MODELING SPATIAL VARIATIONS OF HOUSING PRICES IN TORONTO, ON

MODELING SPATIAL VARIATIONS IN HOUSING PRICES
------AN INVESTIGATION OF THE INDIVIDUAL AND JOINT EFFECTS OF SPATIAL AUTOCORRELATION AND SPATIAL HETEROGENEITY

BY
FEI LONG B.Sc. (Hons)

A THESIS
Submitted to the School of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree
Master of Arts

McMaster University

MASTER OF ARTS (2006)
(School of Geography and Earth Sciences)

McMaster University
Hamilton, Ontario

Title: Modeling Spatial Variations in Housing Prices
------An Investigation of the Individual and Joint Effects of
Spatial Autocorrelation and Spatial Heterogeneity

Author: Fei Long, B.Sc.

Supervisor: Dr. Antonio Páez

Number of Pages: xiii, 137

ACKNOWLEDGEMENTS

The thesis present here could not possibly be finished without help and support from many people. First, I want to express my sincere gratitude to my supervisor, Professor A. Páez, who brings me into this new promising field of spatial data analysis and opens a new door in front of me. My background in economic geography does not provide me sufficient statistic knowledge and programming skills required for this study. Without his understanding, guidance and generous assistance, I would not overcome these problems and stick to the end. I also benefit greatly from his incisive and thorough comments during the process of conducting analysis and writing thesis. If any people think it is a thesis with good quality, credit should go to him.

I would also like to thank my committee members, Professor A. Dickin, Professor P. Kanaraglou and Professor D. Scott. I am grateful for the precious time they devoted to reading the thesis and their intellectual comments and advices. I wish to extend my gratitude to Professor K. Liaw who taught a fantastic course in statistics and rescued me from aversion of it. I appreciate the knowledge and skills I received from him. Also thanks to the wisdom he shared with me about life and about being a researcher.

Thanks also to Patrick Deluca, systems administrator of the Center for Spatial Analysis. He maintained the machine in a perfect condition and always updated the software timely. Without his effort and trouble-shooting skills, the research will not run very smoothly. Special thanks are due to my fellow students especially Karen, Rubern, Dora, Sylvia, Hejun, Justin, Dominic, Kelly. They gave me a warm and friendly welcome when I first came to McMaster University. In the following two years, they offered numerous helps not only in academic but also in other aspects of my life. Without their friendship, I would not lead an enjoyable life in this new country.

Finally, I want to express my gratefulness to my family. Their endless love and unconditional support made me fearless of failure. No matter how far we are apart, they are close to my heart and are always with me. Thanks to my boyfriend for enduring my bad temper during some stressful time and comfort me with great patience.

ABSTRACT

Hedonic analysis is an indispensable tool in the study of urban housing market, particularly in investigating the relationship between residential property values and a set of perceived explanatory variables. However, traditional hedonic modeling based on the use of ordinary least squares (OLS) approach may not be adequate due to spatial effects, namely spatial dependency and spatial heterogeneity.

In this light, Universal Kriging, moving window kriging as well as local regression approaches (including moving window regression and geographically weighted regression) are employed to incorporate spatial effects into hedonic modeling. These approaches not only have the advantage of explicitly modeling spatial process, but their model specifications also allow the analyst to distinguish and identify the individual and joint effects of spatial autocorrelation and spatial heterogeneity.

Using data from the City of Toronto, the objective of this study is to diagnose spatial effects in the process of housing price determination and conduct comparative analysis among different models in terms of out-of-sample prediction accuracy. The results demonstrate that proper incorporation of spatial autocorrelation or spatial heterogeneity improves model performance substantially. In addition, for this particular dataset, spatial heterogeneity plays a larger role in explaining housing price variations; and finally inclusion of one spatial effect may provide adequate control for the other one, therefore, no discernable improvement is gained by incorporating both of them into modeling process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

Hedonic analysis, a method of decomposing the commodity being studied into its constituent characteristics and estimating their implicit prices (Lancaster 1966; Rosen 1974), is an indispensable tool in the study of urban housing market dynamic. Its extensive utility in this field includes: estimating demand for housing and neighborhood attributes; making general improvement in house price indices; analyzing the impact of neighborhood externalities; measurement of housing demand in residential mobility studies; estimation of the benefits to accrue from public investment programs; appraisal of individual housing units; examining the capitalization of a wide range of amenities (Can 1992; Malpezzi 2003).

In the long history of hedonic housing studies, traditional hedonic regression analysis dominated the field until the early 1990's. Since then, with the emerging technology of geographic information systems as well as the continual improvement on spatial statistics, a number of novel and remarkable studies such as Can (1992), Dubin (1992), Basu and Thibodeau (1998) among others have brought hedonic analysis into a promising new area ------ exploiting the spatial nature of residential dataset. As expounded in these studies, given the existence of spatial autocorrelation and spatial heterogeneity, the straightforward use of traditional hedonic model, which is non-spatial in nature, is not sufficient for analysis and modeling of housing datasets. According to a different understanding and emphasis of price determination process, special methods and techniques have been advanced to address the issue of spatial structure of residential properties.

Following these studies, more researchers have realized the limitations of traditional hedonic analysis, its insensitivity to space, which is an inherent nature of housing datasets. Aware of the potential consequence of absent spatial effects upon statistic validity of model estimation, hedonic studies that control for spatial dependence and spatial heterogeneity have proliferated rapidly in the last decade. .

A majority of these studies, however, only deal with one spatial effect explicitly, either spatial dependence or spatial heterogeneity in their modeling framework; very few hedonic applications attempt to incorporate both effects into model specification. Moreover, literatures often look at them separately, that is investigating one spatial effect intensively while neglecting the possible impact of the other one. In light of this, the purpose of this study is to investigate the respective and joint effects of spatial dependence and spatial heterogeneity by isolating and combining them in various spatial models.

While recognizing that there are many paths that can be taken for the fulfillment of incorporating spatial effects, four advanced techniques are deliberately selected in that the special connections among their model formulations make isolation and combination of two spatial effects feasible. Through comparative analysis of prediction accuracy between different pairs of models, some research questions are raised. First, which type of spatial effects, spatial dependence or spatial heterogeneity, plays a larger role upon increasing prediction accuracy? Secondly, when one spatial effect is already embedded into model specification, will the inclusion of the remaining one substantially enhance the prediction accuracy and whether the joint effect is the summing up of their individual effect? These queries will be explored in Chapter 4 using housing price information collected from City of Toronto.

Emphasis is made here that constrained by data limitations, the focus of this study is not to search for a comprehensive hedonic price function with inclusive positive and negative externalities, but rather to investigate the contribution of isolated or joint spatial effects to out-of-sample prediction accuracy. Therefore, little attention will be paid to the filtering of determinants or the plausibility of coefficient estimates.

The thesis is organized into four chapters, following the present introduction.

In chapter 2, the theoretical basis of traditional hedonic price functions is detailed. Although the microeconomic foundation for hedonic price model framework is well articulated in the econometric literature, its functional specification in empirical application has not been investigated systematically. This article seeks to furnish this part by scrutinizing four aspects of hedonic price specification including selection of characteristics vectors, selection of functional forms, assumption on behavior of parameters as well as assumption upon behavior of residuals. The in-depth inspection of these components reveals the need for spatial consideration, including spatial autocorrelation and spatial heterogeneity, in hedonic price structure. Then focus is put on the characteristics of urban housing market, i.e. locational effects and market segmentation. By exploring the relationships between locational effects and spatial dependence, market segmentation and spatial heterogeneity, the rationale for spatial consideration is established. At the end of chapter 2, the consequence of omitting such spatial effects is present with respect to the statistical validity of model inference.

Chapter 3 'Data and Methods' first introduces the dataset employed for empirical analysis conducted in this study. It is a large transaction based dataset containing 33,494 residential sales in City of Toronto, Ontario from January 2001 to December 2003. Section 3.2 and 3.3 describe formal diagnostic tests on spatial autocorrelation and spatial heterogeneity. In accordance with potential diagnosis of spatial effects, alternative spatial hedonic models are proposed, including moving window regression (MWR), geographically weighted regression (GWR), Universal Kriging, and moving window kriging (MWK). Some of the above approaches explicitly address spatial heterogeneity like MWR and GWR. Universal Kriging is more concerned with the specification of spatial dependence. A more recently advanced technique, MWK, takes both spatial dependence and spatial heterogeneity into consideration.

Chapter 4 is 'Results and Discussion'. It first presents initial regression results from a traditional hedonic price function. Based on the initial regression results, rigorous testing on the presence of spatial autocorrelation and spatial heterogeneity is carried out which makes the incorporation of spatial effects necessary for model specification. Then, the link between optimal window size and cross-validation procedure is described, illustrating some of the complexities associated with dealing with spatial process in hedonic price studies. Afterwards, with the purpose of facilitating the application of Universal Kriging in this extremely large dataset, a sample thinning approach is utilized. The stable empirical variogram estimates and constant predictive performances attained by various random samples justify the use of small sample, instead of whole population, for predicting out-of-sample observations. The last sections provide two rounds of comparative assessments of the prediction accuracy of various spatial hedonic specifications. Through the comparative analysis between different pairs of models, the individual and joint effects of spatial autocorrelation and spatial heterogeneity upon model predictive power are distinguished and recognized.

Chapter 5 concludes the study. It summarizes results, provides potential application in property appraisal and taxation and suggests future research directions.

Chapter 2 Background

Hedonic regression analysis is an essential component for urban housing studies. It allows the decomposition of housing expenditure into multiple characteristics internal or external to housing units (Lancaster 1966; Rosen 1974; Sirmans et al. 2005). Therefore hedonic prices, i.e. implicit prices of attributes, can be obtained from multivariate regression analysis between observed prices of differentiated dwellings and the quantity and quality of characteristics associated with them.

The rationale of hedonic regression model is rooted in Lancaster's new consumer behavior theory, which 'breaks away from the traditional approach that treats good as the direct objects of utility, instead, supposes it is the properties or characteristics of the goods from which utility is derived '(Lancaster 1966). Under such framework, 'housing is a multidimensional good differentiated into a bundle of attributes that vary in both quality and quantity' (Can 1990); moreover, accredited to the utility-generating nature of attributes, their values can be priced and total housing expenditure can be essentially broken down into embodied characteristics. Another milestone of hedonic modeling is Rosen's (1974) work, which stresses the interaction between suppliers and consumers through bids and offers for characteristics in the price determination process.

With respect to the original source of hedonic modeling, although Court (1939) is often viewed as the father of hedonic modeling who developed a hedonic price index for automobile, other studies pointed out earlier hedonic work could date back to the 1920's (Malpezzi 2003).

2.1 Model Specification of Traditional Hedonic Model

As with any other econometric approach, the specification of hedonic modeling plays a crucial role in determining the accuracy and precision of modeling results (Can 1997). In the following section, four principal components of modeling specification will be discussed including the selection of characteristics vectors, the functional form, and the assumptions both on parameter vectors and on random error terms.

2.1.1 Selection of Characteristics Vectors

Model specification starts with the selection of explanatory variables or contributing factors. For the fulfillment of estimating the implicit marginal prices of attributes accurately, it requires including a full set of all significant determinants of housing prices into regression; specificly, the characteristics vectors should capture all externalities that yield utility to households and constitute market values of properties (Bowen et al. 2001).

Despite a substantial body of empirical works completed in an attempt to explore the best suite of explanatory variables in terms of explaining most variations in housing sale prices and also satisfying the requirement of parsimony, this optimum set can still not be fully determined a priori. Indeed, the prospects for this endeavor are bleak due to the diversity and complexity of urban housing market dynamics operating in different regions. Therefore, knowledge and experience on local real estate market is important to aid making better judgment regarding the selection of characteristics vectors.

Even though there is no consensus in the literature regarding the specific variables to be included into hedonic price function, characteristics in three basic categories are generally considered. The first category includes the structural characteristics (*S*) of residential property, such as dwelling size, the number of rooms, age of building, housing condition and so on; the second category is comprised of characteristics of immediately surrounding social and natural environment (*N*), such as mean household income, crime rates, quality of social amenities, racial composition of the neighborhood in which the house is located; the third group includes locational characteristics (*L*), such as distance to major employment center, to transportation networks, and proximity to public service, to recreation and shopping facility, etc (Basu and Thibodeau 1998; Bowen et al. 2001).

After selecting appropriate characteristics, the market value of housing unit is generally agreed to be expressed as the following hedonic price function:

$$Y = f(S, N, L) \tag{2.1}$$

Since linear functional form is preferred for this study (the reason will be given in section 2.1.2), the above specification can be further specified as:

$$Y_i = \alpha_i + \sum_k \beta_{ki} S_{ki} + \sum_p \gamma_{pi} N_{pi} + \sum_j \lambda_{ji} L_{ji} + \varepsilon_i \tag{2.2}$$

As stated in the traditional regression model, $Y_i (i = 1, 2, ..., n)$, where n is the number of observations in the dataset, represents housing expenditures which are commonly measured by the market value of property for owners or by annual rent for renters. In this study, to avoid potential bias brought by self-reported appraisals, recent transaction prices obtained from open market sales stand as a proxy for property values. ***S*** is a vector of structural attributes; ***N*** is a vector representing attributes of surrounding natural and social environment; ***L*** is a vector of variables capturing locational or proximity characteristics; ε is a vector composed of random error terms which represents all those factors that affect sale prices but are omitted from the modeling process (Bowen et al. 2001).

Structural characteristics are typically observable and in most cases are easy to quantify. On the contrary, information about property's neighborhood and location characteristics is difficult to acquire since these attributes are both difficult to observe and to measure (Dubin 1992; Basu and Thibodeau 1998). Urban analysts are aware that neighborhood quality and accessibility to social amenities are important predictor in house price determination process in that they yield utility for households and therefore affect housing prices significantly; however, there is no agreement among scholars regarding which variables best proxy neighborhood and accessibility quality. Perhaps, more seriously, even in the best circumstances where intensive empirical work helps determining the most suitable variables, severe measurement problems for both types of attributes may be present that invalidate the efforts.

Specifically, with regard to locational attributes, with the process of urbanization and urban sprawl, the traditional view of monocentric urban structure becomes suspect. Instead, cities have developed into polycentric structure with multiple urban centers (Des Rosiers et al. 2000). Each center plays a dominant role is shaping the nearby area and therefore develops their own zones of influence. As a consequence, the traditional means of capturing accessibility effects by measuring the distance to the CBD area may result in incomplete or even erroneous conclusion (Dubin 1992). For this reason, a more flexible means to describe the polycentric structure and allow for multiple peaks in the rent surface is needed in order to capture the location characteristics faithfully (Dubin 1992).

With respect to neighborhood characteristics, since they are geographic in nature, in order to measure them, the boundaries of neighborhood must be known in advance (Dubin 1992, 1998). However, in practice, relatively little attention has been paid in the process of defining and delineating neighborhood boundaries. The common solution taken by researchers is to assume neighborhood boundaries to implicitly coincide with the boundaries used by data collectors or to simply ignore such issues by using information available at hand (Dubin 1992, 1998; Basu and Thibodeau 1998). For instance, due to the easy access and rich information of census data, it is the most commonly used data resource among researchers; therefore, census tract boundaries are easily treated as being conterminous with neighborhood boundaries. In other cases, if police data are available and applied, then crime reporting area boundaries are taken as representing the neighborhood (Dubin 1992, 1998).

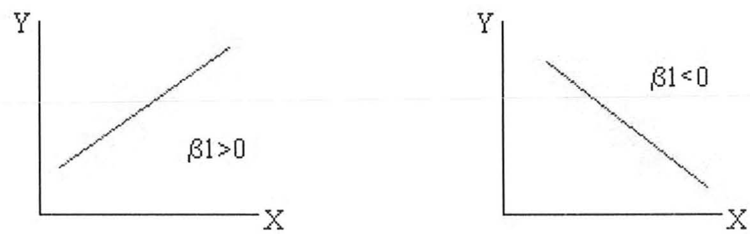
As illustrated above, it is fairly difficult to choose appropriate variables to best proxy neighborhood and locational features (Dubin 1992; Basu and Thibodeau 1998). Additionally, even when attempts of selecting desirable variables succeed, difficulties associated with measuring them pose another obstacle to the proper specification of hedonic modeling. Due to the above reasons, it is not surprising to see weak evidence regarding capitalization of neighborhood and locational effects in the literature (Dubin 1992).

2.1.2 Functional Form of Hedonic Modeling

In spite of a relatively long history of applying hedonic model in various practices, there is no strong theoretical basis for choosing the correct functional form of a hedonic regression. Lancaster and Rosen's studies, for example, present models of housing characteristics without specifying how the numerous characteristics are related to housing prices (Malpezzi 2002). Nevertheless, three functional forms are commonly used which are linear, log-log and semi-log. To clarify the differences among them, simplified models using only one regressor are illustrated as follows:

- *Linear Functional Form.*

This functional form has the equation: $Y = \beta_0 + \beta_1 * X + \varepsilon$, and its graph is shown as below:

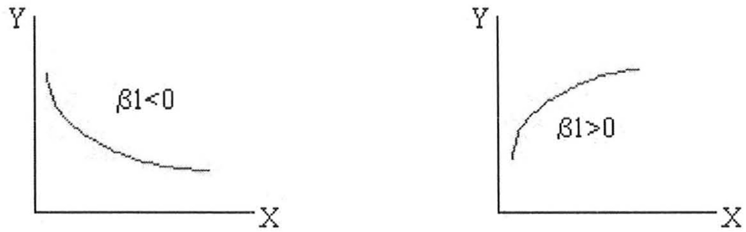


(source: http://www.union.edu/PUBLIC/ECODEPT/schmidsj/funforms.html)

The advantage of the linear functional form is its simplicity. Each time X goes up by 1 unit, Y goes up by β_1 units, and this is true no matter what the values of X and Y are.

- *Logarithmic Functional Form.*

This functional form has the equation: $\log Y = \beta_0 + \beta_1 * \log X + \varepsilon$, and its graph is shown as below:

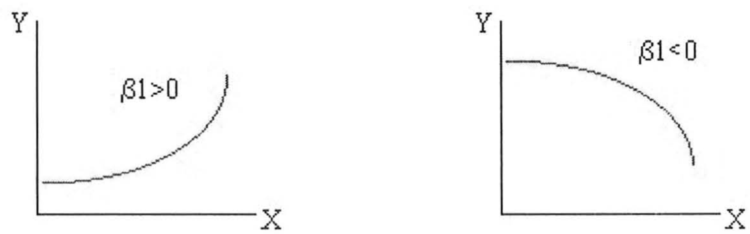


(source: http://www.union.edu/PUBLIC/ECODEPT/schmidsj/funforms.html)

The special attribute of the logarithmic functional form is that if X rises by 1%, then Y will rise by β_1%.

- *Semi-log Functional Form.*

The equation for this functional form is: $\log Y = \beta_0 + \beta_1 * X + \varepsilon$, and its graph is shown as below:



(source: http://www.union.edu/PUBLIC/ECODEPT/schmidsj/funforms.html)

The semi-log functional form has the property that if X rises by 1 unit, Y rises β_1*100%.

Some researchers prefer to model the relationship between housing prices and its characteristics using a semi-logarithmic functional form. This specification of regressing the log of transaction prices on a linear combination of housing characteristics is selected for a variety of economic and statistical reasons.

On economic ground, as pointed out before, in semi-log functional form, the estimated coefficient of a particular attribute can be interpreted as the proportional change on housing values as the attribute rises by 1 unit. Thus, by allowing the marginal price of characteristics to vary with other characteristics in the bundle (Dubin 1998; Basu and Thibodeau 1998), this functional form provides a more faithful depiction of the real world. Unfortunately, as derived from its definition, linear functional form lacks this desirable property of allowing for interaction between independent variables. On the contrary, the marginal price of an attribute in linear functional form remains constant regardless of the condition of other attributes.

On statistical ground, semi-log functional form is preferred since by adjusting the scale of dependent variable into its log form, the estimation problems associated with heteroskedasticity will be reduced or eliminated (Dubin 1998; Basu and Thibodeau 1998). Briefly speaking, heteroskedasticity implies that prediction error tend to be large in absolute value, as property value increases.

In this study, linear functional form is applied for several reasons. Firstly, in linear functional form, the estimated coefficients can be easily interpreted as the implicit marginal prices of attributes. For instance, if number of rooms is one of the explanatory variables used in modeling and the corresponding coefficient for this attribute is 2000, then it can be readily concluded that an additional room will bring 2000 dollars to the market value of properties. However, as semi-log functional form implies, the resulting regression coefficients represent the marginal effect of a unit change in attributes on the transformed price, therefore making them difficult to interpret and clouding their intuitive interpretation as implicit marginal prices (Bowen et al. 2001).

Secondly, as will be seen later, the major interest of this study lies in comparing prediction capacity of a variety of spatial hedonic models. Conditioned by this research goal, the logarithmic functional form can be problematic in the sense that unbiased estimation of regression coefficients in units of log price could be biased when transformed back to original price (Goldberger, 1968).

Moreover, a majority of models explored later belong to the domain of local modeling techniques featured by the utilization of moving window approach. Unlike global models performed at large scaled space, moving window approach use relatively homogeneous subsets of data defined by the close vicinity of their physical locations. Thus, attributed to this special design, the estimation problems associated with heteroskedasticity can also be reduced or eliminated using the linear functional form.

Lastly, most techniques employed in later analyses involve the measurement of spatial autocorrelation among residuals or calculation of variance-covariance structure of residuals. However, transformed residuals obtained in semi-log functional form may have potential undesirable effects upon usage of these models since this transformation will likely obscure the original spatial pattern and subsequently invalidate the measurement of spatial autocorrelations or the construction of variance-covariance structure. Or if it does not, it will at least confound the situation, which is already complex due to the usage of intricate techniques.

2.1.3 Assumption on Behavior of Parameters

As stated in section 2.1.1, the traditional hedonic model, which relates the market value of properties to a number of influential factors, can be specified as:

$$Y_i = \alpha_i + \sum_k \beta_{ki} S_{ki} + \sum_p \gamma_{pi} N_{pi} + \sum_j \lambda_{ji} L_{ji} + \varepsilon_i \tag{2.2}$$

A simplified version of the above function using vector notions can be stated as:

$$Y = X\beta + \varepsilon\,, \tag{2.3}$$

where $\boldsymbol{Y}$ is a $(n*1)$ vector representing the observed sale prices of n housing units on the market; $\boldsymbol{X}$ is a $(n*k)$ vector reflecting K structural, locational and neighborhood characteristics for housing units; $\boldsymbol{\beta}$ is a $(k*1)$ vector of unknown coefficients; $\boldsymbol{\varepsilon}$ is another $(n*1)$ column vector representing the stochastic disturbance term.

Traditional hedonic price function is a typical econometric regression model and ordinary least squares (OLS) is by far the standard technique to estimate the unknown coefficients computed as:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{2.4}$$

According to the above equation, ordinary least squares approach will provide a single set of coefficient estimates for all observations within the dataset, which means traditional hedonic model assumes a set of invariant or “fixed” coefficient over space.

As introduced before, in the linear regression model, regression coefficients can be thought of as implicit marginal prices of attributes, specifically, the change on market value of property when the magnitude of attribute goes up by 1 unit. Such inference is conditioned by the term that regression model used in an application conforms to three criteria: parsimoniousness, plausibility and informativeness (Bowen et al. 2001). To satisfy these requirements, a model should contain the minimum number of parameters required for explaining the key concepts; its variables, concepts, and the stipulated relations between them are justifiable in terms of being coherent with the larger body of knowledge; it should provide situational guidance appropriate to the application at hand (Bowen et al. 2001).

With conformity to the criteria above, a set of universal coefficients also implies invariant marginal prices for attributes over space. In other words, the contributions to the market value of properties of adding one unit of housing service are constant over space. However, the appropriateness of this simplifying assumption has generated considerable debate by posing a simple question: what are the effects of adding one bedroom to a house located in the center of city and to a house in a suburban area? Will this addition bring fairly equivalent monetary worth to market values of these two houses? In most cases, the answer would be no.

The questions about uniform relationship between housing value and its determinants has been echoed by a number of works (Schnare and Struyk 1974; Brunston et al. 1996, 1998; Goodman and Thibodeau 1998; Pavlov 2000; Farber 2004) which posit that the influence of housing attributes on sale prices is characterized by spatial variability. This variation in the behavior of a given process across space is often recognized in urban housing market studies and referred as spatial heterogeneity, which is believed to be caused by the different spatial dynamics operating in local urban housing markets (Can 1990). However, as Kestens et al. (2006) argued, the implicit prices of some property specifics and location attributes are not only defined by property's geographic locality, but also partly linked to individual preference, that is marginal value given to property and location attributes may vary significantly with the household profile, i.e. type of household, age, income, educational attainment, etc. Although it is still arguable whether the heterogeneity identified in parameter estimates refers to household heterogeneity or spatial heterogeneity (Kestens et al. 2006), as well pronounced in the literature and also constrained by data limitation, this study will only focus on spatial heterogeneity and insights into it will be provided in later section.

2.1.4 Assumptions on Behavior of Residuals

As is well-known, in OLS the regression coefficients β are estimated by minimizing the sum of squared errors $\varepsilon^T \varepsilon$, and accordingly the market value of a property with characteristics X_0 is estimated as $X_0 \beta$. To ensure that OLS is the best linear unbiased estimator, there is a set of ideal conditions that should be satisfied: X must be independent of the errors; and the errors themselves must be independent, homoskedastic and normally distributed. The assumption of spherical distributed error terms can be decomposed into the following expressions:

1. $E[\varepsilon] = 0$;

2. $E[\varepsilon \varepsilon^T] = \sigma^2 I$

3. The error terms are multivariate normal.

In accordance with the above assumptions, a variance-covariance matrix of disturbance terms can be constructed: the variances of disturbance terms are positioned along the main diagonal; the covariance between them situate in the off-diagonal positions. In addition, the assumption of homoscedasticity implies that error term is thought to be drawn from the same distribution and then has same variance, therefore the diagonal terms all have the same values; meanwhile, as indicated by the assumption of independence, there is no tendency for error of one observation to be associated with error of another one which makes off-diagonal terms equal to zero (Bowen et al. 2001).

However, given geographic nature of the residential dataset used in hedonic price models, it is natural to anticipate the covariance among errors is not equal to zero, instead, it may be a function of spatial proximity among houses (Can 1997). The plausibility of the assumptions mentioned before is not only suspect from a theoretical perspective, it has also been challenged by many empirical hedonic house price studies which often observe occurrence of interdependence among observations (the cluster of similar or dissimilar values) in geographic space. This commonly recognized phenomenon for cross-sectional data or geographically referenced data is referred as spatial autocorrelation or spatial dependence. The consequence of violating assumptions regarding residuals and the cause of spatial autocorrelation will be discussed subsequently.

2.2 Two Characteristics of Urban Housing Market

As discussed in earlier sections, the assumptions taken by ordinary least squares (OLS) approach on behavior of residuals and on behavior of parameters may not be tenable due to the geographic nature of residential dataset. After careful inspections, two issues regarding spatial effects were raised, namely, spatial dependency and spatial heterogeneity, which are believed to be main factors that violate the assumptions (Can 1990, 1992; Páez et al. 2001)

In this section, a closer look at urban housing market is given. As outlined below, urban housing market is characterized by two features------ locational effects and market segmentation (Basu and Thibodeau 1998), which cause spatial dependency and spatial heterogeneity respectively.

2.2.1 Locational Effects

Locational effects can be defined as attributes associated with the geographic location of properties. To be specific, the geographic location comprises two elements: both the absolute location of a property and the neighborhood in which it is located. Along the same line, Can (1992) distinguished two levels of locational effects: (1) neighborhood effects, which is the array of locational characteristics; and (2) adjacency effects, which are externalities associated with the absolute location of structures. Detailed illustration of both effects will be provided later in this section.

After introducing two forms of locational effects, it is easy to understand why autocorrelation commonly exists among market value of residential properties in geographic space and how the local deviations from the mean value in the housing market "follow" each other in neighboring locations.

With regard to neighborhood effects, first, neighbors tend to develop at the same time, so neighboring properties share similar structural characteristics such as dwelling size, interior and exterior design features (Can 1992; Basu and Thibodeau 1998; Can et al., 1999). Secondly, neighboring properties are close in space, which give them more or less the same proximity to social amenities (Can 1992; Basu and Thibodeau 1998). In addition, a majority of neighboring properties will locate in the same neighborhood, for instance, in the same administrative unit. Therefore, the social environmental variables assigned to them like the mean household income, quality of public service will be identical.

The traditional hedonic model capitalizes the neighborhood effects by including a set of characteristics which account for the socioeconomic and physical make-up of the neighborhood and the accessibility of properties to urban amenities (Can 1992). The conceptualization of neighborhood effects is made operational by entering relevant variables into hedonic price function as direct determinants of housing values.

Can (1990) advanced an alternative approach, based on the spatial expansion model (Casetti 1972), to traditional hedonic price specification as a way to measure and quantify neighborhood externalities. Can argued that unlike traditional hedonic model which treats neighborhood effects as direct determinants and therefore creates constant coefficient estimates for each predictors, spatial expansion model assumes structural attributes produce differing price differentials depending on location. More specifically, it asserts that instead of being fixed, parameters for structural attributes take different values across space according to the socioeconomic and environmental variation present in each neighborhood (Can 1990, 1992). The spatial expansion model is specified as:

$$Y = \alpha + \sum (\beta_{k0} + \beta_{k1} N + \beta_{k2} L) S_k + \varepsilon \qquad (2.5)$$

where neighborhood attributes are not direct entries into model specification, instead, they constitute spatial context which links parameters of structural attributes to the natural and socioeconomic environment of neighborhood and therefore provides unique coefficient estimates of structural attributes for each particular location.

However, with a more cautious consideration of spatial phenomena and a sophisticated model design, spatial expansion model doesn't necessarily generate more accurate housing price estimation. As demonstrated in Can's studies (1992) that traditional hedonic model and spatial expansion model are comparable in terms of explanatory power when using the same set of explanatory variables. Therefore using neighborhood attributes to construct contextual drift doesn't differ much from entering them as direct determinants, at least in terms of model performance. Additionally, generating 'drift parameters' of structures, is a trend-fitting exercise that is of limited use when parameters exhibit complex variation over the space being studied (Brunsdon et al. 1996). Due to the above considerations, spatial expansion model is not applied in this research.

As illustrated above, similarity in prices of nearby houses is partially explained by neighborhood effects, i.e. sharing of similar neighborhood and locational attributes. In addition, adjacency effects, or spillover effects also contribute to similarities among neighboring properties. Intuitively, the effect is similar to a realtor's assessment of a particular house based on the price history of its immediate neighbors, or maintenance /repair decisions of a house affecting the market value of its nearby properties (Can 1990, 1992).

The hedonic price function which takes adjacency effects into model specification was proposed by Can as an alternative to traditional hedonic price function. It is termed spatially autoregressive model (SAR) and defined as:

$$Y = \alpha + \sum_{k} \beta_k S_k + \sum_{i} \gamma_i N_i + \sum_{j} \lambda_j L_j + \rho WY + \varepsilon \tag{2.6}$$

Can (1990, 1992) declared that this model specification not only captured contextual drift, that is including a set of structural, accessibility and neighborhood attributes to account for neighborhood effects, but also acknowledged the presence of spillover effects on the price of a house given the price of nearby dwellings by designing and including an autoregressive term (WY); the coefficient of the spatially lagged dependent variable ρ measures the degree to which nearby houses will have an absolute impact on the price of a given one.

An alternative to the spatial autoregressive model is the autocorrelated error model, which incorporates a spatially autoregressive error term into the functional specification as follows:

$$\begin{aligned} Y &= \alpha + \sum_{k} \beta_k S_k + \sum_{i} \gamma_i N_i + \sum_{j} \lambda_j L_j + \varepsilon \\ \varepsilon &= \lambda W \varepsilon + \mu \end{aligned} \tag{2.7}$$

In essence, there two models are similar in terms of taking adjacency effects into account by including a spatially lagged variable; however, the autocorrelated error model relegates spatial dependence to the error terms and as such it does not offer any behavior explanation of interdependence (Can 1990).

