Spatial Econometrics Revisited: A Case Study of Land Values in Roanoke County

Ioannis K. Kaltsas

Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Economics

Dr. Darrell Bosch, Co-Chair
Dr. Anya McGuirk, Co-Chair
Dr. Leonard Shabman
Dr. Aris Spanos
Dr. Brad Mills

October 31, 2000
Blacksburg, Virginia

Keywords: Spatial Econometrics, Land Values, Misspecification, Residential Settlement

To my father

Konstantinos Kaltsas

ACKNOWLEDGEMENTS

Spatial Econometrics Revisited: A Case Study of Land Values in Roanoke County

Ioannis K. Kaltsas

(ABSTRACT)

An increasing volume of empirical literature demonstrates the possibility of spatial autocorrelation in land value models. A number of objections regarding the methodology followed in those empirical studies have been raised. This thesis examines three propositions. The first proposition states that there is spatial dependence in the land value model in Roanoke County. The second proposition is that mechanical construction of neighborhood effects, or grouping nearby land parcels into neighborhoods, is not always the best way to capture spatial effects. Finally, the third and most important proposition states that by implementing a comprehensive set of individual and joint misspecification tests, one can better identify misspecification error sources and establish a more statistically sound and reliable model than models based on existing spatial econometric practices. The findings of this dissertation basically confirm the validity of those three propositions. In addition, we conclude that based on their development status prices of land parcels in Roanoke County may follow different stochastic processes. Changes in the values of hedonic variables have different implications for different groups of land parcels.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Problem Statement

1.1 Introduction

Land value indices contribute towards better understanding of local and national land markets and can be useful for comparative analysis of price trends, real estate investment analysis and regional development program evaluation. The accuracy and the precision of the indices will be affected by the predictive power of the land value model. Traditionally, economists have constructed hedonic functions to capture the relative importance of a number of land attributes affecting land values (Xu F. *et al*., 1993).

A growing consensus regarding the importance of spatial structure in land and residential value models is reflected in the number of recent studies that also incorporate spatial characteristics as explicit variables (Can and Megboludge 1998, Basu and Thibodeau

1998). Bockstael (1996) emphasizes the importance of including explicit spatial variables, such as proximity to amenities and disamenities, in the hedonic representation of land value. Despite the inclusion of these additional "spatial" variables, it is acknowledged that there may exist more spatial features that are unobserved and thus, omitted from the model (Bockstael and Bell 1998, Irwin 1998, Fleming 1988). As a result, the "true" underlying error of the regression is assumed to be spatially autocorrelated.

Error spatial autocorrelation means that observations that are relatively close in space have error terms that are correlated, while observations that are far apart in space tend not to be correlated. Spatial autocorrelation causes least square estimators to be biased and inefficient, making inference based on them invalid (Basu and Thibodeau, 1998). Spatial error autocorrelation can occur for reasons other than omitted spatially dependent variables. For example, autocorrelation can also occur when the error variance is heteroskedastic (Anselin 1988). Davidson and Mackinnon (1993) conclude that the situation where a model is correctly specified except for a failure to account for error autocorrelation does not account for a very high proportion of the cases in which residuals from a regression model appear to be correlated. Consequently, it is important to determine the source of autocorrelation for respecification purposes; the appropriate "cure" for spatial correlation, depends on its source.

Anselin (1988) and Anselin and Kelejian (1997) have suggested that a spatial autoregressive error model can be used to model the spatial structure of the errors arising from omitted "spatial" effects and generate more precise results. According to Anselin (1988) the spatial autoregressive component corrects predicted values by an estimate of the prediction error's relationship to nearby observations and thus mimics the behavior of real estate appraisers. The absolute influence that nearby properties have on land value is determined using an exogenously determined weight matrix. Different weight matrices represent different hypotheses regarding the structure of spatial dependence and an educated selection is possible through a series of non-nested tests.

Some researchers (Basu and Thibodeau 1998) argue that a spatial autoregressive model may not be effective in cases where the observed spatial dependence is caused by some factors other than omitted variables. Sometimes the chosen functional form does not adequately allow for heterogeneity over space, and the estimated parameters are unstable, usually varying by location. For example, most of the land value models assume that the functional form is the same for both developed and vacant land parcels (Beaton, 1991). However, prices of developed parcels may follow a different stochastic process from prices of vacant land. Part of the observed spatial autocorrelation in the residuals may be attributed to this structural instability. Anselin (1988) believes that the problem of distinguishing the sources of observed spatial autocorrelation is "highly complex" and proposes testing structural stability before creating an autoregressive model. According to this literature, the assumptions of normality and heteroskedasticity should also be tested before the researcher tests the model for spatial autocorrelation (Anselin 1988, Anselin and Kelejian 1997). By first verifying the assumptions of normality, heteroskedasticity and correct functional form, they hope to "ensure" that spatial dependence is the source of autocorrelation and the incorporation of spatial structure in the model improves the properties of the estimators.

However, there are a number of objections to current spatial modeling approaches, which potentially limit their success. First, current spatial econometric studies test only for a subset of the assumptions underlying the statistical model. For example, Anselin (1999) summarizes the findings of a large number of spatial econometric studies. The overwhelming majority of these studies do not report more than two misspecification tests. Typically, there are many more than two assumptions underlying a regression model. Spanos (1986) clearly indicates that testing the validity of all the assumptions underlying the statistical model is an issue of paramount importance. When some of these assumptions are invalid, the statistical inference results are, in general, invalid.

Even if current spatial econometric studies assessed the validity of all the assumptions underlying the statistical model, the approach of testing individual assumptions one-by-one and fixing problems when necessary is very unlikely to lead to a well-specified model. The first reason is that most tests of assumptions are only valid if all other assumptions underlying the model are correct. If one of the other assumptions is invalid the test may be unreliable (Spanos 1986). The test results are unreliable in the sense that they may lead to either too many rejections, or too few rejections. The implication of this fact for the current approach to spatial econometrics is that if there is spatial dependence, the tests conducted to ensure that all assumptions are valid except for the assumption of no spatial autocorrelation will be unreliable, and thus inference based on these tests will be misleading. The second reason the strategy of testing and curing assumptions one-by-one is likely ineffective is that rejection of a particular null hypothesis (that an assumption is valid) should not be interpreted as acceptance of the alternative hypothesis. Tests should not be interpreted one-by-one, but rather contribute to a bigger picture of potential misspecification sources. Model assumptions are very closely related to each other in the sense that the model assumptions are all about a conditional distribution that is derivable from some joint distribution (Spanos 1986). This means that all the model assumptions are really interconnected via reduction assumptions, and if any one of the reduction assumptions is invalid then it is likely that more than one of our model assumptions are invalid. Bockstael and Bell (1998) recognized the possibility that violation of an assumption can lead to violation of other assumptions. They attribute violations of the assumptions of normality and homoskedasticity to the existence of spatial autocorrelation in their land value model. However, they never retest the final model to verify whether their corrected (for spatial autocorrelation) model also "cures" the non-normality and heteroskedasticity problems.

A battery of tests to identify misspecification sources is particularly necessary when violations of the autocorrelation assumption are observed. According to Spanos (1986), autocorrelation type tests are particularly sensitive to violations of other model assumptions. That is, violations of other assumptions often lead to rejection of the no

autocorrelation hypothesis. Thus, the assumption of no spatial autocorrelation is usually rejected when some other assumptions are violated. In addition to individual tests, this battery of tests should also include joint misspecification tests. McGuirk et al. (1993) conclude that misspecification tests of individual assumptions underlying regression models often lead to erroneous conclusions regarding sources of misspecification. Joint misspecification tests have fewer maintained hypotheses and are potentially very helpful in identifying misspecification sources. Consequently, both individual and joint tests are necessary to identify misspecification sources and guide respecification efforts in an attempt towards achieving a statistically adequate model.

Furthermore, the success of the current spatial econometric approaches has been judged solely by the changes in the fitting power of the models. However, it is widely accepted that higher fitting power (higher R^2) alone is not an adequate criterion to judge a model. McGuirk et al. (1993) argue that high fitting power does not guarantee the validity of an econometric model, and an increase in R^2 does not necessarily imply that a model represents an improvement over models with lower R^2. This work suggests that although spatial autoregressive corrections often lead to models with higher fit, the correct model may or may not be an improvement over the original model; one needs to look at more than the overall fit, when assessing the success of the spatial autoregressive models.

Finally, in current approaches to dealing with spatial autocorrelation, the specification of spatial structure in an econometric model is completely arbitrary. There are no specific rules to determine the spatial relationship among individual observations. In addition, there is no way to test the validity of the structure of the exogenously determined spatial weight matrix. Existing non-nested tests only select the weight matrix that maximizes the fitting power of the econometric model. In addition, spatial weight matrices are typically constructed using mathematically computed distances and, thus, geographical proximity is the only criterion to account for neighborhood effects. However, the size of neighborhoods might often be inappropriate for a given case study and proximity might not be always the best criterion for determining neighborhoods.

Despite the fact that the possibility of spatial dependence in a cross-sectional model is widely accepted, it is unclear whether current practices designed to incorporate the spatial structure in cross-sectional data really improve the model and lead to more precise estimates. There is a need to extend current spatial econometric practices to address the above limitations. Indeed, the overall purpose of this dissertation is to address these issues in the context of modeling land values. Land value modeling has been widely used by the "traditional" spatial econometric literature to demonstrate its achievements. In this dissertation we first attempt to evaluate the usefulness of existing spatial autocorrelation models in capturing land values, using data from Roanoke County in Virginia. It turns out for this particular example, existing methods are inadequate in terms of their ability to capture the systematic information of our data. Once we identify the shortcomings of current practices in modeling land values in Roanoke County, we demonstrate a more successful approach, which utilizes the idea of a battery of individual and joint misspecification tests to guide respecification efforts.

1.2 Objectives and Propositions

This thesis is a case study of hedonic land value modeling in Roanoke County, Southwest Virginia. The study is part of a combined effort of 6 academic departments of Virginia Tech to analyze the consequences of alternative residential development on rural areas (Diplas *et al.* 1997). Urban expansion can be achieved with different arrangements of residential lots. Each of these forms has different economic, environmental and social implications for the future of Roanoke County. The construction of a land value model serves the purpose of capturing the economic effects of alternative residential settlement forms. Simulating potential future developments, the model assigns prices to each individual land parcel. In this way, the policy maker can assess the change in the value of land assets and in the property tax revenues for the local government resulting from different development forms.

The contribution of this thesis to the overall study is the construction of a statistically adequate land value model, which will provide reliable land value indices for Roanoke County. The realization of this goal coincides with the necessity to revisit current spatial econometric practices. This thesis aims to: a) demonstrate weaknesses in the current econometric approach to dealing with spatial dependence and b) propose a more reliable way towards achieving a statistically adequate model for analyzing data, which may be characterized by spatial dependence. Thus, the contribution of this study to the economic literature is expected to go beyond the construction of an empirical model for the case of Roanoke County.

This dissertation is built around a set of three propositions. The first proposition is related to the question of whether the appearance of spatial autocorrelation in land value models of Roanoke County is the result of some model misspecification. Our proposition is that there is spatial dependence in the values of land parcels in Roanoke County. This is the starting point of the thesis and justifies the use of the spatial econometric approach. To assess the relevance of this proposition, we initially follow current econometric practices. That is, we test for possible model misspecification using tests of assumptions commonly implemented in this literature. Once we make sure that the statistical model does not violate those assumptions, we then test for spatial autocorrelation. Spatial autocorrelation is tested by using the Moran's I test, which is widely accepted in the spatial statistic literature (Cliff and Ord, 1981), as well as by standard auxiliary regression tests. The results of these tests represent an initial assessment regarding the validity of the first proposition.

The first proposition of the thesis is later reassessed when the problem of spatial autocorrelation is addressed using the alternative approach. A battery of individual and joint misspecification tests of all assumptions is conducted. Based on these tests, our model is respecified and the same battery of tests reconducted. This procedure is repeated until a statistically adequate model is obtained. Based on this model, the validity of the first proposition of the thesis is reassessed; if the final statistically adequate model

includes spatial lags in its structure, we will conclude there is spatial dependence in land values.

The second proposition relates to current practices regarding specification of neighborhoods used for constructing the spatial weight matrix. To date, most spatial empirical studies (Fleming 1998, Bockstael and Bell 1998) use metric distances to create spatial lags to capture effects within neighborhoods, without examining the appropriateness of the particular distance function in the specific case study. The second proposition of this study states that mechanical construction of neighborhood effects, or the grouping of geographically nearby land parcels into neighborhoods, is not always the best way to capture spatial effects. The researcher should also pay close attention to the particularities of the data before defining spatial relationships among empirical observations. In addition to geographical proximity, neighborhood effects may also include morphological and socioeconomic elements of the empirical information.

Finally, the third and most important proposition of the thesis is related to the success of current spatial econometric practices in capturing the spatial aspects of land values. It states that by implementing a comprehensive set of individual and joint misspecification tests, proposed by Spanos (1986), one can better identify misspecification error sources and thus, establish a more statistically sound model, which is more reliable than models based on the existing spatial econometric practices. The current spatial econometric techniques and the alternative comprehensive testing approach will be illustrated in detail in the following chapters. The validity of the third proposition will be assessed by comparing results from both approaches.

Before examining the above propositions, the second chapter of the thesis provides the basic empirical framework in land value modeling. It also presents the empirical information from Roanoke County and the construction of relevant variables. The third chapter of the thesis introduces the concept of spatial autocorrelation, explains the practical reasons for the use of weight matrices, and presents current spatial econometric

practices as well as an alternative approach. The fourth chapter follows the current spatial econometric approach and attempts to ensure that all assumptions of concern in these studies other than spatial autocorrelation are satisfied. Chapter 4 also discusses the optimal construction of a weight matrix to capture neighborhood effects and analyzes the implications of a mathematically computed weight matrix for our case study. In the fifth chapter of the thesis, current spatial econometric techniques are applied to deal with the problem of spatial autocorrelation. Current techniques are then compared and contrasted with the results of an alternative methodology demonstrated in Chapter 6. The sixth chapter will also demonstrate that grouping of geographically nearby land parcels into sets of neighborhoods, is not always the best approach. Conclusions and hypotheses for future research efforts are provided in the last chapter, which also summarizes the findings of the thesis.

1.3 Conclusions

There is an increasing volume of empirical literature, which demonstrates the possibility of spatial autocorrelation in land value models. Current spatial econometric approaches examine the assumptions of normality, heteroskedasticity, and structural stability to make sure that spatial error autocorrelation is not observed due to other misspecification problems. According to this "traditional" approach the problem of true spatial autocorrelation can be corrected successfully by using a spatial error autoregressive model. However, a number of objections regarding the methodology followed in those empirical studies have been raised. Current spatial econometric studies examine only a subset of the underlying model assumptions to determine the source of misspecification problems. Even if they examined all the necessary assumptions, their approach of examining misspecification tests one-by-one and fixing misspecification problems one at a time also probably leads to erroneous conclusions. A battery of individual and joint tests is necessary to identify misspecifications sources and lead respecification efforts. Another objection to current methods is that the success of their results is judged by the changes in

the fitting power of the model. Finally, in the current approach to dealing with spatial autocorrelation, the specification of spatial structure in an econometric model is completely arbitrary.

Using data from Roanoke County, we question the success of current spatial econometric practices in capturing the spatial aspects of land values, and we attempt to present a more reliable approach to model cross-sectional data. This thesis examines three propositions. The first proposition states that there is spatial dependence in the land value model in Roanoke County. The second proposition is that mechanical construction of neighborhood effects, or grouping nearby land parcels into neighborhoods, is not always the best way to capture spatial effects. Finally, the third and most important proposition states that by implementing a comprehensive set of individual and joint misspecification tests, one can better identify misspecification error sources and establish a more statistically sound model, which is more reliable than models based on existing spatial econometric practices.

CHAPTER 2

The Case Study in Roanoke County

2.1 Introduction

This chapter introduces the basic empirical framework of land value modeling, gives details on the available data from Roanoke County, and illustrates the construction of explanatory variables, which will be used to capture land price variations. Existing empirical evidence guided efforts to identify relevant empirical information, while existing empirical studies helped in formulating specific variables.

2.2 Existing Empirical Literature in Land Value Modeling

In 1889, Francis Galton used correlation analysis and paved the way for numerous studies in which researchers attempted to ascertain the importance of various attributes of land. The large body of existing empirical literature suggests that property derives value from three different sources. First, the productivity of land reflects its ability to provide consumers with goods and services. It is assumed that land derives a good portion of its value from characteristics of the parcel itself. The quality of land is usually associated with its innate characteristics such as quality of soil, elevation, erosiveness (which is usually related to slope), and flooding potential. Additional characteristics that have an impact include the size of the parcel, the amount of land available for production purposes, and the level of productivity-enhancing land investment, such as roads and buildings.

Second, location is an important determinant of land value. The price an individual will pay is influenced by access to various goods and services. The proximity of land parcels to towns or malls that offer such amenities and services as shopping, entertainment and educational facilities is expected to increase the value of land. In addition, several studies (Nelson 1986; King and Sinden 1994) conclude that population pressure from nearby towns is expected to increase rural land values. Proximity to good roads will likely have a positive effect on the value of land, as it allows for access to both towns and highways. However, strip development along those same state roads may have a price depressing effect.

Finally, the preferences of both buyers and sellers are usually reflected in the transaction price. If the land attribute of central importance is its production value, its price will reflect relative productivity of alternative parcels (King and Sinden 1994). However, if land is perceived as developable, the land parcel should reflect its speculative value. Policies such as zoning affect productivity, speculative, or consumptive values of the land in positive or negative ways.

2.2 Land Value Empirical Information

The sensitivity of land value indices to changes of geographical patterns depends on the level of disaggregation of the spatial units. To achieve the most detailed level of spatial information possible, this study makes use of information related directly to the individual land parcels. Geographical Information Systems (GIS) can provide detailed information about the exact location and attributes of a land parcel based on a system of coordinates. This is also essential for determining the proximity of the parcels and achieving ordering of spatial observations (Gleeson, 1979). The Roanoke County Planning Department has geocoded the parcels in the tax assessment database, and this study uses the product of their work. Locating the parcels in a GIS means that we can employ our geocoded maps of features of the landscape and road network to describe the parcel characteristics more completely. Most of these characteristics were quantified with the help of the GIS tools.

A random sample of observations used to estimate the model is extracted from the Roanoke County Division of Planning and the Roanoke County Division of Tax and Assessment data base, which includes market prices and transaction dates over the last five years for each privately held parcel in the county. There were 1,844 transactions of vacant and non-vacant land parcels for the period of 1996 to 1997. Our problem is complicated by the fact that we attempt to estimate a model, which will predict the value of the land in any economic use, minus the value of any structure. One way of dealing with this problem is to include in the model as many variables describing the existing structure as possible in the hope that it is feasible to separate the value due to location and size of the lot from the value of the structure. This idea, however, may be problematic for a number of conceptual and econometric reasons, including the absence of structural characteristics for a large portion of the available database. The alternative course of action is to extract the cost of the structure from the locational value by subtracting the

assessed value of the structure from the transaction price (Bockstael and Bell, 1998). This latter course of action is taken in this dissertation.

