METHOD AND APPLICATION OF SPATIAL PROBIT MODEL TO THE BUSINESS RETURN TO NEW ORLEANS AFTER HURRICANE KATRINA

THESIS

Presented to the Graduate Council of
Texas State University-San Marcos
in Partial Fulfillment
of the Requirements

for the Degree

MASTER OF SCIENCE

by

Xingjian Liu, B.S.

San Marcos, Texas
May 2010

ACKNOWLEDGMENTS

Though only my name appears on the cover of this thesis, a great number of people have contributed to its production. I owe my gratitude to all those people who have helped and inspired me during my Master's study at Texas State University-San Marcos and because of whom my stay at San Marcos has been one that I will cherish forever.

My deepest gratitude goes to Dr. James P. LeSage and Dr. F. Benjamin Zhan for their guidance throughout my study at Texas State. As part of their collaboration, I was able to enter an interdisciplinary field of study. This precious experience has widened my horizon and let me have the best of both economics and geography at Texas State. Their enthusiasm in research and pursues in scientific discovery have motivated me, and I had also benefited tremendously from their expertise. In addition, they were always accessible and willing to help me with every single aspect of my study and research, ranging from teaching me research methods and softwares, to exploring scientific questions, to editing and publishing research papers, and to understanding how academia functions. As a result, my academic development at Texas State became smooth and rewarding. I have been amazingly fortunate to work with Dr. LeSage and Dr. Zhan, and I hope that one day I would become as good of an advisor to my students as Drs. LeSage and Zhan have been to me.

I'd like to thank all the faculty members of the Department of Geography for their guidance and encouragement throughout the Master's study. Special thanks go to

Dr. Lawrence E. Estaville and Dr. Yongmei Lu, who have been wonderful mentors. With the breadth and strength of our geography department, I was able to develop towards a true geographer.

I'd also like to express my gratitude to all the Chinese students and scholars in Texas State. Their support and care helped me overcome setbacks, adjust to a new country, and stay focused on my graduate study. I greatly value their friendship and I deeply appreciate their belief in me. Life also became more colorful with my fellow Chinese students and scholars in San Marcos.

Most importantly, none of this endeavor would have been possible without the love and support of my parents, Yaolin Liu, and Yanfang Liu. I am heartily thankful to my parents for introducing me to the discipline of geography. With my parents being professors of geography, I have been long determined to obtain a Ph.D. in this fascinating discipline, and ultimately pursue an academic life. I still remember I told my mother when I was five that "I still need twenty-two years of study to get a Ph.D.". With the completion of my Master's study, I am one step closer to that target.

This manuscript was submitted on 3 April 2009.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

METHOD AND APPLICATION OF SPATIAL PROBIT MODEL TO THE BUSINESS RETURN TO NEW ORLEANS AFTER HURRICANE KATRINA

by

Xingjian Liu, M.S.

Texas State University-San Marcos

May 2010

SUPERVISING PROFESOR: F. Benjamin Zhan

This study employs a theoretical framework from micro-scale retail location studies and implements a spatial autoregressive probit model to account for spatial dependence among firms' decisions and thus identify determinants of business return to New Orleans after Hurricane Katrina. The spatial probit approach allows for interdependence between decisions to reopen by one establishment and those of its neighbors. There is a large literature on the role played by spatial dependence in firm location decisions, and we find evidence of strong dependence in firm's decisions to reopen in the aftermath of a natural disaster such as Katrina. This interdependence has

important statistical implications for how we analyze business recovery after disasters, as well as government aid programs. In order to determine the right model specification, a Monte Carlo experiment is conducted to extends information criteria for selecting alternative model specifications in spatial econometric modeling, and provides some insight about performance of different model selection tools for choosing a spatial weight matrix.

Chapter 1

INTRODUCTION

1.1 Brief background

Hurricane Katrina struck New Orleans, Louisiana, on August 29, 2005. The city suffered severe damage from the hurricane and associated floods caused by levees failures. The city's population, estimated at 485,000 in 2000, declined to fewer than several thousand by the end of the first week of September 2005 (McCarthy, Peterson, Sastry, and Pollard 2006). The city of New Orleans suffers financially as well with only two thirds of the city's original businesses in New Orleans parish reopened even two years after the catastrophe (Arenas and Lam 2009). In the aftermath of a disaster such as Hurricane Katrina, individual firms must make decisions about either investing in repairs necessary to restore operations or going out of the old business, i.e., these firms' decisions are binary. Research about determinants of business return after a disaster, and factors influencing business recovery can produce insights for post-disaster management. Both empirical and theoretical analyses suggest inter-dependence among businesses' decision-making.

Casual observation of how businesses operate suggest that decisions made by an individual business would likely influence decisions of neighboring establishments, and vice versa (Holloway, Shankara, and Rahman 2002, 383-402; Zhou and Kockelman 2009, 321-40). In the case of retail or entertainment firms, traffic generated by neighboring establishments can be an important factor in generating spatial spillover business. In other words, customer traffic on a street may depend on the number of neighboring establishments in operation. Businesses may benefit from goods, marketing activities, promotions, and ultimately customer traffic in the neighborhood. For example, patrons of a restaurant may also patronize neighboring entertainment venues, art galleries, or retail shopping establishments. Spatial spillover business can arise from neighboring establishments that offer competing or complementary products and services. For example, neighboring restaurants (competing businesses) located on the same street may generate spatial spillover business because this clustering attracts patrons to the area. In the meantime, retail shops in the neighborhood (complementary businesses) may also help to attract potential patrons. Moreover, a commercial street may decline when several businesses are closed and a consumer traffic threshold is not sustained. This phenomenon is common in the central business districts of old industrial cities, where decline of population and business closures form a vicious cycle (Keeble 1978). As for business return to New Orleans after Hurricane Katrina, this empirical observation could lead to an extreme question: Would a single firm located on a street decide to re-open knowing that all neighboring firms on the street have decided not to re-open?

From a theoretical perspective, the specific economic mechanism at work here is that some part of the unobserved net profitability associated with the decision to re-open derives from spatial spillover business. The commodity price, sales quantity, and variable costs of operations generally depend on the customer flow, which are affected by spatial spillovers generated by customer traffic from neighboring firms (Ozmen-Ertekin 2005, 293-331). Therefore, the latent unobservable net profitability will depend on neighboring establishments' decisions. If a firm has positive profitability, some sales revenue may reflect spillover traffic and revenue from neighboring firms. The conventional econometric models used for analyzing these binary decision outcomes are probit models (Anselin et al. 2004, 169-92). These models attempt to explain variation in the set of decision outcomes as the dependent variable, where returning to business is indicated by a dependent variable value of one and a decision to go out of business is indicated as a zero value. The independent variables in these models would be a cross-sectional sample of firm and neighborhood characteristics. These probit models can explain the dependent outcomes in terms of independent variables and test the relative strength of relationships among model variables. Still, as the location of firms may exert influences on their decisions, spatial effects needs to be integrated into the probit model. Therefore I will therefore employ spatial econometric models (Anselin 1988) to characterize the relationships between business recovery and the factors.

1.2 Significance of research

This study focuses on some methodological and applied issues on devising a spatial econometrics model to quantify determinants of business return after a disaster. The results can generate significant implications for management, planning, and the recovery of business in New Orleans. More specifically, my study will: (1) solve methodological issues such as model selections among alternative specifications; (2) implement spatial econometric models to characterize the relationships between business recovery and the factors; (3) generate meaningful insights about post-disaster business recovery and management.

1.3 Statement of research problem

This research project aims at implementing a spatial probit model to account for spatial dependence among firms' decisions when exploring determinants of business return to New Orleans after Hurricane Katrina.

Chapter 2

LITERATURE REVIEW

2.1 Critical review of relevant literature

I will review two groups of literature for my research study: studies of retail location at micro-scale, and econometric modeling and estimation of spatial interdependence. The former literature argues for considering spatial effects in retail locations and provides a conceptual framework for econometric modeling, while the latter literature discusses how to incorporate spatial effects into econometric models, including specification of spatial models, estimation of model parameters, as well as model comparisons.

2.1.1 Review of micro-scale retail location studies

Researchers have conducted various studies on business decision-making regarding location in space at various spatial scales. For example, scholars have performed macro-scale studies about regional specialization, meso-scale analysis about urban economics, and micro-scale decision-making about retailer's location. Researchers have also analyzed business location from different perspectives, for example, geography, marketing, economics, and management (Simon 1959; Simon 1979; Bartik 1985; Malmberg 1997; Porter 2000; Krugman 2001, 69-98). However, little research on business decision-making exists after a catastrophic event such as Hurricane Katrina (Birkland and Nath 2000; Runyan 2006). Few studies have attempted to formally model the role of business connectivity and interdependence at the individual firm level in general and after Hurricane Katrina in New Orleans in particular. Most studies regarding economic aspects of disasters focus primarily on understanding the regional and macroeconomic impacts, rather than the street or firm-level economics (Guimaraes, Hefner and Woodward 1992; Tierney and Webb 2001). Understanding spatial dynamics and linkage among individual business decisions in the aftermath of a disaster is critical to planning for redevelopment, city management with both local and regional concerns, and ultimately the recovery of business for sites that have experienced or will experience disasters (Berke, Kartez and Wenger 1993). However, such studies are mainly constrained by limited data availability, the uncertain environment after disasters, and a lack of proper statistical models (McCarthy et al. 2006).

Studies of businesses return to their former locations after a disaster can be treated as special type of retail location study at the micro-scale, because individual firms can only make binary decisions on their former locations: (1) investing in repairs necessary to restore business operations or (2) going out of business. Thus, a research study may draw upon those theoretical and empirical studies on retail location and outlets' interactions. Literature on street-level retail location has produced a consensus on the economic agglomeration of similar or related shop types and dispersal of various types of spatial clusterings throughout the retail districts (Brown 1994; Carter and Haloupek 2000; Ottaviano and Thisse 2002). Various theories and models also identify the interactions between a business, its neighborhood, and nearby businesses, which underpin the observed spatial arrangement of businesses and ultimately affect, if not determine, the location choices of individual businesses. These theories and models follow three major approaches: theoretical, demand-side, and supply-side approaches (Brown 1994).

Literature on street-level retail location (in non-disaster situations) has long emphasized the importance of spatial interaction in firm-level decisions. For example, Reilly (1931) set forth a "law of retail gravitation", drawing on an analogy with Newton's gravitational law as it related to retail shopping behavior and store location decisions. Meanwhile, the principle of minimum differentiation and its modifications are capable of explaining the observed clustering of competitive and compatible retail outlets within shopping districts (Hotelling 1929; Rhee 1989; Fujita and Smith 1990; Hinloopena and van Marrewijk 1999; Brown 2002, 450-67; Liang and Mai 2006). This group of models starts with assumptions concerning market conditions and retailer behaviors and demonstrates that individual firms make location decisions based on other firms' locations to maximize its profit. While the principle of minimum differentiation and models based on 'externalities' account for existence of clustering, the bid-rent theory helps to explain how clusters of different types of businesses spread over space (Alonso 1960; Craig, Ghosh, and McLafferty 1984; Guy 1995; Fujita and Thisse 1996). This theory contends that the need for a more accessible location differs between various types of retail activity that, in turn, affects the rent differential retailers are willing to bid for a location. In bid-rent theory the individual firm's choice on location is the result of interactions between the businesses, e.g., the bid-rent curve of an individual firm and that of neighboring businesses, because competitive bidding among firms takes place for locations with superior market accessibility (O'Roarty, McGreal, and Adair 1997; Carter and Vandell 2005). Another related theoretical framework would be 'Central Place Theory' (Christaller 1963), which describes the distribution and hierarchy of retail locations based on the range and the threshold of a good. Although it is usually applied to national and regional retail patterns, the spatial interaction between clusters discussed in Central Place Theory is still valid for street-level retail. Although these theoretical approaches were initially designed to explain economics of agglomeration, all of them incorporate spatial interaction and interdependence among firms as their core.

Most aforementioned studies explain the observed firms clustering and interaction deductively from a top-down perspective. Approaches that directly generalize street-level market participants' spatial behavior via a 'bottom-up' design also exist (Brown 1994). Analyses from both demand and supply ends suggest that street-level retail location decisions should consider neighboring firms' decisions, and there is evidence that some retail location decisions are made explicitly with consideration of interactions among neighboring competitive or complementary establishments. As for demand-side analysis, general consumer behaviors at street-level retail are investigated. For example, Reilly (1931) set forth a "law of retail gravitation", drawing on an analogy with Newton's gravitational law as it related to retail shopping behavior and store location decisions. Other studies on consumer movement and behavior include analysis of consumer interchange between retail outlets (Nelson 1958), gravitational effects of distances (Reilly 1931; McGoldrick and Thompson 1992) and pedestrian-distance minimization (Garling and Garling 1987), evaluation of spatial convenience in shopping districts (Reimers and Clulow 2004), mental mapping of retail locations (Golledge and Timmermans 1990; Golledge and Stimson 1996), econometric analysis of retail demand regarding the geographic distribution of consumers (Davis 2006) and more recently, computer simulation of grocery shopping behavior (Hanaoka and Clarke 2007; Schenk, Loffler, and Rauh 2007). These analyses arrive at some consensus about consumer behaviors, that can be used to motivate spatial interactions among retail outlets from another perspective: Nelson's 'rule of retail compatibility' and 'theory of cumulative attraction' (Nelson 1958) state that both compatible businesses (those selling related types of goods) and competitive businesses (those selling the same products) in close proximity benefit from customer flows and interchanges drawn by each other. This is becoming increasingly true with customers' preference on multi-purpose or one-stop shopping and customers' increasing desire for comparative shopping prior to making a purchase. It has also been empirically justified that increased market size from agglomeration may offset the fierce competition from clustering (Konishi 2005).

Analysis from the other side the market empirically investigated retailers' spatial strategies finding that micro-scale location decisions often take into account location of competitive or complementary establishments (Berry and Garrison 1958; Berman and Evans 1991; Borchert 1998), and agglomerated retailers are well disposed towards cumulated attraction from clustering (Brown 1987). Although consumer behaviors exert great influence on the extant pattern of retail clustering, it is also true that supply side decision makers, be they in public or private sectors, can affect consumers' behavior via spatial choices of retail locations (Craig, Ghosh, and McLafferty 1984). For example, researches have demonstrated that the location of magnet stores and marketplace entrance/exit affects customer movements (Brown 1994). Therefore, it has also been documented that spatial tools, such as central places theory, spatial interaction theory, and cumulative attraction analysis, have been adopted as pragmatic methods in retail planning (Hernandez, Bennison and Cornelius 1998). Additionally, models in operations research have been implemented to optimally allocate resources over space from a supply side stakeholder's perspective, which explicitly consider spatial interaction among firms (Drezner and Guyse 1999; Drezner and Drezner 2002). Although it has been argued that the extant retail pattern may stem from other non-market spatial processes, for example, historical events and inertia to relocate (Krugman 1992), and local preference (Brown 1989), all of these spatial processes emphasizes the importance of spatial interaction.

The foregoing theories and models arrive at the conclusion that extant patterns of retail activity influence choices of consumers (or patrons) and location decisions by retailers. The decision of a business establishment to re-open after a natural disaster such as Katrina could be viewed as similar to the original location decision made by the firm when it entered business. If the surrounding environment was important for the original decision, it should be equally influential in the decision to re-open in the same or another location after the disaster. This consensus provides a strong motivation for adopting an empirical modeling method that allows one firm's decision to re-open to depend on similar decisions made by neighboring firms.

2.1.2 Review of econometric modeling of spatial interdependence

2.1.2.1 *Spatial Probit Model and its estimation*

Econometric analysis relies heavily on regression models, which test economic relationships between dependent and explanatory variables in general. In my case, such analysis will quantify the relations among various factors contributing to business return decisions in the aftermath of the Katrina disaster (Kennedy 2001, 1-5; Koop 2003, 15). The ordinary linear squares (OLS) model, which is commonly used in econometrics, is not suitable for analyzing data with spatial dependence. The peculiarities caused by space, e.g., spatial interactions between neighboring businesses' decisions violate the Gauss-Markov assumption underlying OLS: all observations are independent (Anselin 1988, 5).

