*.

Table 1. Results of analyses regarding the selection of kernel functions in respective years of analysis

Given years of analysis	Kernel function range (m)	Sum of the raster values
2005	500	**0.006126**
	1000	0.002931
	1500	0.002047
2006	500	**0.004322**
	1000	0.002254
	1500	0.001616
2007	500	**0.002763**
	1000	0.001665
	1500	0.001203
2008	500	**0.000912**
	1000	0.000779
	1500	0.000643
2009	500	**0.001075**
	1000	0.000701
	1500	0.000523
2010	500	**0.001109**
	1000	0.000664
	1500	0.000494
2011	500	**0.002038**
	1000	0.001517
	1500	0.001242
2012	500	**0.002046**
	1000	0.001325
	1500	0.001062

* in line with the assumptions of the procedure, the sum of the values read from the raster regarding the kernel function from 2014 was placed over the points representing respective transactions in 2005, etc.

Source: own study.

The more the density of transactions from the previous year affects transactions in the following year, the higher the sum of the raster, which would attain the highest value if the location of the transactions coincided with each other. The experiments carried out showed that the kernel function range for market data should be relatively small. This is in line with Tobler's (Tobler 1970) principle, which states that interactions occurring between objects in space demonstrate, in many cases, that elements located nearby usually have more similarities than more distant objects. In the research, the range of the kernel function used was 500 m. The specified parameters of the density functions will be used to calculate the probability of transaction occurrence in the future.

The next phase of works aimed to obtain an answer to the question of whether probability measured by transaction density should concern only the previous year, or a few preceding years. Therefore, the area under analysis was covered with evenly distributed control points. The results of estimation in the form of rasters were read at the control points so as to provide the values of density for these points in the respective years.

Correlation coefficients between the values received for the respective years were then calculated (Table 2).

The conducted research shows that all the data has a positive correlation, which means that transaction density in a given year is statistically connected with transaction density from the previous year and from a few preceding years. For example, transactions from 2004 have an impact on the location of transaction occurrence in 2005 and 2006; a relationship between "new" transactions and those which had taken place earlier is also observed in the following years. All correlation coefficients were found to be statistically significant assuming a significance level of less than 0.05. The above results provide sufficient grounds to conclude that there is a relation between the current situation on the property market and future phenomena occurring on this market. Correlations between particular phenomena are not constant and linear, but determined by the specificity of the market and its open and random nature.

Another aim of the analysis concerned searching for a way to include the trend showing the changes in the number of transactions over time. The conducted analyses made it possible to build a model where the explained variable is transaction density in the last year of analysis (2011), while the explanatory variables are the transaction densities in the preceding years. The results of this have been shown in Table 3.

Table 3. Results of estimating a linear model of multiple regression

Transaction density in a given year	Coefficient	Standard error	t	Level p
2004	−0.119	0.032	*−3.704*	0.0002
2005	−0.081	0.031	*−2.636*	0.0084
2006	0.146	0.039	*3.698*	0.0002
2007	0.036	0.035	*1.038*	0.2996
2008	1.280	0.047	*27.456*	0.0000
2009	0.754	0.070	*10.834*	0.0000
2010	−0.046	0.060	*−0.758*	0.4485

Source: own study.

The significance test of the model as an entirety carried out by Fisher-Snedecor statistics showed that dependences are not random at a level of significance lower than 0.001. It was found that the majority of the explanatory variables significantly influence the explained variable. Only for 2007 and 2010 was the significance level greater than 0.05. In the other cases, the analysed variables had very low values.

In the next phase of the research, a simple regression model was built at each of the control points. This resulted in a trend showing whether the number of transactions measured by density increases, decreases, or is maintained at the same

Table 2. Transaction density correlations in respective years.

