THE HEDONIC MODELING OF PROPERTY PRICES: CASE STUDY ON CAPE TOWN, SOUTH AFRICA

Douw Gert Brand Boshoff
University of Cape Town

Joseph Awoamim Yacim
University of Pretoria

Abdullahi Alhassan Ahmed
Federal Polytechnic Nasarawa

Abstract
The real estate and valuation discipline has embraced the recent developments in spatial data analysis as a way of remediating obvious limitations of the ordinary least squares (OLS) approach in handling spatial effects. However, despite the development, the South African property market has yet to embrace spatial analysis when estimating property prices. The objective of this study is to understand the rationale behind the use of spatial hedonic modeling on the Cape Town property market by first testing the data against the existence of spatial effects and using the appropriate techniques to correct the glitches. A spatial error autocorrelation model and geographically weighted regression (GWR) were employed to correct spatial dependence (autocorrelation) and spatial heterogeneity on 3,232 observations. The relative performance of the two spatial modeling techniques as revealed by their goodness-of-fit are quite impressive but the spatial error model marginally outperforms the GWR. Thus, it is recommended that any of these techniques can be used in modeling property prices in the Cape Town market.

Keywords
Cape Town, South Africa

The hedonic price models (HPMs) based on ordinary least squares (OLS) have been used for more than five decades to solve several problems. Court (1939) has been credited with inventing this paradigm that has become famous in property valuation and other related assessments. Drawing from the idea premiered in Court (1939) are studies that provide a simplistic framework for real estate price analysis (Lancaster, 1966; Rosen, 1974). The OLS method has a wide range of applications including to estimate the implicit prices of natural disasters like floods and wildfires (Chivers and Flores, 2002; Mueller and Loomis, 2008), demand for environmental quality, wind power facilities and natural space attractiveness (Brasington and Hite, 2005; Heintzelman and Tuttle, 2012; Daams, Sitjsma, and van der Vlist, 2016), and construction of property price index, portfolio management, and mortgage underwriting (Bourassa, Cantoni, and Hoesli, 2010, 2016).

In particular is the widespread use of OLS in the construction of property prices for taxation purposes in many jurisdictions. Accordingly, McCluskey et al. (1997) provide a list of government departments in Australia, Canada, Hong Kong, Malaysia, New Zealand, Singapore, Sweden, Tasmania, United Kingdom, and United States that use the OLS and other price modeling techniques to assess the market price or value of properties. The usefulness of the OLS in pricing several properties, particularly, in mortgage underwriting and taxation, gave credence to its introduction into the valuation industry of several other property markets in the developed countries. However, in the emerging markets, especially Sub-Saharan Africa, the lack of adequate property data for a value-based assessment remains a challenge for effective implementation of the OLS (Ahmed and Yacim, 2018). Thus, the inefficient market, coupled with a sordid database of property transactions, necessitated the use of a net annual value of properties for academic assessment of OLS in a country like Nigeria (e.g., Megbolugbe, 1988; Ahmed and Yacim, 2018) as manual assessment of properties for tax and related purposes still dominate the market.

In contrast, South Africa has a nation-wide database for all property transactions with detailed attributes required for a computer-assisted mass appraisal (CAMA) (Ward, 2001). Although manual assessment still exists in some jurisdictions, the OLS is probably the most widely used technique in many municipalities (KPMG, 2015). In particular is the use of OLS in the Cape Town property market by the city valuation office (CVO) for property tax assessment and re-assessment. For more than a decade, the CVO has been conducting an assessment of the market value of properties every three years to guard against large variations in property values between valuations subject to review and adjustment by professional valuers. However, cases of objections and appeals from dissatisfied property tax payers relative to the published valuation rolls are concerns that valuation offices in South Africa contend with annually (KPMG, 2015; LexisNexis, 2018). Since most assessments are OLS based, there is a need to revisit the use of this estimator, assess the magnitude of its limitations on Cape Town data, and popularize the use of spatial methods in the South African property market.

Previous studies on the use of spatial methods to control spatial dependence and spatial heterogeneity in property pricing are ubiquitous for developed countries (e.g., Pace and Gilley, 1997; Basu and Thibodeau, 1998; Bowen, Mikelbank, and Prestegaard, 2001; Wilhelmsson, 2002; McCluskey et al., 2013). The dearth of literature on property pricing with these methods in an emerging market such as South Africa is a compelling reason for this study. This will enable policymakers to make appropriate decisions on the suitability of a method(s) relative to the local market. Moreover, Bourassa, Cantoni, and Hoesli (2010) and Zurada, Levitan, and Guan (2011) observe that it is difficult to implement the results of previous studies because of differences in data, thus generating data-driven results. Since results are data-driven, a test must of necessity be performed on a method(s) before its introduction into the local market. Additionally, it is argued that market efficiency is reflected in the quality of data and must be allowed to define the choice of method(s). In this study, we utilize data from the local market of Cape Town, South Africa to comparatively examine the relative performance of spatial methods relative to the widely used OLS estimator.

The explorations in this study have tremendous implications to the CVO for Cape Town and other municipal areas where mass appraisal is being contemplated in South

Africa or other Sub-Saharan African countries. First, the assessment discloses a model that predicts estimates closer to the assessed values, and secondly, the nature of the South African property data discloses the best spatial models applicable to the local market. The remaining aspects of this study are organized as follows. We review the related literature, discuss the underlying philosophies of the various methods and data we use, present the results and discussion, and present concluding remarks.

LITERATURE REVIEW

REVIEW OF PREVIOUS OLS PRICING OF PROPERTY STUDIES

The OLS has been used to estimate property prices by incorporating among others the location indicator variables to account for its influence on price. Incorporating a location dummy into the OLS has the advantage of enhancing the fit and parameter estimates of the model (Pace, Barry, Clapp, and Rodriquez, 1998). The geo-coordinates of a property reflecting its parcel location in space can also be used directly in the OLS model as an expansion series to enhance model performance (e.g., Bitter, Mulligan, and Dall'erba, 2007). However, the OLS has many drawbacks that inhibit its successes in price estimation including functional specification and nonlinearity among variables. Relative to the specification of functional form, there is no specific form the relation between the response and covariates must take. This is the reason why Crooper, Deck, and McConnel (1988) report that several scholars emphasize the outcome of the goodness-of-fit benchmark that Rosen (1974) and Goodman (1978) suggest to identify the right functional form suitable for a data. Additionally, scholars extensively use the Box-Cox transformation, linear, semi-log, and log-log functional forms in property and related analyses (Mok, Chan, and Cho, 1995).

Accordingly, the Box-Cox transformation and linear form are used in Crooper, Deck, and McConnell (1988). The authors note that when all the attributes are fitted into the model, the linear and quadratic Box-Cox transformation provides better results, but when only selected variables are included or substituted by proxy, the linear model performs better than the quadratic Box-Cox function. Similarly, Goodman (1978) finds the most extensively used linear model to be too restrictive and thus supports the use of the Box-Cox transformation. Crooper, Deck, and McConnell (1988) specify six models including the linear, semi-log, log-log, quadratic, linear Box-Cox, and quadratic Box-Cox transformed variables in the property market of Baltimore, Maryland. They examine a sample of 200 observations sold from 1977 to 1978 and their results show a preference towards the linear Box-Cox as the best model.