Researchers, therefore, have developed a group of spatial econometrics models to deal with spatial dependence in a regression context (Anselin 1988, 1-51; Anselin and Bera 1998; LeSage and Pace 2009, 25-42). These models explicitly incorporate spatial dependence using a weight matrix and spatial autoregressive processes allowing investigators to capture the global and local spatial externalities (Anselin 2003). However, many of these models deal only with continuous dependent variables and cannot interpret binary choice outcomes. Consequently, much attention in the spatial econometric literature has been given to models that analyze limited dependent variables. The common rationale behind the developed approaches considers treating the observed binary or categorical outcomes as indicators or reflections that relate to underlying unobserved and continuous latent variable (Franzese and Hays 2008). For example, the unobserved latent variable is the profitability or net utility of individual firms in my case. The models developed for limited dependent variables include spatial tobit model (McMillen 1995), spatial probit model (Smith and LeSage 2000), spatial multinomial-probit model (Bolduc, Fortin, and Gordon 1997), and spatial zero-inflated Poisson model (Rathbun and Fei 2006), each of which focuses on analyzing different types of dependent variables, e.g., binary, categorical, and censored data.

Among these models, the spatial probit model is suitable for analyzing binary outcomes. Many studies have demonstrated the usefulness of spatial probit model in a wide range of applied research. Scholars have devised spatial probit models to study the spatial dependence among various natural, social, and economic phenomena including spatial dependence in adoption of agricultural programs (Hollowy, Shankara, and Rahman 2002); influence of public services and neighborhood characteristics on relocation decisions of home-buyers (Ozturk and Irwin 2001); landuse conversion regarding neighboring property ownership status (Irwin and Bockstael 2004; Zhou and Kockelman, 2008); regeneration of trees in neighboring areas (Rathbun and Fei 2006); spatial externalities in voting preferences in the ”political-economy-of-agriculture” (Holloway, Lacombe, and Shaughnessy 2005); geographic patterns in state lotteries (Coughlin, Garrett, and Hernndez-Murillo 2004). Still, rare studies produce empirical models that quantify the determinants of business decisions in the highly uncertain and spatial interactive environment after a disaster.

Estimating spatial probit model is complicated because of colinearity and heteroskedasticity introduced by spatial interdependence. More specifically, the difference between estimating the nonspatial regression model and the spatial probit model with a spatial autoregressive progress is that the spatial dependence leads to the problem of an n-dimensional integration involving truncated multivariate normal distribution (TMVN) (LeSage and Pace 2009, 283-85). Scholars have proposed several methods to estimate spatial probit model (Flemming 2004); however, most of these efforts do not address the n-dimensional integration problem directly. Methods focus on treating heteroskedasticity induced by spatial dependence produce consistent parameter estimation by grouping elements in the variance-covariance matrix into independent blocks (Case 1992) and Generalized Method of Moments (Pinkse and Slade 1998). These methods do not utilize information in the off-diagonal terms of the variance-covariance matrix and leave the problem of multidimensional integration of the variance-covariance matrix unaddressed. The Expectation Maximimization (EM) algorithm utilizes the complete variance-covariance matrix, and attempts a solution to this n-dimensional integral problem (McMillen 1992). The EM algorithm, however, requires a large sample size to justify the validity of its underlying asymptotic properties and still produces biased estimates (LeSage 2000). Moreover, the EM method focuses on the Spatial Error Model (SEM) instead of the Spatial Autoregressive Model (SAR). The Bayesian method for nonspatial probit model (Albert and Chib 1993) is extended to a spatial probit model using Bayesian Markov Chain Monte Carlo (MCMC) methods (LeSage and Pace 2009, 288-300). Smith and LeSage (2000) have proposed another Bayesian MCMC method for spatial probit models that introduces an additive error specification. The Bayesian methods provide consistent model estimates and combine all relevant information regarding the estimation process including both objective sample data information as well as prior or subjective information about model parameters (Zellner 1971, 35-60; Koop 2003, 16-25). The Bayesian MCMC framework solves the problem of n-dimensional integration and is more flexible and able to work with small samples. Many scholars have argued that the Bayesian MCMC is currently the only feasible approach to estimate spatial probit models with a spatial autoregressive process (Smith and LeSage 2000; Flemming 2004; Franzese and Hays 2008). Beron and Vijverberg (2004) provided an alternative approach to solve this n-dimensional integration problem based on Maximum Likelihood Estimation, which is extremely time-consuming. The computation time increases quadratically when the sample size is doubled, and estimation with moderate sample sizes take a substantial amount of time.

2.1.2.2 *Specifications and comparisons of spatial econometrics models*

As for model specifications, there is a great deal of published research that compares competing model specifications that model dependence in the error structure versus the dependent variable. The former are referred to as spatial error models (SEM) taking the form: $y = X\beta + \mu$, $\mu = \rho W\mu + \varepsilon$. The latter have been labelled spatial autoregressive models (SAR): $y = \rho Wy + X\beta + \varepsilon$. As pointed out by LeSage and Pace (2009), a linear combination of the SEM and SAR models will lead to the spatial Durbin model specification, which subsumes models that incorporate spatial dependence in the dependent variable and error terms, $y = \rho Wy + X\beta_1 + WX\beta_2 + \varepsilon$. The choice of spatial regression model specification, and more specifically, specification of the spatial connectivity structure used, plays an important role in applied work, since changes in these choices lead to differences in estimates and inferences (Anselin and Rey 1991; Florax and Rey 1995; Smith 2008). Spatial connectivity structure, usually reflected in the spatial weight matrix provides a formal expression of connectivity or dependence in space (Anselin 1988). Spatial weight matrices are generated from the locations of observations, and can be generally constructed in three ways (Stakhovych and Bijmolt 2008): (1) as completely exogenous primarily according to the geographical pattern of observations (Cliff and Ord 1981; Anselin 1988; LeSage and Pace 2009); (2) based on existing data and in accordance with the nature of underlying spatial association (Getis and Aldstadt 2004); (3) by treating weigh-matrix as model coefficients (Bhattacharjee and Jensen-Butler 2006). Research has shown that matrix specification issues relating to different aspects of constructing weight matrices (Anselin and Rey 1991; Florax and Rey 1995), mis-specified weight matrices (Griffith and Lagona 1998), and strongly or weakly connected weight matrices (Smith 2008), can impact statistical tests and model estimation.

Identifying the appropriate model specification in general and spatial connectivity structure in particular can be viewed as the selection of model that is "closest" to the underlying data generating process. In conventional econometrics, information criteria such as the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Deviance Information Criterion (DIC) are widely used in model selection, along with other approaches such as RMSE, Mallows' C_p, the coefficient of determination, and Bayesian model comparison approaches. However, few studies have evaluated the performance of various information criterion for selecting spatial weight matrices, and even fewer have evaluated these in the context of spatial econometric modeling. For example, Hoeting et al. (2006), and Lee and Ghosh (2009) have discussed the performance of information criteria for geostatistical models, while discussions on information-based model selection is prevalent for ecological models (Ward 2008; Kissling and Carl 2008; Diniz-Filho et al. 2008; Murtaugh 2009). As for application in spatial econometrics, Stakhovych and Bijmolt (2008) have evaluated the ability of both AIC and a consistent AIC measure to discern different spatial weight matrices, though their results seem to contradict Bayesian model comparison results (LeSage and Pace 2009). Even fewer studies have focused on selection of spatial models in binary or categorical cases, as it is difficult to implement model selection tools, such as the log-marginal likelihood, in situations where the dependent variable is not normally distributed, does not exhibit constant variance, or is discrete rather than continuous. The exact log-marginal likelihood which forms the basis of formal Bayesian inference regarding model comparison has been derived for simple spatial econometric models that involve continuous dependent variables (LeSage and Parent 2007). However, the log marginal likelihood approach requires integration over all model parameters, making this approach difficult to apply in a spatial probit setting. Asymptotically the AIC, BIC, DIC and log marginal likelihood model comparison criteria should produce identical results.

2.2 Theoretical framework

As mentioned above, my research study tries to formally model spatial interdependence in the decision-making process after a disaster at a firm level. Because studies of businesses return to their former locations after a disaster are a special type of retail location study at a micro-scale, my study will be built upon the principle of minimum differentiation.

The modified principle of minimum differentiation with the incorporation of spatial externalities and agglomeration economics from spatial proximity among businesses can model the spatial interdependence among the decision-making of individual firms (Hotelling 1929; Rhee 1989; Fujita and Smith 1990; Liang and Mai 2006). This theory begins with the situation that "two profit maximizing firms, selling identical products from fixed locations in a bounded linear market where transport costs are constant, demand is completely inelastic and identical, utility maximizing consumers are evenly distributed, bear the costs of distribution and patronize outlets solely on the basis of delivery prices" (Hotelling 1929). This basic economic theory then incorporates more assumption including multiple firms, competitive and complementary products, customer interchanges, random utility, customer behavior, and economic externalities (Rhee 1989; Fujita and Smith 1990; Liang and Mai 2006). After mathematical modeling, the complete set of these assumptions arrives at the conclusion that firms selling competitive and complementary products would cluster over space to achieve an economic equilibrium among all firms. During progress toward this equilibrium, the current distribution of retail activity influences choices of consumers and location decisions by retailers.

The principle of minimum differentiation provides a strong motivation in my research study for adopting an empirical modeling method that allows one firm's decision to re-open to depend on similar decisions made by neighboring firms. The decision of one firm may affect the utilities of neighboring firms through a group of interrelated spatial externalities; for example, the total customer flow and interchange in the neighborhood, the competition for local market shares, the investment on restoring the shop front. This interdependence in utilities can ultimately determine firms' choices on returning.

2.3 Specific connections with problem statement

Little research exists about business decision-making after a catastrophic event. In other words, studies on modeling and measuring determinants of the decisions by businesses to return to their former locations after disasters are rare. Spatial relations between a business, its neighborhood, and businesses located nearby can generate significant implications for management, planning, and the recovery of business in New Orleans as well as in other sites of future disasters. To conduct such research, appropriate data observations as well as suitable statistical models are necessary. Theories and models about micro-scale retail location arrive at the conclusion that existing patterns of retail activity exert influence on choices of consumers and location decisions by retailers. This consensus provides a strong motivation and theoretical justification for adopting an empirical modeling method that allows one firm's decision to re-open to depend on similar decisions made by neighboring firms. A spatial probit model with spatial autoregressive processes is most suitable for analyzing binary choice outcomes, like the re-opening decisions of individual firms. More specifically, the Bayesian Markov Chain Monte Carlo approach is currently the only feasible method to estimate a spatial probit model with a spatial autoregressive process (Flemming 2004).

I will therefore utilize a unique survey dataset and undertake the study on method and application of spatial probit models to analyze business return in New Orleans after Hurricane Katrina. This research project uses a theoretical framework from micro-scale retail location studies and can help to answer the overarching problem: how to model and measure determinants of and spatial interdependence between the decisions by businesses to return to their prior location after a disaster, and is the spatial probit model with a spatial autoregressive process suitable for explaining business return to New Orleans after Hurricane Katrina?

Chapter 3

RESEARCH DESIGN

3.1 Working hypotheses

To answer the research question “Can a spatial probit model account for spatial dependence among firms’ decisions and thus identify the determinants of business return to New Orleans after Hurricane Katrina?” I will develop a set of working hypotheses to guide the formal modeling of business connectivity and interdependence in decision-making at the individual firm level.

1. The existing distribution of retail activity at a micro-scale influences choices of consumers and location decisions by retailers.

2. Studies of businesses return to their former locations after a disaster are special types of retail location decision at the micro-scale, and therefore exhibit spatial interdependence among returning decisions of neighboring firms.

3. The spatial probit model with spatial autoregressive specification (SAR) is capable of capturing and quantifying the spatial interdependence among binary outcomes in the neighborhood and, therefore, more consistent with the survey data.

4. Model comparison tools can identify the model, closet to the true underlying data generating process, among alternative spatial probit specifications.

3.2 Study area

New Orleans is located on the banks of the Mississippi River, approximately 105 miles upriver from the Gulf of Mexico. The majority of the metropolitan area spreads along the land between the Mississippi River on the south and Lake Pontchartrain on the north. The city's elevation ranges from 5 feet below mean sea level (MSL) to 15 feet above mean sea level, with 49 percent of the terrestrial surface of the metropolitan area lying below sea level. New Orleans has a humid and subtropical climate with average temperature of 68.1 Fahrenheit degrees and average annual precipitation of 61.88 inches. The city's geographical locations, which is surrounded by large water bodies, and its climate makes it vulnerable to Atlantic hurricanes and accompanying floods (Campanella 2006).

Businesses in New Orleans, as well as the city as a whole, suffered severely from Hurricane Katrina and the consequent drainage failures in late August 2005. Only 26 percent of the city's original businesses were open in December 2005, approximately four months after Hurricane Katrina. In June 2006, the percentage of returned businesses increased to 39 percent; In October 2007, more than two years after Katrina, only about 66 percent of the businesses reopened (Arenas and Lam 2009). More specifically, my research will analyze the reopening statuses of 673 businesses along three commercial streets in New Orleans: St. Claude Avenue, Magazine Street, and Carrollton Avenue. These streets are important commercial corridors in New Orleans, and all of them lie on the north bank of the Mississippi River and south of Lake Pontchartrain. I constrain my analysis to these three streets due to data availability. I will use a survey dataset that reports only about the reopening statuses of businesses on these three streets. Moreover, these streets are representative in terms of flood conditions in the aftermath of Hurricane Katrina, and surveyors believe that the data collected are appropriate for analyzing businesses reopening decisions.

3.3 Study period

I will analyze the reopened status of all firms located on three previously-mentioned streets during the periods from 0 to 3 months, 0 to 6 months and 0 to 12 months after Katrina. While the availability of survey data pre-determined the study period, the length of study period is sufficient for revealing spatial dynamics of businesses reopening decisions.

3.4 Data sources

I will use a unique first-hand survey data set collected by Mr. Richard Campanella, Assistant Research Professor of Earth and Environmental Science at Tulane University, Louisiana. Prof. Campanella is a New Orleans based researcher and has authored several books on physical, human, and environmental geography of New Orleans. Prof. Campanella began surveying businesses open/closed status on the selected streets from October 9, 2005, approximately one month after Hurricane Katrina until November 21, 2006. Prof. Campanella reported the business open/closed status for 673 businesses along three commercial corridors: St. Claude Avenue, Magazine Street, and Carrollton Avenue. He collected the data on a weekly basis during the year following Hurricane Katrina, then seasonally and annually in subsequent years.

Prof. Campanella selected the streets under investigation as representative of streets in the Orleans parish after Hurricane Katrina. Magazine Street is a prosperous commercial corridor. This street lies above the Mean Sea Level and, therefore, did not flood after Hurricane Katrina. Carrollton Avenue is a middle-class avenue that was flooded deeply in many areas and less so in others. With a lower socioeconomic group of residents, St. Claude Avenue was consistently lightly flooded. All 16 miles of the three corridors were surveyed by bicycle weekly starting on October 9, 2005, six weeks after Hurricane Katrina and about two weeks after un-flooded neighborhoods began to repopulate. Businesses of all types that were visible from the street were recorded by (1) address, (2) name, (3) description and category, (4) ownership, (5) general economic status, and (6) size. Finally, and most importantly, the business' status as "still closed," "open,""partially open" (limited hours, by appointment only, etc.), "new"(a new post-Katrina business) or "moved" was recorded and re-recorded with each weekly visit. The weekly pace of surveys was reduced to biweekly in autumn 2006, because the number of reopenings or new businesses did not warrant a weekly revisit. By 2007-2008, conditions had stabilized to the point that only seasonal/annual visits were made (Lam, et al. 2009). The temporal and attribute depth of the original dataset was reduced for the purposes of this exploration.

Furthermore, Dr. Nina Lam, Professor of Geography, and Dr. R. Kelley Pace, Professor of Real Estate, from Louisiana State University add value to the original survey data with follow-up phone calls and local inquiries, integration with external data from the Census Bureau, FEMA, State of Louisiana, and Army Corps of Engineers. The combined street survey and GIS data can thus provide additional information such as block number and street address for individual firms, 1999 median household income of the neighboring census tracts, topographic elevation of the business site, maximum flood depth during Katrina.

