INTERNATIONAL JOURNAL OF REAL ESTATE AND LAND PLANNING

VOL.1 (2018)

eISSN 2623-4807

Available online at https://ejournals.lib.auth.gr/reland

Spatial analysis of housing prices in the Athens Region, Greece

Polixeni Iliopoulou [a], Panagiotis Stratakis [b*],

[a]University of West Attica, Ag. Spyridona, Egaleo, 122 43, Greece

[b] National Technical University of Athens, 42 Patision St., Athens, 106 82, Greece

Abstract

In this paper the housing prices in the Greater Athens region are analyzed employing data for the structural and locational characteristics of dwellings. The sample is derived from the total housing supply in 2017 and includes several thousand dwellings for sale available by online real estate agencies. However only houses for which the approximate location, in terms of geographic coordinates, is available are included in the sample.

At first a description of the houses is provided, in terms of their structural characteristics, such as type of the dwelling, size, floor, number of bedrooms, parking etc. In addition several characteristics relative to the location of dwellings, such as distance from the closest metro station or distance from the city center are calculated in a GIS environment.

Second, the influence of several factors on housing prices, for example the provision of parking or the type of the dwelling are examined, through statistical testing.

Finally, regression analysis is used in order to estimate housing prices in the Greater Athens region. The explanatory factors are structural as well as locational characteristics. Due to the spatial dependency of the residuals in Ordinary Least Squares method, a spatial regression model (Geographically Weighted Regression) is also presented, which improves the accuracy of the prediction.

A general conclusion is that most locational characteristics did not contribute significantly to the explanatory power of a regression model when compared to structural characteristics. However, the importance of locational factors was more evident when some critical distances were incorporated in statistical testing.

Keywords: Housing prices; Geographically Weighted Regression; Greater Athens region; spatial analysis

1. Introduction

The real estate market in Greece has been one of the most important sectors of economic activity, while the rate of home ownership is among the highest in Europe (Eurostat 2017). The real estate market in Greece experienced rapid growth in the period 2001-08 because of the entrance to the euro zone, the Olympic Games and the availability of abundant and low cost mortgages and residential property values increased significantly. Since 2008 the international economic crisis affected severely the Greek economy, investment in housing dropped dramatically and the real estate market experienced intensifying pressures on prices. Apartment prices dropped cumulatively by 40.8% in the period 2007-17. The decrease in prices was higher in the two major urban centers

(Athens: -43.6% and Thessaloniki: -45.2%), when compared to the rest of the country (Bank of Greece, 2018). However, price changes are not uniform across regions, and recent analysis has indicated price increases in the last three years for several areas in the Athens region (Iliopoulou and Stratakis, 2018).

Apart from the general trends, important differences are observed in the residential property prices which can be attributed to structural attributes, neighborhood characteristics and locational influences (Baranzini et al, 2008; Bhattacharjee, Castro and Marques, 2012; Xiao, 2017). Structural attributes describe the physical structure of the property, such as size, age, floor and type of dwelling. Neighborhood characteristics refer to the socioeconomic conditions, crime, quality of schools etc. Locational influences mainly describe proximity to locations of interest such as parks, recreation areas and transportation. For example the effect of open space on residential property values has attracted much attention in the literature (Anderson and West, 2006; Cho, Poudyal and Roberts, 2008; Luttik, 2000; Sander and Polasky, 2008). In the same way good access to transportation networks contributes to higher property values (Efthymiou and Antoniou, 2013; McMillen and McDonald, 2004).

Hedonic regression is the most common technique in order to explore the factors which contribute to property values (Baranzini et al, 2008; Raslanas, Tupenaite and Šteinbergas, 2006; Sander and Polasky, 2009). In these models data for specific regions are often analyzed without employing techniques of spatial analysis. However, several studies employ spatial analysis and GIS techniques in order to study the spatial variation of housing prices (Lake et al, 2000; Mimis, Rovolis and Stamou, 2013; Pace, Barry and Sirmans, 1998).