Mok, Chan, and Cho (1995) explore the effects of location, structural, and neighborhood attributes with the OLS on 1,027 properties to estimate prices in Hong Kong. The authors use a control measure to avoid bias by utilizing properties that are within a five-minute walking distance to the 13 mass transit railway stations and transactions executed within a one-month (August 1990) period. They utilize the linear, log-linear, and Box-Cox transformation on the dependent variable and transformation on both the dependent and independent variables and find the Box-

Cox to be the most preferred. Tse and Love (2000) use hedonic regression to examine 139 properties in four large residential estates in the Tsing Yi district, Hong Kong. The sales transactions took place between January 1, 1999 and March 31, 1999. The log-log form of the hedonic OLS regression is used on both the dependent and independent variables. Because property may be cheaper on account of age, they include a squared log (age) variable in their model. Furthermore, they employ the OLS with different specifications and two weighted least squares (WLS) techniques with weights assigned to the dependent variable and two independent variables namely log (age) and log (area) to account for heteroscedasticity. The authors conclude that the OLS with location variables provides a better fit than a specification without location, although the results of the WLS give a more robust fit as the effect of heteroscedasticity is taken into account.

Yang (2001) uses OLS to estimate the prices of 226 apartments sold in July and August, 1998 in Beijing, China. The author uses a Box-Cox transformation for the dependent variable, as well as the linear and log-linear specifications. The results reveal a similarity between the log-linear model (adjusted R^2 of 66.8%) and the Box-Cox transformation (adjusted R^2 of 66.5%), relative to the linear model adjusted R^2 of 64.4%. Stevenson (2004) uses OLS on 6,441 observations from 1995 to 2000 in Boston, Massachusetts. The model uses structural, locational, and time variables to estimate prices while also controlling for heteroscedasticity. The semi-log form of the OLS is used in all analyses. The results reveal that if the age variable is converted and added to the model in form of cubic and squared, a marginal decline of the influence of heteroscedasticity is achieved and the model provides a better fit (adjusted R^2 of 63%). The marginal improvement in the OLS (from an adjusted R^2 of 62.7% to 63%) in this study shows the need to use a more advanced method that can completely ameliorate the problem of heteroscedasticity.

Shimizu (2014) uses the semi-log OLS on a dataset containing 13,822 single-family dwellings to estimate property prices in 23 wards of Tokyo, Japan. The transactions span a period from January 2000 to December 2000. In all, three models are estimated using structural, land, location, railway/subway, and time variables. However, in specifying the three models, model 1 does not contain neighborhood effects or neighborhood effects by GIS, model 2 contains the neighborhood effect attributes but no neighborhood effect by GIS, and model 3 contains both neighborhood effects. The relative importance of these effects is seen in the goodness-of-fit as model 3 provides a better fit (adjusted R^2 of 66%) than the other two models. Schulz, Wersing, and Werwatz (2014) examine 18,444 single-family property observations sold from 2000 to 2011 in Berlin, Germany. The data are stratified in the development step (i.e., the period of sales from 2000 to 2005; 8,429 sales) and the validation step (i.e., the period from 2006 to 2011; 10,015 sales). The price and log-prices are used as the response variables, while the covariates are left in their linear format. Their model is designed in such a way that it permits a finite set of nonlinear Box-Cox transformations for continuous regressors. The results show that the semi-log model provides a better fit in all scenarios.

McCluskey (2016) examines 40,138 apartment sales from January 2011 to July 2014 in Astana, Kazakhstan. The linear and semi-log form of the OLS is used in the analysis. The initial result is affected by outliers, but after the removal of outliers, the linear model performed better as revealed in the adjusted R^2 of 67%. The semi-log model, however, did not provide a better fit even after outliers were removed from the data.

For comparison of the different functional form specifications, all the studies reviewed provide no evidence of which functional form consistently works better than the others. The behavior of the models, however, reveals that different samples from different countries give varying results. Sirmans, Macpherson, and Zietz (2005) provide an example of why the variation can occur and report that a variable such as a bedroom may be measured in one study as the number of the bedrooms while in another, a binary dummy variable is used. In general, most of the studies find that the semi-log form provides optimal performance over other functional forms. This is because it tends to handle the nonlinearity among variables that reduce bias arising from unusual observations. This might be the reason why despite its failings the method is still used to price several properties. Although it is theoretically and conceptually sound, the OLS faces some other problems including spatial dependence and spatial heterogeneity, which are tackled in spatial hedonic regressions.

REVIEW OF PROPERTY PRICING STUDIES THAT INCORPORATE SPATIAL EFFECTS

Several methods have been developed in the econometric and geostatistical fields to deal with the OLS glitches. These include geographically-weighted regression (GWR) (Brunsdon, Fotheringham, and Charlton, 1996); spatial simultaneous autoregressive (SAR), such as the spatial error, spatial lag and spatial mix models (Anselin, 1988), spatial expansion method (SEM) (Cassetti, 1972, 1997); local kriging and co-kriging (Haas, 1995, 1996) among others are used to correct spatial influences. Consequently, several studies have been undertaken to correct spatial dependence and or spatial heterogeneity in the property data.

Accordingly, Pace and Gilley (1997) used 506 housing data-points from Boston SMSA to correct spatial dependence. The data was previously used in the well-known article by Harrison and Rubinfeld (1978), but Pace and Gilley (1997) correct some of the obvious limitations in the data to accommodate spatial analysis and interaction of geo-coordinates (quadratic expression). The quadratic expression is used to reflect location as suggested in Belsley, Kuh, and Weisch (1980), which enhances model performance from an R^2 of 81.1% to 81.4%. However, when the spatial autoregressive model is used, the maximum likelihood estimator resulted in an increase in R^2 from 81.4% to almost 90%, thus performing better in comparison to the OLS. Similarly, Dubin, Pace, and Thibodeau (1999) use 10 observations to compare the performance of OLS and spatial approaches including SAR, conditional autoregressive (CAR), mixed regressive spatially autoregressive, and Gaussian correlogram. The authors find that only the results of the mixed regressive spatially autoregressive model are different, as all other spatial models give estimates that are closer to the OLS.

Bowen, Mikelbank, and Prestegaard (2001) estimate four models, namely linear, semi-log OLS, and autoregressive spatial lag models, to correct spatial dependence or autocorrelation on the other. The data comprise residential sales in Cuyahoga County, Ohio stratified along the Cuyahoga River with 1,387 observations on the east-side and 1,054 observations on the west-side. The diagnostic test on the two samples suggests ''no discernible spatial dependence in the west-side'' data and thus is not used for modeling spatial dependence. The spatial lag model (in a linear and semi-log form) is used in modeling spatial considerations on the east-side sample. The authors conclude that the OLS estimates for the west-side of Cuyahoga sample are plausible, suggesting a lack of need for spatial processes; the results of the east-side sample suggest the autoregressive model outperforms OLS in price estimation. The study reveals the significance of testing data against the possibility of spatial effects so that the appropriate method of control can be determined for a market. Wilhelmsson (2002) examines 1,377 single-family sales in Stockholm, Sweden during the period from January 2000 and May 2001. Submarket area dummies are used to reflect the location in the OLS, while the spatial lag (SLM) and spatial error models utilize the geo-coordinates to measure the distances between observations. The result reveals the SLM and spatial error models provide better fits (R^2 of 69% and 68%, respectively) in comparison to OLS (R^2 of 66.4%).

Thériault, Des Rosier, Villeneuve, and Kestens (2003) use the SEM to control for spatial heterogeneity in property prices in Quebec, Canada. The authors believe that the influence of property characteristics on price is pushed by the spatial variability of demand that is associated with heterogeneity in the distribution of household types and services within a geographical area. They examine 4,040 observations of sold properties in Quebec City, from January 1990 to December 1991, stratified into 3,633 observations for model building and 407 observations as hold out samples. In their study, various attributes are expanded spatially using the indicators derived from census data, such as property buyers' economic base, status, family cycle, and accessibility to services. A spatial drift occurs in some of the attributes including lot size, age, linkage to the sewer system, inferior ceiling, kitchen cabinet, and number of bathrooms. They find that although interaction does not completely remove autocorrelation, it can nonetheless reduce spatial autocorrelation among the OLS residuals.

