IDENTIFYING GEOGRAPHIC DETERMINANTS OF HOUSING PRICES IN PINELLAS COUNTY, FLORIDA: A GEOGRAPHICALLY WEIGHTED REGRESSION AND MULTISCALE GEOGRAPHICALLY WEIGHTED REGRESSION APPROACH

A RESEARCH PAPER PRESENTED TO

THE DEPARTMENT OF HUMANITIES AND SOCIAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

MASTER OF SCIENCE

By

NICHOLAS HORTON

NORTHWEST MISSOURI STATE UNIVERSITY

MARYVILLE, MISSOURI

APRIL, 2023

**1. Introduction**

Understanding the factors that influence real estate prices is important for planning and development, economic indicators, and making informed investments. These factors include economic conditions, social factors, and geographic factors. While most of the focus of real estate studies is on the structural elements of a house, the economic conditions, and social factors, there needs to be more focus on what geographic factors affect housing prices. These geographic factors include terrain, transportation infrastructure, proximity to amenities, and distance to a city center. The most common approach to modeling the factors that affect housing prices is using a hedonic model. A hedonic model is a statistical model that estimates the value of a good based on its characteristics and surroundings (Rosen 1974). In real estate, this model assesses a property’s value based on the utility of its characteristics given to the buyer. These characteristics can include square footage, number of bedrooms and bathrooms, the size of the lot, and other characteristics that would satisfy the buyer. A hedonic model typically takes the form of a regression model which estimates the relationship between the home’s value and its physical characteristics.

The location of a house can also be used in a hedonic model. For example, Anas *et al.* (1998)found that housing prices gradually declined when the distance from the city center increased. While these models were accurate for estimating real estate values, it assumed that these characteristics are constant for all locations. Typically, these models did not consider the spatial dependence of the data. The value of a home is often spatially autocorrelated, meaning a house’s value is influenced by other houses in nearby locations (Kokot and Gnat 2019). Property value in a neighborhood can be spatially autocorrelated based on several factors such as proximity to schools, access to transportation, crime rates, and proximity to parks. These house values would likely be similar based on their prime locations.

On the other hand, homes that are a large distance from necessary amenities and located in a high crime area would likely share a lower value. Spatial autocorrelation can also occur in coastal communities where property located along the coast is valued much higher than property farther inland due to easy access to the beach and view of the ocean. Spatial autocorrelation can also be influenced by demographic factors. Neighborhoods that are homogenous in income or ethnicity can be of higher value to buyers. Because property values can vary depending on the location, it is essential to incorporate spatial autocorrelation into a hedonic model. Failure to do so can lead to inaccurate and biased results (Helbich *et al.* 2013).

Incorporating spatial autocorrelation can be done by using several global models such as a Spatial Error Model (SLM) or a Spatial Lag Model (SEM). While these models can account for spatial autocorrelation, they assume the relationship between the independent and dependent variables are constant throughout the study area (Helbich *et al.* 2013*).* They do not account for the spatial heterogeneity of the data. This is where local models such as Geographically Weighted Regression (GWR) and Multiscale Geographically Weighted Regression (MGWR) are valuable.

A local spatial model can account for the spatial heterogeneity of the variables across the study area. Local models create neighborhoods to capture the local variations of the relationships between the variables. GWR allows the relationship between the explanatory variables and the dependent variable to vary spatially. For example, the relationship between the value of the house and its physical characteristics can vary depending on the location of the house. GWR can give a better understanding of the factors that influence the value of a house in different regions of the study area. MGWR takes this a step further by allowing the variables to vary at multiple different scales across the study area (Sisman and Aydinoglu 2022). For example, the relationship between distance to public transportation and the value of a house may be different at a neighborhood level than at a city level. Using local spatial models, such as GWR and MGWR, can provide more accurate results based on its ability to capture the spatial heterogeneity of the data (Helbich *et al.* 2013). To determine what geographic features affect housing prices, it is important to understand which factors are more influential in different regions of the study area.

**1.1. Research objective**

The objective of this study is to examine the relationship between geographic features and housing prices with the goal of identifying what geographic features significantly influence housing prices. This study also aims to determine if the influence of geographic features vary across Pinellas County, Florida using GWR and MGWR. Furthermore, this study will compare GWR and MGWR to determine which technique gives more accurate and reliable results in determining what geographic features influence housing prices.

**2. Literature review**

**2.1. *Regression analysis in Geographic studies***

Regression analysis has been used in many geographic studies to assess the relationship between geographic variables and has shown to be adequate for many different scenarios. Rai and Nathaway (2013) used Multiple Linear Regression (MLR), along with other statistical models, to create a malaria susceptibility model. The authors found that while MLR was not the best performing model, but it was still suitable for their study.

Lopes *et al*. (2014) took MLR a step further by incorporating spatial dependence into their models. Multiple MLR models were implemented for transport planning. The results show that by incorporating spatial dependence into the models, they produced better results. This was not the case for Rai and Nathaway (2013). Spatial dependence was not included in the model, and it was outperformed. If this was included in the model, the results may have been similar to what Lopes *et al.* (2014) found in their study.

Wright *et al*. (2014) used logistic regression (LR) to determine if archaeological sites were built with the ability to be seen from the surrounding landscapes. The regression analysis performed in this study aimed the authors in confirming their theory by showing that slope and elevation played an essential role in selecting the site’s locations.

Regression analysis has been shown to be effective in modeling the relationships between geographic features (Rai and Nathaway 2013, Lopes *et al*. 2014). Regression analysis is able to capture the relationship between the dependent and independent variables in many different scenarios and can be applied to many different aspects of geographic studies.

**2.2.** ***Regression analysis in hedonic models***

Regression analysis has also been implemented in hedonic models to model real estate prices. Lins *et al.* (2005) compared MLR to a Double Perspective-Data Envelopment Analysis (DP-DEA) approach to estimating housing prices. The authors concluded that MLR was outperformed by DP-DEA. DP-DEA is based on a data envelopment analysis which is where the factors being investigated comprise the transactions between buyers and sellers. MLR was outperformed due to DP-DEA being a non-linear model. The Index of Relative Distance between Frontiers (IRDF) was used as the model evaluation metric in this study. IRDF is a measure of absolute performance over time. Several iterations, or rounds, were performed and MLR performed adequately but in the 13th round MLR failed to reach the goal of an IRDF of less than 25%.