2. Materials and methods

In this paper the housing market in the Greater Athens region is analyzed employing data for the supply of dwellings in the study region. The study region includes 58 municipalities which roughly correspond to the Greater Athens area. The study region was delineated according to the density of the data on housing prices. Data were obtained by entries concerning property for sale, published on the internet by real estate agencies. However only houses for which the approximate location on a map was available were included in the sample. House sales prices are not generally available in Greece; therefore asking prices were employed for a spatial analysis of property prices. It is known however, that asking prices are in general higher than market values. The sample comprises 5130 dwellings for sale for 2017 (1st quarter) and the location of each dwelling was reported on a map.

The structural characteristics which were derived from real estate web pages were price, size, age, floor, number of bedrooms, bathrooms, type of the dwelling (i.e. apartments, terraced or duplex houses and detached single unit houses), parking, view and fireplace. In addition, distances from locations which might influence positively house prices were calculated in a GIS environment. Euclidean distances from metro stations, the beach, the center of the city, recreational areas, areas of urban green and areas of coniferous forests were calculated for each dwelling. The land uses were derived from Corine Land Cover 2012 and specifically classes 141 (green urban areas), 142 (sport and leisure facilities), and 312 (coniferous forest) (Kosztra, et al., 2017). It has to be noted however, that this classification, especially for recreational areas and areas of urban green is not detailed enough for the purposes of this study.

In addition, the elevation of each dwelling was computed employing the ASTER Global DEM. This variable was selected because some of the more expensive neighborhoods in the city are found either on hills or close to the coast. Finally, the mean value of the zones which are used for tax purposes (tax values) were calculated for each municipality and were assigned to the dwellings. Tax values in previous research (Iliopoulou and Stratakis, 2018) are positively correlated to prices and they represent adequately the locational attributes of dwellings at the municipality level.

At first a description of the houses in the sample is presented, in terms of their structural and locational characteristics.

Second, the influence of several factors on housing prices, for example the provision of parking or the type of the dwelling are examined, through statistical testing. In this section, the influence of several locational attributes are also tested, i.e. distance from the metro station, the city center etc., after classifying distances in two groups according to some critical for proximity value.

Finally, several regression models are tested in order to estimate housing prices in the Greater Athens region. Linear regression is employed with dependent variable either 'Price' or 'Price per square meter'. The explanatory factors in this model are structural as well as locational characteristics. The spatial dependency of the residuals from Ordinary Least Squares (OLS) is tested through autocorrelation coefficients, and a spatial regression model (Geographically Weighted Regression-GWR) is presented (Fotheringham, Brunsdon and Charlton, 2002). When appropriate, spatial regression improves the accuracy of the prediction in comparison to the ordinary least squares model.

3. Results and Discussion

At first, the description of the structural characteristics of the houses in the sample is shown in Table 1. For the variables 'Price', 'Price per sq.m.' and 'size', there is a considerable difference between the mean and the median. This is an indication of a great variability in the data set which is also expressed by the minimum and maximum values in Table 1.

Table 1. Structural characteristics of houses

	Price (€)	Price per sq.m. (€)	Size (sq.m.)	Age (years)	Floor	Bedrooms (number)	Bathrooms (number)
Mean	279132	1806	127	27	1.8	2.4	1.9
Median	150000	1571	95	26	1.0	2	2
Minimum	5500	160	13	0	0	1	1
Maximum	8500000	18519	1500	117	8	10	10

In Table 2 some locational attributes of the houses are presented. Distances from locations which influence property values are presented, together with altitude and the values of the tax zones. As with the structural characteristics, a great variability of values is observed.