3.5 Variables

The specific dependent variable and independent variables in the econometric analysis are listed as follows. Furthermore, the locations of firms will be used to construct spatial weights matrices, and serve as a control variable.

3.5.1 Dependent variable

Opening statuses. Models based on three different dependent variable vectors were estimated, each having the same explanatory variables. One dependent variable vector was constructed to represent the very short period of 0-3 months after Katrina, another reflecting the 0-6 month horizon and a final model for the 0-12 month time period. For each period, firms that had reopened were assigned a dependent variable value of 1 and those not reopened a value of 0. During the 0-3 month time horizon we had 300 of the 673 firms opened, during the 0-6 month period 425 firms were open and in the 0-12 month interval 478 firms. The latter number reflects a re-opening rate of 71% at the 12-month horizon which matches well to the 66% re-opening rate derived from larger samples taken of firms that re-opened.

3.5.2 Independent variable

Median household income. I use logged 1999 median household income of neighboring census tracts to model the neighborhood's socioeconomic statuses. I hypothesize that a decline in income of the neighboring census tracts would correspond to a decrease in business reopening rates.

Flood depth. The maximum flood depth of individual businesses' location was measured in the aftermath of Hurricane Katrina. Zero indicates a site without flood effects, while smaller negative numbers indicate greater flood damage. I hypothesize that flood depth exerts a negative influence on a business' choice of reopening; in other words, stores with deeper maximum flood depth would have a smaller chance of returning.

Ownership. Business can be classified according to its ownership: as belonging to a national chain (labeled as "1"), belonging to a regional chain (labeled as “2”), or locally-owned store (labeled as “3”). Accordingly, I hypothesize that locally-owned, independent businesses reopened faster, compared to regional chains and national chains. These three different categories of ownerships are modeled as two dummy variables for store ownership type in the econometric model, one reflecting sole proprietorships and the other representing national chains with regional chains representing the excluded class.

Economic status. Businesses are classified according to their clientele types: serving functional purpose, serving middle-range customers, and providing high-end and luxury products. I hypothesize that businesses serving middle- to high-end clientele would reopen in the largest numbers. In order to enter economic status into econometric analysis, two dummy variables are used to indicate low and high socioeconomic class of the store clientele with the middle socioeconomic class excluded.

Category. Business codes represent different business categories. The North American Industry Classification System (NAICS) classify individual firms into different business types. Main business types in the study area are hotel, professional service, restaurant, wholesale, and retail.

Size. The survey data contain a categorical variable representing the scale of the business. This variable takes a value “1” for sole proprietorship with five or fewer employees, e.g., a typical restaurant, ”2” for a firm with 6-15 employees, e.g., a wholesale store, and ”3” for a business with scores of employees, e.g., a hotel. I hypothesize that a business with more employees would have a greater chance of returning. In the econometric modeling, two dummy variables will be used to reflect small and large size firms, with medium size firms representing the omitted class.

3.5.3 Controlling variable

Differing from non-spatial models, spatial models explicitly incorporate spatial processes in their specifications. I will use businesses' geographic locations to construct a spatial weight matrix to account for the spatial dependence. This spatial weight matrix specifies the strength of interactions between observations at different locations, transforms a non-spatial model into a spatial model, and impacts the overall model fitting between dependent and independent variables. I will use pair-wise discrete distances between businesses based on k-nearest neighbors criteria to build the spatial weight matrix.

3.6 Operational definitions

Business. I will confine "businesses" in my discussion to the 673 commercial or industrial enterprisers visible from the St. Claude Avenue, Magazine Street, and Carrollton Avenue in New Orleans. I will use "business" interchangeably with "firm," "store," and "retails."

Spatial interdependence. "Spatial interdependence" in my research refers to the situation that one firm's choice regarding returning to its prior location in the aftermath of Hurricane Katrina can exert influence on its neighboring firms, and the firm is in turn affected by the choice of its neighbors. The reason for spatial interdependence is that one firm's decision can influence, if not determine, the decisions of neighboring firms through various spatial externalities, and vice versa.

Spatial probit model. A probit model is an econometric model that explains the relationship between a binary dependent variable and continuous independent variables. Spatial interdependence among observations of dependent variables violates the Gauss-Markov assumption underlying nonspatial probit models. Consequently, researchers develop probit models with explicit consideration of this spatial interdependence and label this group of probit models a "spatial probit model".

Bayesian Markov Chain Monte Carlo (MCMC) method. Bayesian methods involve combining the data distribution embodied in the likelihood function with prior distribution for the parameters assigned by the practitioner, to produce a posterior distribution for parameters. Bayesian inference for model parameters can be summarized as:

$$prob(\theta|D) \propto prob(D|\theta) \times prob(\theta) \tag{3.1}$$

Where $prob$ stands for probability density, θ represents model parameters, and D denotes model data, including dependent and independent variables. $prob(\theta|D)$, the posterior distribution of model parameters, is of fundamental interest in econometrics to learn about model parameters using data. $prob(\theta)$, the prior distribution, contains any non-data information available about model parameters, and $prob(D|\theta)$, the likelihood function, is often referred as the data generating process, which specifies the density of the data conditional on model parameters. By combining the prior distribution and the likelihood function, the posterior distribution contains both sample data information as well as subjective information before looking at the data (Koop 2003, 1-11).

The basic idea of Monte Carlo simulation methods is straightforward: if one can characterize the joint distribution of the quantities of interest, then one can simply sample from that distribution and calculate the desired statistics in those samples. With a sufficiently large number of samples, the sample statistics will converge and approximate the corresponding population parameters. The consecutive draws are independent and the target distribution is specified directly in typical Monte Carlo simulation. In MCMC, each draw is dependent on the previous one in a manner that generates samples with properties mirroring those of the joint population, using just the conditional distribution of each parameter. This procedure is suitable for my case, because the joint distribution of parameters for the spatial probit model is not expressible directly, and the complexity makes direct sampling from the joint distribution prohibitively difficult (LeSage and Pace 2009, 123-52; Franzese and Hays 2008).

3.7 Data processing and analysis

Data analysis consists of three parts, and begins with data pre-processing: Because information regarding reopening statuses in the survey data is in text form, I will quantify this information before conducting statistical analysis. There are seven different re-opening statuses in the original survey data, and these seven opening status will be re-classified into two general categories: (1) Open, including reopen and partially reopen; (2) Closed, including open in another location, still closed, permanently closed. The former category is labeled with "1" and the latter case is labeled with "0". Moreover, I will geocode individual firms' locations. The location of an individual firm is represented by a combination of street name, block number, and street address. I will transform these locations into geographic coordinates, i.e., longitude and latitude, which are essential for constructing a spatial weight matrix in the spatial probit model. Because the geographic coordinates for each street is known and each street is attributed with address ranges, I interpolate the positions of businesses' addresses within the range of address along the corresponding streets. Furthermore, I will treat locations of businesses with the same address carefully. For example, both Martin Wine Cellar and Village Shoe Repair are located at Magazine Street 3502, and they would have the same geographic coordinates after geocoding. The inverse distance between these two stores will be equal to infinity, which will lead to an erroneous spatial weight matrix. Therefore, I introduce a small amount of random errors into locations of businesses with the same address to differentiate these identical locations. In such cases, the inverse distances among stores at the same address would be large numbers instead of infinity.

The second part of data analysis focuses on the performance of model comparison tools in the context of spatial models. In this study, it is important to determine the appropriate number of neighbors to use when forming the spatial weight matrix W used in our model based on the locations of stores. As previously discussed, this involves model comparison tools, which aim at selecting the model that is "closest" to the underlying data generating process. Traditional model comparison tools are not designed for situations where the dependent variable is not normally distributed, does not exhibit constant variance, or is discrete rather than continuous, for example, spatial econometric models. Thus I will conduct a Monte Carlo study to evaluate the performance of different model comparisons tools in the context of spatial models.

The rest of the data analysis will (1) implement a spatial probit model within the Bayesian MCMC framework using data pre-processed in the first step, (2) determine an appropriate model specification based on conclusions drawn in the previous Monte Carlo study, (3) interpret the model estimates to identify the determinants of business return in New Orleans after Hurricane Katrina, and (4) summarize the direct, indirect, and total spatial interactions among firms' decisions regarding re-opening.

Chapter 4

INFORMATION CRITERION FOR COMPARING SPATIAL WEIGHT MATRICES

The choice of spatial regression model specification, and more specifically, specification of the spatial connectivity structure used, plays an important role in applied work, since changes in these choices lead to differences in estimates and inferences (Anselin and Rey 1991; Florax and Rey 1995; Smith 2008). Identifying the appropriate model specification in general and spatial connectivity structure in particular can be viewed as the selection of model that is "closest" to the underlying data generating process. In conventional econometrics, information criteria such as the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Deviance Information Criterion (DIC) are widely used in model selection, along with other approaches such as RMSE, Mallows' C_p, the coefficient of determination, and Bayesian model comparison approaches. However, few studies have evaluated the performance of various information criterion for selecting spatial weight matrices, and even fewer have evaluated these in the context of spatial econometric modeling. Even fewer studies have focused on selection of spatial models in binary or categorical cases, as it is difficult to implement model selection tools, such as the log-marginal likelihood, in situations where the dependent variable is not normally distributed, does not exhibit constant variance, or is discrete rather than continuous.

The exact log-marginal likelihood which forms the basis of formal Bayesian inference regarding model comparison has been derived for simple spatial econometric models that involve continuous dependent variables (LeSage and Parent

2007). However, the log marginal likelihood approach requires integration over all model parameters, making this approach difficult to apply in a spatial probit setting. Asymptotically the AIC, BIC, DIC and log marginal likelihood model comparison criteria should produce identical results. Therefore, the aim of this chapter is threefold: (1) to extend information criteria, such as AIC, BIC, and DIC, for application in a wider range of applied modeling situations; (2) to compare the ability of log-marginal likelihood and alternative information criteria to identify appropriate spatial weight matrices in a continuous dependent variable spatial model setting; and (3) to assess these model comparison critera in a discrete dependent variable spatial probit setting. All of the model selection procedures described in this study were implemented in Matlab using the Spatial Econometrics Toolbox for MATLAB (LeSage 1999).

4.1 Information criteria for spatial models

Information criteria are rooted in the concept of entropy, and involve measures of prediction accuracy and complexity of alternative statistical models. The model that minimizes the information criterion is deemed the best model or the one most consistent with the underlying data generating process among all models under investigation. Three most widely used information criteria are considered in this study: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Deviance Information Criterion (DIC). The definition of these information criterion and their formulation is as follows.

4.1.1 Akaike Information Criterion (AIC)

Akaike Information Criterion (AIC) (Akaike, 1973) is one of the most applied information criteria, calculated using:

$$AIC = -2\log[L(\hat{\theta})] + p \tag{4.1}$$

Where $\log L(\hat{\theta})$ represents the log-likelihood function of the maximum likelihood estimator (MLE) based on the observed data, and p is the size of the parameter vector θ. In the Bayesian setting, the $\log L(\hat{\theta})$ is evaluated using the posterior mean value for θ, and viewed as a measure of model fit. The second term, p, is a "penalty function" so complex models with many parameters will be penalized resulting in a larger information criterion. This is needed since more complex models will frequently result in better model fit by over-fitting the sample data. Competing models are ranked according to their AIC values, with the one having the lowest value deemed as best model.

AIC is well known for its inability to work with small samples, and Hurvich and Tsai (1989) proposed a consistent AIC (labeled as AIC$_c$ here) for work with small samples. AIC$_c$ takes both sample size and number of model parameters into consideration, and is formally defined as:

$$\text{AIC}_c = AIC + \frac{2p(p+1)}{n-p-1} \tag{4.2}$$

4.1.2 Bayesian Information Criterion (BIC)

Bayesian Information Criterion (BIC) (also called the Schwarz Information Criterion (Schwarz, 1978)) is similar to AIC. It is usually viewed as an approximation to standard Bayesian model selection tool: the Bayes Factor (BF). BIC is defined as:

$$BIC = -2\log[L(\hat{\theta})] + p\log(n) \tag{4.3}$$

Where $\log L(\hat{\theta})$ represents the log-likelihood function of the Bayesian (or maximum likelihood) estimator based on the observed data, p is the size the parameter vector θ, and n is the sample size. The only difference between AIC and BIC lies in the degree to which they penalize models with large numbers of parameters relative to the sample size. As usual, the model with minimum BIC value is deemed as the "best" model. We use the Bayesian posterior mean value of θ in place of the MLE estimates to calculate $\log L(\hat{\theta})$.

4.1.3 AIC and BIC for spatial probit model

As discussed in LeSage and Pace (2009), the spatial autoregressive probit model can be summarized as:

$$y_i = 1, \text{if } y_i^* \geq 0 \tag{4.4}$$

$$y_i = 0, \text{if } y_i^* \leq 0 \tag{4.5}$$

$$y_i^* = \rho W y_i^* + X\beta + \epsilon \tag{4.6}$$

Where the continuous latent variable y^* links the binary observation y and underlying spatial process $f(X_i\beta)$. Because that there is spatial dependence in these latent y^* by definition, they are treated as additional set of parameters in the estimation (LeSage and Pace 2009). This means that it is difficult to calculate the effective number of parameter in a spatial probit model setting. For example, if there is perfect spatial dependence (the spatial autocorrelation coefficient is unity), we only need one observation y_i^* to determine all other values in the vector y^*, so the effective number of observations in y^* is one. In contrast, if there is no dependence (a zero spatial autocorrelation coefficient), we will have the conventional sample size of n. This same issue arises with the AIC and BIC criterion making them unsuitable for models where the dependent variable exhibits spatial dependence (e.g., our spatial probit model

(Ward, 2008)). Based on this reasoning, I propose a modification to AIC and BIC for use with spatial probit models, which takes spatial dependence in the vector y^* into consideration. The modified AIC and BIC are termed as AIC_s and BIC_s respectively.

$$\text{AIC}_s = -2\log[L(\hat{\theta})] + (n-1) \times (1-\hat{\rho}) + 1 + p \tag{4.7}$$

$$\text{BIC}_s = -2\log[L(\hat{\theta})] + ((n-1) \times (1-\hat{\rho}) + 1 + p)\log(n) \tag{4.8}$$

where $-2\log[L(\hat{\theta})]$ remains the log-likelihood (evaluated using MLE estimates), p^* is the number of parameters not including the latent dependent variable vector, and $\hat{\rho}$ is the estimated spatial dependence parameter. The effective number of parameters in $y*$ is computed as $(n-1) \times \rho + 1$, which equals the number of observations n when we have no spatial dependence ($\rho = 0$) and 1 when y^* exhibit perfect dependence ($\rho = 1$).

4.1.4 Deviance Information Criterion (DIC)

Deviance Information Criterion (DIC) is developed as a Bayesian equivalent of AIC for models estimated using MCMC draws (Spiegelhalter et al. 2002). The DIC also involves a measure of model fit and a measure of complexity (penalty function).

$$DIC = \overline{D(\theta)} + 2p_D \tag{4.9}$$

$$D(\theta) = -2\log[L(\theta)] + C \tag{4.10}$$

$$p_D = \overline{D(\theta)} - D(\bar{\theta}) \tag{4.11}$$

where $-2\log[L(\hat{\theta}]$ is the log-likelihood function, C is a constant that cancels out in model comparison, and $D(\theta)$ refers to the deviance function. The expected value for the deviance $\overline{D(\theta)}$ is evaluated using sampled MCMC draws for (θ), while $D(\bar{\theta})$ is the deviance function evaluated at the posterior mean, mode, or median of the sample of MCMC draws for θ. DIC is easy to compute in a Bayesian Markov Chain Monte Carlo

(MCMC) framework, as $\overline{D(\theta)}$ can be evaluated using each sample draw of the vector of parameters θ during the MCMC sampling, and $D(\bar{\theta})$ can be calculated using the posterior mean of the parameters θ. In the same spirit as AIC and BIC, the minimum DIC values indicates that the corresponding model is most consistent with the data.