Table 2.1 contains the descriptive statistics for a number of variables essential for the construction of a land value model. The price of the parcels (Price) is measured in dollars per square meter. The average price of our sample is $23.13 while the median is $3 per square meter. The area of the parcels (Area) is measured in square meters. The size of the parcels varies from 0.005 hectares (a parcel close to the urban fringe of Roanoke County dedicated to commercial use) to 216 hectares (a parcel of steep and remote agricultural land). The elevation of the center of the parcel (Elevation) is measured in meters above the average elevation of the GIS map from sea level. The average slope of the parcel (Slope) is measured in geometric degrees. There is a high correlation (r=0.68) between the slope of the parcel and its elevation. Most of the developed parcels are located on relatively flat land with low elevation. The soil quality of the land parcels was classified into three categories according to the permeability to water. The dummy variable representing Soil Quality 1[1] is the less absorbing category of soil, while Soil Quality 2[2] has an intermediate level of penetrability. The ability of the soil to absorb water is related to lower flood risk and soil erosion. The category Soil Quality 1 contains 3% of the parcels, while the category Soil Quality 2 accounts for 87% of the parcels.

Point to point distances of the parcels from the shopping malls, the town of Roanoke and the town of Blacksburg are measured in meters. The minimum distance of the parcels to either of two urban centers is about 3 kilometers and the maximum is close to 15 kilometers. However, the town centers may be less important than the shopping malls in terms of daily commuting. The Planning Department of Roanoke County (PDRC) estimates that several thousand consumers visit daily the two malls of the county. Additionally, these malls have become the center of development of hundreds of small businesses, which offer employment to thousands of Roanoke County residents.

---

[1] Inceptisoils
[2] Ultisoils

According to the PDRC the development rates of the areas close to the shopping malls are expected to be the highest in the county for the next five years.

Table 2.1. Descriptive Statistics of the Empirical Information

Variable	Average	S.D.	Min	Max
Price ($/m^2)	23.13	18.08	0.02	133.40
Area (m^2)	8546.53	75202.91	56.97	2165233
Elevation (m)	379.82	88.69	3.22	1003.00
Slope (degrees)	5.49	3.54	0.00	34.56
Soil Qual. 1	0.03	0.17	0.00	1.00
Soil Qual. 2	0.87	0.33	0.00	1.00
Mall 1 (m)	8861.89	4281.51	2002.89	27024.59
Mall 2 (m)	9246.60	4774.35	435.92	27483.35
Roanoke (m)	8828.68	3818.12	3395.87	28794.36
Blacksburg (m)	39858.68	6786.72	18165.42	51206.84
Road	0.05	0.22	0.00	1.00
Population (p/He)	5.90	4.60	0.05	18.65
Developed	0.88	0.33	0.00	1.00
Coord. Y	16881.90	6022.44	1.81	30585.74
Coord. X	24888.27	6766.91	0.15	36626.16
Year	0.49	0.50	0.00	1.00

About 5% of the parcels are located next to a major road (state or interstate highway). In case of Roanoke County, the existence of a major road (Road) is associated with several disamenities (i.e. noise and air pollution), which may affect negatively the land price. More open space and easier access to natural amenities may also be captured by the measurement of population density of the census blocks in which the parcel belongs. The

average population density (Population) of our sample is about 6 people per hectare. Twelve per cent of the sample consists of vacant parcels, while the remaining 88% of the parcels includes some type of construction. The dummy variable "Developed" equals one if the tax assessors of the Roanoke County indicated the existence of a building on the parcel, otherwise zero. The Coordinates X and Y can be used to identify the exact location of the center of each parcel on a map. The Coordinate X increases as we move West and North in Roanoke County, while the Coordinate Y increases as we move East and North in Roanoke County. These coordinates can calculate direct distances and define the proximity and the neighboring effects of the parcels. Finally the dummy variable "Year" indicates whether a parcel was sold in 1996 (Year=0) or in 1997 (Year=1). According to the U.S. Bureau of Census the average price of rural land in Roanoke County increased by 1.5% in 1997 relative to the previous year.

2.3 Conclusions

Economic theory suggests that location is probably the most important determinant of land values. However, a number of empirical studies also suggest that hedonic attributes (i.e. soil quality, elevation, access to road, development status) can also explain variation in land prices. With the help of GIS, we collected a number of geographical, morphological and socioeconomic characteristics of land parcels in Roanoke County. Before the construction of a land value model, which links those parcel characteristics to their transaction prices, the next chapter introduces some elements of spatial statistics.

CHAPTER 3

Basic Elements of Spatial Statistics

3.1 Introduction

In this chapter, some of the fundamental notions of spatial statistics are illustrated, setting the necessary framework for proceeding with the empirical analysis in the following chapters. We begin by introducing the concept of spatial autocorrelation and explaining the practical reasons for the use of weight matrices and the different ways of constructing a spatial weight matrix. A measure of spatial autocorrelation, the Moran I test, introduces the idea of testing spatial autocorrelation. Then, current spatial econometric modeling practices and an alternative approach proposed by Spanos (1986) are presented.

3.2 Spatial Autocorrelation

Consider a region, which can be exhaustively partitioned into n non-overlapping parcels of land. Let the observed value of land price Y, in the typical parcel i, be y_i. Y describes a single population from which repeated drawings are made to give the sample $\{y_i\}$. If for each pair of parcels i and j, the drawings that yield y_i and y_j are uncorrelated, then we say that there is no spatial autocorrelation of land prices. Conversely, spatial autocorrelation in land prices exists if the drawings are not all pairwise uncorrelated. The problem of spatial autocorrelation or of determining whether geographical data are spatially autocorrelated is fundamentally different from testing autocorrelation in stationary time series. This is due to the fact that the observed variable in a time series is influenced by past values, while spatial dependence potentially extends in all directions.

Formally, in time series analysis a first order autoregressive model can be formulated as:

$$y_t = ay_{t-1} + e_t \tag{3.1}$$

where y_t is the value of Y at period t, a is an unknown parameter, and e_t is a random disturbance. In the spatial situation, the first order autoregressive model for a regular lattice with K rows and Z columns, considering only interactions between cells with common edge, can be stated as:

$$y_{k,z} = a_1 y_{k-1,z} + a_2 y_{k+1,z} + a_3 y_{k,z-1} + a_4 y_{k,z+1} + e_{k,z} \tag{3.2}$$

where a_1, a_2, a_3 and a_4 are parameters. Note, however, that model (3.2) assumes the ideal case where different parcels have the same shape and size. If the shape and size of the different parcels vary then model formulation becomes more complicated. Different shapes and sizes demand different numbers of weights for the parameters. Furthermore, formulating model (3.2) is practically infeasible due to the assumption that each parcel

should be contiguous to a maximum of four other parcels - when in fact parcels may border more than four parcels.

The idea of a weight matrix designed to overcome these problems dates backs to the first days of spatial statistics. A weight matrix can make model (3.2) operational by imposing restrictions on the parameters of the lagged spatial variables. The size and shape of parcels and the relative strength of links between them (road and rail links, for example) can be captured by choosing a set of weights for the matrix. However, the structure of the weight matrix is *ad hoc* and its validity cannot be tested directly.

3.3 Weight Matrices

A weight matrix has the role of ordering observations in space. In a time-series, the ordering is obvious. However, in space it is sometimes preferable instead of ordering one observation after another in an arbitrary fashion to order one observation after the average of more than one observation. The idea is that averaging neighboring y_i's results in a smoother map of land values than the map of the original observations. In this way, the researcher can observe trends in her spatial data and use them in her modeling. The more points included in the moving average, the greater the smoothing will be. In a binary contiguous matrix the underlying structure of the neighbors is expressed by 0-1 values. In other words, the nearby neighbors get a weight of 1 and others 0. If neighbors are characterized as the nearest 4 points, then a four point spatial weight moving average can smooth out the land value map.

The obvious problem with using the binary contiguous matrix is that it does not allow for spatial variations in the distribution of sample sites. For example, there is no discrimination between a site where the closest four points have an average distance of 10 km and another site where the closest four points have an average distance of 50 m. In both cases, the four closest points to a site define a neighborhood. To counter this

problem geologists have developed the notion of a generalized weight matrix, which essentially uses a weighted average of neighboring points. Usually, the weighting factor is chosen to be zero beyond some appropriate distance, downgrading the influence of neighbors beyond some distance from the site under consideration. In a similar manner, the researcher can consider several orders of contiguity by defining a series of concentric bands around the spatial unit under consideration.

In general, we can define a spatial weight matrix W, such that each element, w_{ij} represents a measure of spatial proximity between two parcels i and j. As a rule, the choice of w_{ij} will depend upon the sort of data that the researcher is dealing with and the particular mechanisms through which one expects spatial dependence to arise. According to Griffith (1988), some possible criteria might be:

$w_{ij}=$ 1   centroid of j is one of the k nearest centroids of i,

         0   otherwise

$w_{ij}=$ 1   centroid of j is within some specified distance of i,

         0   otherwise

$w_{ij}=$ z_{ij} if inter-centroid distance $z_{ij}<R$ (R>0),

         0   otherwise

$w_{ij}=$ 1   parcel j shares a common boundary with parcel i,

         0   otherwise

$w_{ij}=$ L_{ij}/L_i where L_{ij} is the length of common boundary between i and j,

                  and L_i is the perimeter of i

Similarly, it is sometimes necessary to specify spatial weight matrices of different orders, often referred to as spatial lags. For example, we might require a series of weight matrices $W^1,\ldots, W^k$ where W^1 represents spatial proximity of the areas at spatial lag 1 (within some distance band) then W^2 represents spatial proximity at spatial lag 2 (within next distance band) and so on.

There is no convention as to which type of weight matrix should be used in spatial statistics analysis. However, it is evident that the structure of spatial dependence incorporated in the spatial weight matrix should be chosen judiciously, and in agreement with general concepts from spatial interaction theory, such as the notion of accessibility and morphological similarity (Anselin, 1988). A spatial weight matrix should be based on the visualization of spatial patterns and of the interaction among spatial units. A theoretical conceptualization of the structure of dependence should not justify the choice of a weight matrix. Rather careful observation of the existing information in each case study could reveal the ordering of the data in space and the appropriate choice of weights.

Practically, the creation of a weight matrix assumes the existence of a map, which depicts the spatial arrangement of the spatial units. In this matrix, each unit is represented both as a row and as a column. In each row, the nonzero column elements correspond to contiguous spatial units. For example, for the nine cells of Figure 3.1, the corresponding 9 by 9 matrix (with the cells numbered from left to right and top to bottom) is given in Table 3.1. For example, the cell Y11 corresponds to the first column and row of the matrix, while the cell Y21 corresponds to the fourth column and row of the matrix.

Figure 3.1 An Example of Spatial Arrangement of Nine Spatial Units

Y11	Y12	Y13
Y21	Y22	Y23
Y31	Y32	Y33

In this case, two cells are defined as contiguous when they are next to each other either vertically or horizontally. By convention, a cell is not contiguous to itself, which results in zero diagonal elements. The weight matrix is also a symmetric matrix, because when a cell A is contiguous to a cell B, then by definition the cell B is also contiguous to a cell A. For example, the cell Y23 is contiguous to the cells Y13, Y22 and Y33, so both the sixth

row and the sixth column of the weight matrix will have values of one only in the third, fifth and ninth positions.

Table 3.1. Weight Matrix Based on the Arrangement of the Spatial Units in Figure 3.1*

0	1	0	1	0	0	0	0	0
1	0	1	0	1	0	0	0	0
0	1	0	0	0	1	0	0	0
1	0	0	0	1	0	1	0	0
0	1	0	1	0	1	0	1	0
0	0	1	0	1	0	0	0	1
0	0	0	1	0	0	0	1	0
0	0	0	0	1	0	1	0	1
0	0	0	0	0	1	0	1	0

*Cells are numbered from left to right and top to bottom. For example, the first column, first row entry corresponds to Y11 in Figure3.1, and second column, first row entry to Y12.

Usually, though not necessarily, the spatial weight matrix is standardized to have row sums of unity. The standardized weight matrix of Figure 3.1 is presented in Table 3.2.

Table 3.2. Standardized Weight Matrix Based on the Arrangement of the Spatial Units in Figure 3.1

0	1/2	0	1/2	0	0	0	0	0
1/3	0	1/3	0	1/3	0	0	0	0
0	1/2	0	0	0	1/2	0	0	0
1/3	0	0	0	1/3	0	1/3	0	0
0	1/4	0	1/4	0	1/4	0	1/4	0
0	0	1/3	0	1/3	0	0	0	1/3
0	0	0	1/2	0	0	0	1/2	0
0	0	0	0	1/3	0	1/3	0	1/3
0	0	0	0	0	1/2	0	1/2	0

* Cells are numbered from left to right and top to bottom. For example, the first column, first row entry corresponds to Y11 in Figure3.1, and second column, first row entry to Y12.

3.4 Testing Spatial Autocorrelation

At first sight, spatial dependence may seem similar to the more familiar time-wise dependence encountered in time series econometrics. However, this is only partially the case. Standard theoretical results from time series econometrics do not always carry over in a straightforward way to the multi-dimensional dependence in space. This realization led researchers to create measures of spatial dependence, which differ from those of serial dependence. The first measure of spatial dependence (or, more precisely, spatial autocorrelation) proposed by Moran (1948) is the best known and most widely used measure to test for spatial autocorrelation. This statistic has been widely adopted in geology and biology because it is also fairly easy to compute.

Formally the Moran's I statistic (using a spatial weight matrix W) measures spatial correlation in attribute values y_i as:

$$I = \Sigma_i \, \Sigma_j \, w_{ij} \, (y_i - \mu) \, (y_j - \mu) / \Sigma_I (y_i - \mu)^2 \qquad\qquad (3.3)$$

where w_{ij} is the element in the spatial weight matrix W corresponding to the observation pair i,j, y_i and y_j are observations of land values for location i and j (with mean μ). Moran's I is similar but not equivalent to the correlation coefficient of $y_i y_j$ and is not centered around 0. In fact, the theoretical mean of Moran's I is -1/(N-1). In other words, the expected value is negative and is only a function of the sample size N. As the sample size increases, however, the mean will tend to zero. A Moran's I coefficient larger than its expected value indicates positive spatial autocorrelation, and a Moran's I less than its expected value indicates negative spatial autocorrelation.

3.5 The Traditional Spatial Autoregressive Model

According to Anselin (1988), the spatial econometric literature was developed to deal with the case of true spatial autocorrelation where an econometric model satisfies all the usual econometric assumptions but exhibits spatial autocorrelation. In this case, the model is correctly specified except for the failure to take spatial autocorrelation into account. Errors in the regression models show spatial dependence and the standard assumption of a spherical error covariance matrix fails to hold. Analogous to the treatment of temporal autocorrelation, the usual method for correcting spatial autocorrelation requires assuming a structure for spatial dependence and estimating one or more parameters of the structure in conjunction with the parameters of the economic model. Following Anselin (1988), the conventional autocorrelation problem is represented as a spatial autoregressive process with the standard regression models revised as:

$$Y = X\beta + u \qquad\qquad\qquad (3.4)$$

where

$$u = \lambda Wu + \varepsilon \qquad\qquad\qquad (3.5)$$

where Y is an Nx1 vector of observations on the dependent variable, X is an NxK matrix of explanatory variables, W is an NxN spatial weight matrix, β is the vector of parameters to be estimated, λ is a scalar to be estimated, ε is an Nx1 vector of random error terms with mean zero and variance-covariance matrix $\sigma^2 I_N$, and u is an Nx1 vector of random error terms with mean zero and non-spherical variance-covariance matrix $\sigma^2 (I_N - \lambda W)^{-1} (I_N - \lambda W')^{-1}$.

As noted earlier, the spatial weight matrix, W, contains information on the spatial dependence between pairs of errors. The i,jth element of the weight matrix, denoted as w_{ij}, represents the relative spatial dependence between the ith and the jth error. All the diagonal

elements of W are equal to zero. Obviously, each observation, i, can have as many as N-1 neighbors, but the relative influence of each neighbor must be specified, a priori, since these elements are impossible to estimate independently.

Based on a joint normal distribution for the error term $u \sim N(0, \Sigma)$ with $\Sigma = \sigma^2 [(I-\lambda W)'(I-\lambda W)]^{-1}$, the estimation of the parameters of interest $(\beta, \lambda, \sigma^2)$ comes from maximizing the likelihood function[3],

$$L(\lambda, \beta, \sigma^2) = \ln|I-\lambda W| - N/2 \ln(2\pi) - N/2 \ln(\sigma_\varepsilon^2) - [(Y - \lambda WY - X\beta + \lambda WX\beta)'$$
$$(Y - \lambda WY - X\beta + \lambda WX\beta)]/2\sigma_\varepsilon^2 \qquad (3.6)$$

The first order conditions yield the familiar generalized least squares estimator :

$$\beta_{ML} = [(X - \lambda WX)'(X - \lambda WX)]^{-1}(X - \lambda WX)'(Y - \lambda WY) \qquad (3.7)$$

and, similarly, the ML estimator for σ_{ML}^2 follows as:

$$\sigma_{ML}^2 = (e - \lambda We)'(e - \lambda We)/N \text{ when } e = Y - X\beta_{ML} \qquad (3.8)$$

It is argued that the above maximum likelihood procedure accounts for the problem of spatial autocorrelation, which existed in the initial model. In this sense, "reliable" results are obtained and the fitting power of the initial model improved. In the case that spatial autocorrelation still exists in the error terms of the autoregressive model, the accepted approach is to look more carefully at the spatial patterns in the data and define another weight matrix which captures better the structure of spatial dependence. The *ad hoc* nature of the selection of an appropriate spatial weight matrix is a widely acknowledged weak point of the traditional spatial econometric approach (Anselin 1988, Bockstael 1996).

---

[3] For analytical derivation of the likelihood function see Breusch (1980)

3.6 The Alternative Approach

The approach to spatial econometrics as just described can be viewed as a simple extension to what Spanos refers to as "traditional" or "textbook" time series econometrics (Spanos 1989, 1994). The criticisms of Spanos (1986) regarding the "textbook" approach are mainly focused on the confusion between theoretical and statistical issues in the "traditional" econometric thinking, the lack of a comprehensive set of statistical assumptions which support the statistical model, and a completely different philosophy in misspecification testing. This section does not attempt to summarize Spanos' views on econometrics, but briefly mentions some basic elements of Spanos' approach that could also apply in the case of spatial autocorrelation.

Spanos (1986) presents a set of consistent statistical assumptions underlying the linear regression model, which forms the backbone of most other statistical models. The validity of those assumptions is essential for the interpretation of the empirical results. The list of the assumptions defining the linear regression model is specified as following:

I] Statistical Generating Mechanism

$Y_s = X_s b + u_s$

1] $\mu_s = E(Y_s/X_s=x_s)$ is the systematic component; and $u_s = Y_s - E(Y_s/X_s=x_s)$ is the non-systematic component.