It is natural to anticipate that spatial dependence will diminish after capitalizing neighborhood effects or adjacency effects or both of them into hedonic model by making use of one of the above approaches. And also a number of studies (Pace and Gilley 1997; Basu and Thibodeau 1998; Bowen et al., 2001; Farber, 2004) have proved the superiorities of spatial autoregressive model and autocorrelated error model over traditional hedonic regression model in terms of improving the explanatory power and reducing spatial autocorrelation.

However, spatial autocorrelation is not completely eliminated after use of advanced approaches since at times correlated or even highly correlated residuals are still present after the modeling. One possible explanation for this phenomenon is that variables which proxy the locational externalities are typically imperfectly measured or simply immeasurable (Basu and Thibodeau 1998; Bowen et al., 2001). As discussed in section 2.1.1, while most structural characteristics are relatively easy to measure and are typically included in publicly available datasets, characteristics about locational externalities or neighborhood attributes such as distance to CBD, proximity to public school, quality of neighborhood environment, crime rate and so on are more difficult to measure and are rarely included in public datasets.

Therefore, due to the difficulty of gathering relevant information on location and neighborhood attributes, some of these important characteristics are simply omitted from many empirical house price studies. If these attributes are essential components in determining housing market value, then it is not surprising to detect spatially autocorrelated residuals in their absence. Additionally, even when a comprehensive dataset including rich information regarding locational and neighborhood characteristics is available, residuals may still correlate since no guarantee is given to the precision and accuracy of the measured variables considering the lack of ideal ways to measure them.

Recalling section 2.1.4, residuals are the stochastic disturbance terms from classical regression theory which represent all those factors that have impact upon house prices but are not incorporated into modeling (Dubin et al. 1999). If any systematic pattern is observed in residuals like the clustering of similar or dissimilar values, then it is likely that potentially valuable information has not been retrieved from the process.

To make full use of this residual information and also to correct for spatially autocorrelated error terms resulting from missing or incorrectly measured variable or other forms of misspecification present in traditional hedonic model, spatially autoregressive model or autocorrelated model, this article will introduce alternative regression technique to OLS, generalized least squares (GLS) and its equivalent technique Kriging in chapter 3 and apply them into the dataset being studied. Under Kriging, rather than eliminating the problem of spatial residual dependence through complicated model specification with comprehensive externalities (Dubin et al. 1999), a single model specification is used, with a small number of variables and neighborhood effects introduced via a weighted average of estimation errors. Details of generalized least squares and Kriging will be provided in the following chapter.

2.2.2 Market Segmentation

A perfectly functioning urban housing market is defined as having divisible quantities of housing services as well as elastic demands and supplies; under this idealized condition, the competitive behavior of households and housing suppliers will ensure that markets achieve equilibrium where the marginal prices of attributes other than land are constant over space and over bundle type (Schnare and Struyk 1976). In other words, the shadow price of each component of housing bundle will normally reflect its underlying market demand and supply.

Specifically, for structural attribute, its marginal price will equal the marginal cost of producing the attribute. With regard to neighborhood and locational attribute, the price represents the utility that household derives from such attribute; if the underlying household utility curve for a particular attribute can be drawn, then the price equals the parameter of this curve (Schnare and Struyk 1976). Unfortunately, housing markets are characterized by two features: the durability of stock and the joint nature of the services produced by the structural attributes of the dwelling unit and by the attributes of the neighborhood in which the unit is located (Maclennan 1976; Schnare and Struyk 1976).

For the durability of stock, once the supply of certain types of housing services is provided, it will be basically fixed for a relatively long period of time due to a variety of factors. First, the high monetary worth of houses restrained the immediate purchase behavior and encouraged the leasing or rental activity, therefore both suppliers and households need a fairly long period to make profit and realize the return on their investment. Additionally, the extremely costly adjustment processes add to the difficulty of altering housing services. Furthermore, suppliers face a large population of households which are geographically dispersed and are also socially, economically and culturally diverse, therefore suppliers lack efficient ways to communicate with buyers as to detect the market demand from pool of preferences. Moreover, even when a fairly representative preference is recognized, it would take years for a particular type of house catering to such preference to be constructed and offered in the market.

In regards to the joint nature of housing attributes, it is the initiative of spatial expansion model that parameters for structural attributes are characterized by spatial contextual variability, i.e. structural attributes produce different price differentials depending on the neighborhood in which the unit it located (Bowen et al. 2001;)

These two characteristics distinguish housing market from other services provided in urban area and also make it hard to function perfectly; as a consequence, the degree of substitution required to the equality of attribute price is unlikely to occur in urban housing markets (Schnare and Struyk 1976). As well, the demand for certain types of house or particular attributes is also likely to be inelastic considering characteristics and preferences of individuals or households are quite different. For instance, households with school age kids would insist on living in a neighborhood with a superior school, and for those households, their options haven been narrowed down to fewer houses located in jurisdiction of appropriate schools (Schnare and Struyk 1976). The inelastic supplies, combined with inelastic demands will segment the market into a number of independent sectors, in which, price of individual housing attributes will be more or less specific to each of those sectors (Schnare and Struyk 1976). In addition, the nature of budget or wealth constraints will also likely segment the general market since individual's age, sex, or occupation will influence access to finance, whilst other characteristics like race or ethnicity will further confound the segmentation (MacLennan 1976; Goodman and Thibodeau 1998)

The preceding discussion confirms the question raised in section 2.1.3 that homogeneity is not an attribute that can be naturally assumed for urban housing market without justification, rather, due to market segmentation, heterogeneity would be a more intrinsic component in this process. Consequently, traditional hedonic analysis of urban housing market, which assumes constant coefficient estimates and accordingly implies stable marginal prices of housing attributes cross the sectors, becomes inappropriate.

A solution to overcome the limitation of invariant coefficient estimates is to define various housing sub-markets, in other words, to stratify the dataset into homogeneous sectors, in which hedonic prices are estimated for each specific sector. By doing so, it allows implicit marginal prices of attributes to vary across sub-markets. Implementing the above procedure is a relatively simple affair if the boundaries of sub-markets or the stratification of independent sectors are known in advance. A popular approach to define independent sectors is to experiment with a number of stratification schemes and define sectors along both neighborhood and structural lines. For example, Schnare and Struyk (1976) used three stratifiers, which were thought to have relatively inelastic demand, including two neighborhood attributes-----average tract income and accessibility to employment center, and one structural attribute------the number of rooms. With these variables, the sample is divided into four neighborhood types, three structural types, therefore resulting in twelve sub-markets. Another simplified and also commonly used method is to view geographic area like census tract, municipalities or school district as potentially distinct sectors within urban housing market.

However, the soundness and appropriateness of these two approaches have generated enormous debates since from an econometric viewpoint, in order to define distinct housing sub-markets objectively, it requires substantive research on the manner that purchasers and sellers perceive the relevance of characteristics and the consequent attribute elasticity (Maclennan 1976). Without explicit exploration of consumer or supplier utility function and market elasticity, the two approaches described above or other criteria used in determining geographic submarkets like the use of central city/suburb, realtors' opinions are all somewhat arbitrary (Schnare and Struyk 1976; Can 1992).

In sum, market segmentation makes traditional hedonic model, an unstratified regression model inappropriate. In order to capture the dynamics operating at local housing market and provide more accurate marginal price estimate for attributes, one needs to stratify the data into different sectors along the segmentation line and fit separate model to each independent sector. However, the boundaries of sub-markets or the segmentation scheme requires excellent knowledge about local markets and is rarely available. In addition, a number of approaches invented so far to define sub-markets or independent sectors have been disputed as being arbitrary and lacking explicit estimate of market elasticity. If these approaches can not ensure the sub-markets delineated under their stratifiers are homogeneous zones, the results will be misleading. Also, the disaggregation into discrete areas may impose unrealistic discontinuities in their effects, for example, the effects of certain neighborhood characteristics go beyond the neighborhood boundaries (Can 1992). In this situation, no set of fixed neighborhood boundaries can accurately describe the urban market structure. Out of such concern, alternative modeling framework------moving window regression (MWR) and geographically weighted regression (GWR) are used in this study as a way to incorporate spatial heterogeneity into hedonic price structure. These two approaches adopt the concept of 'sliding neighborhood' (Can and Isaac 1997) ------moving window, in which no predefined neighborhood boundaries are required. As will be illustrated in chapter 3, under this framework, variations in marginal attribute price can be measured in a continuous rather than a discrete manner across space (Can 1992).

2.3 Implication of Spatial Autocorrelation and Spatial Heterogeneity

As discussed in section 2.2, due to two types of locational effects------neighborhood effects and adjacency effects, dependence in geographic space seems inherent for urban residential properties. If one fails to incorporate locational effects into hedonic modeling due to incomplete information or incorrectly measured attributes or other types of misspecification, spatially autocorrelated residuals after hedonic modeling will be observed.

In addition to the questionable assumption of independent residuals, another assumption in ordinary least squares approach regarding behavior of parameters also rarely holds for urban housing studies. As discussed earlier, urban housing markets are characterized by market segmentation, which implies a number of independent sectors or a number of distinct sub-markets across space. Under such condition, a spatially heterogeneous process is likely to exhibit in which the marginal attribute prices tend to 'drift' over space rather than being 'fixed' (Bowen et al. 2001). Spatial heterogeneity also implies that the mean and variance-covariance structure differ from location to location (Bowen et al. 2001). As a consequence, heteroskedasticity will be present as long as observations exhibit spatial heterogeneity, therefore, another important assumption of constant error variance is also violated (Can 1990).

In summary, in the presence of spatial autocorrelation and spatial heterogeneity, a number of assumptions fundamental to traditional hedonic modeling will be untenable including: observations are independent of each other; residuals are independently, identically and homoskedastically distributed; and the modeling process is homogeneous with stable coefficient estimates and uniform variance-covariance structure.

What is the consequence brought by violating these assumptions to traditional hedonic analysis results? In the presence of these violations, according to statistic theory, the resulting parameter estimates are unbiased but inefficient which means parameters are still estimated accurately in terms of their magnitudes (Basu and Thibodeau 1998; Dubin et al. 1999; Bowen et al. 2001), however, the confidence intervals yielded under untenable assumptions are incorrect. Besides that, the standard tests used to determine the statistic significance of housing characteristics will lead to potentially misleading conclusion, and also a similar impact is possible for testing the overall significance of model (Can 1997; Basu and Thibodeau 1998; Bowen et al. 2001). For instance, the most likely scenario in real estate application is the presence of positive autocorrelation. In this situation, the estimated standard error for parameters will be underestimated; thereby the resulting t-statistics will be biased upward (Basu and Thibodeau 1998; Dubin et al. 1999). In this way, an insignificant determinant, which is not contributing to hedonic housing prices, may be treated as an important factor and be subsequently reserved in model specification.

As outlined above, with respect to model validity, inefficient parameter estimates, misleading significance levels for both housing characteristics and for the whole hedonic price modeling, as well as their potential outcome of insufficient estimates of dependent variable are direct consequences brought by spatial effects. However, caution needs to be paid while incorporating spatial effects into modeling process since violations brought about by spatial autocorrelation and spatial heterogeneity are not necessarily misleading, or if they are, whether consequence will be of sufficient implication to arouse serious doubts as to the believability of the marginal price estimates of attributes or to their statistical significances (Bowen et al. 2001).

This critical consideration about the potential consequence of spatial effects on hedonic price model poses substantial challenges to researchers as to how to diagnose spatial autocorrelation and spatial heterogeneity when they are present, how to correct for them and include them into hedonic price structure as to provide more reliable and accurate estimation, and to what extent incorporation of spatial effects will improve the performance of traditional hedonic price function. Such questions will be explored in subsequent chapters.

Chapter 3 Data and Methods

The study investigates spatial autocorrelation and spatial heterogeneity in a number of spatial hedonic price functions using data from 33,494 transactions of single-family detached houses being sold between January 2001 and December 2003 in the City of Toronto. In order to reflect the true value of properties and avoid potential bias brought by distorted market forces like clearance sales, only open market sale records are included in the dataset.

The primary source of information comes from the Municipal Property Assessment Corporation (MPAC) in Ontario. Every year, MPAC prepares an assessment roll for every Ontario municipality which provides the assessed value of all the properties in a municipality or in the jurisdiction of a school board with taxing authority (www.mpac.ca). The computer file provided by MPAC contains each property's address as well as information of each property's structural characteristics. With ArcView, each transaction is assigned geographic coordinates and is geocoded into the study area in this way.

3.1 Study Area

All the transactions occurred in the City of Toronto, which is the capital and also the largest city of Ontario, Canada with population of 2,481,494 (2001 Census). It is located on the northwest shore of Lake Ontario at Latitude 43.39 N, Longitude 79. 23 W.

Caution need to be paid to the definition of Toronto. In 1998, the cities of Metropolitan Toronto (Toronto, York, East York, North York, Etobicoke, and Scarborough) were merged as City of Toronto, which is the area being studied. It is much smaller than the Toronto CMA (Census Metropolitan Area) and GTA (Greater Toronto Area).

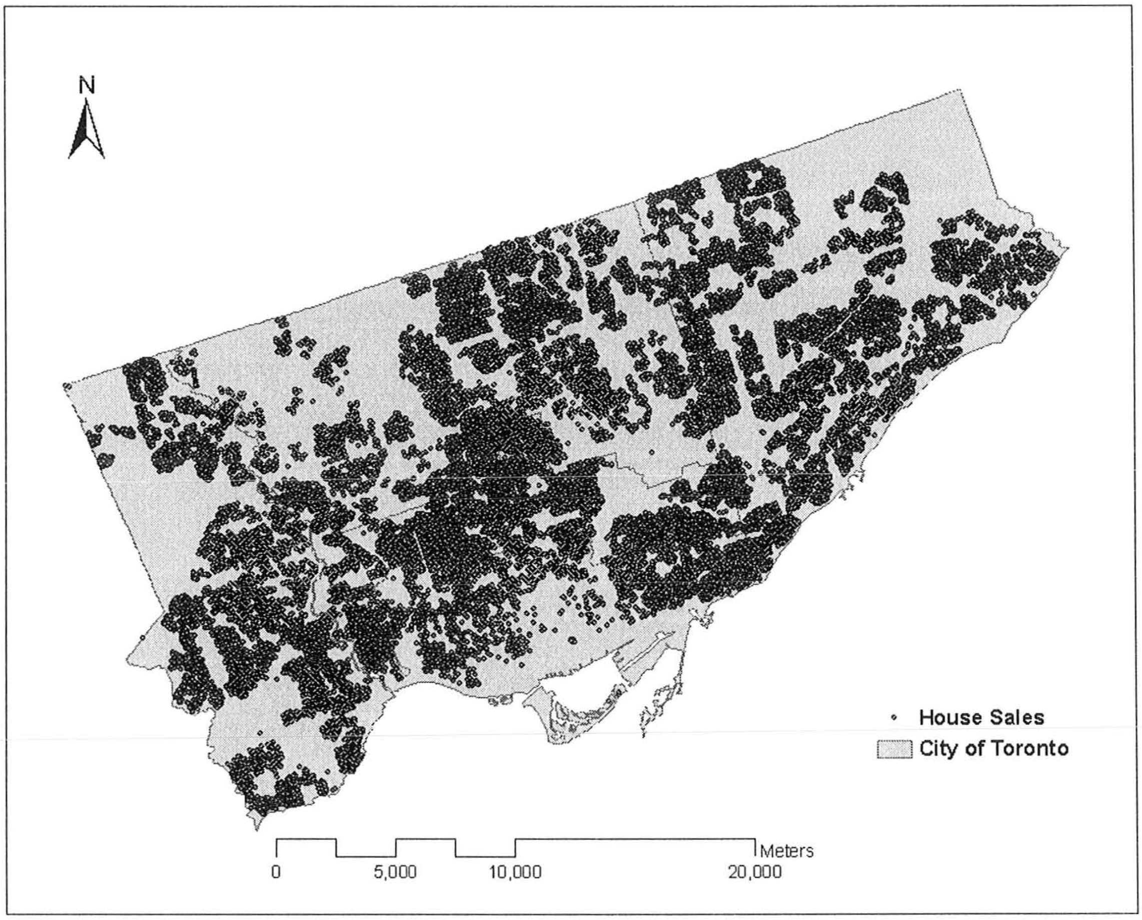


Figure 3.1: Study Area and Sale Locations

3.2 Description of Explanatory Variables

With regard to the explanatory variables used in traditional hedonic analysis, the selection of variables is informed by the study done by Farber (2004) which also looks at residential property value assessment in City of Toronto, combined with results obtained from correlation analysis, multicollinearity test and stepwise regression, 9 variables were selected, and some of them will be applied later in various spatial hedonic model specifications. As mentioned in section 2.1.1, these determinants can be classified into three categories.

The first category is the structural features of properties. Information on those units comes from MPAC's file. The original file consists of a large number of variables, however, many of them are incompletely recorded such as the number of bedrooms, the number of bathrooms, etc. After excluding them from dataset, variables available for modeling specification are more limited. In addition, severe multicollinearity exists among some of the remaining variables that results in a reduced number of variables qualified for the category. Valid variables are:

Area------ Effective site area in square feet

Front------Effective site frontage in feet

HouseAge------The age of dwelling in decades

Saledate------Since transactions occurred during a period of 36 months from January 2001 to December 2003, a variable ranging from 1 to 36 is defined to represent the temporal component, seasonality and inflation.

The second category includes the characteristics of the immediately surrounding natural and social environment. Information is obtained at the census tract level, coming from 2001 Census of Population. Information captured two important dimensions of neighborhood: income level and ethnic composition, which are represented by the following two variables:

MeanIncome------The mean household income at the census tract level; properties located in the same census tract are assigned the same value.

PctImm------The percentage of immigrants in the census tract; all properties in the same census tract share the same attribute value. [1]

The third category is the locational characteristics of properties. In correspondence with polycentric nature of modern cities, locational characteristics are measured by proximity to multiple social amenities rather than being measured simply by distance to the CBD. The variables under this category were derived using ArcView, based on data from DMTI's street network file. They are defined as:

DistShop------The straight line distance of a given property to the nearest community shopping center. [2]

DistSchool-------The straight line distance of a given property to the nearest education institution including school, university and college.

DistTransit------The straight line distance of a given property to the nearest transit including Light Rapid Transit (LRT), subway station or train station.

---

[1] Due to privacy concern, information for census tracts which contain few households is not released from Statistic Canada. Therefore, around 4173 observations have their mean household income and percentage of immigrants set to value of 0. This will influence the modeling results for traditional hedonic model, but doesn't have an impact on other spatial hedonic models since neighborhood attribute is not included in their modeling specification. These observations are still retained in the dataset, considering the main purpose of this study is placed on comparison of different spatial hedonic price equations rather than explaining the spatial phenomena by exploring the contribution of individual factors.

[2] Distance to regional shopping center and neighborhood shopping center is also available, however, community shopping center is more evenly distributed across the space and it is not too dense like neighborhood shopping center or too sparse like regional shopping center.

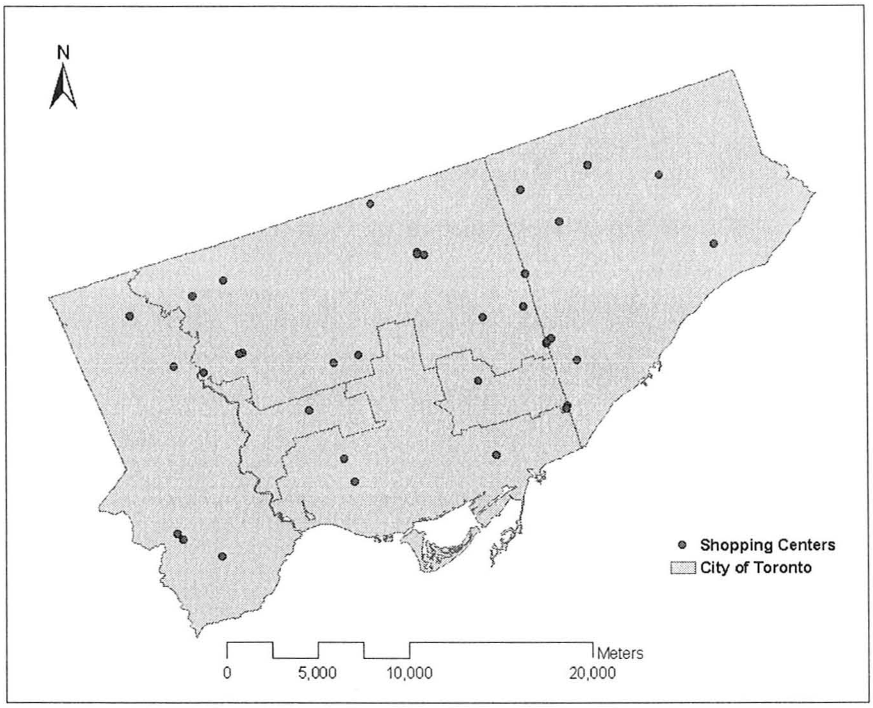


Figure 3.2: Locations of Community Shopping Centers

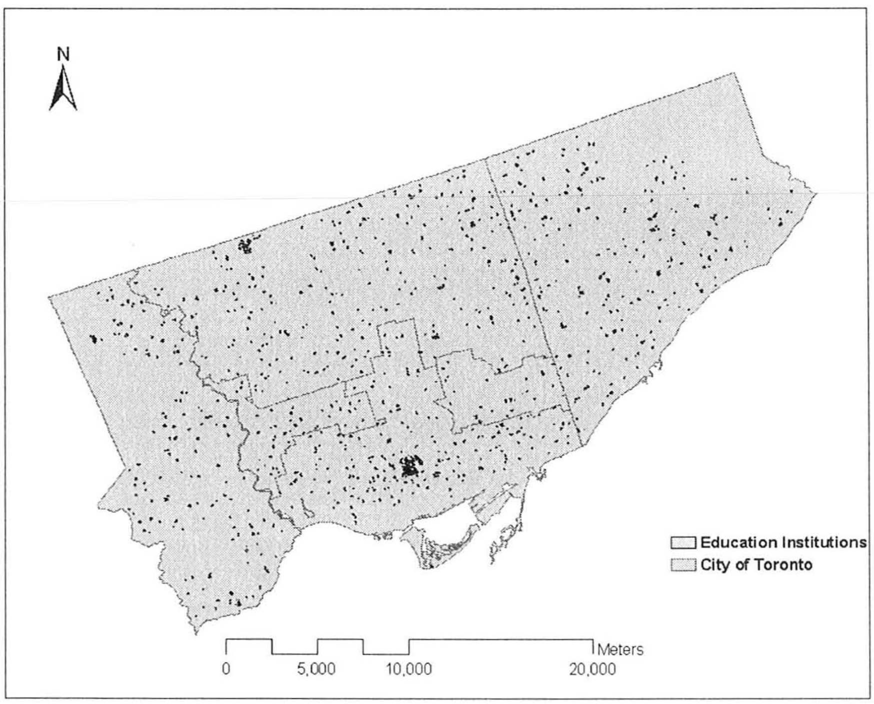


Figure 3.3: Locations of Education Institutions

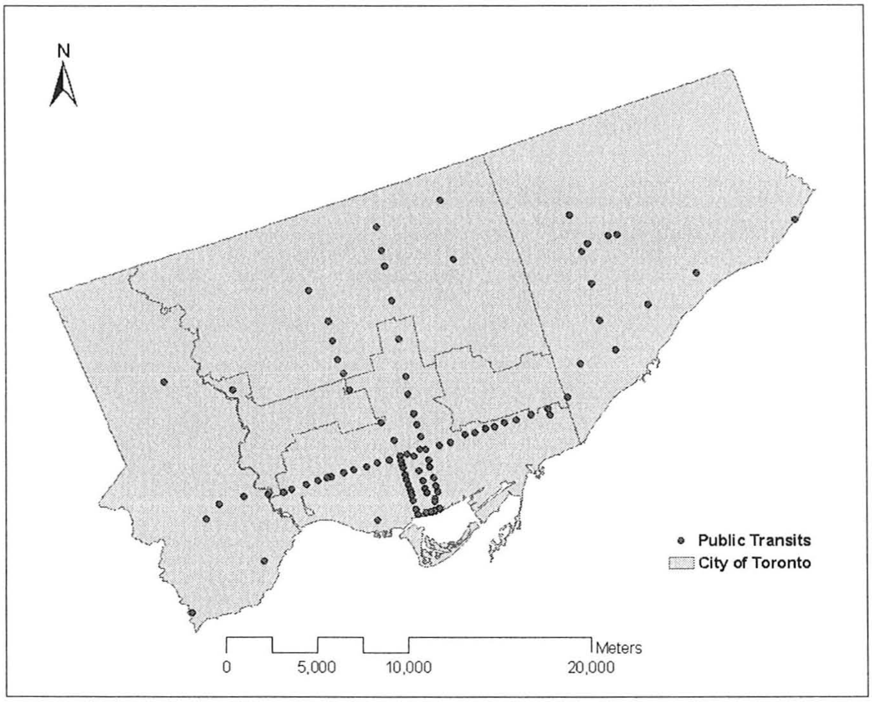


Figure 3.4: Locations of Public Transits (LRT, Train and Subway Station)

Table 3.1 provides summary statistics for sale prices and explanatory variables defined for traditional hedonic price function.

Table 3.1: Summary Descriptive Statistics of Variables[3]

	Mean	Standard Deviation	Range Minimum	Range Maximum
SalePrice ($1000)	384.926	290.793	25.500	7475.000
Structural attributes				
1. Area (square feet/100)	54.064	42.907	6.221	3346.650
2. Frontage (feet)	42.175	16.234	10.090	422.000
3. HouseAge (decade)	4.807	2.451	0	17.700
4. SaleDate	19.631	10.216	1	36

[3] All the information is collected from single family detached houses transacted in the City of Toronto from January 2001 to December 2003.

Neighborhood attributes[4]				
1. MeanIncome ($1000)	74.330	58.305	0	423.989
2. PctImm (%)	37.599	19.399	0	75.500
Locational attributes				
1. DistShop (km)	2.125	1.058	0.016	5.611
2. DistSchool (km)	0.359	0.184	0.023	3.156
3. DistTransit (km)	1.870	1.269	0.005	7.765

3.3 Diagnostic Tests of Spatial Autocorrelation

3.3.1 Moran's I Index

Accompanied with greater awareness of potentially severe consequence brought by spatial autocorrelation, a number of formal diagnostic approaches have been developed by geographers and other scholars in an attempt to detect and capture it. In this section, the most popular and widely used technique, Moran's I index is introduced and will be subsequently applied.

In essence, Moran's I index shares many similarities with Pearson's correlation coefficient: their numerators are the covariance, while their denominators are the sample variance (Oliveau and Guilmoto 2005). According to its definition, Pearson's correlation coefficient measures the degree of correlation between two variables, while slightly different but along the same line, Moran's I index is a measurement of correlation of a single variable between pairs of neighboring observations. Therefore the resulting measurement from this index is regarded as indication of spatial nature of the

[4] The mean value for MeanIncome and PctImm is underestimated due to omitting records. In the same way, their standard deviations are overestimated.

phenomenon being studied: observations may be spatially dependent on their neighbors or they may display a random pattern over space. The precise definition of Moran's I is given as:

$$I = \frac{n\sum_{i-1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{(\sum_{i=1}^{n}(y_i - \bar{y})^2)(\sum\sum_{i \neq j} w_{ij})} \tag{3.1}$$

where y is the value of attribute being studied; n is the number of observations in the dataset; W is a spatial proximity matrix with element w_{ij} between all possible pair of observations (i, j). Unlike from Pearson's correlation coefficient, the values of Moran's I are not constrained to lie in the (–1, 1) range. High positive values imply strong positive spatial autocorrelation, while negative values reveal negative spatial autocorrelation. A value of 0 stands for a random pattern.

3.3.2 Empirical Variogram

A different way of measuring spatial dependence is covariogram or variogram estimation. It is only applicable to spatially continuous data and has gained wide popularity in geosciences. In statistics, the covariance is a measurement of the extent to which two variables vary together. For instance, if one variable increases as the other one does, then there is positive covariance and it is estimated as:

$$Cov(x, y) = E\{[x - E(x)][y - E(y)]\} \tag{3.2}$$

If the covariance is divided by the product of two variables' standard deviations, then Pearson's correlation coefficient is obtained. In a spatial context, especially for spatially continuous phenomena, the same ideas of covariance and correlation are applied and their equivalents are covariogram $C(s_i, s_j)$ and correlogram $\rho(s_i, s_j)$. Instead of computing the covariance or correlation between two variables, they are more concerned with the way in which the deviations of observations from their mean value at different locations on the map co-vary or are correlated (Bailey and Gatrell 1995).

More formally, for a spatial stochastic process $\{Y(s), s \in R\}$ where $E(Y(s))$ is denoted as $\mu(s)$ and $VAR(Y(s))$ is denoted as $\sigma^2(s)$, the covariance of this process at any two particular points s_i and s_j and the corresponding correlation is defined as:

$$C(s_i, s_j) = E((Y(s_i) - \mu(s_i))(Y(s_j) - \mu(s_j))) \tag{3.3}$$

$$\rho(s_i, s_j) = \frac{C(s_i, s_j)}{\sigma(s_i)\sigma(s_j)} \tag{3.4}$$

Since covariogram and correlogram share similar attributes as covariance and correlation, and meanwhile Moran' I index is the geographic manifestation of Pearson's correlation coefficient, then it is fairly easy to conclude that covarigram, correlogram and Moran's I index are comparable measurements in terms of detecting and measuring autocorrelation of spatial process. However, the specific formulations of Moran's Index, covariogram and correlogram are dissimilar and subsequently some distinctions are drawn. For Moran's I index, predefined neighbors are requisite to the fulfillment of estimating spatial autocorrelation (Bowen et al. 2001), but the question of how to define optimal local neighborhood remains unresolved after extensive exploratory work. More problematic is the fact that definition of neighborhood will affect the measured value of spatial autocorrelation. Out of the above concern, covariogram and correlogram appear to be better alternatives to Moran's I index in terms of measuring spatial autocorrelation since they do not demand knowledge about local neighborhood; rather, they convey an initial idea about the approximate range of it.

On the other hand, in order to compute the covariogram, the process is assumed to be stationary, that is, all sample points are taken randomly and independently from one simple probability distribution. When such assumption is tenable, the mean and variance of sample points are independent of location and remain constant through space. In addition, it also assumes $C(s_i, s_j) = C(s_i - s_j) = C(h)$, that is $C(s_i, s_j)$ depends only on the vector difference, h, between s_i and s_j (direction and distance of separation) and not their absolute locations. Moreover, in order to obtain a workable description of the covariance structure, isotropy is often assumed which implies that the dependence is purely a function of distance between s_i and s_j, not the direction.

The advantage of assuming stationarity lies in simplifying the process to ensure that a single covariogram is sufficient to depict the underlying variance-covariance structure, that is, covariance between any two sites can be modeled by a smooth function of distance separating them (Bailey and Gatrell 1995). A weaker assumption of stationarity is intrinsic stationarity which is defined as a constant mean and a constant variance in the differences between values at location separated by a given distance and direction. Such assumption can be described as:

$$\begin{aligned} &E(Y(s+h)-Y(s))=0 \\ &VAR(Y(s+h)-Y(s))=2r(h) \end{aligned} \tag{3.5}$$

For an intrinsic stationary process, the covariogram, correlogram and variogram have the following relationship: $\rho(h)=\dfrac{C(h)}{\sigma^2}$; $r(h)=\sigma^2-C(h)$ (3.6)

These three measurements can be converted into each other and also deliver similar information, thus it would be redundant to compute them all. More importantly, in general, the sample variogram provides more robust estimates of spatial dependence than the sample covarigram in the presence of minor departures from stationarity since it involves a weaker assumption of intrinsic stationarity (Bailey and Gatrell 1995). In this sense, variogram is preferred here as a tool to investigate the spatial dependence structure of the process of interest.