	2004	2005	2006	2007	2008	2009	2010	2011
2004	1	0.669	0.624	0.45	0.403	0.428	0.481	0.245
2005	0.669	1	0.618	0.432	0.284	0.433	0.428	0.191
2006	0.624	0.618	1	0.536	0.282	0.334	0.413	0.228
2007	0.45	0.432	0.536	1	0.25	0.288	0.395	0.197
2008	0.403	0.284	0.282	0.25	1	0.543	0.445	0.638
2009	0.428	0.433	0.334	0.288	0.543	1	0.638	0.638
2010	0.481	0.428	0.413	0.395	0.445	0.638	1	0.353
2011	*0.245*	*0.191*	*0.228*	*0.197*	*0.638*	*0.638*	*0.353*	1

Source: own study.
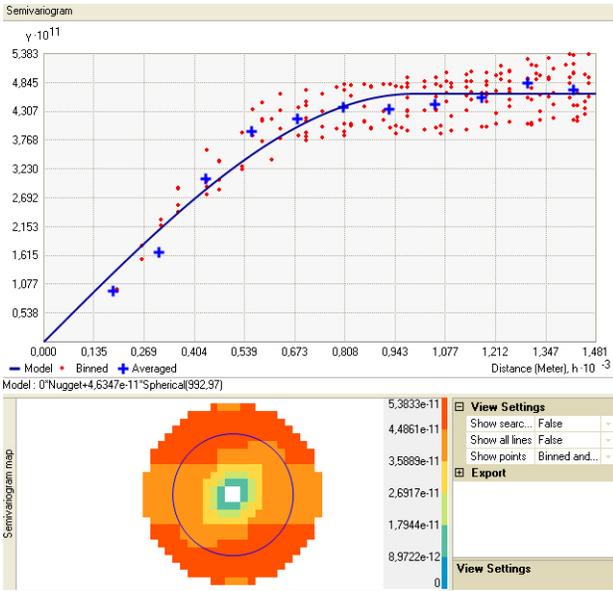
Fig. 5. Semivariogram – spatial structure of
the trend value

Source: own study.

level. The final result is a map of the dynamics of transactions on the market (Fig. 6) created with the use of the ordinary kriging method. The spatial structure of the trend value in the form of a semivariogram has been presented in Figure 5.

The map below shows differentiation in the dynamics of spatial activity on the local property market. The greatest changes concern areas where there are undeveloped lands designated for development (usually single-family detached housing), including the residential areas of: Generałów, Gutkowo, Pieczewo, Nad Jeziorem Długim, and the central part of Redykajny. Low activity areas included city districts where there are no undeveloped lands and characterized by adverse land conditions (lands characterized by poor load bearing conditions), such as the city centre.

Scale 1:128000
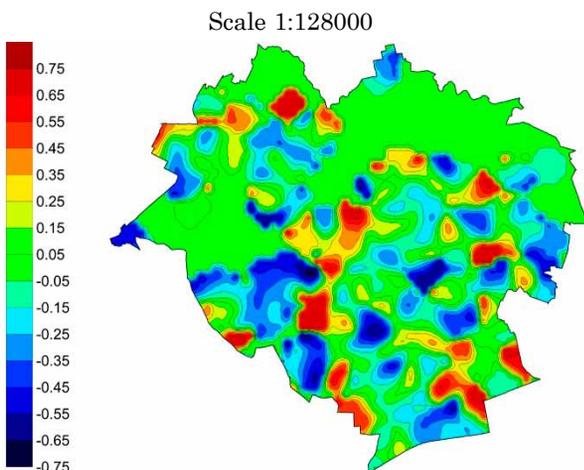
Fig. 6. Map of market transaction number change
dynamics

Source: own study.

Scale 1:128000
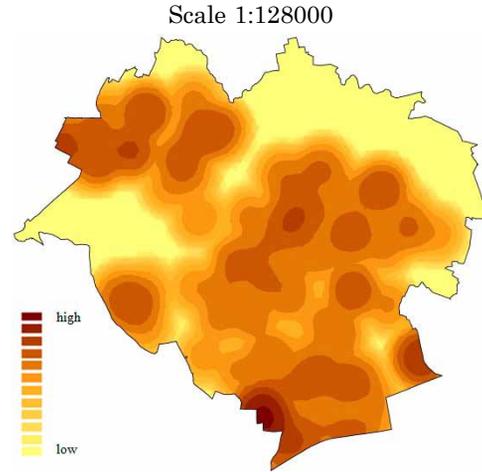
Fig. 7. Map of simulated transaction density
(probability of transaction occurrence)

Source: own study.

Based on the conducted analysis, a map of the simulated transaction density was drawn up for the needs of future research. The map was created on the basis of the dynamics map and the spatial activity of the market in the last year of the analysis (density distribution of the kernel function for $h = 500$ m), which has been presented in the Figure 7.