Table 2. Locational characteristics of houses

	Distance metro station (m)	Distance beach (m)	Distance urban green (m)	Distance recreation (m)	Distance forest (m)	Distance city center (m)	Altitude (m)	Tax value (€)
Mean	2024	6709	4190	1552	4111	7814	98	1501
Median	1452	5493	2949	1292	4099	7344	72	1350
Minimum	9	59	0	0	0	23	0	670
Maximum	9954	17666	15936	9310	8765	18934	473	6500

The geographical distribution of the variables 'Price' and 'Price per sq.m.' are presented in Figures 1 and 2. It is evident that higher prices are observed in the southern and northern suburbs of the study region, as well as in a small area at the center of the city. Both spatial patterns were tested for spatial autocorrelation through the Moran's *I* index, using the inverse distance squared conceptualization of spatial relationships in ArcGIS. The values for the index were *I*=0,325 for 'Price' and *I*=0,448 for 'Price per sq.m.' with very small associated p-values, indicating highly significant clustered patterns.

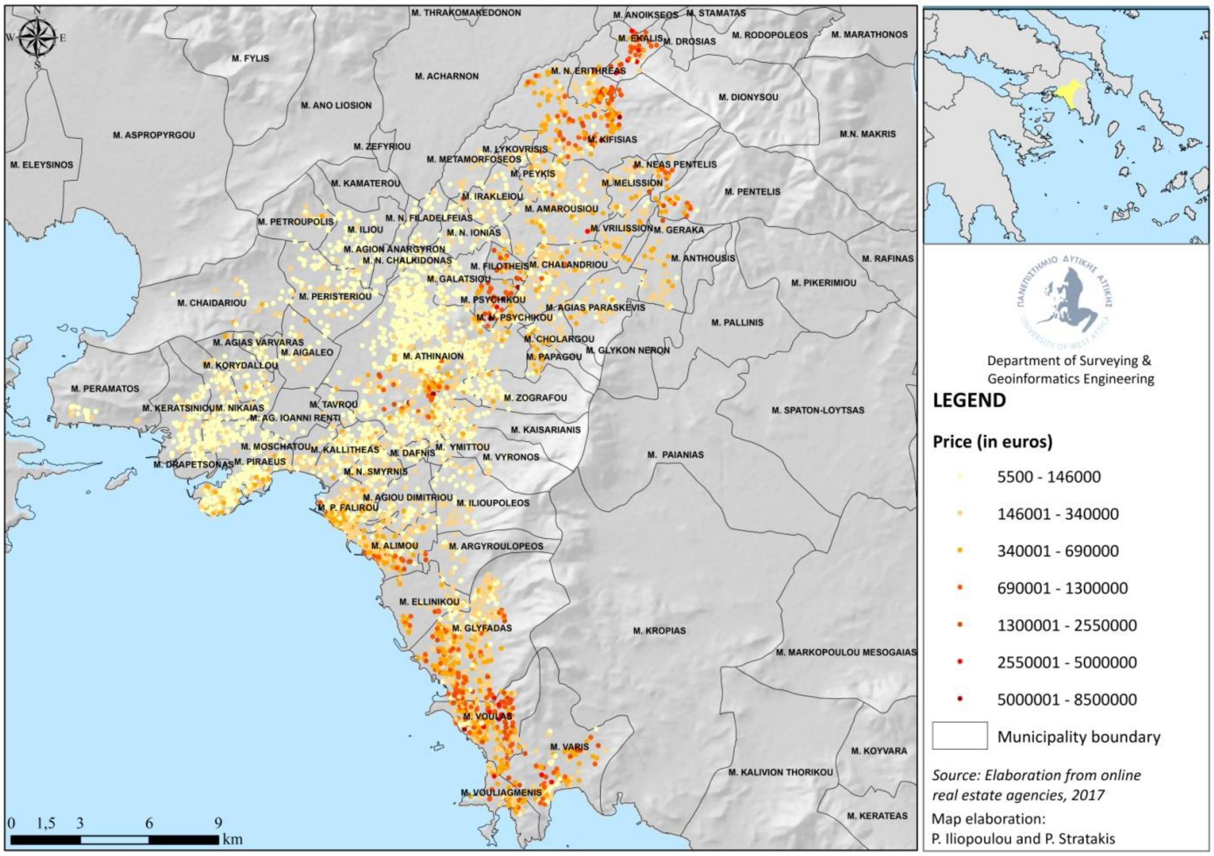

Figure 1. The sample: house prices