2] $\theta = (b, \sigma^2)$ are the parameters of interest

3] X_s is weakly exogenous with respect to θ

4] No a priori information on θ

5] $Rank(X_s) = k$

II] Probability Model

6]   i] $D(Y_s/X_s; \theta)$ is normal;

ii] Linearity in X_s

iii] Homoskedastic Variance

7] θ is invariant in s

III] Sampling Model

8] Identical sample is sequentially drawn from $D(Y_s/X_s; \theta)$

It is important to note that the index *s* refers to spatial ordering in the case of a cross-sectional data set. The above model assumptions (which are more than the assumptions mentioned in traditional spatial econometric studies) are all very closely connected to each other in the sense that the model assumptions are all about a conditional distribution that is derivable from a large joint distribution. If any one of the model assumptions is invalid it is likely that several of the model assumptions are invalid. Spanos (1986) clearly indicates that testing the validity of all the underlying assumptions is an issue of paramount importance, and argues that any particular misspecification test is only valid if all other assumptions underlying the model are valid. A test may result in too many or too few rejections if another assumption underlying the model is invalid. In particular for the case of autocorrelation type tests, Spanos (1986) notes that the assumption of no autocorrelation is usually rejected when other assumptions are violated, and also many other assumptions are usually rejected when there is autocorrelation in the statistical model.

It is primarily for these reasons that Spanos (1986) advocates testing and interpreting the assumptions as a whole. That is, a battery of individual and joint misspecification tests is essential at every step of the misspecification-respecification procedure. Rejection of the null hypothesis of one particular assumption should not be viewed as an acceptance of the alternative. Misspecification tests should provide only a rough guide as to whether or not empirical information supports the null hypothesis. Identification of misspecification sources should lead respecification efforts, until we obtain a statistical model that satisfies all necessary assumptions at the same time.

Although the above elements of Spanos' approach have not been applied to spatial data, they are clearly applicable. This approach to misspecification and model respecification

can potentially help isolate the source or sources of observed autocorrelation. Even when the data are spatially based we still need to do a battery of tests, interpret all results as for or against the null hypothesis, and try to figure out the most likely source of all symptoms highlighted by all different tests including observed spatial autocorrelation. Finally, testing the validity of the underlying model assumptions may help reduce the a*d hoc* nature of choice of the weight matrix. When the choice of the weight matrix leads us towards a statistically adequate model then its choice can be validated.

3.7 Conclusions

In this chapter we have introduced some basic tools that may be used in analyzing and testing patterns of spatial dependence. Fundamental to much of this material is the notion of a weight matrix, W, which captures the spatial relationship between a set of spatial units. A weight matrix has the role of averaging values at neighboring sampled data points. Several alternative definitions of a spatial weight matrix have been discussed. The researcher should use the type of weight matrix that is consistent with the available empirical information, but be aware of its *ad hoc* nature when interpreting her results.

This chapter also contains a description of the Moran *I* test, which is the most widely used test for spatial autocorrelation. Then, we presented the structure and the maximum likelihood estimators of the spatial error autoregressive model. Finally, we briefly introduced some basic elements of an alternative approach for econometric modeling, as proposed by Spanos (1986). This approach has not been applied to spatial data, however, its principles could help us understand sources of spatial autocorrelation and direct our misspecification and respecification efforts.

CHAPTER 4

Neighborhoods and Spatial Autocorrelation

4.1   Introduction

A search in electronic databases for academic work devoted to land value modeling will locate thousands of papers and dozens of books. The overwhelming majority of these documents do not examine the assumption of spatial autocorrelation in their empirical applications. It has only been during the last decade that studies have considered the possibility that land value models may suffer from spatial autocorrelation. This chapter will investigate whether land values in Roanoke County are spatially correlated, demonstrating the possibility that an econometric model, which does not consider the spatial configuration of the data may lead to biased and inconsistent estimators. Following current spatial econometric techniques, we initially estimate a land value model for Roanoke County, and then we examine whether spatial autocorrelation is

present. Prior to testing for spatial autocorrelation, the land value model for Roanoke County is tested for the assumptions of normality, heteroskedasticity and structural stability. In this way, we follow current spatial econometric practices and attempt to verify that the observed autocorrelation is "true"; and not due to violation of an underlying assumption (i.e. normality, heteroskedasticity and structural stability).

As we saw in the previous chapter, the construction of an appropriate weight matrix to account for neighborhood effects is necessary for testing spatial autocorrelation. A weight matrix can be based either on the Euclidean distances between centers of parcels or on existing information about designated neighborhoods in Roanoke County. Current spatial econometric studies construct weight matrices, which are based on mathematical computation of distances, and pay little attention to the particularities of the empirical information in the study area. The choice of a weight matrix can not be based on statistical tests, and the *ad hoc* nature of the matrix is widely acknowledged as a serious drawback of spatial econometric modeling. In the following sections, we will see that in our case study, the mechanically constructed matrix makes no sense and thus, would probably lead to erroneous conclusions. At the end of this chapter two different tests are conducted, both of which reject the hypothesis that the residuals of the model are spatially independent.

4.2   A Statistical Model That Assumes Spatial Independence

In this section, we try to create the best possible model assuming that the sample observations are not spatially correlated. In other words, we try to develop a land value model, which satisfies the "relevant" assumptions of current spatial econometric studies (Anselin 1988). Most of the "traditional" spatial econometric studies do not report misspecification tests results. Some papers, which report results of misspecification tests (Bockstael and Bell 1998), find violations of fundamental assumptions. However, they often justify these violations as due to the presence of spatial autocorrelation.

After performing a number of misspecification tests (see Table 4.1) for the assumptions of normality, heteroskedasticity and structural stability, the land value model is specified as follows:

$$\text{Log(Price)} = A_0 + A_1[\text{Log(Size)}] + A_2[\text{Log(Size)}]^2 + A_3[\text{Log(Elevation)}] + A_4[\text{Log(Elevation)}]^2 + A_5(\text{Soil1}) + A_6(\text{Soil2}) + A_7(\text{Population}) + A_8(\text{Population})^2 + A_9[\text{Log(Mall)}] + A_{10}[\text{Log(Mall)}]^2 + A_{11}[\text{Log(Town)}] + A_{12}(\text{Developed}) + A_{13}(\text{Road}) + A_{14}(\text{Year}) + A_{15}[\text{Log(X)}] + A_{16}[\text{Log(Y)}] + A_{17}\text{Log(X)Log(Y)} + u \qquad (4.1)$$

where Price is the price of the parcel per square meter, Size is the area of the parcel, Elevation is the average elevation of the parcel, Soil1 and Soil2 are dummy variables capturing soil quality, Population is the population density in the U.S. census block containing the parcel, Mall is the minimum distance to an existing mall, Town is the minimum distance to the closest town, the dummy variable Developed indicates whether the parcel is vacant, Road is another dummy variable which reveals whether the parcel is adjacent to a major Road, the dummy variable Year shows if the parcel was sold in 1996 or 1997, the Coordinates X and Y determine the exact location of the parcel and finally u represents the error term of the model.

Table 4.2 contains the OLS estimates of the land value model (4.1) for Roanoke County. If one were to assume neither spatial autocorrelation nor any other misspecification problems the results of the OLS estimation suggest this model explains approximately 80% of the variation in the land transaction prices. The value of a land parcel per square meter is expected to be lower for larger parcels. Parcels, which already have some type of residential or commercial development, have higher transaction prices. Lower water permeability (and consequently higher flood risk) affects negatively the parcel value, while a parcel sold in 1997 is expected to have higher value than a similar parcel sold in 1996. A careful analysis of the non-linear relations of the model and the value range of the variables indicates that longer distance from the closest mall as well as higher elevation and lower population density affect positively the land transaction prices but at

a decreasing rate. The effect of land size on price is opposite to that found in some previous empirical studies (Xu et al, 1993).

Table 4.1 includes the misspecification tests that are usually performed in current spatial econometric studies. The Jacque-Bera test rejects the null hypothesis that the errors are normally distributed. However, it is well known that this test is very sensitive to outliers. After dropping some of the extreme sample observations (about 2%) we realized that outliers were driving the rejection of the normality hypothesis. The P-value of the White test supports the hypothesis of homoskedasticity. A series of Chow tests also indicates that the model is unlikely to have some structural break. Model observations are ordered by neighborhood according to the numerical code of the Roanoke County Planning Commission. Following the existing spatial econometric literature (Anselin 1988), we can conclude that if we find observed spatial autocorrelation in the econometric model then this spatial autocorrelation is the result of omitting relevant spatial characteristics.

Table 4.1 Misspecification Tests for the Land Value Model for Observations in Roanoke County

Test	Null Hypothesis	Specification	P-Value
Jacque-Bera (Normality)	Residuals are normally distributed	$JB = (N-k)(4S^2 + (K-3)^2)/24$ S is the skewness, K is the Kurtosis, and N-k are the degrees of freedom	0.000000
White Test (Heteroskedasticity)	Homoskedasticity	$u^2 = c + bx^2 + dy$ u is the vector of residuals, c is a constant, x is the vector of variables, y is the vector of cross-product variables	0.762358
Chow Test (Structural Break)	Existence of structural change (Breakpoint n = 300)	F statistic based on the comparison of restricted and unrestricted sum of square residuals.	0.853459
Chow Test (Structural Break)	Existence of structural change (Breakpoint n = 600)	F statistic based on the comparison of restricted and unrestricted sum of square residuals	0.800440
Chow Test (Structural Break)	Existence of structural change (Breakpoint n = 900)	F statistic based on the comparison of restricted and unrestricted sum of square residuals	0.281454
Chow Test (Structural Break)	Existence of structural change (Breakpoint n = 1200)	F statistic based on the comparison of restricted and unrestricted sum of square residuals	0.521240

*For analytical discussion of misspecification tests see Appendix 1.

Table 4.2 OLS Estimates for the Land Value Model in Roanoke County

Variable	Coefficient	Std. Dev.	t-ratio
Constant	-17.46850	3.214461	5.434
Log(Size)	-0.483947	0.069485	6.964
$[Log(Size)]^2$	-0.030618	0.009440	3.243
Log(Elevation)	0.337926	0.274165	1.233
$[Log(Elevation)]^2$	-0.106225	0.065750	1.616
Soil1	-0.056682	0.019007	2.982
Soil2	-0.091607	0.036173	2.532
Population	0.004845	0.004217	1.149
$(Population)^2$	-0.000059	0.000023	2.571
Log(Mall)	1.402944	0.417148	3.363
$[Log(Mall)]^2$	-0.220563	0.057922	3.808
Log(Town)	0.250346	0.068201	3.671
Developed	0.094025	0.015405	6.103
Road	-0.070932	0.022242	3.189
Year	0.056391	0.009418	5.987
LogX	4.190094	0.732566	5.719
LogY	3.811132	0.695302	5.481
(LogX)*(LogY)	-0.930265	0.167058	5.569
R^2		0.809	
Adjusted R^2		0.807	

4.3 Neighboring Effects and Weight Matrices

The testing of spatial independence requires the creation of at least one spatial lag. In the case of a land value or a housing model the notion of a spatial lag is synonymous with the neighborhood around the parcel or the house, and spatial effects are usually referred to as neighboring effects. The magnitude of the neighborhood depends on the population density and the structure of observations. For example, the area of a neighborhood in New York City is expected to be much smaller than a neighborhood in rural Arizona. The purpose of a weight matrix is to capture the neighboring effects of the land parcels. Spatial proximity, which may generate neighboring effects, can be identified through many ways. Perhaps the most popular is the use of interactive GIS maps and defining contiguity between spatial units as a function of the distance that separates them. In this case, the distance is usually computed between the geometrical centers of the parcels. Two units are then considered contiguous if these points are less than a pre-specified distance apart. Descriptive statistics of the distances among the centers of the parcels are usually computed to enhance our understanding of the study area. Table 4.3 presents the descriptive statistics of parcel distances in Roanoke County.

Table 4.3 Descriptive Statistics of Distances between Parcel Centers in Roanoke County

Number of parcels	1,844.00
Average Distance(meters)	11,076.39
Standard Deviation(meters)	6,439.35
Minimum Distance(meters)	6.53
Maximum Distance(meters)	37,078.10
Minimum Contiguity(meters)	2,781.95

Perhaps the most important information in Table 4.3 is related to the minimum distance required for the creation of a spatial weight matrix, which is indicated by the minimum contiguity. Minimum contiguity is defined as the minimum distance necessary for every parcel to have at least one neighbor. The minimum diameter of a spatial lag in Roanoke County should be at least 2,782 meters. A smaller distance would result in a lack of contiguity for a number of parcels and thus, in the creation of discontinuous spatial variables. This is because each parcel must be contiguous with another parcel in order to create a spatial lag, and each row of a weight matrix must have at least one non-zero element.

However, a neighborhood of approximately 3 km diameter may be not sensitive enough to capture neighborhood effects among land parcels. In the case of large distances among parcels, Anselin (1988) suggests the use of a non-linear relationship between distance and neighborhood effects for the creation of the appropriate spatial weight matrix. *Ad hoc* determinations of different functional forms can lead to the construction of different weight matrices. The optimal weight matrix is usually selected through a series of non-nested tests. A major disadvantage of this procedure is that it fails to take into consideration the natural borders of different neighborhoods as well as other spatial particularities of the study area.

Fortunately, the Tax Department of Roanoke County has classified the land parcels of the study area into different neighborhoods. The criteria used for this classification are: the geographic proximity of spatial units, level of economic development, geographical and morphological elements as well as conventional and administrative definitions of "neighborhood" from other departments of the local government. There are 164 different neighborhoods in our sample, and each neighborhood contains an average of 12 land parcels included in the sample. The magnitude of these neighborhoods is not the same. Some of these neighborhoods close to the town of Roanoke have a diameter smaller than 0.3 Km to capture different types of residential development, while neighborhoods at the borders of the Roanoke County are large enough to capture similar characteristics of

remote parcels. These neighborhoods can be used to define spatial lags for our case study, and a weight matrix can average the value of the neighboring land parcels in each defined neighborhood.

4.4 Testing Spatial Autocorrelation in Roanoke County

Let W be the weight matrix based on the existing neighborhood information. Two different tests have been selected to test the hypothesis of no spatial autocorrelation in the land value model (4.1) for Roanoke County. The first test is the Moran's I test, which is the most common in spatial statistics and is demonstrated in the previous chapter of the thesis. The results of the Moran's I test indicate the existence of spatial autocorrelation. The statistic I equals 0.928 and Z_I equals 5.54, which provides evidence for the rejection of our hypothesis of no spatial autocorrelation (P-value less than 0.001).

Spatial autocorrelation can also be tested through the use of the following auxiliary regression test:

$$u = Xb + kWu + \varepsilon \qquad (4.2)$$

where u and X are the residuals and the explanatory variables of equation (4.1), b and k are the estimated coefficients, W is the weight matrix and ε is the error term of equation (4.2). The null hypothesis is that H_0: k = 0 against H_1: k $\neq$ 0. The F-test for H_0 provides evidence against the null hypothesis (F-statistic = 17.98, P-value less than 0.001). Consequently, both tests indicate the existence of spatial autocorrelation and that the model (4.1) is not well specified.

4.5 Conclusions

This chapter seems to provide evidence in support of the first proposition of this dissertation, which states that there is spatial dependence in the value of land parcels in Roanoke County. Following current spatial econometric studies we tested the hypotheses of normality, heteroskedasticity and structural stability in the land value model of Roanoke County, and then two additional tests provided evidence of spatial autocorrelation. In the next chapter, we will attempt to "cure" spatial autocorrelation following existing econometric practices. However, the sixth chapter of the thesis will show that the observed spatial autocorrelation may be the result of other model misspecification.

There is one more interesting observation regarding the results of this chapter. Using the statistical package Spacestat, we calculated that the diameter neighborhood in Roanoke County should be at least 2.8 km. However, such a large neighborhood would be clearly inappropriate to capture spatial effects in Roanoke County. In addition this neighborhood would not consider natural frontiers and other socioeconomic parameters essential to determine spatial relationships. Neighborhoods defined by the Planning Department of Roanoke County are built based on the geographical particularities of Roanoke County. Some of these neighborhoods have diameter smaller than 0.3 km, while others are big enough to include common types of residential development at the county borders. One should be very careful using neighborhoods defined numerically via Spacestat, Gauss or Matlab; it may well be that the implied specification of W is completely inappropriate for the data in question and, thus, the results obtained with this specification questionable. It is worth noticing that the overwhelming majority of existing spatial econometric studies use weight matrices calculated by the above statistical packages, while no discussion is included about the appropriateness of the weight matrix.

CHAPTER 5

The Traditional Spatial Econometric Approach

5.1 Introduction

Given the existence of observed spatial autocorrelation in the land value model (4.1) for Roanoke County, this chapter attempts to account for the existence of spatial autocorrelation using the latest accepted practices. The third chapter of the thesis described the use of maximum likelihood techniques to account for error spatial autocorrelation. The existing spatial econometric literature argues that this methodology alleviates the problem of spatial autocorrelation and produces reliable estimates (Anselin, 1988). During the last decade, maximum likelihood estimation for spatial autoregressive models has become common-place for empirical applications. The first studies utilizing this approach included public finance and socio-economic models of the behavior of different geographical regions and the spillover effects between them. This empirical work was characterized by a small numbers of observations and the spatial relationships were generally those of adjacent neighbors (i.e. counties with common borders). The development and increasing availability of GIS data sets, however, has changed the nature of spatial econometric applications. In the 1990's studies of houses, firms and land

parcels dominated the spatial econometric literature. In contrast to the initial applications, these studies are characterized by large data sets and declining functions of distances are used to measure spatial relationships.

While these changes may seem inconsequential, the size of sample has had an important impact on both the properties and the feasibility of the maximum likelihood estimator. Several studies (Pinkse and Slade, 1998; Kelejian and Prucha, 1999) have observed that maximum likelihood techniques become increasingly problematic as sample size grows. This is because the calculation of the eigenvalues of the spatial weight matrix may become infeasible as the size of the matrix increases. In addition, Kelejian and Prucha (1997) proved that eigenvalues of spatial weight matrices of dimension over 400 could not be calculated reliably. This means that even if we manage to derive the maximum likelihood estimates of a spatial autoregressive model, the solution may not be reliable.

In recent papers, alternative estimation techniques of the spatial econometric model are proposed. These techniques can be categorized into parametric and non-parametric approaches. Parametric techniques focus on creating an alternative autoregressive model, which achieves estimators with optimal properties through the use of instrumental variables. The parametric techniques are characterized by simplicity, they require limited computing capacity and they are less mathematically complicated in comparison to non-parametric techniques. The only disadvantage of the parametric techniques seems to be the arbitrary choice of the instrumental variables. Studies that use two or three stage least squares techniques (Land and Deane, 1992; Kelejian and Prucha, 1998) recognize that the efficiency of their instrumental variable estimator relies on the proper choice of the instruments.

The success of non-parametric techniques is based on a "low cost" means of obtaining parameter estimates, which are close to the maximum likelihood estimates. In other words, estimates with values very close to the actual maximum likelihood estimates can be achieved faster and with limited computing capacity, while the calculation of the

actual maximum likelihood estimates requires calculating the eigenvalues (which are not always reliable and demand very high computing capacity) of very large matrices. The validity of the parametric and non-parametric approaches can be demonstrated by the similarity of the maximum likelihood estimates to the non-parametric estimates (Anselin, 1999; Bockstael and Bell, 1998). Kelejian and Prucha (1998) provide Monte Carlo results for contiguity-type weight matrices and samples up to 400 observations, which suggest that non-parametric estimators, for these problems, are virtually as efficient as the maximum likelihood estimates.