Militino, Ugarte, and Garcia-Reinaldos (2004) use 293 sale transactions to test the performance of the OLS, SAR, CAR, and geostatistical models in Pamplona, Spain sold in 2000. They find that SAR and CAR provide adequate correction for spatial dependence as shown in their inferences relative to other spatial methods. However, despite the high performance of the autoregressive models in remediating the problem of spatial dependence among observations in a local market, Bourassa, Cantoni, and Hoesli (2007) find a less persuasive result. Using 4,880 property sales from Auckland, New Zealand, they conclude that the autoregressive models (SAR and CAR) do not give more accurate predictions than the global OLS and geostatistical models.

Kestens, Thériault, and Des Rosier (2006) examine 761 single-family properties sold from 1993 to 2001 in Quebec City, Canada. The study was undertaken to address the problem of spatial heterogeneity in the data. They employ two methods, namely the SEM and GWR. The results suggest that both methods are appropriate for the pricing of properties in Quebec City. However, the spatial expansion captured both the spatial and non-spatial heterogeneity of the parameter estimates and identifies the causes of parameter drift among variables. The GWR gives detailed information through local regression and identifies the causes of non-stationarity, thereby aiding the differentiation of the complex relation that affects property values.

Farber and Yeates (2006) use 19,007 property sales in Toronto, Canada to compare the relative performance of the OLS, GWR, SLM, and the moving window regression. The observation took place from July 2000 to June 2001. Several variables are used to search for an unbiased and stable global OLS but there is little multicollinearity in the best global OLS, which consists of five property structural characteristics, two neighborhood characteristics, and two accessibility characteristics. Although the R^2 of the OLS is a good fit (83%) and the residual errors normally distributed, there is spatial autocorrelation in the data as shown by the Moran I coefficient (0.24). To reduce the problem, a spatial autoregressive term is added to OLS, which increased the model fit from an R^2 of 83% to R^2 of 88% and reduced the spatial bias. The problem of the non-stationary regression coefficient is also addressed within the GWR and yields an impressive model fit of R^2 = 90%. However, due to the statistical limitation of the evaluation criteria, the authors could not support the application of GWR in the Toronto property market.

Bitter, Mulligan, and Dall'erba (2007) use the OLS (without *x*, *y* coordinates), OLS (with *x*, *y* coordinates to control for location), SEM, spatial expansion with a lagged variable, and GWR to control for spatial heterogeneity in the property data of Tucson, Arizona. They examine 11,732 single-family properties sold in 2000. The model with location control in the form of third-order polynomial expansion provides a better fit (adjusted R^2 of 88%) and a reduction in standard error than the OLS without location control. The inclusion of the lagged term in the expansion yields a higher model fit (adjusted R^2 of 91%) more than the spatial expansion techniques (adjusted R^2 of 89%). The spatial lag term captures the effect of externality, suggesting the presence of spatial heterogeneity in the data. Thus the GWR is used to vary parameter estimates of the regression model and yields an adjusted R^2 of 92%. The hold out sample (1,163) they use to test the predictive accuracy of all models in the study reveals the GWR as the most preferred method. However, both the spatial expansion and GWR methods yield better results than those of the stationary coefficient OLS method, and the GWR outperforms the expansion method in terms of predictive and explanatory powers.

Páez, Long, and Farber (2008) use an estimation sample of 30,145 observations and a validation sample of 3,349 observations in the city of Toronto, Canada to test the predictive power of OLS, moving windows regression (MVR), GWR, kriging, and moving windows kriging. The sales transactions take place between January 2001 and December 2003. The property characteristics include the lot size in square feet, effective frontage in feet, age, date of sale, characteristics of the immediate surroundings, and social environment, among others. The neighborhood attributes influence the fit of the global OLS model, but the problem of censoring, particularly variables set to zero, in the sparsely populated area affects the performance of the model. If omitted, this might be a problem but if included, the effect may not be profound. Their results, however, show that in terms of predictive power, the GWR performs marginally better than the MVR but substantially outperforms the other models.

Borst and McCluskey (2008) and McCluskey and Borst (2011) use the GWR to identify market segments and model large-scale variations in property value relative to the global OLS method. Property sales data from three counties in the U.S.: Catawba County, North Carolina, Sarasota County, Florida, and Fairfax County, Virginia. They conclude that the resultant segments tend to enhance predictive precision and reduce spatial autocorrelation in the residual errors. McCluskey et al. (2013) use a sample of 2,694 residential properties sold between 2002 and 2004 in Northern Ireland to detect spatial dependence and spatial heterogeneity. Several approaches including the OLS, artificial neural networks (ANNs), GWR, and SAR are used. The GWR, however, performs better than other models relative to cost efficiency, ease of use, and prediction accuracy. This is one study that utilizes several methods to conclude which technique performed optimally and can be used within the mass appraisal environment. As noted earlier, all related studies reviewed above are undertaken in mature property markets in developed countries. Thus far, the only known study in South Africa and indeed the whole of Sub-Saharan Africa that controls for spatial effects is Yacim and Boshoff (2019). They utilize GWR to control for spatial heterogeneity in the Cape Town property data. In this study, we extend the literature by incorporating a control for spatial dependence, as suggested in de Graaff, Florax, Nijkamp, and Reggiani (2001) for an emerging market. This allows us to make appropriate decisions relative to the Cape Town property market.

However, while it is acceptable to deal with spatial dependence separately, dealing with both spatial dependence and heterogeneity is an essential condition to permanently remove the glitches in the data used for property pricing. This is because the property markets are vulnerable to the occurrence of both effects (Bitter, Mulligan, and Dall'erba, 2007). Again as suggested in de Graaff, Florax, Nijkamp, and Reggiani (2001), several reasons are adduced for considering spatial dependence and spatial heterogeneity together including: (1) there may be no difference between autocorrelation (dependence) and non-stationarity (heterogeneity) in an observation because specific clustering of low or high property prices may be the result of dependence among properties (see Anselin, 2000); (2) the specification of spatial dependence in the models may lead to a particular form of heteroskedasticity (see also Kelejian and Robinson, 2000); and (3) because the tests may exhibit inappropriate results. It is therefore difficult to differentiate or separate between the two spatial effects.

Methods

In this section, we describe the data and different techniques we use to price properties in the Cape Town property market. We begin with the traditional HPMs based on the OLS estimation paradigm. We then describe the advance techniques used to control spatial dependence and spatial heterogeneity. We use SAR and GWR modeling to control for spatial effects. Lastly, we discuss the data we use to test all the models.

THE HEDONIC PRICE MODELS

The OLS estimator measures the contribution of individual characteristics (structural, temporal, and neighborhoods) of a property to the total price. To determine the price of a property, the structural (e.g., the number of bedrooms, size of the property, swimming pool, etc.), time of sale (e.g., reverse month of sale and dummy representing quarter/semi-annual indicator), and neighborhoods (e.g., environmental amenities, location of properties, etc.) are regressed against price. The general formulation of the OLS is given as:

$$\gamma = X\beta + \varepsilon \tag{1}$$

or

$$\gamma = \beta_0 + \sum_{k=1}^{K} X_k \beta_k + \varepsilon,$$

where ε represents an independent and normally distributed (iid) error term; γ denotes a $1 \times n$ vector of property price; and β denotes a parameter vector corresponding to structural, temporal, and neighborhood attributes (X_k). The major limitation of using these models lies in their inability to fully capture all location features, making them susceptible to autocorrelation among the residuals. Militino, Ugarte, and Garcia-Reinados (2004) report that the results of autocorrelation in the residuals makes the OLS provide inefficient parameter estimates, which although is unbiased gives a biased variance estimate that causes an unreliable inference. The OLS that largely ignore spatial effects has over the years being relegated by models that are designed to capture all effects. In the spatial models, we use the Bayesian estimation, general-method-of-moments (GMM), and the maximum likelihood (ML) approaches for estimating prices. The most widely employed estimator to date in spatial hedonic analysis is the ML technique.