4.2 Experiment design

To evaluate the performance of these different information criterion in a spatial econometric model setting with a focus on how these criterion work to identify specifications based on different spatial weight matrices, two Monte Carlo simulation experiments were carried out for models involving both continuous and binary dependent variables.

4.2.1 Monte Carlo simulations with continuous dependent variable

The Monte Carlo experiment for the case of a continuous dependent variable involved the following. First, a spatial connectivity pattern based on one of three real-world datasets (having different sample sizes) was used to construct a spatial weight matrix based on six nearest neighbors (which we label W_6) that was used to generate a dependent variable vector y using a Spatial Durbin model specification. The experiment considered eight alternative spatial weight matrices based on incorrect numbers of nearest neighbors as well as the true six nearest neighbors weight matrix W_6. The experiment allowed for variation in the signal-to-noise ratio by using independent variables X with different variances relative to the disturbance term variance. The correct model based on W_6 was estimated using the correct Spatial Durbin model specification via Bayesian Markov Chain Monte Carlo (MCMC) along with the nine candidate weight matrices. The different model selection criterion were calculated in order to identify the "best" model weight matrix. This process was replicated 100 times for each combination of design factors (i.e., sample size, variances of the signal in X versus the noise in the error term, spatial dependence parameters, and weight matrices). For reporting purposes, the probability of recovering the true weight matrix W_6 was obtained for different scenarios by averaging over the 100 replications.

In the following discussion each of these aspects of the Monte Carlo experimental setup are discussed in more detail.

4.2.1.1 *Alternative weight matrices*

At the beginning of each replication, I generate spatial weight matrices based on real-world datasets. There are design considerations regarding construction of spatial weight matrices. One issue is whether a simulated regular lattice should be used or a real-world irregular tessellation. Researchers often use a simulated regular lattice, while applied work typically involves an irregular tessellation based on the spatial data sample being analyzed (Florax and Rey 1995; Farber et al. 2009). Although simulated spatial patterns have been prevalent in experimental designs used in spatial statistics, a recent study suggests that the randomness in simulated datasets may cause underestimation of the ability of information criterion to distinguish between different spatial weight matrices (Stakhovych and Bijmolt 2008). Therefore, I adopted three irregular tesellations based on actual spatial datasets in my experiments. These datasets contained latitude and longitude coordinates for: 49 census tracts from the Anselin's Columbus, Ohio dataset (Anselin 1988), 98 census tracts from the Toledo, Ohio dataset (LeSage 1999), and 506 Boston census tract observations from Harrison and Rubinfeld (1978). These dataset have been used in other econometric analyses (Belsley, Kuh and Welsch 1980; Gilley and Pace 1996), and will be used here to represent small, medium and large samples sizes respectively. The latitude and longitude coordinates were used to construct weight matrices based on varying numbers of nearest neighbors weight matrices. These were constructed using the latitude-longitude coordinates to calculate distances of each observation to all other observations from which nearest neighboring regions can be found.

Hence I have three different spatial patterns and sample sizes for use in the Monte Carlo experiments. There exist numerous ways of constructing spatial weight matrices in the spatial econometric literature, which range from simple contiguity of borders, inverse distances, k-nearest neighbors, to more complicated methods that combine measures of distance and contiguity. For model comparison purposes consideration of varying numbers of nearest neighbors weight matrices seems the most straightforward. Using each set of latitude-longitude coordinates, nine spatial weight matrices based on one to nine nearest neighbors was constructed, with the six nearest neighbor weight matrix (W_6) used to generate the experimental dataset. We note that the commonly used contiguity-based weight matrix is typically equivalent to a six nearest-neighbor weight matrix when dealing with a regular lattice configuration (LeSage and Pace 2009). Inverse-distance based weight matrices are typically dense, which gives rise to heavier computational burdens. The generated spatial weight matrices were row-standardized to have row sums of unity.

4.2.1.2 *Alternative signal-to-noise ratios*

The second step involved generating random explanatory variables and error terms based on differing variances to control the signal versus noise in the model. Two explanatory variable vectors X_1, X_2 were drawn from a standard normal distribution $N(0, 1)$, and another two spatially lagged explanatory variable vectors were constructed using W_6X_1, W_6X_2, where W_6 is the six nearest neighbor weight matrix. The spatial lags of the explanatory variables are needed for the spatial Durbin model specification. The coefficients $\beta_1 = \begin{pmatrix} \beta_{1,1} & \beta_{1,2} \end{pmatrix}'$ for $X = \begin{pmatrix} X_1 & X_2 \end{pmatrix}$ and $\beta_2 = \begin{pmatrix} \beta_{2,1} & \beta_{2,2} \end{pmatrix}'$ for W_6X in the Spatial Durbin Model were both set to [1, -1]. In addition, an intercept vector with associated parameter of zero was used. In assessing the information-based model selection, the signal-noise ratio in the data becomes important. The signal-noise ratio can be varied using different ratios between the variance of the explanatory variables and the error term (for a given sample size). The error term ε is drawn from a normal distribution $N(0, \xi)$, with variances ξ set to 0.25, 0.5, 1, 2, and 5.0 to produce varying signal-to-noise ratios. Smaller noise variances reflect larger signal-to-noise ratios which should produce more accurate inferences regarding the correct model.

4.2.1.3 *Alternative levels of spatial dependence*

The dependent variable vector y is constructed using the spatial Durbin model (SDM) specification: $y = \rho W_6 y + X\beta_1 + WX\beta_2 + \varepsilon$ in conjunction with the six nearest neighbors spatial weight matrix W_6. This is done for each of the varying signal-to-noise ratios described in the previous section. To control the degree of spatial dependence in the experiment, the dependent variable is generated using nine different values for ρ ranging from 0.1 to 0.9 in increments of 0.1. Low levels of spatial dependence are reflected by values near 0.1 and high levels of spatial dependence are associated with values near the 0.9 value. As noted previously, the Spatial Durbin Model can be viewed as a linear combination of the Spatial Autoregressive Model and Spatial Error Model. Thus I use only the Spatial Durbin Model specification in the Monte Carlo experiment, ignoring other types of spatial regression specifications developed to account for spatial dependence.

4.2.1.4 *Estimation method*

The spatial Durbin model based on the various generated values for the vector y and matrices X and WX were estimated using a Bayesian Markov Chain Monte Carlo (MCMC) approach (LeSage 1997). The model estimation process is carried out for each of the nine alternative spatial weight matrices, including the "true" weight matrix used to generate the experiment data. The model estimates were based on 1,250 draws with the first 250 omitted for the burn-in process required for MCMC sampling. Estimation relied on exact computation of the log-determinant term $\log|I_n - \rho W|$. The information criteria discussed previously were calculated using posterior estimates, and the performance of these information criteria were judged according to their ability to discriminate between the true six nearest neighbor spatial weight matrix and the alternative matrices for varying levels of spatial dependence, signal-to-noise ratios, and sample sizes. The log-marginal likelihood criterion typically used for Bayesian model comparison purposes was also calculated for each experimental dataset as a benchmark against which to judge performance of the various information criterion.

For each combination of experiment factors (5 levels of signal-noise ratio, 9 different spatial dependence levels, and 3 different sample sizes), the entire data generation and model estimation process was repeated 100 times using the seven different weight matrices. The experiment thus consisted of $5 \times 9 \times 3 \times 100 \times 9 = 121,500$ simulations. Summarizing across these 121,500 simulations, I evaluated the performance of different model selection criteria in terms of the probability of finding the true weight matrix in different model settings.

4.2.2 Monte Carlo experiments for the case of a binary dependent variable

Aiming at assessing the performance of various model selection criterion in the discrete spatial probit model setting, I performed another set of Monte Carlo experiments based on a generated binary dependent variable. These Monte Carlo experiments mirror those for the continuous dependent variable case, but were simplified due to the heavy computational burden of estimating spatial probit models. The simplifications involved the following. Spatial weight matrices were based only on the 506 Boston census tract sample with only five alternative weight matrices constructed using five to nine nearest neighbors, with the six nearest neighbor weight matrix (W_6) the "true" weight matrix used in the data generating process. The independent variables X were generated in the same manner as in the continuous model. Only three different levels of signal-to-noise ratio were employed based on setting the variance of the error term to 0.25, 1, and 5.0. As in the continuous dependent variable simulations, a continuous dependent variable y was generated using W_6 and different spatial dependence levels and signal-to-noise ratios based on the spatial Durbin model specification. The simplification involved only three different ρ values: 0.3, 0.6, and 0.9, which correspond to weak, medium, and strong spatial dependence respectively.

To maintain comparability of the information criterion performance comparison with results found for the continuous dependent variable and the spatial probit models, the continuous dependent variable y was truncated to $(0,1)$ values when the continuous y value was negative or positive, resulting in a binary dependent variable that we label y_b.

Since the settings employed for the parameters, signal and noise variance and so on were devised to ensure a normal distribution of the continuous dependent variable, that is centered on zero, truncation of the continuous dependent variable y based on negative and positive values to the (0,1) values to produce a binary dependent variable y_b is consistent with the way we typically think of an idealized probit model data generating process.

This truncation of a continuous dependent variable from the non-binary dependent variable experimental setting helps assess how information loss impacts the information criterion model selection performance in a ceteris paribus setting.

4.2.3 Implementation of the binary dependent variable Monte Carlo experiments

As noted, the spatial Durbin model for both continues and binary dependent variables were estimated using Bayesian MCMC based on the five weight matrices. For the case of the binary dependent variable based on the large sample of 506 observations, the estimation of the spatial probit model used a one-step Gibbs sampler to construct the multivariate draws for the latent utility parameters (see LeSage and Pace 2009, chapter 10). Other estimation options such as number of draws and number of burn-ins remained the same as those in the case of the continuous dependent variable experiments. The experiment involved 100 replicated for each combination of design factors: 3 signal-to-noise ratios, 3 different spatial dependence levels, and 5 alternative spatial weight matrices, as in the case of the continuous model experiments. This experiment results in $3 \times 3 \times 100 \times 10 = 9,000$ replications. Finally, the probability of finding the true weight matrix for each data generated dependent variable vector was obtained for different scenarios by averaging over these 100 replications, which allows us to compare the various information criterion performance with that found for the continuous models. The calculation of AIC, BIC, DIC, and log-marginal likelihood requires the exact closed-form log-likelihood of the corresponding model. It is possible to calculate this value for the case of models involving a continuous dependent variable y, but we do not have an expression for the exact log-likelihood in the case of a binary dependent variable model such as the spatial Durin probit model, since the latent variable y_b^* has a multivariate truncated normal distribution. Therefore, computation of information criterion for the spatial probit model requires an approximation. One reasonable approach is to substitute latent y_b^* draws into the log-likelihood function during calculation of the DIC.

Intuitively the y_b^* should be very close to the continuous y values if the MCMC sampling produces accurate draws reflecting these values which are treated as latent unobservable parameters in the spatial probit model. Following this reasoning, the posterior mean of the latent variable y_b^* draws were used in place of the binary values in the vector y_b to calculate AIC, BIC, and log-marginal likelihood values.

4.3 Results and discussions

It is important for researchers to know which model selection tool is most appropriate in a given modeling scenario. For example, given a certain sample size, level of spatial dependence, or type of dependent variable. The effects of varying these experimental factors in the Monte Carlo experiments on the ability of the various model selection criterion to find the true model can be summarized as follows.

4.3.1 Weight matrix identification for the case of a continuous dependent variable

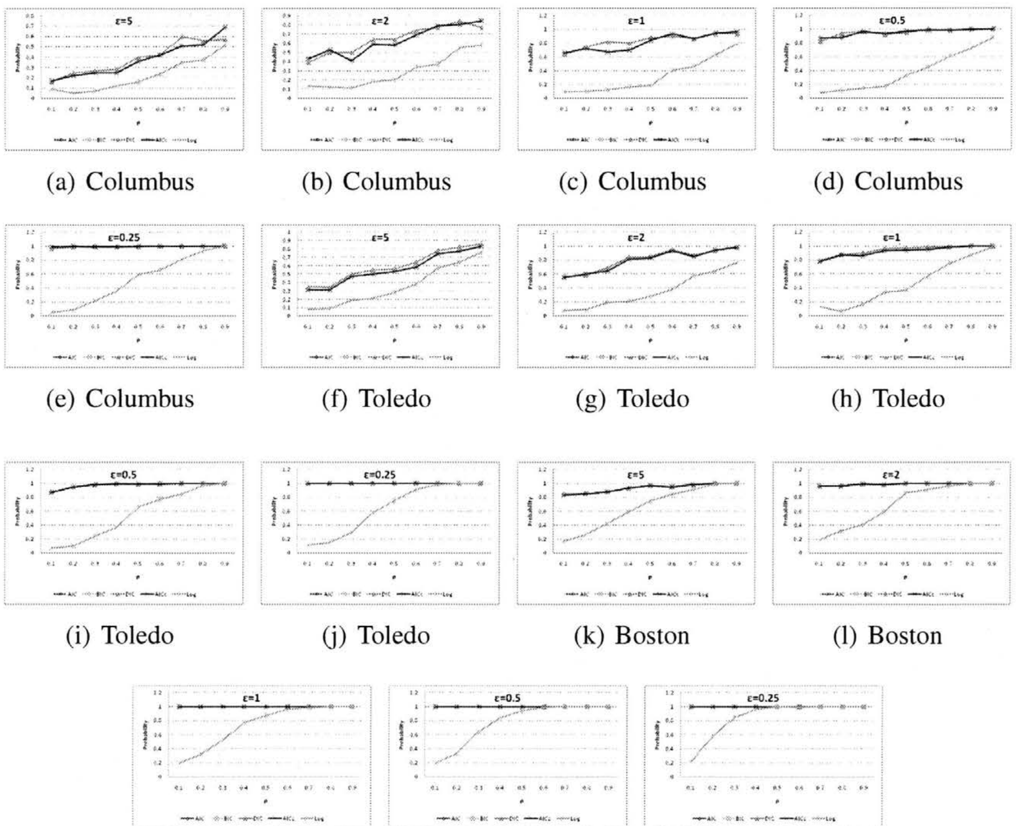
Intuitively, different model comparison tools may have different power in identifying the true model, as their performance can be influenced by other design factors such as spatial dependence, signal-to-noise ratio, and sample size. In line with Stakhovych and Bijmolt (2008), I summarize the direction and strength of variation in these model design factors rather than present detailed parameter estimates. The summary was constructed using analysis of variance (ANOVA) techniques. The main and interaction effects of these design factors on the probability of finding the true model are illustrated in Table 4.1. The ANOVA model fits the data well with an adjusted $R^2 = 0.918$. All main and two-way interaction effects are significant at the 0.01 level when the associated partial-η^2 is larger than 0.140. A plot of the probability of finding the true model specification with different model selection tools and design factors in shown in Figure 4.1. The analysis of Table 4.1 and Figure 4.1 point to the following conclusions.

Table 4.1: Effects on the probability of identifying the true model with continuous dependent variable

Source	Df	F	Sig.	η_p^2
Intercept	1	68178.977	0.000	0.992
Sample size (sample)	2	603.569	0.000	0.689
Spatial dependency (rho)	8	146.293	0.000	0.683
Variances of error term(var)	4	435.436	0.000	0.762
Model selection tool (criteria)	4	489.354	0.000	0.783
rho $\times$ criteria	32	21.712	0.000	0.561
sample $\times$ criteria	8	11.056	0.000	0.140
sige $\times$ criteria	16	5.685	0.000	0.143
sample $\times$ rho	16	6.061	0.000	0.151
rho $\times$ sige	32	4.583	0.000	0.212
sample $\times$ sige	8	62.459	0.000	0.479

Notes: $R^2 = 0.934$ (Adjusted $R^2 = 0.918$)

First, model selection tools perform differently as the factor $criteria$ in Table 4.1 is significant with $\eta_p^2 = 0.783$. However, the differences lie mainly between the performance of the log-marginal likelihood and the various information criteria, while different information criterion reveal similar capabilities of identifying the true model. The similar performance of information criterion can be ascribed to their definitions. For example, AIC and BIC perform identically in terms of the



(a) Columbus (b) Columbus (c) Columbus (d) Columbus

(e) Columbus (f) Toledo (g) Toledo (h) Toledo

(i) Toledo (j) Toledo (k) Boston (l) Boston

(m) Boston (n) Boston (o) Boston

Figure 4.1: Probability of finding the true weight matrix

probabilities of finding the true weight matrix in the continues dependent variable setting. The reason for this is that the difference between their definitions is in the "penalty function", which involves sample and parameter sizes, which do not relate to the spatial weight matrix.