This chapter applies both parametric and non-parametric techniques to the land value model of Roanoke County in an attempt to capture the observed spatial autocorrelation. The results indicate that both techniques achieve higher fitting power (based on R^2) than the initial model estimated in the last chapter. Higher fitting power implies that by using spatial lags, the model explains a larger portion of the variance of the dependent variable. However, the statistical validity of these models (as used in current spatial econometric studies) has not been tested, and the results rely on the assumption that the model is well specified. That is, we assume the models adequately capture the existing spatial autocorrelation and that no other specification problems exist.

Before ending this chapter, we attempt to make sure that the parametric and non-parametric models are well specified by conducting the misspecification tests applied in the existing spatial econometric literature. The results indicate that there is still spatial autocorrelation and in addition the assumption of homoskedasticity is now violated. Given these assumption violations, we can conclude that by applying the tools of current spatial econometric methodology we have not adequately modeled land values. Anselin (1988) argues that a failure of spatial econometric techniques to account for spatial autocorrelation can sometimes be attributed to the choice of an inappropriate weight matrix. He also suggests that the researcher should continue trying alternative weight matrices until the misspecification problem is solved. However, we saw in Chapter 4 that the weight matrix used in this chapter relies on detailed information provided by the

Planning Department of Roanoke County, and matrices based on distances among the parcel centers make no sense in our case study. Thus, instead of trying various *ad hoc* definitions of weight matrices, we will attempt modeling land values using an alternative approach; a more comprehensive set of misspecification tests which will guide our respecification efforts.

5.2 Parametric Techniques

Research (Land and Deane 1992, Kelejian and Robinson 1993 and Kelejian and Prucha 1998) suggests that two stage least squares (2SLS) estimators of the following spatial autoregressive model are consistent and asymptotically normal[4]:

$$Y = c\,W\,Y + X\,b + e \qquad\qquad (5.1)$$

and

$$W\,Y = k\,W\,X\,b + u \qquad\qquad (5.2)$$

where W is the weight matrix, c is the spatially autoregressive coefficient, WY is the spatial lag of land prices, WX is the set of instruments and e is a vector of error terms. Anselin (1988) proves that the spatial lag term WY in (5.1) is always correlated with the error term e. Consequently, the spatial lag should be treated as an endogenous variable and estimation should account for this endogeneity (OLS estimators will be biased and inconsistent due to simultaneity error). Most spatial econometric studies (for literature review see Kelejian and Prucha 1999) agree that a theoretically sound choice of an instrument for WY would include the set of lagged independent variables WX. According

---

[4] Notice that parametric techniques do not use a spatial error autoregressive model, but rather an autoregressive model of the dependent variable (Kelejian and Robinson, 1993), and thus model specification varies slightly from the spatial autoregressive model proposed in Chapter 3.

to Anselin (1999), the spatial two stage least square estimator based on this set up is consistent and asymptotically normal.

Table 5.1 summarizes the spatial two stage least squares estimates for the land value model in Roanoke County. The signs and the values of almost all coefficients in this model are similar to the sign and values of coefficients of model (4.1) in the previous chapter. Larger undeveloped parcels with impermeable soil, located far from a shopping mall and close to town center are expected to have lower transaction prices per square meter. A major highway attached to the lot or high population density also reduces land values in Roanoke County. Transaction prices also increased in 1997 relative to the previous year. Finally, the positive sign of WPRICE (average value of other parcels in the "neighborhoods defined by the Roanoke County Planning Department) indicates the value of a parcel will increase as the land values of neighboring parcels increase. This parametric model captures spatial dependence among prices in Roanoke County and its fitting power exceeds 81%.

However, a set of misspecification tests summarized in Table 5.2 implies serious violations of fundamental statistical assumption for the estimated parametric model. The White test indicates that the error terms of the model are not homoskedastic, despite the fact that the assumption of homoskedasticity was satisfied in the initial model. In addition, auxiliary regression test indicates that there is almost no support for the assumption of no spatial autocorrelation. Given that this model was constructed to deal with the appearance of spatial correlation in model (4.1) of the previous chapter, it obviously failed its mission. Table 5.2 implies that the parameter estimates described in Table 5.1 must be interpreted with caution.

Table 5.1 Spatial Two Stage Least Squares Estimates for the Spatial Autoregressive Land Value Model in Roanoke County

Variable	Coefficient	Std. Dev.	t-ratio
Constant	-14.025794	3.342787	4.201
WPRICE	0.224864	0.021740	10.18
Log(Size)	-0.358954	0.073127	4.943
$[Log(Size)]^2$	-0.040581	0.009814	4.214
Log(Elevation)	0.186547	0.282630	0.663
$[Log(Elevation)]^2$	-0.048758	0.067642	0.842
Soil1	-0.030005	0.023742	1.672
Soil2	-0.066687	0.037231	1.830
Population	-0.002354	0.004381	0.450
$(Population)^2$	-0.000186	0.000241	0.808
Log(Mall)	0.835478	0.436488	1.894
$[Log(Mall)]^2$	-0.112562	0.060764	2.189
Log(Town)	2.845821	0.621488	4.160
Developed	0.135478	0.015608	8.157
Road	-0.058667	0.022836	2.653
Year	0.053485	0.009699	5.442
LogX	3.316587	0.755211	4.393
LogY	3.026457	0.716423	4.217
(LogX)*(LogY)	-0.655475	0.172219	4.278
pseudo-R^2		0.815	

Table 5.2 Misspecification Tests for the Spatial Autoregressive Land Value Model in Roanoke County Using the Two Stage Least Squares Estimator

Test	Null Hypothesis	Specification	P-Value
Jacque-Bera (Normality)	Residuals are normally distributed	$JB = (N-k)(4S^2 + (K-3)^2)/24$ S is the skewness, K is the Kurtosis, and N-k are the degrees of freedom	0.000000
White Test (Heteroskedasticity)	Homoskedasticity	$u^2 = c + bx^2 + dy$ u is the vector of residuals, c is a constant, x is the vector of variables, y is the vector of cross-product variables	0.000000
Auxiliary Regression (Spatial Autocorrelation)	No spatial autocorrelation (ordering according to neighborhoods)	$u = c + ax + bWu$ u is the vector of residuals, c is a constant, x is the vector of variables, W is the weighting matrix	0.000000
Chow Test (Structural Break)	Existence of structural change (Breakpoint n = 600)	F statistic based on the comparison of restricted and unrestricted sum of square	0.697544

*For analytical discussion of misspecification tests see Appendix 1.

5.3 Non-Parametric Methods

The Generalized Moments Estimator (GME) developed by Kelejian and Prucha (1998), should not be confused with the Generalized Methods of Moments (GMM) estimator developed by Hansen (1982). From the perspective of Kelejian and Prucha, the GME should be used in large samples when the traditional maximum likelihood estimator for a spatial autoregressive model is problematic. Theoretically, the generalized moments estimator achieves the optimal properties of the maximum likelihood estimator. It requires some matrix multiplication but it involves neither the calculation of the determinant nor the eigenvalues of the weight matrix, and so is accessible to most applied econometricians[5]. Bockstael and Bell (1988) argue that the generalized moments estimator has the additional advantage that it is consistent irrespective of whether the errors follow a normal distribution. The GME is based on the three moments of the error term, u, in the following error autoregressive model:

---

[5] Most of the statistical packages can do the estimation in microcomputers of memory capacity.

$$Y = X\,b + e \tag{5.3}$$

and

$$e = c\,We + u \tag{5.4}$$

Kelejian and Prucha (1998) derive the following moment conditions from assumptions underlying the model (5.3) and (5.4) and the fact that the diagonal elements of W are always set to zero:

$$E\,[u'u\,/\,N] = \sigma^2 \tag{5.5}$$

$$E\,[u'W'Wu] = \sigma^2\,N^{-1}\,Tr(W'W) \tag{5.6}$$

$$E\,[u'W'u\,/\,N] = 0 \tag{5.7}$$

Those moments can be rewritten as:

$$E\,[e'(I - c\,W)'(I - c\,W)\,e\,/\,N] = \sigma^2 \tag{5.8}$$

$$E\,[e'(I - c\,W)'W'W(I - c\,W)\,e\,/\,N]] = \sigma^2\,N^{-1}\,Tr(W'W) \tag{5.9}$$

$$E\,[e'(I - c\,W)'W'(I - c\,W)\,e\,/\,N]] = 0 \tag{5.10}$$

A three-equation system can then be specified in terms of predictors of e, denoted ê. For simplicity, we note ė = W ê and ë = WWê. The system is presented below:

$$2\hat{e}'\ddot{e}/N \qquad c \qquad - \dot{e}'\dot{e}/N\ c^2 + \qquad\qquad \sigma^2 - \hat{e}'\hat{e}/N + v_1 = 0$$

$$2\ddot{e}'e/N \qquad c \qquad - \ddot{e}'\ddot{e}/N\ c^2 + \mathrm{tr}(W'W)/N\ \sigma^2 - \dot{e}'\dot{e}/N + v_2 = 0 \qquad\qquad (5.11)$$

$$(\hat{e}'\ddot{e} + e'\dot{e})/N \quad c \qquad - \dot{e}'\dot{e}/N\ c^2 \qquad\qquad\qquad - e'\dot{e}/N + v_3 = 0$$

where c, c^2, and σ^2 are parameters to be estimated, $v_1, v_2,$ and v_3 are residual terms of the system and N are the sample observations. Using the OLS residuals as predictors of e, the system can be solved using non-linear least squares, imposing the necessary additional restriction between the parameters c and c^2. Once the estimate for c is obtained, estimates for b (*b*) and σ^2 (*σ^2*) are derived using feasible generalized least squares. So, for example,

$$b = (X^{*'}X^*)^{-1}(X^{*'}Y^*) \qquad\qquad (5.12)$$

where $X^* = (I - c\ W)X$ and $Y^* = (I - c\ W)Y$. In case that $c = 0$, then the OLS estimator coincides with the GM estimator.

This approach is more appealing to researchers because it does not depend on the normality of errors and the estimators are equivalent to the maximum likelihood estimator under normality[6]. In fact, the generalized moments estimator procedure results are similar to those attained from maximum likelihood estimation, but the two methods are based on different assumptions. The maximum likelihood is derived under the assumption of normally distributed errors, while the generalized moments method assumes that the three moments of the error term, e, of the classic autoregressive model (5.3) and (5.4) exist. Kelejian and Prucha (1998) underline the fact that the generalized moments method follows the "reverse" procedure from the maximum likelihood methodology to arrive at similar results. In other words, the generalized moments method estimates *c* by first calculating the terms in the system of equation (5.11), which depend on the elements of

---

[6] Spanos (1986) argues that econometric methods which include the use of moments should be approached with caution. A first reason is that these moments may not exist, and secondly the derivation of equations (5.5) to (5.7) may rely on different distributional assumptions than equations (5.8) to (5.10).

the weight matrix W and the residuals from the OLS regression. In this case study, the weight matrix W was calculated by using neighborhoods indicated by the Roanoke County Planning Department. Then, the estimation of *c* is easy by using non-linear least squares procedures. In the case of maximum likelihood estimation, the estimate of *c* is obtained by first creating the likelihood function, and then searching for the *c* that maximizes its value.

Table 5.3 summarizes the generalized moments estimates for the spatial autoregressive land value model in Roanoke County. This model explains about 88% of the variation in land value prices. The fitting power of this model is higher than the model estimated using two stage least squares however, the signs of almost all parameter coefficients remain the same. The only exception is that the results now indicate higher population density leads to higher land transaction prices, after evaluating linear and non-linear terms for the range of data in Roanoke County. This result is more compatible with empirical evidence from other land value studies and in this model the coefficients of population density are statistically significant (Xu F. *et al*., 1993). The value of parameter c, which could be interpreted crudely as a measure of autocorrelation, is around 0.78 and confirms the presence of spatial error autocorrelation in model (4.2)[7]. According to the traditional spatial econometric methodology (Kelejian and Robinson, 1999), the success of this model can be demonstrated by the fact that it captures existing spatial dependence (caused by omitted variables) and increases its fitting power.

However, the set of misspecification tests summarized in Table 5.4 suggests that the above model is statistically inadequate and thus the parameter estimates should be treated with caution. This model was constructed to "cure" spatial autocorrelation, however the statistical tests indicate no improvement over the initial model (4.1). Given the violation of fundamental assumptions, the fitting power of the model should not be a criterion of model improvement (McGuirk and Driscoll 1995).

---

[7] The value of c should be higher than zero and less than one. When c equals to zero there is no spatial autocorrelation, while c equal to one indicates the existence of a spatial unit root (Anselin 1988).

Table 5.3 Generalized Moments Estimates for the Spatial Autoregressive Land Value Model in Roanoke County

Variable	Coefficient	Std. Dev.	t-ratio
Constant	-8.587449	0.495045	7.398
Log(Size)	-0.318225	0.057487	5.864
$[Log(Size)]^2$	-0.047040	0.008206	7.818
Log(Elevation)	0.419223	0.108988	3.831
$[Log(Elevation)]^2$	-0.145284	0.022863	6.461
Soil1	-0.059056	0.014836	3.981
Soil2	-0.061578	0.030554	2.110
Population	0.009975	0.003375	2.980
$(Population)^2$	-0.00090	0.000185	4.926
Log(Mall)	1.772356	0.319612	5.585
$[Log(Mall)]^2$	-0.275935	0.044456	6.207
Log(Town)	2.135478	0.473028	4.484
Developed	0.047611	0.016766	2.920
Road	-0.142565	0.026328	4.665
Year	0.014578	0.015241	1.022
LogX	4.129325	0.528205	7.818
LogY	3.758745	0.496919	7.564
(LogX)*(LogY)	-0.925445	0.118783	7.781
c	0.784251		
σ^2	0.003854		
pseudo-R^2	0.882		

Table 5.4 Misspecification Tests for the Spatial Autoregressive Land Value Model in Roanoke County Using the Generalized Moments Estimates

Test	Null Hypothesis	Specification	P-Value
Jacque-Bera (Normality)	Residuals are normally distributed	$JB = (N-k)(4S^2 + (K-3)^2)/24$ S is the skewness, K is the Kurtosis, and N-k are the degrees of freedom	0.000000
White Test (Heteroskedasticity)	Homoskedasticity	$u^2 = c + bx^2 + dy$ u is the vector of residuals, c is a constant, x is the vector of variables, y is the vector of cross-product variables	0.000000
Auxiliary Regression (Spatial Autocorrelation)	No spatial autocorrelation (ordering according to neighborhoods)	$u = c + ax + bWu$ u is the vector of residuals, c is a constant, x is the vector of variables, W is the weighting matrix	0.000000
Chow Test (Structural Break)	Existence of structural change (Breakpoint n = 600)	F statistic based on the comparison of restricted and unrestricted sum of squares	0.236569

*For analytical discussion of misspecification tests see Appendix 1

5.4. Conclusions

Using both parametric and non-parametric techniques, we attempted to correct for the problem of spatial autocorrelation in the land value model (4.1) for Roanoke County. The results indicate that these techniques did not manage to "cure" the problem of spatial autocorrelation. The next chapter uses an alternative approach, which looks more carefully into the sources of spatial autocorrelation.

CHAPTER 6

An Alternative Approach for Land Value Modeling

6.1 Introduction

In the previous chapters, current spatial econometric techniques were used to model land values in Roanoke County. Violation of fundamental statistical assumptions indicates that the proposed modeling techniques have not captured the relevant systematic information in our data. In this chapter, we follow an alternative approach (proposed by Spanos (1986)) in an attempt to derive a statistically adequate model for land values in Roanoke County. This approach does not seek to "cure" the problem of spatial autocorrelation or any other individual assumption violation one at a time. Individual and joint misspecification tests are conducted to identify misspecification sources. Rejection of a hypothesis is not interpreted as an acceptance of the alternative, and more importantly the

results of misspecification tests are interpreted as a whole. In other words, we use misspecification tests as a rough guide towards identifying problems and particularities of the empirical information rather than trying to fix problems from misspecification assumptions one-by-one. Our first step in this chapter is to go back to model (4.1) specified in the fourth chapter and conduct a more comprehensive set of individual and joint misspecification tests. An iterative procedure of respecification and testing will lead us to the adoption of our final models.

6.2 Land Value Model Revisited

Table 6.1 summarizes the results of a set of individual and joint misspecification tests for the land value model (4.1). These tests examine all testable assumptions for the linear regression model defined in the section 3.6. In the fourth chapter, we saw that individual tests provided strong support for the assumptions of homoskedasticity and structural stability, while there was no evidence that the residual terms were spatially independent. Also, extreme observations were probably responsible for the low P-value of the Jacque-Bera test for normality. Table 6.1 contains two new tests for linearity, which indicate that non-linear (squared and cross-product) variables are not essential for the land value model. The Ramsey test also confirms that the functional form of the model is adequate for our data. The ARCH test provides no support for the null hypothesis that there is no second-order spatial dependence. Thus, the residual terms of the land value model seem to exhibit first (of the means) and second (of the variance) order spatial dependence. The joint tests in Table 6.1 confirm that the hypotheses of linearity, structural stability and no spatial dependence do not hold jointly. Similarly, the joint variance test indicates the hypotheses of homoskedasticity, structural stability and second order dependence are not supported by our data. The parcels are ordered by neighborhood and then by development status (vacant parcels and then developed parcels), while developed parcels are also ordered using the assessed value of the construction on the parcels. Both individual and joint misspecification tests (Auxiliary Regression, ARCH, First Joint Mean and Joint Variance test) indicate that there is no support for the existence of structural breaks in the

structure of the land value model (4.1). Despite the evidence of structural stability, parameters (b and σ^2) may vary across neighborhoods. Table 6.1 includes the results of both individual and joint misspecification tests (Fixed Effects from Neighborhoods and Second Joint Mean Test) for the null hypothesis that parameters vary across neighborhoods.

The results of the above misspecification tests indicate that spatial autocorrelation is probably the most serious problem in the land value model (4.1). In the joint mean test for the hypotheses of linearity, no spatial autocorrelation and structural stability, we notice that spatial autocorrelation has the lowest P-value in the joint test. Similarly, second order dependence seems to be the main reason for the rejection of the joint hypothesis in the joint variance test. At the same time the low P-values of the no neighborhood fixed effects hypothesis and the joint hypothesis of no spatial autocorrelation and no neighborhood fixed effects provide evidence against the hypothesis that parameters are stable across neighborhoods (Second Joint Mean Test). In the joint mean test of no spatial autocorrelation and no neighborhood fixed effects, we notice that both individual hypotheses probably lead to the violation of the joint hypothesis. Given that several spatial econometric studies (Anselin 1999) underline that missing neighborhood specific variables are often the source of spatial autocorrelation, it seems appropriate to add a set of neighborhood dummies in the land value model (4.1). After estimating the fixed effects land value model accounting for neighborhood effects we retest the model using the same set of misspecification tests.

