Urban Residential Land Value Analysis: The Case of Potenza

Benedetto Manganelli, Piergiuseppe Pontrandolfi, Antonello Azzato, and Beniamino Murgante

University of Basilicata, 10, Viale dell'Ateneo Lucano, 85100 Potenza, Italy
beniamino.murgante@unibas.it

Abstract. Urban real estate property values are mainly conditioned by several aspects, which can be summarized in two main classes: intrinsic and extrinsic ones. Intrinsic characters are specific goods while extrinsic features are related to a diversity of goods. Therefore, there is an extremely close correlation between "rigidity location" of property (fixed location) and its value. Possibilities offered by recent developments of statistical techniques, principally Geographically Weighted Regression (GWR), in analyzing housing market have given a new impetus in mass appraisal of urban property. More particularly, Geographically Weighted Regression has been adopted in analyzing housing market, in order to identify homogeneous areas and to define the marginal contribution that a single location (outlined by these areas) gives to the market value of the property. The model has been built on a sample of 280 data, related to the trades of residential real estate units occurred between 2008 and 2010 in the city of Potenza (Basilicata, southern Italy). The results of territory zoning into homogeneous market areas, in addition to the undoubted usefulness in the field of real estate valuations, has useful implications in terms of taxation, programming territorial transformations and checking ongoing or ex post planning decisions.

Keywords: Real estate market, Urban planning Geographically Weighted Regression, Spatial autocorrelation, Spatial statistics.

1 Introduction

The value of urban real estate property is influenced by several factors. Excluding general aspects (macro-economic situation, security profitability, general and real estate taxation, accessibility to credit), all others are classified into two main categories: intrinsic and extrinsic features.

While the former ones are specific goods and are positional (e.g. orientation, front), technological or productive, extrinsic characters are related to a diversity of goods, especially in urban areas, where the value function is indifferent to location variable. In other words, in urban areas. the balance between supply and demand is not conditioned by position, because spatial context, in the opinion of market actors, is uniform in terms of infrastructure (accessibility to public services, accessibility to public transport, the presence of basic commercial services, etc.) and environment (social context, absence of noise, building density, clean air, etc.).

B. Murgante et al. (Eds.): ICCSA 2013, Part IV, LNCS 7974, pp. 304–314, 2013.

Extrinsic features are the determinants of the market price which geographically identify the property. The market area related to a residential property can be defined as the area wherein all units are linked by a substitutability [1].

Therefore, an extremely close correlation between "rigidity location" of property (fixed location) and its value exists.

This study aims to formalize and test a procedure, which adopts Geographically Weighted Regression (GWR) in analyzing housing market, in order to identify homogeneous areas and to define the marginal contribution that a single location (outlined by these areas) gives to the market value of the property.

Traditional statistical methods, such as multiple regression analysis (MRA), are normally adopted in mass appraisal of real estate property. But often they are ineffective, because the property position is not integrated in the analysis [2]. In MRA, position effects cause residues that are spatially correlated, infringing one of the basic requirements of the analysis.

In last decades the development of spatial statistical techniques made possible detailed analysis of real estate data, with geographical reference giving a new impetus in mass appraisal of urban property.

In this work GWR has been adopted to locate zones with homogeneous value within urban areas. More particularly, zones of urban areas with homogeneous environmental, structural and positional variables have been defined, excluding variables connected to specificity of property. An attempt to isolate and highlight the phenomenon of urban rent has been developed.

This approach differentiates this work by elaborations already carried out adopting spatial statistical inference in analyzing real estate market, where the phenomenon can be analyzed in two ways: on the basis of sample data or by means of a random decomposition of space characterized by a predefined grid [3] or deductively defined during the first approximation analysis [4].

2 An Overview of Geographically Weighted Regression

Geographically Weighted Regression (GWR) [5] [6] [7] [8] [9] [10] is a method which allows to analyze how a phenomenon spatially changes within a particularly place. Starting from Tobler [11] first law of geography "Everything is related to everything else, but near things are more related than distant things", GWR can be considered as a spatial extension of multiple linear regression. GWR is not limited to global parameters, but it considers also local parameters. Also, the mathematical formulation is very similar to the typical regression analysis (equations 1, 2).

