ABSTRACT

Purpose - This study examines and documents spatial heterogeneity in Istanbul housing market using Geographically Weighted Model (GWR).

Methodology - A GWR model with a Gaussian kernel and an adaptive bandwidth based on cross-validation is employed on a cross-sectional housing listing data set. Additional analysis is provided using geographically weighted Spearman’s rank correlation measure between prices and variables.

Findings - GWR model substantially boosts goodness of fit in our pricing model compared to a standard hedonic regression model. The variation within GWR coefficients is high and of micro nature. Median GWR coefficients often differ from standard hedonic regression coefficients. The variability of coefficients is plotted on map.

Conclusion - Findings suggest the existence of spatial non-stationarity in standard hedonic regressions and favor the use of models appropriate for spatial heterogeneity. Findings encourage further research in hedonic models applications such as in quality adjustments to price indices.

Keywords: Spatial heterogeneity, spatial non-stationarity, geographically weighted regression, Istanbul housing market, quality adjusted price index

JEL Codes: C31, R30, R31

1. INTRODUCTION

Real estate properties are location fixed and therefore real estate markets are a natural domain for incorporating spatial models. An important distinction in the effects of space is between spatial dependence (spatial autocorrelation) and spatial heterogeneity. The spatial dependence is defined as “the lack of independence which is often present among observations in cross-sectional data sets” and spatial heterogeneity as “structural instability over space” such as “heteroskedasticity, random coefficient variation and switching regressions” (Anselin, 1988). This study employs a geographically weighted regression (GWR) model to examine the spatial heterogeneity in Istanbul housing market.

GWR (Brunsdon et al., 1996; Fotheringham et al., 1997) is a technique used to incorporate a space definition to regressions and enables the exploration of spatial heterogeneous processes. Using an adaptive bandwidth and a Gaussian kernel, we employ a GWR model to our data in order to explore spatial heterogeneity. We use a cross-sectional listing database of housing sales in Istanbul and additional data regarding the city that are collected from open sources.

We find that GWR model substantially boosts goodness of fit in our pricing model. The coefficients of all variables exhibit large variation across Istanbul over space and median GWR coefficients often differ from standard hedonic regression coefficients. Additional analysis is provided using geographically weighted Spearman’s rank correlation measure between prices and variables. Our findings suggest spatial non-stationarity in Istanbul housing market and our results are aligned with
those reported in other housing markets in terms of spatial heterogeneity (Bitter et al., 2007; Huang et al., 2010; Lu et al., 2014; Wu et al., 2014).

Spatial heterogeneity constitutes a problem for global standard hedonic regressions as it violates its assumptions. While hedonic models can be applied in many different methods, using them without sufficient regard of spatial heterogeneity might result in misleading results. One application of hedonic models worth focusing on is the quality adjustments made on price indices to differentiate price increases associated with quality increases (Statistical Office of the European Communities et al., 2013). For this reason, accounting for spatial heterogeneity in price indices has been suggested in literature (Bárcena et al., 2014; Brunauer et al., 2012; Olszewski et al., 2017), Our findings similarly encourage consideration of spatial heterogeneity when using hedonic models in housing markets.

The next section of this paper provides a literature review on GWR and its use on housing markets. Data is explained in Section 3. Section 4 states findings and a discussion of the findings, and Section 5 presents brief conclusion of this study.

2. LITERATURE REVIEW

In order to analyze the spatial heterogeneity in housing prices, we use the GWR model by Bitter et al. (2007) as GWR models differ from spatial dependence models in being able to explore spatial heterogeneous processes. While spatial dependence models accommodate residuals, GWR explicitly allows parameter estimates to vary over space. It fits many weighted regressions in a moving window and it could be used as an exploratory tool for coefficient variation across a surface.

Following the presentation of GWR by Bitter et al (2007), consider a standard hedonic regression, in a representation where the subscripts are explicitly denoted:

\[ y_i = \alpha_i + \sum_k \beta_k x_{ik} + \epsilon_i \]  

where \( \alpha \) is the intercept, \( \beta_k \) is the coefficient of \( k^{th} \) independent variable, \( x_{ik} \) is the \( i^{th} \) observation of \( k^{th} \) independent variable, and \( \epsilon \) is the random error term.