Second, the level of the spatial dependence (reflected in the parameter ρ) has a substantial effect on the probability of identifying the true weight matrix, as suggested by the significance of the factor rho and associated η_p^2 of 0.683. For all models selection tools, the probability of selecting the true weight matrix among alternative matrices is higher with larger values of ρ, and the converse is also true. Although the performances of all model selection tools under investigation converge with strong spatial dependence, the information criterion outperform the log-marginal likelihood with weak spatial dependence. The log-marginal likelihood's inability to identify the true weight matrix for small values of ρ is consistent with findings reported in LeSage and Pace (2009).

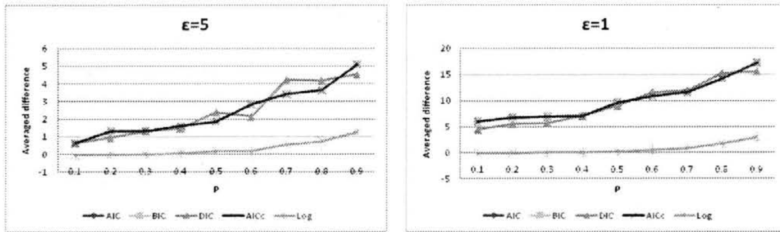
Third, the sample size and signal-to-noise ratio also significantly impact the probability of finding the true weight matrix with substantial effects sizes reflected in η_p^2 of 0.689 and 0.762. Both sample size and different signal-to-noise ratio act similarly in the experiments as we would expect. As the signal-to-noise ratio increases, i.e., with larger sample size and/or smaller variances of the error terms, the performance of all models selection tools improve. This type of impact for the signal-to-noise ratio confirms previous studies by Stakhovych and Bijmolt (2008). Even with a large signal-to-noise ratio, the log-marginal likelihood remains incapable of identifying the true weight matrix with low spatial dependence levels. In contrast, the information criterion recover the true weight matrix nearly 100% of the time when the signal-to-noise ratio is high (for example, sample size = 500 and/or a variance of the error term ε = 0.25), for all levels of spatial dependence in the experiment. In these cases, the signal-to-noise ratio is so large, i.e., so much information, that low spatial dependence no longer hinders performance of the information-based model selection tools.

Researchers may be interested in the probability of recovering the true model as well as the discriminating power of the various model selection tools. That is, how substantial is the difference between the "true" model" and competing models in terms of AIC, AIC_c, BIC, DIC and log-marginal likelihood values? For example, a DIC difference of more than 7-10 units is regarded as "strong evidence" in favor of the model with the smaller DIC, while a different of 1.5 between log-marginal likelihoods provides only "substantial support" for model with a smaller value. A plot of the minimum differences between the true model and alternative models is shown for the various model selection criterion as we vary spatial dependence levels, signal-to-noise ratios, and sample size in (Figure 4.2). This figure illustrates that the discriminating power of all model selection tools increases with larger sample size and signal-to-noise ratios as we would expect. It also points out that stronger spatial dependence is associated with larger differences in AIC, BIC, DIC and log-marginal likelihood values.

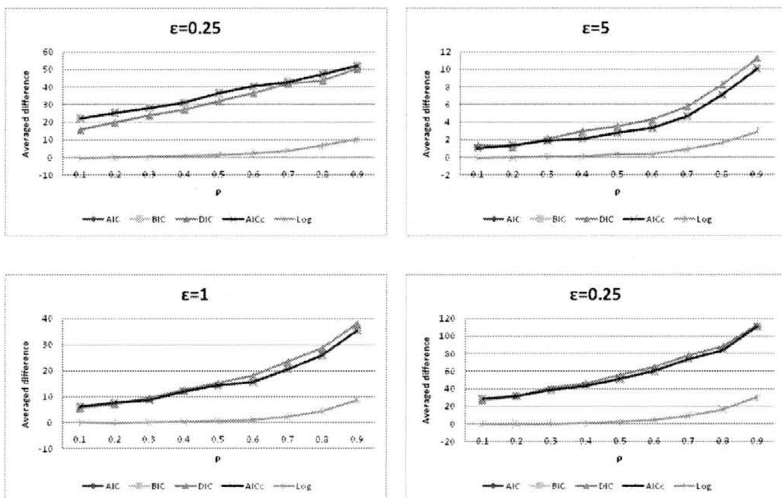
4.3.2 Weight matrix identification for the case of a binary dependent variable

The spatial dependence, signal-to-noise ratio, and sample size are also influential factors for performance of the various model selection tools in the case of the spatial probit model. For example, with a binary dependent variable, the main and interaction effects of the various design factors on the probability of finding the true model are shown in Table 4.2 using the ANOVA model, which fits the data with an adjusted R^2 of 0.984, and all main and two-way interaction effects are significant at the 0.01 level except the factor var which is significant at the 0.05 level. These effects are also substantial with associated partial-η^2 larger than 0.140. In order to compare performance of the information criteria in the case of continuous and probit models, a plot of the probability of finding the true model specification in both cases for different model selection tools and experimental design factors is shown in Figure 4.3.

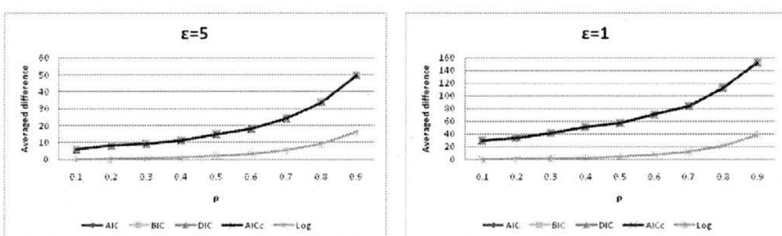
Model selection tools have a distinctly different level of performance for the case of a binary dependent variable, as indicated by the fact that the factor $criteria$ is significant with η_p^2 close to 1 at 0.993. Again, AIC and BIC perform identically in terms of the probabilities of finding the true weight matrix, consistent with the fact that the difference in their definitions does not include the weight matrix. The AIC,



(b) Columbus



(e) Toledo



(h) Boston

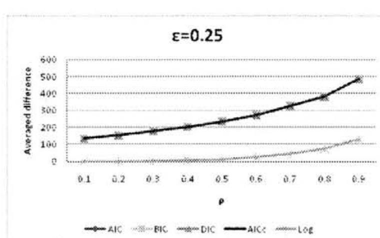


Figure 4.2: Model selection tools' discriminating power

Table 4.2: Effects on the probability of identifying the true model with binary dependent variable

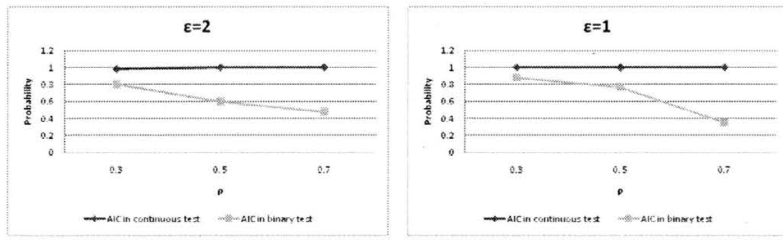
Source	Df	F	Sig.	η_p^2
Intercept	1	3293.250	0.000	0.993
Spatial dependency (rho)	2	13.136	0.000	0.523
Variances (var)	2	3.598	0.043	0.231
Model selection tool (criteria)	6	591.924	0.000	0.993
rho $\times$ criteria	12	16.331	0.000	0.891
var $\times$ criteria	12	10.700	0.000	0.843
rho $\times$ var	4	5.468	0.000	0.477

Notes: R^2 = 0.994 (Adjusted R^2 = 0.984)

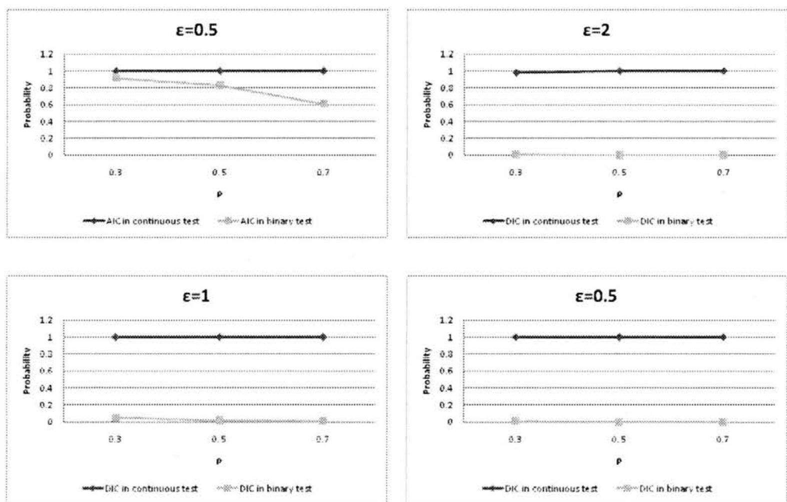
BIC and log-marginal likelihood have better performance than DIC. This can be ascribed to the fact that the model estimation procedure produced very accurate draws for the latent variable y_b^* (the correlation coefficient between y_b^* and y_b is around 0.9) and the posterior mean of y_b^* was used in place of the binary values in the vector y_b for calculating the AIC, BIC, and log-marginal likelihood.

As suggested by Figure 4.3, AIC, BIC, and DIC perform worse in the spatial probit model than in the case of a continuous dependent variable, while the performance of the log-marginal likelihood is similar in both cases. Even within the group of information criteria, AIC and BIC perform significantly better than DIC, as the latter select the true weight matrix with very low probability in the spatial probit model setting. The degradation of performance in the information criterion is consistent with the notion that truncation of a continuous dependent variable should lead to a significant loss of information.

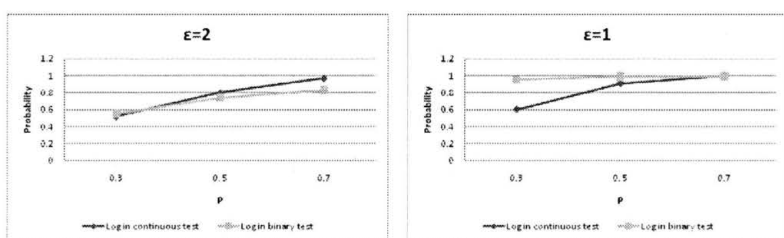
A second finding is that the level of the spatial dependence (ρ) has a significant and substantial effect (η_p^2 = 0.523) on the probability of identifying the true weight matrix. As for AIC, the probability of being correct decreases when the spatial



(b) AIC



(e) DIC



(h) Log-marginal

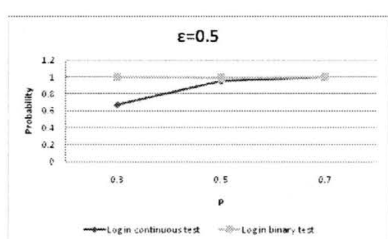


Figure 4.3: Comparison of performance in the continuous and binary settings

dependence increases. Correlation, such as spatial dependence in this case, will cause a decrease in effective sample size, and become problematic when calculating statistics in the presence of missing data values. One can view this as the case, since we use latent variable parameter estimates arising during the estimation procedure to construct the information criterion measures for the spatial probit model. In addition, the probability of selecting the correct weight matrix using DIC remains very low under all conditions. In both binary and continuous cases, the performance of the log-marginal likelihood improves for larger ρ values (more spatial dependence).

A third result pertains to the impact of the signal-to-noise ratio. The variance of error term var exerts a significant and substantial impact on the probability of finding the true weight matrix. However, the size of this effect is smaller than in the previous analysis involving the continuous dependent variable. This is because a large data sample containing 506 observations was used for this probit model test. A larger sample of course leads to a higher signal-noise ratio, so changing the variance of the error term in this case is less influential.

Table 4.3: Model selection tool performances averaging out all experiment factors

Weight matrices	W_4	W_5	W_6	W_7	W_8
AIC	0.0000	0.0078	0.6933	0.2000	0.0989
BIC	0.0000	0.0078	0.6933	0.2000	0.0989
DIC	0.0389	0.0045	0.0111	0.0977	0.8478
Log-marginal	0.0167	0.0233	0.8933	0.0445	0.0222

Considering the overall probability of finding the correct weight matrix across all experimental factors for the model selection tools in the case of the probit test provides some interesting results (Table 4.3). First, as previously discussed, the log-marginal likelihood performed best as it identified the true weight matrix nearly

90% of the time. AIC and BIC are second best, with a 0.69 probability of recovering the true model. The worst criterion is the DIC which selected the true weight matrix with a probability of only 1%. Second, all model selection tools tended to select a weight matrix with more nearest neighbors than the true number. Even though the log-marginal likelihood chose a weight matrix with six nearest neighbors (the true model in the experiments) 90% of the time, it selected weight matrices with seven and eight nearest neighbors more frequently (these having a 0.06 probability) than weight matrices with four and five nearest neighbors (where we see a 0.04 probability). The performance of AIC and BIC are slightly biased upwards as they identify weight matrices with seven or more nearest neighbors around 30% of the time. DIC is severly biased as it "hits the boundary" of the experimental range of weight matrices and suggests a weight matrix with eight nearest neighbors to be the true weight matrix with a probability of 0.85.

I do not include AIC and BIC for the spatial probit model (AIC_s and BIC_s) in the following analysis as they consistently produced poor model selection results. As simple modifications in "penalty function" do not improve the information criterion performance, other approaches needs to be derived with a more solid statistical basis. A possible solution would be the incorporation of the log-determinant term $\log|I_n - \rho W|$. Raftery (1996) suggested that the calculation for BIC should include the expected information matrix, which in our case involves the log-determinant term.

4.4 Summary and implications

This Monte Carlo study extended the AIC, BIC, and DIC information criterion for application to spatial econometric modeling situations where the dependent variable exhibits spatial dependence. Several experiments were used to provide insights regarding performance of these different model selection tools for the task of selecting the appropriate spatial weight matrix. Performance of course varied depending on a host of factors used in the experimental design such as: continuous versus discrete dependent variables, the level of spatial dependence, and signal-to-noise ratio. The results indicate that the information criterion and log-marginal likelihood do not perform uniformly under different conditions.

As for selection of the correct spatial weight matrix in the case of a model involving a continuous dependent variable, all model selection tools were capable of recovering the model based on the true weight matrix and the power to discriminate between models increased with larger sample sizes, increased signal-to-noise ratios, and stronger spatial dependence. The AIC, BIC and DIC information criteria produced similar model results in the continuous model setting. Although performance of all model selection tools under investigation converged in the presence of high levels of spatial dependence, the information criteria outperformed the log-marginal likelihood in cases involving weak spatial dependence. Therefore, information-based model selection criteria seem suitable for selecting the appropriate spatial weight matrix in spatial econometric models involving a continuous dependent variable. In contrast, for the case of the spatial probit model where the dependent variable was discrete, the log-marginal likelihood would be the best method for choice of the appropriate spatial weight matrix. The log-marginal likelihood performed similarly in both continuous and discrete dependent variable models. Performance of the information-based model selection criterion was worse in the discrete/spatial probit model setting. Moreover, DIC consistently produced the worst weight matrix selection results. Another finding was that all model selections tools were biased upwards in terms of selecting weight matrices with a larger number of nearest neighbors than used to generate the experimental dependent variable vector. The DIC had the largest upward bias with

AIC/BIC second. Another finding was that adjustments to the degrees of freedom used in the AIC and BIC criterion to take into account spatial dependence that arises in spatial models did not produce fruitful results.

Therefore, I will use the log-marginal likelihood to determine the appropriate number of neighbors to use when forming the spatial weight matrix W used in analyzing the business re-opening decisions where the dependent variable was binary.