Ibeas *et al*. (2012) also used MLR along with spatial regression models to estimate housing prices in metropolitan areas. The spatial regression models produced a better goodness-of-fit to the data. It was found that when there was a strong spatial dependence in the residuals, MLR did not perform as well as the other models. This is consistent with what Kokot and Gnat (2019) found in their study. The authors ultimately concluded that MLR is a viable model for estimating real estate prices, but there must be a sufficient number of transactions (i.e., large sample size) and a market where housing prices mirror the market attributes. They also raised several problems with MLR in hedonic models. The explanatory variables need to be chosen carefully. While it can be necessary to have numerous explanatory variables, including variables that do not influence housing prices is not good practice. Furthermore, the quality of the data plays a large role in the outcome of the model. For example, modeling house prices for a small market with few transactions can lead to a poor performing model. It was also stated that the relationship between housing prices and the explanatory variables may not be linear. This would indicate that MLR would not be sufficient to model this relationship since MLR assumes a linear relationship of the data. Finally, real estate markets are not homogeneous in nature and spatial autocorrelation is likely to be present. The authors use an example of more valuable houses that are more likely to be near other valuable houses. MLR is not able to capture the spatial dependence of the data which can lead to unreliable results (Kokot and Gnat 2019).

These previous studies showed the complexity of building a hedonic model and that it is difficult for MLR to capture the variation in the dependent variable. While MLR has its limitations in hedonic modeling, it has also shown to be effective in other studies such as Zhang *et al*. (2019). The authors used MLR to predict housing prices in Boston, Massachusetts. It was concluded that MLR is able to analyze and predict housing prices but is limited in its prediction accuracy. This model still needs to be researched more and suggests implementing a more advanced modeling technique such as machine learning. Li (2022) also found that MLR is effective for modeling housing prices. The author used a generalized linear regression approach to predict real estate prices. This study plotted the fitting curve on the regression model against the real curve of the data and found them to be relatively close. The author concluded that the generalized regression model has high prediction accuracy. Furthermore, Liu (2022) also implemented MLR for housing price prediction in four Chinese cites from 2011 to 2017. It was concluded that MLR was able to accurately predict housing prices with an error of price prediction of 7.6%.

Hedonic modeling is not an easy process. There are many factors that influence the real estate market and capturing this in a statistical model can be a difficult process; it is complex and can vary widely in different study areas. For example, Helbich *et al*. (2013) suggested that the real estate market in Austria is impacted by national policies such as government subsidies which was also confirmed by Liu (2022). Each study area has different geographic factors that affect the real estate market. For example, in coastal communities, distance to the beach can be the most influential determinate on the real estate market (Chen and Fik 2017). However, commercial centers can be the most influential factor on real estate prices in urban areas (Zhang *et al*. 2019). A study area can exhibit many different types of geographic features that affect housing prices meaning these features are non-stationary across the study area. For example, a study area may contain an urban center and coastal communities which will have different factors that affect their housing prices. Housing prices can also be spatial autocorrelated. While the non-stationarity and spatial autocorrelation of the data are crucial aspects that affect housing prices, they are not accounted for in global models.

**2.3.** ***Local spatial regression models***

To capture the spatial dependence and the non-stationarity of the data in a linear regression model, GWR and MGWR are commonly implemented in hedonic models. Iliopoulou and Feloni (2022) used Kriging techniques and hot spot analysis in conjunction with GWR to model housing prices in Athens, Greece. Kriging, which is a spatial interpolator, was used to create a geovisualization of the data. This gave the authors the ability to describe the spatial pattern of the house characteristics. This was then compared to the hot spot analysis to assess the spatial patterns. These two geovisualization techniques gave the authors a clear visual way to see the spatial differentiation within the study area. Finally, several regression models were created to model housing prices within the study area. Ordinary Least Squares (OLS) was first used to test if the residuals were random or clustered. If they were observed to be clustered, the authors would implement GWR. A stepwise method was used to select the appropriate variables to use in OLS and it was found to be a moderate fit to the data with a R squared of .583. GWR was found to be a much better fit to the data with an R squared of .848. The higher R squared is due to GWR being a local regression model that allows the regression coefficients to vary spatially. It was suggested that spatial regression techniques, such as GWR can improve the accuracy of hedonic models. Wang *et al.* (2023) also found similar results in their study. GWR and Geographically Temporally Weighted Regression (GTWR) were used with OLS for real estate appraisal in urban centers. GTWR was used to assess real estate prices over several years. OLS was first used to determine the statistically significant variables. Then GWR and GTWR models were created with the statistically significant variables identified by OLS. GTWR was shown to be the best performing model with a R squared of .8192. GWR did not perform as well due to the temporal factor of the data. It was concluded that a local spatial regression model produces better results than global non spatial models.

These two studies (Iliopoulou and Feloni 2022, Wang *et al.* 2023) incorporated GWR and that it was more capable than the other models GWR was compared against. Helbich *et al*. (2013) and Sisman and Aydinoglu (2022) took GWR a step further by applying MGWR. Both studies compared GWR and MGWR to non-spatial and spatial global models and it was found that the local spatial models produced better results, specifically MGWR. Helbich *et al*. (2013) concluded that both GWR and MGWR were a better fit to the data than the other models they were compared against. MGWR was the best performing model used in this study. Due to the flexibility of MGWR, it was able to capture the spatial variation of the explanatory variables across the study area better than GWR. Sisman and Aydinoglu (2022) concluded that both GWR and MGWR were able to better explain the relationship between the dependent variable and independent variables. GWR and MGWR showed very similar results. GWR had a slightly higher R squared, while MGWR produced a lower Akaike Information Criterion (AIC).

Regression analysis is a useful statistical technique to examine the relationship between the dependent variable and the explanatory variables. While it has been useful in many geographic studies, traditional regression analysis, such as MLR, has had its downfalls when used in hedonic models as shown by Kokot and Gnat (2019). This is in part due to these models not accounting for the spatial dependence of the data. Spatial regression models, specifically local models, have been shown to be more adequate at explaining the variation of the dependent variable in hedonic models. Although local spatial models have shown to be the best performing models, they are often used with global models, such as OLS, to assess what explanatory variables to implement in the model and examine if multicollinearity issues exist between the variables (Hong and Yoo 2020, Wang *et al.*, 2020, Iliopoulou and Feloni 2022, Sisman and Aydinoglu 2022).

**3. Methodology**

This study is a quantitative analysis of determining what geographic features affect housing prices using regression analysis techniques. The methodology used in this study started with determining the study area and obtaining a dataset that included the price the house sold for. The data was then pre-processed and cleaned for further analysis. This included filling in missing or null values, removing outliers, and removing records that do not pertain to this study. Global Moran’s I was then used to determine if the houses were spatially autocorrelated. Next, the clusters of high and low values were mapped for selected variables. Global and local regression models were then applied to the final dataset. Ordinary Least Squares (OLS) was used as the global model and GWR and MGWR were used as the local models. The global and local model’s results were then compared and analyzed from mapping the coefficients. All processes and analysis for this study were conducted in ArcGIS Pro version 3.0.2. An overview of the methodology used in this study can be seen in Figure 1.