GWR differs from standard hedonic regression by making coefficients local estimates instead of a global estimate:

\[ y_i = \alpha_i + \sum_{1,m} \beta_{ik} x_{ik} + \epsilon_i \]  

where \( \alpha \) is the intercept, \( \beta_{ik} \) is the coefficient of \( k^{th} \) independent variable at location \( i \), \( x_{ik} \) is the \( i^{th} \) observation of \( k^{th} \) independent variable, and \( \epsilon \) is the random error term.

The variation in a parameter’s coefficient estimates across observations comes from a weighted least squares approach such that different emphasis is put on different observations during estimation of a parameter at different locations. Thus, the estimator of GWR in matrix notation is:

\[ \hat{\beta}_i = (X'G_i X)^{-1}X'G_i y \]  

where \( X \) is a set of independent variable vectors, and \( G_i \) is a spatial weighting matrix.

Note that the spatial weighting matrix in GWR represents the spatial definition through which the weights of observations around a locality in a weighted regression are determined. This is in contrast to spatial weights matrices in spatial dependence models, where it represents the spatial definition among observations through which a spatial process occurs and spatial interaction is observed in the form of error structure or spillovers. Both matrices, on the other hand, serve the same purpose of favoring nearby observations in a prediction setting. One possible specification for the spatial weighting matrix in GWR could again be a Gaussian form:

\[ G_{ij} = \left( \frac{-d_{ij}}{h} \right)^2 \]  

where \( d_{ij} \) is the distance between observations \( i \) and \( j \), and \( h \) is the bandwidth parameter

The bandwidth parameter is a tool to control the window of the sliding regressions over the space. The larger the bandwidth gets, the larger the range of the nearby observations are included and the closer the GWR model gets to an OLS model. The model becomes more local as the bandwidth gets smaller; however, the locality is bounded by the fact that the number of observations should be high enough for each regression to provide enough degrees of freedom for estimation. While a priori specification of the spatial weighting schema, the kernel, and the bandwidth is required in GWR, there are adaptive procedures through which a bandwidth is selected in an automatic manner. In this study, such an automated procedure with the Gaussian form is used and the distances are interpreted as adaptive distances.
Proposed as an exploratory spatial technique, GWR has more recently been applied to various areas of real estate and urban studies, such as in spatiotemporal analysis of housing price variation (Huang et al., 2010; Wu et al., 2014; Yao & Fotheringham, 2016), examining urban growth (Salvati, 2019), in housing segmentation (Manganelli et al., 2014) and in differentiating determinants of rent (Bera & Kangalli Uyar, 2019). According to a simulation study by Páez et al. (2011), GWR can indeed be used to determine spatial nonstationarity; however, inferring spatial relationships with GWR is error prone, particularly in small samples, due to spurious relations. In addressing spatial heterogeneity, it has more explanatory power and predictive accuracy than spatial expansion techniques (Bitter et al., 2007).


There has been a few studies on the Istanbul housing market about the necessity or the use of the spatial models. Keskin (2008) reports the need for the inclusion of spatial factors while Ozus et al. (2007) employs different regressions for different districts of Istanbul. Moreover, Koramaz and Dokmeci (2012) performs spatial interpolation and presents a spatial distribution of the regression residuals on map. Kangalli Uyar and Yayla (2016) compares performance of spatial models to find better performance of spatial Durbin model.

3. DATA AND METHODOLOGY

3000 observations of housing listings across Istanbul as of October 15, 2017 are used in this analysis. The data set is provided by HurriyetEmlak.com, a well-known marketplace for real estate listings in Turkey. The data set is a random sample of the active listings at the cross-section; therefore, is not exposed to inflationary time effects of the data collection period. In addition to provided data, a primary data set of significant locations, such as public transportation stops and hospitals, schools or universities in the city is used.

Each observation consists of an attribute set regarding the structural characteristics and locational characteristics of the housing unit and the location of an observation is given by the geolocation of the property. Structural characteristics are age, area, number of bedrooms, number of living rooms and number of bathrooms. Locational characteristics of a housing unit consists of distances to closest social services or significant points such as shopping malls, universities or major transportation stops.