Chapter 5

MODEL SPECIFICATIONS

Another Monte Carlo study which is similar to those in last chapter is conducted to compare the SAR versus spatial Durbin (SDM) model specifications over varying numbers of neighbors. Some initial results suggest that the SAR model specification to be more consistent with the sample data. Therefore, I will start elaborating on the spatial autoregressive probit model, then analyze the dependent and independent variables with a SAR probit model, and finally interpret the model estimates.

5.1 A spatial autoregressive probit model

Let the $n \times 1$ vector y be a 0,1 binary vector reflecting the closed/open status of the n firms at some point in time (say three months after Katrina). A conventional probit model would attempt to explain variation in the binary vector y using an $n \times k$ matrix of firm-specific explanatory variables X and associated $k \times 1$ vector of parameters β, under the assumption that each observed decision is independent from all others. LeSage and Pace (2009) set forth a spatial autoregressive (SAR) variant of the conventional probit model that takes the form shown in (5.1).

$$y = \rho W y + X\beta + \varepsilon, \quad \varepsilon \sim N(0, I_n) \tag{5.1}$$

The spatial lag of the dependent variable Wy involves the $n \times n$ spatial weight matrix W that contains elements consisting of either $1/m$ or 0, where m is some number of nearest-neighbors. If observation/firm j represents one of the $m-$nearest neighboring establishments to firm i, then the i, jth element of W contains the value $1/m$. All elements in the ith row of the matrix W that are not associated with neighboring observations take values of 0. By construction, W is row-stochastic (non-negative and each row sums to 1). This results in the $n \times 1$ vector Wy consisting of an average of the m neighboring firms closed/open status, creating a mechanism for modeling interdependence in firm decisions to reopen in the aftermath of a disaster. The scalar parameter ρ measures the strength of dependence, with a value of zero indicating independence. It should be clear that when $\rho = 0$, we have a conventional non-spatial probit model.

As previously mentioned, the common rationale behind the developed approaches considers treating the observed binary outcomes as indicators or reflections that relate to underlying unobserved and continuous latent variable (Franzese and Hays 2008). The Bayesian estimation approach to these models is to replace the unobserved latent profit with *parameters* that are estimated. For the case of a SAR probit model, given estimates of the $n \times 1$ vector of missing or unobserved (parameter) values that we denote as y^*, we can proceed to estimate the remaining model parameters β, ρ by sampling from the same conditional distributions that are used in the continuous dependent variable Bayesian SAR models (see Chapter 10 in LeSage and Pace, 2009). More formally, the unobserved profits in this study can be modeled as: $(\pi_{1i} - \pi_{0i}), i = 1, \ldots, n$, where π_{1i} represents profits (of firm i) in the open state and π_{0i} in the closed state. The probit model assumes this difference, $y_i^* = \pi_{1i} - \pi_{0i}$, follows a normal distribution. Because we do not observe y_i^*, only the choices made, i.e., re-opening decisions, which are reflected in:

$$
\begin{aligned}
y_i &= 1, \quad \text{if } \; y_i^* \geq 0 \\
y_i &= 0, \quad \text{if } \; y_i^* < 0
\end{aligned}
$$

If the vector of latent profits y^* were known, we would also know y, which led Álbert and Chib (1993) to conclude: $p(\beta, \sigma_\varepsilon^2 | y^*) = p(\beta, \sigma_\varepsilon^2 | y^*, y)$. The insight here is that if we view y^* as an additional set of parameters to be estimated, then the (joint) conditional posterior distribution for the model parameters $\beta, \sigma_\varepsilon^2$ (conditioning on both y^*, y) takes the same form as a Bayesian regression problem involving a continuous dependent variable rather than the problem involving the discrete-valued vector y. This approach was used by LeSage and Pace (2009) to implement a Bayesian Markov Chain Monte Carlo estimation procedure for the SAR model in (5.1).

5.2 Interpreting the model estimates

Interpreting the way in which changes in the explanatory variables in the matrix X impact the probability of a firm reopening in the spatial autoregressive (SAR) probit model requires some care. The expressions in (5.2) make it clear that the probability (of a 0,1 event outcome) is a non-linear function $F()$ (the probability rule) of a function $(I_n + \rho W + \rho^2 W^2 + \ldots) X\beta$ of the explanatory variables in the model X.

$$
\begin{aligned}
y &= \rho W y + X\beta + \varepsilon \qquad (5.2)\\
y &= S(\rho) X\beta + S(\rho)\varepsilon \\
X\beta &= x_{(1)}\beta_1 + x_{(2)}\beta_2 + \ldots + x_{(k)}\beta_k \\
S(\rho) &= (I_n - \rho W)^{-1} = (I_n + \rho W + \rho^2 W^2 + \ldots) \\
Pr(S(\rho)X\beta) &= F(S(\rho)X\beta)
\end{aligned}
$$

We first consider the simpler case of a non-probit spatial autoregressive model shown in (5.3), where z denotes a continuous $n \times 1$ dependent variable vector.

$$\begin{aligned} z &= \alpha\iota_n + \rho W z + X\beta + \varepsilon \qquad (5.3) \\ \partial z/\partial x_k' &= (I_n - \rho W)^{-1} I_n \beta_k \\ &= S(\rho)\beta_k \end{aligned}$$

LeSage and Pace (2009) proposed using an average of the diagonal elements from the $n \times n$ matrix: $\partial z/\partial x_k'$ to produce a scalar summary of the *direct effects*, which are derived from the own partial derivatives: $\partial z_i/\partial x_{k,i}$.

They also use an average of the (cumulated) off-diagonal elements from the $n \times n$ matrix: $\partial z/\partial x_k'$ to produce a scalar summary of the (cumulative) *indirect effects* associated with the cross-partial derivatives: $\partial z_i/\partial x_{k,j}$. This scalar summary measure cumulates the spatial spillovers falling on neighboring establishments as well as neighbors to these neighbors, and so on.

When we allow for dependence among observations/firms, changes in the explanatory variables associated with one firm, say firm j will influence the dependent variable value of firm j as well as other firms, say i. The spatial autoregressive model collapses to an independence model when the scalar spatial dependence parameter ρ takes a value of zero. In this case, the cross-partial derivatives are all zero.

For the case of spatial dependence, the (non-zero) cross-partials represent what are commonly thought of as *spatial spillover* impacts. Changes in the value of a single observation j explanatory variable can (potentially) influence all $n - 1$ other observations. This is true for all $j = 1, \ldots, n$ explanatory variable values leading to the $n \times n$ matrix of own- and cross-partial derivatives. This motivates the need for scalar summary measures that average across the sample of observations similar in spirit to the way we interpret conventional least-squares regression estimates. An important point is that the scalar summary measure of *indirect effects* cumulates the spatial spillovers falling on all other observations, but the magnitude of impact will be greatest for nearby neighbors and decline in magnitude for higher order neighbors.

The sum of the two effects (direct and indirect) represent the (cumulative) *total effect* associated with a change in an observation for that explanatory variable. As for the more complicated case of the spatial autoregressive probit model, LeSage et al. (2010) develop scalar summary measures of spatial effects in the case of the probit model developed by incorporating the normal Cumulative Probability Function $Pr()$, and devising a matrix version of own- and cross-partial derivatives. With this newly derived effects summary for the spatial probit model, I will be able to calculate how the direct, indirect, and total effects of changes in the explanatory variables at one firm location impact the probability of the firm's reopening and its neighbors' reopening.

5.3 A simple illustration

To provide a concrete illustration, let us consider the case of seven firms located in a line along (one side of) a street. Figure 5.1 shows seven regions located from west to east along a single highway. We use a simple spatial weight matrix that identifies a single left- and right-neighbor to each observation. Let y be a 0,1 binary vector, and:

F1	F2	F3	F4 Firms	F5	F6	F7
			Street			

Figure 5.1: Seven regions along a commercial street

$$
\begin{aligned}
y &= \rho W y + X\beta + \varepsilon \\
\text{Pr}(y = 0, 1) &= F[S(\rho)(X\beta + \varepsilon)] \\
S(\rho) &= (I_n - \rho W)^{-1}
\end{aligned}
$$

We generated the vectors of flood depth and firm size. Subsequently, we calculated the probability of reopening using a model based on two explanatory variables, flood depth and firm size, using a value of the spatial dependence parameter $\rho = 0.8$ and assumed that the parameters associated with flood depth and size equaled -0.25 and $+0.5$.

$$
Pr(\text{Reopening}) = F[S(\rho = 0.8)(-0.25 \text{ flood depth} + 0.5 \text{ firm size})] \tag{5.4}
$$

We show with the resulting values shown in Table 5.1. We note that using the conventional practice of interpreting predicted probabilities less than 0.5 as implying $y = 0$ and greater than 0.5 as $y = 1$, the model perfectly predicts the pattern of observed 0,1 values.

Table 5.1: Illustration based on $n = 7$ firms

Firms	$y-$value	Pr($y = 1$)	flood depth	firm size
obs 1	0	0.0036	40	1
obs 2	0	0.0231	30	2
obs 3	0	0.1964	20	3
obs 4	1	0.5131	25	4
obs 5	1	0.8569	20	4
obs 6	1	0.9907	20	8
obs 7	1	0.9968	20	8

To illustrate the effect of changing a single observation on the probabilities we increased the flood depth at the location of firm number 3 from 20 to 60, *ceteris* *paribus*. Table 5.2 shows the original predicted probabilities $Pr(y = 1)$, the new probabilities $Pr(y' = 1)$ and the change in probabilities/predictions implied by the spatial autoregressive probit model $Pr(y = 1) - Pr(y' = 1)$.

The first point to note is that the change in flood depth at the location of firm number 3 leads to a decrease in the probability that this firm will reopen by 14.14%. However, neighboring firms number 4 and number 5 are also impacted by the higher level of flooding at firm number 3, leading to a lower probability that these firms will reopen. Specifically, the probability that the immediately neighboring firm number 4 will reopen decreases by 9.28%, and that for the neighbor to this neighboring firm number 5 by 1.9%. These represent an illustration of spatial spillover impacts that arise when we allow for interdependence of other firms decisions on firm number 3's decision to open/close.

Table 5.2: The impact of changing a single observation

Firms	$y-$value	Pr($y = 1$)	Pr($y' = 1$)	Pr(y=1)$-$Pr($y' = 1$)	flood depth
obs 1	0	0.0036	0.0098	0.0061	40
obs 2	0	0.0231	0.0241	0.0010	30
obs 3	0	0.1964	0.0550	-0.1414	60
obs 4	1	0.5131	0.4203	-0.0928	25
obs 5	1	0.8569	0.8375	-0.0194	20
obs 6	1	0.9907	0.9904	-0.0003	20
obs 7	1	0.9968	0.9968	0.0000	20

The direct effect arising from the change in observation number 3's flood depth is -0.1414, whereas the cumulative indirect effect is -0.1054, the sum of all non-zero changes, with the (cumulative) total effect being -0.2468. We also note that the change in flood depth at firm number 3 location leads the model to predict that firm number 4 would not reopen, since the probability of reopening has fallen from 0.51 to

0.42 as a result of increased flooding at the neighboring location. This directly lowers the chances that firm number 3 will reopen by 14% and indirectly influences the reopening decision of firm number 4 its neighbor (as well as firm number 5 the neighbor to its neighbor).

The effects presented in Table 5.2 represent only the impact of changing a single observation number 3, whereas the scalar summary measures based on the expressions for own- and cross-partial derivatives (described earlier) would average over changes in all observations (for each explanatory variable in the model).

Chapter 6

MODELING RESULTS

6.1 Model coefficients

As already noted, we wish to determine the appropriate number of neighbors to use when forming the spatial weight matrix W used in our model. This model comparison can be done based on log-marginal likelihoods values, which is calculated by the continuous dependent variable log-likelihood in conjunction with the posterior mean of latent draws for $y^{\star}$. Models based on varying numbers of nearest neighbors were estimated and the log-marginal likelihood values were calculated, which reflects the posterior probability of a certain model specification. The log-marginal likelihoods are converted into posterior probabilities, which represent the chance of a certain model specification being identical with the "true" model. The results point to a model with 13 nearest neighbors for the 3 and 12 month horizons and 15 neighbors at the 6 month horizon (Table 6.1) .

In addition to using log-marginal likelihood values to explore the number of nearest neighboring stores to employ when forming the spatial weight matrix, the same approach is used to compare the SAR versus spatial Durbin (SDM) model specifications over varying numbers of neighbors. The results pointed to the SAR model as more consistent with the sample data and justify the adoption of SAR probit model.

The coefficient estimates (posterior means, standard deviations and Bayesian p-levels) for the model parameters β, ρ are shown in Table 6.2 for the three different time horizons. As already noted, the coefficient estimates β from the spatial

Table 6.1: Posterior probabilities for varying neighbors and three time horizons

# Neighbors/Probsc	3 month	6 month	12 month
2	0.0007	0.0006	0.0959
3	0.0026	0.0003	0.0397
4	0.0061	0.0027	0.0229
5	0.0014	0.002	0.0185
6	0.0431	0.0017	0.0227
7	0.0022	0.0014	0.0198
8	0.0165	0.0024	0.03
9	0.0471	0.004	0.0131
10	0.0424	0.0203	0.0106
11	0.0169	0.0198	0.031
12	0.0871	0.0799	0.1452
13	0.0434	0.0136	0.1476
14	0.521*	0.0882	0.3075*
15	0.0158	0.0037	0.0234
16	0.0565	0.4388*	0.0147
17	0.0712	0.0299	0.0068
18	0.0259	0.2908	0.0505

autoregressive probit model cannot be interpreted as representing how changes in the explanatory variables affect the probability of stores reopening. One point to note is that the coefficient ρ associated with the spatial lag of the dependent variable Wy is more than four standard deviations away from zero in all three sets of results. The magnitudes ranging between 0.42 and 0.59 point to significant positive spatial dependence in firm decisions regarding open/close status so that firms located nearby exhibit similar decision outcomes regarding reopening.

The coefficient of the flood depth variable are statistically significant at 99% level for all three time horizons, which suggests a long-time impact from Hurricane Katrina. Other statistically significant coefficients (99% confidence interval) for 0-3 month horizon include those for logged median income, low socioeconomic status, and sole proprietorships. low socioeconomic status, and sole proprietorships are significant using the 95% confidence interval for the 0-6 month horizon, while the former remains statistically significant for 0-12 month horizon.

The signs (negative/positive) of the reported estimates in Table 6.2 reflect the signs of the direct effects estimates as well as the indirect and total effects which have the same signs in our SAR probit model. The reported signs suggest that some variables exhibited consistent signs for all three time horizons. Specifically, flood depth has a negative influence on the probability of firms reopening, (logged) median household income has a positive impact, small and large firms relative to medium sized firms have a negative impact, low social economic status of customers has a negative effect (relative to middle social economic status), and sole proprietorships exhibited a positive effect on the probability of stores reopening.

Two variables exhibited a change in sign between the first two time horizons and the last, high social economic status and ownership by national chains, which both had a positive impact on reopening at the first two time horizons but a negative impact in the last time horizon. These changes of sign suggest the role played by high social economic status and ownership by national chains are different as time elapsed.

6.2 Direct, indirect, and total spatial effects

The effects estimates for models covering the three time horizons are shown in Tables 6.3 to 6.5. These are the basis for inference regarding the impact of changes in the various explanatory variables on the probability of stores reopening as well as the (cumulative) spatial spillover impacts on neighboring stores. Unless noted otherwise, the effects estimate discussed here are statistically significant based on the 99% confidential interval.

In the 0-3 month horizon, the flood depth variable exerts a negative direct and (cumulative) indirect effect on the probability of stores reopening implying a direct effect of 4.4% decrease in probability of reopening for every one foot of flood depth, an indirect effect around 3.1%, combining for a total effect around 7.6%. Over time, the direct effects of flooding decrease to 2.3% by the 0-12 month horizon, whereas the indirect or spillover impacts decline by a smaller amount to 3.0% by the 0-12 month horizon. Total effects decline from 7.6% to around 5.3% over time. This seems consistent with the notion that disaster assistance as well as market forces work over time to produce a move back towards equilibrium, ameliorating the impact of flooding with the passage of time.

