Erciyes Üniversitesi
İktisadi ve İdari Bilimler Fakültesi Dergisi
ISSN: 1301-3688/e-ISSN: 2630-6409

Araştırma Makalesi | Research Article
Yayın Tarihi: 30 Aralık 2022 | Published: December 30, 2022
doi: 10.18070/erciyesiibd.1122568

# SPATIAL HETEROGENEITY IN HOUSING MARKET: ANKARA METROPOLITAN AREA

***TUĞBA GÜNEŞ[1*] & AYŞEN APAYDIN[2]***

*[1] Dr., Ankara Üniversitesi, gunest@ankara.edu.tr; https://orcid.org/ 0000-0002-7472-1017, [2] Prof. Dr., Ankara Üniversitesi, Uygulamalı Bilimler Fakültesi, Sigortacılık ve Aktüerya Bilimleri Bölümü, aapaydin@ankara.edu.tr; https://orcid.org/ 0000-0003-4683-0459*

## ABSTRACT

Advanced statistical models have been widely used in real estate valuations for various purposes over the last fifty years, and hedonic approaches with their simple and easy interpretable features are still the most popular among these models. However, spatial heterogeneity and spatial autocorrelation are the two major features of the housing markets, and traditional regression cannot reflect these locational effects into the model sufficiently. This study employs a Geographically Weighted Regression (GWR) model to explore the spatial heterogeneity in the metropolitan area housing market in the city of Ankara. By applying a Gaussian kernel weighting function with adaptive bandwidth based on cross-validation approach on a house listing dataset, it is found that the GWR fit the data better than the traditional ordinary least squares regression which mostly ignore the spatial effects, and there is spatial heterogeneity in the housing market. Explanatory power of the GWR model and parameter estimations are non-stationary over the geographical area. The variations in the coefficients of the variables are depicted on the map and is supported with the spatial correlations between the housing prices and attributes as well.

**Keywords:** spatial heterogeneity, housing market, geographically weighted regression

***Editör / Editor:***
*Hatun KORKMAZ,*
*Erciyes Üniversitesi, Türkiye*

***Hakemler / Referees:***
*Emre EKİNCİ,*
*Yozgat Bozok Üniversitesi, Türkiye*
*Mehmet MARIN,*
*Mardin Üniversitesi, Türkiye*

****Sorumlu Yazar / Corresponding Author:***
*Tuğba Güneş*
*gunest@ankara.edu.tr*

***JEL:***
*C31, R30, R31*

***Geliş:** 28 Mayıs 2022*
***Received:** May 28, 2022*
***Kabul:** 8 Ekim 2022*
***Accepted:** October 8, 2022*
***Yayın Tarihi:** 30 Aralık 2022*
***Published:** December 30, 2022*

***Atıf Önerisi / Cited as (APA):***
*Güneş, T. & Apaydın A. (2022), Spatial Heterogeneity in Housing Market: Ankara Metropolitan Area, Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 63, 9-16, doi: 10.18070/erciyesiibd.1122568*

# KONUT PİYASASINDA MEKANSAL HETEROJENLİK: ANKARA METROPOLİTEN ALANI

## ÖZ

İleri istatistiksel modeller, çeşitli amaçlarla gerçekleştirilen gayrimenkul değerleme çalışmalarında son elli yıldır yaygın olarak kullanılmakta olup, hedonik yaklaşımlar basit ve kolay yorumlanabilir özellikleri sebebiyle bu modeller arasında popüler hale gelmiştir. Ancak konut piyasalarında mekânsal heterojenlik ve mekânsal otokorelasyon durumları söz konusu olabilmektedir ve geleneksel regresyon analizinde bu konumsal etkiler modele yeterince yansıtılamamaktadır. Bu çalışmada Coğrafi Ağırlıklandırılmış Regresyon (CAR) analizi kullanılarak, Ankara ilinin metropoliten alanındaki konut piyasasında mekânsal heterojenlik incelenmiştir. Konut fiyatları ve özelliklerinden oluşan veri seti üzerinde Gauss kernel ağırlık fonksiyonu ve çapraz doğrulama yöntemine dayalı olarak belirlenen değişken (adaptif) bant genişliği kullanılmış, mekânsal etkileri çoğunlukla göz ardı eden en küçük kareler yöntemine dayalı geleneksel regresyon modeline kıyasla CAR modelinin daha başarılı sonuçlar elde ettiği ve konut piyasasında mekânsal heterojenlik olduğu görülmüştür. CAR modelinin konut fiyatlarını açıklama gücünün ve parametre tahminlerinin coğrafi olarak durağan olmadığı anlaşılmıştır. Parametrelerdeki bu değişimler harita üzerinde gösterilerek açıklanmış ve konut fiyatları ile özellikleri arasındaki mekânsal korelasyonlar yardımıyla bu sonuçlar desteklenmiştir.

**Anahtar Kelimeler:** mekânsal heterojenlik, konut piyasası, coğrafi ağırlıklandırılmış regresyon

## INTRODUCTION

Real estate markets have been either the reason or trigger of various economic and financial crisis in the history. Great Depression in the late 1920s, savings and loan crisis in the US in 1980s, and the global financial crisis occurred in 2007 are some of the well-known downturns that triggered by real estate markets and affected severely many countries' economies (Mooya, 2016). Property price trends can provide significant suggestions for market actors, policy developers and decision makers in all related fields ranging from financial markets to urban planning, taxation, and social aids (Hu et al., 2019; Schulz, Wersing, & Werwatz, 2014). Therefore, property price indexes and real estate valuation have become one major interest area of many practitioners, governments, and researchers.