SPATIAL MODELING TECHNIQUES

Simultaneous Autoregressive. The autoregressive process can occur in three different areas including at the response or dependent variable, at both the response and predictor variables, and in the error term. Consequently, different types of simultaneous autoregressive models are used to control for such effects including the spatial lag model (SLM), the spatial mix model (SMM), and the spatial error model. However, we use the SLM and spatial error model to control for spatial dependence among observations in the property market. The SLM is used to control effects when there is dependence between the response variable of property at observation i and the response variable of property at observation j. Can (1992) uses adjacency effects to give a more full explanation of the dependency among observations in which the prices of nearby properties are partly explained by shared neighborhood externalities, as well as the property's physical quality and uses. When these adjacency effects shared by properties in the neighborhoods are capitalized into their prices, a spatial dependence will occur in the process. Thus, the "adjacency effects" influence the decision of real estate agents/valuers in determining the prices of properties in the local market.

The spatial dependence that occurs between observations in the property market makes it necessary to include the functional interdependence among the prices of neighboring properties in the OLS (Can, 1992). The resultant model is the SLM represented in equation (2):

$$\gamma = \rho W\gamma + X\beta + \varepsilon, \tag{2}$$

where ρ is the parameter coefficient, W is the weight matrix, and $W\gamma$ is the spatial lag variable introduced into the OLS. The spatial error or spatially correlated error exists if the error term in the OLS displays spatial autocorrelation as a result of the omission of variable bias relative to the location of a property parcel (Mueller and Loomis, 2008). If the property location is not properly captured in the OLS, measurement error is likely to occur in the process. To tackle this problem, Anselin and Bera (1998) suggest that a spatial process should be properly specified for the disturbance term. Accordingly, the usual specification is to have a spatial autoregressive process calibrated in the error terms as follows:

$$\gamma = X\beta + \varepsilon \tag{3}$$

$$\varepsilon = \lambda W\varepsilon + \mu, \tag{4}$$

where λ is a spatial autoregressive coefficient for error lag $W\varepsilon$ and μ is an uncorrelated and homoscedastic error term.

To establish and correct spatial dependence requires a priori specification of a spatial weights matrix. Two weighting regimes namely border contiguity (rook and queen contiguity) or distance-based matrices are typically used. However, in property-related analysis, the distance-based weights specification is the most commonly used regime because sales of adjacent properties are unlikely to take place in the same period and information about parcel borders is not always provided in the sales data (Mueller and Loomis, 2010). Since the Cape Town data we use does not contain information on parcel borders and sizes, we use the distance-based weights specification with a threshold value of 300 meters. According to Mueller and Loomis (2008), the weights matrix (W) captures similarities between properties in a given jurisdiction or neighborhood, which is likely to be ignored in non-weighted OLS estimation techniques.

The spatial weight is an $N \times N$ (positive and symmetric W) matrix that models the relationship of neighbors for every observation within the sample as nonzero elements. Therefore, if property i and property j are neighbors, then $W_{ij} = 1$; conversely, if property i and property j are not neighbors, then $W_{ij} = 0$. More so, since property cannot be a neighbor to itself, the diagonal elements of the weights matrix element are conventionally set to zero, i.e., $W_{ii} = 0$ or $W_{jj} = 0$. The weights matrix is usually row standardized for simplicity of comparison and interpretation so that each row will have weights that sum to 1. Accordingly, the presence of spatial dependence or autocorrelation in the data can be investigated using the Moran's I statistics. The Moran's I test suggests the existence of spatial autocorrelation in the OLS residuals, 0.3730 (mi/df), and 47.4742 (value). The conclusion, therefore, is that the null hypothesis of no spatial effects in the data is hereby rejected. Consequently, the Lagrange Multiplier (LM) and robust LM tests reveal a preference for the spatial error model as the proper model specification for controlling spatial dependence in the Cape Town data.

Geographically-Weighted Regression. Some property data contain geostatistical elements resulting in a biased estimation when modeling with the OLS. In such a case, GWR or other geostatistical techniques are essentially used to model relationships among property variables. According to Wheeler (2007), the variables are used at fixed points that have spatial coordinates and the variable values are usually mean measures of aggregate data. The GWR fits a model at each of the point locations and permits the parameter estimates to vary locally across space. The model is designed to utilize a specification that is flexible in which an undefined function of location in space (u, v) provides the regression parameters (β). The GWR for each location is given as:

$$\gamma = \beta_0(u, v) + \sum_{k=1}^{K} X_k \beta_k(u, v) + \varepsilon. \tag{5}$$

Weighted least squares are used to weight and estimate the regression coefficients for each location, thus assigning greater weights on observations that are located nearer to the calibrated location in geographic space and lesser weights to observations that are located further away. The weights are specified by:

$$\hat{\beta}(u, v) = [X^T W(u, v) X]^{-1} X^T W(u, v) \gamma, \tag{6}$$

where $W(u, v) = \text{diagonal}[W_1(u, v), \ldots, W_n(u, v)]$ is a diagonal weights matrix, which changes (increases or decreases) relative to distances of each location i in the geographical space. $\hat{\beta}(u, v)$ is the vector of regression coefficients at each location and the superscript T is the matrix transpose. The X vector is as defined previously—the matrix of property attributes. There are two acceptable and widely-used weighting specifications in GWR: Gaussian function and bi-square function. However, in this study, the Gaussian function given in equation (7) is used.

$$w_j(u, v) = \exp\left(\frac{-d_{ij}}{h}\right)^2, \tag{7}$$

where the Euclidian distance specified as d_{ij} separates the location of observation at the property i and the property at location j, and h denotes a scalar quantity, generally known as the bandwidth. The bandwidth (h) can be defined by a user (bandwidth parameter), estimated using the cross-validation (CV) or corrected Akaike Information Criteria (*AICc*), which allows a user to employ an automatic method that finds the best bandwidth with the optimal predictions. In this study, the *AICc* is preferred,

Exhibit 1
Descriptive Statistics of Variables

Variable	Mean	Median	Std. Dev.	Skewness
Assessed value	4483474	3600000	3117754	3.282
Log of assessed value	15.154	15.1	0.54	0.535
Number of bedrooms	3.558	3.00	0.99	0.804
Property quality	2.496	2.00	0.61	0.914
Property condition	2.509	2.00	0.62	0.408
Sales month	14.885	15.0	8.17	0.010
Property view	2.628	3.00	0.85	0.440
Property style	2.035	2.00	0.43	5.251
Number of floors	1.518	1.00	0.55	0.423
Size of living area	177.48	168	79.0	0.992
Swimming pool	13.97	0.00	18.4	1.533

although there is not much difference between the two since both methods allow the GWR tool to select an optimal bandwidth (Charlton, Fotheringham, and Brundson, 2005). The technique is primarily designed to explore the spatial non-stationarity of the parameter estimates; nonetheless, it is also useful in prediction processes (Leung, Mei, and Zhang, 2000).

DATA

The CVO supplied our cross-sectional data involving 3,526 single-family residences sold at arms-length between January 2012 and May 2014 in the city of Cape Town, South Africa. The CVO is the body charged with the responsibility of preparing the valuation rolls in the city of Cape Town. The sample contains incomplete, negative, extreme, and unrealistic transactions that were discarded from the original data to avoid potential bias. We recoded some of the attributes from non-numeric to numeric values making them suitable for multivariate analysis. The selection of 11 property variables is based on a series of preliminary regression assessments that reveal them as value significant. This approach is consistent with Bitter, Mulligan, and Dall'erba (2007), who utilize regressions to select the 13 property attributes they use in their study. Additionally, previous multivariate analysis of this nature reveals a preference for these and other variables depending on their availability in the data.