The administrative divisioning of Istanbul consists of districts and neighborhoods. Figure 1 shows the administrative units on map. The map on the left in Figure 1 shows the district and the map on the right shows the neighborhoods. In comparison, Figure 2 shows our data distribution on map. The map on the left in Figure 2 are the housing units on map and the map on the right reflects the density of observations.

**Figure 1: Districts (left) and neighborhoods (right) of Istanbul**
Figure 2: Locations (left) and the density (right) of the observations on map

4. FINDINGS AND DISCUSSIONS

A median housing unit in Istanbul housing market for sale in 2017 is a recently built 3 bedrooms 120 square meters’ apartment with a price tag of 540,000TRY. There is schooling, a police department and a major transportation stop within a kilometer and it can reach to a fire department, a shopping mall and a university within one and half kilometers. On the other hand, the housing unit can vary substantially, especially in prices. The summary statistics of all variables are given in Table 1.

Table 1: Summary Statistics of the Variables

<table>
<thead>
<tr>
<th>Statistic</th>
<th>N</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>price</td>
<td>3000</td>
<td>1022191</td>
<td>1646082</td>
<td>105000</td>
<td>345000</td>
<td>540000</td>
<td>998500</td>
<td>22425600</td>
</tr>
<tr>
<td>logprice</td>
<td>3000</td>
<td>13.4</td>
<td>0.9</td>
<td>11.6</td>
<td>12.8</td>
<td>13.2</td>
<td>13.8</td>
<td>16.9</td>
</tr>
<tr>
<td>age</td>
<td>3000</td>
<td>7.8</td>
<td>11.3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>12</td>
<td>50.0</td>
</tr>
<tr>
<td>area</td>
<td>3000</td>
<td>130.9</td>
<td>56.9</td>
<td>31</td>
<td>90</td>
<td>120</td>
<td>152</td>
<td>475</td>
</tr>
<tr>
<td>bathrooms</td>
<td>3000</td>
<td>1.5</td>
<td>0.7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>livingrooms</td>
<td>3000</td>
<td>1.1</td>
<td>0.3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>bedrooms</td>
<td>3000</td>
<td>2.6</td>
<td>1.0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>km.2.closest.fire.department</td>
<td>3000</td>
<td>1.67</td>
<td>0.97</td>
<td>0.01</td>
<td>0.94</td>
<td>1.48</td>
<td>2.28</td>
<td>5.17</td>
</tr>
<tr>
<td>km.2.closest.hospital</td>
<td>3000</td>
<td>0.93</td>
<td>0.79</td>
<td>0.01</td>
<td>0.46</td>
<td>0.74</td>
<td>1.15</td>
<td>11.07</td>
</tr>
<tr>
<td>km.2.closest.shopping.mall</td>
<td>3000</td>
<td>1.68</td>
<td>1.28</td>
<td>0.02</td>
<td>0.85</td>
<td>1.53</td>
<td>2.30</td>
<td>25.44</td>
</tr>
<tr>
<td>km.2.closest.police.department</td>
<td>3000</td>
<td>1.05</td>
<td>0.73</td>
<td>0.04</td>
<td>0.52</td>
<td>0.87</td>
<td>1.39</td>
<td>6.71</td>
</tr>
<tr>
<td>km.2.closest.transportation</td>
<td>3000</td>
<td>1.55</td>
<td>2.28</td>
<td>0.02</td>
<td>0.56</td>
<td>0.98</td>
<td>1.68</td>
<td>38.63</td>
</tr>
<tr>
<td>km.2.closest.priv.preschool</td>
<td>3000</td>
<td>0.41</td>
<td>0.49</td>
<td>0.01</td>
<td>0.19</td>
<td>0.32</td>
<td>0.52</td>
<td>13.82</td>
</tr>
<tr>
<td>km.2.closest.priv.primschool</td>
<td>3000</td>
<td>0.80</td>
<td>0.86</td>
<td>0.02</td>
<td>0.40</td>
<td>0.68</td>
<td>1.02</td>
<td>25.43</td>
</tr>
<tr>
<td>km.2.closest.priv.secondarieschool</td>
<td>3000</td>
<td>0.78</td>
<td>0.84</td>
<td>0.02</td>
<td>0.41</td>
<td>0.68</td>
<td>0.98</td>
<td>25.42</td>
</tr>
<tr>
<td>km.2.closest.priv.highschool</td>
<td>3000</td>
<td>0.88</td>
<td>0.90</td>
<td>0.01</td>
<td>0.44</td>
<td>0.73</td>
<td>1.13</td>
<td>25.42</td>
</tr>
<tr>
<td>km.2.closest.public.preschool</td>
<td>3000</td>
<td>2.07</td>
<td>1.80</td>
<td>0.04</td>
<td>0.97</td>
<td>1.61</td>
<td>2.61</td>
<td>13.79</td>
</tr>
<tr>
<td>km.2.closest.public.primschool</td>
<td>3000</td>
<td>0.41</td>
<td>0.31</td>
<td>0.01</td>
<td>0.21</td>
<td>0.34</td>
<td>0.52</td>
<td>3.18</td>
</tr>
<tr>
<td>km.2.closest.public.secondarieschool</td>
<td>3000</td>
<td>0.43</td>
<td>0.33</td>
<td>0.01</td>
<td>0.21</td>
<td>0.35</td>
<td>0.53</td>
<td>4.36</td>
</tr>
<tr>
<td>km.2.closest.public.highschool</td>
<td>3000</td>
<td>0.59</td>
<td>0.41</td>
<td>0.01</td>
<td>0.32</td>
<td>0.50</td>
<td>0.75</td>
<td>3.53</td>
</tr>
<tr>
<td>km.2.closest.university</td>
<td>3000</td>
<td>1.64</td>
<td>1.49</td>
<td>0.02</td>
<td>0.74</td>
<td>1.34</td>
<td>1.95</td>
<td>12.03</td>
</tr>
</tbody>
</table>