Under the assumption of isotropy, an omnidirectional empirical variogram is obtained using the following equation:

$$2r(h) = \frac{1}{n(h)} \sum_{|si-sj|=h} (y_i - y_j)^2 \tag{3.7}$$

where (y_i, y_j) is pair of observed data points with a vector separation of h and $n(h)$ is the number of pairs. By plotting $r(h)$ to h, a typical figure of empirical variogram will be like:

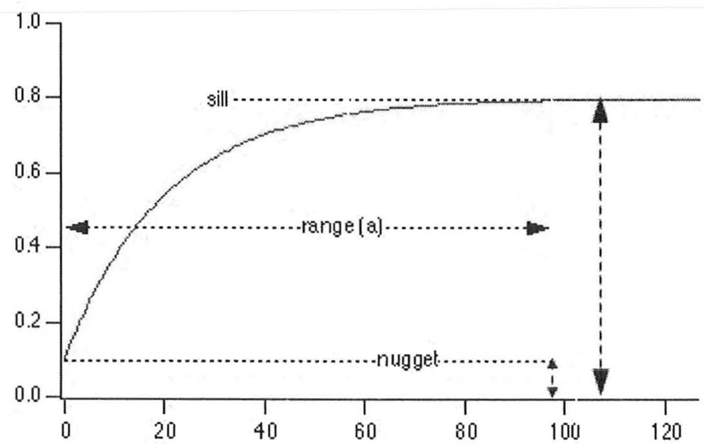


Figure 3.5: Typical Image of Empirical Variogram

source:http://www.wavemetrics.com/products/igorpro/imageprocessing/imagetransforms/imageinterpolationpix/variogram.jpg

Three important parameters can be inferred from this figure:

Sill ------ upper bound of the variogram. (variance of the sample)

Range ------ the distance at which sill occurs. It indicates the distance where the spatial phenomena are no longer autocorrelated.

Nugget ------ the value of r (h) when h=0 which is caused by measurement error, micro-scale variation or spatial discontinuities relevant in the case of house studies.

3.4 Diagnostic Test of Spatial Heterogeneity

As illustrated before, spatial homogeneity or stationarity is not an attribute that can be assumed or taken for granted; especially ascribed to market segmentation, it is more frequent to observe spatial heterogeneity in a given process. Therefore, even the variogram has the merit of providing preliminary information on local neighborhood rather than requesting it as a prior, its strong assumptions on stationarity and isotropy somewhat impair its credibility.

An intuitive way to diagnose spatial heterogeneity would be to apply empirical variogram estimation in each sub-region and examine whether the corresponding local variograms substantially differentiate from each other or from the whole-area variogram. This approach was utilized by Haas in 1990 to detect spatial nonstationarity in the deposition of heightened levels of sulfuric and nitric acid in the United States and was extensively adopted later by other scholars. In his study, Haas (1990a) selected six estimation locations in conterminous United States arbitrarily, and then six circular sub-regions centered at each estimation site were defined in which local variogram for each sub-region was calculated and compared. This exploratory work showed obvious dissimilarity among variance-covariance structure of different sub-areas.

Basu and Thibodeau (1998) computed variograms of transaction prices for 9 zones in the City of Dallas, Texas whose boundaries are defined by major highways. Also Walter et al. (2001) conducted similar experiment on topsoil salinity in the Chelif Valley. The study area was arbitrarily divided into two subregions, and variograms for each sub-area and whole area were estimated and compared. Similar conclusions were drawn for both cases.

To conduct a relatively inclusive comparison and rule out the possibility of incomplete specification of sub-areas, in this research they are defined along two schemes: one is to delineate sub-region to ensure all of them have the equal size, i.e., a circular district with radius of 2km; the other is to allow their sizes to vary to accommodate around 1000 observations[5]. The relative locations of sub-areas cross space as well as their sizes are shown in Figure 3.6 and Figure 3.7.

Results for sub-market variogram modeling as well as diagnostic test of spatial autocorrelation using both Moran's I index and empirical variogram will be provided in Chapter 4.

---

[5] The selection of 2 km or 1000 observations as the sub-area delineation scheme is to ensure the number of sub-areas defined under two schemes is manageable and also approximates to each other.

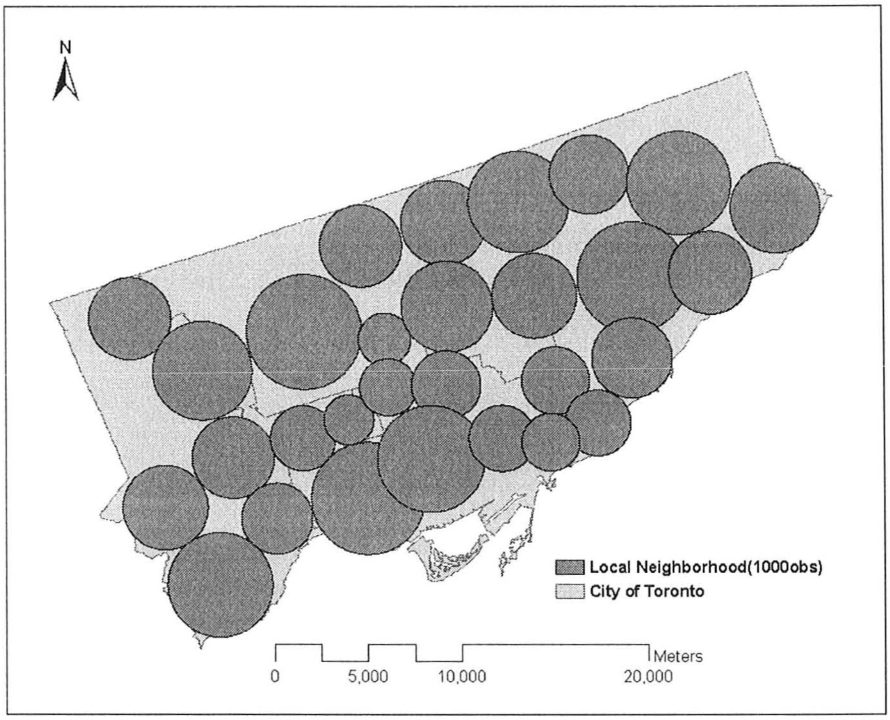


Figure 3.6: Local Neighborhoods with 1000 Observations

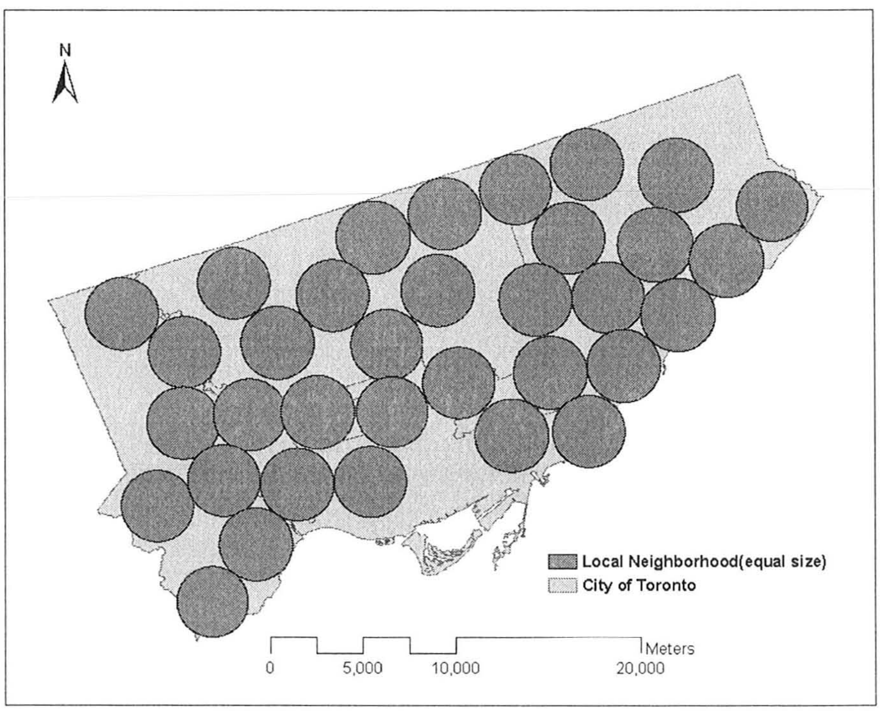


Figure 3.7: Local Neighborhoods with Equal Size

3.5 Spatial Hedonic Modeling

"Methodological developments in spatial statistics and econometrics have shown that the straightforward use of traditional hedonic models may not be adequate for the analysis and modeling of geographically referenced data due to spatial effects, namely spatial dependence and spatial heterogeneity."(Can 1990)

In Chapter 2, the basic components of traditional hedonic model were reviewed including its origin, underlying economic theory as well as model specification. It also explored several spatial issues, basically spatial autocorrelation and spatial heterogeneity, inherent to residential property values. It is noted here, there is an emerging consensus that the traditional hedonic price function, the most commonly used approach to study urban housing market, is incompatible with the spatial nature of residual dataset. Being inherently non-spatial, it fails to take spatial variation of cross-sectional dataset into account by making a series of unrealistic assumptions about parameters and residuals. In this section, a number of alternative spatial techniques and models, which aim to overcome the above limitations and incorporate spatial effects into housing price estimation will be introduced and elaborated.

3.5.1 Moving Window Regression and Geographically Weighted Regression

As discussed earlier, ascribed to market segmentation, the price of individual housing attributes will not be necessarily constant over space. It is likely that prices will exhibit variations cross space according to the environmental and locational characteristics of the specific neighborhood in which the property is located (Can 1992; Bowen et al. 2001). Circumscribed by such condition, a single set of parameters derived from traditional hedonic model may not be suitable for all observations, which are widely dispersed in the study area and have great chances of locating in distinctive submarkets.

An intuitively appealing idea to solve the fixed relationship between variables implied by traditional hedonic model is to estimate coefficients of attributes for each site by only using its neighboring observations for model calibration. More specifically, the idea is to impose a window centered at the point of interest, and only observations situated within the window will be regarded as "neighbors" and be consequently included into regression for estimating coefficient of the given point. Then the window will move to reach the next point, where a new set of parameters is established using only neighboring observations. Such procedure will iterate until the moving window visits all the points of interest in the study area. The above method is called moving window regression (MWR). Essentially, it composes a series of locally linear regressions and generates unique parameter estimates for every regression point. Thus MWR allows the model to alter over space to reflect the varying structure within the data.

Another technique, termed geographically weighted regression (GWR) was proposed by Brunsdon, Fotheringham and Charlton in 1996. The improvement made by GWR is: after determining the local neighbors for regression point i, rather than treating them equally, GWR applies different weight to each observation in accordance with its proximity to regression point i. The spatial weighting function employed in this study is the adaptive bi-square function which is defined as:

$$w_{ij} = \left[1 - d_{ij}^2 / d^2\right]^2 \text{ if } d_{ij} < d;$$

$$w_{ij} = 0 \text{ otherwise}; \qquad (3.8)$$

where d is the distance between i to its n th nearest neighbors. The value of n will be determined by cross-validation procedure discussed in the following chapter to achieve the best model performance. This adaptive bi-square function ensures that every moving window contains same number of observations, and accordingly every local regression works with equal size of sample.

Equation 3.8 illustrates how observations gain different emphasis and then contribute differently to parameter estimates: firstly, points outside the moving window, i.e. with distance to regression point i larger than d, will be excluded from model calibration; the weightings of points inside the window are inversely proportional to their distance from point i; the weight starts from unity from the point itself, then it falls down to zero when distance equals d (Brunston et al. 1996). According to the above illustration, Waldo Tobler's first law of geography is embedded in the formulation of GWR, that is, "Everything is related to everything else, but near things are more related to each other".

In fact, moving window regression can be regarded as a simplified GWR using a less complicated weighting scheme defined as:

$$w_{ij} = 1 \text{ if } d_{ij} < d;$$

$$w_{ij} = 0 \text{ otherwise}; \qquad (3.9)$$

where d is again the distance between i to its n th nearest neighbors.

In ordinary least squares, the coefficient estimates are calculated by minimizing the sum of the squared differences between predicted and actual value of dependent variables. In MWR and GWR, to cater to their special model design, weighed least squares is used for coefficient estimation, where a weighting factor w_{ij} is applied to each squared difference before minimizing. If W is a diagonal matrix of weights w_{ij}, then the coefficients can be estimated as:

$$\beta = (X'WX)^{-1} X'WY \tag{3.10}$$

Based on the definitions described above, moving window regression (MWR) and geographically weighted regression (GWR) are in essence similar approaches: they both endeavor to model spatial nonstationarity by performing regression for each site of interest at local level through the use of moving window.

3.5.2 Generalized Least Squares

All the hedonic price functions described so far including traditional hedonic price function, moving window regression (MWR), and geographically weighted regression (GWR) employ ordinary least squares or its modification------weighted least squares for coefficient estimates and accordingly predict housing values by βX. However, ordinary least squares approach yields best estimates of parameters and best linear unbiased estimation only when residuals are independent and residual variances are constant.

Unfortunately, the assumption of uncorrelated residuals is not always tenable. Empirical studies (Can 1990, 1992; Pace and Gilley 1997; Basu and Thibodeau 1998;

Bowen et al. 2001; Farber 2004; Case et al. 2004) have found that spatial autocorrelation exists in residuals after traditional hedonic modeling or even after spatial hedonic modeling like spatial autoregressive model or autocorrelated error model are performed. A plausible explanation is that in order to capture all scope of spatial dependency, it should specify all the positive and negative spatial externalities affecting housing prices into modeling consideration and also capture spillover effects by a proper design of spatial dependency structure. However, the first condition is rarely satisfied: as illustrated before, there is no consensus in the literature regarding which variables best proxy these externalities and even if agreement could be achieved, it is likely that this information is difficult to measure or simply immeasurable. For spillover effects, very few models explicitly devise a spatial dependency structure to conceptualize them into model specification or if they do like spatial autoregressive model or autocorrelated error model, the dependency structure is generally pre-determined without resource from investigating a real process.

GWR and MWR, on the other hand, are expected to eliminate part of spatial heterogeneity and the resulting spatial heteroskedasticity ascribed to price structure instability; the part originating from missing variables or other forms of misspecification may still, however, be present. This would, as already discussed, invalidate the assumptions about error independence and constant error variance.

Thus, even after substantial efforts of incorporating spatial autocorrelation and spatial heterogeneity into traditional hedonic modeling specification, it is still possible to observe interdependent residuals after OLS estimation. In this respect, generalized least squares (GLS), an alternative regression technique, is introduced which generates more reliable estimates for parameters when error terms from OLS display non-random patterns such as autocorrelation. The model can be written using vector notations as:

$$Y(s) = x^T(s)\beta + U(s) \tag{3.11}$$

The above equation shares similar definitions of variables as in equation 2.3: $Y(s)$ is the house price value in process at the point s; the $(k*1)$ vector $x(s)$ consists of k attributes of this given house; β is a $(k*1)$ vector of parameter to be estimated. The difference is that although $U(s)$ are still zero mean errors as in OLS regression, they are not necessarily independent of each other; rather, they have a covariance structure $C()$, a $n*n$ matrix of covariance $C(s_i, s_j)$ between $U(s_i)$ and $U(s_j)$, for each possible pair of observations (i, j). The diagonal elements of this matrix are $VAR\ (U(s_j))$, which are assumed to be constant as in OLS. However, for off diagonal elements, unlike OLS which assigns 0 to all positions, GLS does not assume error independence but allows finite positive values. In this situation, the previous ordinary least squares estimates for β and the corresponding standard errors are not appropriate and their generalized least square equivalent are:

$$\begin{aligned} \hat{\beta} &= (X^T C^{-1} X)^{-1} X^T C^{-1} Y \\ VAR(\hat{\beta}) &= (X^T C^{-1} X)^{-1} \end{aligned} \tag{3.12}$$

Based on the above equations, the most challenging aspect of GLS is the estimation of $C()$, which is seldom known a priori. In addition, the error terms should be known first in order to calculate the covariance matrix, however these error terms can not be obtained unless there is already an assessment about the parameters which requires the estimation of covariance matrix. Therefore, it seems that situation has returned to the original point and will start to circulate again unless this chain can be broken. Two methods have been created that solve the above problem: (1) estimated generalized least squares (EGLS) and (2) maximum likelihood (ML).

EGLS (estimated generalized least squares) divides the problem of incorporating spatial information residing in residuals into a number of steps as described below:

(1) Fit the model into observed data using ordinary least squares regression; subtract predicted values from actual prices to get residuals.

(2) Use residuals obtained from step 1 to estimate a semi-variogram.

As introduced in section 3.3.2, a very important tool for diagnosing spatial autocorrelation is the empirical variogram:

$$2r(h) = \frac{1}{n(h)} \sum_{|si-sj|=h} (y_i - y_j)^2 \tag{3.7}$$

This empirical variogram is fitted by a parametric autocorrelation function which models the variance of the difference between values of regionalized variable as function of a separation distance. Since the covariance can be estimated as $r(h)+\sigma^2$, a smooth, continuous description of the covariance structure can be obtained in which covariance between any two houses is simply the function of intervening distance.

There are several options for parametric models, among which the most widely used ones are spherical, exponential and Gaussian model. Durbin has recommended exponential model for GLS estimation of residential property values, however, after scrutiny of empirical variogram estimation carried out at various local levels and various locations (details will be provided in section 4.3), the spherical model is selected here to fit the empirical variogram which is defined as:

$$r(h)=\begin{cases} a+(\sigma^2-a)\left(\dfrac{3h_{ij}}{2r}-\dfrac{h_{ij}^3}{2r^3}\right) & 0<h\le r \\ 0 & h=0 \\ \sigma^2 & otherwise \end{cases} \qquad (3.13)$$

where σ^2, a and r are parameters sill, nugget and range that help defines a variogram. The value of these parameters will be determined through the spherical model.

(3) The variogram model from the above step gives rise to an equivalent covariogram model $\hat{C}()$. This covariogram model makes it feasible to construct a covariance matrix $\hat{C}$ between sample sites, with elements $\hat{C}(s_i, s_j)$.

(4) Then refit the original model using generalized least squares with the estimated covariance matrix, $\hat{C}$. This provides more reliable parameter estimates and the corresponding standard errors.

(5) If necessary, iterate the whole process until stable estimates of β and $C()$ are obtained.

Another way to generate a covariance matrix is maximum likelihood estimation. In this approach, the errors are also spatially autocorrelated, $E\{\varepsilon\varepsilon'\}=\sigma^2 K = C$ where K (the correlation matrix of the error term) have nonzero off-diagonal elements and value of 1 for diagonal elements. Like modeling empirical variogram, there are several candidates proposed for modeling this spatial correlation matrix K . In accordance with

EGLS, only spherical model is applied here. The estimation for β and its standard errors still remain the same for maximum likelihood function.

$$\begin{aligned} \hat{\beta} &= (X^T C^{-1} X)^{-1} X^T C^{-1} Y \\ VAR(\hat{\beta}) &= (X^T C^{-1} X)^{-1} \end{aligned} \tag{3.12}$$

The difference lies in that maximum likelihood approach estimates covariance matrix C and parameters β simultaneously, that is combining the separate steps performed in EGLS by choosing the values of unknown parameters that maximize the following log likelihood function:

$$\tfrac{1}{2}\log\left|\hat{C}^{-1}\right| - \tfrac{n}{2}\log(Y - X\hat{\beta}(\hat{C}))^T \hat{C}^{-1}(Y - X\hat{\beta}(\hat{C})) \tag{3.14}$$

A more detailed description of this maximum likelihood approach can be found in Lark's (2000) and Dubin's (1998, 1999) papers.

EGLS is a more traditional approach that has already gained wide popularity especially in geological studies for interpolation purposes. However, accompanied with the conspicuous enhancement of computational power in the last decade, maximum likelihood approach seems more appealing by estimating all unknown elements simultaneously.

3.5.3 Universal Kriging

In OLS, it is assumed spatially independent residuals where error terms are randomly distributed in space. However, if residuals exhibit any systematic pattern, like spatial clusters of residuals of the same sign, it implies OLS doesn't incorporate all those factors affecting house prices into modeling (Dubin 1998; Dubin et al. 1999).

In order to explicitly capitalize information contained in residuals, GLS was advanced which as illustrated before uses spatial correlation of residuals to obtain more efficient estimates of parameters. In this section, a technique referred to as Kriging, is introduced, which is an equivalent to GLS with a primary interest in prediction of values for spatially continuous variable such as sale prices in this dataset. Its mathematical expression is as follows:

$$Y(s) = x^T(s)\beta + U(s)$$
$$\hat{U}(s) = \sum_{i=1}^{n} \lambda_i(s)U(s) \tag{3.15}$$

where $U(s)$ is still a zero mean process with covariance function $C()$ as defined in equation 3.11. Generalized least squares approach is again employed to estimate β. However, the values of $U(s)$ are not entirely unpredictable; an estimate $\hat{U}(s)$ is computed as a weighted linear combination of nearly observed residuals to approximate $U(s)$. According to the above definition, Kriging is an optimal spatial interpolation technique, which aims to improve the accuracy of prediction in two ways: firstly, like GLS, it considers spatial correlation in residuals and specifies it into parameter estimates; second, it adjusts the prediction values by adding a local component (or predicted error term denoted as $\hat{U}$) to the contextual drift, which is obtained from nearby properties as a weighted average of observed residuals $U(s)$ (Bailey and Gatrell 1995; Dubin et al., 1999)

After following the same step as described in GLS estimation, the parameter can be computed as $(X^TC^{-1}X)^{-1}X^TC^{-1}Y$ by either employing EGLS or ML approach. Now the challenge encountered here is how to determine the weights assigned to nearby properties as a function of proximity and spatial dependence.

As seen from equation 3.15, $\hat{U}(s)$ is a sum of random variables, therefore itself is another random variable. It is intuitively sensible to design weights λ_i, to ensure that $\hat{U}(s)$ is as close as possible to the true residual $U(s)$. The degree of closeness between them is measured by the expected mean square error between values of $\hat{U}(s)$ and values of $U(s)$. This expected mean square error is calculated as:

$$E\left(\left(\hat{U}(s)-U(s)\right)^2\right)=E\left(\hat{U}^2(s)\right)+E\left(U^2(s)\right)-2E\left(U(s)\hat{U}(s)\right)$$
$$=\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i(s)\lambda_j(s)C(s_i,s_j)+\sigma^2-2\sum_{i=1}^{n}\lambda_i(s)C(s_i,s_j) \tag{3.16}$$
$$=\lambda^T(s)C\lambda(s)+\sigma^2-2\lambda^T(s)c(s)$$

(Bailey and Gatrell 1995)

where C is defined as in GLS, a $(n*n)$ matrix of covariance, $C(s_i,s_j)$, between all possible pairs of sample sites and $c(s)$ is an $(n*1)$ column vector of covariance, $C(s,s_i)$, between the prediction site and each sample site. Minimizing this mean square errors provides the solution as follows : $\lambda = c^T(s)C^{-1}$,

therefore $\hat{U}(s)=\sum_{i=1}^{n}\lambda_i(s)U(s)=c^T(s)C^{-1}U$ [6] (3.17)

Since the weights assigned to observed residuals are determined under the optimality condition that weights should have such values that estimation variance is minimized, Kriging estimate is thus regarded as the best linear unbiased estimate (Lam,

---

[6] The above deductions are from 'Interactive Spatial Data Analysis' wrote by Bailey and Gatrell, page 184-185.

1983). Kriging was invented by South African mining geologist D.G. Krige, whom its name is derived from, and later it was refined by French geostatisticians. Kriging has since become a major tool in the field of geostatistics, especially in the field of mining, for interpolation purpose in the last decades (Lam 1983). It was only in 1992, when it was first introduced into urban housing studies by Dubin and still remains a rarity in this field.

3.5.4 Moving Window Kriging

Recall the definition of variogram in section 3.3.2, a fundamental assumption is spatial stationarity of attributes, that is the mean and variance of each distribution is the same at all locations and the correlation between observations is simply a function of distance separating them. Since the variogram is built into Kriging as an essential component, the assumption of spatial stationarity is also applied to residuals from OLS estimation. Only under such condition, a single empirical variogram established using all observations can be sufficient to describe the underlying covariance structure of spatial process (Bailey and Gatrell 1995). However, in reality, many if not most spatial phenomenon would display both mean nonstationarity and location-dependent variance. In addition, it is also implausible that correlation of any pair of observations can be expressed by a universal parametric function.

As will be seen in section 4.3, sale prices will be examined for spatial stationarity by conducting local variogram estimate under two schemes of submarket delineation. Variography analysis demonstrates parameters from variogram modeling display considerable variations cross space which confirmed the existence of differing spatial structures among sub-regions (details will be provided in section 4.3). It is evident now that a more faithful depiction of urban housing market as well as corresponding hedonic price specification should echo a spatial heterogeneous process by accommodating location-dependent residual variance and inconstant covariance structure.

A modification to Kriging, namely Moving Window Kriging, was advanced by Haas in 1990, aiming at adapting to non-uniform spatial structure and improving prediction accuracy. Such goal is accomplished by estimating and modeling a new variogram for each prediction site, using only observations within a neighborhood immediately surrounding it. As discussed in GWR and MWR, this neighborhood is called a "window", which is either defined as a circle centered at the prediction point with a given radius or is determined by containing a certain number of nearest observations. Therefore, differing from conventional Kriging, where a unique variogram is inferred and modeled for all points providing a universal spatial covariance structure, Moving Window Kriging allows for the establishment of local spatial covariance structure for each site where estimate is desired by estimating and modeling local variogram with observations inside the window. Credited to a customized, locally derived variogram calculated for each prediction point, the variance-covariance structure now varies from site to site, which is believed to be a more faithful portrayal of the observed spatial phenomena (Hass 1989, 1990).

The rationale for Moving Window Kriging is derived from Journel and Huijbregts' definition and discussion of quasi-stationarity, that is a nonstationary process defined over large region will exhibit smaller absolute trend and a more homogeneous covariance structure over subregions. As the process is quasi-stationary within subregions, Kriging performed at each moving window is robust to nonstationary process exhibited over whole study area.

3.6 Summary

As discussed in Chapter 2, due to two types of locational effects, namely neighborhood effects and adjacency effects, it is almost inevitable to observe the presence of spatial autocorrelation in the process. To avoid potential misleading results regarding standard inferential test and property value estimation, there is need to remove spatial autocorrelation from modeling process and providing more reliable assessment for parameter estimates and price predictions.

Besides spatial interdependence, urban housing markets are characterized by another important feature------ market segmentation. The existence of spatial heterogeneity would invalidate the assumption of homoscedasticity held by traditional hedonic model and also poses problems to 'global' models which fail to alter over space to reflect varying relationship between variables or non-uniform variance-covariance structure among observations.

Accompanied with the theoretical speculation of spatial effects, formal diagnostic tests upon spatial autocorrelation and spatial heterogeneity were introduced in this chapter. In addition, on the purpose of incorporating spatial autocorrelation or modeling spatial heterogeneity, a number of spatial techniques or approaches have been discussed. These approaches are designed differently in correspondence with various understandings or emphasis of the spatial process. MWR and GWR attempt to model a spatially heterogeneous process by carrying out regression for each prediction site using neighboring observations. In addition, the two approaches also eliminate spatial autocorrelation partially by calibrating model within moving window, which is believed to be a more homogeneous area. Compared to MWR and GWR, Kriging is focused on exploration of spatial dependency structure of residuals. Through the construction of variance-covariance matrix, Kriging improves prediction accuracy by adjusting parameter estimate and adding predicted error term. With respect to MWK, it accounts for both spatial autocorrelation and spatial heterogeneity. The former part of MWK is the same as MWR, that is, performing regression within each moving window surrounding the prediction site. The latter part can be regarded as a localized kriging, i.e. estimating and establishing a unique spatial dependency structure, therefore providing adjusted parameter estimate and predicted error term for each prediction site. By doing that, MWK models a spatially varying process with flexible relationship between some sets of variables and also exploits spatial information remained in residuals via the prediction of error terms.

In the next chapter, these advanced spatial hedonic models as well as traditional hedonic model will be applied. Their modeling performance in terms of out-of-sample prediction accuracy will be evaluated and discussed.

Chapter 4 Results and Discussion

4.1 Traditional Hedonic Regression Results

Parameter estimates and other statistics for traditional hedonic model using all the variables specified in section 3.2 are presented in Table 4.1.

Table 4.1: Basic Hedonic Model Regression Results[7]

R-squared:	**0.575**	**Degrees of freedom:**	**33483**
Residual SS:	**1.2036E+09**	**Residual Standard Errors:**	**189.6**
F-statistic:	**4530**	**Probability of F:**	**0.000**

Variable	Estimate	Std Error	t-value	Prob> |t|
Constant	308.5317	6.4694	47.6908	0.0000
Area	1.3689	0.0367	37.2803	0.0000
Front	3.2388	0.1064	30.4293	0.0000
Saledate	2.7904	0.1016	27.4754	0.0000
HouseAge	- 96.0981	1.4451	- 66.5014	0.0000
Age^2	8.7340	0.1414	61.7663	0.0000
DistTransit	- 20.4508	0.9147	- 22.3585	0.0000
DistSchool	- 15.8665	5.8146	- 2.7287	0.0064
DistShop	- 9.7137	1.0356	- 9.3797	0.0000
MeanIncome	2.6846	0.0192	139.5746	0.0000
PctImm	- 3.0920	0.0542	- 57.0449	0.0000

[7] Semi-log function using same set of variables is performed which generates similar results in term of R square.

Table 4.2: Explanatory Power of Each Variable and Each Category

Variable	Estimate	Prob> |t|	R-square[8]	R-square[9]
Area	2.6815	0.0000	0.1566	0.2577
Front	6.2341	0.0000	0.1211	
Saledate	2.3116	0.0000	0.006595	
HouseAge	- 80.7099	0.0000	0.04272[10]	
Age^2	7.1603	0.0000		
DistTransit	- 34.6114	0.0000	0.02285	0.04917
DistSchool	168.7017	0.0000	0.01144	
DistShop	41.0558	0.0000	0.02231	
MeanIncome	3.0505	0.0000	0.3741	0.4171
PctImm	- 2.7005	0.0000	0.03246	

Based on preliminary experiment, the square of housing age is included in the model specification to capture the nonlinear relationship between property value and its age. It is observed that HouseAge has a negative sign and Age^2 has a positive sign. This confirms the general expectation that house prices will vary with age at a decreasing rate while very old houses may actually be quite valuable due to their historical or cultural significance.

[8] This column records the value of R^2 when the corresponding variable was used as the single regressor.

[9] This column records the value of R^2 when the corresponding group of attributes were put into regression as determinants.

[10] Since HouseAge and Age^2 are describing the same aspect of residential properties, the R^2 value was reported when both of them are used as regressors.

Coefficients of other characteristics also have plausible signs and magnitudes. For example, house price has a tendency of increasing by \$ 2311 after one month period which is compatible with the long term pattern of housing market. The results also reveals that all accessibility attributes contribute significantly in a negative manner to the housing values as expected that close proximity to social amenities like school, shopping center or public transit will add monetary value to properties.

In Table 4.2, the parameter estimate and the respective explanatory power for each variable when it is entered into regression as the single determinant as well as the relative explanatory power for each category are reported. As seen in Table 4.2, the highest R-square is achieved by MeanIncome which indicates the substantial impact that neighborhood characteristics have upon units' values. Structural characteristics, as a whole, only account for 25.77% of the variations exhibited in housing market; its relatively poor performance may be caused by the limited information available in this category. With respect to locational characteristics, although their significance is confirmed again by P-values, their individual or joint explanatory power is trifling as revealed by low values of R-square in Table 4.2. In addition, as single determinant, the coefficients of DistShop and DistSchool both have implausible positive sign, which is contradictory to initial regression results in Table 4.1 and is also incompatible with general expectation. In this light, DistShop and DistSchool will be excluded from model specification.