Appraisers traditionally use three approaches to estimate a property's value: the cost, income, and sales-comparison approaches. On the other hand, valuation approaches based on advanced statistical methods have been used since 1970s. Since these approaches, known as "mass appraisal" or "computer assisted mass appraisal - CAMA", enable determining the values of too many properties in a short-period of time, they have been used for taxation purposes in many countries such as the US, Germany, Denmark, and the Netherlands (Gloudemans & Almy, 2011; Grover, 2016). The quest of mortgage market professionals to use these advanced statistical models in also collateral valuation led to developing automated valuation models (AVMs) since 1980s (RICS, 2017).

Hedonic approach that is based on the principle of "the price (or value) of real properties is a function of their various observable and measurable attributes" has gained wide acceptance in predicting property prices as in developing AVMs. The relationship between the real estate values and attributes is usually formulated with a regression equation (Mooya, 2016). Multiple Regression Analysis (MRA) (generally with ordinary least squares - OLS) with its simplicity and easy interpretation has become the most common method adopted in automated property valuations (Matysiak, 2017; Steurer, Hill, & Pfeifer, 2021). However, standard OLS regression assumptions may be violated in real life problems, and the risk of multicollinearity and the difficulty of reflecting the location to the model are the two major issues in real estate analyses. Multicollinearity, strong correlations among independent variables in a regression model, can become a more severe phenomenon in real estate analysis because properties built in same periods of time and/or in close proximities may have similar attributes (Grover, 2016).

Location has been emphasized as the most fundamental driver of property prices in the related literature (Chica-Olmo, 2007; Davis, Bidanset, McCord, & Cusack, 2019; Tchuente & Nyawa, 2021). This effect of the location is divided into two by Páez, Long, and Farber (2008) as neighborhood and adjacency effects. Neighborhood attributes can be reflected in a regression model by incorporating area-based indicators such as socio-economic conditions, education level, accessibility of transportation while the adjacency is related to the issues of spatial dependency and heterogeneity. Spatial dependence is a term related to the spatial autocorrelation. It refers that the closer the distance among the properties is, the stronger the similarities of property attributes and prices. Properties in close proximities to each other share common local amenities and were generally built in the same time periods (Crawford, 2009; Militino, Ugarte, & Garcia-Reinaldos, 2004; Tchuente & Nyawa, 2021). However, marginal prices of property attributes may vary as the location changes throughout an area or space, and this indicates the spatial heterogeneity (Tchuente & Nyawa, 2021).

This study employs a Geographically Weighted Regression (GWR) model to explore the spatial heterogeneity in the metropolitan area housing market of Ankara. GWR is basically a parametric model based on traditional regression but also considers the spatial heterogeneity. By applying a Gaussian kernel weighting function with adaptive bandwidth based on cross-validation approach on a house listing dataset, we found that there is spatial heterogeneity in Ankara housing market. Therefore, spatial heterogeneity should be considered in the analyses and the GWR can outperform than the traditional OLS models.

Section I of this study presents a summary of the related research studies in the literature. Section II provides an overview of the GWR methodology and the OLS regression. Study area and the data used in this paper are explained in Section III, the findings of the empirical analysis are provided in Section IV, and finally a brief conclusion of this study is given in the last section.

## LITERATURE REVIEW

Spatial heterogeneity and spatial autocorrelation are two major components of hedonic modelling approaches in assessing real estate markets. They violate the assumptions of the traditional OLS models which also cannot account for effects of location accurately. Therefore, spatial econometric or local modelling techniques that allow to consider the spatial heterogeneity and spatial autocorrelation in real estate market analysis have been widely used in the literature (P. E. Bidanset, Lombard, Davis, McCord, & McCluskey, 2017; Wang & Chen, 2020).

The Geographically Weighted Regression (GWR) is a local modelling technique that allows parameter estimates to differ over a space (Brunsdon, Fotheringham, & Charlton, 1996; Fotheringham, Brunsdon, & Charlton, 2003; Fotheringham, Charlton, & Brunsdon, 1998). Rather than specifying a single global model for the whole study area, the GWR estimates separate models for each regression point (housing sale point) and give greater weights to the nearby observations around this point. Thus, marginal prices of housing attributes may show variations based on the housing locations. In other words, the GWR can account for the non-stationarity relationship between house prices and their attributes (Fullerton & Bujanda, 2018).

There is strong evidence in the literature that the GWR models can outperform than the traditional OLS models in housing market analysis (P. E. Bidanset & Lombard, 2014; P. E. Bidanset et al., 2017; Bitter, Mulligan, & Dall'erba, 2007; Páez et al., 2008). For instance, Hanink, Cromley, and Ebenstein (2012) used the GWR to analyze the spatial heterogeneity and determinants of the housing prices and apartment rents in China. A similar study were performed by Yu (2007) to explore the spatial non-stationarity between single family house values and their predictors in the US. More specifically, Fullerton and Bujanda (2018) tested the impact of accessibility and proximity to the transportation conditions on the commercial real properties in Texas metropolitan area, and found that GWR estimations were superior to the outputs for the traditional OLS regression. Wang and Chen (2020) explored the impact of both global and local built-environment conditions on house price increases.

By adding the time component into the weighting scheme, spatiotemporal GWR models were introduced, which allows giving higher weights to the transactions occurred more recently. Huang, Wu, and Barry (2010), Fotheringham, Crespo, and Yao (2015), P. Bidanset, McCord, Lombard, Davis, and McCluskey (2018), and Soltani, Pettit, Heydari, and Aghaei (2021) are some of the examples in the growing literature that exploring spatiotemporal variations in residential property prices in various countries.

There is a limited literature directly using locational modelling techniques to investigate the spatial heterogeneity in housing prices in Ankara. By applying the GWR technique, Morali and Yilmaz (2020) found the existence of spatial heterogeneity in İstanbul housing market in their recent study. Sayın (2021) showed the superiority of the spatial models in modeling housing prices in the city of İzmir. Sisman and Aydinoglu (2022) studied in a specific district in İstanbul by using the GWR and several global spatial models to show the geographic variations of property market determinants.