Two types of regression, a standard hedonic regression and geographically weighted regression is employed using the same log-linear pricing model. This log-linear model includes distances to significant places in the city as direct locational variables.
in addition to a few structural characteristics of the housing unit. The standard hedonic regression model has an \( R^2 \) of 0.5978 (with an adjusted \( R^2 \) of 0.5952). The GWR model substantially improves the goodness of fit with an \( R^2 \) of 0.9252 (and an adjusted \( R^2 \) of 0.8952). The residuals of standard hedonic model (left) and GWR models (right) are plotted on the map in Figure 3. A visual inspection suggests that GWR successfully pulls the residuals towards zero. Comparing both maps, there is a decrease in coloring in GWR map and the extreme values are reduced.

**Figure 3: Standard Hedonic Regression Residuals (left) and GWR Residuals (right) on Map**

The bandwidth of GWR is chosen adaptively by an automated cross-validation process and has a value of 35, meaning that the geographical weighting is done in bandwidths of nearest 35 neighbors based on great circle distances. The results of both models are summarized side by side in Table 2. First five statistics in the table columns present the quantiles of the coefficients of the GWR model and the last two statistics in the columns are coefficients and standard errors in standard hedonic regression model.

For all of the 19 variables in the regression, the GWR coefficients show great variability. For instance, the coefficients for variable age vary from -0.0260 to 0.08529 and area vary from 0.0012 to 0.0133 across space. Such variability is not only of orders of magnitude but also include a sign change.

Most of the median coefficients in GWR are not similar to the standard hedonic regression coefficients. Moreover, there are often sign differences between the OLS coefficient and the median GWR coefficient. For instance, the age of a housing unit has a statistically significant coefficient of 0.01 in OLS, while the median GWR coefficient for age is -0.01. Furthermore, the median GWR coefficient is outside of three standard deviation range of the OLS coefficient.

The variability of selected GWR coefficients are plotted on map in Figure 4. As some coefficients have sharp variation in extreme values, such as from -12.69 to 60.62 as in distances to a shopping mall, it is not always possible to show relatively smaller but still large variation on map without selective plotting, the distribution of GWR coefficients on map changes on variable and the pattern they exhibit is not immediately noticeable.

A similar way of exploring spatial heterogeneity is looking at the direct correlation between the prices and the variables. This can be done by using geographically weighted summary statistics (Gollini et al., 2015). Using the same adaptive specification as above, we compute geographically weighted Spearman’s Rank Correlation between natural logarithm of price and selected variables over space. The results, presented in Figure 5, again shows substantially changing correlations over micro areas and suggests spatial heterogeneity.