**3.1. *The Study Area***

The study area chosen for this study is Pinellas County, Florida. Pinellas County is located on the west central coast of Florida, just to the west of Tampa Bay. It is approximately 280 square miles with 588 miles of shoreline and is made up of 24 municipalities. While it is not the largest county in the state of Florida, it is the most densely populated with a population of

Diagram

Description automatically generated

Figure 1: Methodology flowchart

over 950,000 residents (Pinellas County 2022). St. Petersburg is the county’s largest city located in the southern portion of the county. Clearwater, Largo, and Dunedin are other notable cities. Pinellas County is a geographically rich area that offers residents and tourists a wide range of amenities such as beaches, parks, and natural areas. The county is known for its beautiful beaches and its barrier island that runs the length of the county and provides abundant area for waterfront properties. Along with the shoreline, Pinellas County also provides inland features such as lakes, parks, trails, and natural areas. This gives residents access to a multitude of outdoor activities. The county also has an extensive public transportation system with bus routes provided by the Pinellas Suncoast Transit Authority (PSTA). PSTA runs 41 routes with 4,395 bus stops throughout the county and has an average of 9.1 million service miles per year (Pinellas Suncoast Transit Authority 2023). The county also includes several major highways including I-275 which runs from St. Petersburg to Tampa Bay, and US-19 which runs from north to south through the county.

The county is made up of suburban and urban areas that each have their own distinct characteristics and housing market. Clearwater Beach is located on the barrier island and is near Caladesi Island and Honeymoon Island State Park. This area offers different characteristics from urban areas within the county such as downtown St. Petersburg which gives way to these areas having different housing markets. The study area can be seen in Figure 2.

Map

Description automatically generated

Figure 2. The study area (Pinellas County, Florida)

**3.2. *Data sources and dataset***

The housing dataset was obtained from Pinellas County Property Appraiser (PCPA) website (https://www.pcpao.gov/). An advanced data search was used based on residential houses for the property type, and all sales in 2022 were selected for transactions. For the selected output fields, the fields that were not relevant to this study were left out of the data set such as, owner and tax information. The final fields that were selected can be seen in Table 1. The data was then downloaded as a Comma-Separated Values (CSV) file.

Table 1. Housing data set acquired from Pinellas County Property Appraiser (PCPA)

|  |  |  |
| --- | --- | --- |
| Field Name | Description | Data Type |
| Parcel Number | Parcel identification number for each house | Numeric |
| Property Address | Address of each house | Text |
| City and zip code | City and zip code each house is located in | Text |
| Land Area (Acres) | How many Acres the house possesses | Numeric |
| Property use description | The type of residential property. i.e., condo, single-family home | Text |
| Total gross area in square feet | The total square feet of the house | Numeric |
| Total living (heated) area in square feet | The total living area for each house | Numeric |
| Year built | The year the hose was built | Numeric |
| House price | The price the house sold for | Numeric |

To obtain geographic features within the study area, shapefiles were downloaded from multiple GIS portals: Florida’s Geospatial Open Data (https://geodata.floridagio.gov/), Pinellas County Enterprise GIS (https://new-pinellas-egis.opendata.arcgis.com/), and Florida Fish and Wildlife Conservation Commission (https://geodata.myfwc.com). These features were used as explanatory variables along with the total living square feet, acreage, and the age of the home. These variables can be seen in Table 2.

To map the location of the houses, a parcel shapefile was also obtained from the PCPA website. This is a point shapefile that contains every parcel within the study area. The housing

Table 2. The variables used in this study with their description and data type

|  |  |  |  |
| --- | --- | --- | --- |
| Explanatory Variables | Description | Source | Data Type |
| Acreage | Acreage of house | Pinellas County Property Appraiser | Continuous variable |
| House age | Age of house | Pinellas County Property Appraiser | Continuous variable |
| Total living square footage | Square footage of house | Pinellas County Property Appraiser | Continuous variable |
| Located in St. Petersburg | Describes if the house is located in the city of St. Petersburg | Created in ArcGIS Pro | Dummy variable |
| Located on barrier island | Describes if the house is located on the barrier island | Created in ArcGIS Pro | Dummy variable |
| Proximity to beach access | Distance to public beach access points (U.S. feet) | Florida’s Geospatial Open Data | Continuous variable |
| Proximity to beach | Distance to beach (U.S. feet) | Florida Fish and Wildlife Conservation Commission | Continuous variable |
| Proximity to boat ramp | Distance to boat ramps (U.S. feet) | Pinellas County Enterprise GIS | Continuous variable |
| Proximity to bridges | Distance to bridges on barrier island (U.S. feet) | Created in ArcGIS Pro | Continuous variable |
| Proximity to bus stops | Distance to bus stops (U.S. feet) | Pinellas County Enterprise GIS | Continuous variable |
| Proximity to downtown St. Petersburg | Distance to city center of St. Petersburg (U.S. feet) | Created in ArcGIS Pro | Continuous variable |
| Proximity to Golf Course | Distance to golf courses (U.S. feet) | Pinellas County Enterprise GIS | Continuous variable |
| Proximity to health care sites | Distance to health care sites (U.S. feet) | Pinellas County Enterprise GIS | Continuous variable |
| Proximity to historic districts | Distance to historic districts (U.S. feet) | Pinellas County Enterprise GIS | Continuous variable |
| Proximity to lakes | Distance to lakes (U.S. feet) | Florida’s Geospatial Open Data | Continuous variable |
| Proximity to marinas | Distance to marinas (U.S. feet) | Florida Fish and Wildlife Conservation Commission | Continuous variable |
| Proximity to parks | Distance to parks (U.S. feet) | Pinellas County Enterprise GIS | Continuous variable |
| Proximity to schools | Distance to schools (U.S. feet) | Pinellas County Enterprise GIS | Continuous variable |
| Proximity to shoreline | Distance to shorelines other than the beach. (U.S. feet) | Florida Fish and Wildlife Conservation Commission | Continuous variable |
| Proximity to sports facilities | Distance to sports facilities (U.S. feet) | Pinellas County Enterprise GIS | Continuous variable |
| Proximity to major roads | Distance to major roads (U.S. feet) | Pinellas County Enterprise GIS | Continuous variable |
| Proximity to Tampa Bay | Distance to Tampa Bay (U.S. feet) | Created in ArcGIS Pro | Continuous variable |
| Proximity to trails and paths | Distance to trails and paths (U.S. feet) | Pinellas County Enterprise GIS | Continuous variable |

dataset was joined to the parcel shapefile using the Parcel ID number. A new feature class was then created of residential homes that were sold in 2022. To properly clean and prepare the data, several issues needed to be addressed for the dataset. The dataset included houses that were foreclosed on and had a sales price of zero dollars. These records were excluded from the final data set as they are not relevant for this study. The dataset also included condos and apartments that were sold in units of 2-9, 10-49, and 50 or more. These records were listed as a single transaction, but this transaction included multiple houses or condos. These features were also excluded from the final dataset as they are outside the scope of this study. The year the home was built was listed in the dataset instead of the age of the home. A new field was created to calculate the age of each house. There were several fields that contained missing values. This was addressed by filing in the missing values with the average of the field. The average of the field was calculated, then the missing values were updated accordingly.