Table 6.1. Misspecification Tests* for the Land Value Model (4.1) for Roanoke County

Test	Null Hypothesis	Specification	P-Value
Jacque-Bera (Normality)	Residuals are normally distributed	$JB = (N-k)(4S^2 + (K-3)^2)/24$ S is the skewness, K is the Kurtosis, and N-k are the degrees of freedom	0.000000
Linearity (Squares)	Redundancy of non-linear (squared) variables	$u = c + ax + bx^2$ u is the vector of residuals, c is a constant, x is the vector of variables	0.863646
Linearity (Cross-Products)	Redundancy of non-linear (cross- product) variables	$u = c + ax + by$ u , c, x as described above and y is the vector of cross-product variables	0.154613
White Test (Heteroskedasticity)	Homoskedasticity	$u^2 = c + bx^2 + dy$ u, c, x, as described above and y is the vector of cross-product variables	0.762358
Auxiliary Regression (Spatial Autocorrelation)	No spatial autocorrelation (ordering according to neighborhoods)	$u = c + ax + bWu$ u , c, x as described above and W is the weight matrix	0.000000
Ramsey RESET (Incorrect Functional Form)	Correct specified functional form of the model	$u = c + ax + bm$ u, c, x as described above, and m is the vector of fitted values of x	0.932831
ARCH Test (Dependence in Variance)	No dependence in residual variance	$u^2_z = c + au^2_{z-1} + bu^2_{z-2} + du^3_{z-3}$ u, c, x as described above and z is the ordering factor	0.000000
Chow Test (Structural Break)	Existence of structural change (break point n = 213, 715)	F statistic based on the comparison of restricted and unrestricted sum of square residuals	> 0.2
Fixed Effects from Neighborhoods	No neighborhood fixed effects	$u = c + au_s$ u, c defined as above, and u_s is the residual average at a given neighborhood	0.000000
First Joint Mean Test	Linearity, no spatial autocorrelation and structural stability. (break point n = 213)	$u = c + ax + bx^2 + dWu + kT$ u, x, and W as described above and T is a binary variable with 0 before the break point and 1 after	0.000000
- No Spatial Autocorrelation	No spatial autocorrelation (in the joint mean test)	$u = c + ax + bx^2 + kT$ u, x, W, and T as described above	0.000000
- Structural Stability	Existence of structural change	$u = c + ax + bx^2 + dWu$ u, x, and W as described above	0.086541
- Linearity	Redundancy of non-linear variables (break point as above)	$u = c + ax + dWu + kT$ u, x, W, and T as described above	0.401531
Joint Variance Test	Homoskedasticity, no second order dependence and structural stability (break point as above)	$u^2_z = c + ax + bx^2 + du^2_{z-1} + kT$ u, c, x, W, z, and T as described above	0.000000
- No Second Order Dependence	No dependence in residual variance (in the joint variance test)	$u^2_z = c + ax + bx^2 + kT$ u, c, x, W, z, and T as described above	0.000000
- Structural Stability	Existence of structural change (in the joint variance test)	$u^2_z = c + ax + bx^2 + du^2_{z-1}$ u, c, x, W, z, and T as described above	0.165318
- Homoskedasticity	Homoskedasticity (in the joint variance test)	$u^2_z = c + ax + bx^2 + du^2_{z-1} + kT$ u, c, x, W, z, and T as described above	0.555664
Second Joint Mean Test	No spatial autocorrelation, no neighborhood fixed effects	$u = c + au_s + bWu$ u, c, us, and W defined as above	0.000000
- No Spatial Autocorrelation	No spatial autocorrelation (in the joint mean test)	$u = c + au_s$ u, x, W, and T as described above	0.000000
- Fixed Effects from Neighborhoods	No neighborhood fixed effects (in the joint mean test)	$u = c + bWu$ u, c, us, and W defined as above	0.000000

*For analytical discussion of misspecification tests see Appendix 1.

The OLS estimates of the fixed effects model and the conclusions of the new set of misspecification tests are presented in Table 6.2 and Table 6.3, respectively. The fixed effects model accounts for neighborhood effects, based on neighborhoods indicated by the Planning Department in Roanoke County. The fixed effects model was created by deducting from all variables their average values within each neighborhood. In Table 6.3, we can see an improvement in the P-value (Auxiliary Regression and Joint Mean Test) of the hypothesis of no spatial autocorrelation. An improvement can also be noticed in the P-value of the ARCH test, however there is still significant evidence of second order dependence. The Chow tests using the same ordering (by development status, assessed value of constructions, and by neighborhood) now indicate the existence of structural breaks. In the joint mean test, we examine the joint hypothesis of linearity, no spatial autocorrelation and structural stability. The results indicate that there is no support for this joint hypothesis, while the low P-value of the structural stability hypothesis seems to be the reason for this rejection. Similarly, the low P-value of structural stability is probably again the reason for the low P-value in the joint variance test. Given the results of the joint tests, it seems that structural instability may be the major source of misspecification in the fixed effects model.

In Table 6.3, we notice that in the fixed effects model there is strong evidence for a structural break between developed and vacant parcels (P-value of the Chow test for n = 213 is close to zero). Several plots of recursive OLS estimates that appear in Figure 6.1 also provide additional evidence for structural instability. In Figure 6.1, c(1), c(2),… represent the value of the coefficient estimates of the parameters as ordered in Table 6.2, and the value of a coefficient estimate at a specific point (i.e. n = 213) in the graph represents the value of the value of this coefficient estimate when we use the first n observations from our sample[8]. We can see in the graphs of several coefficients, substantial change in the magnitude of the coefficient estimates for several variables after the first 213 observations of the vacant parcels.

---

[8] Specifically, the coefficient estimates shown are: C(1) – Log(Size), C(2) – [Log(Size)]^2, C(5) - Log (Elevation), C(7) – Soil 2, C(8) – Log (Mall), C(9) – Log(Town), C(10) - Year, C(11) - Road, C(14) – LogY.

The plots in Figure 6.1 also indicate the possibility of structural instability in the developed parcels when we order them according to the assessed value of their construction. Land parcels with expensive construction may follow a different stochastic

Table 6.2 OLS Estimates for the Fixed Effects Land Value Model

Variable	Coefficient	Std. Dev.	t-ratio
Log(Size)	-0.782231	0.015734	49.7
$[Log(Size)]^2$	-0.007844	0.015939	0.49
$[Log(Size)]^3$	0.034182	0.014357	2.38
Population	0.001238	0.002914	0.42
Log(Elevation)	0.019486	0.062747	0.31
Soil1	0.050458	0.030326	1.66
Soil2	-0.059731	0.044928	1.32
Log(Mall)	-0.052942	0.148750	0.35
Log(Town)	0.201436	0.232659	0.86
Year	0.050337	0.006757	7.44
Road	-0.036436	0.018225	2.00
Developed	0.172657	0.012719	13.6
LogX	0.264301	0.036362	7.26
LogY	0.016592	0.050194	0.33
LogX*LogY	0.947712	0.966546	0.98
R^2		0.7303	
Adjusted R^2		0.7282	

Table 6.3 Misspecification Tests for the Fixed Effects Land Value Model

Test	Null Hypothesis	Specification	P-Value
Jacque-Bera (Normality)	Residuals are normally distributed	$JB = (N-k)(4S^2 + (K-3)^2)/24$ S is the skewness, K is the Kurtosis, and N-k are the degrees of freedom	0.000000
Linearity (Squares)	Redundancy of non-linear (squared) variables	$u = c + ax + bx^2$ u is the vector of residuals, c is a constant, x is the vector of variables	0.831094
Linearity (Cross-Products)	Redundancy of non-linear (cross- product) variables	$u = c + ax + by$ u , c, x as described above and y is the vector of cross-product variables	0.261101
White Test (Heteroskedasticity)	Homoskedasticity	$u^2 = c + bx^2 + dy$ u, c, x, as described above and y is the vector of cross-product variables	0.179652
Auxiliary Regression (Spatial Autocorrelation)	No Spatial Autocorrelation (Ordering according to Neighborhoods)	$u = c + ax + bWu$ u , c, x as described above and W is the weight matrix	0.097708
Ramsey RESET (Incorrect Functional Form)	Correct Specified Functional Form of the Model	$u = c + ax + bm$ u, c, x as described above, and m is the vector of fitted values of x	0.865346
ARCH Test (Dependence in Variance)	No Dependence in Residual Variance	$u^2_z = c + au^2_{z-1} + bu^2_{z-2} + du^3_{z-3}$ u, c, x as described above and z is the ordering factor	0.029708
Chow Test (Structural Break)	Existence of Structural Change (Breakpoint n = 213, 750)	F statistic based on the comparison of restricted and unrestricted sum of square residuals	<0.001
Joint Mean Test	Linearity, no spatial autocorrelation and structural stability. (break point n = 213)	$u = c + ax + bx^2 + dWu + kT$ u, x, and W as described above and T is a binary variable with 0 before the break point and 1 after	0.000000
- No Spatial Autocorrelation	No spatial autocorrelation (in the joint mean test)	$u = c + ax + bx^2 + kT$ u, x, W, and T as described above	0.042568
- Structural Stability	Existence of structural change	$u = c + ax + bx^2 + dWu$ u, x, and W as described above	0.000000
- Linearity	Redundancy of non-linear variables (break point as above)	$u = c + ax + dWu + kT$ u, x, W, and T as described above	0.334561
Joint Variance Test	Homoskedasticity, no second order dependence and structural stability. (break point as above)	$u^2_z = c + ax + bx^2 + du^2_{z-1} + kT$ u, c, x, W, z, and T as described above	0.000000
- No Second Order Dependence	No dependence in residual variance (in the joint variance test)	$u^2_z = c + ax + bx^2 + kT$ u, c, x, W, z, and T as described above	0.006277
- Structural Stability	Existence of structural change (in the joint variance test)	$u^2_z = c + ax + bx^2 + du^2_{z-1}$ u, c, x, W, z, and T as described above	0.000000
- Homoskedasticity	Homoskedasticity	$u^2_z = c + ax + bx^2 + du^2_{z-1} + kT$	0.294615
Chow Forecast Test	Structural Stability (model estimated for observations 214 to 750 and then the estimated model predicts land values for 751 to 1804	F statistic based on the comparison of residual sums of squares when the equation is fitted to all sample observations with residual sum of squares for part of the sample	0.000000

*For analytical discussion of misspecification tests see Appendix 1

Figure 6.1 Recursive OLS Estimates for the Fixed Effects Model in Roanoke County
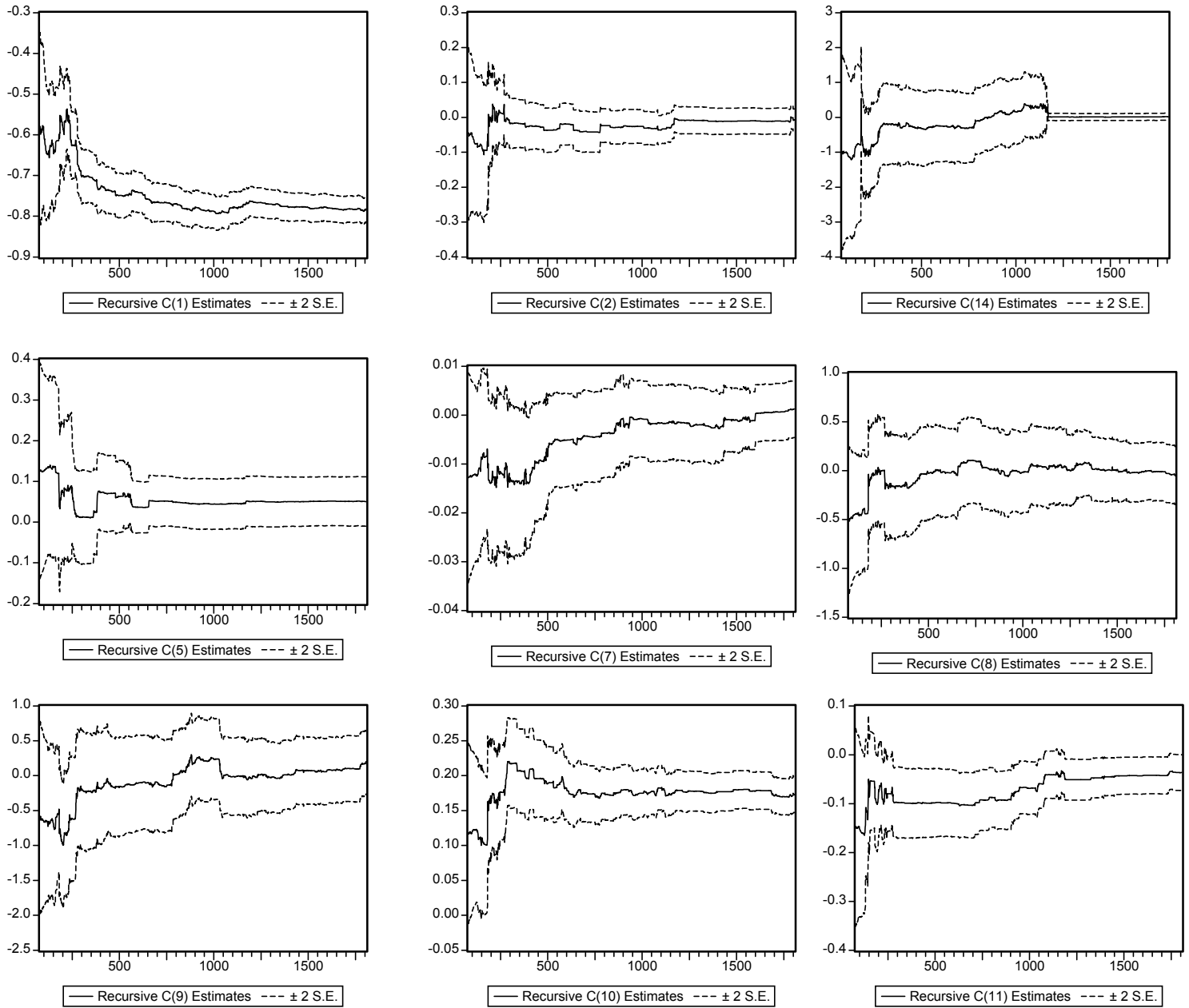
process than parcels with inexpensive constructions. Despite the fact that we can not clearly distinguish structural breaks at specific points of the plots, it seems that almost all of these plots have some type of "jump" around the 750th observation, when the assessed value of the construction is about $60 per square foot. In addition, window OLS (estimating the value of coefficients for observations 214 until 750 and comparing them with the values of the same coefficients for observations 751 until 1803) do not support the hypothesis that the parameter estimates for developed parcels are the same before and after the 750th observation. This can be demonstrated by the low P-value of the Chow forecast test. The Chow forecast test estimates the fixed effects model for the subsample of the observations 214 until 750, and then examines the difference between actual and predicted land values for the observations 751 to 1804. Thus, in addition to dividing our sample into vacant and developed parcels, we also use two subgroups of developed parcels to deal with the problem of structural instability. The first group contains parcels with inexpensive constructions (an assessed value below $60 per square foot), while the second group has parcels with expensive constructions (parcels with an assessed value of $60 per square foot or higher). If the leading cause of misspecification in the estimated land value model is structural instability, it seems possible that dividing the data into more spatially homogeneous groups will improve the statistical validity of the model.

6.3 Land Values of Developed Parcels

In this section, we estimate fixed effects models (used also in the previous section) for the group of developed parcels with expensive constructions and the group of developed parcels with non-expensive construction. Model (6.1) is estimated for the group of developed parcels with expensive constructions (greater than \$60 per square foot), and model (6.2) is estimated for the developed parcels with non-expensive constructions[9].

$Log(Price) = A_1[Log(Size)] + A_2[Log(Size)]^2 + A_3[Log(Size)]^3 + A_4[Log(Elevation)] + A_5(Population) + A_6(Soil1) + A_7(Soil2) + A_8[Log(Mall)] + A_9[Log(Town)] + A_{10}(Road) + A_{11}(Year) + A_{12}[Log(X)] + A_{13}[Log(Y)] + A_{14}[Log(X)][Log(Y)] + u$ (6.1)

$Log(Price) = A_1[Log(Size)] + A_2[Log(Elevation)] + A_3(Population) + A_4(Soil1) + A_5(Soil2) + A_6[Log(Mall)] + A_7[Log(Mall)] + A_8[Log(Town)] + A_9(Road) + A_{10}(Year) + A_{11}[Log(X)] + A_{12}[Log(Y)] + A_{13}[Log(X)][Log(Y)] + u$ (6.2)

Table 6.4 contains the results of misspecification tests for model (6.1), and Table 6.5 contains the results of misspecification tests for model (6.2). A careful look at these two tables indicates a clear improvement in the P-values of assumptions with limited support in the previous models. In fact, the P-values of individual and joint misspecification tests in both tables indicate that there is adequate support for all the underlying model assumptions. Note that the number of misspecification tests reported in Table 6.4 and Table 6.5 are fewer than the misspecification tests reported in the previous sections. More misspecification tests were necessary in the previous sections to investigate violations of individual and joint assumptions, while in Table 6.4 and Table 6.5 we see no evidence of any assumption violation.

---

[9] Note that although estimated as fixed effects models, for simplicity we do not explicitly report all the neighborhood effects.

Specifically, in Table 6.4 we see that the Jacque-Bera test provides adequate support for the assumption of normality. The assumption of linearity is also supported by the P-values in Table 6.4, while the Ramsey RESET test provides additional evidence that our data supports the choice of the functional form. Both individual and joint tests provide evidence for the acceptance of the homoskedasticity assumption. Relatively high P-values confirm that the problem of spatial autocorrelation does not exist in this model, while the results also indicate that there is no second order dependence. In addition, various break point tests (n = 400 and n = 800) and the Joint Mean Test provide support for the structural stability of the model. Similarly, in Table 6.5 we can see even higher P-values for the Jacque-Bera and the linearity tests in comparison to the results in Table 6.4. There is also support for the assumption of homoskedasticity, coming from both individual (White test) and the Joint Variance tests. In Table 6.5, we can also see that individual and joint misspecification tests provide evidence that there is no first or second order dependence in the land value model.

Table 6.4 Misspecification Tests for the Fixed Effects Land Value Model for Observations in the Expensive Constructions Group

Test	Null Hypothesis	Specification	P-Value
Jacque-Bera (Normality)	Residuals are normally distributed	$JB = (N-k)(4S^2 + (K-3)^2)/24$ S is the skewness, K is the Kurtosis, and N-k are the degrees of freedom	0.377409
Linearity (Squares)	Redundancy of non-linear (squared) variables	$u = c + ax + bx^2$ u is the vector of residuals, c is a constant, x is the vector of variables	0.548247
Linearity (Cross-Products)	Redundancy of non-linear (cross- product) variables	$u = c + ax + by$ u is the vector of residuals, c is a constant, y is the vector of cross-product variables	0.128269
White Test (Heteroskedasticity)	Homoskedasticity	$u^2 = c + ax + bx^2 + dy$ u is the vector of residuals, c is a constant, x is the vector of variables, y is the vector of cross-product variables	0.092861
Auxiliary Regression (Spatial Autocorrelation)	No spatial autocorrelation (ordering according to neighborhoods)	$u = c + ax + bWu$ u is the vector of residuals, c is a constant, x is the vector of variables, W is the weighting matrix	0.176971
Ramsey RESET (Incorrect Functional Form)	Correctly specified functional form of the model	$u = c + ax + bz$ u is the vector of residuals, c is a constant, x is the vector of variables, z is the vector of fitted values of x	0.483672
ARCH Test (Second Order Dependence)	No second order dependence	$u^2_z = c + au^2_{z-1} + bu^2_{z-2} + du^2_{z-3}$ u is the vector of residuals, c is a constant, x is the vector of variables, z is the ordering factor	0.084552
Chow Test (Structural Break)	Existence of structural change (Breakpoint n = 400, 800)	F statistic based on the comparison of restricted and unrestricted sum of square residuals	>0.1
Joint Mean Test	Linearity, no spatial autocorrelation and structural stability. (Break point n = 400, 800)	$u = c + ax^2 + bWu + dT$ u, x, and W as described above and T is a binary variable with 0 before the break point and 1 after	> 0.1
Joint Variance Test	Homoskedasticity, no second order dependence and structural stability. (Break Points as above)	$u^2_z = c + ax^2 + bu^2_{z-1} + dT$ u, c, x, W, z, and T as described above	>0.1
Redundancy Test	Variables "Road" and "LogX*LogY" are essential for the land value model	F-test comparing residual sums of squares for the land value model with and without these variables	0.000000

*For analytical discussion of misspecification tests see Appendix 1.