In the 0-3 month horizon, (logged) median household income of the Census block group in which the store was located had a positive and significant direct effect that would raise the probability of stores reopening by 19.7% for every 1% increase in income, an indirect effect of 13.9% for a total effect of 0.34. Over the next two time horizons the direct, indirect and total effect of income remained positive, but diminished in size to produce a total effect that was insignificant (using the 0.05 and 0.95 credible intervals).

In the 0-3 month period, small size stores (measured by employment) had a negative direct effect, reducing the probability of reopening by around -7.7% (relative to the omitted category of medium sized/employment stores) with an indirect/spillover effect around -5.6% for a total effect around -13.3%. As in the case of the income variable, this variable became insignificant at the 0-6 and 0-12 month horizons. In comparison, the large size stores had a slightly smaller but still negative direct effect of -0.09, an indirect effect of -0.07, and a total effect of -0.16 in the 0-3 month period. The effects estimates for the first period are significant with 0.1 α-level, and become insignificant as time elapses.

Low social economic status of the store clientele had a negative and significant impact for all three time horizons. During the 0-3 month period, the direct effects were equal to -0.091, with indirect effects of -0.066 for a total effect of -0.157 suggesting a decrease in probability of reopening relative to the omitted class of middle social economic status. Over time the direct effects remained about the same during the next two periods, while the indirect effects increased to -0.11 and -0.10 in these latter two periods. We note that high social economic status clientele had a positive but insignificant effect (relative to the omitted class of middle social economic status) for all three time horizons.

The other variable exhibiting significant effects was the ownership dummy variable representing sole proprietorships (relative to the excluded class of regionally owned chains). This variable had a positive direct effect of 0.16 and indirect effects of 0.11, suggesting a positive total effect around 0.27 on the probability of reopening in the 0-3 month horizon relative to the excluded class of regionally owned chains (the total effects for national chains are 3%). This estimate seems consistent with the notion that sole proprietorships would be more likely to reopen quickly since this likely represents their sole source of economic support.

Over time, the direct effect declined to 0.09 and 0.03, becoming insignificant during the 0-12 month period. The indirect effect remained high during the 0-6 month period having a magnitude of 0.12, but declined to 0.05 and became insignificant in the 0-12 month period. This pattern of relationships lead to a total effect in the 0-3 month period of 0.27 declining to 0.21 at the 0-6 month horizon and an insignificant 0.08 in the 0-12 month period. Again, the notion that as time passes the influence of this ownership type might diminish as recovery leads back to an equilibrium situation seems intuitively plausible.

These scalar summary effects estimates produce a valid global inference regarding the direction and comparative magnitude of influence for the model variables on the probability of stores' re-opening decisions. Deeper flood, smaller business size, lower socioeconomic status of the store cliente, sole proprietorships, and higher median household income of the Census block group at a particular firm's location, would lead to a larger probabilities of reopening of that firm, as well as positive spatial spillover effects which increase neighboring firms' probability of re-opening.

Table 6.2: SAR Probit model estimates for three time horizons

Part 1: 0-3 months time horizon			
Variables	Posterior Mean	Std Deviation	p-level
constant	-7.158531	2.677524	0.004000
flood depth	-0.155659	0.043976	0.000000
log(median income)	0.686685	0.260414	0.004000
small size	-0.266117	0.143268	0.031000
large size	-0.323012	0.330496	0.161000
low status customers	-0.318269	0.165219	0.023000
high status customers	0.097288	0.133581	0.227000
sole proprietorship	0.548108	0.195808	0.003000
national chain	0.063874	0.372836	0.435000
Wy	0.417119	0.098652	0.000000
Part 2: 0-6 months time horizon			
Variables	Posterior Mean	Std Deviation	p-level
constant	-2.943112	2.587916	0.133000
flood depth	-0.110292	0.032870	0.000000
log(median income)	0.306807	0.254534	0.116000
small size	-0.107887	0.146387	0.228000
large size	-0.356865	0.318280	0.126000
low status customers	-0.355684	0.165377	0.016000
high status customers	0.027769	0.155180	0.442000
sole proprietorship	0.375451	0.181157	0.020000
national chain	0.286679	0.381303	0.232000
Wy	0.577877	0.076106	0.000000
Part 3: 0-12 months time horizon			
Variables	Posterior Mean	Std Deviation	p-level
constant	-3.825662	2.999146	0.096000
flood depth	-0.101883	0.032870	0.000000
log(median income)	0.433212	0.295146	0.068000
small size	-0.188510	0.159744	0.115000
large size	-0.269888	0.326356	0.198000
low status customers	-0.347205	0.162231	0.018000
high status customers	-0.103846	0.165248	0.248000
sole proprietorship	0.155128	0.188673	0.212000
national chain	-0.123218	0.383073	0.365000
Wy	0.586851	0.079149	0.000000

Table 6.3: SAR Probit model effects estimates for 0-3 month time horizon

Variables	Lower 0.05	Posterior Mean	Upper 0.95
Part 1: Direct effects			
flood depth	-0.069600	-0.044884	-0.021917
log(median income)	0.053467	0.197126	0.332591
small size	-0.160497	-0.076981	0.004433
large size	-0.273167	-0.092907	0.094821
low status customers	-0.184635	-0.091560	-0.002830
high status customers	-0.052098	0.028067	0.105184
sole proprietorship	0.051317	0.157991	0.266298
national chain	-0.190847	0.018582	0.224714
Part 2: Indirect effects			
flood depth	-0.057033	-0.031440	-0.012224
log(median income)	0.028678	0.139540	0.291662
small size	-0.137540	-0.055895	0.004233
large size	-0.228540	-0.069595	0.069056
low status customers	-0.176937	-0.066379	-0.001096
high status customers	-0.036890	0.020239	0.085859
sole proprietorship	0.027357	0.115004	0.254768
national chain	-0.163904	0.012313	0.178825
Part 3: Total effects			
flood depth	-0.116964	-0.076323	-0.040174
log(median income)	0.091016	0.336666	0.571705
small size	-0.282904	-0.132876	0.008180
large size	-0.487373	-0.162502	0.155165
low status customers	-0.333515	-0.157939	-0.003922
high status customers	-0.089262	0.048306	0.180701
sole proprietorship	0.083613	0.272995	0.487063
national chain	-0.345965	0.030895	0.381417

Table 6.4: SAR Probit model effects estimates for 0-6 month time horizon

Variables	Lower 0.05	Posterior Mean	Upper 0.95
Part 1: Direct effects			
flood depth	-0.043040	-0.027292	-0.012419
log(median income)	-0.055334	0.076065	0.195196
small size	-0.101645	-0.027560	0.042505
large size	-0.249673	-0.089563	0.066283
low status customers	-0.167701	-0.088025	-0.007667
high status customers	-0.068971	0.007314	0.091351
sole proprietorship	0.006258	0.094298	0.191935
national chain	-0.114009	0.071363	0.263166
Part 2: Indirect effects			
flood depth	-0.057026	-0.034932	-0.016612
log(median income)	-0.081650	0.095816	0.259106
small size	-0.153473	-0.036527	0.054500
large size	-0.355161	-0.116735	0.083706
low status customers	-0.245362	-0.114874	-0.013511
high status customers	-0.096922	0.007633	0.112239
sole proprietorship	0.006338	0.124584	0.280572
national chain	-0.148355	0.094969	0.383949
Part 3: Total effects			
flood depth	-0.092647	-0.062224	-0.030839
log(median income)	-0.136406	0.171881	0.433592
small size	-0.255375	-0.064088	0.100316
large size	-0.598493	-0.206298	0.143543
low status customers	-0.395030	-0.202899	-0.020548
high status customers	-0.163803	0.014947	0.200101
sole proprietorship	0.014259	0.218882	0.459625
national chain	-0.270969	0.166333	0.625045

Table 6.5: SAR Probit model effects estimates for 0-12 month time horizon

Variables	Lower 0.01	Posterior Mean	Upper 0.99
Part 1: Direct effects			
flood depth	-0.038171	-0.023219	-0.008507
log(median income)	-0.028286	0.098796	0.232998
small size	-0.118636	-0.043713	0.027404
large size	-0.234627	-0.062016	0.094639
low status customers	-0.158440	-0.079941	-0.006709
high status customers	-0.104160	-0.023621	0.054097
sole proprietorship	-0.049391	0.035359	0.122883
national chain	-0.198385	-0.029019	0.150640
Part 2: Indirect effects			
flood depth	-0.049707	-0.030394	-0.013516
log(median income)	-0.042144	0.128706	0.315660
small size	-0.185028	-0.059541	0.041219
large size	-0.323216	-0.083896	0.11303
low status customers	-0.226731	-0.105796	-0.008666
high status customers	-0.143146	-0.031171	0.074168
sole proprietorship	-0.066163	0.048810	0.179643
national chain	-0.279671	-0.038422	0.211160
Part 3: Total effects			
flood depth	-0.082558	-0.053613	-0.023535
log(median income)	-0.071296	0.227502	0.534166
small size	-0.298076	-0.103254	0.068205
large size	-0.525085	-0.145912	0.196055
low status customers	-0.368830	-0.185738	-0.015330
high status customers	-0.238708	-0.054792	0.125534
sole proprietorship	-0.112203	0.084169	0.291942
national chain	-0.466476	-0.067441	0.366107

Chapter 7

CONCLUSIONS

This study employs a theoretical framework from micro-scale retail location studies and implements a spatial autoregressive probit model to account for spatial dependence among firms' decisions to study the determinants of business return to New Orleans after Hurricane Katrina. Spatial dependence of household or firm decisions when these economic agents are located nearby is likely to be a frequently encountered situation in applied spatial modeling. Existing literature on micro-scale retail locations suggest that extant patterns of retail activity influence choices of consumers (or patrons) and location decisions by retailers. The decision of a business establishment to re-open after a natural disaster such as Katrina could be viewed as similar to the original location decision made by the firm when it entered business. If the surrounding environment was important for the original decision, it should be equally influential in the decision to re-open in the same or another location after the disaster. This consensus provides a strong motivation for adopting an spatial probit model which allows for interdependence among observations.

Along with estimating the spatial probt model, the study adopts scalar summary measures for the total effects (or impacts) associated with changing the explanatory variables. This is analogous to the situation arising in non-spatial probit models where 'marginal effects' are calculated in an effort to consider the nonlinearity of response in decision probabilities with respect to changes in the magnitude of the explanatory variables.

The findings with regard to factors influencing business return in New Orleans after Katrina suggests a significant positive spatial dependence in firm decisions regarding open/close status so that firms located nearby exhibit similar decision outcomes regarding reopening. This statistically significant spatial dependence justifies the adoption of a spatial probit model. As expected, flood depth plays a negative and significant role in affecting firms' decision regarding reopening, as the associated coefficients are significant for all three time horizons. A scalar summary of the spatial spillover effects among firms produces a valid global inference regarding the direction and comparative magnitude of influence for the model variables on the probability of stores' re-opening decisions. Deeper flood, smaller business size, lower socioeconomic status of the store cliente, sole proprietorships, and higher median household income of the Census block group at a particular firm's location, would lead to larger probabilities of reopening for that particular firm, as well as positive spatial spillover effects which increase neighboring firms' probability of re-opening.

The results also show that the effects of different variables change, in terms of both direction and magnitude, with passage of time after the disaster. For example, two variables exhibited a change in sign between the first two time horizons and the last, high social economic status and ownership by national chains, which both had a positive impact on reopening during the short time horizons but a negative impact in the longer time horizon. These changes of sign suggest the role played by associated variables are different as time elapsed. As for effects' magnitude, we found that in the short-term (0-3 months) the sole proprietorship ownership category had a positive impact on the probability of these firms reopening, as well as a positive impact on the probability that neighboring establishments reopened. With the passage of time (0-6 months), the direct impact of this ownership type (relative to regional chains) diminished while the spatial spillover impacts on neighboring firms grew, with both remaining positive and significant. As one might expect, very high levels of flooding at store locations tended to mitigate the positive impacts associated with this type of ownership. In the longer-time (0-12 months) as the impact of disaster aid and other forces bringing the economic climate back towards normality, factors that influenced the probability of reopening in the short-term diminished to the point of insignificance in many cases.

Another contribution was a Monte Carlo study of extended information criteria, such as the AIC, BIC, and DIC, for application in spatial econometric modeling. This provides some insights about performance of different model selection tools for choosing a spatial weight matrix, conditioning on the type of dependent variable, level of spatial dependence, and signal-noise ratio. The experiment indicated that different information criteria as well as log-marginal likelihood do not perform uniformly under different conditions.

As for selection of the correct spatial weight matrix in the case of a model involving a continuous dependent variable, all model selection tools were capable of recovering the model based on the true weight matrix and the power to discriminate between models increased with larger sample sizes, increased signal-to-noise ratios, and stronger spatial dependence. The AIC, BIC and DIC information criteria produced similar model results in the continuous model setting. Although performance of all model selection tools under investigation converged in the presence of high levels of spatial dependence, the information criteria outperformed the log-marginal likelihood in cases involving weak spatial dependence. Therefore, information-based model selection criteria seem suitable for selecting the appropriate spatial weight matrix in spatial econometric models involving a continuous dependent variable. In contrast, for the case of the spatial probit model where the dependent variable was discrete, the log-marginal likelihood would be the best method for choice of the appropriate spatial weight matrix. The log-marginal likelihood performed similarly in both continuous and discrete dependent variable models. Performance of the information-based model selection criterion was worse in the discrete/spatial probit model setting. Moreover, DIC consistently produced the worst weight matrix selection results. Another finding was that all model selections tools were biased upwards in terms of selecting weight matrices with a larger number of nearest neighbors than used to generate the experimental dependent variable vector. The DIC had the largest upward bias with AIC/BIC second. Another finding was that adjustments to the degrees of freedom used in the AIC and BIC criterion to take into account spatial dependence that arises in spatial models did not produce fruitful results.

BIBLIOGRAPHY

Akaike, Hirotugu. 1973. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19, no. 6: 712-23.

Albert, J. H., and S. Chib. 1993. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 88, no. 422: 669-79.

Alonso, William. 1960. A theory of the urban land market. *Papers in Regional Science* 6, no. 1: 149-57.

Anselin, Luc. 1988. *Spatial econometrics : Methods and models*. Kluwer Academic Publishers.

Anselin, Luc, and S. Rey. 1991. Properties of tests for spatial dependence in linear-regression models. *Geographical Analysis* 23: 112-31.

Anselin, Luc, and Anil K. Bera. 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In Ullah A., and D. E. Giles (eds) *Handbook of applied economic statistics*: 237-89. New York: Marcel Dekker.

Anselin, Luc. 2003. Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review* 26, no. 2: 153-66.

Arenas, Helbert, and Lam Nina. 2009. Evaluating predictors for business recovery in New Orleans after Hurricane Katrina. Paper presented at the 105th Association of American Geographers annual meeting, Las Vegas, NV.

Bartik, Timothy J. 1985. Business location decisions in the United States: Estimates of the effects of unionization, taxes, and other characteristics of States. *Journal of Business and Economic Statistics* 3, no. 1: 14-22.

Belsley, D.A., E. Kuh, and R.E. Welsch. 1980. *Regression Diagnostics: Identifying influential data and sources of collinearity*. New York: John Wiley.

Berke, Philip R., Jack Kartez, and Dennis Wenger. 1993. Recovery after disaster: Achieving sustainable development, mitigation and equity. *Disaster* 17, no. 2: 93-109.

Berman, B., and J. R. Evans. 1991. *Retail management: A strategic approach.* New York: Macmillan.

Beron, Kurt J., and Wim P. M. Vijverberg. 2004. Probit in a spatial context: A Monte Carlo analysis. In Anselin Luc, Raymond J. Florax, and Sergio J. Rey (eds) *Advances in spatial econometrics: methodology, tools and applications*: 145-69. Berlin Heidelberg: Springer-Verlag.

Berry, B. J. L., and W.L. Garrison 1958. A note on central place theory and the range of a good. *Economic Geography* 34: 304-311.

Bhattacharjee, A., and C. Jensen-Butler. 2006. Estimation of spatial weight matrix, with an application to diffusion in housing demand.

available at: http://www.ucergy.fr/IMG/A._Bhattacharjee_020206.pdf.