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_m x_{mi} + \varepsilon_i \quad \text{with } i = 1 \ldots n \tag{1}$$

Where:

y_i = Dependent variable

x_i = Independent (also the term Explanatory is adopted) variables

β_0 = Coefficients (sometimes the term Parameters is used) expressing the relationship between dependent and independent variables.

ε_i = Residuals, i.e. the part of dependent variable not explained in the model

$$y_i = \beta_0(u_i,v_i) + \beta_1 (u_i,v_i)\, x_{1i} + \beta_2 (u_i,v_i)\, x_{2i} + \ldots + \beta_m (u_i,v_i)\, x_{mi} + \varepsilon_i \qquad (2)$$

In Geographically Weighted Regression the term (u_i,v_i) is also considered, which represents coordinates of point i in the space.

It is possible to have positive or negative relationships between dependent and independent variables: according to the kind of relationship, a sign (+/-) is associated to the coefficients.

In order to model in the best way the phenomenon to be investigated it is fundamental to define all factors which may influence the analyses. The central point is to find the main variables in phenomenon modeling, defining the dependent variable and identifying the possible independent variables. It is also important, before analyzing data with GWR, to test with Ordinary Least Squares the possible independent variables to adopt.

Two main measures of Ordinary Least Squares are useful in understanding if the variables adopted in the analysis are meaningful: r^2 or adjusted r^2 and Akaike. r^2 results are generally included between 0 and 1. A better predictive performance has been highlighted by values close to 1. Akaike Information Criterion (AIC) [12] [13] has not an absolute scale of measure, but it is useful in comparing two models, with the same dependent variable, in order to assess which of them fits better the phenomenon. Smaller values of the AIC indicate a better simulation, if the difference is not big, less than 3, two models can be considered equivalent.

Another important check in model performance concerns Residuals. It is fundamental to analyse that spatial dependence does not occur in residuals, verifying a random spatial distribution. Residuals have to be analyzed by Moran Index I. Moran Index I [14] is a global measure of spatial autocorrelation and its values can be included between -1 and 1. If Moran Index I is close to zero data are randomly distributed, if the term is higher than zero, autocorrelation is positive, otherwise it is negative.

Regression coefficients are estimated using nearby feature values. Consequently, main parameters are kernel and bandwidth which provide a definition of nearby.

There are two kinds of kernel, fixed and adaptive: the first one defines nearby according to determined fixed distance band; while adaptive kind defines nearby according to determined number of neighbours.

Fixed kernel is adopted if observation points are regularly located, otherwise, if observation points are clustered, adaptive kernel is more suitable.

Bandwidth controls the size of kernel and can be defined in three ways: directly by the analyst (it is possible to directly define distance or neighbours number), by means of AICc method, which minimises Akaike Information Criterion (AIC), or by using CV, which minimises the CrossValidation score.

3 The Model Structure

The model is built on a sample of 280 data, related to the trades of residential real estate units occurred between 2008 and 2010 in the city of Potenza (Basilicata, southern Italy).

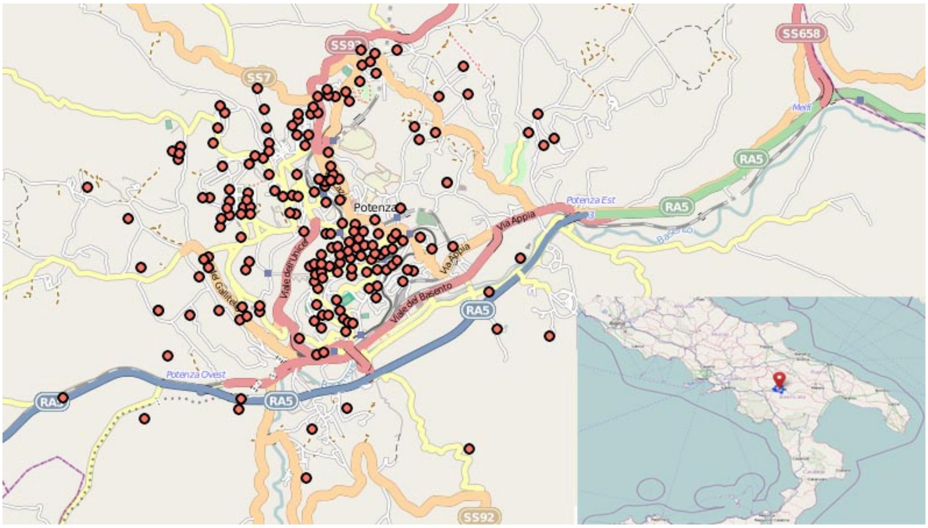

Fig. 1. Study area and location of 280 sample data in Potenza municipality

The 280 data used in this study were selected from a database of approximately 650 prices found in documents registered at "Agenzia delle Entrate", a branch of the Italian Ministry of Economy and Finance, which includes the Territorial Agency. The Agency provided, for each transaction, trading act details (date and price), cadastral data of property, cadastral area and details about area (i.e. distinguishing residential use and different types of outbuildings).