Table 6.5 Misspecification Tests for the Fixed Effects Land Value Model for Observations in the Non-Expensive Constructions Group

Test	Null Hypothesis	Specification	P-Value
Jacque-Bera (Normality)	Residuals are normally distributed	$JB = (N-k)(4S^2 + (K-3)^2)/24$ S is the skewness, K is the Kurtosis, and N-k are the degrees of freedom	0.510699
Linearity (Squares)	Redundancy of non-linear (squared) variables	$u = c + ax + bx^2$ u is the vector of residuals, c is a constant, x is the vector of variables	0.863646
Linearity (Cross-Products)	Redundancy of non-linear (cross- product) variables	$u = c + ax + by$ u is the vector of residuals, c is a constant, y is the vector of cross-product variables	0.154613
White Test (Heteroskedasticity)	Homoskedasticity	$u^2 = c + ax + bx^2 + dy$ u is the vector of residuals, c is a constant, x is the vector of variables, y is the vector of cross-product variables	0.112854
Auxiliary Regression (Spatial Autocorrelation)	No spatial autocorrelation (ordering according to neighborhoods)	$u = c + ax + bWu$ u is the vector of residuals, c is a constant, x is the vector of variables, W is the weighting matrix	0.425345
Ramsey RESET (Incorrect Functional Form)	Correctly specified functional form of the model	$u = c + ax + bz$ u is the vector of residuals, c is a constant, x is the vector of variables, z is the vector of fitted values of x	0.932831
ARCH Test (Second Order Dependence)	No second order dependence	$u^2_z = c + au^2_{z-1} + bu^2_{z-2} + du^2_{z-3}$ u is the vector of residuals, c is a constant, x is the vector of variables, z is the ordering factor	0.098172
Chow Test (Structural Break)	Existence of structural change (break point n = 200, 400)	F statistic based on the comparison of restricted and unrestricted sum of square residuals	>0.1
Joint Mean Test	Linearity, no spatial autocorrelation and structural stability. (break point n = 200, 400)	$u = c + ax^2 + bWu + dT$ u, x, and W as described above and T is a binary variable with 0 before the break point and 1 after	> 0.1
Joint Variance Test	Homoskedasticity, no dependence in residual variance and structural stability. (break points as above)	$u^2_z = c + ax^2 + bu^2_{z-1} + dT$ u, c, x, W, z, and T as described above	>0.05
Redundancy Test	Variables "Road" and "LogX*LogY" are essential for the land value model	F-test comparing residual sums of squares for the land value model with and without these variables	0.000001

*For analytical discussion of misspecification tests see Appendix 1.

The OLS estimates for models (6.1) and (6.2) are presented in Table 6.6 and Table 6.7, respectively. These results contain the coefficient estimates that are significantly different from zero for at least one of the two models. In other words, the estimated models do not contain the variables "Road" and "LogX*LogY", because their coefficients are statistically equal to zero in both models. The F-tests reported in Table 6.4 and Table 6.5 indicate the low contribution of these two variables to the land value models. The omission of these variables does not alter the conclusions of the misspecification tests.

In Table 6.6, we can see that the fixed effects land value model for the group of parcels with expensive constructions explains about 73% of the variation in land transaction prices in this category. The results indicate that the size of the parcel is an important determinant of the land value in this group. Larger land parcels are associated with higher land values. There is strong evidence higher elevation is associated with higher land values, while weaker evidence indicates that impermeable soils are associated negatively with land values. Higher elevation and soil permeability are usually two proxies, which indicate lower flood risk. Roanoke County has experienced several floods in the last fifty years (Planning Report of Roanoke County, 1994). The model results indicate that lower flood risk areas have higher land values. Land parcels far from the two major malls are less expensive in this group of observations. Given that these large malls accommodate shopping facilities and entertainment amenities (theaters, restaurants, etc) the negative sign should be anticipated. The service industry of Roanoke County is also located close to these malls and not in the town center, and housing locations close to these malls is attractive to upper-middle class residents of Roanoke County. Finally, the average price of land parcels sold in 1997 was higher than those sold in 1996.

Table 6.6 OLS Estimates for the Fixed Effects Land Value Model for Observations in the Expensive Constructions Group

Variable	Coefficient	Std. Dev.	t-ratio
Log(Size)	-0.829923	0.021288	39.3
$[Log(Size)]^2$	0.056520	0.030224	1.37
$[Log(Size)]^3$	0.073669	0.040547	2.59
Population	-0.002805	0.003200	0.87
Log(Elevation)	0.288472	0.167098	1.82
Soil1	-0.020531	0.034965	0.58
Soil2	-0.086192	0.049628	1.74
LogX	-0.109929	0.177109	0.62
LogY	0.155010	0.128656	1.20
Log(Mall)	-0.192311	0.010076	1.97
Log(Town)	0.024088	0.251096	0.09
Year	0.044958	0.007231	6.21
R^2		0.7316	
Adjusted R^2		0.7286	

Table 6.7 summarizes the OLS estimates for the fixed effects land value model for observations in the non-expensive constructions group. This model explains about 65% of the variance in land transaction prices in this group. It is important to note that the significance of the variables differs between the two models for the developed land parcels. In the non-expensive construction group, larger parcels have lower land value per square meter. Lack of water permeability to soil (and consequently higher flood risk as indicated by the Soil1 and Soil2 dummies) is expected to affect negatively land prices. The sign of the elevation parameter is again positive but not statistically important. There

is weak evidence that population density may affect negatively the price of the lot in relatively inexpensive areas. The negative sign of population density may reflect the willingness of the residents in Roanoke County to live in less populated areas and enjoy open space amenities. The negative relationship of land values with distance from the nearest town reflects both distance to amenities and the potential of the neighborhood for faster residential and commercial development. The quadratic term of the distance to the nearest town indicates that the parcel value increases at a decreasing rate when a parcel is closer to the town center. The importance of location is also reflected by the statistical significance of the coordinate X. As already mentioned in Chapter 2, the direction of the X coordinate is from south-east to north-west of Roanoke County. The price of the lots sold in 1997 was higher than those sold during the previous year.

Table 6.7 OLS Estimates for the Fixed Effects Land Value Model for Observations in the Non-Expensive Constructions Group

Variable	Coefficient	Std. Dev.	t-ratio
Log(Size)	-0.747182	0.027792	26.9
Population	-0.004161	0.002680	1.56
Log(Elevation)	0.054530	0.070985	0.77
Soil1	-0.102809	0.045168	2.27
Soil2	-0.153847	0.078481	1.97
Log(Town)	-0.369564	0.183270	2.06
$[Log(Town)]^2$	-2.118983	0.855956	2.25
Log(Mall)	0.019926	0.137524	0.14
LogX	0.230214	0.035913	6.27
LogY	-0.097929	0.113661	0.85
Year	0.061557	0.012926	4.76
R^2		0.6556	
Adjusted R^2		0.6497	

6.4 Land Values of Undeveloped Parcels

In this section, we apply the fixed effects model to the group of undeveloped parcels. As we saw in the previous section, the misspecification test results justified the application of the fixed effects model in the two groups of developed parcels. Model (6.3) was estimated for the undeveloped parcels and Table 6.8 contains the results of misspecification tests for this model.

Log(Price) = A_1[Log(Size)] + A_2[Log(Size)]^2 + A_3[Log(Elevation)] + A_4(Population) + A_5(Soil1) + A_6(Soil2) + A_7[Log(Mall)] + A_8[Log(Town)] + A_9(Road) + A_{10}(Year) + A_{11}[Log(X)] + A_{12}[Log(Y)] + A_{13}[Log(X)][Log(Y)] + u   (6.3)

In Table 6.8, we notice that individual and joint misspecification tests provide support for the assumptions of linearity, homoskedasticity and structural stability. The low P-value in the Jacque-Bera test suggests possible violation of the normality assumptions. However, the Jacque-Bera tests is sensitive to extreme observations and when we exclude some observations (less than 1%) from our sample the P-value of the Jacque-Bera tests exceeds 0.1, and provides support for the assumption of normality. However, there is serious evidence (coming from the Auxiliary Regression test, the ARCH test and the Joint Mean and Variance tests) that the assumptions of no first and second order spatial dependence are violated. Both individual and joint misspecification tests for these two assumptions provide low P-values. Despite the fact that misspecification tests indicate that second order spatial dependence may be the main source of model misspecification (lower P-values), it may well be the case that violations of both assumptions have the same source. This group of observations contains no information regarding whether a parcel is located in an area of expensive or non-expensive constructions, and this subgroup of parcels is probably less homogeneous than the two subgroups of developed parcels.

Table 6.8 Misspecification Tests for the Fixed Effects Land Value Model for the Group of Undeveloped Parcels

Test	Null Hypothesis	Specification	P-Value
Jacque-Bera (Normality)	Residuals are normally distributed	$JB = (N-k)(4S^2 + (K-3)^2)/24$ S is the skewness, K is the Kurtosis, and N-k are the degrees of freedom	0.000000
Linearity (Squares)	Redundancy of non-linear (squared) variables	$u = c + ax + bx^2$ u is the vector of residuals, c is a constant, x is the vector of variables	0.495375
Linearity (Cross-Products)	Redundancy of non-linear (cross- product) variables	$u = c + ax + by$ u , c, x as described above and y is the vector of cross-product variables	0.176429
White Test (Heteroskedasticity)	Homoskedasticity	$u^2 = c + bx^2 + dy$ u, c, x, as described above and y is the vector of cross-product variables	0.138545
Auxiliary Regression (Spatial Autocorrelation)	No Spatial Autocorrelation (Ordering according to Neighborhoods)	$u = c + ax + bWu$ u , c, x as described above and W is the weight matrix	0.000964
Ramsey RESET (Incorrect Functional Form)	Correct Specified Functional Form of the Model	$u = c + ax + bm$ u, c, x as described above, and m is the vector of fitted values of x	0.766772
ARCH Test (Dependence in Variance)	No Dependence in Residual Variance	$u^2_z = c + au^2_{z-1} + bu^2_{z-2} + du^3_{z-3}$ u, c, x as described above and z is the ordering factor	0.000000
Chow Test (Structural Break)	Existence of Structural Change (Breakpoint n = 100)	F statistic based on the comparison of restricted and unrestricted sum of square residuals	0.281194
Joint Mean Test	Linearity, no spatial autocorrelation and structural stability. (break point n = 213)	$u = c + ax + bx^2 + dWu + kT$ u, x, and W as described above and T is a binary variable with 0 before the break point and 1 after	0.035488
- No Spatial Autocorrelation	No spatial autocorrelation (in the joint mean test)	$u = c + ax + bx^2 + kT$ u, x, W, and T as described above	0.000000
- Structural Stability	Existence of structural change	$u = c + ax + bx^2 + dWu$ u, x, and W as described above	0.686541
- Linearity	Redundancy of non-linear variables (break point as above)	$u = c + ax + dWu + kT$ u, x, W, and T as described above	0.561565
Joint Variance Test	Homoskedasticity, no second order dependence and structural stability. (break point as above)	$u^2_z = c + ax + bx^2 + du^2_{z-1} + kT$ u, c, x, W, z, and T as described above	0.000000
- No Second Order Dependence	No dependence in residual variance (in the joint variance test)	$u^2_z = c + ax + bx^2 + kT$ u, c, x, W, z, and T as described above	0.000000
- Structural Stability	Existence of structural change (in the joint variance test)	$u^2_z = c + ax + bx^2 + du^2_{z-1}$ u, c, x, W, z, and T as described above	0.264588
- Homoskedasticity	Homoskedasticity	$u^2_z = c + ax + bx^2 + du^2_{z-1} + kT$	0.110168

*For analytical discussion of misspecification tests see Appendix 1

In the previous chapters, we saw that the traditional "remedy" for the problem of autocorrelation in spatial econometric studies is the specification of an autoregressive error term model. However, the success of the error autoregressive model in Chapter 5 was limited. Spanos (1986) also suggests that instead of modeling "invisible" variables by using lags of the error term, it is better to use spatial lags of the dependent and independent variables. Following the alternative approach, proposed by Spanos (1986), we estimate a fixed effects model for the vacant parcels, which also allows spatial lags of the dependent and independent variables, and we retest the validity of our results.

Table 6.9 summarizes the results of these misspecification tests. Table 6.9 indicates that by adding spatial lags in model (6.3) there is an obvious improvement in the statistical validity of the model. There is still strong support for the assumptions of linearity, homoskedasticity and structural stability, while we also see higher P-values (Auxiliary Regression test and ARCH test) for the no spatial autocorrelation assumptions (both first and second order, respectively). However, there is still limited support for the hypothesis of no second order spatial dependence (ARCH test). The coefficients of some variables and their respective spatial lags are not statistically different from zero, and the joint F-test recommends dropping these variables from our model (Year and LogX*LogY). The final model estimated is (6.4). Table 6.10 contains the OLS estimates of model (6.4), while there is no change in the conclusions of the misspecification tests.

$$\text{Log(Price)} = A_1[\text{Log(Size)}] + A_2[\text{Log(Size)}]^2 + A_3(\text{Soil1}) + A_4(\text{Soil2}) + A_5[\text{Log(Mall)}] + A_6[\text{Log(Town)}] + A_7(\text{Year}) + A_8[\text{Log(X)}] + A_9[\text{Log(Y)}] + A_{10}(\text{Road}) + A_{11}[\text{WLog(Price)}] + A_{12}[\text{WLog(Size)}] + A_{13}(\text{WSoil1}) + A_{14}(\text{WSoil2}) + A_{15}[\text{WLog(Town)}] + A_{16}(\text{WYear}) + A_{17}[\text{WLog(X)}] + A_{18}[\text{WLog(Y)}] + A_{19}(\text{WRoad}) + u$$
$$(6.4)$$

In Table 6.10, we can see that the size of the parcel is again a significant determinant of land prices, while there is some weak support for a quadratic relation between the size of the parcel and the land transaction price. The quadratic form indicates that the value of

the parcel per square meter decreases at a declining rate with increases in parcel size. The results also indicate that higher land values should be expected for land parcels which are closer to the shopping malls, but far from the town centers. The value of the land is also lower when the parcel is next to an interstate highway. The importance of the parcel location is also underlined by the statistical significance of the determinants X and Y. The "Year" variable again has a positive sign but its value is not statistically important. Finally, the autoregressive variables account for spatial effects in the model. It is important to note that this model explains about 95% of the variance in land transaction prices. The high R^2 suggests that spatial lags are capable of capturing additional variation of the dependent variable in our case study. However, the high R^2 value would have no meaning if the model were not well specified.

Table 6.9 Misspecification Tests for the Fixed Effects Land Value Model with Spatial Lags for the Group of Undeveloped Parcels.

Test	Null Hypothesis	Specification	P-Value
Jacque-Bera (Normality)	Residuals are normally distributed	$JB = (N-k)(4S^2 + (K-3)^2)/24$ S is the skewness, K is the Kurtosis, and N-k are the degrees of freedom	0.000000
Linearity (Squares)	Redundancy of non-linear (squared) variables	$u = c + ax + bx^2$ u is the vector of residuals, c is a constant, x is the vector of variables	0.999761
Linearity (Cross-Products)	Redundancy of non-linear (cross- product) variables	$u = c + ax + by$ u is the vector of residuals, c is a constant, y is the vector of cross-product variables	0.646551
White Test (Heteroskedasticity-Squares)	Homoskedasticity	$u^2 = c + ax + bx^2 + dy$ u is the vector of residuals, c is a constant, x is the vector of variables, y is the vector of cross-product variables	0.863646
White Test (Heteroskedasticity-Cross-Products)	Homoskedasticity	$u^2 = c + ax + bx^2 + dy$ u is the vector of residuals, c is a constant, x is the vector of variables, y is the vector of cross-product variables	0.169623
Auxiliary Regression (Spatial Autocorrelation)	No spatial autocorrelation (ordering according to neighborhoods)	$u = c + ax + bWu$ u is the vector of residuals, c is a constant, x is the vector of variables, W is the weighting matrix	0.425345
Ramsey RESET (Incorrect Functional Form)	Correct specified functional form of the model	$u = c + ax + bz$ u is the vector of residuals, c is a constant, x is the vector of variables, z is the vector of fitted values of x	0.487500
ARCH Test (Second Order Dependence)	No second order dependence	$u^2_z = c + au^2_{z-1} + bu^2_{z-2} + du^3_{z-3}$ u is the vector of residuals, c is a constant, x is the vector of variables, z is the ordering factor	0.001703
Chow Test (Structural Break)	Existence of structural change (break point n = 100)	F statistic based on the comparison of restricted and unrestricted sum of square residuals	0.394885
Joint Mean Test	Linearity, no spatial autocorrelation and structural stability. (break point n = 100)	$u = c + ax^2 + bWu + dT$ u, x, and W as described above and T is a binary variable with 0 before the break point and 1 after	0.163581
Joint Variance Test	Homoskedasticity, no second order dependence and structural stability. (break point as above)	$u^2_z = c + ax^2 + bu^2_{z-1} + dT$ u, c, x, W, z, and T as described above	0.076762
Redundancy Test	Variables which enter model 6.3 and not enter 6.4 are essential for the model	F-test comparing residual sums of squares for the land value model with and without these variables	0.000000

*For analytical discussion of misspecification tests see Appendix 1.

Table 6.10 OLS Estimates for the Fixed Effects Land Value Model for the Group of Undeveloped Parcels

Variable	Coefficient	Std. Dev.	t-ratio
Log(Size)	-0.695926	0.028389	24.5
$[Log(Size)]^2$	0.019661	0.012441	1.58
Log(Mall)	-0.313128	0.106184	2.95
Log(Town)	1.722006	0.376349	4.57
Road	-0.210169	0.057640	3.64
LogX	0.437769	0.139052	3.15
LogY	-0.195595	0.099372	1.97
Year	0.023158	0.020734	1.12
WLog(Price)	-1.725382	0.075915	22.7
WLog(Size)	-1.197150	0.069400	17.2
WLog(Town)	3.371006	0.844649	3.99
WRoad	-0.343883	0.085956	4.00
WLogX	0.959943	0.423241	2.27
WLogY	-0.555855	0.272661	2.04
WYear	0.084640	0.047181	1.79
R^2		0.9516	
Adjusted R^2		0.9482	

6.6 Conclusions

In this chapter, three models were created to explain the variation in land transaction prices in Roanoke County. Our initial step was to revisit the land value model estimated in the fourth chapter of the thesis and conduct a more comprehensive set of individual and joint misspecification tests. The misspecification test results indicate the possibility of spatial autocorrelation and that neighborhood dummies were needed in the land value model. After deriving the OLS estimates for the fixed effects model (which accounts for neighborhoods) for Roanoke County, we retested the statistical validity of our model. In spite of more support for the hypothesis of no spatial autocorrelation, tests for structural instability as well as plots of recursive OLS estimates indicated the need for more homogeneous groups of observations. There was strong evidence that land values of vacant parcels might follow a different stochastic process than the land values of developed parcels. Some additional evidence indicated that separating parcels with expensive constructions from parcels with non-expensive constructions might also help deal with the problem of structural instability.