The large variability, sign changes and map distributions of GWR coefficients as well as the differences between median GWR and OLS coefficients suggest that the hedonic regression results does not seem to be robust across space and indicates spatial heterogeneity. One alternative cause for this could be multicollinearity among independent variables. According to Wheeler & Tiefelsdorf (2005) local estimates can be collinear even though independent variables are uncorrelated and this is a potential cause for numerical instability. On the other hand, Fotheringham & Oshan (2016) argue based on simulation work that GWR is robust to multicollinearity effects, except for the most extreme cases. The variance inflation factors (VIF’s) for the global variables in our analysis are provided in Appendix A and they suggest there is no multicollinearity in global regression as none of the VIF values exceed 10.
ski et al., -
-
- poses. In comparison, our results suggest the median property, could be misaligned depending on the
-
larly CBRT (Hülagü et al., 2016) uses hedonic models in performing quality
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
Figure 4: GWR Residuals on Map
5. CONCLUSION

This paper documents spatial heterogeneity in Istanbul housing market. We employ a GWR model with a Gaussian kernel and an adaptive bandwidth based on cross-validation. Using a log-linear pricing model having structural and locational housing characteristics in a GWR model allowing for spatial heterogeneity, the $R^2$ is significantly increased from 0.60 to 0.93. The variation in GWR coefficients is high and of micro nature. Median GWR coefficients often differ from standard hedonic regression coefficients. The findings suggest spatial non-stationarity in house pricing regression coefficients. Results are confirmed using a geographically weighted Spearman’s rank correlation. The results are comparable to those reported in the literature for other housing markets in terms of performance, coefficient variability and sign changes (Bitter et al., 2007; Huang et al., 2010; Lu et al., 2014; Wu et al., 2014).

An implication of spatial non-stationarity is regarding the use of hedonic models in house price index quality adjustments. Hedonic models are used to differentiate between quality price increases and inflationary house price increases (Hülagü et al., 2016; Statistical Office of the European Communities et al., 2013). However, our findings support the idea that it could be beneficial to incorporate local spatial heterogeneity in price indices. (Bárcena et al., 2014; Brunauer et al., 2012; Olszewski et al., 2017)
The results calls for caution in the use of hedonic pricing in policy making. The use of statistics based on hedonic pricing in policy decisions requires the spatial heterogeneity of the market to be kept in mind and further discussion on spatial non-stationarity is needed before applying them to decision making processes of far reaching decisions.

REFERENCES


APPENDIX A

Table 3: Global VIF Values

<table>
<thead>
<tr>
<th>Statistic</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>1.13</td>
</tr>
<tr>
<td>area</td>
<td>4.41</td>
</tr>
<tr>
<td>bathrooms</td>
<td>2.35</td>
</tr>
<tr>
<td>livingrooms</td>
<td>1.24</td>
</tr>
<tr>
<td>bedrooms</td>
<td>3.33</td>
</tr>
<tr>
<td>km.2.closest.fire.department</td>
<td>1.28</td>
</tr>
<tr>
<td>km.2.closest.hospital</td>
<td>2.26</td>
</tr>
<tr>
<td>km.2.closest.shopping.mall</td>
<td>1.96</td>
</tr>
<tr>
<td>km.2.closest.police.department</td>
<td>1.48</td>
</tr>
<tr>
<td>km.2.closest.transportation</td>
<td>2.45</td>
</tr>
<tr>
<td>km.2.closest.priv.preschool</td>
<td>2.79</td>
</tr>
<tr>
<td>km.2.closest.priv.primaryschool</td>
<td>8.77</td>
</tr>
<tr>
<td>km.2.closest.priv.secondarieschool</td>
<td>3.56</td>
</tr>
<tr>
<td>km.2.closest.priv.highschool</td>
<td>8.83</td>
</tr>
<tr>
<td>km.2.closest.public.preschool</td>
<td>1.66</td>
</tr>
<tr>
<td>km.2.closest.public.primaryschool</td>
<td>1.84</td>
</tr>
<tr>
<td>km.2.closest.public.secondarieschool</td>
<td>1.54</td>
</tr>
<tr>
<td>km.2.closest.public.highschool</td>
<td>1.95</td>
</tr>
<tr>
<td>km.2.closest.university</td>
<td>2.43</td>
</tr>
</tbody>
</table>