The opportunity of having a so large database in Italy, is the consequence of law n. 266 of 2005 (2006 Budget Law) which produced a more transparent market in residential property trading between privates. This changed perspective could also allow the repetition in time of the procedure developed in this paper for the analysis of housing market dynamics in order to verify the efficiency of planning decisions (e.g. urban regeneration programs).

The static nature of buildings contrasts with the relative mobility determined by socio-economic context change or the variation of city dimensions. To properties with known price, a measure of a series of independent variables, useful to describe qualitative and quantitative features of flats and buildings, has been associated.

First results and elaborations of statistical tests suggested to exclude some variables from the model. Other tests have suggested the aggregation of some parameters. Ancillary dwelling units have been summed to residential areas adopting weighting coefficients that express an ordinary commercial relationship. Even the period of construction has been aggregated to last year restructuring. Other variables included in the function of the price were outdoor common areas, parking areas and presence of elevators.

4 Results

The five variables illustrated in the previous chapter have been tested using Ordinary Least Squares in order to understand in which measure they are reliable.

More particularly r^2 was 0.48 and Akaike Information Criterion (AIC) was 7068. Ordinary Least Squares results were useful for residual spatial distribution.

A first summary assessment can be made looking at location standardized residual values (figure 2) and it is noticeable that spatial distribution of residuals is completely random.

Fig. 2. Spatial distribution of standardized residual values

More particularly, standardized values of residuals, calculated by means of Ordinary Least Squares, have been used as input data in calculating spatial autocorrelation, in order to understand if residuals were autocorrelated or not.

Spatial autocorrelation has been calculated adopting Moran scatter plot and considering standardized variables of residuals as abscissa and spatial weighted standardised variable of residuals as ordinate. In the graph, Moran Index corresponds to direction coefficient of linear regression, which represents the scatter plot. Positive autocorrelation corresponds to spatial clusters in upper right and lower left quadrants. Lower right and upper left quadrants can be classified as spatial outliers [15] [16]. Figure 3 shows that the slope of Moran Index is equal to zero coinciding with abscissas axis.

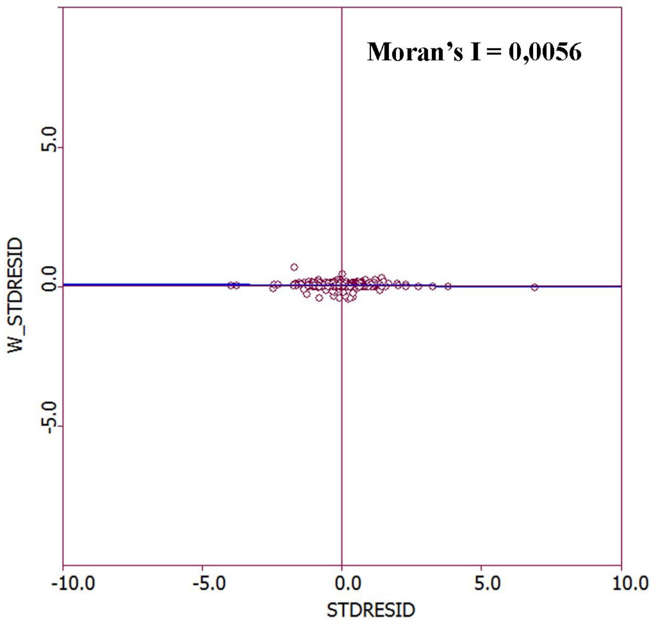

Fig. 3. Moran scatter plot of residuals standardized variable

As expected GWR results are better than those achieved with OLS: namely, r^2 was 0.64 and Akaike Information Criterion (AIC) was 7061.