## METHODOLOGY

Hedonic modelling approach has been widely used in real estate price analysis to identify the marginal contribution of property characteristics on their prices (Matysiak, 2017; Sirmans, Macpherson, & Zietz, 2005). The hedonic models are mostly expressed in traditional linear regression models based on the ordinary least squares (OLS). Due to its practicality in making interpretations of the price elasticity of property characteristics, the semi-log model is applied in these analyses (Davis et al., 2019). The conventional OLS model specification is as follows:

$$\ln y_i = \beta_0 + \sum_k \beta_k x_{ik} + \varepsilon_i \tag{1}$$

where $\ln y_i$ is the natural logarithm of the housing price of the $i$. sale, $\beta_0$ is the intercept, $\beta_k$ is the $k$. coefficient (estimated parameter), $x_{ik}$ is the $k$. variable for the i. housing sale, and $\varepsilon_i$ is the error term of the i. housing sale.

The OLS model includes all available data in one equation and assumes that this formula is constant everywhere in the study area. However, property attributes, prices and price drivers may show significant variations over a geographic space. Dummy variables might be included in the global OLS model to capture locational effects, but this will be insufficient to consider the spatial heterogeneity and spatial autocorrelation in property markets. The GWR, developed by Brunsdon et al. (1996) and Fotheringham, Brunsdon, and Charlton (2000); Fotheringham et al. (1998), is a technique allowing to take into account the spatial heterogeneity within the entire space of interest (Bitter et al., 2007).

By following the Tobler (1970)'s First Law of Geography stating that *"Everything is related to everything else, but near things are more related than distant things"*, the GWR model gives a higher weight to the observations (houses) that are closer to a subject property (regression point) than to those are further away. Rather than specifying one single formula, the GWR estimates a separate model for each housing point and weight a subset of observations by their distances to this regression point (Brunsdon et al., 1996; Fotheringham et al., 2003). In essence, parameter estimations of a GWR are conducted by an enhanced form of weighted least squares (WLS) approach. The GWR model is expressed as

$$\ln y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) + \varepsilon_i \tag{2}$$

where $u_i$ and $v_i$ denote longitude and latitude of house $i$, $\beta_k(u_i, v_i)$ is the parameter of the $k_{th}$ explanatory variable of house $i$ to be estimated. To produce local parameters for each observation, a spatial weighting matrix is determined including all observations falling within a specific bandwidth around a regression point $(u_i, v_i)$. The bandwidth can be determined by number of neighbors, by distance or by a kernel function which can be fixed or adaptive. As the bandwidth increases, the number of observations to be included in each regression increases and the GWR model gets closer to the global OLS model. On the other hand, if the bandwidth is identified narrower, then the GWR model becomes more local; however, the risk of insufficient subsample size may arise in this case. Kernel functions optimize the bandwidth to find the most appropriate number of neighboring points to be included in each regression. Minimization of the Akaike Information Criterion (AIC) and a cross-validation (CV) are the two most common methods in identifying the optimal bandwidth (Fotheringham et al., 2003).

GWR model estimations are undoubtedly sensitive to the choice of bandwidth and kernel function (Guo, Ma, & Zhang, 2008). Bujanda and Fullerton (2017) stated that using adaptive kernel bandwidths is recommended when regression points are not uniformly distributed in the study area because an adaptive kernel tries to find a certain number of neighboring points to ensure a constant subsample size in each location while a fixed kernel include all observations within the specified fixed radius around each regression point. In this study, we use a Gaussian kernel weighting function with adaptive bandwidth by adopting both CV and AIC approaches.

## DATA AND STUDY AREA

Real estate transaction prices in the official records in Türkiye do not reflect the actual prices. According to the current legislations, buyers and sellers must report the actual price during the transaction, and this price cannot be lower than the tax value of the real property. However, buyers and sellers usually report the tax values as the actual prices to avoid paying high land registry fees because real estate tax values are significantly lower than their market values. Furthermore, housing attributes are also not available in the official records. Due to lack of reliable data for house prices and attributes, we used a cross-sectional dataset provided by hepsiemlak.com, one of the most popular websites for all types of real estate listings in Türkiye. The dataset includes listing prices and attributes of 2784 residential properties as of November 13, 2021 in the Ankara metropolitan area where the majority of the real estate transactions in the city occur according to the official records of the Land Registry and Cadastre Agency (Tapu ve Kadastro Genel Müdürlüğü). As seen in Table 1, in the last five years annually more than 80 percent of the property sales transactions were made in this particular area consisting of eight counties. Figure 1 shows the Ankara metropolitan districts included in this study (a), and distribution (b) and density (c) of dwellings on the map. In addition to our main data set, POI (points of interest) data we considered in this study was retrieved from the OpenStreetMap.

**TABLE 1 |** Distribution of Real Estate Transactions in Ankara

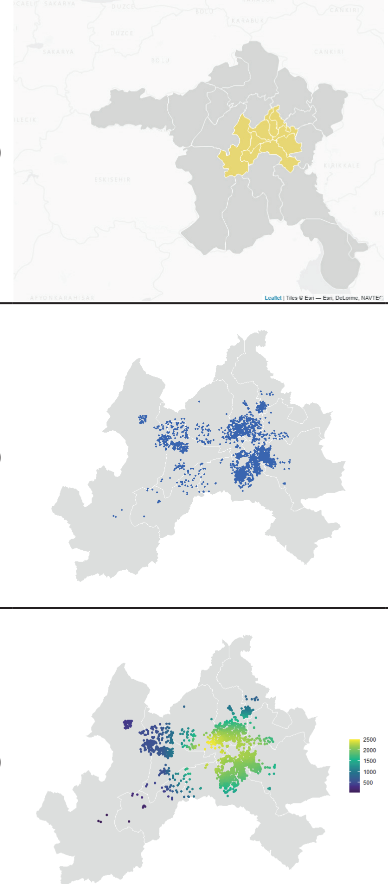
| Districts | 2017 | 2018 | 2019 | 2020 | 2021 |
|---|---|---|---|---|---|
| Çankaya | 13.25% | 13.80% | 14.47% | 15.00% | 15.57% |
| Keçiören | 13.59% | 13.25% | 13.04% | 13.44% | 12.97% |
| Mamak | 12.54% | 12.24% | 12.10% | 11.67% | 11.62% |
| Etimesgut | 10.20% | 9.50% | 10.02% | 11.13% | 10.97% |
| Yenimahalle | 10.45% | 10.49% | 11.46% | 10.94% | 10.33% |
| Sincan | 9.48% | 8.66% | 8.93% | 9.22% | 10.06% |
| Altındağ | 8.48% | 8.91% | 8.63% | 7.98% | 7.54% |
| Pursaklar | 3.33% | 3.55% | 3.58% | 3.52% | 3.38% |
| **Total** | **81.31%** | **80.40%** | **82.24%** | **82.88%** | **82.43%** |