In Tables 3 and 4 some qualitative characteristics of the houses are presented, such as the provision of parking, fireplace and view. In addition, in Table 3 distances from locational features are grouped using a value of 500m or 1000m, according to the accessibility characteristics of each feature. For all variables the means for 'Price' and 'Price per sq.m.' were compared using t-test and the mean differences are shown on Table 3. All variables, with the exception of 'distance from areas of urban green', present statistically significant mean differences. The largest difference in mean price is observed for the group of dwellings with a fireplace, while the largest difference in mean price per sq.m. is observed for the group of dwellings with provision of parking. Distance from the metro station shows opposite than expected behavior, i.e. price is higher away from metro stations, probably because metro lines do not cross the most expensive neighborhoods.

In addition an analysis of variance for the mean price and mean price per sq. m. according to the type of houses is presented in Table 4. Descriptive statistics indicate major differences among the three types but the results were significant only when apartments were compared either with single unit houses or terraced and duplex houses.

.

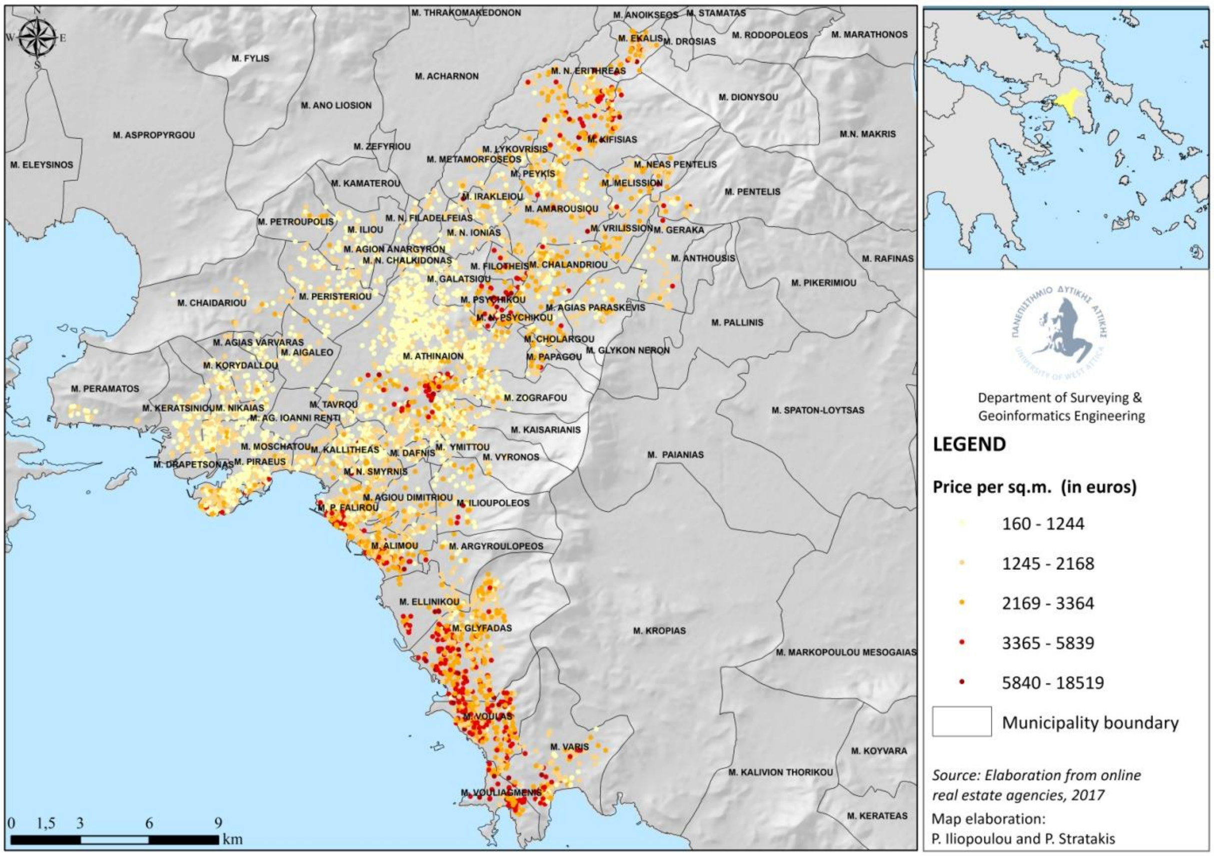

Figure 2. The sample: Price per sq.m.