5. CONCLUSION AND FUTURE WORK

A detailed analysis of a property market in the context of its activity resulted in finding an answer to the questions preoccupying scientists as regards the possible tendencies forming in the structure of this market. The research results show that the range of the selected function is of significant relevance for market spatial activity research with the use of kernel estimation. This confirms the principle that objects located closer to each other in space influence each other to a greater extent. In addition, the density of transactions in a given year is reflected both by the transactions in the next year as well as in subsequent years. The map of the dynamics of transactions on the market will be the basis for further simulations (i.e. it will be a forecast model with known probabilities – errors). Based on the results of the present work, a map of simulated transaction density was created.

Further research will be aimed at carrying out a transaction simulation process as a tool, *inter alia*, for obtaining additional information on the property market. The conducted research on creating property market models will allow simulation experiments that are pioneering on the property

market and have tremendous research potential to be carried out. The awareness of a variety of possibilities provided by simulations and the results of preliminary tests provide grounds to conclude that the planned research will offer significant input which will help bridge the existing gap with respect to the property market. The formulated algorithms should be the basis for the development of GIS tools for purposes of property management.

An alternative approach to the process of estimating the density of transactions in space for purposes of simulating the location of transactions is focusing attention on factors which are attractors of demand. The problem of assessing future market activity can be solved, among others, by the carrying out an inventory of areas in terms of the attractiveness of their location (access to technical and transportation infrastructure, green areas, etc.) additionally accounting for the level of prices and potential supply. Therefore, testing the spatial distribution of supply and demand can serve, along with kernel density estimation, as an alternative basis for simulating the location of transactions.

REFERENCES

Armstrong, J. S. 2001. *Principles of forecasting. A handbook for researcher and practitioners*. Boston: Kluwer Academic Publisher. https://doi.org/10.1007/978–0-306–47630–3

Bao, H.; Chong, A.; Wang, H.; Wang, L.; Huang, Y. 2012. Quantitative decision making in land banking: a Monte Carlo simulation for China's real estate developers, *International Journal of Strategic Property Management* 16(4): 355–369. http://dx.doi.org/10.3846/1648715X.2012.735272

Battaglia, F.; Borruso, G.; Porceddu, A. 2010. Real estate values, urban centrality, economic activities. A GIS analysis on the city of Swindon (UK), in *Computational Science and Its Applications – ICCSA 2010*, 23–26 March 2010, Fukuoka, Japan, *Lecture Notes in Computer Science* 6016: 1–16. http://dx.doi.org/10.1007/978–3-642–12156–2_1

Bełej, M. 2011. Teoria przejść nieciągłych jako metoda modelowania rynku nieruchomości [Theory of discontinous change as a method for modeling real estate market], *Journal of the Polish Real Estate Scientific Society-Studia i Materiały Towarzystwa Naukowego Nieruchomości* 19(1): 59–66.

Cellmer, R. 2010. Analiza przestrzenna dynamiki zmian cen nieruchomości lokalowych z wykorzystaniem regresji ważonej geograficznie [Spatial analysis of dynamic changes in real estate prices premises using geographically weighted regression], *Acta Scientiarum Polonorum, Administratio Locorum* 9(3): 5–14.

Cressie, N. 1993. *Statistics for spatial data*. New York: John Willey & Sons.

Diappi, L.; Bolchi, P. 2008. Smith's rent gap theory and local real estate dynamics: a multi-agent model, *Computers, Environment and Urban Systems* 32(1): 6–18. http://dx.doi.org/10.1016/j.compenvurbsys.2006.11.003

Feige, A.; Mcallister, P.; Wallbaum, H. 2013. Rental price and sustainability ratings: which sustainability criteria are really paying back?, *Construction Management and Economics* 31(4): 322–334. http://dx.doi.org/10.1080/01446193.2013.769686

Filatova, T. 2015. Empirical agent-based land market: integrating adaptive economic behavior in urban land-use models, *Computers, Environment and Urban Systems* 54: 397–413. http://dx.doi.org/10.1016/j.compenvurbsys.2014.06.007

Fishman, G. S. 1981. *Symulacja komputerowa. Pojęcia i metody* [Computer simulation. Terms and methods]. Warszawa: PWE.