Misspecification tests of the fixed effects land value models for the expensive and non-expensive constructions, indicate that these models satisfy the underlying statistical assumptions. In the group of undeveloped land parcels, there is evidence that spatial autocorrelation is the most important source of model misspecification. By using spatial lags of dependent and independent variables instead of spatial lags of error terms proposed by the traditional spatial econometric approach, we improve the statistical validity of the model and explain approximately 95% of the land transaction price variation. Misspecification tests also provide support for the validity of the model.

The major finding of this chapter is that based on their development status prices of land parcels in Roanoke County may follow different stochastic processes. Changes in the values of hedonic variables have different implications for different groups of land parcels. For example, larger parcels are expected to have higher value per square meter when they accommodate non-expensive constructions, and lower value per square meter when they accommodate expensive constructions. The signs of most other hedonic characteristics are consistent in the different land value models, however we can notice that some variables seem to be more significant than others. For example, the elevation of the land parcel is significant for the group of parcels with expensive constructions, while it is not statistically different from zero in the group of parcels with non-expensive constructions.

CHAPTER 7

Summary & Conclusions

After more than two decades of applied spatial econometric studies, the possibility of spatial autocorrelation in a cross-sectional model is widely accepted. However, it is unclear whether current spatial econometric studies have really enhanced our ability to take into consideration the spatial structure in cross-sectional modeling and improve the validity of our estimates. Typically, current spatial econometric approaches examine the assumptions of normality, heteroskedasticity, and structural stability to make sure that spatial error autocorrelation is not observed due to misspecification problems other than omission of relevant spatial variables. According to this "traditional" approach the problem of true spatial autocorrelation can be corrected successfully by using a spatial error autoregressive model. However, a number of objections regarding the methods followed in those empirical studies have been raised. Current spatial econometric studies examine only a subset of the underlying model assumptions to determine the source of misspecification problems. Even if current spatial econometric studies assessed the

validity of all the assumptions underlying the statistical model, the approach of testing individual assumptions one-by-one and fixing problems when necessary is very unlikely to lead to a well-specified model. A battery of individual and joint tests is necessary to identify misspecifications sources and lead respecification efforts. Another objection to the methodology of current spatial econometric studies is that the success of their results is judged by the changes in the fitting power of the model. Finally, in the "traditional" approach to dealing with spatial autocorrelation, the specification of spatial structure in an econometric model is completely arbitrary.

In the second chapter, we briefly review basic elements of existing empirical studies of land values. Location is probably the most important determinant of land values, while a number of empirical studies also suggest that hedonic attributes can explain variation in land prices. Chapter 2 also describes our empirical information, which is based on a number of geographical, morphological and socioeconomic characteristics of land parcels in Roanoke County.

In the third chapter we introduce some basic tools that may be used in analyzing and testing patterns of spatial dependence. Fundamental to much of this material is the notion of a weight matrix, W, which captures the spatial relationship between a set of spatial units. A weight matrix has the role of averaging values at neighboring sampled data points. Several alternative definitions of a spatial weight matrix are also discussed. The researcher should use the type of weight matrix that is consistent with the available empirical information, but be aware of its *ad hoc* nature when interpreting her results.

The fourth chapter investigates whether land values in Roanoke County are spatially correlated, demonstrating the possibility that an econometric model, which does not consider the spatial configuration of the data may lead to biased and inconsistent estimators. Following current spatial econometric techniques, we initially estimate a land value model for Roanoke County, and then we examine whether spatial autocorrelation is present. Prior to testing for spatial autocorrelation, the land value model for Roanoke

County is tested for the assumptions of normality, heteroskedasticity and structural stability. In this way, we attempt to verify that the observed autocorrelation is "true", meaning due to omitted spatial variables and not due to violations of other assumptions.

The fifth chapter applies both parametric and non-parametric techniques to the land value model of Roanoke County to deal with the problem of spatial autocorrelation. The results indicate that both techniques achieve higher fitting power (based on R^2) than the initial model estimated in the fourth chapter. Higher fitting power implies that by using spatial lags, the model explains a larger portion of the variance of the dependent variable. However, results of misspecification tests indicate that both the parametric and non-parametric models violate essential underlying statistical assumptions. Given these assumption violations, we conclude that by applying the tools of current spatial econometric methodology we have not adequately modeled land values. Anselin (1988) suggests that the researcher should continue trying alternative weight matrices until the misspecification problem is solved. However, we saw in Chapter 4 that the weight matrix used in this chapter relies on detailed information provided by the Planning Department of Roanoke County, and matrices based on distances among the parcel centers make no sense in our case study. Thus, instead of trying various *ad hoc* redefinitions of weight matrices, we attempt modeling land values using an alternative approach.

In the sixth chapter, ultimately three models are estimated to explain the variation in land transaction prices in Roanoke County. We revisit the land value model estimated in the fourth chapter of the thesis and conduct a more comprehensive set of individual and joint misspecification tests. Following an iterative procedure of respecification and testing, we concluded that vacant parcels follow a different stochastic process than developed parcels. In addition, two subgroups of developed parcels were created to account for a lack of homogeneity across expensive and non-expensive construction land parcels. Misspecification tests provide evidence that the final models satisfy the underlying statistical assumptions, and thus, provide more statistically reliable estimates than the models derived earlier using current spatial econometric techniques.

The findings of this dissertation basically confirm the validity of the three propositions stated in the introductory chapter. The first proposition of the thesis is that there is spatial dependence in the values of land parcels in Roanoke County. Using current spatial econometric techniques, we initially evaluated this proposition by estimating a land value model, which did not violate normality, heteroskedasticity and structural stability, and then tested for the assumption of spatial autocorrelation. This procedure led us to the conclusion that there is indeed spatial dependence in the values of land parcels in Roanoke County. However, in the sixth chapter of the thesis, we applied a more comprehensive set of misspecification tests, which indicated that lack of structural stability among land parcels may be one of the sources of spatial autocorrelation. Using the assessed value of land parcel constructions, we classified developed land parcels in two more homogeneous groups. The estimated land value models presented no evidence of spatial autocorrelation. However, in the land value model of vacant parcels, there was persistent evidence of spatial autocorrelation and spatial lags of dependent and independent variables were used to solve this problem. In conclusion, the land value models of more homogeneous groups of observations (developed parcels classified based on the value of their construction) do not exhibit evidence of spatial autocorrelation, while spatial dependence does exist in the group of vacant parcels in Roanoke County.

The second proposition of the thesis is that mechanical construction of neighborhood effects, or grouping of geographically nearby land parcels into neighborhoods, is not always the best way to capture spatial effects. Using the statistical package Spacestat, we calculated that the neighborhood diameter in Roanoke County should be at least 2.8 km. However, such a large neighborhood would be clearly inappropriate (much too large) to capture spatial effects in Roanoke County. In addition this neighborhood specification would not consider natural frontiers and other socioeconomic parameters essential to determine spatial relationships. Neighborhoods defined by the Planning Department of Roanoke County are built based on the geographical particularities of Roanoke County. Some of these neighborhoods have a diameter smaller than 0.3 km, while others are big

enough to include common types of residential development at the county borders. Thus, we reject the use of mechanically constructed neighborhood effects for our case study and instead we use neighborhoods based on the empirical information provided by local government agencies.

The third and most important proposition of the thesis is related to the success of current spatial econometric practices in capturing the spatial aspects of land values. It states that by implementing a comprehensive set of individual and joint misspecification tests proposed by Spanos (1986) one can better identify misspecification error sources and thus establish a statistically adequate model, which is more reliable than models based on the existing spatial econometric practices. In the fifth chapter of the thesis we saw violation of fundamental assumptions in the empirical results generated with current spatial econometric techniques. Parametric and non-parametric techniques did not solve the problem of spatial autocorrelation. In Chapter 6, we used a battery of individual and joint misspecification tests to derive statistically adequate models of land values. The results of this case study suggest that applied econometricians could be more assured of the validity of their results by following the alternative approach described in Chapter 6 rather than using current methods to deal with the problem of spatial autocorrelation. However, further studies are needed in order to generalize these findings.

Finally, Table 7.1 summarizes the statistical estimates derived from the different models for land values in Roanoke County. A first conclusion from this table is that different hedonic attributes have different effects on land values depending on their development status. For example, while the first three models indicate that there is a quadratic relationship between parcel size and land value, the models derived in the sixth chapter indicate a linear relationship for the developed parcels with non-expensive constructions. Figure 7.1 presents the effects of parcel size on the land value per square meter for the different models. The vertical axis indicates the value of the parcel in US dollars per square meter, while the horizontal axis presents the size of the parcel in hectares. The graphs in Figure 7.1 assume mean values for all other variables except for parcel size.

Table 7.1 Model Estimates for Land Values in Roanoke County

Variable	OLS[a]	2SLS[b]	GME[b]	OLS[c]-Expensive Construction	OLS[c]-Non Expensive Construction	OLS[c]-Vacant Parcels
Constant	-17.46850	-14.02579	-8.58745			
Log(Size)	-0.48395	-0.35895	-0.31823	-0.82992	-0.74718	-0.69593
$[\text{Log(Size)}]^2$	-0.03062	-0.04058	-0.04704	0.05652		0.01966
$[\text{Log(Size)}]^3$				0.07367		
Log(Elevation)	0.33793	0.18655	0.41922	0.28847	-0.05453	
$[\text{Log(Elevation)}]^2$	-0.10623	-0.04876	-0.14528			
Soil1	-0.05668	-0.03001	-0.05906	-0.02053	-0.10281	
Soil2	-0.09161	-0.06669	-0.06158	-0.08619	-0.15385	
Population	0.00485	-0.00235	0.00998	-0.00281	-0.00416	
$(\text{Population})^2$	-0.00006	-0.00019	-0.00090			
Log(Mall)	1.40294	0.83548	1.77236	-0.19231	0.01993	-0.31313
$[\text{Log(Mall)}]^2$	-0.22056	-0.11256	-0.27594			
Log(Town)	0.25035	2.84582	2.13548	0.02409	-0.36956	1.72201
Log(Town)^2					-2.11898	
Developed	0.09403	0.13548	0.04761			
Road	-0.07093	-0.05867	-0.14257			-0.21017
Year	0.05639	0.05349	0.01458	0.04496	0.06156	0.02316
LogX	4.19009	3.31659	4.12933	-0.10993	0.23021	0.43777
LogY	3.81113	3.02646	3.75875	0.15501	-0.09793	-0.19560
(LogX)*(LogY)	-0.93027	0.17222	0.11878			
WLog(Price)		0.22486				-1.72538
WLog(Size)						-1.19715
WLog(Town)						3.37101
WRoad						-0.34388
WLogX						0.95994
WLogY						-0.55586
WYear						0.08464
R^2	0.8090	0.8154	0.8821	0.7316	0.6556	0.9516

[a] Estimated in Chapter 4

[b] Estimated in Chapter 5

[c] Estimated in Chapter 6

The results present clear evidence that the effect of parcel size differs depending on development status of the parcel; this finding highlights the importance of modeling these properties separately from a policy point of view. The need to separate these properties was only discovered by carefully examining the model assumptions using the approach proposed by Spanos. In the group of vacant parcels we can see that land value increases with the parcel size at a declining rate. Assuming average values for the other characteristics, irrespective of the parcel size the value of vacant parcels will be lower than parcels with some type of residential or commercial development. Smaller parcels have higher values in the group of developed parcels with non-expensive constructions, while the opposite relationship is indicated for the group of parcels with expensive constructions. Parcels with expensive constructions are relatively more expensive than those with non-expensive constructions. However, as shown by Figure 7.1 expensive construction values decrease with size while non-expensive values increase with size. The value differential between parcels with expensive and non-expensive constructions decreases with the size of the parcel.

The first three models have relatively small differences in their estimated coefficients. However, we saw in the previous chapters that these models violated the underlying statistical assumptions and their implications should not be trusted. This means that the policymaker should use the correctly specified models estimated in the sixth chapter to derive land value indices for Roanoke County and study the effects of different hedonic variables on land values. Comparing the results of the first three models to those derived in the sixth chapter we can see that all models agree, however, that land value per square meter increases with parcel size but at a decreasing rate, except for the model estimated in Chapter 6 for the expensive constructions. Higher elevation and permeable soil are two proxies indicating lower flood risk in Roanoke County. Most of the models indicate that higher elevation increases the value of the parcel at a decreasing rate while impermeable soil qualities (Soil1 and Soil2) are related to lower land values. The results are mixed in terms of the effects of population density on land values. The correctly specified models for developed parcels indicate that population density is related to lower land values. In

other words, parcels which accommodate some type of construction are relatively more expensive in less populated areas. The 2SLS model agrees with the correctly specified models about the negative effect of higher population density on land values, while the initial (misspecified) OLS and GME models suggest that land values increase with higher population density but at a decreasing rate. The first three models in Table 7.1 suggested that land values increase with distance from town and mall, however the models estimated in the sixth chapter suggest this conclusion is not accurate. Longer distance from a mall is related to lower land values for parcels with expensive constructions and vacant parcels, while longer distance from the nearest town is related to lower land value in the group of non-expensive parcels.

The first three models indicate that the presence of a highway next to a land parcel affects negatively its value. This is also the case for vacant parcels, according to the models estimated in the sixth chapter, while the presence of a road does not affect parcel values in the groups of developed parcels. However, all models agree that land transaction prices were higher in 1997 relative to the previous year. The location determinants of the parcels (X and Y) are significant in almost all models, indicating that location is a very important attribute of the parcel value even after accounting for neighborhood effects. Finally, spatial lags are used even after accounting for neighborhood effects to deal with the problem of spatial autocorrelation for vacant parcels. The size of the coefficients of spatial lags is higher than the coefficients of the respective explanatory variables. This implies that neighborhood hedonic characteristics may have stronger effects than individual parcel characteristics on its value. The signs of spatial lags are consistent with the signs of their respective explanatory values. For example, an increase in the size of a parcel and increases in the sizes of the parcels in a neighborhood affect in the same direction the price of the land parcel. The first three models yield relatively higher R^2 values than the estimated models for developed parcels estimated in Chapter 6. However, higher fitting power can be misleading if the model is not well-specified. In addition, the models for developed parcels were estimated using smaller more homogeneous samples

and thus, lower variability in the dependent variable is likely to cause a decline in the fitting power of the models.

Davidson and Mackinnon (1993) conclude that the situation where a model is correctly specified except for a failure to account for error autocorrelation does not account for a very high proportion of the cases in which residuals from a regression model appear to be correlated. Future research may indicate that this is also the case in cross-sectional studies that present spatial autocorrelation. More work is also needed to corroborate whether following the approach to develop a statistically adequate model (Spanos 1986) will make unnecessary inclusion of arbitrary specified weight matrices to account for influence of surrounding parcels. In case that spatial lags are needed, more research would be also useful to examine how a simple linear distance performs as a spatial weight matrix relative to other neighborhood boundaries based on socioeconomic and morphological characteristics. Finally, it may be a good idea to re-examine existing spatial econometric studies, under the light of a more complete set of misspecification tests, to validate the choice of the weight matrix in those studies.

From a policy point of view, our results have implications for urban expansion to rural areas in Roanoke County as well as the existing zoning policy of the local government. Specifically, we saw that the type of residential development (expensive versus non-expensive constructions) affects the stochastic process of land values in Roanoke County. The recent planning policy of Roanoke County (Roanoke County Planning Department, 1994) is restricted to the choice of area for development, and does not schedule the type of development in different areas. For example, our results indicate that the value of the land and consequently the revenue of the local government will increase not only by further construction in vacant parcels, but also with redevelopment with more expensive construction in already developed areas with non-expensive constructions. In addition, more research is necessary to examine how parcel size affects land value, an issue quite important to local governments who contemplate changes in their zoning policies. Our results indicate that changes in parcel size have different implications for land values
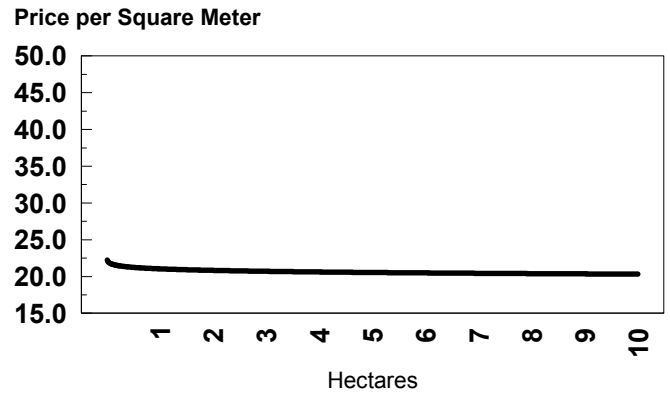
according to their development status. Smaller parcels may result in higher values (and tax income for the local government) in areas with expensive construction while larger parcels are slightly more expensive in areas with vacant parcels or non-expensive constructions.