Table 3. Comparison of means: t-test

	Mean difference 'Price' (in euros)	Mean difference 'Price per sq.m.' (in euros)
Parking (Yes, No)	299305	1127
View (Yes, No)	256456	959
Fireplace (Yes, No)	350999	1106
Distance from recreational areas (<500m, >500m)	73955	446
Distance from beach (<500m, >500m)	100399	827
Distance from coniferous forest (<1000m, >1000m)	214655	202
Distance from CBD (<1000m, >1000m)	263723	1088
Distance from areas of urban green (<500m, >500m)	not significant	not significant
Distance from metro station (<500m, >500m)	-107770	-346

Table 4. Comparison of means: ANOVA

		N	Mean
Price	Single unit house	589	815261
	Appartment	4044	171244
	Terraced-duplex	497	521626
	Total	5130	279132
Price per sq.m.	Single unit house	589	2368
	Appartment	4044	1650
	Terraced-duplex	497	2401
	Total	5130	1806

In order to examine the factors influencing housing values, price was associated with the structural and locational attributes of the houses (Tables 5a and 5b). Correlations are presented for 'Price' and 'Price per sq.m.' with 13 variables representing size, age, floor, bedrooms and bathrooms of the dwellings, altitude, distances from locations of interest and the values of tax zones. The dwelling size has the most powerful positive correlation with price. Age and floor show weak negative correlation with price. The negative correlation between 'Price" and 'Floor' might be related to the fact that the most expensive dwellings are single unit houses (11.5% of the sample) and terraced houses (9.4% of the sample). For these types of houses the floor is recorded as ground floor. Locational factors, as expressed by distances from metro stations, the beach, the center of the city, recreational areas, areas of urban green and areas of coniferous forests, as well as altitude, show weak correlations with 'Price'. Distances from metro stations, the center of the city and areas of urban green present somewhat stronger correlations with 'Price per sq.m.'. In addition, the sign of the correlation coefficients is not always the expected. For example, distance from the metro station is positively correlated with both variables representing price. As mentioned previously, a possible explanation is that the metro stations are extended into a small part of the study area, while they are not located in the most expensive neighborhoods, such as the southern and northern suburbs. Finally, both 'Price' and 'Price per sq.m.' present moderate positive correlation with tax values, indicating that the values used for tax purposes are not proportional to market values.

Table 5a. Correlations

		Size	Age	Floor	Bedrooms	Bathrooms
Price	Pearson Correlation	,794**	-,145**	-,186**	,599**	,690**
	Sig. (2-tailed)	0,000	0,099	0,000	0,000	0,000
Price per sq.m.	Pearson Correlation	,368**	-,392**	-0,004	,359**	,434**
	Sig. (2-tailed)	0,000	0,000	0,775	0,000	0,000

Table 5b. Correlations

		Distance metro station	Distance beach	Distance urban green	Distance recreation	Distance forest	Distance city center	Altitude	Tax value
Price	Pearson Correlation	,273**	-0,023	,189**	,168**	-,140**	,311**	,132**	,440**
	Sig. (2-tailed)	0,000	0,099	0,000	0,000	0,000	0,000	0,000	0,000
Price per sq.m.	Pearson Correlation	,333**	-,141**	,290**	,028*	-,107**	,360**	0,024	,455**
	Sig. (2-tailed)	0,000	0,000	0,000	0,048	0,000	0,000	0,088	0,000