Gatrell, A.; Bailey, T.; Diggle, P.; Rowlingson, B. 1996. Spatial point pattern analysis and its application in geographical epidemiology, *Transactions of the Institute of British Geographers* 21(1): 256–274. http://dx.doi.org/10.2307/622936

Giulioni, G.; Bucciarelli, E. 2008. Understanding the price dynamics of a real market using simulations: the Dutch auction of the pescara wholesale fish market, *Lecture Notes in Economics and Mathematical Systems* 614: 15–25. http://dx.doi.org/10.1007/978–3-540–70556–7_2

Gordon, G. 1974. *Symulacja systemów* [Systems simulation]. Warszawa: WNT.

Hanappi, D. 2011. Economic action, fields and uncertainty, *Journal of Economic Issues* 45(4): 785–803. http://dx.doi.org/10.2753/JEI0021–3624450402

Haß, L. H.; Johanning, L.; Rudolph, B.; Schweizer, D. 2012. Open-ended property funds: risk and return profile – diversification benefits and liquidity risks, *International Review of Financial Analysis* 21: 90–107. http://dx.doi.org/10.1016/j.irfa.2011.11.001

Heilperen, S. 1998. *Dynamika i niepewność w modelowaniu ekonomicznym* [Dynamics and uncertainty in economic modeling]. Wrocław: AE.

Ho, K. H. D.; Addae-Dapaah, K. 2014. Real estate market cyclical dynamics: the prime office sectors of Kuala Lumpur, Singapore and Hong Kong, *International Journal of Managerial Finance* 10(2): 241–262. http://dx.doi.org/10.1108/IJMF-10–2013–0108

Hurst, N. 2012. Energy efficiency rating systems for housing: an Australian perspective, *International Journal of Housing Markets and Analysis* 5(4): 361–376. http://dx.doi.org/10.1108/17538271211268501

Journel, A. G.; Huijbregts, C. J. 1978. *Mining geostatistics*. London: Academic Press Inc.

Kaklauskas, A.; Kelpsiene, L.; Zavadskas, E. K.; Bardauskiene, D.; Kaklauskas, G.; Urbonas, M.; Sorakas, V. 2011. Crisis management in construction and real estate: conceptual modelling at the micro-, meso- and macro-levels, *Land Use Policy* 28(1): 280–293. http://dx.doi.org/10.1016/j.landusepol.2010.06.008

Kaklauskas, A.; Zavadskas, E.; Andruskevicius, A. 2005. Cooperative integrated web-based negotiation and

decision support system for real estate, *Lecture Notes in Computer Science* 3675: 235–242. http://dx.doi.org/10.1007/11555223_26

Kiel, K. A.; Zabel, J. E. 2008. Location, location, location: the 3L Approach to house price determination, *Journal of Housing Economics* 17(2): 175–190. http://dx.doi.org/10.1016/j.jhe.2007.12.002

Kouwenberg, R.; Zwinkels, R. 2014. Forecasting the US housing market, *International Journal of Forecasting* 30(3): 415–425. http://dx.doi.org/10.1016/j.ijforecast.2013.12.010

Kulczycki, P. 2005. *Estymatory jądrowe w analizie systemowej* [Nuclear estimators in the systems analysis]. Warszawa: WNT.

Law, A. M.; Kelton, W. D. 1991. *Simulation modeling and analysis*. McGraw-Hill, Inc.

Longley, P.; Goodchild, M.; Maguire, D.; Rhind, D. 2005. *Geographic information systems and science*. Chichester: John Wiley & Sons Ltd.

Matheron, G. 1967. Kriging or polynomial interpolation procedures, *CIMM Transactions* 70: 240–244.

Mc Breen, J.; Goffette-Nagot, F.; Jensen, P. 2010. Information and search on the housing market: an agent-based model, *Progress in Artificial Economics, Lecture Notes in Economics and Mathematical Systems* 645: 153–164. http://dx.doi.org/10.1007/978–3-642–13947–5_13

Miller, N. G. 2014. Workplace trends in office space: implications for future office demand, *Journal of Corporate Real Estate* 16(3): 159–181. http://dx.doi.org/10.1108/JCRE-07–2013–0016

Özbaş, B.; Özgün, O.; Barlas, Y. 2014. Modelling and simulation of the endogenous dynamics of housing market cycles, *Journal of Artificial Societies and Social Simulation* 17(1): 1–20. http://dx.doi.org/10.18564/jasss.2353

Penn, A.; Turner, A. 2003. Space layout affects search efficiency for agents with vision, in *Proceedings of the 4th International Space Syntax Symposium*, 17–19 June 2003, London, United Kingdom.