In addition to the exclusion of DistShop and DistSchool, the use of census data as proxy to neighborhood characteristics also precludes the use of MeanIncome and PctImm from local modeling techniques, i.e. MWR, GWR and MWK, in that some of the estimation areas------the moving windows, are contained entirely in one or very few census tracts. This poses a problem in model estimation since the $X'X$ matrix may not be invertible in this situation. Therefore, in model comparison conducted in section 4.7, there will be two suites of independent variables: the first suite of determinants includes Area, Front, SaleDate, HouseAge, Age^2 and DistTransit which is applicable to all model specifications; the second set adds neighborhood attributes, i.e. MeanIncome and PctImm to the first one, which is only available to global modeling techniques including traditional hedonic model and Unviersal Kriging. Parameter estimates and other statistics of traditional hedonic model using different sets of regressors are present in Table 4.3 and Table 4.4.

Table 4.3: Traditional Hedonic Model Regression Results (OLS1)*

R-squared:	**0.2917**	**Degrees of freedom:**	**33487**
Residual SS:	**2.0059E+09**	**Residual Standard Errors:**	**244.7**
F-statistic:	**2299**	**Probability of F:**	**0.000**

Variable	Estimate	Std Error	t-value	Prob> |t|
Constant	351.7978	7.1152	49.4434	0.0000
Area	1.6319	0.0473	34.5268	0.0000
Front	5.0239	0.1355	37.0764	0.0000
Saledate	2.6194	0.1311	19.9828	0.0000
HouseAge	- 116.9609	1.8517	- 63.1648	0.0000
Age^2	11.3280	0.1809	62.6048	0.0000
DistTranist	-45.9607	1.1450	-40.1402	0.0000

Table 4.4: Traditional Hedonic Model Regression Results (OLS2) **

R-squared:	**0.5737**	**Degrees of freedom:**	**33485**
Residual SS:	**1.2074E+09**	**Residual Standard Errors:**	**189.9**
F-statistic:	**5633**	**Probability of F:**	**0.000**

Variable	Estimate	Std Error	t-value	Prob> |t|
Constant	282.4466	5.8252	48.4874	0.0000
Area	1.3725	0.0367	37.3819	0.0000
Front	3.1740	0.1064	29.8354	0.0000
Saledate	2.7868	0.1017	27.3986	0.0000
HouseAge	- 96.0504	1.4439	- 66.5238	0.0000
Age^2	8.7717	0.1414	62.0185	0.0000
DistTransit	- 19.4807	0.9060	- 21.5017	0.0000
MeanIncome	2.6424	0.0188	140.5247	0.0000
PctImm	- 3.0300	0.0539	- 56.1787	0.0000

* *without neighborhood attributes: MeanIncome and PctImm*

** *with neighborhood attributes: MeanIncome and PctImm*

Dubin (2004) argued that the exclusion of neighborhood variables from local models won't impair their predictive ability for the reason that neighborhood attributes obtained at census level haven't varied enough in the small geographic areas used for local regression. To further investigate whether neighborhood attributes are essentially indispensable to hedonic housing price analysis and whether the exclusion of them will subsequently affect the validity of local modeling techniques, section 4.7 will compare the prediction performance of local models not only to traditional hedonic model and Universal Kriging with same set of determinants, but also to them with extra neighborhood attributes.

4.2 Diagnostic Tests of Spatial Autocorrelation

4.2.1 Visual Inspection

An intuitive way to inspect the underlying pattern of spatial process is to visualize it, using symbol to represent each observation while the color or size of symbol will indicate the magnitude of observed values.

This preliminary approach gives a hint regarding potential existence of spatial autocorrelation underlying the spatial phenomena: clusters of high values or clusters of positive values of residuals are referred as hot spots; clusters of low values or clusters of negative values of residuals are referred as cold spots. If it is observed that hot spots or cold spots exhibit pattern over space, then positive spatial autocorrelation may occur in this process. Otherwise, if sites of high values or positive residuals are surrounded by low valued observations or negative residuals, or sites of low values or negative residuals are neighboring to high valued observations or positive residuals, then this will bring cautions as to the prevalence of negative spatial autocorrelations over space.

In the scenario of urban housing market, due to adjacency effects and neighborhood effects discussed before, it is more natural to anticipate positive spatial autocorrelation prevailing over space. Such hypothesis is supported by Figure 4.1 which depicts the spatial distribution of housing transaction prices and reveals strong spatial dependence.

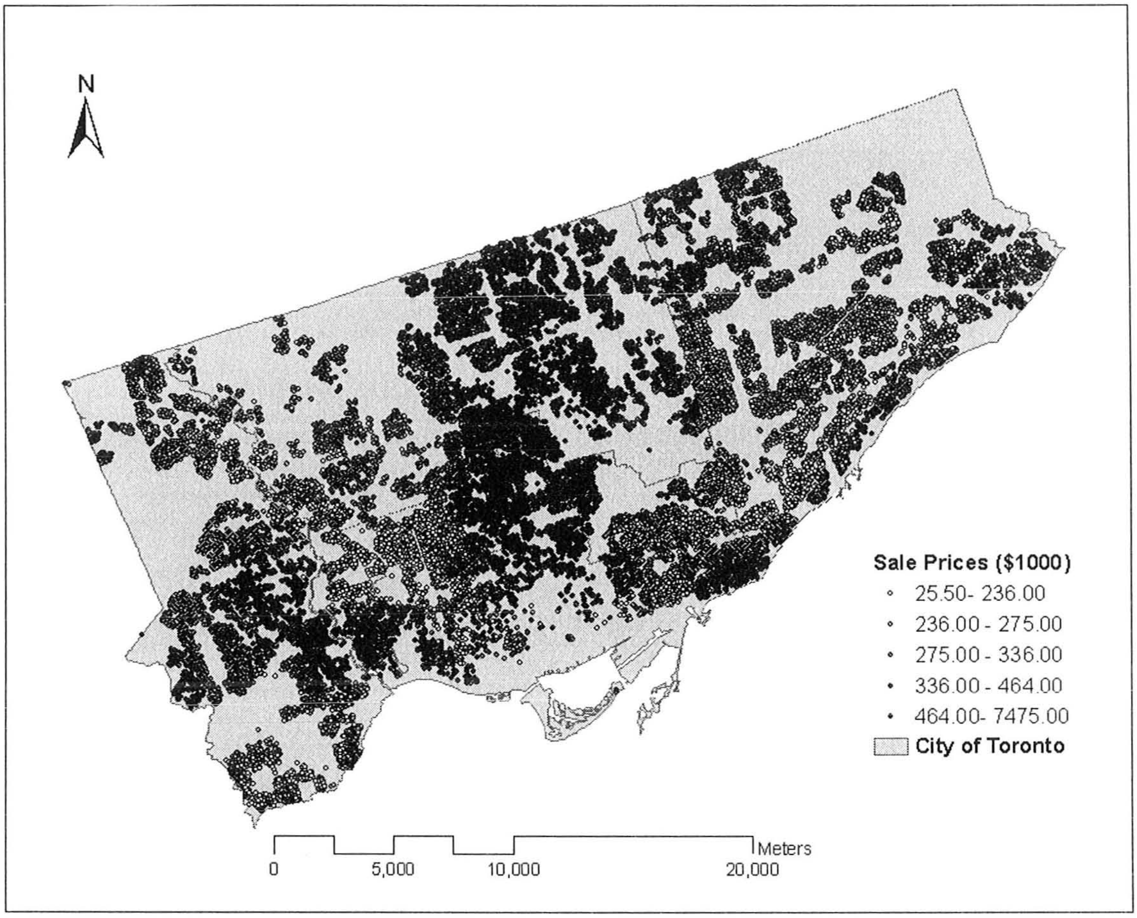


Figure 4.1: Spatial Distribution of Housing Prices

Even with inclusion of a number of explanatory variables into traditional hedonic model to account for influential factors, there is still the possibility of observing occurrence of spatial autocorrelation in residuals. This might be caused by failure of incorporating adjacency effects, or simply due to model misspecification such as incomplete information in this specific case. Such phenomenon is especially evident in Figure 4.2 that graphs geographic distribution of residuals from OLS1 (traditional hedonic model without neighborhood attributes). The visual examination reveals almost the same spatial pattern as observed in Figure 4.1. As regards OLS2 (traditional hedonic model with neighborhood attributes), visualization of its subsequent residuals might also suggest the existence of spatial dependence; however, unlike transaction prices or residuals from OLS1, residuals are more evenly distributed in Figure 4.3 and display less concentration of hot spots.

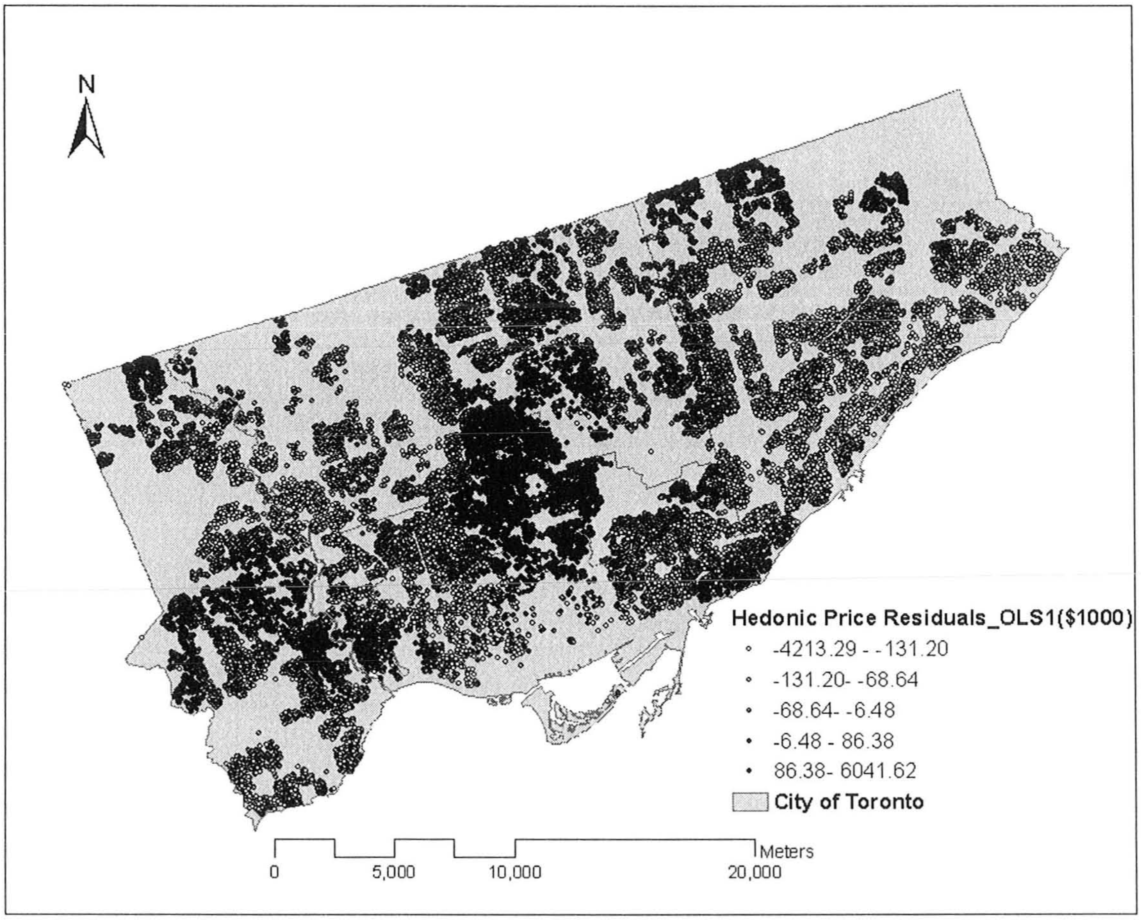


Figure 4.2: Spatial Distribution of Traditional Hedonic Price Residuals (OLS1) *

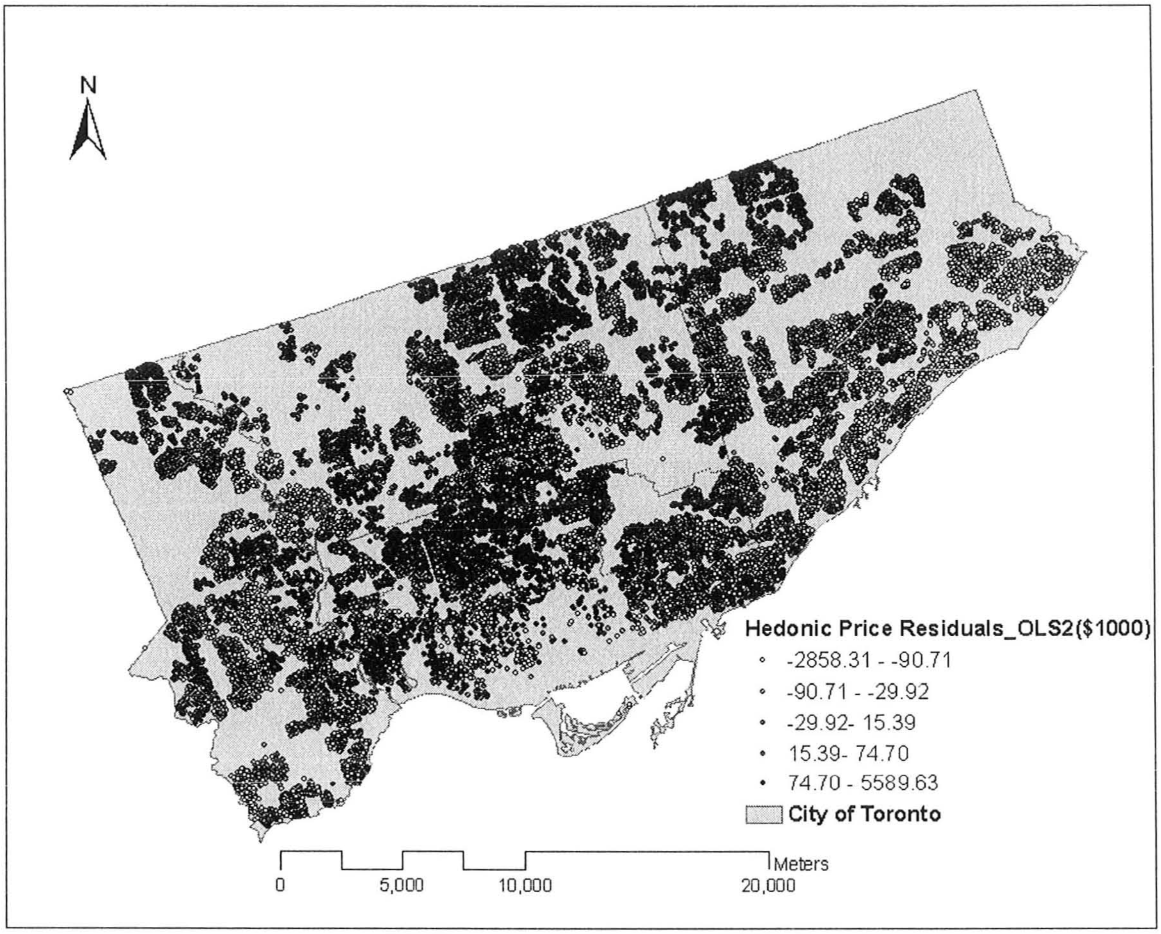


Figure 4.3: Spatial Distribution of Traditional Hedonic Price Residuals (OLS2) **

* *without neighborhood attributes: MeanIncome and PctImm*

** *with neighborhood attributes: MeanIncome and PctImm*

4.2.2 Moran's I index

Visual inspection is helpful in that it provides a vivid image about the spatial arrangement of observations and gives insights into the potential spatial pattern. However, intuitive impression obtained from visualization may be misleading or ambiguous, especially when the underlying process is not apparent enough to arouse sufficient attention to it. Therefore formal diagnostic test of spatial autocorrelation Moran's I index is conducted in this section. Recalling the definition of Moran's I from section 3.3.1, it estimates the correlation of a single variable between pairs of neighboring observations. Thus, a crucial step concerning its application is the definition of neighbors which is given by W in equation 3.1. Can and Megbolugbe (1997) argued that the spatial dependency occurring in the Miami property market is contained within a 3km radius. On the other hand, Farber (2004) suggested that due to the density of sales in Toronto housing market, a 3km neighborhood radius would probably be too broad to define local neighbors.

Adopting Farber's suggestions combined with exploration of semivariogram at local housing submarket which indicates that approximately 50% local markets are within 1000 to 1500 meters range (details will be provided in section 4.3), the following weight matrices W are created:

W_{1000} ------ W_{ij} equals 1 if the distance between i and j is less than or equal to 1000 meters, otherwise W_{ij} equals 0; the design of W implies that if observations are within 1000 meters, they are regarded as neighbors.

W_{1500} ------ W_{ij} equals 1 if the distance between i and j is less than or equal to 1500 meters, otherwise W_{ij} equals 0; the design of W implies that if observations are within 1500 meters, they are regarded as neighbors.

The value of Moran's I and its corresponding P-value after 999 permutation using Monte-Carlo approach is displayed in Table 4.5:

Table 4.5: Moran's I index of Spatial Autocorrelation for Sale Prices

	Moran's I	P-value
W_{1000}	0.4818	0.001
W_{1500}	0.4055	0.001

Despite of usage of two different weight matrices to define neighbors, it is still far from confident to assert their soundness of depicting the reality. Especially when a number of studies (Anselin 1995; Oliveau and Guilmoto 2005) have pointed out that the strength of spatial autocorrelation is greatly influenced by the choice of local neighbors.

As Oliveau and Guilmoto (2005) observed in their study of exploring India's demographic pattern that the variable of interest may be strongly autocorrelated at local level, but display no correlation over a larger extent, it is not surprising to observe a radically altered Moran's I value which may range from significantly positive value to insignificant value when different weight matrix is applied.

Another motivation to perform Moran's I autocorrelation analysis using various definitions of local neighbors rises from the hypothesis of searching for possible connections between degree of spatial autocorrelation and performance of MWR (moving window regression). It will be discussed later that for spatial regression and spatial interpolation techniques like MWR, GWR or MWK, an inherent difficulty is the determination of window size or the definition of neighbors. If MWR or other models perform best when Moran's I attained the highest value, it would be natural to choose the moving window which displays the strongest autocorrelation.

Attending the preceding considerations, the following weight matrices W will also be applied:

$Wn - Wij$ equals 1 if the distance between i and j is less than or equal to the distance between i and its nth nearest neighbor. (n takes a series of values: 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 1000, 1500)

The values of Moran's I and their respective P-values employing Monte-Carlo approach with 999 randomizations are reported in Table 4.6. By plotting Moran's I value to the corresponding number of nearest neighbors, Figure 4.4 is obtained which shows that the degree of spatial autocorrelation among transactions decreases substantially as the number of nearest neighbors increases.

Table 4.6: Moran's I index for Sale Prices Using Different Weighting Matrix

ID	Num of Neighbors	Moran's I	P-value
1	50	0.566	0.001
2	100	0.523	0.001
3	150	0.493	0.001
4	200	0.471	0.001
5	250	0.453	0.001
6	300	0.437	0.001
7	350	0.424	0.001
8	400	0.412	0.001
9	450	0.401	0.001
10	500	0.399	0.001
11	1000	0.332	0.001
12	1500	0.303	0.001

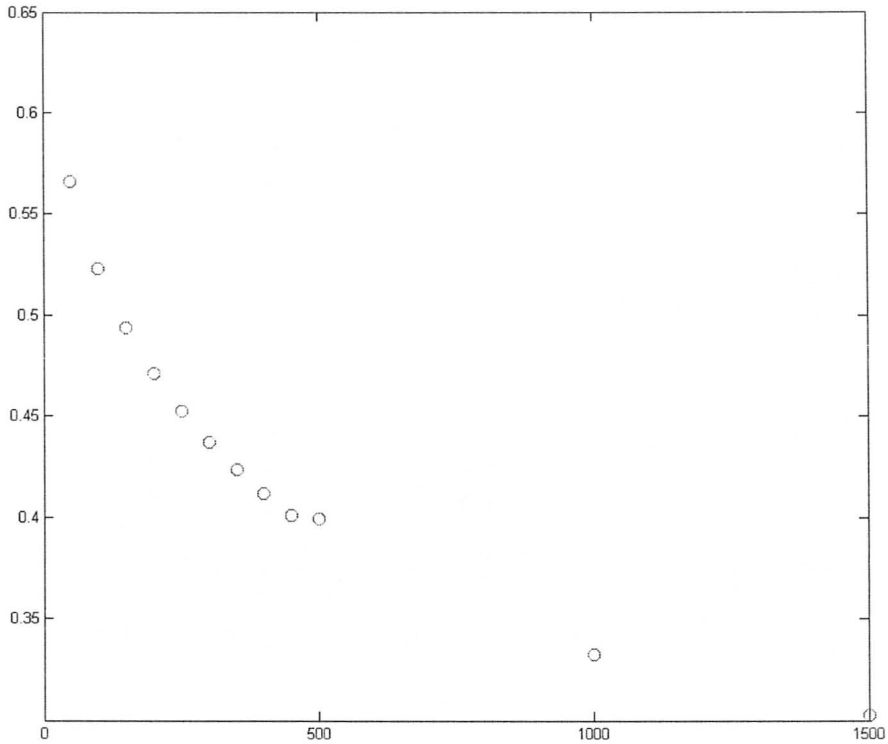


Figure 4.4: Scatter Plot of Moran's I versus Number of Neighbors

Unfortunately, the pattern shown in the above figure doesn't coincide with the trend revealed in section 4.4.2 between size of neighborhood and performance of local modeling technique. Thus, no relationship can be derived to shed lights on the selection of optimum window size based on the corresponding spatial autocorrelation strength.

As observed in Figure 4.2 and Figure 4.3, when the first set of determinants is applied in traditional hedonic modeling, spatial autocorrelation is prevalent in the residuals; on the other hand, when controlling for neighborhood characteristics, residuals might still exhibit spatial dependency, but in a minor way. The visual inspection is confirmed by formal diagnostic test and results are displayed in Table 4.7:

Table 4.7: Moran's I index of Spatial Autocorrelation for Hedonic Price Residuals

	Moran's I	P-value
* W_{1000}	0.4480	0.001
* W_{1500}	0.4011	0.001
** W_{1000}	0.1657	0.001
** W_{1500}	0.1389	0.001

* *without neighborhood attributes: MeanIncome and PctImm*

** *with neighborhood attributes: MeanIncome and PctImm*

The sharp difference between strength of autocorrelation displayed by residuals from OLS1 and OLS2 implies that the inclusion of neighborhood attributes diminishes spatial autocorrelation substantially for this particular case. However, whether this improvement is sufficient to control for spatial dependency or whether incorporating spatial autocorrelation is still necessary or beneficial after accounting for neighborhood characteristics still lacks of explicit answer. These questions will be further explored in section 4.4.4.

4.2.3 Empirical Variogram

As discussed in section 3.3.2, the empirical variogram not only helps diagnose potential pattern in the spatial process, but also conveys a rough impression about the geographic extent of spatial dependency structure. Under such purpose, semivariogram for transaction prices is graphed in Figure 4.5, while Figures 4.6 and 4.7 provide semivariograms for traditional hedonic price residuals from OLS1 and OLS2.

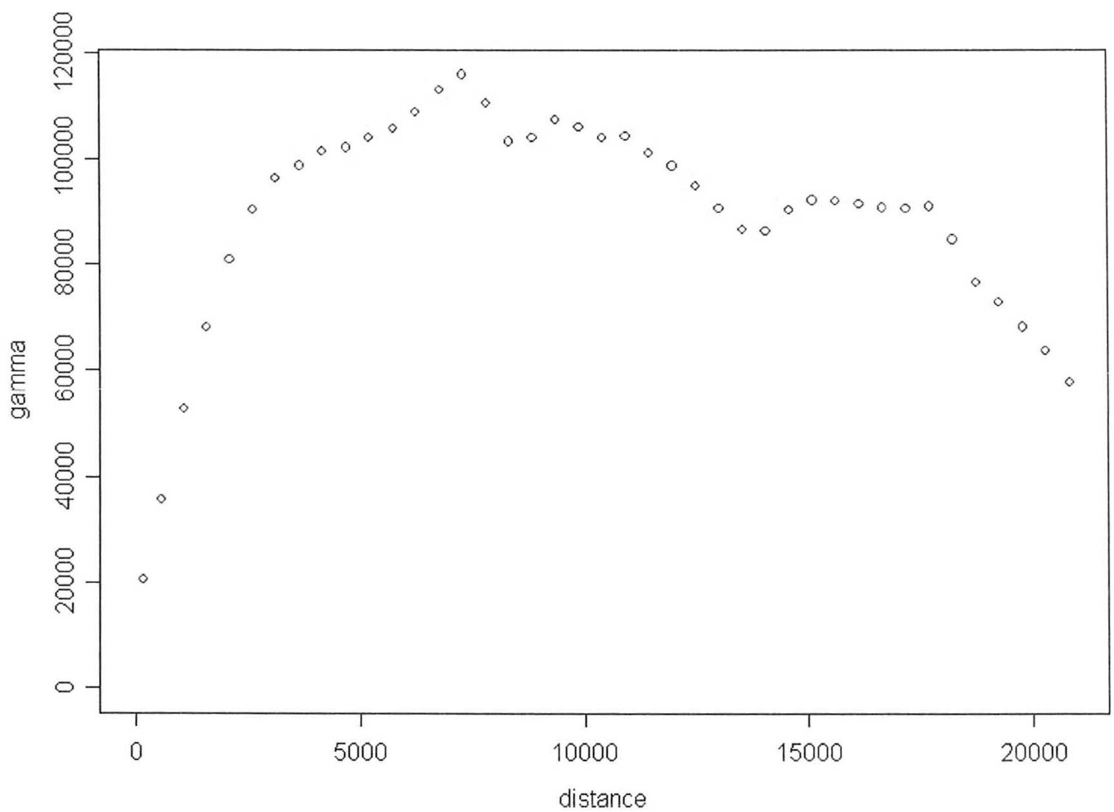


Figure 4.5: Empirical Variogram for Sale Prices[11]

In Figure 4.5, the value of $r(h)$ goes up steadily until it reaches a sill at 3.5 kilometers, after which there is small fluctuation but it still remains more or less constant until it levels off at 15 kilometers, afterwards the value of r decreases slowly. At first glance, the shape of the observed empirical variogram is out of ordinary in that property values start to re-correlate when they are 15 kilometers apart or more distant. Undoubtedly, such correlations of transaction prices for properties apart by long distance is not caused by neighborhood effects or adjacency effect as observed for properties

[11] . For Figure 4.5, the x axis is the distance lag which is measured by meters, and the y axis is the variogram estimation whose unit is 10^6 dollars. The definitions and units for x, y axis are also applicable to other variogram plots in this chapter.

located in vicinity. If the polycentric nature of urban structure is taken into consideration, this phenomenon has a reasonable explanation: local housing markets are developing in multiple urban centers; for properties sited at different local markets, even they are far apart, if they are characterized by similar proximity to local market center or having similar structural attributes (which is common due to the standardization of modern architecture), then it is not surprising to observe similar property values at distant space. However, it would be more appropriate to name such relationship as 'similarity' rather than 'correlation'.

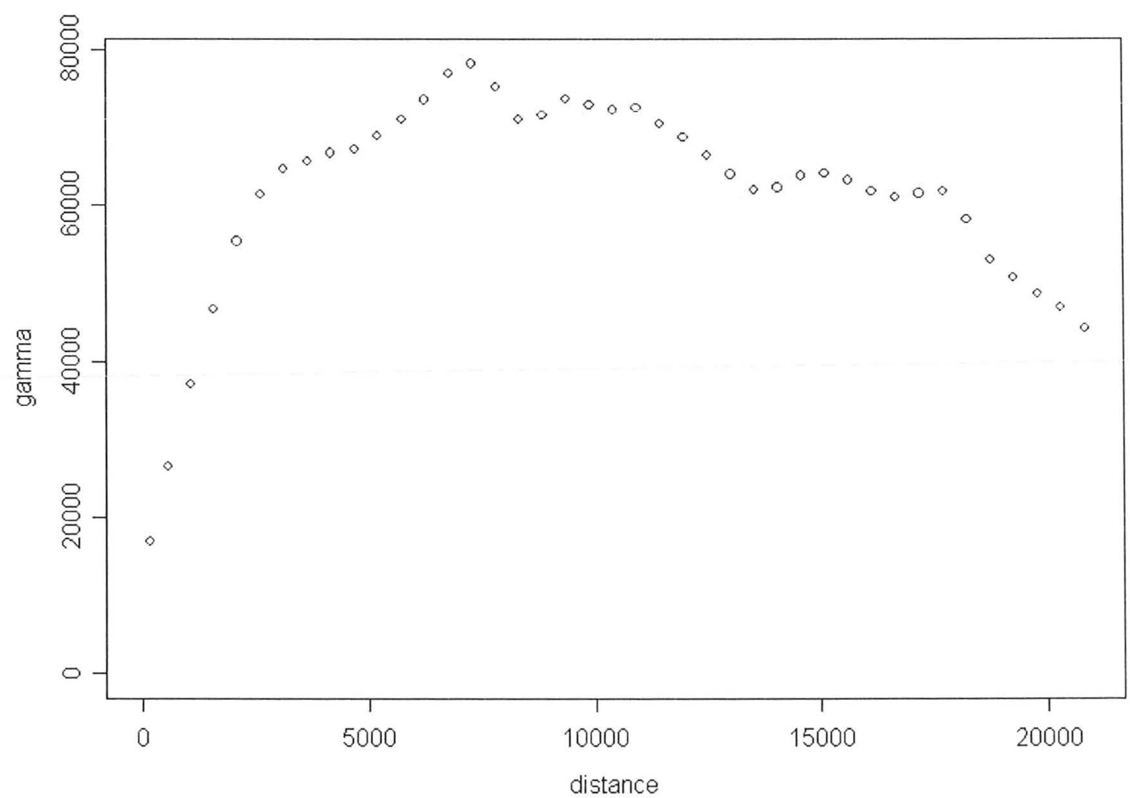


Figure 4.6: Empirical Variogram for Traditional Hedonic Price Residual (OLS1) *

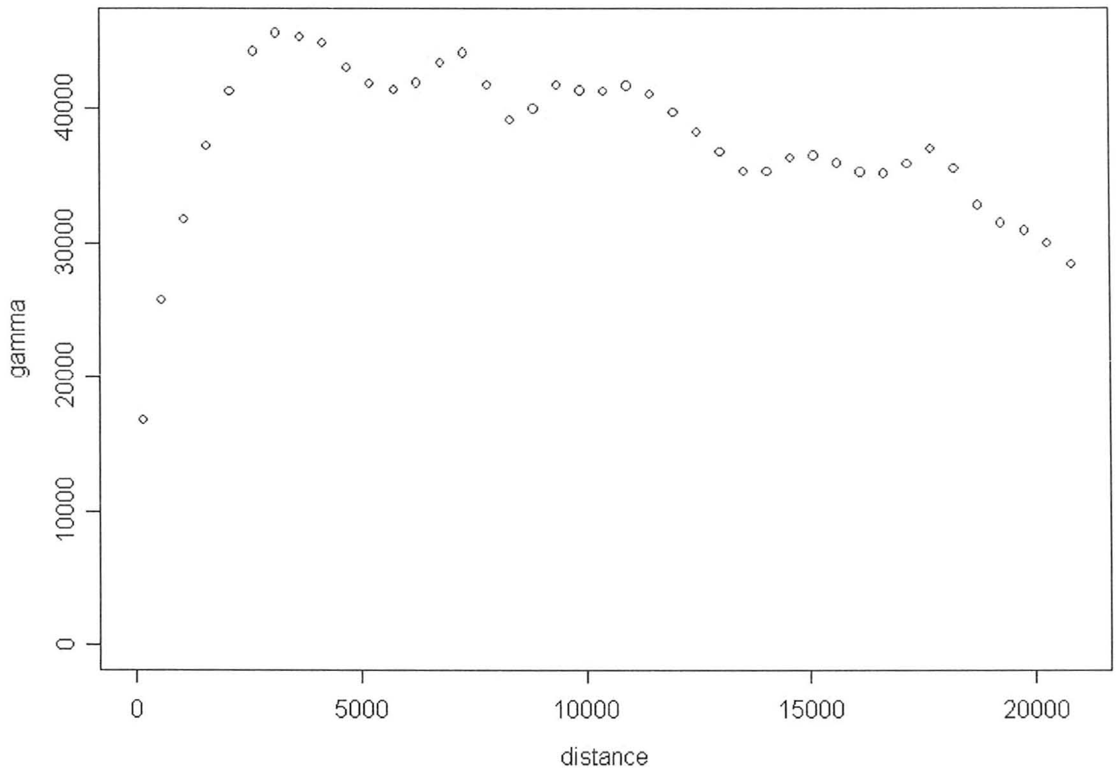


Figure 4.7: Empirical Variogram for Tradional Hedonic Price Residuals (OLS2) **

* *without neighborhood attributes: MeanIncome and PctImm*

** *with neighborhood attributes: MeanIncome and PctImm*

The semivariogram for traditional hedonic price residuals obtained from OLS1 displays analogous trend as observed in Figure 4.5 for transaction prices. The value of $r(h)$ also increases and reaches a sill at the range of approximate 3.5 kilometers, where since a trend was removed from the process, residuals have smaller variance compared to original prices. Then, like Figure 4.5, it exhibits small fluctuations until leveling off at 15 kilometers, after which it decreases slowly. On the other hand, Figure 4.7 presents a different pattern where the semivariogram for hedonic house price residuals from OLS2 increases more sharply until reaching 2.7 kilometers, after which it levels off with small variation. More importantly, it has a higher nugget effect which is also an indication of less autocorrelation.