The proximity of each house to the geographic features was then calculated for the dataset. This was done by calculating the distance of each house to the nearest geographic feature. The result is a table that contains a feature ID number and the distance between the house and the geographic feature. This process was performed for each geographic feature. Each table was then joined to the housing dataset. A logarithmic transformation was performed on the variable price to ensure it follows a normal distribution.

**3.3. *Spatial autocorrelation and spatial clustering***

The second process in this study is to identify spatial clusters from the measurement of spatial autocorrelation. Spatial autocorrelation is the tendency of values to be more related in nearby locations than in faraway locations. In most cases the values at one location are most likely to be influenced by the values at locations that are close by. Because spatial statistics can account for spatial autocorrelation, it can be used to find geographic patterns and trends in the data (Environmental Systems Research Institute n.d. b.). Spatial autocorrelation can either be positive or negative. If the data shows evidence of positive spatial autocorrelation, it implies there are clusters that have a similar value of the variable being analyzed. If there is negative spatial autocorrelation found in the data, the variable being analyzed has dissimilar values. Global Moran’s I statistic is one of the most common ways to measure spatial autocorrelation. Global Moran’s I is presented in Equation (1).

Equation (1)

where n is the number of spatial units, xi is the value of the spatial unit, is the average of the observations in the spatial units, and is the spatial weight matrix (Wang *et al*. 2023). It gives a numerical value to indicate the strength of spatial autocorrelation. The values range from -1 to 1. Values close to 1 indicate strong positive spatial autocorrelation whereas values close to -1 indicate strong negative spatial autocorrelation. Values that are close to zero indicate no spatial autocorrelation meaning the values are randomly distributed throughout the study area. An important part of Global Moran’s I is the construction of the spatial weights. The spatial weight accounts for the locations at smaller distances being expected to have similar values of the variable that is being analyzed. This can be done by assigning a distance-based weight for the distance between locations (Iliopoulou and Feloni 2022). This is commonly used for point data such as the inverse weighted distance method. Furthermore, Local Moran’s I and the Getis-Ord Gi\* are measures of local spatial autocorrelation. They indicate clusters of high or low values. This is also referred to as hot spot analysis. Both spatial autocorrelation and spatial clustering are similar. They are both concerned with finding spatial patterns in the data, but spatial clustering identifies statistically significant areas of high or low clusters for a particular variable. It uses a distance and similarity measure to group locations based on their proximity to each other and the similarity of the value of the variable being analyzed.

In this study, global Moran’s I was used to determine if spatial autocorrelation exists. If positive spatial autocorrelation was found, hot spot analysis will be implemented to map the clusters of high and low values. This is important to gain a better understanding of the data. These techniques can give a visual presentation as to where housing prices are high and where they are low.

**3.4. *Regression models***

The last step of this study is constructing a regression model to determine how geographic features affect house prices within the study area. Regression analysis examines the relationship between a dependent variable and one or more independent variables. Linear regression is perhaps the most common form of regression analysis. Linear regression assumes a linear relationship between the dependent and independent variables. The slope of the line in the equation represents the change in the dependent variable to the change in the independent variable. When more than one independent variable is used in the model, it is known as multiple linear regression. The equation is presented in Equation (2).

y = B0 + b1 \* X1 + B2 \* X2 + . . . Bn \* Xn + e Equation (2)

where Y is the dependent variable, B are the coefficients, X are the explanatory variables, and e is the random error term (Environmental Systems Research Institute n.d. b.). The coefficients explain the effect of the independent variables on the dependent variable.

**3.4.1. *Ordinary Least Squares (OLS)***

A common form of multiple linear regression is OLS regression. OLS is a global model that estimates the difference between the predicted and actual values of the dependent variable using the least squares method. OLS is presented in Equation (3).

Equation (3)

where is the dependent variable, is the constant, Xi are the independent variables, is regression coefficients, and is the error term (Sisman and Aydinoglu 2022). OLS is commonly used for its simplicity, flexibility, and ease of interpretation. While OLS has its advantages, it is inaccurate when multicollinearity is present. Multicollinearity is when two or more independent variables are highly correlated with each other. OLS also suffers from its assumption of the independence of the residuals. Meaning the data are assumed to be independent from each other across the study area. This is a common problem when modeling geographic features since the data are often spatially autocorrelated. (Iliopoulou and Feloni 2022, Sisman and Aydinoglu 2022).

OLS is used in this study to examine the relationship between housing prices and geographic features. The coefficients will be examined to determine if they are statistically significant and to determine their strength and type of relationship with the dependent variable. The VIF of each explanatory variable will also be assessed to ensure that no variable has a VIF higher than 7.5. If a variable is found to have a 7.5 or higher VIF, it will be removed from the model. If the residuals are found to be clustered, a local spatial model will be implemented to further investigate the dependent variable.

**3.4.2. *Local spatial regression models***

To overcome this problem, spatial regression models are commonly applied along with OLS in hedonic models (Helbich *et al*., 2013, Zang *et al*., 2019, Hong and Yoo 2020). Global models, such as OLS, assume that the relationship between the dependent and independent variables do not vary across the study area. However, local regression models, such as Geographically Weighted Regression (GWR), model this relationship while allowing for the geographic features to be nonstationary*.* GWR is an extension of linear regression that uses a kernel weight to create local coefficients for each location within the study area (Brunsdon *et al.* 1996). This gives features that are father away less weight while features that are closer together are given more weight. This is done through the local weight of each feature.

GWR creates a neighborhood, also known as bandwidth, for the construction of the local regression equation. The neighborhood can be defined by the number of neighbors or by a distance band. In GWR, there are two weighting schemes: Gaussian and bi-square. These are distance decay functions that control how weight decreases and distance increases. A Gaussian weighted scheme does this by gradually decreasing the weight as distance increases. Features that are far away will have a smaller weight, but the weight cannot reach zero. This is where Gaussian differs from bi-square. The bi-square weighted scheme is the same as Gaussian but features outside of the neighborhood are given a weight of zero. This means that the features outside of the neighborhood will have no effect on the results (Brunsdon *et al.* 1996). For housing price models, the bi-square method provides better results as shown by Yang *et al*. (2016). GWR can provide better results over global regression models when analyzing geographic relationships. However, GWR assumes that the local relationships of the features are constant throughout the study area and are measured at the same spatial scale (Fotheringham *et al.* 2017). Multiscale Geographically Weighted Regression (MGWR) expands on GWR by varying the bandwidth across the study area. MGWR is shown in Equation (4).

Equation (4)

where xjj are the explanatory variables, is the bandwith, are the coefficients, i is the dependent variable, and is the error. MGWR allows each explanatory variable to have a different bandwidth. This is important to capture the changing spatial scale over the study area between the explanatory variables and the dependent variable. This can give MGWR an advantage over GWR to accurately estimate a local regression model. In this study, GWR and MGWR are used with the bi-square weighted scheme. A map of the coefficients for each explanatory variable will be examined to determine their significance across the study area. This is important to identify where each geographic feature has a significant impact on housing prices.