**FIGURE 1 |** Study Area (a), Housing Units Distribution (b) and Density (c) Maps

(a)

(b)

(c)



Our final data set includes the listing prices (Turkish Lira) of each housing unit and their structural and locational characteristics including the geolocations. Variables were selected primarily on the basis of literature review and previous research. Structural variables are net surface area of the dwelling in square meters (net.area), numbers of

**TABLE 2 |** Summary Statistics

| Variables | N | Mean | Std. Dev. | Min | Q1 | Median | Q3 | Max |
|---|---|---|---|---|---|---|---|---|
| price | 2784 | 647650 | 525446 | 88000 | 325000 | 490000 | 775000 | 5820000 |
| net.area | 2784 | 135.85 | 48.66 | 40 | 105 | 120 | 150 | 430 |
| bedroom | 2784 | 3.22 | 0.87 | 1 | 3 | 3 | 4 | 7 |
| livingroom | 2784 | 1.02 | 0.14 | 1 | 1 | 1 | 1 | 3 |
| total.floor | 2784 | 5.94 | 5.29 | 1 | 3 | 4 | 6 | 49 |
| age | 2784 | 13.05 | 11.18 | 0 | 3 | 12 | 20 | 60 |
| num.views | 2784 | 2.07 | 0.82 | 1 | 2 | 2 | 3 | 4 |
| dist.transport | 2784 | 2.16 | 3.06 | 0.00 | 0.44 | 0.95 | 2.46 | 24.48 |
| dist.school | 2784 | 1.52 | 3.16 | 0.00 | 0.14 | 0.26 | 0.47 | 24.62 |
| dist.mall | 2784 | 3.19 | 3.32 | 0.01 | 1.23 | 1.97 | 3.63 | 26.76 |
| dist.hospital | 2784 | 2.31 | 3.08 | 0.00 | 0.60 | 1.07 | 2.35 | 25.39 |
| dist.university | 2784 | 2.96 | 2.96 | 0.00 | 1.14 | 1.93 | 3.66 | 25.21 |
| dist.police | 2784 | 3.28 | 3.94 | 0.02 | 0.87 | 1.57 | 3.73 | 25.46 |

| Categorical Variables | | N | Percentage |
|---|---|---|---|
| floor | top.floor | 681 | 24.50% |
| | middle.floor | 1365 | 49.00% |
| | ground.floor | 453 | 16.30% |
| | basement.floor | 285 | 10.20% |
| bath | more than 1 | 1174 | 42.20% |
| carpark.indoor | yes | 752 | 27.00% |
| carpark.outdoor | yes | 1906 | 68.50% |
| gated. community | yes | 724 | 26.00% |

bedrooms (bedroom) and living rooms (livingroom), floor number in four categories (basement level, ground floor, top floor, and middle floors between the top and ground floors), number of views (east, west, south, north) of the dwelling (num.views), a categorical variable representing whether the dwelling has one or more than one bathroom (bath), total number of floors in the building (total.floor), age of the building in years (age), whether or not the building has an indoor or outdoor parking area (carpark.indoor and carpark.outdoor), and whether it is located in a gated community (gated.community). Several locational variables to reflect the attractiveness of the building by the distances to selected points are also included: distances of the building in kilometers to the closest transportation station (dist.transportation), primary school (dist.school), shopping mall (dist.mall), hospital (dist. hospital), university (dist.university), and police station (dist.police). Dependent variable is natural logarithm (to correct for skewing) of asking prices (price in Turkish Lira) of each housing unit. An overview of the included variables is given in Table 2.

## EMPRICIAL ANALYSIS

We first performed a standard ordinary least squares (OLS) regression model that is a global regression and serves as the benchmark model. Afterwards, we employed geographically weighted regression (GWR) to lay down that parameter estimations are not constant over the space and spatial data is not stationary. All analyses were performed in R software.

The traditional OLS regression results are shown in Table 3. Thirteen out of independent variables were found statistically significant. Adjusted R2 of the model is 0.695 which means that the explanatory variables can explain 69.5% of the variation of the housing prices (dependent variable). Multicollinearity does not seriously exist as all global variance inflation factor values (GVIF) are lower than five, except for the variable of the distance to the closest primary school. However, global OLS regression model assumes that marginal prices of property attributes are constant throughout the study area. Also, Moran's I value was calculated 0.149 (p-value < 0.000), which indicates that the residuals generated from the global OLS regression are not statistically random. In other words, housing prices in this study have significant agglomeration characteristics and positive spatial autocorrelation. To explore the spatial heterogeneity, we perform a GWR for further analysis.