The above results confirms the impression gained from visual examination and Moran's I diagnostic tests. First, Figure 4.6 illustrates that even traditional hedonic specification controls for differentials in structural and proximity attributes of units, there is still discernable spatial autocorrelation remaining in residuals which may necessitate the efforts of incorporating it into model specification like Universal Kriging and moving window kriging (MWK) do. Secondly, traditional hedonic specification with inclusion of neighborhood attributes diminishes residual spatial dependency substantially, but whether such control appear to be sufficient to release the tension of modeling spatially autocorrelated error terms still remains as question. This issue will be further investigated in section 4.4.4.

4.3 Diagnostic Test of Spatial Heterogeneity

As discussed in section 3.2, spatial heterogeneity will be diagnosed through sub-area variogram modeling in which sub-areas are delineated along two schemes: one is to ensure sub-areas have equal size and the other one is to ensure sub-areas contain more or less the same number of observations. This experiment is conducted for transaction prices. Summary statistics for parameters of local variograms are reported in Table 4.8 and Table 4.9.

Table 4.8: Local Variogram for Sale Price with 1000 Observations

Fixed Obs	Nugget (10^2)	Sill (10^2)	Variance(10^2) [12]	Range (m)	Nugget [13]Effect (%)
Min	4.82	5.90	12.51	200.00	0.050
1st Qu	15.16	24.84	40.00	700.00	0.203
Mean	144.93	494.36	639.30	1156.91	0.358
Median	30.17	47.04	80.34	1200.00	0.329
3rd Qu	90.74	200.66	264.50	1550.00	0.519
Max	1527.57	4428.10	5228.55	2100.35	0.744
Std Dev.	313.72	1085.11	1371.65	545.90	0.186

Table 4.9: Local Variogram for Sale Price with Fixed Radius

Fixed radius	Nugget (10^2)	Sill (10^2)	Variance(10^2)	Range (m)	Nugget Effect (%)
Min	1.16	2.29	8.79	120.00	0.052
1st Qu.	13.44	15.82	29.63	767.51	0.251
Mean	86.67	404.58	491.26	1104.47	0.360
Median	27.66	44.35	65.72	1100.00	0.313
3rd Qu	59.32	126.54	184.24	1475.00	0.468
Max	646.85	4208.99	4715.05	2117.34	0.765
Std Dev.	155.24	1034.03	1168.23	523.29	0.189

According to the descriptive analysis of local variogram parameters, sub-areas exhibit distinct variance-covariance structure in terms of shape of empirical variogram as well as parameter magnitudes. A different way to show the substantial variations among sub-regions is to plot histogram of parameters and place it in the corresponding subdivision. Figure 4.8 to Figure 4.11[14] are the spatial histogram maps of parameters for transaction price using subregions with fixed radius or fixed observations.

[12] variance=sill+nugget

[13] nugget effect=nugget/sill+nugget. It measures the degree of spatial dependency.

[14] Since the original values of nugget, sill and variance exhibit tremendous variations cross sub-areas in which most values become negligible compared to the upper bound, Figure 4.9 and Figure 4.11 plot the parameter nugget effect.

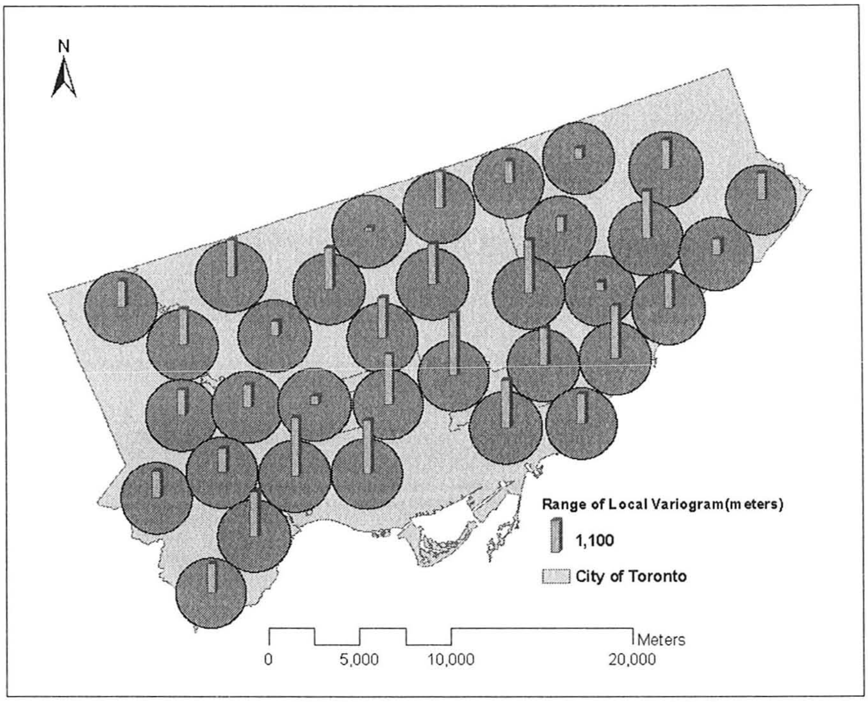


Figure 4.8: Spatial Histogram Map for 'Range' of Sales Prices (a)

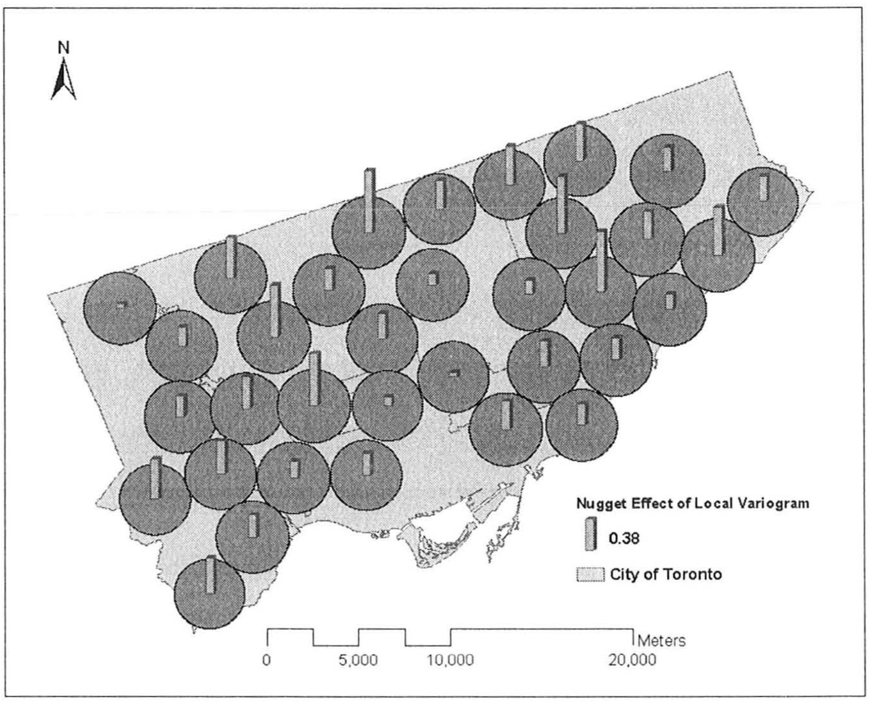


Figure 4.9: Spatial Histogram Map for 'Nugget Effect' of Sales Prices (a)

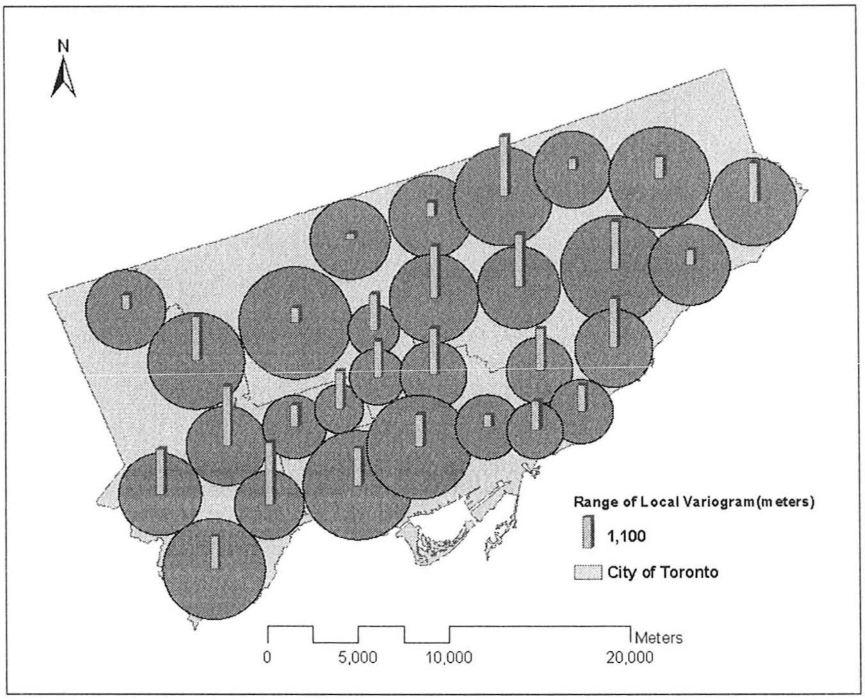


Figure 4.10: Spatial Histogram Map for 'Range' of Sales Price (b)

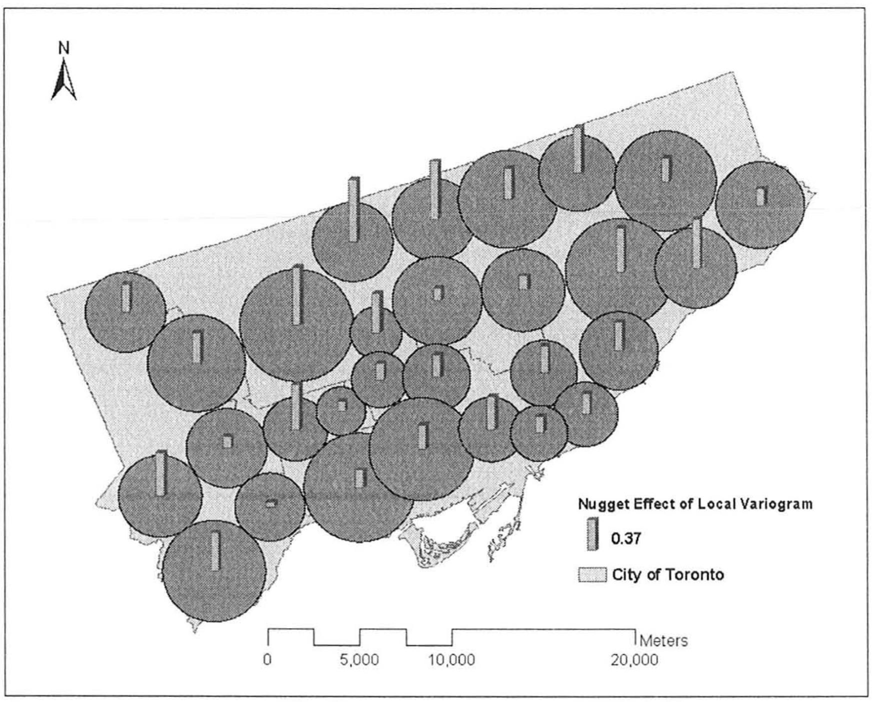
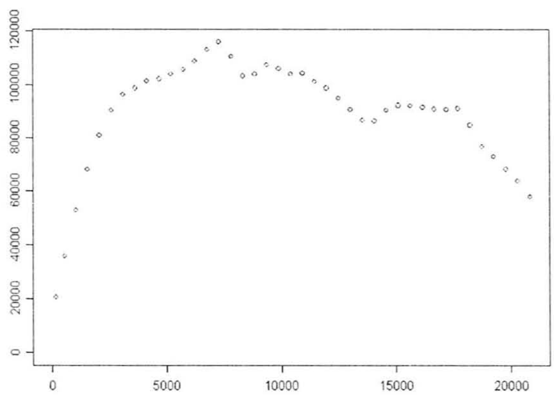
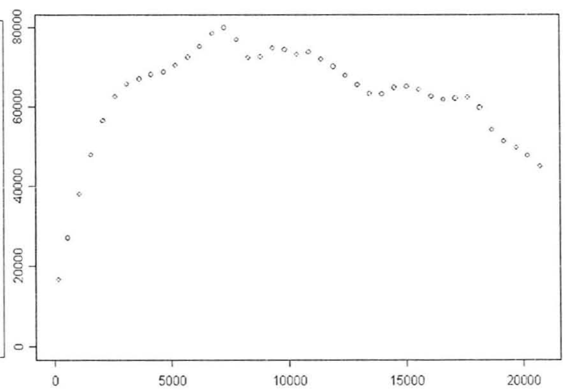


Figure 4.11: Spatial Histogram Map for 'Nugget Effect' of Sales Price (b)

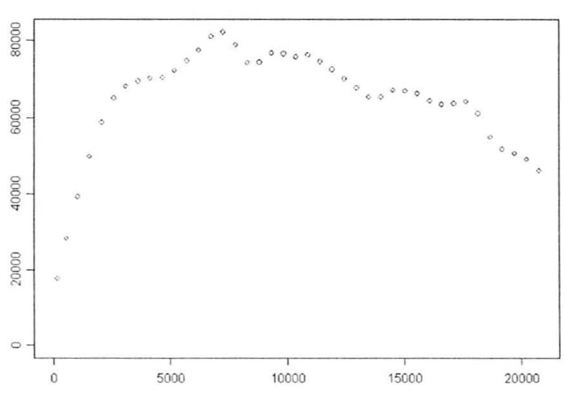
Despite the descriptive analysis and visualization of local variogram estimation, it is not certain that whether the difference between local variogram and whole-area variogram is raised by distinctive spatial dependency structure cross local markets or it is simply caused by smaller sample size used for local variogram. To distinguish these two effects, sample thinning approach is employed here which uses randomly selected sample to estimate empirical variogram for the whole area. The size of random sample created for comparison purpose ranges from 32,449 to 1,000. The corresponding semivariograms are displayed in Figure 4.12 with the sequence of decreasing sample size.



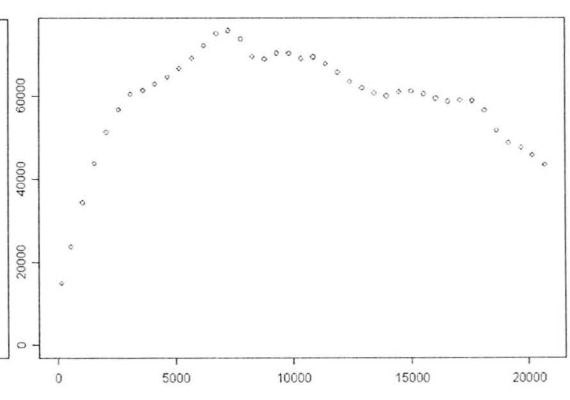
sample size: 33,494



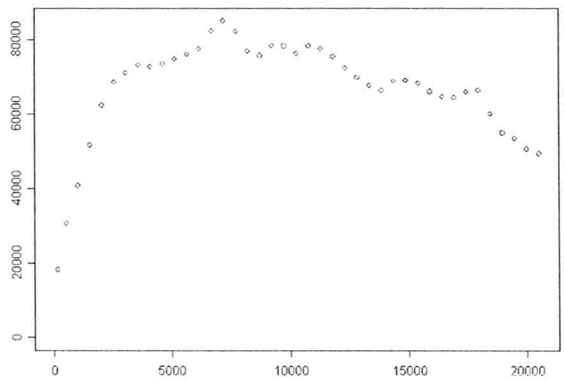
sample size: 30,000



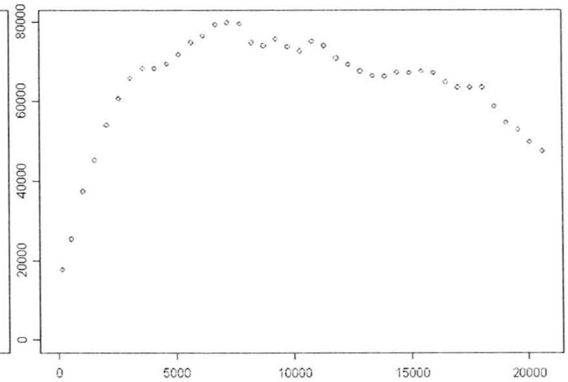
sample size: 25,000



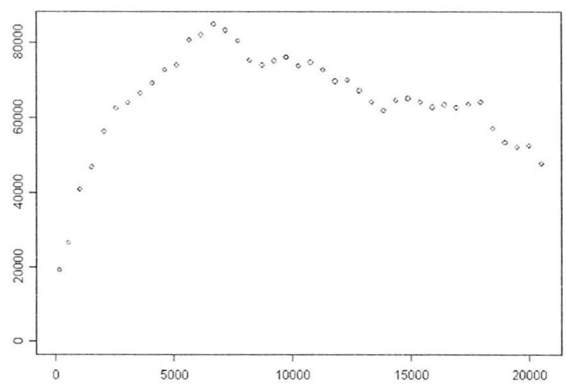
sample size: 20,000



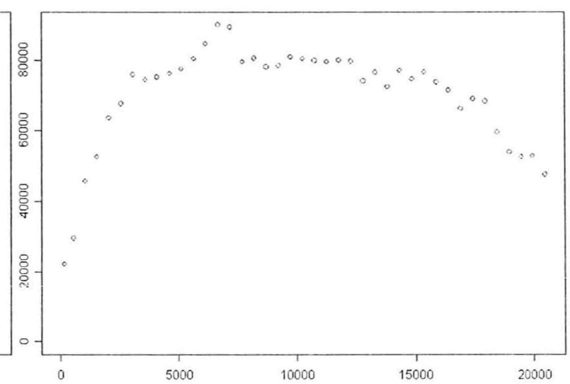
sample size: 15,000



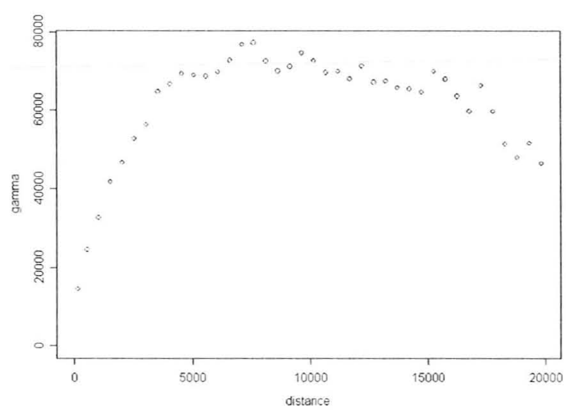
sample size 10,000



sample size: 7,000



sample size : 3500



sample size: 1000

Figure 4.12: Empirical Variograms for Different Sized Samples

It is observed from the above figures that even as sample size varies dramatically, the shape and magnitude of empirical variogram remains relatively stable. As demonstrated by this experiment and the earlier local variogram modeling, there are clear differences between random sampling (aspatial sampling) and spatial sampling approach. Whilst the former one retained the spatial dependence structure with stable variogram estimates despite varying sample size, the latter one, variograms computed using spatial subsets, exhibited considerable variations cross space.

Therefore, it is safe to conclude that such variation is not generated by different sample size used for global variogram and local variogram but simply caused by diverse spatial structure across urban housing market. Since the price determination processes operating at sub-markets are dissimilar, modeling spatial heterogeneity becomes an essential concern for this particular case. In the coming sections, moving window regression (MWR), geographically weighted regression (GWR) and moving window kriging (MWK) which intend to portray the varying relationship between property value and the determinants are applied and their predictive performance are also evaluated.

4.4 Model Comparisons

As illustrated in chapter 3, in moving window regression (MWR) and geographically weighted regression (GWR), spatial heterogeneity is formally built into model specification by performing regression at neighborhood level and subsequently allowing varying relationship between a number of factors; the other fundamental feature of urban housing market ------ spatial dependency is explicitly included into Universal

Kriging via the variance-covariance matrix of residuals; in moving window kriging (MWK) which can be thought of as the moving window version of Universal Kriging, both spatial dependence and spatial heterogeneity are considered and specified into modeling process. In this section, these spatial models are applied using the dataset provided by MPAC and their predictive performances are compared to traditional hedonic model which includes no spatial effects.

The idea of comparing the performance of models in terms of their predictive power derives from experiment conducted in 2004 on modeling spatial and temporal components of house prices (Case et al. 2004). In said paper, the data was divided into in-sample observations and out-of-sample observations: the in-sample observations were distributed to 4 independent researchers and the out-of-sample observations are withheld and their values are used to estimate the performance of various models submitted by these participants.

Since it is desirable to test the predictive power of models using a different sample than the one used for parameter estimates, in this study, the original dataset, which consists of 33,449 observations, is also divided into two groups ------ an estimation sample, i.e. in-sample observations and a prediction sample, i.e. out-of-sample observations. The estimation sample contains 30,145 observations, which is a 90% random sample, whereas the prediction sample is comprised of 3,349 observations, which is comprised of 10% transactions.

The prediction to be executed later treats the in-sample observations as sampled sites with known values and attributes, while the out-of-sample observations can be thought of as unsampled sites whose values will be predicted by the estimated model. Summary statistics for estimation sample and prediction samples are reported in Table 4.10 and 4.11. Their spatial distributions are displayed in Figure 4.13 and 4.14.

Table 4.10: Summary Statistics for Estimation Sample

	Mean	Standard Deviation	Range	
			Minimum	Maximum
SalePrice ($1000)	384.406	293.100	25.500	7475.000
Area (feet^2/100)	54.098	43.895	6.222	3346.65
Front (feet)	42.198	16.323	10.090	422.000
HouseAge (decade)	4.808	2.447	0.000	17.100
SaleDate	19.613	10.219	1.000	36.000
DistTransit (km)	1.872	1.269	0.005	7.591
MeanIncome ($1000)	74.166	58.098	0.000	423.989
PctImm (%)	37.670	19.369	0.000	75.500

Table 4.11: Summary Statistics for Prediction Sample

	Mean	Standard Deviation	Range	
			Minimum	Maximum
SalePrice ($1000)	389.603	269.132	70.000	3870.000
Area (feet^2/100)	53.753	32.694	6.660	494.000
Front (feet)	41.964	15.414	15.090	140.000
HouseAge (decade)	4.805	2.491	0.000	17.700
SaleDate	19.798	10.183	1.000	36.000
DistTransit (km)	1.856	1.270	0.046	7.765
MeanIncome ($1000)	75.811	60.118	0.000	423.989
PctImm (%)	36.969	19.665	0.000	75.500

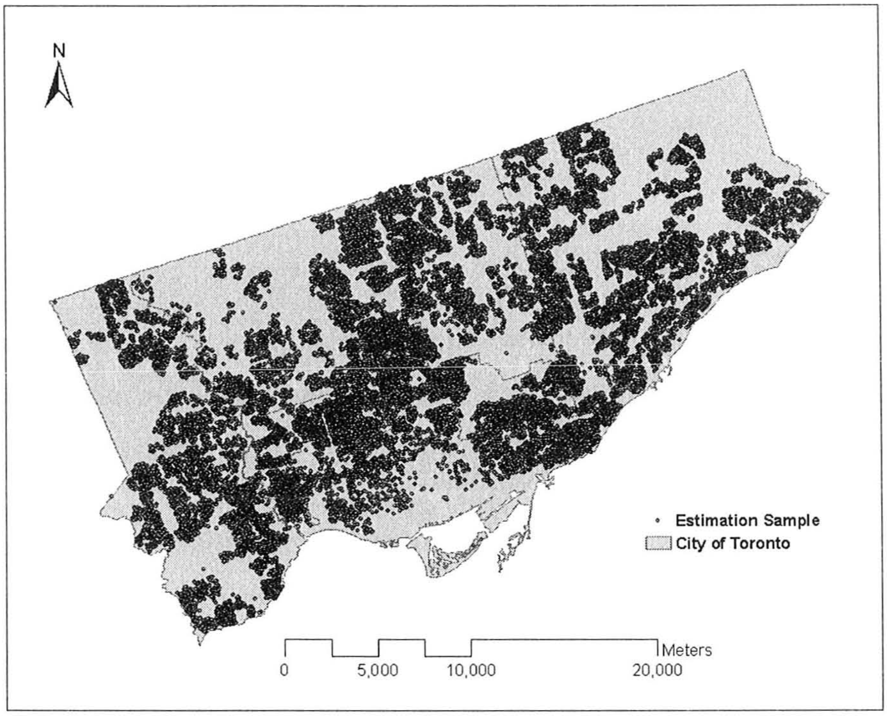


Figure 4.13: Spatial Distribution of Estimation Sample

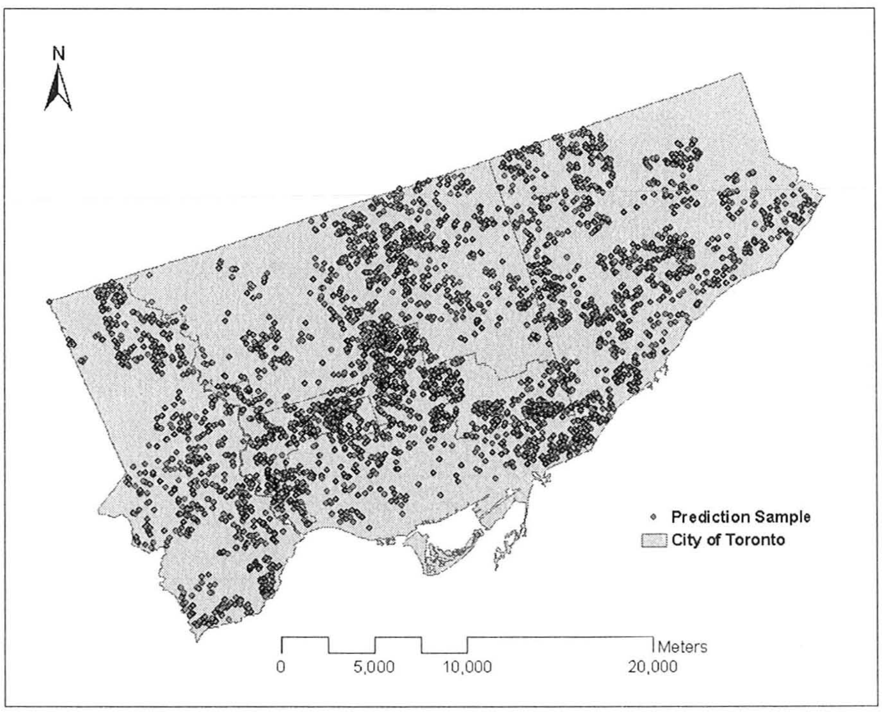


Figure 4.14: Spatial Distribution of Prediction Sample

Concerned with the evaluation of prediction capacity of various models employed in this study, the estimation sample is used for estimating the parameters of hedonic price functions, modeling the spherical semivariogram and therefore constructing the variance-covariance structure. Accordingly, the parameter estimations or the spatial dependency structure obtained from estimation sample are then used for predicting out-of-sample observations. By comparing the predicted and actual values of prediction sample, several statistics will be computed and the predictive performances of various spatial hedonic models will be derived in section 4.4.4.

As will be elaborated in the following section, the estimation sample is used not only for estimating parameters or spatial dependency structure, it is also used to solve the biggest mystery regarding the size of moving window required for local modeling approaches including moving window regression (MWR), geographically weighted regression (GWR) and mowing window kriging (MWK).

4.4.1 Optimally Defined Window Size

A common feature shared by moving window regression (MWR), geographically weighted regression (GWR) and mowing window kriging (MWK) is the usage of moving window approach. As previously discussed, by conducting regression analysis within the moving window, GWR and MWR are able to model the spatially varying relationship between dependent variables and a set of explanatory attributes; meanwhile, by calculating a local variogram function for each prediction site with observations situated in the window, MWK models location-dependent covariance structure and therefore is expected to generate more accurate assessment.

Based on the definition of these localized regression or interpolation techniques, it is evident that the moving window, in essence, enables them to take spatial heterogeneity, either varying relationship between variables or nonstationary covariance structure of residuals into account explicitly. However, a problem inherent with this approach is the determination of optimal moving window size.

Even with rapidly growing applications of these localized spatial techniques, how to determine an optimal moving window size as to yield the best regression or interpolation result still remains a mystery. Particularly, whether various definitions of moving window will affect modeling outcome, and if it does, to what extent the influence will be, still lack theoretical examination and exploratory inspection.

Despite that, recent empirical studies agree upon the manner in which the moving window is defined, in other words, whether to use circular moving window with fixed radius or use moving window which contains certain number of nearest neighbors regardless of shape or size. In general, the latter one is favored as to avoid extremely dense or sparse observations in the modeling process, a desirable advantage for unevenly distributed spatial phenomena.

With respect to the specific number of nearest neighbors contained in the moving window, Fotheringham et al. (2002) and Haas (1995) suggest using cross-validation to determine an appropriate size for moving window if there is no prior justification. Cross-validation is a technique in which the optimal number of nearest neighbors is the one that minimizes the following function:

$$CV = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

where n is the number of data points in the study area, y_i is the observed value for i th data point and $\hat{y}_i$ is the predictive value of i th point obtained when itself is omitted from the computation. More specifically, the i th observation is removed from the dataset and now the task is to predict its value using the remaining $n\text{-}1$ observations; procedure is repeated for all observations in the dataset. Subtracting the prediction value from the actual value generates prediction errors. The sum of the squared errors provides an effective and direct measurement of performance of models at different window sizes.

Since cross-validation is already embedded into the GWR software developed by Fotheringham and his colleague in University of Newcastle, it may seem easy to use the optimal number of nearest neighbors obtained by cross-validation in GWR software for use in MWR. However, as will become evident, the optimum size of moving window for GWR does not necessarily perform best for MWR.

For Moving Window Kriging, Haas (1990a, b) proposed that the window should be just large enough to accommodate enough observations to generate the variogram with accuracy sufficient for the intended uses of the process estimates. As indicated by Haas, the above statement regarding optimum window size comprises two conflicting criteria: on the one hand, the moving window is expected to be small as to approximate a stationary process; on the other hand, a larger moving window will contain more available observations for variogram estimates whose accuracy is partly a function of the number of couples used in each distance lag.

In attempt to reach equilibrium between the two contradictory criteria, most geologists choose a moving window of 100 observations, which is suggested by Webster and Oliver (1992). Dubin (2004) used 200 ~300 closet observations for Moving Window Kriging when modeling spatial variation of house prices, though, no justification is given for this decision. Haas proposed cross-validation to search for the optimum size of neighborhood, however, his advice is not adopted by most researchers who worry that the improved accuracy provided by optimum moving window may not justify the computational effort required for cross validation process.