**4. Results**

**4.1. *Spatial autocorrelation***

Using Global Moran’s I, the spatial autocorrelation was assessed with price as the input field and inverse distance as the conceptualization of spatial relationships. The null hypothesis for this test is that features area not spatially correlated and are randomly distributed. The Moran’s Index was .521306, the z-score was 169.406376, and the distance threshold was set to 2472.27 U.S. feet. These values led to the rejection of the null hypothesis, indicating that the values are spatially autocorrelated.

**4.2. *Hot spot analysis***

Hot spot analysis gives the ability to identify hot spots and cold spots for a selected variable. Price was used for the input field and the inverse distance was used as the conceptualization of spatial relationships. Inverse distance decreases the importance of an observation when distance increases. Because of this trait, inverse distance was deemed to be the most appropriate for this study. This method was also used by Iliopoulou and Feloni (2022). The results can be seen in Figure 3. First, hot spot analysis (Getis-Ord-Gi\*) was examined for each house in the study area. The hot spots are located primarily in three locations: Downtown St. Petersburg, Clearwater Beach, and on the barrier island. There are also several cold spots identified throughout the county. These cold spots were identified as either condos or model homes. This is due to their small size in square footage, acreage, and being older homes. For example, the cold spot that was identified located just north of St. Petersburg is a mobile home park. These houses have an average age of 40 years old, zero acres, and an average of just under 1,000 square feet. Another example can be seen in southwest of Clearwater with a condominium complex. These homes have an average age of 47 years, zero acres, and an average of 775 square feet.

For the hot spots, it was expected that the barrier island would be a hot spot due to its proximity to the beach. In Clearwater Beach the average house price is just over $1,000,000 and an average square footage of 1,654.7. The average age in this region is 44 years old due to older condos in this area. This is consistent with the rest of the barrier island with an average age of 43 years old, 1,578.6 square feet, and an average price of just under $1,000,000. The hot spot in St. Petersburg is similar in age with an average of 48 years old but is higher in both price and square footage. The average price is $1,239,139 and the average size is 2,014.8 square feet.

**4.3. *Result of OLS regression***

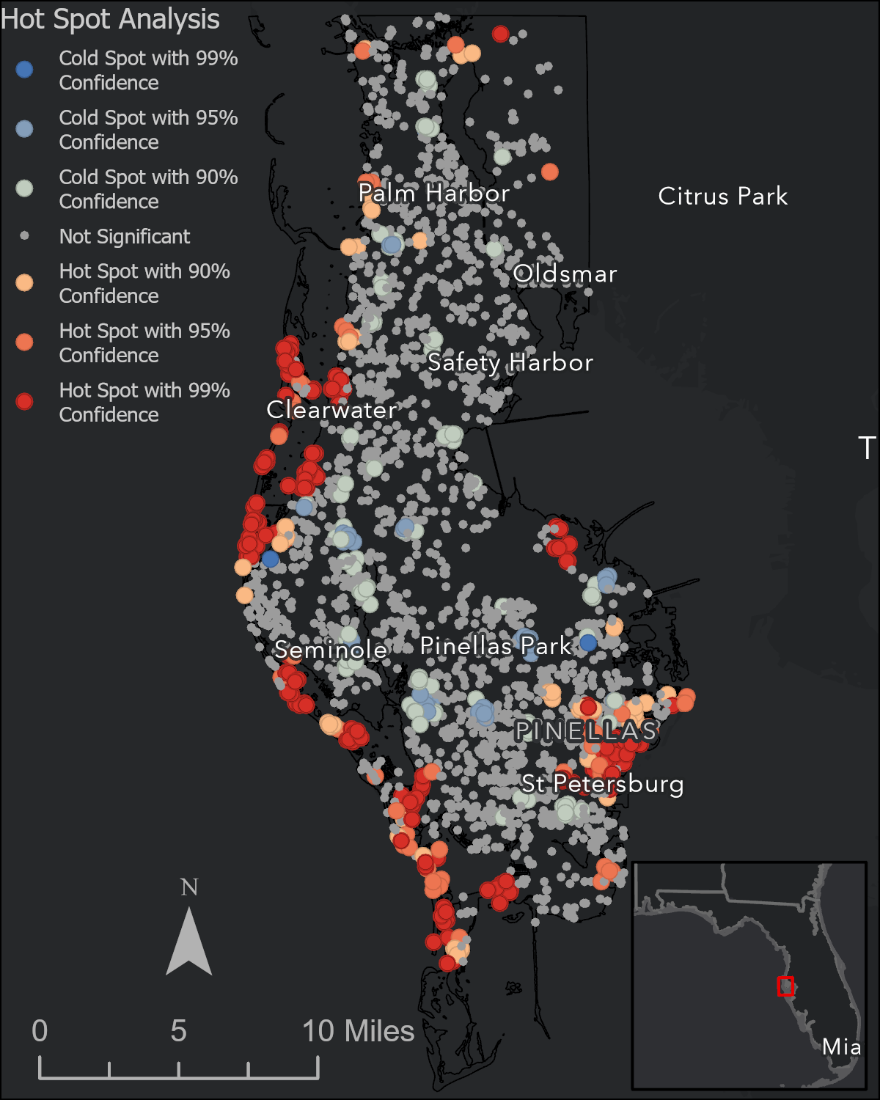


Figure 3: Hot-spot analysis results

OLS was performed before GWR for two reasons. First, OLS is more capable of identifying multicollinearity problems in the data. Secondly, the output of OLS gives the VIF, probability, robust probability, and the coefficient for each variable entered into the model. This made it easier to address any multicollinearity issues and to identify the importance of each explanatory variable.

OLS also outputs several diagnostic tests to evaluate the model. Histograms and scatterplots are also given for each explanatory variable entered into the model. Initially, every explanatory variable was included into the model and the variable price was set as the dependent variable. Due to multicollinearity issues, the variables of proximity to public beaches, located on barrier island, located in St. Petersburg, proximity to Tampa Bay, and major roads were removed from the final model.

The variables public beaches and located on the barrier island were highly correlated with proximity to the beach. After several iterations, it was found that proximity to the beach produced a better fit to the data than public beaches located on the barrier island. Proximity to Tampa Bay was also highly correlated due to St. Petersburg being located on the bay. The variable located in St. Petersburg was also highly correlated with these variables. Because of this correlation, a new feature class of proximity to downtown St. Petersburg was created to address these issues. This feature class was found to be a better fit to the data than the variable located in St. Petersburg. Major roads was highly correlated with paths and trials, therefore it was removed from the final model. The results of the final OLS model are presented in Figure 4.