**TABLE 3 |** Global OLS Regression Results

| Variables | Coef. | Std. Error | t value | p value | GVIF |
|---|---|---|---|---|---|
| (Intercept) | 13.015*** | 0.065 | 199.772 | 0.000 | - |
| net.area | 0.004*** | 0.000 | 13.863 | 0.000 | 2.051 |
| bedroom | -0.013 | 0.015 | -0.862 | 0.389 | 1.932 |
| livingroom | -0.048 | 0.052 | -0.916 | 0.360 | 1.059 |
| bath: more.than.one | 0.286*** | 0.019 | 14.870 | 0.000 | 1.409 |
| floor: ground.floor | -0.212*** | 0.020 | -10.461 | 0.000 | |
| floor: top.floor | -0.073*** | 0.019 | -3.912 | 0.000 | 1.070 |
| floor: basement.floor | -0.447*** | 0.024 | -18.323 | 0.000 | |
| total.floor | 0.024*** | 0.002 | 15.813 | 0.000 | 1.206 |
| age | -0.007*** | 0.001 | -9.138 | 0.000 | 1.347 |
| carpark.indoor: yes | 0.170*** | 0.018 | 9.570 | 0.000 | 1.172 |
| carpark.outdoor: yes | -0.035* | 0.015 | -2.328 | 0.020 | 1.042 |
| gated.community: yes | -0.010 | 0.016 | -0.663 | 0.507 | 1.016 |
| num.views | -0.007 | 0.008 | -0.879 | 0.379 | 1.019 |
| dist.transportation | -0.015 | 0.010 | -1.518 | 0.129 | 4.396 |
| dist.school | 0.210*** | 0.011 | 18.965 | 0.000 | 5.189 |
| dist.mall | -0.160*** | 0.008 | -19.614 | 0.000 | 4.018 |
| dist.hospital | -0.091*** | 0.011 | -8.619 | 0.000 | 4.832 |
| dist.university | -0.049*** | 0.008 | -6.395 | 0.000 | 3.339 |
| dist.police | 0.060*** | 0.007 | 8.394 | 0.000 | 4.169 |

| # of Observations | 2784 | Global Moran I for regression residuals | |
|---|---|---|---|
| $R^2$ | 0.6974 | Moran's I Index | 0.149 |
| Adj. $R^2$ | 0.6953 | Expected Index | -0.001 |
| RSS | 350.0301 | Variance | 0.000 |
| F-statistic | 335.2 | p-value | 0.000 |
| p-value | 0.000 | | |

*Note: ***, **, and * represent significance at 1%, 5% and 10% respectively*

Two GWR models were estimated with a Gaussian kernel weighting function with adaptive bandwidth. Only difference between the two models is the adopted approach to find the optimal bandwidth. The cross-validation (CV) method in the first GWR model and the Akaike Information Criterion (AIC) in the second were used to obtain an optimum size of nearest neighbors for the adaptive kernel. The resulting optimal bandwidths for the first and second models were found as 45 and 59 observations respectively. Together with the global OLS regression, both GWR models results are summarized in Table 5. However, the GWR 1 model with cross-validation approach fit the data much better than both GWR 2 model with AIC approach and the OLS model. As seen in Table 4, among the three models, the GWR 1 model has the lowest AIC and residual sum of squares (RSS), and the highest R2 and adjusted R2 values. Therefore, all evaluations are made based on the results of the GWR model with CV approach in the rest of the paper.

**TABLE 4 |** Comparison of the OLS and GWR Models

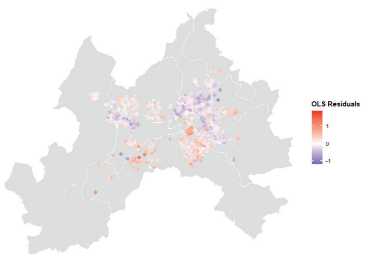
| | Approach | Adaptive Bandwidth | $R^2$ | Adj. $R^2$ | AIC | RSS |
|---|---|---|---|---|---|---|
| OLS | | | 0.697 | 0.695 | 2.169.678 | 350.030 |
| **GWR 1** | **CV** | **45** | **0.901** | **0.866** | **-436.021** | **114.824** |
| GWR 2 | AIC | 59 | 0.889 | 0.860 | -232.054 | 128.918 |

The findings showed that the GWR model improve the adjusted $R^2$ values from 0.69 to 0.86. The spatial distribution of the residuals of the standard OLS and GWR models depicted in Figure 2 suggests that the GWR residuals are closer to zero compared to the OLS residuals.

The local $R^2$ values are higher than 0.76, indicating that the selected variables in this study have quite able to explain the housing prices in the study area. Figure 3 shows the spatial distribution of the goodness to fit between the housing attributes and prices based on the results obtained from the GWR model. The local explanatory powers are relatively higher in the west areas far from the city center. These results might present that the selected variables can explain the variation in the housing prices on the west side while in the central business area further determinants might also impact the house prices. For instance, majority of the buildings in the central business district are older than those on the west side, where the city tends to grow towards, and gated communities are more popular. Another example is that rising oil prices might be a significant driver in people's housing choices, and thus on the house prices, as commuting has become a significant component of household expenditures recently (Akkoç, Akçağlayan, & Kargın Akkoç, 2021). The variation in the local $R^2$ values is an indicator of the existence of spatial heterogeneity in the study area.

**FIGURE 2 |** Model Residuals

(OLS)




(GWR)

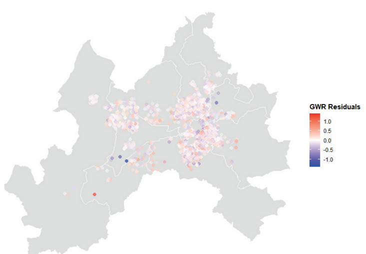
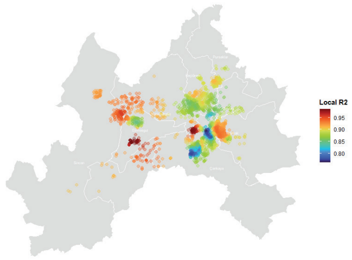



**FIGURE 3 |** Local $R^2$ values derived from the GWR model




Estimated coefficients from the GWR model are generated for each point differently as the influence of independent variables on dwelling prices is varied over the space. Minimum, maximum and quartile values of these estimations are shown in Table 5 together with the OLS regression results. In addition to that the distribution of selected regression coefficients obtained from the GWR model, and by following Morali and Yilmaz (2020), Spearman Correlations between the logarithm of house prices and these variables are depicted on Figure 4. The geographically weighted Spearman Correlations were computed by using the same bandwidth specification in the GWR Model. As illustrated in Figure 4, that the spatial patterns of the parameter estimations for each variable and the correlations differ over the space suggest the spatial heterogeneity in the study area.