In this research, given the main interest in comparing the performance of different spatial models in terms of prediction accuracy, it is of greatest importance to ensure that MWR, GWR and MWK are equipped with the optimum moving window and in accordance exhibit the maximum prediction accuracy. Otherwise, misleading conclusion would be derived since it will lay the blame for disagreeable prediction outcome on the model design rather than an improperly defined window size.

As described earlier, the dataset is divided into two subsets using random sampling. The estimation sample, consisting of 30,145 observations, will be used to determine the optimal sized neighborhood for MWR, GWR and MWK. The cross validation results will be provided in the following section.

4.4.2 Cross-validation Results

Before presenting cross-validation results, a note about that the explanatory variables used in local models is in order. First, compared to the variables specified in section 3.2, accessibility attributes DistShop and DistSchool are excluded from model specification since their marginal contribution is trivial as revealed by minor decrease in R-square from 0.575 to 0.5737 after leaving them out. Also, as reported in Table 4.2, their coefficients don't have expected sign when using them as single determinants. As well, neighborhood attributes obtained at census level are not applicable due to the use of moving window which might cause non-invertible $X'X$ matrix when MeanIncome and PctImm are included into model specification.

The first step is cross validation for moving window regression (MWR). 24 different window sizes ranging from 50 to 700 nearest neighbors are applied and the CV (cross validation score) is computed for each experiment. By plotting the value of CV to the window size measured by the number of nearest neighbors, Figure 4.15 is graphed in which X axis is the number of nearest neighbors for each trial and Y axis reports the corresponding value of CV. As implied by its definition, in general, the lower the CV is, the closer the predicted values are to the observed values. It is quite noticeable that the lowest values of CV are obtained within the range of 100 to 200 nearest neighbors. To take a closer look, the suspect area was zoomed out in Figure 4.16 which clearly shows that MWR performs best when the moving window contains 190 nearest neighbors.

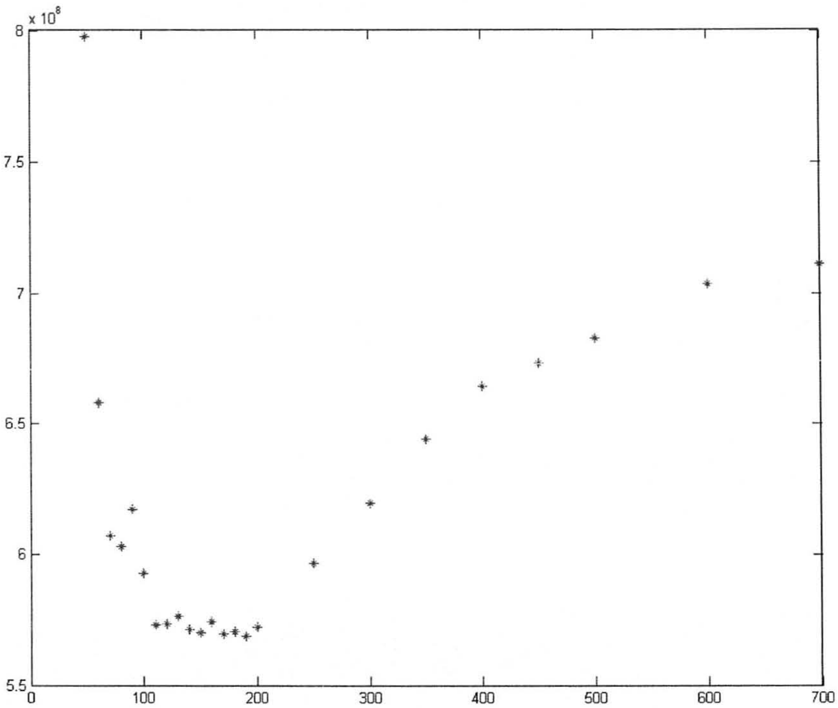


Figure 4.15: Cross Validation Results for MWR

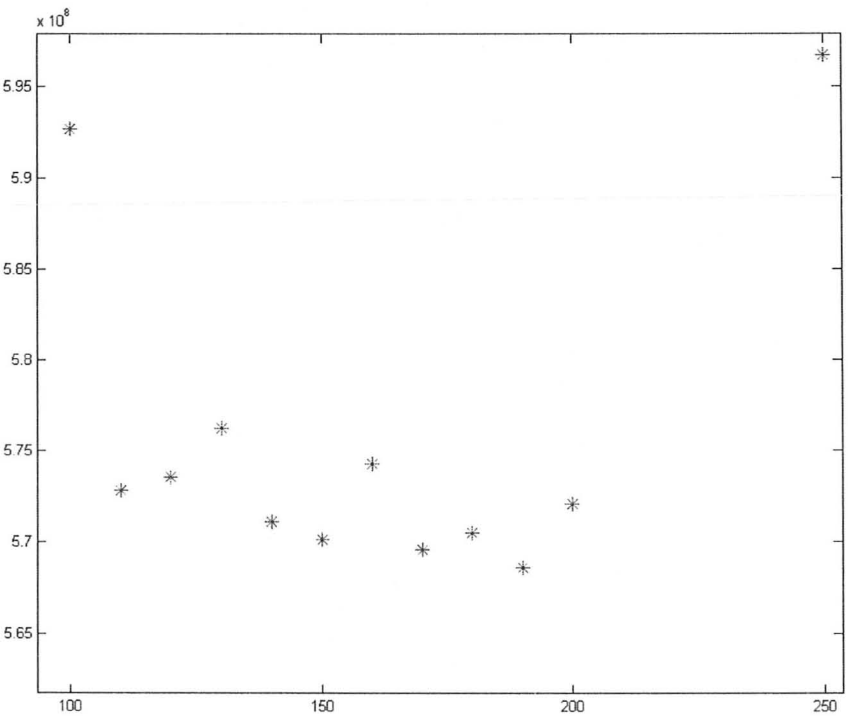


Figure 4.16: Cross Validation Results for MWR at A Closer Look

The same experiment is carried out for geographically weighted regression (GWR). The graph of CV versus window size is seen in Figure 4.17 showing that CV is minimized when 250 nearest neighbors were included into moving window.

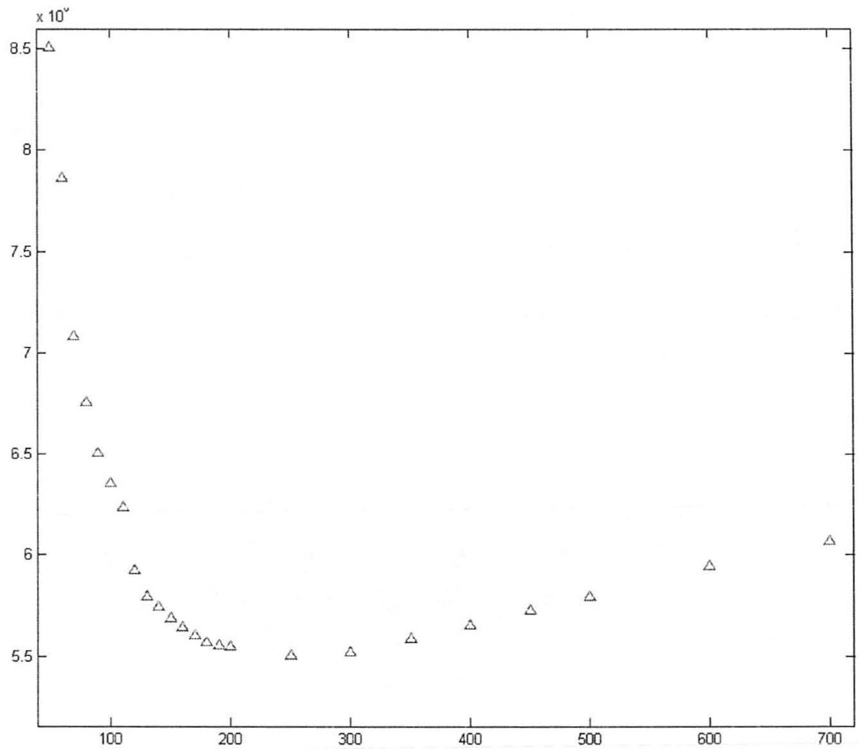


Figure 4.17: Cross Validation Results for GWR

Figure 4.18 and Figure 4.19 are the cross validation results for moving window kriging (MWK). Due to computation burden of the estimation procedure entailed in MWK, fewer moving window sizes are experimented here, ranging from 50 to 450. Nonetheless, it still captures the general trend and attains the lowest value at the bandwidth of 150 nearest neighbors.

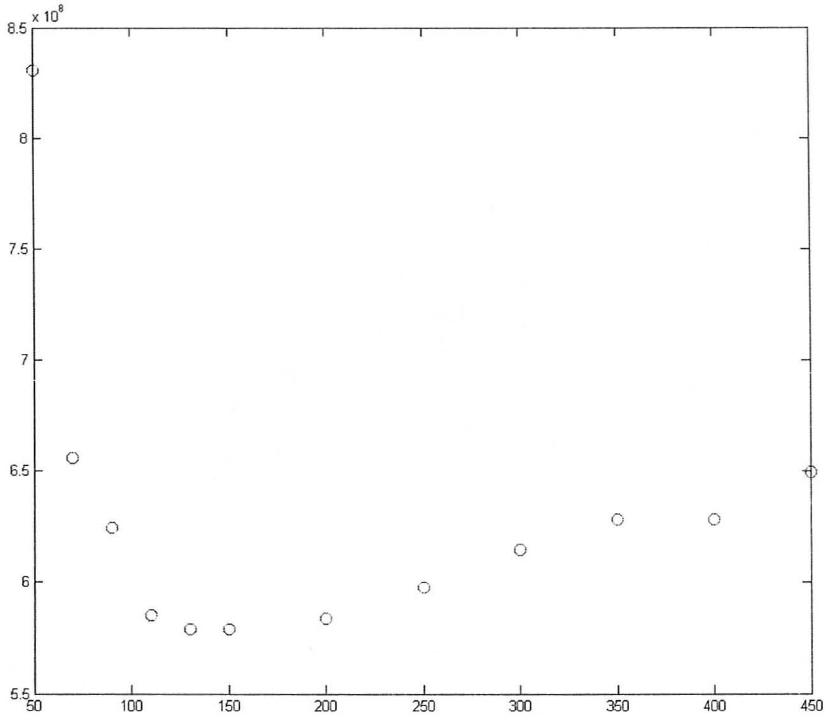


Figure 4.18: Cross Validation Results for MWK

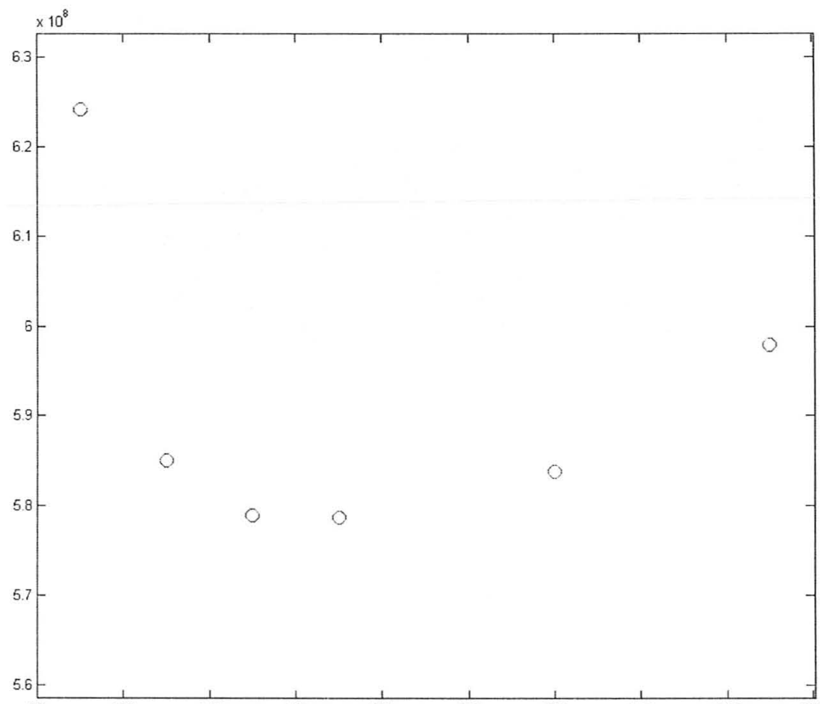


Figure 4.19: Cross Validation Results for MWR at A Closer Look

Recalling the definition of MWR and GWR, they originate from the same notion of modeling spatially varying relationship facilitated by conducting local regression for each point (Fotheringham et al. 1996). The special element of GWR that distinguishes it from MWR is the use of a distance weighting scheme. But how effective this weighting scheme is or to what extent that GWR will differ from MWR still remains as question. Given the main interest of this study on model performance, such inquiry is not the highlight of this study and won't be investigated in depth. However, by taking a glance at their cross-validation results, some relevant and interesting findings are disclosed.

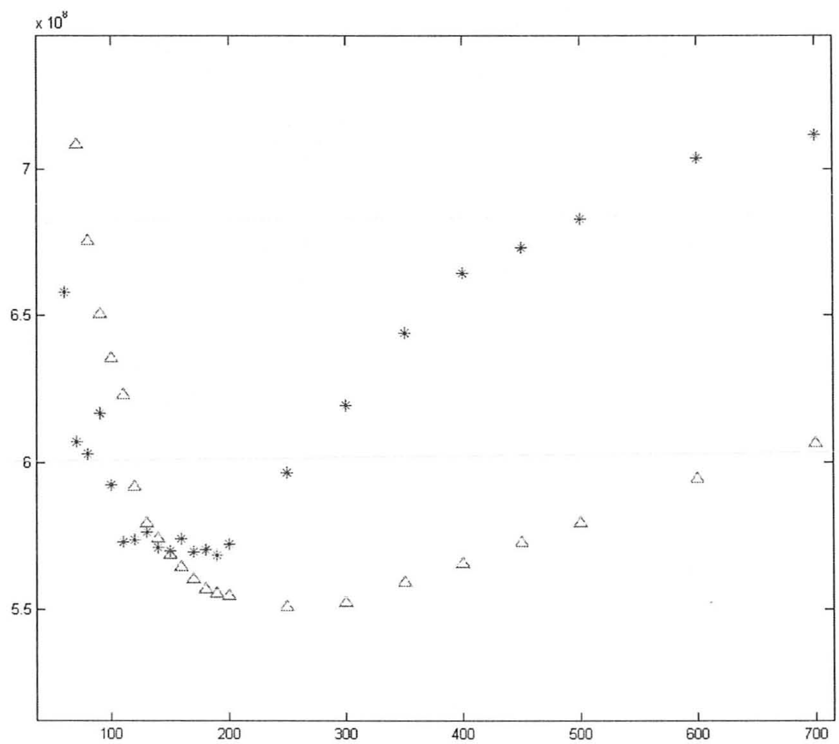


Figure 4.20: Cross Validation Results for MWR and GWR

First, their figures display similar trend that with the increase of window size, CV decreases sharply at the beginning until it reaches the lower bound, then it stays relatively stable for a certain range, after which it keeps going up without reaching a sill. However, as shown in Figure 4.20, MWR goes up more rapidly than GWR which implies that by employing the distance weighting scheme, GWR has a relatively steady performance and is less sensitive to the choice of moving window size. With respect to the contribution of distance weighing scheme, according to Figure 4.20, beyond the bandwidth of 200 nearest neighbors, a substantial and rapidly increasing gap is observed between MWR and GWR from which the efficacy of weighting scheme can be conferred. However, if the comparison is made both at their lowest points (minimized CV score at the optimal window size), the difference is not substantive any more which demonstrates the robustness of MWR when an optimal window size is used. In addition, based on the same graph, it is apparent that the optimum window for MWR ranges from 100 to 200 nearest neighbors, which is quite different from GWR whose best performance is achieved when 200 to 300 nearest neighbors are used. Therefore, researcher should be cautious towards using the identical window size for both models, especially when the interest is on assessing their performance.

As indicated by its name, moving window kriging (MWK) incorporates a variance-covariance matrix of residuals and therefore adds local component to the trend obtained from moving window regression. Bearing this in mind, it is easy to understand that besides modeling spatial heterogeneity as MWR does, MWK also makes use of residual spatial autocorrelation. Therefore, comparing the cross-validation results of MWR and MWK can help detecting the isolated and combined effects of spatial autocorrelation and spatial heterogeneity upon model performance.

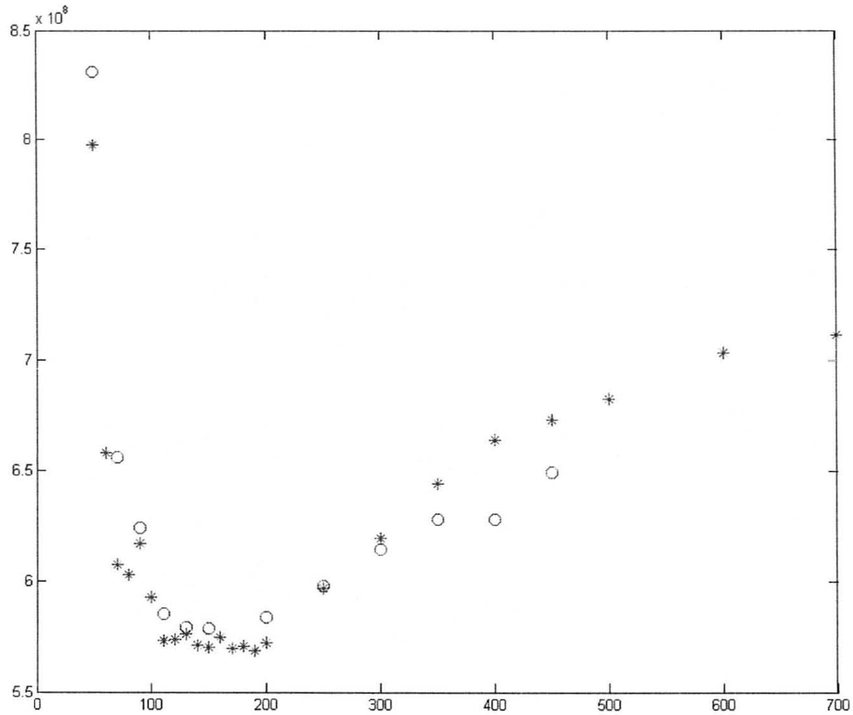


Figure 4.21: Cross Validation Results for MWR and MWK

It is observed in Figure 4.21 that MWK shares the same tendency as MWR, or to be more detailed, they are very much analogous in both the direction of trend and the magnitude of CV. Naturally, one would anticipate that MWK will surpass MWR since spatial heterogeneity and spatial dependency are both conceptualized into its model specification. However, what is shown here indicates that it is not necessarily the case. In fact, as will be seen in section 4.4.4, the incorporation of spatial autocorrelation after local regression sometimes could bring the model performance downward tremendously.

In the interest of testing the robustness of local models under the limitation of lacking neighborhood attributes, i.e. MeanIncome and PctImm for this specific case, Figure 4.22 adds cross validation results for OLS1 (traditional hedonic model using the same set of variables as MWR, GWR and MWK) and OLS2 (traditional hedonic model using additional neighborhood characteristics). The line at the top stands for the cross validation score of OLS1, whereas the dash line describes the cross validation score of OLS2. As manifested in Figure 4.22, the spatial hedonic functions including MWR, GWR and MWK perform much better than OLS1 and OLS2. Recalling the initial regression result in which neighborhood attributes explains 41.71 % variations occurred in housing market while the whole set of determinants with inclusion of other structural and accessibility attributes only accounts for 57.5% of total variations, it is remarkable that MWR, GWR and MWK make great progress given limited information.

More importantly, this experiment can help researchers alleviate the dilemma that choice must be made between the use of neighborhood attributes and the use of moving window approach since their properties determined they are mutually exclusive. As shown in Figure 4.22, the adoption of moving window approach at the expense of sacrificing neighborhood attributes achieved better prediction performance than the opposite action. Such inference will be confirmed in later sections through comparative analysis among various hedonic price functions.

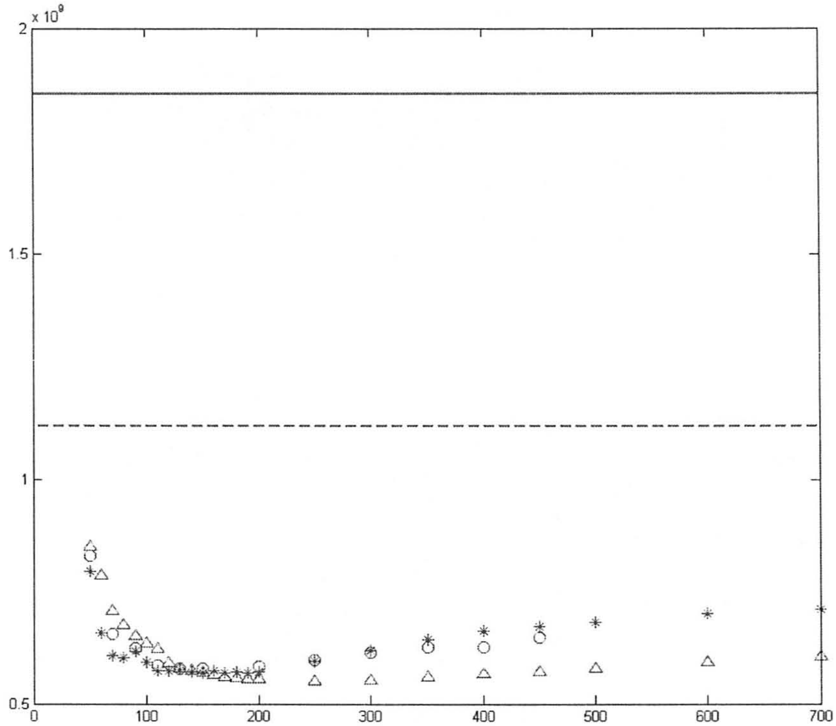


Figure 4.22: Cross Validation Results for All Hedonic Models

For the sake of searching optimum window size for MWR, GWR and MWK, which is essential to out-of-sample estimation, cross validation is applied in this section. This experiment is more than an optimization process. It also helps recover the predictive power of different spatial hedonic functions and disclose the individual effects of spatial dependency and spatial heterogeneity. These issues will be further explored in section 4.4.4 and conclusions will be drawn in section 4.5.

Before moving into the next section, attention is called onto Figure 4.4 in section 4.2.2 which plots the values of Moran's I to the size of neighborhood defined by the number of nearest neighbors. As mentioned in section 4.2.2, another impetus of diagnosing spatial autocorrelation at various neighborhood levels is to seek out potential relationship between degree of spatial autocorrelation and optimal window size. Unfortunately, Figure 4.4 and cross validation graphs display quite different patterns and no connection can be inferred from it. Therefore the determination of optimum window size still remains reliant on computationally expensive cross validation process.

4.4.3 Sample Thinning for Universal Kriging

As illustrated before, Universal Kriging is a global modeling technique which uses all data points within the study area to generate a single variance-covariance structure. Therefore, unlike MWR, GWR and MWK, it doesn't require efforts towards determining an optimal window size; however, it is afflicted by another problem associated with matrix manipulation. Recalling the computation of variance-covariance matrix in Universal Kriging, it involves both the inverse and determinants of a $N*N$ matrix where N is the sample size. Therefore, for out-of-sample estimation purpose, applying Universal Kriging means manipulating a 30,145 by 30,145 matrix. Hence, the computational burden of the estimation procedure is extremely large which may be beyond computer memory limitation. For example, the spatial interpolation function in Splus is only capable of manipulating matrix smaller than $7000*7000$; in Matlab, which is the software employed in this study for most estimations, inverting a 3500 by 3500 matrix is already a time consuming process. As well recognized the computational burden increases exponentially with the size of matrix, it is not feasible to use the whole estimation sample consisting 30,145 data points to predict the out-of-sample observations. Thus, constrained by the memory limitation, a sample containing 3500 points is randomly selected to represent the estimation sample and is accordingly used to predict out-of-sample observations.

The justification behind such practice is rooted in the ability of random sample to retain the shape and magnitude of empirical variogram and consequently generate almost identical variance-covariance structure as the whole population does. Such hypothesis is supported by the experiment conducted in section 4.3 which clearly shows that even the sample size is decreasing speedily from 30,145 to 3,500, the general trend and typical features of the empirical variogram are still withheld. This notion is confirmed again here by conducting similar experiment upon residuals after removing the trend in Universal Kriging[15]. Figure 4.23 is the empirical variogram computed using the whole estimation sample, while Figure 4.24 plots semivariogram obtained using 3,500 randomly selected observations. Two graphs are quite similar in terms of the magnitude and the tendency of variogram estimation. Therefore, it is reasonable to anticipate that the spatial dependency structure revealed by different sized samples would be much alike and the prediction accuracy produced by them would be more or less the same.

[15] The trend is composed of structural and locational attributes including Area, Front, Age, Age^2, SaleDate, DistTransit.

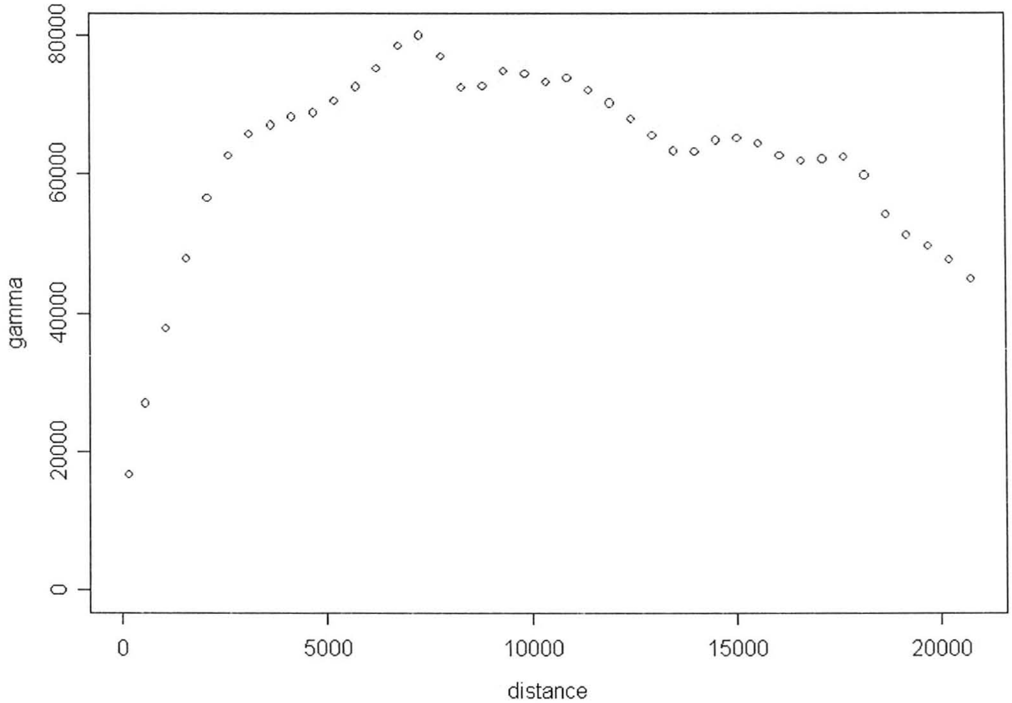


Figure 4.23: Empirical Variogram Using All in-sample Observations

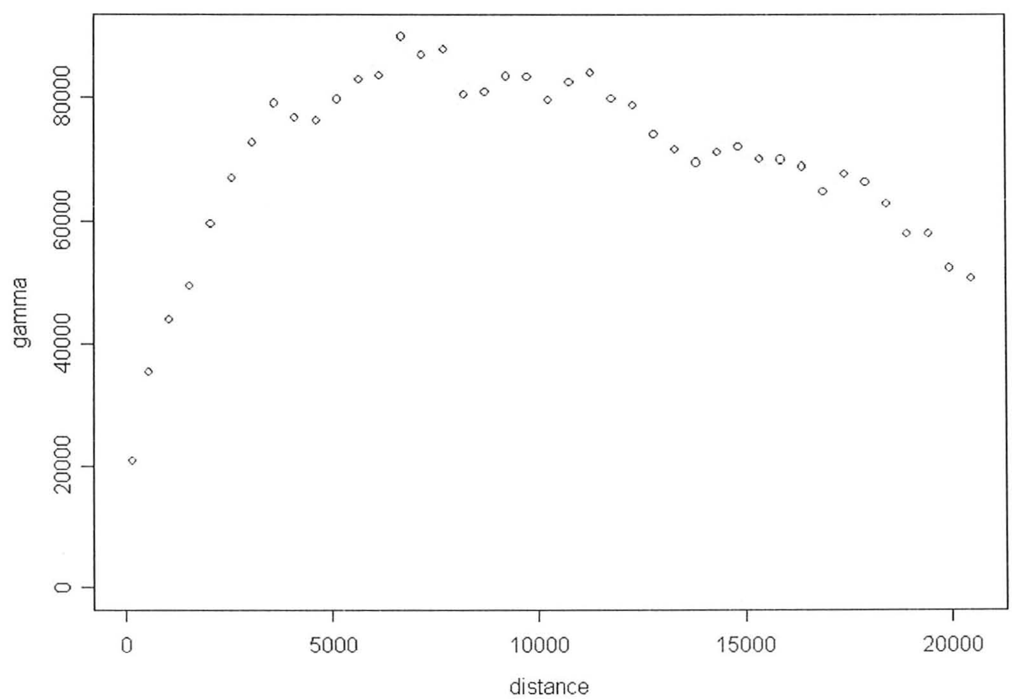


Figure 4.24: Empirical Variogram Using 3500 Randomly Selected Observations

To further examine the impact of sample size on the prediction accuracy of out-of-sample estimation, samples ranging from 500 to 3500 are tried. The prediction results are displayed in Table 4.12. The reported statistics like Pseudo R^2 and Root MSE clearly indicate that even sample size varies, their modeling behavior, especially the predictive performance is quite similar. It is noted that the coefficients computed under different samples are dissimilar, but this dissimilarity tends to vanish at large sample sizes.

Table 4.12: Sample Thinning Results for Universal Kriging

Number of Obs	Coefficient estimates							Pseudo R^2 [16]	Root MSE[17]
	CONST	AREA	FRONT	AGE	AGE^2	SALE DATE	DIST TRANSIT		
500	295.22	3.92	3.32	-114.93	9.77	2.90	-35.14	0.6011	185.31
1000	317.84	0.80	4.56	-84.26	6.35	2.95	-18.57	0.5819	178.75
1500	227.66	2.12	5.06	-80.07	5.81	3.11	-19.61	0.6148	171.05
2500	219.73	1.22	5.71	-86.91	6.93	2.72	-5.74	0.6181	171.76
3500	216.48	3.59	2.14	-74.79	5.74	3.21	-4.24	0.6314	169.29

Figure 4.25 provides the scatter plots between predicted prices and actual transaction prices which again confirms the notion that the varying sizes of estimation sample do not have deterministic influence upon the prediction accuracy of out-of-sample observations. Therefore, it is believed that the predictions obtained using 3500 random sample approximate well the results that would be theoretically produced by the whole population of 30,145 observations. Therefore, Universal Kriging predictions using 3500 randomly selected samples are retained in the following section for model comparison.

[16] Pseudo R^2 is the squared correlation between predicted and observed prices in prediction sample.
[17] Root MSE is the rooted mean squared error. The definition is used in other parts of the article.