The final part of this study is to predict the house prices employing their attributes. Several sets of independent variables were entered at first in an Ordinary Least Square (OLS) regression model and subsequently in a spatial regression model. Analysis was performed employing statistical software (SPSS) and GIS software (ArcGIS). At first a linear regression model was produced employing the stepwise method in SPSS. The dependent variable is price and 15 independent variables were tested which included: (a) all the structural characteristics of Table 1, (b) all locational attributes of Table 2 (with the exception of tax value) and (c) three dummy variables, i.e. parking, fireplace and view. Only eight independent variables were selected by the stepwise method, i.e. size, distance from recreational areas, age, parking, view, number of bedrooms, number of bathrooms and distance from metro station. The results are presented in Table 6 (OLS model 1). This model accounts for 65.1% of total variation and the residuals were clustered, as the spatial autocorrelation Moran's *I* indicated. This model however, has some problems with the interpretation of the coefficients for two variables: bedrooms and distance from metro station. In addition, bedrooms and bathrooms are positively correlated with size and each other as well. Furthermore, the corresponding GWR model did not yield reliable results.

After examining a very large number of combinations of independent variables, it was evident that the coefficient of determination for the OLS models was around 65%. In addition, three independent variables would appear in all models: size, age and distance from recreational areas. With only these three independent variables, the linear regression model (OLS model 2 in Table 6) resulted to a small loss of explanatory power (0.9%) in comparison to the OLS model 1. The residuals of the OLS model 2 were also clustered. Consequently a Geographically Weighted Regression (GWR) model was calculated and the coefficient of determination increased to 0.769 which is a significant improvement relative to both the OLS models (Table 6). The Akaike Information Criterion has decreased relative to both OLS models indicating an improvement of the model fit. The coefficients of the independent variables for the GWR model are not reported in Table 6 because they are different for each observation, since the method produces local estimations of the dependent variable.

The residuals of the GWR model are presented In Figure 3 and result to a random spatial pattern according to the Moran's *I* index of spatial autocorrelation. However, some residuals remain, mostly in the most expensive areas: the southern and the northern suburbs, the city center (Kolonaki) and Filothei – Psychiko.

Moreover, in many trials of regression analysis, several independent variables describing locational attributes could not be incorporated in a GWR model, due to local collinearity issues.

Table 6. Regression models

	OLS MODEL 1	OLS MODEL 2	GWR
DEPENDENT VARIABLE 'PRICE'	Coefficients and	Coefficients and	Coefficients and
CONSTANT	-70065	-69373	
Size	3618	3504	
Distance from recreational	-38	-36	
Age	-588	-1579	
Parking	45881		
View	23473		
Bedrooms	-62766		
Bathrooms	35089		
Distance from metro station	5		
R^2	0.651	0.642	0.769
AIC	143282	143403	141700