The model produced an R squared of .555079, an adjusted R squared of .551890, and an AIC of 3820.121397. The R squared and adjusted R squared diagnostics express the goodness- of-fit of the model. The goodness-of-fit measures how well the explanatory variables can explain the variation of the dependent variable. This means that the model was able to explain 55% of the variation in the dependent variable. While the R squared and adjusted R squared are similar tools for assessing model performance, they are different. The R squared statistic does not consider the number of independent variables entered into the model. R squared can have the tendency to increase as more independent variables are used in the model. The adjusted R squared statistic considers the number of independent variables entered into the model. It will only increase if the added independent variable improves the overall fit of the model. Adjusted R squared is a more accurate measure of the goodness-of-fit of the model, especially with many

Map

Description automatically generated

Figure 4: OLS results

independent variables (Environmental Systems Research Institute n.d. a.). Since this study used many independent variables, the adjusted R squared was used to assess the model’s performance.

The VIF indicates multicollinearity between the variables. If a VIF of 7.5 or higher was found, it was removed from the model. The VIF was relatively low for each variable with the highest being 4.21 for marinas. The variables that were statistically significant were then identified from the robust probability statistic, which was used in leu of the probability statistic since the Koenker (BP) statistic was found to be significant. The Koenker (BP) statistic, also known as the Breusch-Pagan test, is used to determine if the variables are non-stationary. The null hypothesis is defined as the variables that are stationary, which means they are constant throughout the study area. A statistically significant variable indicates that it has a significant influence on the dependent variable.

The coefficient of each significant explanatory variable was then assessed to evaluate their relationship with the dependent variable. The coefficient reflects the strength of the independent variable as well as the type of relationship the explanatory variables have with the dependent variable. A positive coefficient indicates a positive relationship. This signifies when the independent variable increases, the dependent variable also increases.

Square feet was found to have the highest positive significant coefficient of 400 and trails and paths was second with 23.96. A negative coefficient indicates a negative relationship between the independent variable and the dependent variable. Therefore, when the independent variable decreases, the dependent variable increases. The age of the houses had the lowest coefficient with -622.87 and parks was second with -12.40. The AIC was used as a measure of model performance when addressing multicollinearity issues. When determining which variables to remove from the model, such as historic districts and Tampa Bay, AIC was used to identify which variable was a better fit. The variables that produced a lower AIC score were used in the final model. A lower AIC score indicates the model is a better fit to the data. The Jarque-Bera statistic was also found to be significant.

The Jarque-Bera statistic evaluates model bias. This test measures if the residuals are normally distributed. The null hypothesis assumes that the residuals are normally distributed. Since this test was found to be significant in the model, the residuals do not follow a normal distribution, meaning the residuals are spatially autocorrelated (Environmental Systems Research Institute n.d. a.). To confirm this, Global Moran’s I was performed on the residuals. The residuals were found to be clustered with a Moran’s Index of .323153 and a z-score of 103.087583. Overall, OLS produced a moderate fit to the data, but it is not an appropriate model. Since the residuals are spatially autocorrelated and both the Koenker (BP) and Jarque-Bera Statistic were both significant, GWR would be a better fit to the data.

**4.4. *Result of GWR***

All the variables from the OLS model were used in the GWR except for historic districts. This was due to multicollinearity issues with the variable downtown St. Petersburg. It was found that the variable St. Petersburg produced a better adjusted R squared and lower AIC than the variable historic districts, therefore it was left out of the final model. Since GWR is a local model, it is sensitive to local multicollinearity. Even though OLS helped to determine multicollinearity, it is a global model and does not consider local multicollinearity as GWR does. This was also found by Wang *et* al. (2020) and Iliopoulou and Feloni (2022).

To address these issues, the variable historic districts was removed and the optimal neighborhood, or bandwidth, needed to be selected. Multiple iterations were performed for both the distance band and number of neighbors method. It was concluded that the number of neighbors method gave the best results with 150 neighbors being the optimal number. This number was found from starting with 50 neighbors and increasing the number by 50 until the optimal number of neighbors were found. The model was also used with both the bi-square and Gaussian local weighted scheme to determine which was the best performing. Bi-square produced a higher adjusted R squared and a lower AIC which is consistent with what Yang *et al*. (2016) found in their study.

GWR showed itself to be a much better fit to the data than OLS with a R squared of .7788, an adjusted R squared of .7008, and an AIC of 5022.5627. The results can be seen in Figure 5. The distribution of the standardized residuals produced a mean of -.00808. This was examined to ensure they follow a normal distribution. The plot of the standardized residuals vs. the predicted values shows a slight positive linear trend with an R squared of .00256.

Map

Description automatically generated

Figure 5: GWR standardized residuals

**4.5. *Result of MGWR***

All the parameters used in MGWR were the same as GWR apart from the neighborhood method. For MGWR, the golden search method was applied to identify the optimal number of neighbors for the model. The golden search method is an iterative process that searches for the optimal number of neighbors using multiple combinations. The number of neighbors with the lowest AIC score is then used in the model. Since MGWR allows each explanatory variable to vary across the study area, the optimal number of neighbors can differ for each variable. The results of the golden search method can be seen in Table 3. While this method can be helpful in determining the optimal bandwidth, it takes an extensive amount of time to process. MGWR produced an R squared of .7263, an adjusted R squared of .6937, and an AIC of 4757.4628. The coefficient estimates can be seen in Table 4 and the standardized residuals can be seen in Figure 6.

Table 3: The optimal number of neighbors for each explanatory variable

|  |  |
| --- | --- |
| Explanatory Variables | Neighbors (% of Features) |
| Intercept | 37 (1.39) |
| Acreage | 74 (2.77) |
| House Age | 2671 (100.00) |
| Total living square feet | 1186 (44.40) |
| Proximity to Airports | 2671 (100.00) |
| Proximity to Beach | 2286 (85.59) |
| Proximity to Boat Ramp | 2671 (100.00) |
| Proximity to Bus Stops | 2671 (100.00) |
| Proximity to Downtown St. Petersburg | 2671 (100.00) |
| Proximity to Golf Course | 2671 (100.00) |
| Proximity to Health Care Sties | 2671 (100.00) |
| Proximity to Lakes | 2671 (100.00) |
| Proximity to Marinas | 2671 (100.00) |
| Proximity to Parks | 2671 (100.00) |
| Proximity to Schools | 1753 (65.63) |
| Proximity to Shoreline | 2671 (100.00) |
| Proximity to Sports Facilities | 2671 (100.00) |
| Proximity to Trails and Paths | 2671 (100.00) |