The characteristics of the property sample are displayed in Exhibit 1. The Exhibit reports varied statistics of the variables for the Cape Town data. The median sale of R3,600,000 (250,632 USD), is quite high (a South African rand exchange for 0.070 USD), depicting that some of the areas in the sample are high priced. Accordingly, the prices of properties in the sample range between R824,000 (57,367 USD) and R38,000,000 (2,645,558 USD), with a standard deviation of R3,117,754 (217,058 USD), suggesting a wide dispersion of property prices. This is expected because neighborhoods such as Bantry Bay, Bishops Court, Camps Bay, Bakoven, and Clifton are characterized by high-priced properties, while lower-priced properties in neighborhoods including Blougerg, Blaauwbergstran, and Claremont among others are included in the sample (see Map of Cape Town in Exhibit 2 and spatial distribution of sales in Exhibit 3). The property prices are highly skewed (3.282). Pavlov (2000) reports that although an uncommonly high transaction price is an important element in an example data, it can partly be remediated through the log transformation of the property prices (dependent variable). The average bedroom in the sample is approximately 4 while a typical property has an average size of 177 square meters. The variables are left in their categorical and continuous state when modeling property prices with the GWR, while the binary format is used in modeling prices with the autoregressive and OLS models. Following Liu (2013), we aggregate some variables (that are in the categorical state but converted into binary format) due to an infrequent number of occurrences. This approach is also used by Guan, Zurada, and Levitan (2008) to avoid dimensionality problems that could impede the performance of the regression technique.

Three functional form specifications including linear, semi-log, and log-log are common but we utilize the semi-logarithmic form in which the assessed value is log-transformed and all covariates are left in their linear form. The distances between properties in the spatial models are calculated through the use of latitude and longitude information provided in the sample. As previously mentioned, the latitude and longitude characteristics of the properties are used to control for spatial effects in autoregressive and geostatistical techniques. The submarkets/market segments defined by the city of Cape Town Tax Assessors based on the neighborhood structure and similarity of selling prices are used in the OLS to explain the influence of location on property prices. In the HPMs literature, time of sale is treated in several ways including incorporating a time variable in the regression, reverse month of sale, and creating dummy variables to reflect the time (year, semi-annual, quarter, and month) the sales or assessment took place. We use the effect of time as reflected in the form of semi-annual dummies in the OLS and autoregressive models while the reverse month of sale is used in the local model (GWR). The use of dummies to reflect an input variable in the models is not without a correction such that all frequently occurring variables are excluded from the analysis to avoid the "dummy variable trap" described in Greene (2003) and Borst (2007).

Results and Discussion

In this section, we discuss the empirical tests of all models. We begin with an analysis of the stationary coefficients OLS techniques with a naïve specification (without location elements) and specifying the location indicator elements (submarkets) to observe the influence of variables on prices. This is closely followed by an analysis of the spatial error model that reveals its influence in tackling spatial dependence in the data. The next analysis is GWR, which reveals the influence of non-stationarity and varying parameters in the local market. We conclude with a detailed comparison of model performance using acceptable statistics provided in the software (GeoDa and GWR4).

Exhibit 2
Map of Cape Town Neighborhoods

Exhibit 3
Spatial Distribution of Sales (x, y coordinates)

Source of Map: https://www.roomsforafrica.com/dest/south-africa/western-cape/cape-town.jsp?tab=3.

RESULTS OF SPATIAL AND NON-SPATIAL HEDONIC MODELING

Exhibit 2 shows the results for three model specifications. The OLS specification without and with control for spatial effects is reported in OLS1* and OLS2**. The estimates are a priori expected with appropriate signs. The parameter estimates in OLS1 without control for location are generally larger than OLS2 with control for location and the autoregressive models that utilize geo-coordinates as its measure of location. The bigger estimates in OLS1 can be attributable to bias caused by the absence of location indicators defined by covariate submarket dummies and the x, y coordinates. The inclusion of submarkets dummies to control for location in OLS2 enhances the predictive power as revealed by the adjusted R^2, which increases from 42% to 69% and the decrease in standard error from 0.412 to 0.302.

Additionally, we find that all location dummies (segments or submarket) have a significant intercept-shifting influence on the assessed values. The negative sign in 86% of the location dummies implies the positive effects of location on the assessed values decline relative to other locations. This variation is best captured in the GWR model due to its moving window ability (Páez, Long, and Farber, 2008) and localized comprehension of the distinct contributions of specific variables (McCord et al., 2012). Generally, the results suggest the OLS2 as the baseline model for the Cape Town data, thus further explanation is relative to the global OLS2 model. However, the global Moran's I test on the residuals of the OLS2 (estimation data) finds it to be spatially dependent (0.3624), suggesting the presence of spatial autocorrelation. The likelihood ratio test results further confirm the existence of spatial autocorrelation in the data and thus rejects the null hypothesis of no spatial effects. Again the results supports earlier findings that neighborhoods/submarkets indicator variables we use to control for location do not completely remove spatial effects (see Páez, Uchida, and Miyamoto, 2001; Wilhelmsson, 2002; Bitter, Mulligan, and Dall'erba, 2007).

The spatial error model we use to correct the spatial dependence in the OLS increases the R^2 from 69% to 80%, as shown in the last two columns of Exhibit 4. This result reveals that spatially correlated error exists in the OLS model for the Cape Town property market. The parameter estimates of the property variables in the spatial error model generally drop in comparison to the baseline OLS2 model. Interestingly, the number of significant variables increases relative to the OLS2 model. The key variable in this assessment, namely the size of a living area (dwelling), is significant in all the three model specifications. The estimate on size, interpreted as the elasticity of assessed values relative to changes in dwelling size, is positive, showing that larger properties have higher assessed values in the study area. This implies that for a 1% increase in the size of the dwelling, assessed values increase by 0.0024%, 0.0019%, and 0.0018%, for the OLS1, OLS2, and spatial error model, respectively. Again, as expected the parameter estimate of the swimming pool is positive and significant for the three models, revealing that a bigger swimming pool adds more to the assessed value of a property. The coefficient estimate on poor property condition is also as expected, negative and not significant, revealing that poor property condition does not contribute to property assessed values in the OLS1 and OLS2 models. In the spatial error model, poor property condition although negative is a significant indication that it has a positive effect on assessed value but the effect declines as the condition

Exhibit 4
Regression Coefficients of Global OLS and Autoregressive Model ($N = 3,232$)