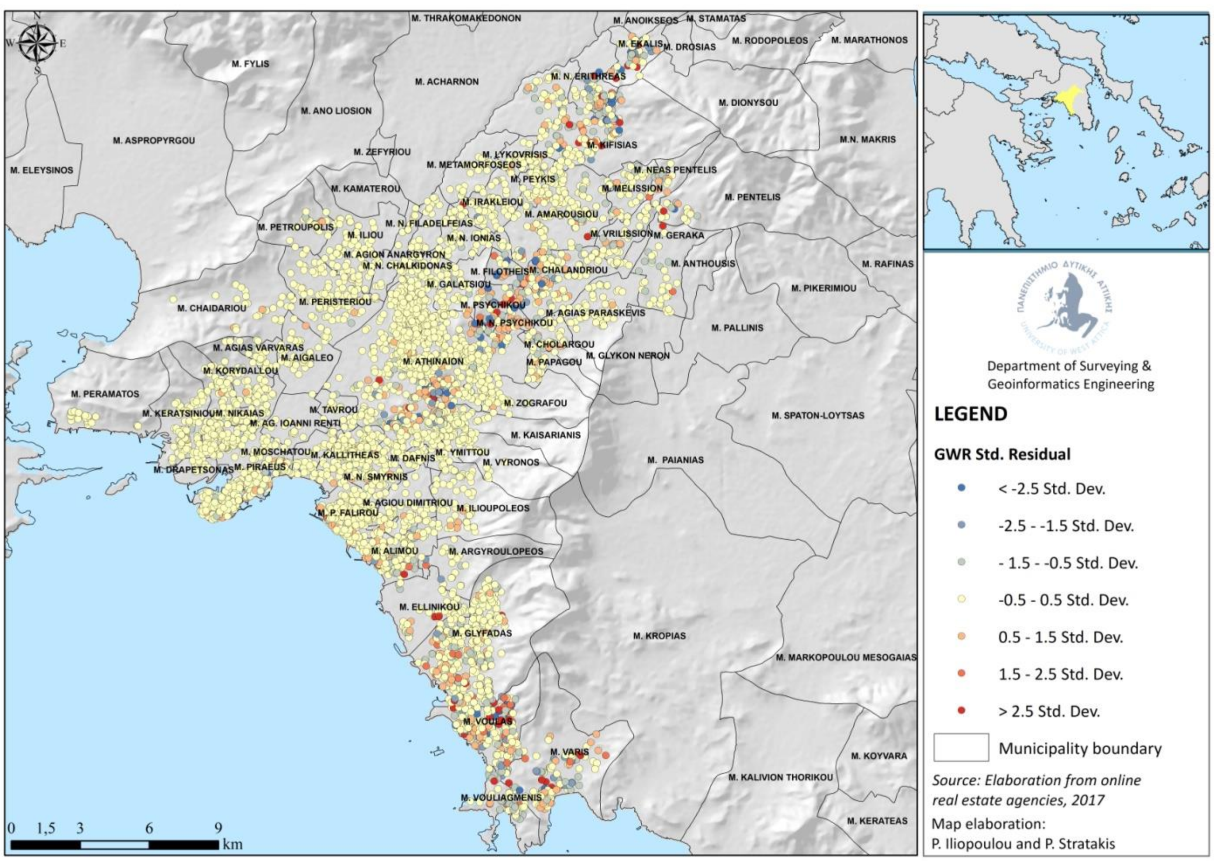

Figure 3. Residuals Geographically Weighted Regression

Finally several trials with dependent variable 'Price per sq. m.' were performed. The linear regression model with 13 independent variables (only floor and distance from the forest were excluded by the stepwise method) produced a coefficient of determination of 38.6%. The logarithmic model was more appropriate in this case and an OLS model with dependent variable ln(Price per sq.m.) and 14 independent variables (only the bedrooms were excluded) resulted to a coefficient of determination of 53.3%, while the GWR model did not yield reliable results due to local autocorrelation of some independent variables.

Conclusions

The real estate market in Greece is under significant pressure since the beginning of the economic crisis. However, apart from the general economic environment, significant differences in property values are observed within the Greater Athens region. A large number of dwellings for sale were analyzed in order to indicate the factors which contribute to the differences in housing prices. Both structural and locational factors were examined and emphasis was attributed to the spatial variation of prices within the study region. The higher prices are observed in the northern and southern suburbs of the Greater Athens region as well as in the city center; the lowest are found in some western suburbs. Statistical tests and correlations showed that structural characteristics, for example size, parking and view, had stronger impact on prices when compared to locational characteristics, mostly distances from locations of interest. This is possibly due to the concentration of transportation infrastructure, recreational facilities etc. into certain areas within the study region and also to data limitations.

Three regression models were produced in order to estimate housing prices in the Greater Athens Region. If eight independent variables are selected, the OLS model results to a rather moderate explanatory power. If only three independent variables are selected, i.e. size, age and distance from recreational areas, a spatial regression model can increase significantly the model fit. In terms of the selection of the model it has to be noted that often locational attributes could not be incorporated in a GWR model, due to local collinearity issues, therefore the selection of independent variables was constrained mostly to structural characteristics. The impact of locational attributes on house prices was better described when some critical distances, in terms of proximity, were used in statistical testing.