Table 4: Results of MGWR

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Explanatory Variables | Mean | Standard Deviation | Minimum | Median | Maximum |
| Intercept | -0.0096 | 0.2897 | -1.2675 | 0.0024 | 0.9256 |
| Acreage | 0.2906 | 0.2247 | -0.4529 | 0.2766 | 0.9558 |
| House age | -0.0725 | 0.0038 | -0.0781 | -0.0727 | -0.0673 |
| Total living square feet | 0.4738 | 0.0438 | 0.3905 | 0.4802 | 0.5329 |
| Proximity to Airports | -0.0433 | 0.0038 | -0.0504 | -0.0427 | -0.0367 |
| Proximity to Beach | -0.1206 | 0.0158 | -0.1434 | -0.1248 | -0.0869 |
| Proximity to Boat Ramp | -0.0719 | 0.0017 | -0.074 | -0.0725 | -0.068 |
| Proximity to Bus Stops | 0.0619 | 0.0008 | 0.0595 | 0.062 | 0.0642 |
| Proximity to Downtown St. Petersburg | -0.1034 | 0.0102 | -0.1161 | -0.106 | -0.0883 |
| Proximity to Golf Course | 0.0632 | 0.0051 | 0.0558 | 0.0635 | 0.0705 |
| Proximity to Health Care Sties | 0.0308 | 0.0055 | 0.021 | 0.0335 | 0.0375 |
| Proximity to Lakes | 0.0566 | 0.0067 | 0.0477 | 0.0574 | 0.0669 |
| Proximity to Marinas | -0.0905 | 0.0008 | -0.0922 | -0.0907 | -0.0883 |
| Proximity to Parks | -0.0225 | 0.0047 | -0.0308 | -0.0207 | -0.0171 |
| Proximity to Schools | 0.0282 | 0.0194 | -0.0059 | 0.0336 | 0.0562 |
| Proximity to Shoreline | -0.0504 | 0.0029 | -0.0539 | -0.0517 | -0.0439 |
| Proximity to Sports Facilities | 0.0603 | 0.0055 | 0.0505 | 0.0628 | 0.0666 |
| Proximity to Trails and Paths | 0.0361 | 0.0019 | 0.0327 | 0.0357 | 0.0414 |

**4.6. *Comparison of GWR and MGWR***

The results of the regression models show that GWR had the highest adjusted R squared. While OLS had the lowest AIC of the models used in this study. A comparison of each model can be seen in Table 5. These results are also similar to what Sisman and Aydinoglu (2022) found in their study. While GWR had a higher adjusted R squared, MGWR produced a lower AIC. Both GWR and MGWR were shown to be adequate for this study. The results of MGWR can provide more information than the output of GWR. Understanding the scale at which the variables operate is helpful in identifying the importance of each variable. It is of not that the

Map

Description automatically generated

Figure 6: MGWR standardized residuals

Table 5: Comparison of OLS, GWR, and MGWR

|  |  |  |
| --- | --- | --- |
| OLS | GWR | MGWR |
| R squared: .555079 | R squared: .7788 | R squared: .7263 |
| Adjusted R squared: .551890 | Adjusted R squared: .7008 | Adjusted R squared: .6937 |
| AIC: 3820.121397 | AIC: 5022.5627 | AIC: 4757.4628 |

processing time for MGWR is much more intensive than GWR. MGWR may not be appropriate for large datasets due to its processing time.

**4.7. *Discussion***

It was concluded that 14 of the 20 geographic features entered into the regression models had a statistically significant effect on housing prices. These variables were identified from the results of OLS and can be seen in Table 6. While OLS was not the best performing model used in this study, it was crucial in identifying which factors had a significant influence on housing prices. OLS was also useful in eliminating multicollinearity. Multicollinearity was identified by the VIF with a VIF of 7.5 or higher being removed from the model. Eliminating multicollinearity was essential to ensure that multicollinearity did not exist in the data before the variables were used in GWR and MGWR. Both GWR and MGWR produced a better fit to the data than OLS indicating that local spatial models were more capable of capturing the variance of the dependent variable.

The results of the regression models showed the distance to the beach to be the most influential variable with a mean value of -.1206. The distance to downtown St. Petersburg had the second highest influence with a mean value of -.1034, and marinas was third with a -.0905 mean value. However, distance to bus stops and distance to golf courses had the highest positive relationship with the dependent variable. While 14 of the 20 variables used in this study were found to be significant, it is evident that distance to beach and distance to downtown St. Petersburg are the most influential geographic factors within the study area.

The results of MGWR were analyzed to determine which geographic features varied throughout the study area. The results of OLS and GWR were useful in this study however, these results do not show where these explanatory variables have the most influence and if they have a global, regional, or local influence. MGWR gives estimates for what scale each variable operates at across the study area. Every explanatory variable was found to operate at a global scale apart from two variables: distance to the beach, and distance to schools as seen in Table 3.

Table 6: Statistically significant geographic features

|  |
| --- |
| Statistically Significant Variables |
| Proximity to Airports |
| Proximity to Beach |
| Proximity to Boat Ramp |
| Proximity to Bus Stops |
| Proximity to Downtown St. Petersburg |
| Proximity to Golf Course |
| Proximity to Health Care Sties |
| Proximity to Lakes |
| Proximity to Marinas |
| Proximity to Parks |
| Proximity to Schools |
| Proximity to Shoreline |
| Proximity to Sports Facilities |
| Proximity to Trails and Paths |

Both variables operate at a regional level. The optimal number of neighbors was also found to be lower for both variables.

Although GWR produced a slightly higher adjusted R squared than MGWR, MGWR was the preferred model just in this study. This was due to MGWR’s ability to identify which explanatory variables vary across the study area. This ability provides greater insight into what causes the dependent variable to vary across the study area. However, MGWRs processing time is much more extensive than GWR. GWR can be completed in several minutes while MGWR can take several hours or days depending on the size of the dataset. Even though MGWR proved to be the preferred model, GWR was still able to adequately capture the variance of the dependent variable. These results are also similar to what Sisman and Aydinoglu (2022) found in their study.

The regression models used in this study were not perfect. They did not show a great fit to the data. This was expected due to other explanatory variables being left out of the analysis as they are out of the scope of this study. These include economic and neighborhood variables that can have a large influence on housing prices.

**5. Conclusion**

Hedonic models are a valuable tool for assessing housing prices. While there are many features that affect housing prices such as the structural elements of the house and neighborhood characteristics, this study focused on what geographic features significantly influence housing prices in Pinellas County, Florida. A dataset was obtained with housing sold in the study area in 2022. The data was pre-processed and cleaned to remove outliers, fill in missing or null values, and remove unnecessary records. The dependent variable, price, was logarithmically transformed to ensure a Gaussian distribution. The distance to the geographic features was calculated for each house to create the final dataset.

First, Global Moran’s I was used to determine if the house prices were spatially autocorrelated. Second, since the house prices were found to be spatially autocorrelated, hot spot analysis was performed to determine where clusters of high and low values were located within the study area. Third, a global regression model, OLS, was performed which helped to eliminate multicollinearity and to identify which explanatory variables were statistically significant. The residuals of OLS were found to be clustered from the results of Global Moran’s I indicating that a local regression model would be a better fit to the data. Finally, MGR and MGWR were used with the statistically significant variables identified by OLS. Of the statistically significant geographic features found in this study, the distance to the beach and distance to downtown St. Petersburg were found to have the most influence on housing prices. This indicates that houses near the beach and downtown St. Petersburg are more valuable. Of the 14 geographic features, distance to the beach and distance to schools were identified by MGWR as being nonconstant across the study area. Distance to the beach and distance to schools were identified as regional variables whereas the other 12 variables were identified as global variables. It was demonstrated that OLS was shown not to be preferred model in this study due to its inability to account for spatial autocorrelation. The residuals of OLS were found to be clustered from the results of Global Moran’s I, indicating that a spatial model would be a better fit to the data. GWR and MGWR proved to be better performing models than OLS. While GWR and MGWR produced similar results, MGWR was the preferred model. This was due to MGWR being able to identify which explanatory variables vary throughout the study area.