From GWR, many different price functions ensue for each geographical location of initial data, obtaining also different coefficients of explanatory variables functions. These coefficients represent the implicit marginal prices, or, in other words, the marginal contribution that each selected variable provides to price explanation.

Including only intrinsic features in the function, allows us to say that values of the coefficient related to the area reflect only effects of localization.

It may have occurred that residential units with identical building features have been traded in different parts of the city. In this condition the difference relative to the equilibrium price is an expression of urban rent. In particular, it is an expression of rent rate defined as differential. Differential rent depends on land position occupied in urban aggregation and it is distinguished from absolute rent, more dynamic, which is mainly a function of resource scarcity compared to demand. Absolute rent is determined by the planning choice or, simply, by the presumption of market actors about the future planned development. Urban land rent is a value that belongs to land owner and not to entrepreneur investing his capital in the transformation process. It is therefore a distinctive feature of the area. According to this consideration, in this work the delimitation of homogeneous urban market areas is attributed to the exiting difference in land rent. Values ranges adopted in identifying and classifying different homogeneous urban areas (figures 4 and 5) have been chosen using the guidelines provided by the Decree of the President of Italian Republic (D.P.R. 138/98) "Rules for the review of census areas and valuation prices executing Law 662/96"

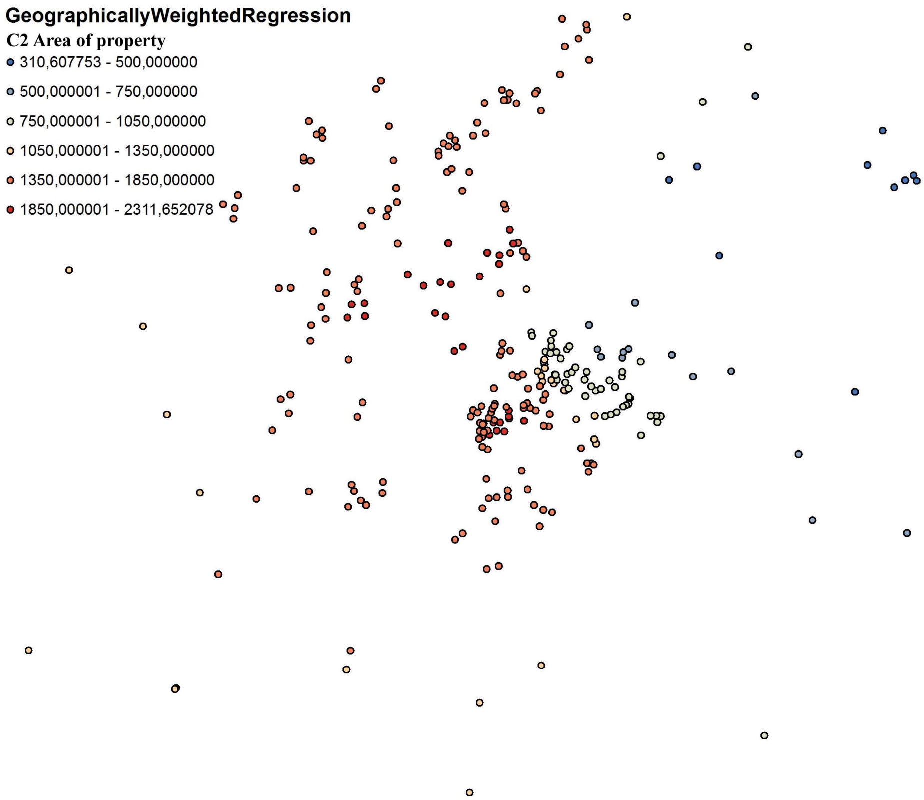

Fig. 4. Classification of area variable coefficient of 280 sample data according to D.P.R.138/98

Known the minimum and the maximum value of marginal implicit price of the area variable (figures 4 and 5), the corresponding interval has been classified according to the following rules:

- within each class, the ratio between maximum and minimum market values (square meter) is not higher than two;
- percent deviation between mean values (square meter) of each class of property units in two adjacent areas is not less than 30%.

The physical limit of market homogeneous areas, corresponding to previous classification, has been defined overlapping data points to cadastral map, zoning maps of Land use planning and Chorography.

Limits of the areas have been achieved reproducing main boundaries of these maps, including in each area data points corresponding to the previously identified six classes.