References

Anderson, S.T., West, S.E., (2006) Open space, residential property values, and spatial context, *Regional Science and Urban Economics*, 36, p773–789.

Bank of Greece (2018) Indices of residential property. Available at http://www.bankofgreece.gr/BoGDocuments/%CE%9D%CE%AD%CE%BF%CE%B9_%CE%A0%CE%AF%CE%BD%CE%B1%CE%BA%CE%B5%CF%82_%CE%A4%CE%B9%CE%BC%CF%8E%CE%BD_%CE%9A%CE%B1%CF%84%CE%BF%CE%B9%CE%BA%CE%B9%CF%8E%CE%BD_full.pdf. (Accessed: 1 March 2018).

Baranzini, A., Ramirez, J., Schaerer, C., Thalmann, P., eds. (2008) *Hedonic Methods in Housing Markets, Pricing Environmental Amenities and Segregation*, New York: Springer.

Bhattacharjee, A., Castro, E. Marques, J., (2012) 'Spatial Interactions in Hedonic Pricing Models: The Urban Housing Market of Aveiro, Portugal', *Spatial Economic Analysis*, 7(1), p133-167.

Cho S.H., Poudyal N.C. Roberts R.K., (2008) Spatial analysis of the amenity value of green open space. *Ecological Economics*, 66, p403-416.

Efthymiou, D., Antoniou, C., (2013) How do transport infrastructure and policies affect house prices and rents? Evidence from Athens, Greece. *Transportation Research*, Part A 52, p1–22.

Eurostat, (2017) Housing conditions, Available at: http://ec.europa.eu/eurostat/statistics-explained/index.php/Housing_statistics (Accessed: 1 March 2018).

Fotheringham,S. A., Brunsdon, C., Charlton, M., (2002) *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, Wiley: Chichester.

Iliopoulou, P., Stratakis, P., (2018) 'The Geography of housing prices in the greater Athens region, Greece: Patterns, correlations and trends' *11th International Conference of the Hellenic Geographical Society (ICHGS)*, 'Innovative Geographies. Understanding and connecting our world', Lavrion, Greece, April 12-15-2018, available at: http://hellenicgeosociety.org/el/11th-conference-proceedings?field_topic_tid=All&title=&title_field_value_1=ILIOPOULOU

Kosztra, B., Büttner, G., Hazeu, G., Arnold, S., (2017) *Updated CLC illustrated nomenclature guidelines,* Wien: European Environmental Agency.

Lake, I. R., Lovett A.A., Bateman I.J., Day B., (2000) Using GIS and large-scale digital data to implement hedonic pricing studies, *International Journal of Geographical Information Science*, 14(6), p521-541.

Luttik, J., (2000) The value of trees, water and open space as reflected by house prices in the Netherlands, *Landscape and Urban Planning* 48, p161-167.

McMillen, D. P., McDonald, J. (2004) Reaction of House Prices to a New Rapid Transit Line: Chicago's Midway Line, 1983–1999, *Real Estate Economics*, 32: p463–486.

Mimis, A., Rovolis, A. Stamou, M., (2013) Property valuation with artificial neural network: the case of Athens, *Journal of Property Research*, 30(2), p128-143.

Pace, R. K., Barry, R., Sirmans, C.F., (1998) Spatial Statistics and Real Estate, *Journal of Real Estate Finance and Economics,* 17(1), p5-13.

Raslanas, S., Tupenaite, L., Šteinbergas, T., (2006) Research on the prices of flats in the South East London and Vilnius. *International Journal of Strategic Property Management*, 10(1), p51-63.

Sander, H.A., Polasky, S., (2009) The value of views and open space: Estimates from a hedonic pricing model for Ramsey County, Minnesota, USA, *Land Use Policy*, 26, p837–845.

Xiao, Y., (2017) *Urban Morphology and Housing Market*. Shanghai, China: Springer.