As seen in Table 5, there are variabilities in coefficient estimations for all variables in terms of their impact of both magnitude and direction. For instance, coefficient estimations for the net area of the dwellings vary from -0.001 to 0.012; similarly, this variation range is between -0.036 and 0.003 for the building age variable and -0.069 and 0.104 for the number of views of the dwellings. The GWR coefficients being positive and negative reveals the fact that drivers of the housing prices have an obvious spatial heterogeneity and instability over the study area. Furthermore, many of the median coefficients differ from the OLS coefficients, and even the signs of the three out of coefficients in the OLS regression are not the same as the GWR's.

Majority of parameter values for the distance measures have negative signs, which means that the house prices decrease as the points of interests are getting further away them. On the other hand, their effect on the house prices is larger in the west areas where are far away from the central business area. Consistent with the related literature (e.g. Wang and Chen (2020), Morali and Yilmaz (2020), and Soltani et al. (2021)), this result supports the significant impact of accessibility to various points of interests such as hospitals and schools on residential property prices.

Similarly, there is a negative correlation between the building age and house prices. As seen on Figure 4 (b), the effect of the building age is less notable in the older neighborhoods where are closer to the central business area than it is in those areas on the far western and north-eastern side of the city.

By contrast, the dwelling size variables (i.e., number of bedrooms, net surface area, and extra bathroom) and existence of a car park have positive impacts on the house prices. Compared to the overall picture, it is noteworthy that the effects of the number of bedrooms and the net surface area seem stronger in a particular micro area in the district of Çankaya, as seen on Figure 4 (a and c).

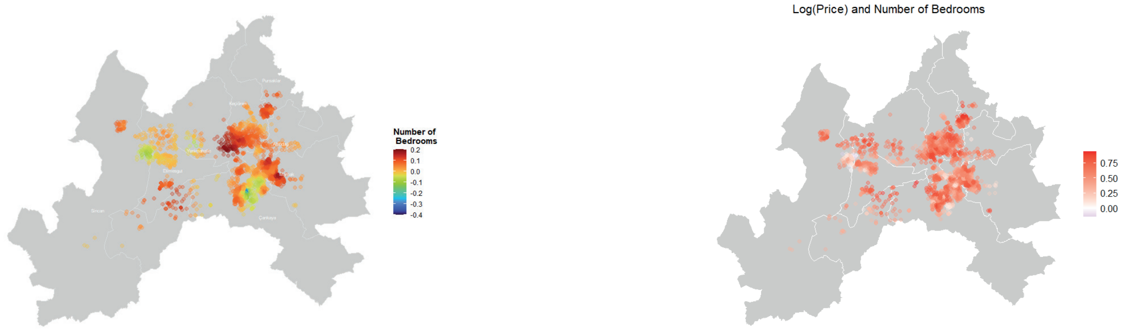
**TABLE 5 |** GWR Models and Global OLS

| | GWR 1 (CV approach) | | | | | GWR 2 (AIC approach) | | | | | Global OLS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Min. | 1st Qu. | Median | 3rd Qu. | Max. | Min. | 1st Qu. | Median | 3rd Qu. | Max. | |
| (Intercept) | 2.119 | 12.551 | 12.828 | 13.160 | 78.088 | 11.047 | 12.624 | 12.832 | 13.184 | 32.088 | 13.015 |
| net.area | -0.001 | 0.002 | 0.003 | 0.004 | 0.012 | 0.000 | 0.002 | 0.003 | 0.004 | 0.009 | 0.004 |
| bedroom | -0.401 | 0.007 | 0.060 | 0.093 | 0.205 | -0.234 | 0.005 | 0.053 | 0.094 | 0.192 | -0.013 |
| livingroom | -14.027 | -0.140 | 0.039 | 0.184 | 0.969 | -0.999 | -0.111 | 0.040 | 0.165 | 0.431 | -0.048 |
| bath: more.than.one | -0.070 | 0.094 | 0.147 | 0.210 | 0.400 | -0.027 | 0.104 | 0.155 | 0.207 | 0.396 | 0.286 |
| floor: ground.floor | -0.472 | -0.348 | -0.271 | -0.205 | 0.225 | -0.444 | -0.347 | -0.270 | -0.209 | 0.093 | -0.212 |
| floor: top.floor | -0.278 | -0.105 | -0.073 | -0.034 | 0.119 | -0.199 | -0.106 | -0.075 | -0.040 | 0.060 | -0.073 |
| floor: basement.floor | -0.856 | -0.566 | -0.481 | -0.410 | 0.679 | -0.774 | -0.564 | -0.476 | -0.413 | -0.097 | -0.447 |
| total.floor | -0.023 | 0.018 | 0.025 | 0.031 | 0.053 | -0.005 | 0.020 | 0.025 | 0.030 | 0.050 | 0.024 |
| age | -0.036 | -0.016 | -0.011 | -0.008 | 0.003 | -0.034 | -0.015 | -0.011 | -0.008 | 0.002 | -0.007 |
| carpark.indoor: yes | -0.105 | 0.026 | 0.071 | 0.125 | 0.384 | -0.068 | 0.035 | 0.075 | 0.123 | 0.360 | 0.170 |
| carpark.outdoor: yes | -0.195 | -0.019 | 0.005 | 0.036 | 0.175 | -0.177 | -0.016 | 0.004 | 0.032 | 0.137 | -0.035 |
| gated.community: yes | -0.141 | -0.047 | -0.021 | 0.009 | 0.112 | -0.096 | -0.043 | -0.019 | 0.006 | 0.052 | -0.010 |
| num.views | -0.069 | -0.021 | -0.002 | 0.013 | 0.104 | -0.061 | -0.019 | -0.002 | 0.012 | 0.058 | -0.007 |
| dist.transportation | -76.786 | -0.141 | -0.033 | 0.068 | 2.473 | -1.223 | -0.109 | -0.020 | 0.070 | 1.890 | -0.015 |
| dist.school | -552.540 | -0.112 | 0.083 | 0.262 | 3.246 | -1.420 | -0.076 | 0.127 | 0.301 | 2.545 | 0.210 |
| dist.mall | -2.735 | -0.208 | -0.096 | -0.019 | 304.215 | -2.140 | -0.213 | -0.118 | -0.045 | 1.430 | -0.160 |
| dist.hospital | -1.741 | -0.139 | -0.056 | 0.054 | 409.446 | -1.448 | -0.138 | -0.071 | 0.014 | 0.850 | -0.091 |
| dist.university | -2.022 | -0.158 | -0.021 | 0.068 | 38.315 | -1.666 | -0.135 | -0.024 | 0.055 | 3.673 | -0.049 |
| dist.police | -170.900 | -0.103 | -0.007 | 0.089 | 3.612 | -5.637 | -0.084 | -0.013 | 0.074 | 1.923 | 0.060 |