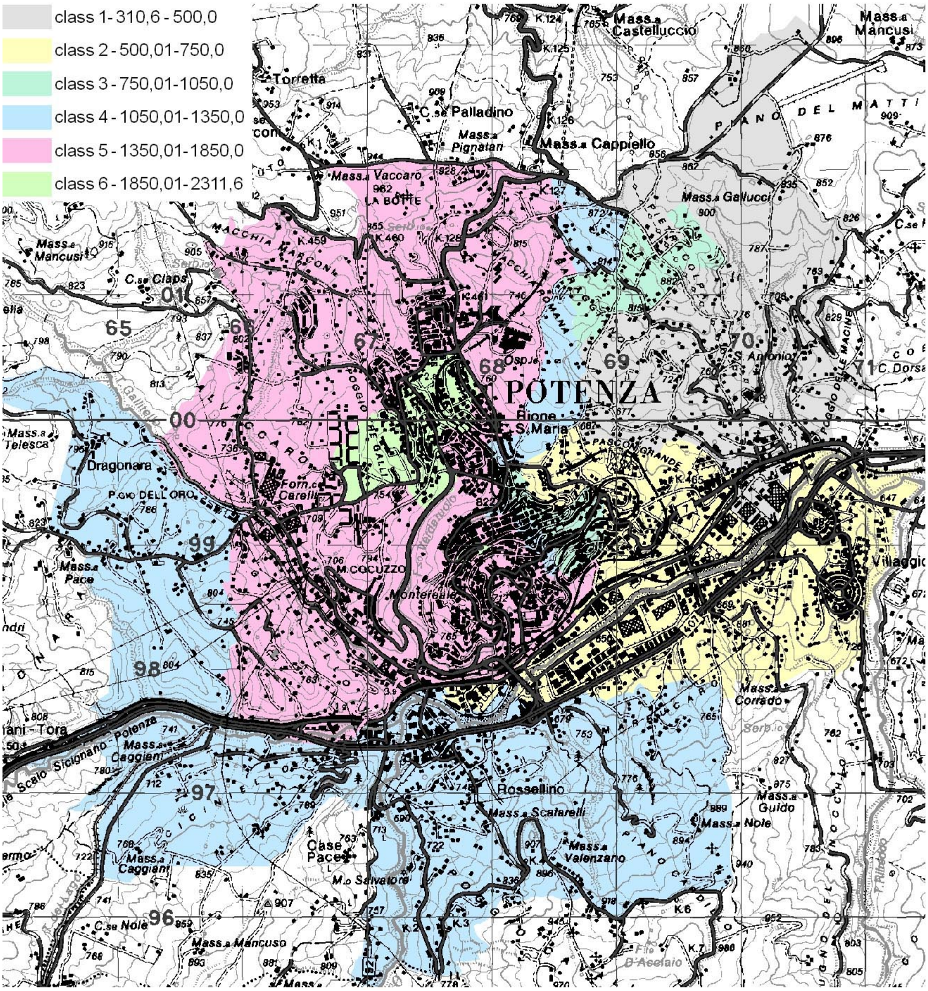

Fig. 5. Six classes of homogeneous market areas in Potenza municipality

Results of this processing have been verified running a new GWR elaboration, considering other variables related to building location. Such parameters have been defined on the basis of hypotheses formulated in a deductive way, to identify market homogeneous areas. Proximity of a property to urban elements, such as infrastructures, facilities, services or natural elements with environmental value, leads to a possible value increase (e.g. proximity to park, health care facility, railway station, university campus, etc.) or, in some cases, to a probable depreciation (e.g. proximity to road junctions, environmental pollution sources or noise generators).

The measure of these real estate advantages (or disadvantages) generated by the presence in the proximity of a natural or anthropic element is measured by geographical distance.

However, when market assigns an advantage to the property in relation to the possibility of direct use of an urban element, walking (travel) time defines the proximity to this element. The walking (travel) time is obviously closely related to roads, and to the presence of anthropic or natural barriers (e.g. a railway, a cliff, a river).

Consequently, all resources and possible sources of differentials value in property residential use have been located in a GIS platform.

These elements constitute the centre of isochrones that are traced starting from several geographic information: 1) the road network, 2) the presence of barriers in connections between urban areas; 3) the presence of interruptions to these barriers (e.g. , pedestrian paths, elevators, escalators, etc.); 4) pedestrian travel times.