Porta, S.; Strano, E.; Iacoviello, V.; Messora, R.; Latora, V.; Cardillo, A.; Wang, F.; Scellato, S. 2009. Street centrality and densities of retail and services in Bologna, Italy, *Environment and Planning B-Planning & Design* 36(3): 450–465. http://dx.doi.org/10.1068/b34098

Renaud, B. 1997. The 1985 to 1994 global real estate cycle: an overview, *Journal of Real Estate Literature* 5(1): 13–44. http://dx.doi.org/10.1023/A:1008623304689

Renigier-Biłozor, M.; Wiśniewski, R. 2012. The effectiveness of real estate market versus efficiency of its participants, *European Spatial Research and Policy* 19(1): 95–110. http://dx.doi.org/10.2478/v10105–012–0008–5

Sarma, D. D. 2009. *Geostatistics with applications in earth sciences*. Springer. https://doi.org/10.1007/978–1-4020–9380–7

Silverman, B. W. 1986. *Density estimation for statistics and data analysis*. New York: Chapman and Hall. https://doi.org/10.1007/978–1-4899–3324–9

Spencer, J.; Angeles, G. 2007. Kernel density estimation as a technique for assessing availability of health services in Nicaragua, *Health Services and Outcomes Research Methodology* 7(3–4): 145–157. http://dx.doi.org/10.1007/s10742–007–0022–7

Stępień, B. 2010. *Estymacja niepewności wskaźników zagrożeń hałasowych środowiska* [Uncertainty Estimation of noise environmental hazards indicators]: PhD Thesis. Kraków, AGH University of Science and Technology.

Thurstain-Goodwin, M.; Unwin, D. J. 2000. Defining and delimiting the central areas of towns for statistical modelling using continuous surface representations, *CASA Working Paper Series*, London, UK.

Tobler, R. W. 1970. A computer movie simulating urban growth in the Detroid region, *Economic Geography* 46(2): 234–240. http://dx.doi.org/10.2307/143141

Von Malottki, C. 2008. Geomodelling and geovisualizations in urban planning und real estate industry: the example of office market research, in H. Hagen; M. Hering-Bertram; C. Garth (Eds.). *Proceedings of the Second workshop of the DFG's International Research Training Group on Visualization of Large and Unstructured Data Sets – Applications in Geospatial Planning, Modeling, and Engineering*, Lecture Notes in Informatics (LNI) – Seminars, 9–11 September 2007, Kaiserslautern, Germany. Gesellschaft fur Informatik, 138–147.

Vorel, J. 2014. Residential location choice modelling: a micro-simulation approach, *AUC Geographica* 49(1): 83–97. http://dx.doi.org/10.14712/23361980.2014.8

Wiśniewski, R. 2007. *Multidimensional forecasting of the value of a real estate:* Thesis for the degree of associated professor. Olsztyn: University of Warmia and Mazury in Olsztyn.

Wyman, D.; Seldin, M.; Worzala, E. 2011. A new paradigm for real estate valuation?, *Journal of Property Investment and Finance* 29(4): 341–358. http://dx.doi.org/10.1108/14635781111150286

Xie, X.; Liu, Y.; Hou, J. 2014. An analysis on behaviors of real estate developers and government in sustainable building decision making, *Journal of Industrial Engineering and Management* 7(2): 491–505. http://dx.doi.org/10.3926/jiem.1042

Xiong, Y. 2011. Uncertainty evaluation of the coordinated development of urban human settlement environment and economy in changsha city, *Journal of Geographical Sciences* 21(6): 1123–1137. http://dx.doi.org/10.1007/s11442–011–0905-y

Zoppi, C.; Argiolas, M.; Lai, S. 2015. Factors influencing the value of houses: estimates for the city of Cagliari, Italy, *Land Use Policy* 42: 367–380. http://dx.doi.org/10.1016/j.landusepol.2014.08.012