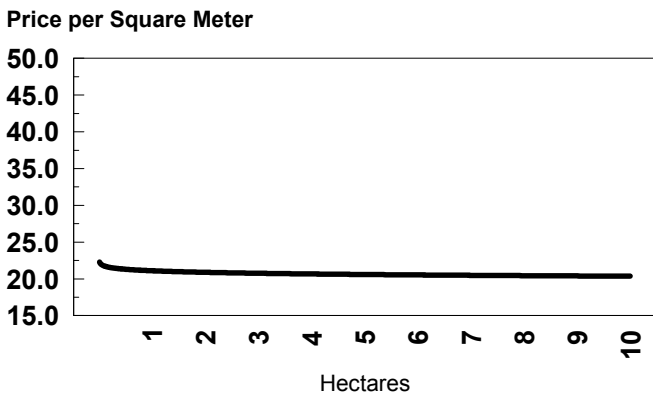Figure 7.1 Effects of Parcel Size to Land Values in Roanoke County
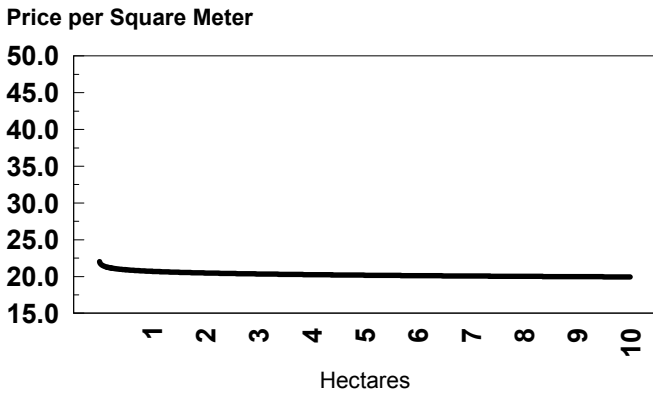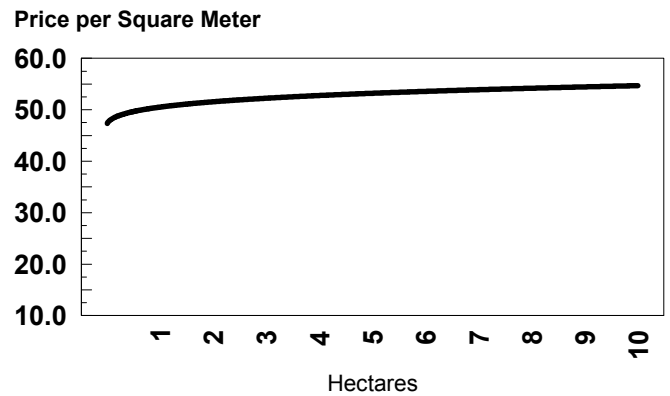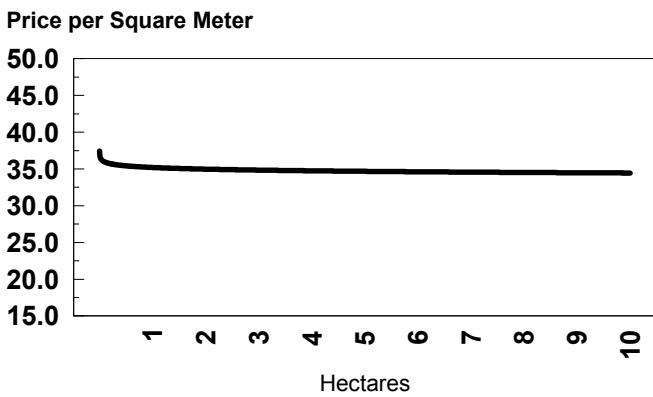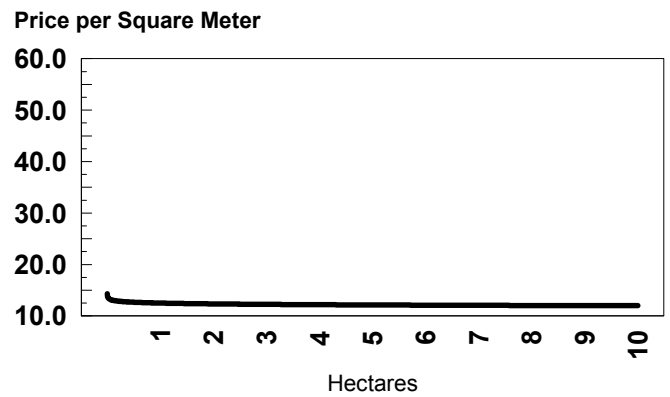
BIBLIOGRAPHY

1. Anselin, L., (1988). *Spatial Econometrics: Methods and Models*, Kluwer Academics.

2. Anselin, L., (1994). *Testing for Spatial Dependence in Linear Regression Models: A Review*,West Virginia University, Regional Research Institute, Research Paper 9416.

3. Anselin, L., (1999). *Spatial Econometrics*, School of Social Sciences, University of Dallas in Texas, Working Paper.

4. Anselin, L. and H. Kelejian (1997). "Testing for Spatial Autocorrelation in the Presence of Endogenous Regressors", *International Regional Science Review*, 20, pp.153-82.

5. Basu, S. and T. G. Thibodeau (1998). "Analysis of Spatial Correlation in House Prices", *Journal of Urban Economics*, 46, pp. 207-18.

6. Beaton, W.P., (1991). "The Impact of Regional Land-Use Controls on Property Values", *Land Economics*, 67, pp.172-94.

7. Bockstael, N.E., (1996). "Modeling Economics and Ecology: The Importance of a Spatial Perspective", *American Journal of Agricultural Economics*, 78, pp. 1168-80.

8. Bockstael, N. E. and K. Bell, (1998). *Land Use Patterns and Water Quality: The Effect of Differential Land Management Controls*, Boston: Kluwer Academic Publishers.

9. Can, A. and I. Megboludge, (1998). "Spatial Dependence and House Price Index Construction", *Journal of Real Estate Finance and Economics*, 14, pp. 203-22.

10. Breusch (1980)

11. Chomitz, K. and D. Gray, (1996). "Roads, Land Use, and Deforestation: A Spatial Model Applied to Belize", *World Bank Economic Review*, 10, pp. 487-512.

12. Cliff, A. D. and Ord, J.K. (1981). *Spatial Processes: Models and Applications*, Pion.

13. Clifton, I. and S. Spurlock, (1983). "Analysis of Variations in Farm Real Estate Prices of a Homogeneous Market Areas in Southeast", *Southern Journal of Agricultural Economics*, 15, pp. 89-96.

14. Cressie, N. (1991) *Statistics for Spatial Data: Part I*, Geostatistical Data, Oxford University Press.

15. Diplas, P. (1998) *From Landscapes to Waterscapes: Integrating Framework for Urbanizing Watersheds*, Project Proposal for the Environmental Protection Agency and the National Science Foundation.

16. Davidson, R. and MacKinnon J. G. (1993) *Estimation and Inference in Econometrics*, Oxford University Press.

17. Fleming, M, M., (1998). *Sample Selection and Spatial Dependence in Hedonic Land Value Models*, SEA Annual Meeting, University of Maryland at College Park.

18. Gleeson, M., (1979). "Effects of an Urban Growth Management System on Land Values", *Land Economics*, 55, pp.350-65.

19. Griffith, D. (1988). *Advanced Spatial Statistics*, Kluwer Academics.

20. Hansen, L.P. (1982). "Large Sample Properties of Generalized Methods of Moments Estimators", *Econometrica*, 50, pp. 1029-54.

21. Irwin, E, G., (1998). *Do Spatial Externalities Matter? The Role of Endogenous Interactions in the Evolution of Land Use Pattern*, University of Maryland, College Park.

22. Kelejian, H.H. and Robinson D.P., (1992). "Spatial Autocorrelation: A New Computationally Simple Test with an Application to Per Capita County Policy Expenditures", *Regional Science and Urban Economics*, 22, pp. 317-31.

23. Kelejian, H. H. and D. P. Robertson (1993). "A Suggested Method of Estimation for Spatial Interdependent Models with Autocorrelated Errors, and an Application to a Country Expenditure Model", *Papers in Regional Science*, 72, pp. 297-312.

24. Kelejian, H. H. and I. Prucha (1997). "Estimation of Spatial Regression Models with Autoregressive Errors by Two Stage Least Squares Procedures: A Serious Problem", *International Regional Science Review*, 20, pp. 103-11.

25. Kelejian, H. H. and I. Prucha (1998). "A Generalized Spatial Two Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances", *Journal of Real Estate Finance and Economics*, 17, pp. 99-121.

26. Kelejian, H. H. and I. Prucha (1999). "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model", *International Economic Review*, 31, pp.68-79.

27. King, D. A. and J.A. Sinden, (1994). "Price Formation in Farm Land Markets", *Land Economics*, 70, pp. 38-52.

28. Land, K. and G. Deane (1992). "On the Large-Sample Estimation of Regression Models with Spatial or Network-Effects Terms: A Two Stage Least Square Approach", *Sociological Methodology*, 12, pp. 221-48.

29. McGuirk, A., Driscoll, P. and J. Alwang, (1993). "Misspecification Testing: A Comprehensive Approach", *American Journal of Agricultural Economics*, 75, pp. 1044-1055

30. McGuirk, A. and P. Driscoll (1995). "The Hot Air in R^2", *American Journal of Agricultural Economics*, 79, pp. 319-328.

31. Moran, P. A. P. (1948). "The Interpretation of Statistical Maps", *Biometrica*, 37, 17-23.

32. Nelson, A. C. (1986). "Towards a Theory of the American Rural Residential Land Market", *Journal of Rural Studies*, 2, pp. 309-19.

33. Palmquist, R. and L. Danielson, (1989). "A Hedonic Study of Effects of Erosion Control in Drainage on Farmland Values", *American Journal of Agricultural Economics*, 71, pp. 55-62

34. Pinkse, J. and I. R. Slade (1998). "Contracting in Space: An Application of Spatial Statistics to Discrete-Choice Models", *Journal of Econometrics*, 85, pp.125-54.

35. Quandt, R. (1972). "A New Approach to Estimating Switching Regressions", *Journal of the American Statistical Association*, 67, 206-10.

36. Roanoke County Planning Department (1994). *Annual Planning Report, 1994*, Roanoke Planning Department.

37. Schmidt, P. (1976). *Econometrics,* New York: Marcel Dekker

38. Spanos, A. (1986). *Statistical Foundation of Econometric Modeling*, Cambridge University Press

39. Spanos, A. (1989). "On Rereading Haavelmo: A Retrospective View of Econometric Modeling", *Econometric Theory*, 5, pp. 405-29.

40. Spanos, A. (1990). "The Simultaneous Equations Model Revisited: Statistical Adequacy and Identification", *Journal of Econometrics*, 44, pp. 87-105.

41. Spanos, A. (1994). "On Modeling Heteroscedasticity: The Student's 'T' and Elliptical Linear Regression Models", *Econometric Theory*, 10, 1994, 286-315.

42. Von Thunen, J., (1826). *Der Isolierte Staat in Beziehung der Landwirtschaft und Nationalokonomie.* Translated in Peter Hall, ed. Von Thunen's Isolated State,Pergamon Press, 1966.

43. Xu F. *et al*., (1993). "Measuring the Contributions of Site Characteristics to the Value of Agricultural Land", *Land Economics*, 69, pp.356-69.

APPENDIX A: Description of Misspecification Tests

Normality

The Jacque-Bera test is based on the difference of the skewness and kurtosis of the series with those from the normal distribution. The statistic is computed as:

$$JB = 4(N-k)(4S^2 + (K-3)^2)/6 \qquad (1)$$

Where S is the skewness, K is the kurtosis, and k represents the number of estimated coefficients used to create the series. Under the null hypothesis of normal distribution, the Jacque-Bera statistic is distributed as chi-squared with 2 degrees of freedom. The reported P-value is the probability that a Jacque-Bera statistic exceeds (in absolute value) the observed value under the null – a small probability value leads to the rejection of the null hypothesis of a normal distribution.

Functional Form

Functional form tests include test for Linearity as well as the Ramsey RESET test. The Linearity tests are based on the significance of the parameter d, in the following regressions:

$$u = c + ax + dx^2 \qquad (2)$$

and

$$u = c + ax + dz \tag{3}$$

where u is the vector of residuals, c is a constant, x is the vector of variables, and z is the vector of cross-products of the explanatory variables. For the Ramsey test, $z = [y^2 \ y^3]$, where y is the vector of fitted variables of y on x, and the superscripts indicate the powers to which these predictions are raised. The significance of b is examined by using the F-statistics for the null hypothesis that $d = 0$.

Heteroskedasticity

The heteroskedasticity test is based on the significance of d and k in the auxiliary regression:

$$u^2 = c + dx^2 + kz \tag{4}$$

where u is the vector of residuals, c is a constant, x is the vector of variables, and z is the vector of cross-products of the explanatory variables. The significance of d and k is examined by using the F-statistics for the null hypothesis that $d = 0$ and $k = 0$.

Spatial Autocorrelation

The spatial autocorrelation test is based on the significance of d in the auxiliary regression:

$$u = xb + dWu + \varepsilon \tag{5}$$

where u is the vector of residuals, x is the vector of explanatory variables, b is the vector of estimated coefficients, W is the weight matrix and ε is the error term of equation (5). The significance of d is examined by using the F-statistic for the null hypothesis that d = 0.

Second Degree Dependence

The second-degree dependence test that we use is known as ARCH test, which is an auxiliary regression test for autoregressive conditional heteroskedasticity in the residuals. The test is based on the significance of a, d and k in the following auxiliary regression:

$$u^2_z = c + au^2_{z-1} + du^2_{z-2} + ku^2_{z-3} \tag{6}$$

where u is the vector of residuals, c is a constant, and z is the ordering factor. The observations are ordered based on their development status (first vacant then developed parcels), on the assessed value of their constructions, and also ordered by neighborhood. The significance of a, d and k is examined by using the F-statistic for the null hypothesis that a = 0 and d = 0 and k=0.

Structural Stability

Two individual tests were used to evaluate structural stability. Both the Chow test and the Chow Forecast test are based on the same principals and their evaluation uses F-tests. The idea behind the Chow test is to fit the equation separately for each subsample and to see whether there are significant differences in the estimated equations. The F-statistic is

based on the comparison of the restricted and unrestricted sum of squared residuals and in the simplest case involving a single breakpoint, is computed as:

$$F = [(u'u - u_1'u_1 - u_2'u_2)(T - 2k)] / [(u_1'u_1 + u_2'u_2) k] \qquad (7)$$

where $u'u$ is the restricted sum of squared residuals, $u_i'u_i$ is the sum of squared residuals from subsample I, T is the total number of observations, and k is the number of parameters in the equation.

The Chow Forecast test estimates the model for a subsample comprised of the first T1 observations. The estimated model is then used to predict the values of the dependent variable in the remaining T2 data points. A large difference between the actual and predicted values cast doubt on the stability of the estimated relation over the two subsamples. The F-statistic is computed as:

$$F = [(u'u - u_1'u_1)(T_1 - k)] / [u_1'u_1 T_2] \qquad (8)$$

where $u'u$ is the residual sum of squares when the equation is fitted to all T observations of the sample, $u_1'u_1$ is the residual sum of squares when the equation is fitted to T_1 observations, and k is the number of estimated coefficients.

Joint Mean Test

The joint mean test simultaneously checks the appropriateness of functional form, independence, and structural stability as each of these assumptions refers to aspects of conditional mean. The test is based on the auxiliary regression:

$$u = c + ax + dx^2 + kWu + qT \tag{9}$$

where u is the vector of residuals, x is the vector of explanatory variables, W is the weight matrix, and T is a binary variable with 0 before the break point and 1 after. The significance of a, d and k is examined by using the F-statistics for the null hypothesis that $d = 0$ and $k = 0$ and $q=0$. The most likely cause of rejection, if it results, can be investigated by assessing the significance of d, k, and q individually.

Joint Variance Test

The joint variance test simultaneously checks for first and second order spatial dependence as well as for structural stability. It is based on the auxiliary regression:

$$u^2_z = c + ax + dx^2 + ku^2_{z-1} + qT \tag{10}$$

where u is the vector of residuals, x is the vector of explanatory variables, T is a binary variable with 0 before the break point and 1 after, and z is the ordering factor. The significance of a, d and k is examined by using the F-statistic for the null hypothesis that $d = 0$ and $k = 0$ and $q=0$. The most likely cause of rejection, if it results, can be investigated by assessing the significance of d, k, and q individually.

Redundancy Test

The redundancy test checks whether the entrance of some explanatory variables in the model is necessary. It is based on the regression:

$$y = c + ax + dz + u \tag{11}$$

where y is the vector of dependent variables, x, z are vectors of explanatory variables and u is the vector of residuals. Testing whether the vector z should enter the model, we examine the significance of coefficient d, by using the F-statistic for the null hypothesis that $d = 0$.

Ioannis Kaltsas

20 Fleming Street, Trikala
TK 42100

Education

Doctor of Philosophy	*Virginia Polytechnic Institute & State University, USA*
	Ph.D. in Economics
	Blacksburg, December 2000 (expected**)**
	Major: Applied Econometrics
	Minor: Environmental Economics
Master of Science	*Virginia Polytechnic Institute & State University, USA*
	MSc in Applied Economics
	Blacksburg, December 1988
	Major: Econometrics
	Minor: Environmental Economics
Master of Arts	*University of East Anglia, UNITED KINGDOM*
	MA in Industrial Economics
	Norwich, December 1996
	Major: Industrial Economics
	Minor: Quantitative Methods
Ptychio (BSc + MSc)	*Aristotle University of Thessaloniki, GREECE*
	BSc+MSc in Agricultural Economics, Rural Sociology
	Thessaloniki, July 1995
	Thesis: Economic development and environmental management of the French Champaign industry. Lessons for the future.

Recent Professional Experience

Graduate Research Associate	*University of Nottingham, UNITED KINGDOM*
	Nottingham, Oct. 1997 – Jul. 1998

- Econometric Modeling of Industrial Efficiency and Productivity Differentials
- Establishing Methodology of Measuring Technical & Allocative Efficiency in UK Potato Industry

Stageur du Service Economique	*Comité Interprofessionnel du Vin de Champagne, FRANCE*
	Epernay, Sep. 1994 – Dec. 1994

- Statistical Analysis on Intra-European Trade of Champagne

Consultant	*Ministry of Agriculture & Rural Affairs, GREECE*
	Trikala, Jul. 1994 – Aug. 1994

- Advisor on European Union Development Projects

Teaching Experience_____

Visiting Lecturer   *Textile International School, Lubbock TX*
                           International Cotton School, Rhodes College, TN
Teaching Assistant/   *Virginia Polytechnic Institute & State University*
   Instructor   **1998/99 Econometrics Graduate Level, Intermediate Microeconomics**

Awards, Scholarships_____
- *1990*: Greek National Scholarship (5 years)
- *1994*: European Community COMET program (6 months)
- *1997:* European Union and NSF Scholarship (for 3.5 years)

Languages_____
- Greek: Native Speaker
- French: Fluent
- English: Fluent
- Spanish: Beginner

Selected Literature_____

1. "Short and Long Term Forecasting Methodology of Commodity Prices", *20th International Symposium on Forecasting*, Lisbon, June 2000 (with John Baffes, Don Mitchel and Randy Verbrugge)
2. "The Effects of Clandestine Immigrant Workers on the Technical Efficiency: An Empirical Application on the Greek Agriculture", SAEA Annual Meeting, (Abstract) *Journal of Agricultural and Applied Economics*, **28**, pp. 67
3. "Explaining Technical Inefficiencies in Greek Food and Beverage Industries", *Journal of Food Distribution Research*, **30**, pp. 48-62
4. "Comparison of Production Frontier Methods and Alternative Functional Forms: A Study of Greek Cotton Farms", *American Agricultural Economics Association Annual Meeting*, August 2-5, 1998, Salt Lake City, Utah
5. "Measuring and Explaining Technical Efficiency in UK Potato Production", *Journal of Agricultural Economics*, **49**(3), pp.22-31(with S. Ramsden and P. Wilson)
6. "Restructuring Agricultural Production in Albania: A Bootstrap Approach", *Quarterly Journal of International Agriculture*, Forthcoming (with E. Zaloshja, D. Taylor and A. McGuirk)

Computer Skills_____

- *Familiar with*: UNIX, NOVELL, DOS and Windows environment
- *Statistic Packages*: SAS, LIMDEP, TSP, SPSS, FRONTIER, GAUSS, EVIEW, SPACESTAT, MATLAB

Personal Data_____

- *Name*:   Ioannis Kaltsas
- *Permanent Address*: 20 Fleming Street, Trikala, TK 42100, GREECE
- *Telephone*:   ++3 – 0431 - 25410
- *Date of Birth*   18/03/1973