	OLS 1[a]		OLS 2[b]		Spatial Error Model	
Variable	Estimate	*P*-value	Estimate	*P*-value	Estimate	*P*-value
Constant	14.3575	0.0000	14.7202	0.0000	14.6258	0.0000
Bedroom1	0.0491	0.5744	−0.0914	0.1521	−0.1169	0.0188
Bedroom2	0.0567	0.0424	−0.0353	0.0850	−0.0681	0.0002
Bedroom4	0.0919	0.0000	0.0396	0.0025	0.0663	0.0000
Bedroom5	0.1335	0.0000	0.0712	0.0004	0.1249	0.0000
Bedroom6	0.1755	0.0001	0.0931	0.0053	0.0879	0.0007
Bedroom7	0.1062	0.2785	0.1194	0.0949	0.1726	0.0020
Bedroom8	0.3968	0.0028	0.4046	0.0000	0.3302	0.0000
Poor property quality	−0.0116	0.8902	0.0027	0.9652	0.0079	0.8669
Good property quality	0.1245	0.0000	0.0905	0.0000	0.0768	0.0000
V/Good property quality	0.2845	0.0002	0.2676	0.0000	0.2234	0.0000
Excel. property quality	0.4921	0.0000	0.3523	0.0000	0.1713	0.0000
Poor property condition	−0.0224	0.7388	−0.0141	0.7740	−0.0944	0.0136
Good property condition	0.0312	0.1737	0.0484	0.0043	0.0634	0.0000
Excel property condition	−0.1758	0.0000	−0.0669	0.0241	0.0155	0.5210
2 floors	0.1612	0.0000	0.1468	0.0000	0.1547	0.0000
3 floors	0.4456	0.0000	0.3428	0.0000	0.2477	0.0000
Unconventional style	0.3116	0.0000	0.1658	0.0000	0.0965	0.0008
Georgian victor style	0.1011	0.1243	0.1181	0.0141	0.0415	0.3040
Cape Dutch style	−0.1594	0.1683	−0.0751	0.3736	−0.0179	0.7863
Maisonette style	−0.1349	0.0380	0.0233	0.6265	−0.0476	0.2323
Below average view	0.1679	0.0000	0.0859	0.0005	0.0737	0.0003
Above average view	0.1627	0.0000	0.1155	0.0000	0.0415	0.0000
Panoramic view	0.3471	0.0000	0.2425	0.0000	0.1641	0.0000
Excellent view	0.4982	0.0000	0.2871	0.0000	0.2429	0.0000
Semi-annual 1	−0.0086	0.7006	−0.0154	0.3506	0.0093	0.4718
Semi-annual 2	−0.0310	0.1563	−0.0341	0.0330	−0.0094	0.4537
Semi-annual 3	−0.0292	0.1880	−0.0207	0.2013	−0.0155	0.2202
Semi-annual 5	0.0074	0.7436	0.0042	0.8011	0.0272	0.0368
Size of living area	0.0024	0.0000	0.0019	0.0000	0.0018	0.0000
Swimming pool	0.0035	0.0000	0.0015	0.0000	0.0013	0.0000
Lambda	—	—	—	—	0.8448	0.0000
R^2	0.419		0.693		0.795	
Adj. R^2	0.413		0.689		—	
Log-likelihood	−1715		−682		−291	
Standard Error	0.413		0.301		0.244	

Note:
[a]Without locational dummies
[b]With locational dummies

Exhibit 5
GWR (local) Coefficients

Variable	Minimum	Lower Quartile	Median	Upper Quartile	Maximum
Intercept	12.0585	13.3346	13.6587	14.0384	15.5524
Bedroom	−0.1646	0.0223	0.0454	0.0702	0.1722
Property quality	−0.1235	0.0783	0.1291	0.1703	0.3354
Property condition	−0.3339	−0.0129	0.0503	0.0985	0.3522
Number of floors	−0.2539	0.1337	0.1905	0.2421	0.3675
Building style	−0.5192	−0.1238	−0.0363	0.0067	0.2788
Property view	−0.0142	0.0517	0.0848	0.1178	0.4428
Month of sale	−0.0067	−0.0011	0.0015	0.0035	0.0070
Size of living area	0.0003	0.0021	0.0024	0.0030	0.0054
Swimming pool	−0.0033	0.0004	0.0021	0.0035	0.0279
R^2	0.7846				
Adj. R^2	0.7482				
−2 Log likelihood	221.1825				
Standard Error	0.2504				

deteriorates further. Thus, for any 1% increase in poor property condition, assessed values fall by 0.11%.

Exhibit 5 illustrates the results of the GWR model that capture spatial variation in property prices in Cape Town. Again, like the autoregressive model, the local model performs better than the global stationary coefficient OLS model as evidenced in the adjusted R^2 of 78% and reduction in the standard error from 0.302 to 0.250. The decrease in the standard error of estimates in the spatial error model and GWR relative to OLS indicate they give a better fit for the data. However, relative to the results of the spatial error model, GWR produces an R^2 of 78%, which implies similarity in their explanation of variance in assessed property values in the study area. Thus, the performance of GWR in this study reveals that there exist significant localized spatial effects in the Cape Town property market. This is demonstrated by the spatial variation in their parameter estimates across the geographic location. The minimum, lower quartile, median, upper quartile, and maximum coefficients in Exhibit 5 show a non-stationarity and variation of property prices relative to different locations. Though not identical, the parameter estimates of the size and swimming pool variables in the OLS and spatial error models are similar to the median estimates in GWR (Borst, 2007). Again, due to different model specifications, explanations of coefficient estimates could simultaneously not be done, but the specification of the variables "dwelling size" and "swimming pool" applies to global and local models.

In the GWR model, the coefficient estimates for dwelling size range from a minimum of 0.0003 to a maximum of 0.0054, while the median value is 0.0024. The local model has varied coefficients, whereas the global OLS and spatial error model have very similar parameter estimates (0.0019 and 0.0018, respectively) for size (i.e., the

Exhibit 6
Summary of Predicted Results of Different Hedonic Models (N = 3,232)

Model	Log-Likelihood	Standard Error	AIC	BIC	R^2	Rank
OLS1	−1715.2	0.4134	3492.4	3680.9	42%	4
OLS2	−681.7	0.3008	1453.5	1727.1	69%	3
Spatial error model	−291.2	0.2443	644.4	832.9	80%	1
GWR	−221.2	0.2503	940	3129	78%	2

average dwelling size of all observations within the geographic region). Accordingly, all else being equal, a 1% increase in dwelling size would increase the assessed value or property selling price by as little as 0.0003% in one regression point (location) and 0.005% more at another location of the property market. Similarly, the parameter estimates for swimming pool range from a minimum of −0.0033 to a maximum of 0.0279, which suggests that a property with a swimming pool sells for 0.003% less in one location and 0.03% more in another precinct of Cape Town. The negative values in all but one of the coefficients of the independent variables indicate counter-intuition, implying a scenario where a property with inadequate buyer or user requirements can sell for a higher amount or have a higher assessed value in one location than those with a more modern or better buyer or user requirements in other locations. This scenario sometimes occurs because of buyers' preference for a location and thus they trade-off other requirements for location.

COMPARATIVE ANALYSIS OF PREDICTIVE ACCURACY OF MODELS

There are several indices (statistics) for comparing the predictive accuracy of models, but we use R^2, log-likelihood, standard error, AIC, and BIC. These are used to appraise the predictive accuracy of the four models (OLS1, OLS2, spatial error model, and GWR). The results are summarized in Exhibit 6.

The results reveal that the spatial error model outperforms all other models in correcting spatial effects in the property market. In comparison to GWR, the spatial error model marginally outperforms the non-stationarity coefficients model. This shows that the inclusion of advances in spatial statistics in the hedonic regression modeling significantly strengthens property price estimation. Bowen, Mikelbank, and Prestegaard (2001) report that capturing spatial effects in the HPMs would enhance the accuracy of the models. However, as noted earlier, assumption could not be made without testing the example data against available methods. In this study, the relevance of Moran's *I*, LM, and robust LM tests reveal the presence of autocorrelation and the need to correct. Although, the OLS2 provides acceptable results relative to spatial models, the reasoning is weakened by our diagnostic tests. Since diagnostic tests suggest the necessity of controlling spatial effects, relegating this would result in implementing results with biased coefficients in the local market.

Thus the performance of the two spatial models we use to control spatial dependence and spatial heterogeneity is quite impressive but the spatial error model has more strength in capturing spatial dependence than GWR in capturing spatial heterogeneity. Like Páez *et al*. (2008) despite the absence of location characteristics (submarkets) in the spatial models, prediction of prices was more accurate than the OLS in this study. To further assert their relevance, it thus follows that location-related attributes (submarkets) are not indispensable to hedonic price modeling in price estimation with non-stationarity and autoregressive methods. This is because their contribution to price is marginal (Basu and Thibodeau, 1998; Páez, Long, and Farber, 2008) as also supported in this analysis.