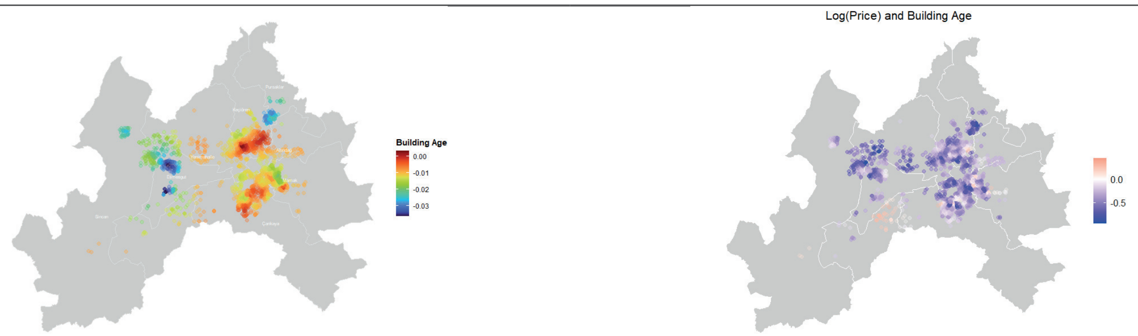
| | GWR 1 | GWR 2 | Global OLS |
|---|---|---|---|
| # of Observations | 2784 | 2784 | 2784 |
| $R^2$ & Adj. $R^2$ | 0.900 & 0.866 | 0.889 & 0.860 | 0.697 & 0.695 |
| Adaptive Bandwidth | 45 | 59 | |
| Effective # of Parameters | 715.615 | 565.573 | |
| Residual Sum of Squares | 114.824 | 128.918 | 350.030 |

**FIGURE 4 |** Spatial Distribution of GWR Coefficients and Spearman Correlations (1/2)

Coefficient | Spearman Correlation




(a) Number of Bedrooms (bedroom)




(b) Building Age (age)

**FIGURE 4 |** Spatial Distribution of GWR Coefficients and Spearman Correlations (2/2)

Coefficient | Spearman Correlation


(c) Net Surface Area (net.area)


(d) Distance to Transportation (dist.transport)

## CONCLUSIONS

It is well known that real estate markets are characterized by both spatial heterogeneity and spatial autocorrelation. This paper accounts for the spatial heterogeneity in the housing market in Ankara metropolitan area by using geographically weighted regression. Taking several explanatory variables representing the structural and locational attributes of dwellings, the GWR model with a Gaussian kernel weighting function with adaptive bandwidth based on cross-validation approach fit the data much better than the traditional OLS in explaining the housing prices and raised the adjusted R2 significantly from 0.695 to 0.866. Spatial variations in parameter estimations and R2 values confirmed the spatial heterogeneity in the market, which was also supported by the spatial correlations depicted on the map.

Hedonic models are still the most prominent approach in automated valuations of real estates due to their simplicity and easy interpretability (Matysiak, 2017). However, as seen in the previous studies in the literature summarized in Section I, and suggested in this study as well, the models considering spatial dependence and spatial heterogeneity can give more accurate estimations in real estate market analysis. Furthermore, developing automated valuation models by applying machine learning algorithms has been under the focus of both professionals and academics recently to make more accurate price predictions for various purposes ranging from collateral valuations to portfolio assessments (e.g. Steurer et al. (2021) and Tchuente and Nyawa (2021)). Thus, using machine learning methods that consider spatial heterogeneity and spatial autocorrelation could be recommended for further studies on housing price predictions.