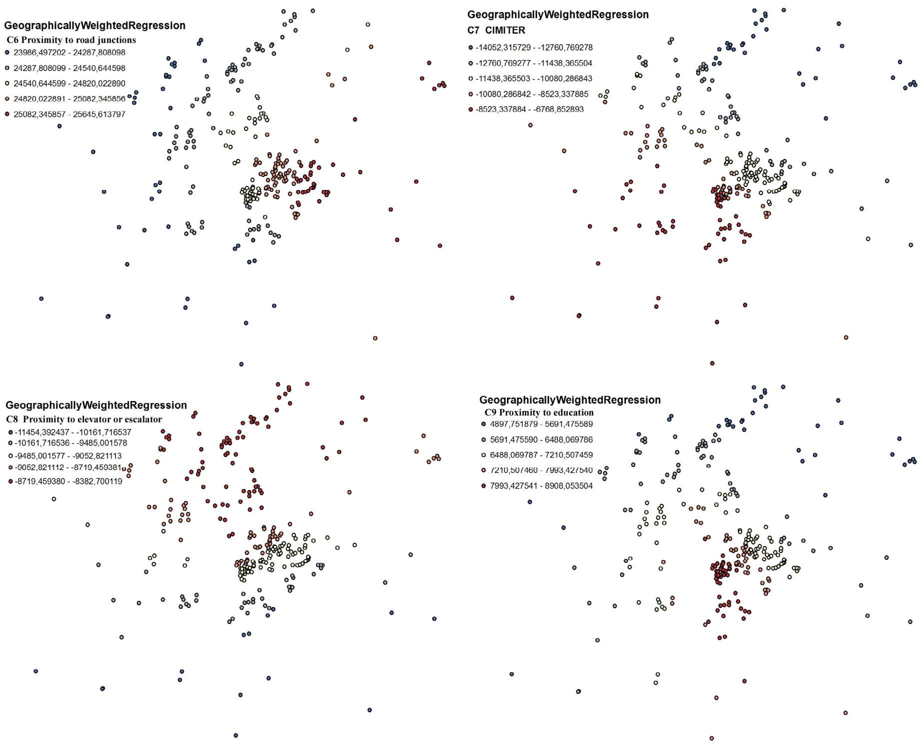

Fig. 6. GWR coefficients of new variables

Time ranges have been defined and isochrones have been calculated according to the different travel times needed to reach the previously defined element. The weighted overlap of all isochrones for the same element category allowed us to assign a rate for each extrinsic feature reference to each property.

Also for this elaboration, preliminary statistical tests led to the selection of the most significant variables and the final calibration of the model.

The new variables included were: proximity to equipments for education, proximity to elevators or escalators, proximity to road junctions, proximity to cemeteries.

The new elaboration provided an implicit verification of the previously achieved zoning. In fact, it reduces variability range of implicit marginal price to a percentage, measured on the average value of the same interval, of approximately 8% (min. €/m^2 1,286, max €/m^2 1,400).

The introduction of the new variables actually made negligible differences in marginal prices values of the area. The new variables can almost completely explain the phenomenon of differential rent.

Coefficients of these variables help to understand their relative weight. It is obvious that the phenomenon of urban differential rent does not depend only on them, but can be a valid approximation.

5 Conclusions

Spatial heterogeneity is a factor that makes fallible an already complex process of real estate assessment. The development of statistical techniques, such as Geographically Weighted Regression has given a new impetus in mass appraisal of urban property and in analyzing real estate market.

The implemented procedure adopts GWR in order to analyze the value of urban land for residential use and to define market homogeneous areas. The subdivision of the urban territory in homogeneous market areas is useful both for predictive purposes and interpretation of actual mechanism of making real estate values.

Territory zoning in homogeneous areas can be a good support in planning process and it can also be conducted with both compensatory tools and transfer of development rights [17] for fiscal purposes and investment location choices. The model presented in this paper, if applied repeatedly over time can highlight dynamics of city development. It also allows to balance possible previous wrong choices or to contrast undesirable imbalances in urban fabric [18].

The results of territory zoning into homogeneous market areas, in addition to the undoubted usefulness in the field of real estate valuations, has useful implications in terms of taxation, programming territorial transformations and checking ongoing or ex post planning decisions [19].