Future research should be conducted to identify what socio-economic factors have a significant impact on housing prices within this study area. These factors could include education, income, and employment. This would help in providing a more complete picture of the real estate market in Pinellas County, Florida.

**References**

Anas, A., Arnott, R., Small, K., 1998. Urban spatial structure. *Journal of Economic Literature*, 36(3), pp. 1426-1464.

Brunsdon, C., Fotheringham, A., Charlton, M., 1996. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis,* 28(4), pp. 281-298. Available at: [https://doi.org/10.1111/j.1538-4632.1996.tb00936](https://doi.org/10.1111/j.1538-4632.1996.tb00936x)

Chen, Y. and Fik, T., 2017. Housing-market bubble adjustment in coastal communities- A spatial and temporal analysis of housing prices in Midwest Pinellas County, Florida. *Applied Geography*, 80, pp. 48-63. Available at: [https://doi.org/10.1016/j.apgeog.2017.01.007](https://doi-org.ezproxy.nwmissouri.edu/10.1016/j.apgeog.2017.01.007)

Environmental Systems Research Institute, n.d. a. *How OLS regression works* [online]. ArcGIS Pro Documentation. Available at: https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/how-ols-regression-works.htm [Accessed 3 April 2023].

Environmental Systems Research Institute, n.d. b. *Regression analysis basics* [online]. ArcGIS Pro Documentation. Available at: https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/regression-analysis-basics.htm. [Accessed 27 February 2023].

Fotheringham, S., Yang, W., Kang, W., 2017. Multiscale Geographically Weighted Regression (MGWR). *Annals of the American Association of Geographers*, 107(6), pp. 1247-1265. Available at: <https://doi.org/>10.1080/24694452.2017.1352480

Helbich, M.*,* Brunauer, W., Vaz, E., Nijkamp, P., 2013. Spatial heterogeneity in hedonic house price models: The case of Austria. *Urban Studies*, 51(2), pp. 390–411. Available at: <https://doi.org/10.1177/0042098013492234>.

Hong, I. and Yoo, C., 2020. Analyzing spatial variance of Airbnb pricing determinants using multiscale GWR approach. *Sustainability*, 12(11), p. 4710. Available at: https://doi.org/10.3390/su12114710.

Ibeas, A., Cordera, R., Olio, L., Coppola, P., Dominguez, A., 2012. Modelling transport and real-estate values interactions in urban systems. *Journal of Transport Geography*, 24, pp. 370-382. Available at: [https://doi.org/10.1016/j.jtrangeo.2012.04.012](https://doi-org.ezproxy.nwmissouri.edu/10.1016/j.jtrangeo.2012.04.012).

Iliopoulou, P. and Feloni, E., 2022. Spatial modelling and geovisualization of house prices in the Greater Athens region, Greece. *Geographies*, 2(1), pp. 111–131. Available at: <https://doi.org/10.3390/geographies2010008>

Kokot, S. and Gnat, S., 2019. Simulative Verification of the Possibility of using Multiple Regression Models for Real Estate Appraisal. *Real Estate Management & Valuation*, 27(3), pp. 109-123.

Li, X., 2022. Prediction and Analysis of Housing Price Based on the Generalized Linear Regression Model. *Computational Intelligence & Neuroscience*, 2022, pp. 1-9.

Lins, M., Novaes, L., Legey, L., 2005. Real Estate Appraisal: A Double Perspective Data Envelopment Analysis Approach. *Annals of Operations Research*, 138, pp. 79-96.

Liu, G., 2022. Research on Prediction and Analysis of Real Estate Market Based on the Multiple Linear Regression Model. *Scientific Programming*, 2022, pp. 1-8.

Lopes, S., Brondino, N., Rodrigues da Silva, A., 2014. GIS-based analytical tools for transport planning: spatial regression models for transportation demand forecast. *ISPRS International Journal of Geo-Information*, 3(2), pp. 565–583.

Pinellas County, 2022. *Fast facts about Pinellas County* [online]. Available at: https://pinellas.gov/about-pinellas-facts/. [Accessed: March 17, 2023].

Pinellas Suncoast Transit Authority, 2023. *PSTA history & fact* [online]. Available at: https://www.psta.net/about-psta/history-and-facts/. [Accessed: March 18, 2023].

Rai, K. and Nathaway, M. S., 2013. Application of GIS and statistical methods to select optimum model for malaria susceptibility zonation: a case study. *Scientific Annals of the 'Alexandru Ioan Cuza' University of Iasi: Geography Series*, 59(2), pp. 73-94.

Rosen, S., 1974. Hedonic prices and implicit markets: product differentiation in pure competition. *Political Economy*, 82(1), pp. 34-55. Available at: <https://doi.org/10.1086/260169>

Sisman, S. and Aydinoglu, A.C., 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, p.106183. Available from: [https://doi.org/10.1016/j.landusepol.2022.106183](https://doi-org.ezproxy.nwmissouri.edu/10.1016/j.landusepol.2022.106183)

Wang, D., Li, V.J., Yu, H., 2020. Mass appraisal modeling of real estate in urban centers by geographically and temporally weighted regression: A case study of Beijing’s core area. *Land*, 9(5), p. 143. Available at: <https://doi.org/10.3390/land9050143>

Wang, Y., Wangyong, Y., Wang, M., Chen, X., Li, Y., 2023. Application of improved Moran’s I in the evaluation of Urban Spatial Development. *Spatial Statistics*. 54, p. 100736. Available at: [https://doi.org/10.1016/j.spasta.2023.100736](https://doi-org.ezproxy.nwmissouri.edu/10.1016/j.spasta.2023.100736)

Wright, D.K., MacEachern, S., Lee, J., 2014. Analysis of feature intervisibility and cumulative visibility using GIS, Bayesian, and Spatial Statistics: A study from the Mandara Mountains, northern Cameroon. *PLoS ONE*, 9(11), pp. 1-15.

Yang, Yi., Liu, J., Shenghua, X., Zhao, Z., 2016. An Extended Semi-Supervised Regression Approach with Co-Training and Geographical Weighted Regression: A Case Study of Housing Prices in Beijing. *International Journal of GeoInformation*, 5(1), p. 4. Available at: <https://doi.org/10.3390/ijgi5010004>

Zhang, S., Wang, L., Lu, F., 2019. Exploring housing rent by mixed geographically weighted regression: A case study in Nanjing*. ISPRS International Journal of Geo-Information*, 8(10), p. 431. Available at: <https://doi.org/10.3390/ijgi8100431>