CONCLUSION

The consensus that the location of a property influences price is not novel. However, capitalizing the influence of location into a hedonic property price framework has been a herculean task despite the use of indicator dummy variables. The uncertainty surrounding the effective measurement of spatial dimensions in the property data has given rise to advances in spatial analysis. Thus, if spatial variables are omitted in the models, the dependency problem leading to bias and inefficient parameter estimates will ensue. Moreover, the existence of spatial heterogeneity cannot be completely removed from property data. This study demonstrates the use of hedonic modeling in property price estimation with an example of data from Cape Town and takes into account advances in spatial analysis. The diagnostic tests reveal the presence of spatial dependence and spatial heterogeneity that require correction. This is consistent with Wilhelmsson (2002), who observes that despite the inclusion of location indicator variables in the data, the null hypothesis of no spatial autocorrelation is rejected.

These concerns impel our use of spatial models namely the spatial error model and GWR relative to the OLS techniques in this study. We adopt the spatial error model because the LM and robust LM tests on the Cape Town data suggest it is suitable for controlling spatial error autocorrelation. The GWR is a technique designed to control spatial heterogeneity in the property data. The performance of spatial models in property price prediction is compared among themselves and with the OLS1 and OLS2 specifications. Apart from the detailed results shown in the differing contributions of variables to price, the goodness-of-fit criteria are used to assess model performance. The results reveal that the spatial models (spatial error model and GWR) constantly performed better than the OLS with and without location indicator variables. This is an indication that the spatial error model and GWR did a good job in tackling the problems of spatial dependence and spatial heterogeneity in the data. However, the spatial error model marginally outperforms the GWR, suggesting that the model did better in correcting spatial error autocorrelation in the Cape Town property market. These two models are therefore recommended for use in the pricing of properties in the Cape Town market.

An area of limitation in this study that provides an opportunity for further research is in the treatment of data we use for modeling prices. We utilize a different specification for the GWR relative to the global models. For instance, in the calibration of variables, a categorical and continuous format is used in the local model, while the binary and continuous formats are used in the global models. It is hoped that in a future study, the same model specification should be used in all models to observe their relative performance. Additionally, a similar study could be undertaken in other property markets within South Africa and or other Sub-Saharan African countries to generalized findings.

Our results provide a platform for the CVO to begin the practical application of spatial modeling in the local market. The GWR and spatial error model as this study reveals to be best suited can be applied in practice with readily available software packages. There are open source and commercial license software packages that the CVO should make concerted efforts to acquire for use within the local market. Additionally, the CVO must liaise with jurisdictions in the United States and the United Kingdom, which are using spatial analysis for pricing properties for taxation purposes, to give basic training on these methodological approaches.

References

Anselin, L. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publisher, 1988.

——. Spatial Econometrics. In: *Companion to Econometrics*, B. Baltagi (ed.). Oxford: Blackwell, 2000.

Anselin, L. and A.K. Bera. Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics. In: *Handbook of Applied Economic Statistics*, Ullah, A. and D.E.A. Giles (eds.). New York: Marcel Dekker, 1998.

Ahmed, A.A. and J.A. Yacim. The Hedonic Modeling of Rental Values of Properties in Nasarawa Nigeria. *Journal of Civil Engineering and Architecture*, 2018, 12, 382–90.

Basu, S. and T.G. Thibodeau. Analysis of Spatial Autocorrelation in House Price. *Journal of Real Estate Finance and Economics*, 1998, 17:1, 61–85.

Belsley, D.A., E. Kuh, and R.E. Welsch. *Regression Diagnostics: Identifying Influential Data and Source of Collinearity*. New York: John Wiley, 1980.

Bitter, C., G.F. Mulligan, S. Dall'erba. Incorporating Spatial Variation in Housing Attribute Prices: A Comparison of Geographically Weighted Regression and the Spatial Expansion Method. *Journal of Geographical Systems*, 2007, 9, 7–27.

Borst, R.A. Discovering and Applying Locational Influence Pattern in the Mass Valuation of Domestic Real Property. Ph.D. thesis submitted to the faculty of engineering, University of Ulster, U.K., 2007.

Borst, R.A. and W.J. McCluskey. Using the Geographically Weighted Regression to Detect Housing Submarkets: Modeling Large-Scale Spatial Variations in Value. *Journal of Property Tax Assessment & Administration*, 2008, 5:1, 21–54.

Bourassa, S.C., E. Cantoni, and M. Hoesli. Spatial Dependence, Housing Submarkets, and House Prediction. *Journal of Real Estate Finance & Economics*, 2007, 35, 143–60.

——. Predicting House Prices with Spatial Dependence: A Comparison of Methods. *Journal of Real Estate Research*, 2010, 32:2, 143–60.

——. Robust Hedonic Price Indexes. *International Journal of Housing Markets and Analysis*, 2016, 9:1, 47–65.

Bowen, W.M., B.A. Mikelbank, and D.M. Prestegaard. Theoretical and Empirical Considerations Regarding Space in Hedonic Housing Price Model Applications. *Growth and Change*, 2001, 23:4, 466–90.

Brasington, D.M. and D. Hite. Demand for Environmental Quality: A Spatial Hedonic Analysis. *Regional Science and Urban Economics*, 2005, 25:1, 57–82.

Brunsdon, C., A.S. Fotheringham, and M.E. Charlton. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, 1996, 28:4, 281–98.

Can, A. Specification and Estimation of Hedonic Housing Price Models. *Regional Science and Urban Economics*, 1992, 22, 453–74.

Casetti, E. Generating Models by the Expansion Method: Applications to Geographical Research. *Geographical Analysis*, 1972, 4:1, 81–91.

——. The Expansion Method, Mathematical Modeling and Spatial Econometrics. *International Regional Science Review*, 1997, 20:1–2, 9–33.

Charlton, M., S. Fotheringham, and C. Brunsdon. Geographically Weighted Regression. NCRM Methods Review Papers. Available at: http://ncg.nuim.ie/ncg/GWR/index.htm, 2005.

Chivers, J. and N. Flores. Market Failure in Information: The National Flood Insurance Program. *Land Economics*, 2002, 78, 515–21.

Court, A.T. Hedonic Price Indexes with Automobile Examples. In: *The Dynamics of Automobile Demand*. New York, 1939.

Crooper, L.M., L.B. Deck, and K.E. McConnell. On the Choice of Functional Form for Hedonic Functions. *The Review of Economics and Statistics*, 1988, 70:4, 668–78.

Daams, M.N., F.J. Sijtsma, and A.J. van der Vlist. The Effect of Natural Space on Nearby Property Prices: Accounting for Perceived Attractiveness. *Land Economics*, 2016, 92:3, 389–410.

de Graaff, T., R.J.G.M. Florax, P. Nijkamp, and A. Reggiani. A General Misspecification Test for Spatial Regression Models: Dependence, Heterogeneity, and Nonlinearity. *Journal of Regional Science*, 2001, 41:2, 255–76.

Dubin, R., K.R. Pace, and T.G. Thibodeau. Spatial Autoregression Techniques for Real Estate Data. *Journal of Real Estate Literature*, 1999, 7, 79–95.

Farber, S and M.A. Yeates. A Comparison of Localized Regression Models in a Hedonic House Price Context. *Canadian Journal of Regional Science*, 2006, 29:3, 405–20.

Goodman, A.C. Hedonic Prices, Price Indices, and Housing Markets. *Journal of Urban Economics*, 1978, 5, 471–84.

Greene, H.W. *Econometric Analysis*. Fifth Edition. Upper Saddle River, NJ: Prentice-Hall, 2003.

Guan, J., J. Zurada, and A.S. Levitan. An Adaptive Neuro-Fuzzy Inference System-Based Approach to Real Estate Property Assessment. *Journal of Real Estate Research*, 2008, 30:4, 395–420.

Haas, T.C. Local Prediction of Spatio-Temporal Process with an Application to Wet Sulphate Deposition. *Journal of the American Statistical Association*, 1995.