Acknowledgements. Authors are grateful to "Agenzia delle Entrate" agency of Potenza, a branch of the Italian Ministry of Economy and Finance for providing data analyzed in this paper.

References

1. The Institute for Urban Land Use and Housing Studies, Columbia University, Housing Market Analysis: A Study of Theory and Methods. Housing and Home Finance Agency, Washington, DC (1993)

2. Gaoa, X., Asamib, Y., Chungc, C.F.: An empirical evaluation of spatial regression models. Computers & Geosciences 32, 1040–1051 (2006); Toffler A.: Future Shock. Random House Publishing Group, New York (1970)
3. Kulczycki, M., Ligas, M.: Spatial Statistics For Real Estate Data, Strategic Integration of Surveying Services. FIG Working Week, Hong Kong SAR, China, May 13–17 (2007)
4. Manganelli, B., Murgante, B.: Spatial Analysis and Statistics for Zoning of Urban Areas. World Academy of Science, Engineering and Technology 71, 783–788 (2012) ISSN: 2010-376X
5. Casetti, E.: Generating models by the expansion method: applications to geographic research. Geographical Analysis 4, 81–91 (1972)
6. Jones, J.P., Casetti, E.: Applications of the expansionsion method. Routledge, London (1992)
7. Fotheringham, A.S., Brunsdon, C., Charlton, M.: The geography of parameter space: an investigation of spatial non-stationarity. International Journal of Geographical Information Systems 10, 605–627 (1996)
8. Fotheringham, A.S., Charlton, M., Brunsdon, C.: Two techniques for exploring non-stationarity in geographical data. Geographical Systems 4, 59–82 (1997)
9. Fotheringham, A.S., Brunsdon, C., Charlton, M.: Geographically Weighted Regression: the analysis of spatially varying relationships. Wiley, Chichester (2002)
10. Brunsdon, C., Fotheringham, A.S., Charlton, M.: Geographically weighted regression: a method for exploring spatial non-stationarity. Geographical Analysis 28(4), 281–298 (1996)
11. Tobler, W.: A computer movie simulating urban growth in the Detroit region. Economic Geography 46(2), 234–240 (1970)
12. Hurvich, C.M., Simonoff, J.S., Tsai, C.-L.: Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. Journal of Royal Statistical Society, Series B 60, 271–293 (1998)
13. Akaike, H.: Information theory and an extension of the maximum likelihood principle. In: Petrov, B., Csaki, F. (eds.) 2nd Symposium on Information Theory, Budapest, Akadé miai Kiadó, pp. 267–281 (1973)
14. Moran, P.: The interpretation of statistical maps. Journal of the Royal Statistical Society (10) (1948)
15. Hsieh, B.M.: A study on Spatial Dependence of Housing Prices and Housing Submarkets in Tainan Metropolis, Taiwan, Territorio Italia – Land Administration. Cadastre, Real Estate 2(1), 9–22 (2012)
16. Murgante, B., Rotondo, F.: A Geostatistical Approach to Measure Shrinking Cities: The Case of Taranto. In: Montrone, S., Perchinunno, P. (eds.) Statistical Methods for Spatial Planning and Monitoring. Contributions to Statistics, vol. 158, pp. 119–142. Springer, Berlin (2012), doi:10.1007/978-88-470-2751-0_6
17. Manganelli, B.: Stime e valutazioni economiche nella perequazione urbanistica. In: Las Casas, G., Pontrandolfi, P., Murgante, B. (eds.) Informatica e Pianificazione Urbana e Territoriale, vol. 3, pp. 173–188. Libria, Melfi (2010) ISBN: 9788896067475
18. Murgante, B., Las Casas, G., Danese, M.: Analyzing Neighbourhoods Suitable for Urban Renewal Programs with Autocorrelation Techniques. In: Burian, J. (ed.) Advances in Spatial Planning, pp. 165–178. InTech — Open Access (2012) ISBN: 978-953-51-0377-6, doi:10.5772/33747
19. Murgante, B., Danese, M.: Urban versus Rural: the decrease of agricultural areas and the development of urban zones analyzed with spatial statistics. International Journal of Agricultural and Environmental Information Systems (IJAEIS) 2(2), 16–28 (2011) ISSN 1947-3192, doi:10.4018/jaeis.2011070102