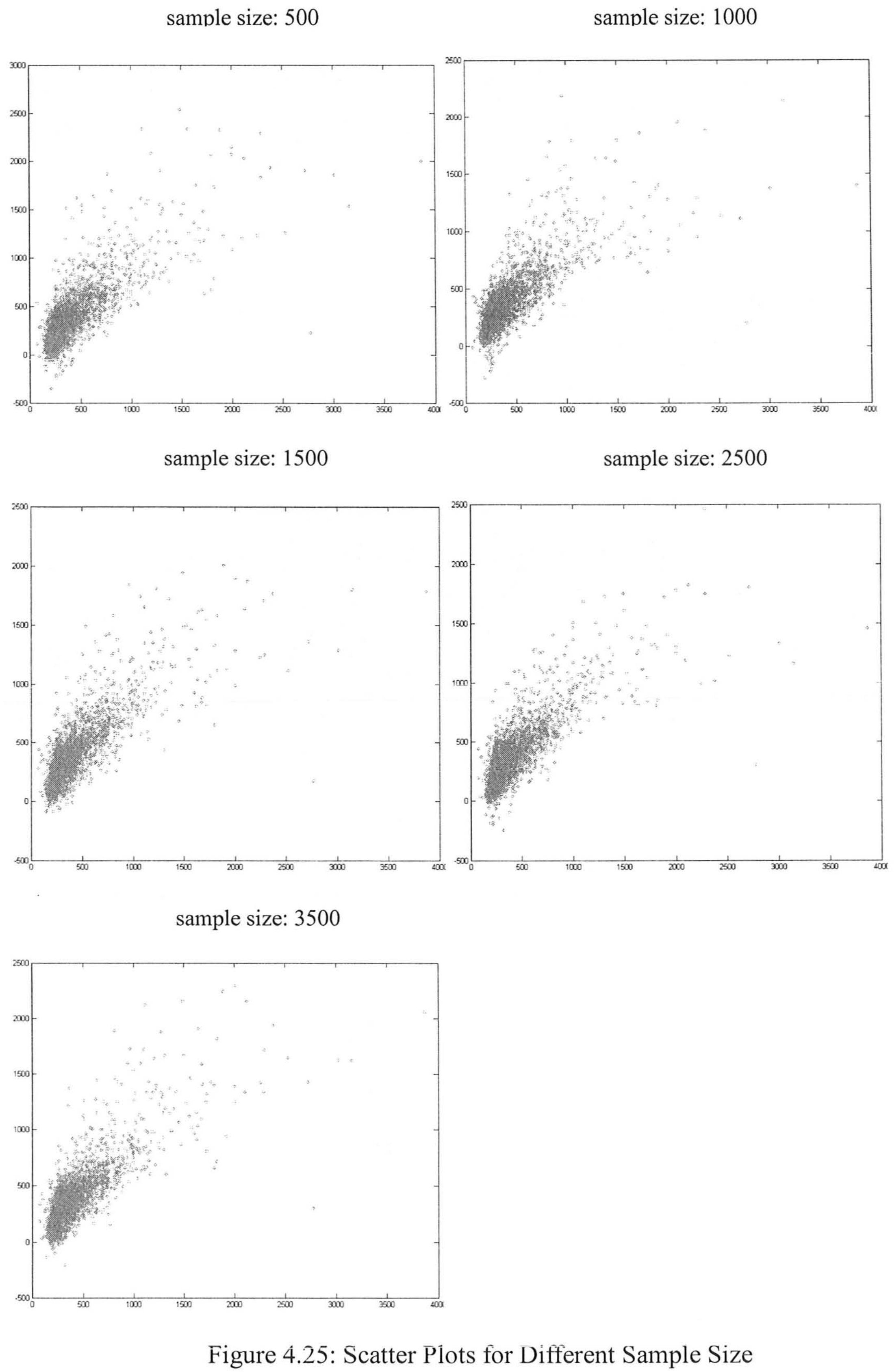


Figure 4.25: Scatter Plots for Different Sample Size

4.4.4 Comparative Analysis of Predictive Power of Various Hedonic Price Functions

As a quantitative measure of predictive performance for different hedonic price functions, a number of indexes describing the prediction error of out-of-sample observations------ the difference between predicted price and actual transaction price are computed. Table 4.13 presents these statistics which enable researchers to quantify and evaluate various models' performance.

Table 4.13: Summary Statistics for Prediction Errors

Model	Mean Absolute	Root MSE	Minimum	Maximum	Standard Deviation
OLS1*	139.37	227.58	-1767.26	2626.91	227.59
OLS2**	102.22	174.14	-1171.27	2464.75	174.15
MWR	66.08	143.52	-1845.57	2525.50	143.47
GWR	64.03	139.97	-1630.32	2526.55	139.91
Kriging1*	109.76	169.29	-1079.57	2468.58	167.98
Kriging2**	99.3882	169.17	-1015.86	2595.54	169.13
MWK	199.49	376.59	-3475.57	3006.38	376.38

* *without neighborhood attributes: MeanIncome, PctImm*

** *with neighborhood attributes: MeanIncome, PctImm*

According to Table 4.13, the best predictive power is achieved by GWR in terms of the mean absolute error and the root mean squared error; MWR also attains agreeable results with key statistics only marginally different from GWR; Kriging2 using extra neighborhood attributes performs slightly better than Kriging1, however, the difference is not quite noticeable with similar reported statistics; as expected, the performance of

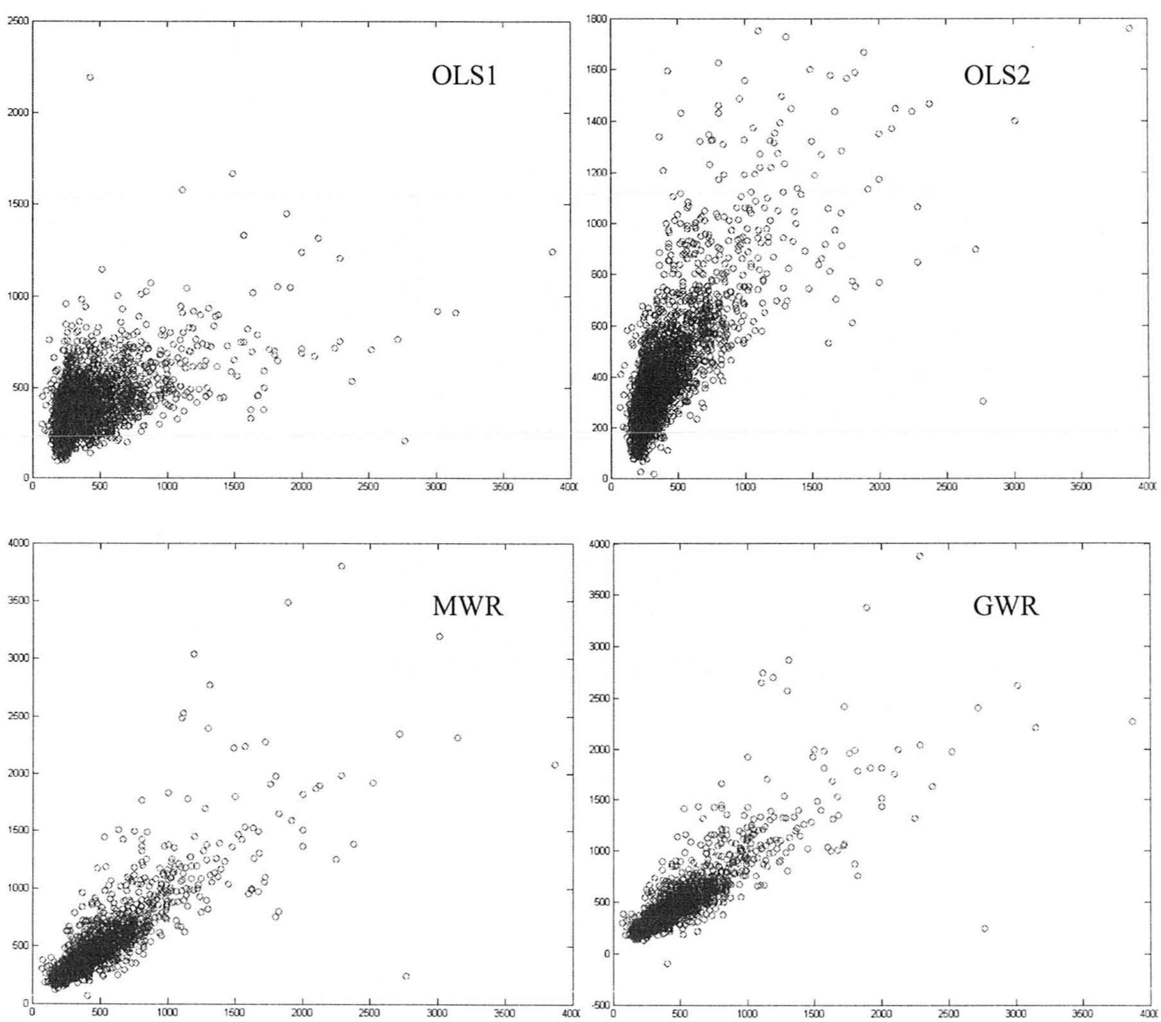
OLS1 is not comparable to the performance of OLS2 due to the omission of neighborhood variables from model calibration; surprisingly, MWK has the lowest predictive power, which is even worse than the traditional hedonic price function OLS1.

Table 4.14 represents some other statistics where the column termed 'correlation' is the Pearson's correlation coefficients between predicted and observed values of prediction sample and the column entitled '% better than OLS' is a percentage of the number of times that the model in question predicts better than traditional hedonic price function. The results are visually represented by plotting the predicted prices against the observed sales prices as displayed in Figure 4.26. The table and figures again confirm the initial impression gained from Table 4.13.

Moreover, scatter plots of predicted values to actual transaction prices give more insight of the prediction procedure. As seen in scatter plots for GWR, Universal Kriging and MWK, negative predictive values are observed which is unacceptable for housing price estimation. It indicates that when distance weighting scheme or residual spatial dependence structure can improve the overall prediction accuracy, it may not be applicable to some individual observations and consequently distorts their estimations. At the extreme situation, negative predictions appear. This should arouse cautions of applying these spatial techniques into appraisal or assessment of housing units in business practice.

Table 4.14: Other Statistics for Model Prediction Performance

Case	Correlation	Pseudo R^2	% Better than OLS1	% Better than OLS2
OLS1	0.5339	0.2850		33.41%
OLS2	0.7648	0.5849	66.59%	
MWR	0.8594	0.7386	77.66%	70.50%
GWR	0.8669	0.7515	78.29%	71.25%
Kriging1	0.7946	0.6314	58.17%	45.98%
Kriging2	0.7779	0.6051	65.33%	49.66%
MWK	0.4058	0.1647	47.51%	37.47%



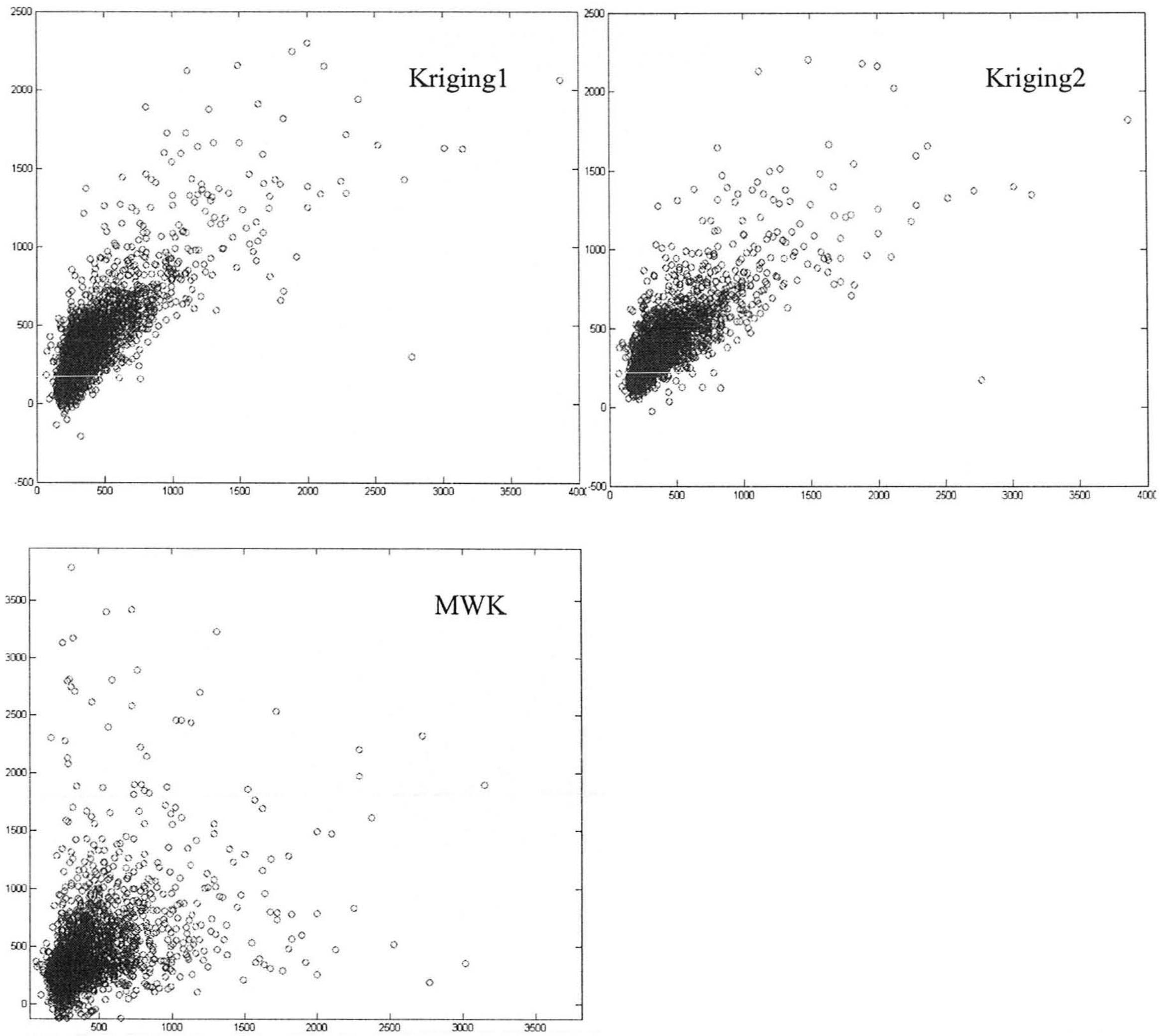


Figure 4.26: Scatter Plot of Observed Price VS Predicted Price for Prediction Sample

As discussed at the beginning of section 4.4, MWR and GWR use local regression techniques to integrate spatial heterogeneity into traditional hedonic price function; Universal Kriging takes spatial dependence into account via the construction of residual variance-covariance matrix; MWK, by performing regression within moving window and adjusting the estimation with predicted error term, incorporates both spatial heterogeneity and spatial autocorrelation into model specification. The relationship among these models can be illustrated through Figure 4.27.

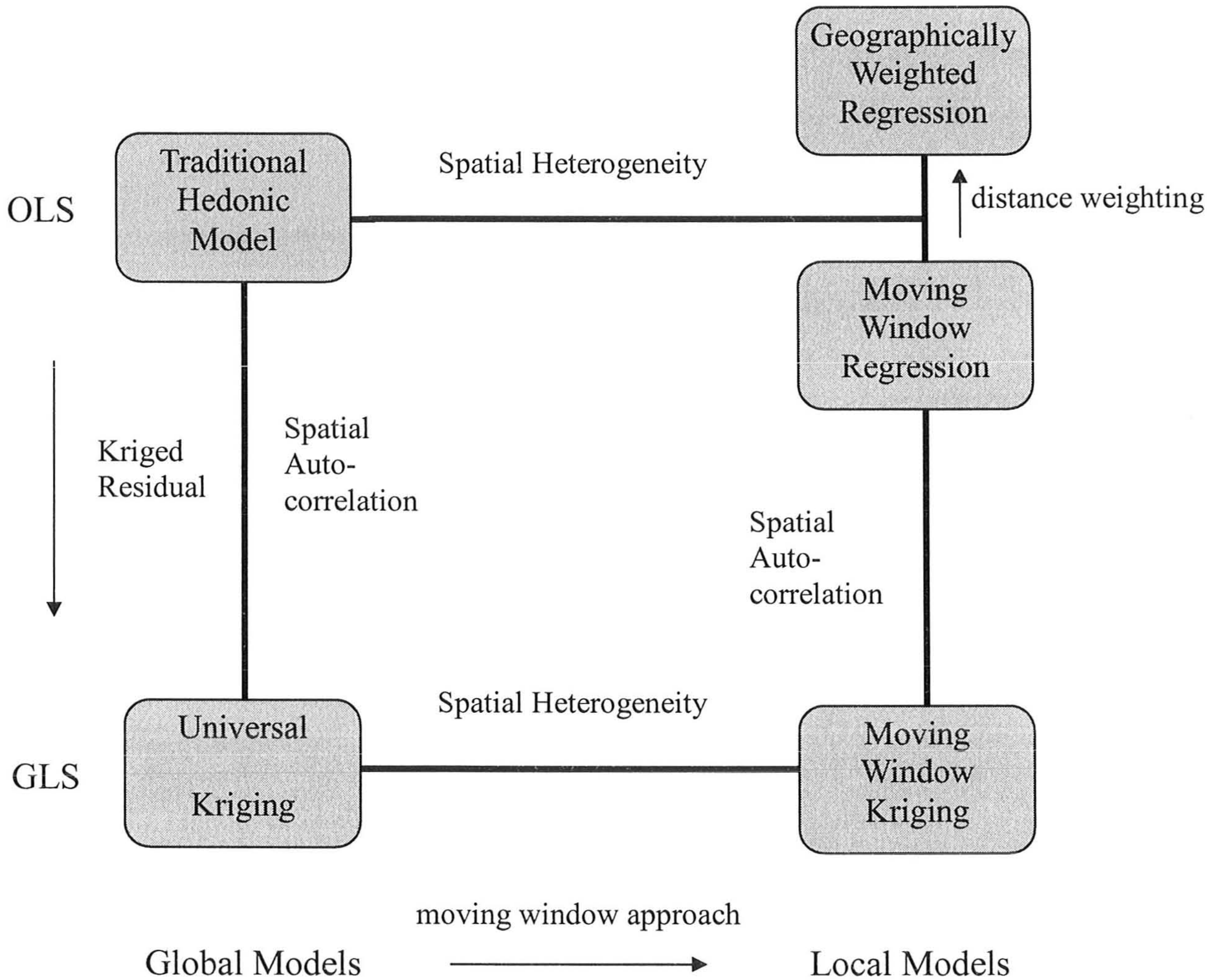


Figure 4.27: Model Relationship Map

As seen in Figure 4.27, the use of moving window approach can divide models into two groups------global models and local models where traditional hedonic model and Universal Kriging belong to global models while GWR, MWR and MWK are regarded as local modeling techniques; moving window approach creates a set of customized coefficient estimates for each prediction site and therefore realizes modeling spatial heterogeneous process. Also seen in the above figure, another way to categorize the models is the statistical approach used for coefficient estimates, specifically, the use of ordinary least squares (OLS) or generalized least squares (GLS). According to this classification, traditional hedonic model, MWR and GWR are put into the same category, whereas Universal Kriging and MWK are grouped into another class. By adding a local component, i.e. predicted error term through GLS procedure, Universal Kriging adds spatial dependence into traditional hedonic model, while MWK incorporates the same effect into MWR. Based on the above understanding, the contributions of two types of spatial effects ------ spatial dependency and spatial heterogeneity can be detected and measured respectively by comparing traditional hedonic price function OLS1 and other spatial hedonic models since they are using the same set of explanatory variables.

Firstly, the improvement in model performance made by Kriging1 over OLS1 can be attributed to the presence of predicted error terms at a global level which captures remaining spatial dependence in traditional hedonic residuals. As discussed earlier, traditional hedonic price function used in this study, especially OLS1, may suffer from misspecification due to limited information regarding housing structural attributes and exclusion of neighborhood characteristics. As a result, the error is not randomly distributed over space; instead, one would observe clusters of positive or negative residuals. Such speculation is already verified in section 4.2 by conducting diagnostic tests upon residuals from OLS1 before splitting the population into two samples. It is again confirmed by visualizing the spatial distribution of OLS1 prediction errors. In Figure 4.28, the circles present the predictions within 25% of the actual transaction prices, the upward triangles indicate negative errors (where the price is overestimated) over 25%, and the downward triangle are positive errors (where the price is underestimated) in excess of 25%.

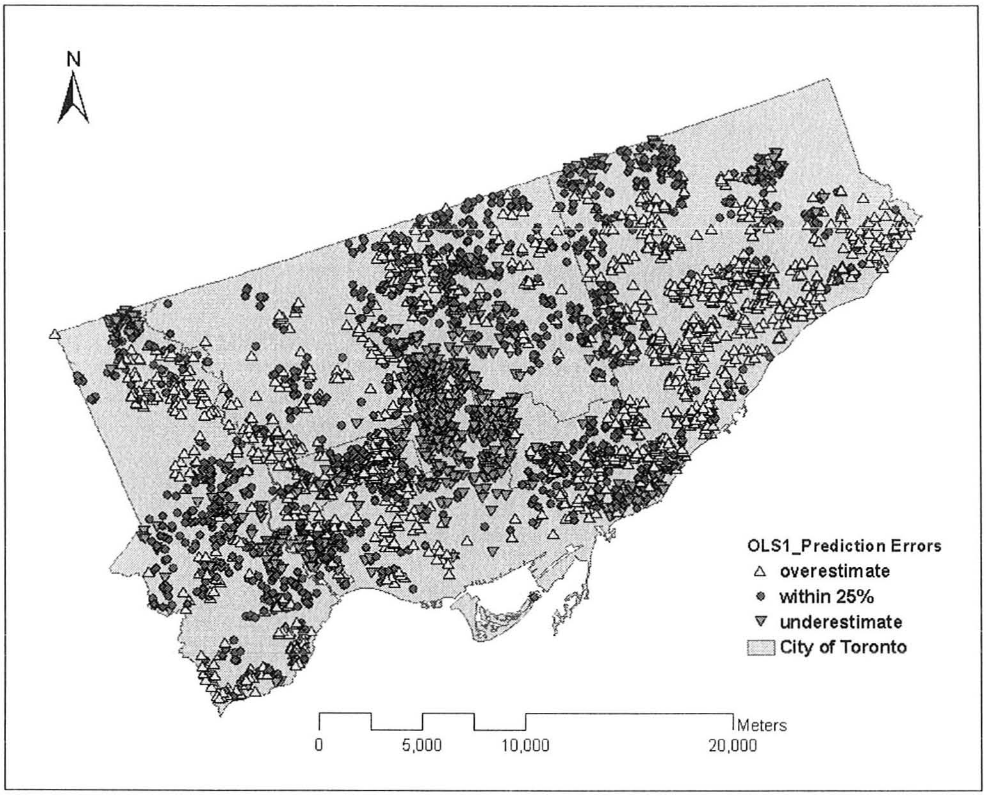


Figure 4.28: Spatial Distribution of Prediction Errors for OLS1

If the error is completely randomly distribute over space, the best predicted value of observations would rely completely upon the characteristics of property (Dubin et al 1999), thus traditional hedonic model would generate best linear unbiased prediction by multiple the obtained coefficients to the characteristic vectors $X\beta$. However, after observing patent spatial cluster of residuals of the same sign, the traditional way is not appropriate since when underestimation or overestimation prevails in space, it seems appealing to correct for this tendency by assigning weights to nearby observed residuals (Dubin et al 1999). This explains the improvement gained by Kriging1 which captures the effect of omitted neighborhood attributes through adding predicted error term into the normal trend.

Secondly, as seen in Figure 4.29 and Figure 4.30 MWR and GWR achieve best prediction results in which around 85% prediction errors are within an acceptable range of 25% of the original sale prices while in Kriging1 the percentage is around 52 %. Thus it can be derived that for this particular case, spatial heterogeneity captured by moving window approach, has a larger impact upon model performance. Also compatible with cross validation results, the marginal improvement gained through distance weighting scheme is small as indicated by similar statistic values for MWR and GWR. However, whether the extra accuracy obtained by GWR validates the efforts of incorporating weighting scheme is subject to individuals' judgments.

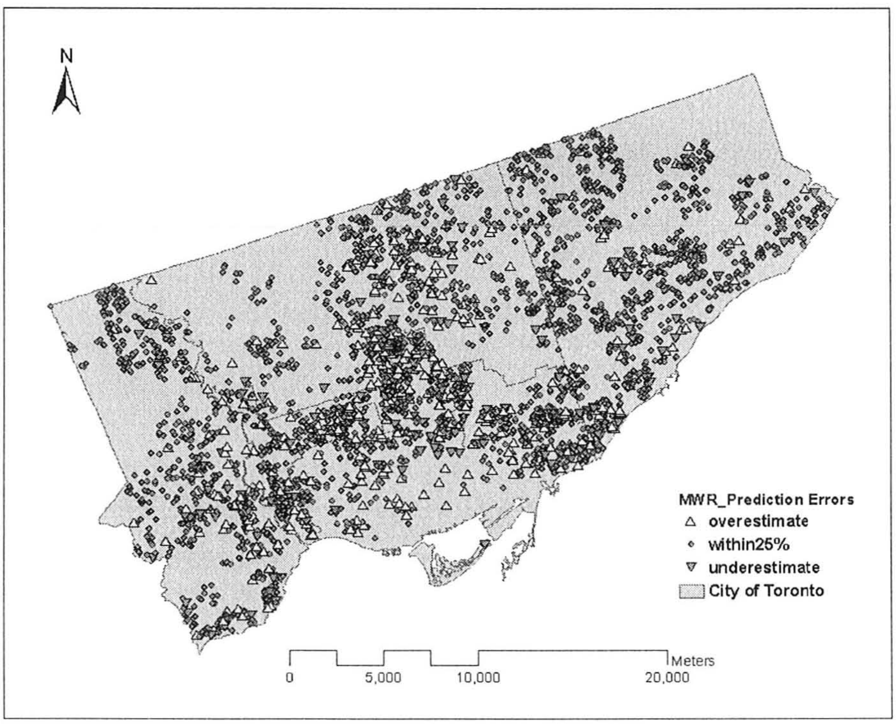


Figure 4.29: Spatial Distribution of Prediction Errors for MWR

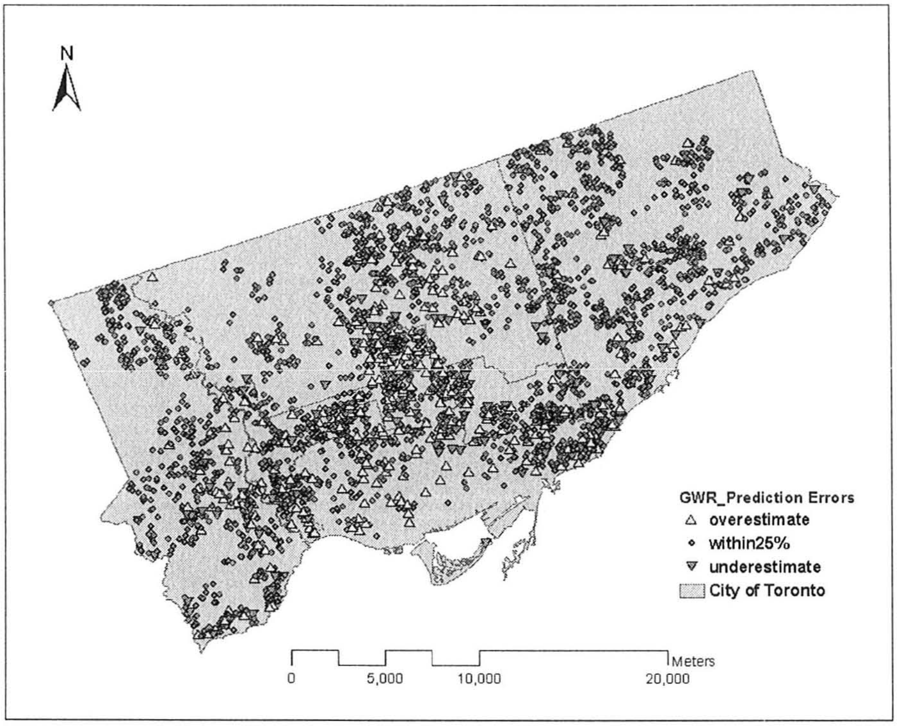


Figure 4.30: Spatial Distribution of Prediction Errors for GWR

The above comparisons between Kriging1 and OLS1, MWR, GWR and OLS1 can help distinguish the individual effect of spatial autocorrelation and spatial heterogeneity on hedonic price estimation. Subsequently, comparative analysis between MWK and MWR can help identify the effect of spatial autocorrelation after spatial heterogeneity is built into modeling process. Also comparing MWK and Kriging1 can shed lights on the effect of spatial heterogeneity in case of embedded spatial autocorrelation in estimation procedure.

As noticed in section 4.2.2, there in no distinguishable difference between MWK and MWR in terms of cross validation results. In the out-of-sample prediction, on the other hand, it is observed that the worst predictive performance is made by MWK. However, it is not that surprising by looking at Figure 4.29 and Figure 4.31; in the latter figure, light dots represent negative residuals, whereas dark dots indicate positive residuals. Both figures clearly show that unlike traditional hedonic price function, errors generated by local regressions exhibit random pattern over space. As illustrated earlier, under such circumstance, the prediction of out-of-sample observations would rely completely on property attributes which means ordinary least squares approach is able to produce best linear unbiased prediction.

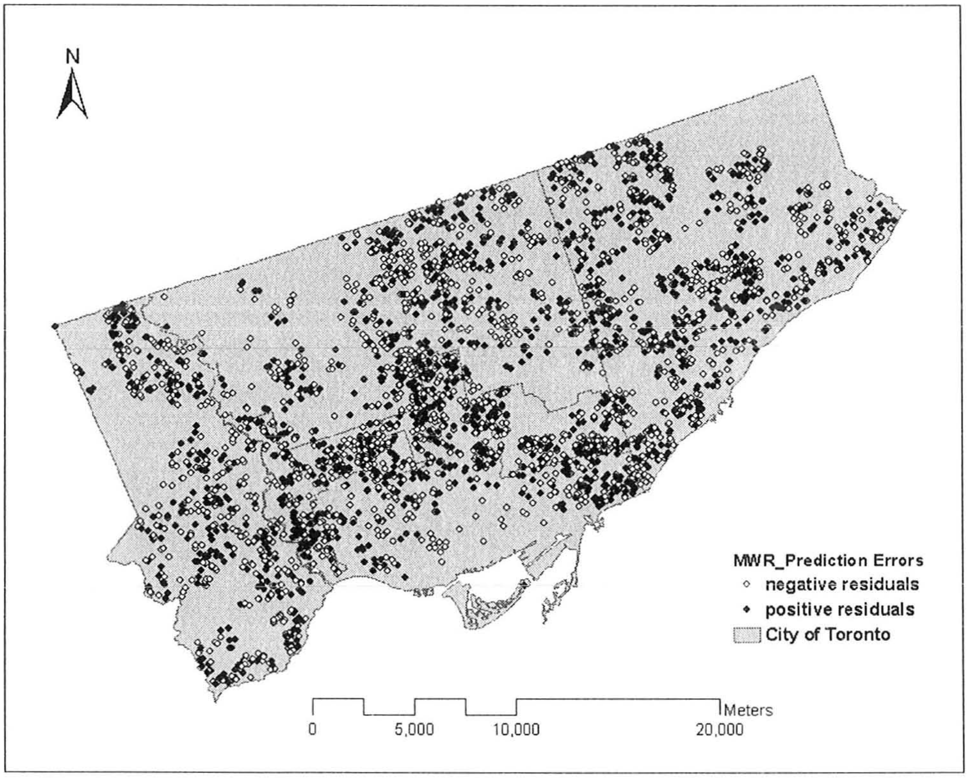


Figure 4.31: Spatial Distribution of Residuals of Same Sign for MWR

As also pointed out by Base and Thibodeau (1998) in their studies on Dallas County, for submarkets where the hedonic residuals are spatially uncorrelated, kriging will either (1) have no influence on prediction accuracy or (2) reduce prediction accuracy. Since this study performs local kriging procedure for 3,349 prediction points, it is not an easy task to conduct formal diagnostic test of spatial autocorrelation as Base and Thibodeau did by computing semivariogram for 9 zones. But based on visual inspection, there are no obvious clusters of residuals after MWR, which implies that within each moving window------a relatively homogeneous area, inclusion of structural and accessibility attributes provides adequate control for spatial autcorrelation. Therefore, what is observed confirms the findings of Base and Thibodeau that modeling spatial autocorrelation with kriged residuals when it is absent would add noise into process, cloud the relationship between housing prices and its influential factors, and consequently produce unsatisfactory estimations.