——. Multivariate Spatial Prediction in the Presence of Non-Linear Trend and Covariance Non-Stationarity. *Environmetrics*, 1996, 7, 145–65.

Harrison, D. and D.L. Rubinfeld. Hedonic Housing Prices and the Demand for Clean Air. *Journal of Environmental Economics and Statistics*, 1978, 77, 81–102.

Heintzelman, M.D. and C.M. Tuttle. Values in the Wind: A Hedonic Analysis of Wind Power Facilities. *Land Economics*, 2012, 88:3, 571–88.

Keledjian, H. and D.P. Robinson. The Influence of Spatially Correlated Heteroskedasticity on Tests for Spatial Correlation. In: *Advances in Spatial Econometrics*, L. Anselin and R.J.G.M. Florax (eds.). Berlin: Springer-Verlag, 2000.

Kestens, Y., M. Thériault, and F. Des Rosier. Heterogeneity in Hedonic Modeling of House Prices: Looking At Buyers' Household Profiles. *Journal of Geographical Systems*, 2006, 8:1, 61–96.

KPMG. Exploring the Benefits of Technology to Government and Society: A Study of Thomson Reuters Aumemtum in the City of Cape Town, South Africa. KPMG, LLP, 2015, 1–26.

Lancaster, K. A New Approach to Consumer Theory. *The Journal of Political Economy*, 1966, 74:2, 23–9.

Leung, Y., C-L. Mei, and W-X. Zhang. Statistical Tests for Spatial Nonstationarity Based on the Geographically Weighted Regression Model. *Environment and Planning A*, 2000, 32, 9–32.

LexisNexis. Using Technology to Ensure Accurate Property Assessments. *LexisNexis News*. Retrieved from https://www.polity.org.za/article/using-technology-to-ensure-accurate-property-assessments-2018-07-31.

Liu, X. Spatial and Temporal Dependence in House Price Prediction. *Journal of Real Estate Finance and Economics*, 2013, 47:2, 341–69.

Megbolugbe, I.F. A Hedonic Index Model: The Housing Market of Jos, Nigeria. *Urban Studies*, 1988, 26:5, 486–94.

McCluskey, W.J. Real Property Taxation in the Republic of Kazakhstan. *Land Tenure Journal*, 2016, 2, 119–38.

McCluskey, W.J. and R.A. Borst. Detecting and Validating Residential Housing Submarkets: A Geostatistical Approach for Use in Mass Appraisal. *International Journal of Housing Market and Analysis*, 2011, 4:3, 290–318.

McCluskey, W., W. Deddis, A. Mannis, D. McBurney, and R. Borst. Interactive Application of Computer-Assisted Mass Appraisal and Geographic Information Systems. *Journal of Property Valuation and Investment*, 1997, 15:3, 448–65.

McCluskey, W.J., M. McCord, P.T. Davis, M. Haran, and D. McIlhatton. Prediction Accuracy in Mass Appraisal: A Comparison of Modern Approaches. *Journal of Property Research*, 2013, 30:4, 239–65.

McCord, M., P.T. Davis, M. Haran, S. McGreal, and D. McIlhatton. Spatial Variation as a Determinant of House Price: Incorporating a Geographically Weighted Regression Approach Within the Belfast Housing Market. *Journal of Financial Management of Property and Construction*, 2012, 17:1, 49–72.

Militino, A.F., M.D. Ugarte, and L. Garcia-Reinados. Alternative Models for Describing Spatial Dependence among Dwelling Selling Prices. *Journal of Real Estate Finance and Economics*, 2004, 29:2, 193–209.

Mok, H.M.K., P.P.K. Chan, and Y.S. Cho. A Hedonic Price Model for Private Properties in Hong Kong. *Journal of Real Estate Finance & Economics*, 1995, 10, 37–48.

Mueller, J.M. and J.B. Loomis. Spatial Dependence in Hedonic Property Models: Do Different Corrections for Spatial Dependence Result in Economically Significant Difference in Estimated Implicit Prices? *Journal of Agricultural and Resource Economics*, 2008, 33:2, 212–31.

Pace, R.K and O.W. Gilley. Using the Spatial Configuration of the Data to Improve Estimation. *Journal of Real Estate Finance and Economics*, 1997, 14:3, 333–40.

Pace, R.K., R. Barry, J.M. Clapp, and M. Rodriquez. Spatiotemporal Autoregressive Models of Neighborhood Effects. *Journal of Real Estate Finance and Economics*, 1998, 17:1, 15–33.

Páez, A., F. Long, and S. Farber. Moving Windows Approach for Hedonic Price Estimation: An Empirical Comparison of Moving Techniques. *Urban Studies*, 2008, 45:8, 1565–81.

Páez, A., T. Uchida, and K. Miyamoto. Spatial Association and Heterogeneity Issues in Land Price Models. *Urban Studies*, 2001, 38:9, 1493–1508.

Pavlov, A.D. Space-Varying Regression Coefficients: A Semi-Parametric Approach Applied to Real Estate Markets. *Real Estate Economics*, 2000, 28:2, 249–83.

Rosen, S. Hedonic Prices and Implicit Markets, Product Differentiation in Pure Competition. *Journal of Political Economy*, 1974, 82, 218–33.

Schulz, R., M. Wersing, and A. Werwatz. Automated Valuation Modeling: A Specification Exercise. *Journal of Property Research*, 2014, 31:2, 131–53.

Shimizu, C. Estimation of Hedonic Single-Family House Price Function Considering Neighborhood Effect Variables. *Sustainability*, 2014, 6, 2946–60.

Sirmans, G.S., D.A. Macpherson, and E.N. Zietz. The Composition of Hedonic Pricing Models. *Journal of Real Estate Literature*, 2005, 13:1, 3–43.

Stevenson, S. New Empirical Evidence on Heteroscedasticity in Hedonic Housing Models. *Journal of Housing Economics*, 2004, 13, 136–53.

Thériault, M., F. Des Rosier, P. Villeneuve, and Y. Kestens. Modeling Interactions of Location with a Specific Value of Housing Attributes. *Property Management*, 2003, 21:1, 25–48.

Tse, R.Y.C. and P.E.D. Love. Measuring Residential Property Values in Hong Kong. *Property Management*, 2000, 18:5, 366–74.

Ward, R.D. Demonstration of CAMA in South Africa. *Assessment Journal*, 2001, 8:3, 33–43.

Wheeler, D.C. Diagnostic Tools and a Remedial Method for Collinearity in Geographically Weighted Regression. *Environment and Planning A*, 2007, 39:10, 2464–81.

Wilhelmsson, M. Spatial Models in Real Estate Economics. *Housing, Theory and Society*, 2002, 19, 92–101.

Yacim, J.A. and D.G.B. Boshoff. A Comparison of Bandwidth and Kernel Function Selection in Geographically Weighted Regression for House Valuation. *International Journal of Technology*, 2019, 10:1, 58–68.

Yang, Z. An Application of the Hedonic Price Model with an Uncertain Attribute: The Case of the People's Republic of China. *Property Management*, 2001, 19:1, 50–63.

Zurada, J., A.S. Levitan, and J. Guan. A Comparison of Regression and Artificial Intelligence Methods in a Mass Appraisal Context. *Journal of Real Estate Research*, 2011, 33:3, 349–87.

Douw Gert Brand Boshoff, University of Cape Town, Cape Town 7701, South Africa or drdgbboshoff@gmail.com.

Joseph Awoamim Yacim, University of Pretoria, Hatfield 0028, South Africa and Federal Polytechnic Nasarawa, Nasarawa 962101, Nigeria or joseph.yacim@gmail.com.

Abdullahi Alhassan Ahmed, Federal Polytechnic Nasarawa, Nasarawa 962101, Nigeria or aaahmednas@gmail.com.