## REFERENCES

1. Akkoç, U., Akçağlayan, A., & Kargın Akkoç, G. (2021). The impacts of oil price shocks in Turkey: sectoral evidence from the FAVAR approach. *Economic Change and Restructuring,* 54(4), 1147-1171.
2. Bidanset, P., McCord, M., Lombard, J. A., Davis, P., & McCluskey, W. (2018). Accounting for locational, temporal, and physical similarity of residential sales in mass appraisal modeling: the development and application of geographically, temporally, and characteristically weighted regression. *Journal of Property Tax Assessment and Administration,* 14(2), 5-13.
3. Bidanset, P. E., & Lombard, J. R. (2014). Evaluating spatial model accuracy in mass real estate appraisal: A comparison of geographically weighted regression and the spatial lag model. *Cityscape,* 16(3), 169-182.
4. Bidanset, P. E., Lombard, J. R., Davis, P., McCord, M., & McCluskey, W. J. (2017). Further evaluating the impact of kernel and bandwidth specifications of geographically weighted regression on the equity and uniformity of mass appraisal models. In *Advances in Automated Valuation Modeling* (pp. 191-199): Springer.
5. Bitter, C., Mulligan, G. F., & Dall'erba, S. (2007). Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method. *Journal of geographical systems,* 9(1), 7-27.
6. Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis,* 28(4), 281-298.
7. Bujanda, A., & Fullerton, T. M. (2017). Impacts of transportation infrastructure on single-family property values. *Applied Economics,* 49(51), 5183-5199.
8. Chica-Olmo, J. (2007). Prediction of housing location price by a multivariate spatial method: Cokriging. *Journal of Real Estate Research,* 29(1), 91-114.
9. Crawford, T. W. (2009). Scale Analytical. In R. Kitchin & N. Thrift (Eds.), *International encyclopedia of human geography* (pp. 29-36): Elsevier.
10. Davis, P., Bidanset, P., McCord, M., & Cusack, M. (2019). Nationwide Mass Appraisal Modeling in China: Feasibility Analysis for Scalability Given Ad Valorem Property Tax Reform. *Working Paper WP20PD1, Lincoln Institute of Land Policy.*
11. Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2000). *Quantitative geography: perspectives on spatial data analysis:* Sage.
12. Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). *Geographically weighted regression: the analysis of spatially varying relationships:* John Wiley & Sons.
13. Fotheringham, A. S., Charlton, M. E., & Brunsdon, C. (1998). Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and planning* A, 30(11), 1905-1927.
14. Fotheringham, A. S., Crespo, R., & Yao, J. (2015). Exploring, modelling and predicting spatiotemporal variations in house prices. *The Annals of Regional Science,* 54(2), 417-436.
15. Fullerton, T. M., & Bujanda, A. (2018). Commercial property values in a border metropolitan economy. *Asia-Pacific Journal of Regional Science,* 2(2), 337-360.
16. Gloudemans, R. J., & Almy, R. R. (2011). *Fundamentals of mass appraisal:* International Association of Assessing Officers.
17. Grover, R. (2016). Mass valuations. *Journal of Property Investment & Finance, Emerald Group Publishing, Vol. 34(2),* pages 191-204.
18. Guo, L., Ma, Z., & Zhang, L. (2008). Comparison of bandwidth selection in application of geographically weighted regression: a case study. *Canadian Journal of Forest Research, 38*(9), 2526-2534.
19. Hanink, D. M., Cromley, R. G., & Ebenstein, A. Y. (2012). Spatial variation in the determinants of house prices and apartment rents in China. *The Journal of Real Estate Finance and Economics, 45*(2), 347-363.

20. Hu, L., He, S., Han, Z., Xiao, H., Su, S., Weng, M., & Cai, Z. (2019). Monitoring housing rental prices based on social media: An integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies. *Land Use Policy, 82,* 657-673.
21. Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science, 24*(3), 383-401.
22. Matysiak, G. A. (2017). The accuracy of automated valuation models (AVMs). *Report for TEGoVA.*
23. Militino, A. F., Ugarte, M. D., & Garcia-Reinaldos, L. (2004). Alternative models for describing spatial dependence among dwelling selling prices. *The Journal of Real Estate Finance and Economics, 29*(2), 193-209.
24. Mooya, M. M. (2016). *Real Estate Valuation Theory:* Springer.
25. Morali, O., & Yilmaz, N. (2020). Spatial Heterogeneity in Istanbul Housing Market: A Geographically Weighted Approach. *Journal of Economics Finance and Accounting, 7*(4), 298-307.
26. Páez, A., Long, F., & Farber, S. (2008). Moving window approaches for hedonic price estimation: an empirical comparison of modelling techniques. *Urban Studies*, 45(8), 1565-1581.
27. RICS. (2017). *The Future of Valuations: The relevance of real estate valuations for institutional investors and banks – views from a European expert group, London, UK.* Retrieved from RICS.
28. Sayın, Z. M. (2021). *House price dynamics in izmir's neighborhoods: A comprehensive approach, Unpublished master's thesis,.* Izmir Institute of Technology,
29. Schulz, R., Wersing, M., & Werwatz, A. (2014). Automated valuation modelling: a specification exercise. *Journal of Property Research, 31*(2), 131-153.
30. Sirmans, S., Macpherson, D., & Zietz, E. (2005). The composition of hedonic pricing models. *Journal of Real Estate Literature, 13*(1), 1-44.
31. Sisman, S., & Aydinoglu, A. (2022). A modelling approach with geographically weighted regression methods for determining geographic variation and influencing factors in housing price: A case in Istanbul. *Land Use Policy, 119*, 106183.
32. Soltani, A., Pettit, C. J., Heydari, M., & Aghaei, F. (2021). Housing price variations using spatio-temporal data mining techniques. *Journal of Housing and the Built Environment, 36*(3), 1199-1227.
33. Steurer, M., Hill, R. J., & Pfeifer, N. (2021). Metrics for evaluating the performance of machine learning based automated valuation models. *Journal of Property Research,* 1-31.
34. Tchuente, D., & Nyawa, S. (2021). Real estate price estimation in French cities using geocoding and machine learning. *Annals of Operations Research,* 1-38.
35. Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography, 46*(sup1), 234-240.
36. Wang, C.-H., & Chen, N. (2020). A geographically weighted regression approach to investigating local built-environment effects on home prices in the housing downturn, recovery, and subsequent increases. *Journal of Housing and the Built Environment, 35*(4), 1283-1302.
37. Yu, D. (2007). Modeling owner-occupied single-family house values in the city of Milwaukee: A geographically weighted regression approach. *GIScience & Remote Sensing, 44*(3), 267-282.