Assessments of Universal Kriging and MWK prediction results suggest that for this particular case, the local variance-covariance structure used by MWK does not improve the prediction accuracy as expected; on the opposite, lower predictive power of MWK is observed in model evaluation. As indicated by Walter et al. (2001), Moving Window Kriging can greatly increase the precision of estimation as seen by minimized kriging variance, however, with respect to predictive accuracy, no distinction can be draw between the two approaches. It should be noted that the drifts employed in their studies are different ------ either using geographic coordinates to build a quadratic or cubic trend or using ordinary kriging without any explanatory variables. When Ordinary Kriging and its local counterpart are applied in this dataset, similar conclusion is obtained: moving window kriging does reduce the kriging variance associated with the prediction, but no discernable improvement is gained in terms of prediction accuracy. Therefore, as suggested by this specific study, when a relatively complex drift is included into the kriging procedure, residual information is not very rich. Accordingly, a single semivariogram would be sufficient to capture the pattern and exploit this part of information.

In addition, comparison between OLS2 and spatial hedonic price functions also reveals other interesting findings. As observed in Table 4.13 and Table 4.14, unlike OLS1 and Kriging1, Kriging2 does not improve the performance of OLS2 by taking use of residual spatial dependency; in fact, the opposite effect might occur since some key statistics reported for Kriging2 are not as good as OLS2. This implies that the adjustment made by Universal Kriging by adding prediction error term to the trend established in OLS2 is not helpful. This answers the questions posed in section 4.2 that for this particular dataset, adding neighborhood attributes provides sufficient control for spatial dependency, and consequently the enterprise of modeling spatially autocorrelated error terms is not necessarily beneficial.

It is also noticed that while adding neighborhood attributes into traditional hedonic model can improve prediction accuracy substantively, no discernable difference is observed between Kriging1 and Kriging2. The explanation is that the omission of influential determinants from model specification can be compensated by modeling spatial autocorrelation of residuals. Dubin (1992) argued that since there are severe measurement problems inherent in neighborhood and accessibility variables, she preferred to exclude these variables from regression and model spatial autocorrelation of residuals coming from omission of these important determinants. Her justification lies in that even omitting relevant variables from model specification will cause biased and inconsistent estimates; however, incorrectly measured explanatory variables also cause biased and inconsistent estimates. When facing such a dilemma, Dubin used Kriging to take the spatial relationship into account explicitly rather than suffering from unclear measurement errors. The results uncovered in this study further support this modeling approach for the reason that credited to the capacity of modeling residual spatial dependency, Kriging is not largely affected by the omission of neighborhood attributes and will generate equally good results.

Lastly, consistent with cross validation results, MWR and GWR achieve more accurate prediction than OLS2 despite their lack of neighborhood characteristics. This solves the questions raised at the beginning of this chapter that whether neighborhood attributes are indispensable to hedonic housing price analysis and whether the inability of using this information will bring down the value of local models. As clearly shown in this study, it is not the issue for MWR and GWR for the reason that they perform regression within moving window, a relatively homogeneous area, in which neighborhood attributes don't vary enough to contribute to housing prices. As well, the minor difference between Kriging1 and Kriging2 also points out that omission of neighborhood information from model specification can be compensated by modeling residual spatial dependency. Therefore, it can be concluded that although neighborhood characteristics are the most significant determinants in price determination process and explain the largest proportion of variation in housing market, they are not indispensable to hedonic housing studies as long as spatial modeling techniques are applied.

4.4.5 Another Combination of Drift and Variogram

"How closely the variogram of the estimate residuals corresponds to the true but unknown variogram depends upon the appropriateness of the function selected to represent the drift, the function selected to represent the variogram, and the size of the chosen neighborhood." (Lam 1983)

In previous sections, the selection of the size of neighborhood was justified through the cross validation process; also as mentioned earlier, the spherical function is selected based on extensive exploratory work of conducting semivariogram at various scales and location. After that, if the validity of MWK on this dataset is still suspect, the only possibility is that an inappropriate drift is defined for this approach which results in disagreeable prediction accuracy.

Bearing this in mind, recall the comparative analysis between MWR and MWK in the preceding section: after including structural and accessibility attributes, MWR controls for most spatial autocorrelation at moving window level which leaves little valuable information in residuals and therefore invalidates the efforts of MWK to estimate the predicted error term. In the same way, with a relatively complex drift composed of six determinants, the traditional hedonic model captures the primary features of the modeling process; as a consequence, it allows for a single variance-covariance structure to describe the small variations left in residuals. In this situation, the locational dependent variance-covariance matrix generated by MWK is redundant and obscures the process.

Due to the above considerations, a simple drift formulated by fewer variables------ Area and DistTransit is applied in the second round of model evaluation as to leave more space for MWK to exert its potency of utilizing spatial autocorrelated errors. OLS2 and Kriging2 using extra neighborhood variables are also included in model comparison. Cross validation results are displayed in Figure 4.32 and Figure 4.33; values of the same set of statistics for the second round model comparison as the first round are reported in Table 4.15 and Table 4.16.

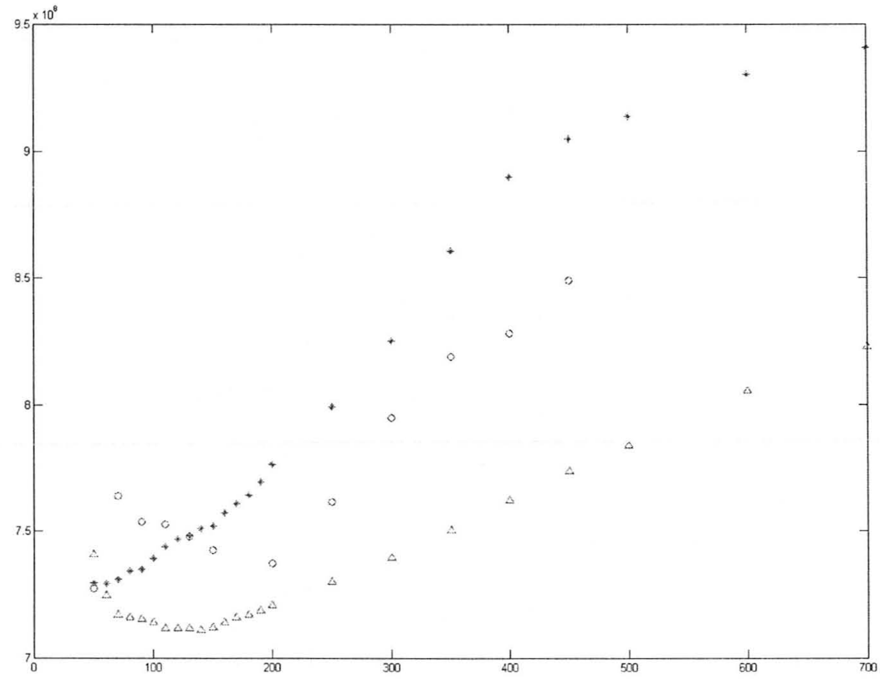


Figure 4.32: Cross Validation Results for MWR, GWR and MWK (2nd round)

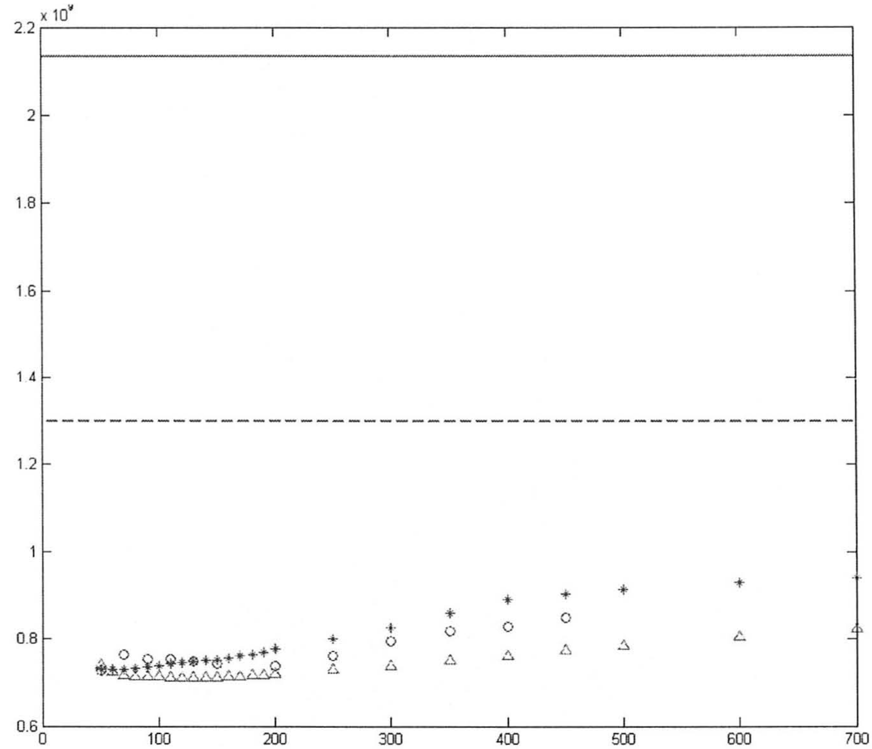


Figure 4.33: Cross Validation Results for All Models (2nd round)

Table 4.15: Summary Statistics of Prediction Errors for Second Round Comparison

Case	Mean Absolute	Root MSE	Minimum	Maximum	Standard Deviation
OLS1	150.0953	241.2678	-742.56	2835.93	241.24
OLS2	110.3593	186.1266	-1013.84	2451.56	186.15
MWR	77.6374	156.7824	-2385.62	2486.02	156.79
GWR	76.3025	154.2950	-2064.23	2487.15	154.31
Kriging1	106.3207	173.5379	-1290.82	2421.28	173.09
Kriging2	101.3898	176.2319	-1155.66	2549.85	176.25
MWK	146.5522	287.7564	-3938.73	2934.15	287.47

Table 4.16: Other Statistics for Second Round Model Comparison

Case	Correlation	Pseudo R^2	% Better than OLS1	% Better than OLS2
OLS1	0.4513	0.2037		31.74%
OLS2	0.7230	0.5227	68.26%	
MWR	0.8256	0.6816	75.72%	67.99%
GWR	0.8300	0.6889	76.53%	68.32%
Kriging1	0.7789	0.6067	64.23%	52.05%
Kriging2	0.7557	0.5710	71.42%	53.96%
MWK	0.5233	0.2738	58.67%	47.51%

The figures and tables confirm previous findings that GWR and MWR perform best, while Kriging follows right after them. Slightly different from the first round, MWK prediction performance improves in a minor fashion this time, with cross validation result is a little better than MWR and some key statistic for out-of-sample predictions are slightly better than OLS1.

This experiment demonstrates that MWK does improve the model performance of hedonic price function when considerable spatial autocorrelation exists among residuals. However, its performance appears to be inconsistent as revealed by its varying predictive power exhibited in cross validation and out-of-sample estimation. This weakness is also verified by the descriptive analysis of its prediction errors which has the highest standard deviation and broadest range among all candidate models as shown in Table 4.13 and Table 4.15.

As also been noticed, in correspondence with a relatively simple drift, rich information resides in residuals after traditional hedonic modeling which enables

Kriging2 to take advantage of spatially autocorrelated error term and improve the prediction accuracy of OLS2. However, excluding neighborhood attributes from OLS1 leaves more space for modeling residual spatial dependency, thus more noticeable improvement is observed between OLS1 and Kriging1.

4.5 Summary

In all, the diagnostic tests of spatial autocorrelation and spatial heterogeneity detect the existences of two spatial effects in the study process which necessitate the use of advanced spatial hedonic models. After experimenting with a number of sophisticated hedonic price functions and evaluating their respective predictive power, some interesting findings are made as follows.

Firstly, based on two rounds of model comparisons, GWR and MWR largely improve traditional hedonic model, which substantiates the validity of moving window as an effective approach to model a heterogeneous process. In addition to improved prediction accuracy, by performing regression within each moving window ------ a more homogeneous area, MWR and GWR also control for most spatial autocorrelation in spite of limited variables. Also, similar to findings made by Farber (2004), this research also suggests that the distance weighting scheme employed in GWR may not bring noticeable enhancement to model performance as long as the optimal window sizes are used for each approach.

Secondly, Universal Kriging which aims to account for residual spatial dependency not always improves model performance. As noticed in diagnostic test of spatial autocorrelation, Moran's I index computed from residuals after traditional hedonic modeling without neighborhood attributes is 0.4011, then it goes down to 0.1389 after including MeanIncome and PctImm into model specification. It indicates that a simple drift composed of limited structural and proximity variables leaves substantive amount of spatial autocorrelation in residuals, however, after control for neighborhood characteristics, it diminishes considerably. In accordance, while Kriging1 generates more accurate price predictions than OLS1, relatively little difference is observed between the performances of Kriging2 and OLS2.

Based on the above observations, it may be concluded that incorporation of spatial effects------spatial autocorrelation or spatial heterogeneity into traditional hedonic price function will both improve prediction accuracy if they are properly recognized by the model. However, when the individual effect of spatial autocorrelation and spatial heterogeneity are certain and evident, there is controversy regarding their combined outcome as revealed by this study. Unlike what would be expected for MWK, inclusion of both spatial autocorrelation and spatial heterogeneity into model specification doesn't contribute to prediction accuracy. On the contrary, the predictive power of MWK is weak, sometimes even worse than traditional hedonic model. As discussed earlier, the underlying reason lies in that MWK models spatial heterogeneity first by regressing housing prices on determinants within each moving window, which is the equivalent of MWR; as shown by Figure 4.29 and 4.31, there is no obvious spatial autocorrelation in residuals after local regression, therefore the second step taken by MWK, that is kriging the residuals, simply brings in noise and messes the process. Additionally, the model performance of MWK is not stable as pointed earlier.

Lastly, as shown in prior sections, even OLS2 uses extra neighborhood attributes which account for 41.71 % of total variations, MWR and GWR still perform much better than OLS2. The reason that MWR and GWR are robust to this condition is that when regression is performed at neighborhood level, neighborhood attributes do not vary enough to contribute to housing values. Therefore, inability of using neighborhood characteristics does not reduce the value of local regression modeling techniques. Similarly, Kriging1 behaves almost as well as OLS2 which provides empirical support to Dubin (1992)'s modeling approach of omitting neighborhood attributes and modeling spatial autocorrelation of residuals. The competency of spatial modeling techniques to absence of neighborhood characteristics reinforces their integrity since in the literature no consensus is achieved upon selection of variables that best proxy neighborhood quality or if there is any agreement, as discussed before, due to the difficulty of defining boundaries, severe measurement problems might exist.

Chapter 5 Conclusion and Suggestions

5.1 Conclusion

As detailed in chapter 2, traditional hedonic price functions make a series of unrealistic postulates about both residuals and parameters in modeling process. Due to locational effects and market segmentation which characterize urban housing markets, it is unlikely that these assumptions will be tenable in most circumstances. While the potential implications of violated assumptions are well understood on strictly theoretical grounds, how they actually affect the validity of results from traditional hedonic modeling, and how to adjust hedonic price function to correct for these violations still remain as questions in concrete situations. Thus, it is of great importance to conduct formal diagnostic tests to identify the existence of any particular assumption violation in order to determine the appropriate adjustment needed in model specification.

As discussed in chapter 4, when local variogram modeling indicates the existence of spatial heterogeneity, dramatic differences are observed in prediction accuracy between traditional hedonic model and local regression models MWR and GWR. Moreover, visual inspection and Moran's I test reveal strong spatial autocorrelation in traditional hedonic price residuals after accounting for structural and locational differentials. In correspondence with that, Kriging1 behaves much better than OLS1 in terms of prediction accuracy. On the other hand, diagnostic tests upon traditional hedonic price residuals using the second set of determinants imply the control of spatial autocorrelation after incorporation of neighborhood characteristics. In accordance, no discernable difference is detected between OLS2 and Kriging2. Thus it confirms previously reported studies regarding the importance of conducting spatial diagnostic test before any formal modeling procedure in hedonic price analysis.

As indicated in chapter 1, the main purpose of this study was to identify the individual and combined impact of two types of spatial effects upon hedonic price estimation by comparative analysis of various spatial hedonic specifications. After two round assessments of model prediction accuracy, two questions are addressed------ (1) whether housing characteristics account for all the variations occurred in housing market and leave randomly distributed error terms; and (2) whether inclusion of spatial effects, i.e. spatial dependency and spatial heterogeneity will uniformly improve hedonic price prediction accuracy.

For the first question, diagnostic tests and comparative analysis between OLS2 and Kriging2 all suggest that traditional consideration of modeling structural, locational and neighborhood characteristics provides sufficient control for spatial autocorrelation. However, this conclusion may be debatable for the reason that on the one hand constrained by data limitations, little information is available for housing structures, which discounts the prediction performance of OLS2 and on the other hand, owing to the computation burden of implementing Universal Kriging to this large dataset, a random sample is created to represent the in-sample observations. Although the sample thinning approach has been justified in section 4.4.3, it might still deduct the credit of Kriging in a minor way. Since traditional hedonic price function and Universal Kriging both suffer from misspecification, it can not guarantee that indiscernible difference between model performance of OLS2 and Kriging2 is ascribed to the elimination of spatial autocorrelation after control for structural, proximity and neighborhood characteristics.

Despite the uncertainty associated with the former argument, it is well recognized that inclusion of mere structural and locational attributes generates spatially clustered error terms which is verified by diagnostic tests as well as the remarkable improvement of predication accuracy gained by Kriging1 over OLS1. Interestingly, when substantial spatial autocorrelation exists in traditional hedonic price residuals, MWR, controlling for limited structural attributes and proximity to transportation network, has eliminated spatial autocorrelation significantly. Even though constrained by sample size no formal diagnostic test is conducted, visual inspection as well as comparative analysis between MWR and MWK supports this deduction. The opposite inferences obtained from OLS1 and MWR regarding the existence of spatial autocorrelation in residuals demonstrate that controlling for structural characteristics and selected locational variable alone, its presence is almost unavoidable at the global level. However, residual spatial dependency can be diminished substantially when local modeling technique is applied.

This finding validates the effectiveness of moving window as a sub-market delineation approach to define spatially homogeneous zones and capture spatial heterogeneity. Additionally, the computation effort required for MWR only involves cross validation procedure as a way to determine the optimum window size. The computation expense is not as substantial as other stratification approaches like cluster analysis. Moreover, the moving window approach is designed in a simple and straightforward manner with sound theoretical foundation. Researchers do not need extensive knowledge about local market and neither are they forced to make arbitrary decisions to implement this method. These advantages make MWR a favorable candidate when spatial heterogeneity is present in the process being studied.

With respect to the second question, which is also the highlight of this study, a number of interesting results are obtained. As detailed in chapter 4, MWR and GWR achieve most accurate predictions through incorporation of spatial heterogeneity; despite the absence of neighborhood characteristics, Universal Kriging also improves prediction accuracy substantially after utilizing spatial autocorrelation in residuals via a variance-covariance matrix; MWK, in spite of inclusion of both spatial effects, does not bring essential difference to traditional hedonic model with plain model performance. Therefore, spatial hedonic models do not uniformly dominate traditional hedonic model regarding out-of-sample prediction. The incorporation of spatial dependence or spatial heterogeneity into hedonic modeling is necessary and makes substantive difference only if its existence is well recognized. Thus, caution needs to be paid for exploiting the spatial nature of hedonic modeling since, on the one hand, control for spatial dependence or spatial heterogeneity when they are absent would introduce noise into process and complicate the modeling, while on the other hand, neglecting appropriate adjustments to correct for spatial effects when they exist would bring the plausibility of non-spatial model into doubt.

Lastly, this study is centered on the rationale and steps taken to integrate spatial process in hedonic price function. While the theoretical or conceptual implication of neglecting spatial process in model estimation is well documented, in practice, it depends upon the empirical situation being studied and would vary from one particular housing market to another.

Case et al. (2004) used hedonic price model with homogeneous district and nearest-neighbor residuals to predict out-of-sample transactions for Farifax County's single-family properties. After experimenting with several alternative versions, the relative importance of nearest-neighbor residuals and homogeneous within-country districts is identified. According to the results, the conclusion is that the use of nearest neighbor residuals makes greater contribution in improving the predictive power of model than does the use of homogeneous within-country district. Following this, Case et al. suggest that for future hedonic analysis, focus should be placed on the enterprise of incorporating geographic correlation in that it yields more benefit than the endeavor of defining discrete local areas.

However, based on the results of the present research, a different conclusion is reached. Despite Case's use of cluster analysis for defining discrete local areas which differs from the moving window approach used in MWR and GWR, both approaches reach the same goal of partitioning the study area into homogeneous zones by different routes. Therefore the suggestion inferred from the study of Toronto's metropolitan housing market would be the opposite that improvement gained by describing spatial heterogeneous process is much greater than the improvement achieved by incorporating spatial autocorrelation.

In this light, in practical application, due to the diversity and complexity of price dynamics operating at local market, there is neither a consistent answer to the implication of spatial process nor uniform solution to the adjustment of spatial effects. In order to achieve the best modeling performance, diagnosis of spatial process as well as trial of different paths for incorporating spatial effects are essential.

5.2 Application

It is important for business and academic investigation to accurately predict the metropolitan housing prices. The most typical method is to use the characteristics of housing stock to formulate a hedonic regression, where coefficient estimates obtained from ordinary least squares approach are then used to produce the predicted house prices. The above approach is well known as traditional hedonic price regression. As stressed in literature, the straightforward use of the traditional hedonic function is limited by its neglect of geographic nature of dwellings.

In this research, different ways to incorporate spatial structure are discussed and are also proved to generate more accurate predictions in hedonic analysis. This is relevant in the context of predicting or appraising individual housing units. As indicated in the comparative analysis, the awareness of spatial nature of properties will improve professional assessment, in particular, mass appraisal of property for taxation or other public services. However, it is important to note that for some spatial modeling techniques like GWR, Kriging, when overall prediction benefit from the use of distance weighting scheme or residual spatial dependency structure, undesirable or even intolerable estimations would be obtained for individual observations. The underlying reason might be that there are always 'outliers' in a spatial process, for which distance weighting scheme is not appropriate since their connections with nearby properties are loose, or spatial dependency structure generalized from the entire dataset is not applicable or even act towards an opposite direction. Therefore, for spatial hedonic analysis, especially under the purpose of prediction, caution needs to be paid to those observations that do not follow the general rule as their counterparts do. These 'outliers' may also give valuable information to improve the base model.

Since the investigation is conducted using transactions occurred in the City of Toronto, special advices to this area rather than common generalization can be recommended: (1) defining relatively homogeneous area for hedonic study is essential and will bring significant improvement to housing units' prediction or appraisal; (2) searching for nearby comparable properties for adjustment will also make great contribution when analysis is performed or house price index is constructed on 'global' scale; (3) inclusion of major structural characteristics provide adequate control for spatial autocorrelation in homogeneous sub-markets, therefore rectification or price modification based on nearby transactions is redundant and sometime can even be deleterious.

5.3 Suggestions for Future Research

Due to the extreme computation burden, moving window kriging (MWK) with drift composed of variables other than geographic coordinates, has not been widely applied, especially in hedonic housing studies. Moreover, circumscribed by the condition that no commercial software has a ready-made function for MWK, it application has been further restricted. However, as revealed in this study, inconsistent with its consideration on both spatial autocorrelation and spatial heterogeneity, MWK does not exhibit any extraordinary prediction performance as initially expected. Therefore, more extensive work is needed on the examination of its theoretical soundness and empirical application in order to find out whether the plain performance of MWK is an exception under this particular context or a universal phenomenon.

A possible way to examine the validity of MWK is to further investigate the cross validation procedure. As revealed in chapter 4, MWK has an unstable performance and also yields a relatively large number of negative predictions; therefore, it is of great interest to explore the details of cross validation procedure to understand how the optimum window size is actually defined, whether it is largely affected by poorly performing points, or more specifically, whether the cross validation score carries too much penalty from these points and accordingly the optimum window size leads towards these 'outliers' rather than maximizing the overall prediction accuracy.

Another potential direction is the methodological modification to moving window kriging (MWK). According to its definition, MWK attempts to control for spatial autocorrelation using observed nearby residuals after undertaking regression within each moving window. As indicated in chapter 4, the explanation for undesirable prediction performance of MWK is that moving window regression including structural and proximity characteristics provides adequate control for spatial autocorrelation, therefore, the second step taken by MWK to correct for spatially autocorrelated errors simply adds noise and clouds the process. In order to fix the possibly unwanted step 2, a 'spatial autocorrelation' detector would be helpful which diagnoses the existence and degree of spatial autocorrelation after moving window regression and then, according to the threshold set by modelers, determines whether step 2 will be executed. However, a potential problem involved with this modified MWK would be that no guarantee can be given to the satisfaction of marginal improvement at the expense of costly computation efforts.

Lastly, spatial autocorrelation and spatial heterogeneity has plagued hedonic price analysis and other spatial phenomena studies for a long time. Conditioned by the complexity of specifying spatial effects into model framework, conventionally modelers avoid dealing with both of them at the same time. While no justification is given to the incorporation or declination of a particular one in their studies, some inspiring findings are discovered in this investigation. According to the comparative analysis of prediction performance, it is found that the inclusion of one spatial effect may provide sufficient control for the other one. In other words, when spatial autocorrelation is embedded into model specification, the attempt of modeling heterogeneous process is futile or even harmful for model estimation and vice versa. Inspired by that, future studies should explore how the theoretical basis of spatial hedonic studies can be enhanced by the efforts of distinguishing the individual and joint effects of spatial autocorrelation and spatial heterogeneity. If consistent results are revealed in other studies, it would be legitimate to abstract one spatial effect from the modeling process and focus on the other one. Otherwise, clear statement of implication of ignoring one type of spatial effect will at least minimize the danger of discarding it for model estimation.

BIBLIOGRAPHY

Anselin L (1995). "Local Indicators of Spatial Association—LISA." *Geographical Analysis* 27(2), 93–115.

Bailey TC and Gatrell AC (1995). *Interactive Spatial Data Analysis*. Longman Scientific and Technical, Essex, England.

Basu S and Thibodeau TG (1998). "Analysis of Spatial Autocorrelation in House Prices." *Journal of Real Estate Finance and Economics* 17 (1), 61-85.

Bowen WM, Mikelbank BA and Prestegaard DM (2001). "Theoretical and Empirical Considerations Regarding Space in Hedonic Housing Price Model Applications." *Growth and Change* 32(4), 466-490.

Brunsdon CF, Fotheringham AS and Charlton ME (1996). "Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity." *Geographical Analysis*, 28(4), 281-298.

Brunsdon CF, Fotheringham AS and Charlton ME (1998). "Geographically Weighted Regression --- Modelling Spatial Non- stationarity." *Journal of the Royal Statistical Society, Series D-The Statistician* 47(3), 431-443.

Can A (1990). "The Measurement of Neighborhood Dynamics in Urban House Prices." *Economic Geography* 66(3), 254–272.

Can A (1992). "Specification and Estimation of Hedonic Housing Price Models." *Regional Science and Urban Economics* 22, 453–477.

Can A and Isaac M (1997). "Spatial Dependence and House Price Index Construction." *Journal of Real Estate Finance and Economics* 14, 203–222.

Case B, Clapp J, Dubin RA and Rodriguez M (2004). "Modeling Spatial and Temporal House Price Patterns: A Comparison of Four Models." *Journal of Real Estate Finance and Economics* 29(2), 167-191.

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

Des Rosiers F, Thériault M and Villeneuve P (2000). "Sorting Out Access and Neighbourhood Factors in Hedonic Price Modeling." *Journal of Property Investment and Finance*18(3), 291-315.

Dubin RA (1992). "Spatial Autocorrelation and Neighborhood Quality." *Regional Science and Urban Economics* 22, 433–452.

Dubin RA (1998). "Predicting House Prices Using Multiple Listings Data." *Journal of Real Estate and Financial Economics* 17(1), 35–59.

Dubin RA, Pace RK and Thibodeau TG (1999). "Spatial Autoregression Techniques for Real Estate Data." *Journal of Real Estate Literature* 7(1), 79 – 96.

Farber S (2004) "A Comparison of Localized Regression Models in an Hedonic House Price Context." (Ryerson University, research paper).

Fotheringham AS, Charlton ME and Brunsdon CF (1996), "The Geography of Parameter Space: An Investigation into Spatial Non-Stationarity." *International Journal of Geographic Information Systems* 10(5), 605-627.

Fotheringham AS, Brunsdon CF and Charlton ME (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: Wiley

Goldberger AS (1968). "The Interpretation and Estimation of Cobb-Douglas Functions." *Econometrica* 36, 464-472.

Goodman AC and Thibodeau TG (1998). "Housing Market Segmentation." *Journal of Housing Economics* 7, 121-143.

Haas TC (1990*a*). "Kriging and Automated Variogram Modeling Within a Moving Window." *Atmospheric Environment* 24(A), 1759–1769.

Haas TC (1990*b*). "Lognormal and Moving Window Methods of Estimating Acid Deposition." *Journal of the American Statistical Association* 85, 950–963.

Haas TC (1995). "Local Prediction of a Spatio-Temporal Process with an Application to Wet Sulfate Deposition." *Journal of the American Statistical Association* 90, 1189–1199.

Kestens Y, Thériault M and Des Rosiers F (2006). "Heterogeneity in Hedonic Modeling of House Prices: Looking at Buyers' Household Profiles." *Journal of Geographical Systems* 8, 61-96.

Lam N (1983). "Spatial Interpolation Methods – A Review." *American Cartographer* 10(2),129-149.

Lancaster KJ (1966). "A New Approach to Consumer Theory." *Journal of Political Economy* 74, 132–157.

Lark RM (2000). "Regression Analysis with Spatially Autocorrelated Error: Simulation Studies and Application to Mapping of Soil Organic Matter." *International Journal of*

Geographical Information Science 14(3), 247-264.

Maclennan D (1976). "Some Thoughts on the Nature and Purpose of House Price Studies." *Urban Studies* 14 (1), 59-71.

Malpezzi S (2003). "Hedonic Pricing Models: A Selective and Applied Review." Prepared for Housing Economics and Public Policy: Essays in Honor of Duncan Maclennan.

Oliveau S and Guilmoto CZ (2005). "Spatial Correlation and Demography: Exploring India's Demographic Patterns." *XXV International Population Conference.*

Pace RK and Gilley OW (1997). "Using the Spatial Configuration of the Data to Improve Estimation." *Journal of the Real Estate Finance and Economics* 14(3), 333–340.

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

Pavlov AD (2000). "Space-Varying Regression Coefficients: A Semi-parametric Approach Applied to Real Estate Markets." *Real Estate Economics* 28 (2), 249-283.

Rosen S (1974). "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy* 82 (1), 34–55.

Schnare AB and Struyk RJ (1976). "Segmentation in Urban Housing Markets." *Journal of Urban Economics* 3, 146-166.

Sirmans D, Macpherson DA and Zietz EN (2005). "The Composition of Hedonic Pricing Models." *Journal of Real Estate Literature* 13 (1), 1-44.

Walter C, McBratney AB, Douaoui A and Minasny B (2001). "Spatial Prediction of Topsoil Salinity in the Chelif Valley, Algeria, Using Local Ordinary Kriging with Local Variograms versus Whole-area Variogram." *Australian Journal of Soil Research* 39 (2), 259-272.

Webster R and Oliver MA (1992). "Sample Adequately to Estimate Variogram of Soil Property." *Journal of Soil Science* 43, 177-192.

MPAC: Municipal Property Assessment Corporation http://www.mpac.ca/. Accessed on April 15, 2006.