Birkland, Thomas A., and Radhika Nath. 2000. Business and political dimensions in disaster management. *Journal of Public Policy* 20: 275-303.

Bolduc, Denis, Benard Fortin, and Stephen Gordon. 1997. Multinomial probit estimation of spatially interdependent choices: An empirical comparison of two new techniques. *International Regional Science Review* 20: 77-101.

Borg, Walter R., and Meredith D. Gall. 1989. *Educational research: An introduction.* New York: Longman.

Borchert, Johan G. 1998. Spatial dynamics of retail structure and the venerable retail hierarchy. *GeoJournal* 45, no. 4: 327-336.

Brown, S. 1987. A perceptual approach to retail agglomeration. *Area* 19, no. 2: 131-140.

Brown, S. 1989. Retail location theory: The legacy of Harold Hotelling. *Journal of Retailing* 65, no. 4: 450-70.

Brown, S. 1994. Retail location at the micro-scale: inventory and prospect. *Service Industries Journal* 14, no. 4: 542-76.

Brown, S. 2002. Retail location theory-the legacy of Harold Hotelling. In Findlay Anne M., and Leigh Sparks (eds) *Retailing: The evolution and development of retailing* 450-70. London: Taylor and Francis Group.

Campanella, Richard. 2006. *Geographies of New Orleans: Urban fabrics before the storm.* New Orleans, LA: Center for Louisiana Studies.

Carter, Charles C., and Kerry D. Vandell. 2005. Store location in shopping centers: Theory and estimates. *Journal of Real Estate Research* 27, no. 3: 237-66.

Carter, Charles C., and William J. Haloupek. 2000. Spatial autocorrelation in a retail context. *International Real Estate Review 3*, no. 1: 34-48.

Case, Anne. 1992. Neighbourhood influence and technological change. *Regional Science and Urban Economics* 22: 491-508.

Casetti, E. 1972. Generating models by the expansion method: Applications to geographic research. *Geographical Analysis* 4: 81-91.

Christaller, W. 1966. *Central places in southern Germany*, trans. C. W. Baskin, Englewood Cliffs. Prentice Hall, NJ.

Cliff, A.D., and J.K. Ord. 1981. *Spatial processes: Models and applications.* London: Pion.

Coughlin, Cletus C., Thomas A. Garrett, and Ruben Hernandez-Murillo. 2003. Spatial probit and the geographic pattern of state lotteries. St. Louis Federal Reserve Bank Working Paper. available at: http://research.stlouisfed.org/wp/2003/2003042.pdf.

Craig, Samuel, Avijit Ghosh, and Sara McLafferty. 1984. Models of the retail location process: A review. *Journal of Retailing* 60, no. 1: 5-36.

Cressie, N. 1993. *Statistics for spatial data.* New York: John Wiley.

Davis, Peter. 2006. Spatial competition in retail market: Movie theatres. *The RAND Journal of Economics* 37, no. 4: 964-982.

De Palma, A., V. Ginsburgh, Y. Y. Papageorgiou, and J.-F. Thisse. 1985. The principle of minimum differentiation holds under sufficient heterogeneity. *Econometrica* 53, no. 4: 767-781.

Diniz-Filho J.A.F., T.F.L.V.B. Rangel, and L.M. Bini. 2008. Model selection and information theory in geographical ecology. *Global Ecology and Biogeography* 17: 479-88.

Drezner, T., and Z. Drezner. 2002. Retail facility location under changing market conditions. *IMA Journal of Management Mathematics* 13: 283-302.

Drezner, Z., and J. Guyse. 1999. Application of decision analysis techniques to the Weber facility location problem. *European Journal of Operational Research* 116: 69-79.

Duranton, Gilles and Diego Puga. 2003. Micro-foundations of urban agglomeration economies. CEPR Discussion Paper. available at: http://ssrn.com/abstract=468960.

Eaton, Curtis B., and Richard G. Lipsey. 1975. The principle of minimum differentiation reconsidered: Some new developments in the theory of spatial competition. *The Review of Economic Studies* 42, no. 1: 27-49.

Fleming, Mark M. 2004. Techniques for estimating spatially dependent discrete choice models. In Anselin Luc, Raymond J. Florax, and Sergio J. Rey (eds) *Advances in spatial econometrics: methodology, tools and applications* 145-69. Heidelberg, Germany: Springer-Verlag.

Florax, R.J.G.M., and S. Rey. 1995. The impact of misspecified spatial structure in linear regression models. In: Anselin L., and R.J.G.M. Florax (eds) *New Directions in Spatial Econometrics*. Berlin: Springer-Verlag.

Fotheringham, A.S., C. Brunsdon, and M. Charlton. 2002. *Geographically weighted regression: The analysis of spatially varying relationships*. West Sussex, U.K: John Wiley & Sons Ltd.

Franzese, Jr., Robert, and Jude Hays. 2008. The spatial probit model of interdependent binary outcomes: Estimation, interpretation, and presentation. Paper presented at the annual meeting of the MPSA Annual National Conference, Chicago, IL.

available at: http://www.allacademic.com/meta/p265722_index.html.

Fujita, M., and T.E. Smith. 1990. Additive-interaction models of spatial agglomeration. *Journal of Regional Science* 30, no. 1: 51-74.

Fujita, M., and Jacques-Francois Thisse. 1996. Economics of agglomeration. *Journal of The Japanese and International Economics* 10: 339-378.

Garling, T., and E. Garling. 1987. Distance minimization in downtown pedestrian shopping. *Environment and Planning A* 20, no. 4: 547-554.

Getis, A., and Jared Aldstadt. 2004. Constructing the spatial weight matrix using a local statistic. *Geographical Analysis* 36: 90-104.

Gelfand, A.E., and A.F.M. Smith. 1990. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85: 398-409.

Gelfand, A.E., S.E. Hills, A. Racine-Poon, and A.F.M. Smith. 1990. Illustration of Bayesian inference in normal data models using Gibbs sampling, Journal of the American Statistical Association 85, 972-985.

Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin. 1995. *Bayesian data analysis.* New York: Chapman & Hall.

Gelman, Andrew, and Jennifer Hill. 2007. *Data analysis using regression and multilevelhierarchical models.* New York: Cambridge Press.

Gilley, O.W. and R.K. Pace. 1996. On the Harrison and Rubinfeld Data. *Journal of Environmental Economics and Management* 31, no. 3: 403-5.

Golledge, R.G., and Robert J. Stimson. *Spatial behavior: A geographic perspective.* NY: Guilford Press.

Golledge, R.G., and H. Timmermans. 1990. Applications of behavioural research on spatial problems. *Progress in Human Geography* 14, no. 1: 57-99.

Griffith, Daniel, and F. Lagona. 1998. On the quality of likelihood-based estimators in spatial autoregressive models when the data dependence structure is misspecified. *Journal of Statistical Planning and Inference* 69: 153-74.

Guimaraes, Paulo, Frank L. Hefner, and Douglas P. Woodward. 1993. Wealth and income effects of natural disasters: An econometric analysis of Hurricane Hugo. *Review of Regional Studies* 23, no. 3: 97-114.

Guy, Clifford. M. 1995. *The retail development process: Location, property and planning.* UK: Routledge.

Hanaoka, Kazumasa, and Graham P. Clarke. 2007. Spatial microsimulation modelling for retail market analysis at the small-area level. *Computers, Environment, and Urban Systems* 31, no. 2: 162-187.

Harrison, D., and D.L. Rubinfeld. 1978. Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management* 5: 81-102.

Hernandez, T., and D. Bennison. 2000. The art and science of retail location decisions. *International Journal of Retail and Distribution Management* 28, no. 8: 357-67.

Hinloopenaand, J., and C. van Marrewijk. 1999. On the limits and possibilities of the principle of minimum differentiation. *International Journal of Industrial Organization* 17, no. 5: 735-50.

Hoeting J.A., R.A. Davis, A.A. Merton, and S.E. Thompson. 2006. Model selection for geostatistical models. *Ecological Applications* 16, no. 1: 87-98.

Holloway, G., B. Shankara, and S. Rahman. 2002. Bayesian spatial probit estimation: A primer and an application to HYV rice adoption. *Agricultural Economics* 27, no.3: 383-402.

Holloway, Garth J., Donald Lacombe, and Timothy O'Shaugnessy. 2005. How large is congressional dependence: Bayesian spatial probit analysis of congressional voting on the Farm Bill. Working paper of Dr. Garth J. Holloway.

available at: http://www.pubchoicesoc.org/papers2005/Lacombe.pdf.

Hotelling, H. 1929. Stability in competition. *The Economic Journal* 39, no. 153: 41-57.

Hunt, M.A., and J.L. Crompton. 2008. Investigating attraction compatibility in an East Texas city. *Progress in Tourism and Hospitality Research* 10, no. 3: 237-46.

Irmen, Andreas, and Jacques-Francois Thisse. 1998. Competition in multi-characteristics spaces: Hotelling was almost right. *Journal of Economic Theory* 78: 76-102.

Keeble, David. 1978. Industrial Decline in the Inner City and Conurbation. *Transactions of the Institute of British Geographers* 3, no. 1: 101-14.

Kennedy, Peter. 2001. *A guide to econometrics.* Cambridge, MA: The MIT Press.

Kevin McCarthy, D.J. Peterson, Narayan Sastry, and Michael Pollard. 2006. The repopulation of New Orleans after Hurricane Katrina. *Rand Gulf States Policy Institute Technical Report.* available at: http://www.rand.org/pubs/technical_reports/2006/RAND_TR369.pdf.

Kissling W.D., and G. Carl. 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography* 17: 59-71.

Konishi, Hideo. 2005. Concentration of competing retail stores. *Journal of Urban Economics* 58: 488-512.

Koop, Gary. 2003. *Bayesian econometrics.* The Atrium, Southern Gata, Chichester, West Sussex, UK: John Wiley & Sons Ltd.

Krugman, Paul R. 1991. Increasing return and economic geography. *Journal of Political Economics* 99, no. 3: 483-499.

Krugman, Paul R. 1992. *Geography and trade*. Cambridge, MA: MIT Press.

Lam, Nina, Kelley Pace, Richard Campanella, James LeSage, and Helbert Arenas. 2009. Business return in New Orleans: Decision making amid Post-Katrina uncertainty. *Public Library of Science (PLoS ONE).* available at: http://dx.plos.org/10.1371/journal.pone.0006765.

Lee H., and S.K. Ghosh. 2009. Performance of information criteria for spatial models. *Journal Statistical Computing and Simulation* 79, no. 1: 93-106.

LeSage, J.P. 1997. Bayesian estimation of spatial autoregressive models. *International Regional Science Review* 20: 113-129.

LeSage, James P. 1999. Applied econometrics using Matlab. Econometrics toolbox for Matlab. available at: http://www.spatial-econometrics.com/html/mbook.pdf.

LeSage, James P. 2000. Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis* 32, no. 1: 19-35.

LeSage, J.P., and Olivier Parent. 2007. Bayesian model averaging for spatial econometric models. *Geographical Analysis* 39: 241-267.

LeSage, J.P. and R.K. Pace. 2009. *Introduction to Spatial Econometrics*. New York: Taylor and Francis/CRC Press.

LeSage, J.P., R.K. Pace, N., Lam, R. Campanella, and X. Liu. 2010. New Orleans business recovery in the aftermath of Hurricane Katrina. Working paper.

Liang, Wenjung, and Chaocheng Mai. 2006. Validity of the principle of minimum differentiation under vertical subcontracting. *Regional Science and Urban Economics* 36, no. 3: 373-84.

Malmberg, Anders. 1997. Industrial geography: Location and learning. *Progress in Human Geography* 21, no. 4: 573-82.

McGoldrick, P.J., and M.G. Thompson. 1992. *Regional shopping centres: In-town versus out-of-town*. Aldershot, UK: Avebury Press.

McMillen, Daniel P. 1992. Probit with spatial autocorrelation. *Journal of Regional Sciences* 32, no. 3: 335-48.

McMillen, Daniel P. 1995. Spatial effects in probit models: A Monte Carlo investigation. In Anselin Luc, and Raymond J. Florax (eds) *New directions in spatial econometrics* 189-229. Heidelberg, Germany: Springer-Verlag.

McMillen, Daniel P. 1997. Multiple regime bid-rent function estimation. *Journal of Urban Economics* 41, no. 2: 301-19.

Murtaugh, P.A. 2009. Performance of several variable-selection methods applied to real ecological data. *Ecology Letters* 12: 1061-1068.

Nelson, R.L. 1958. *The selection of retail cocations.* New York: Dodge.

Netz, Janet S., and Beck A. Taylor. 2002. Maximum or minimum differentiation? Location patterns of retail outlets. *The Review of Economics and Statistics* 84, no.1: 162-175.

O'Roarty, Brenna, Stanley McGreal, and Alastair Adair. 1997. The impact of retailers' store selection criteria on the estimation of retail rents. *Journal of Property Valuation and Investment* 15, no. 2: 119-30.

Ottaviano, Gianmarco, and Thisse, Jacques-Francois. 2003. Agglomeration and economic geography. CEPR Discussion Paper. available at: http://ssrn.com/abstract=410026.

Ozturk, Erdogan, and Elena G. Irwin. 2001. Explaining household location choices using a spatial probit model. Paper presented at the 2001 Agricultural Economics Association Meeting, Chicago, IL. available at: http://ageconsearch.umn.edu/bitstream/20626/1/sp01oz01.pdf.

Papageorgious Y.Y., and Jacques-Francois Thisse. 1985. Agglomeration as spatial interdependence between firms and households. *Journal of Economic Theory* 37, no. 1: 19-31.

Pinkse, Joris, and Margaret E. Slade. 1998. Contracting in space: An application of spatial statistics to discrete choice models. *Journal of Econometrics* 85: 125-54.

Porter, Michael E. 2000. Location, competition, and economic development: Local clusters in a global economy. *Economic Development Quarterly* 14, no. 1: 15-34.

Rathbun, Stephen L., and Songlin Fei. 2006. A spatial zero-inflated Poisson model for oak regeneration. *Environmental and Ecological Statistics* 13, no. 4: 406-26.

Raftery, A.E. 1996. Hypothesis testing and model selection. In W.R. Gilks, D.J. Spiegelhalter and S. Richardson (eds) *Markov Chain Monte Carlo in Practice* London: Chapman and Hall, 163-188 .

Rhee, Byongduk. 1989. Restoring the principle of minimum differentiation. Working paper of Division of Research, School of Business Administration, University of Michigan. available at: http://hdl.handle.net/2027.42/36021.

Reilly, W.J. 1931. *The Law of Retail Gravitation.* New York: Knickerbocker Press.

Reimers, Vaughan, and Val Clulow. 2004. Retail concentration: A comparison of spatial convenience in shopping strips and shopping centre. *Journal of Retailing and Consumer Service* 11, no. 4: 207-221.

Runyan, Rodney C. 2006. Small business in the face of crisis: Identifying barriers to recovery from a natural disaster. *Journal of Contingencies and Crisis Management* 14, no. 1: 12-26.

Schenk, Tilman A., Gunter Loffler, and Jurgen Rauh. 2007. Agent-based simulation of consumer behavior in grocery shopping on a regional level. *Journal of Business Research* 60: 894-903.

Schwarz, Gideon. 1978. Estimating the Dimension of a Model. *The Annals of Statistics* 6, no. 2: 461-464.

Simon, Herbert A. 1959. Theories of decision-making in economics and behavioral science. *The American Economic Review* 49, no. 3: 253-83.

Simon, Herbert A. 1979. Rational decision making in business organizations. *The American Economic Review* 69, no.4: 493-513.

Smith, T., and J.P. LeSage. 2000. A Bayesian probit model with spatial dependencies. In LeSage, James P., and R. Kelley Pace (eds) *Advances in econometrics: Spatial and spatiotemporal econometrics* 127-60. Oxford, UK: Elsevier.

Smith, T. 2009. Estimation bias in spatial models with strongly connected weight matrices. *Geographical Analysis* 41, no. 2: 307-332.

Spiegelhalter, D., N. Best, B. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* 64: 583-639.

Stakhovych, S., and T.H.A. Bijmolt. 2008. Specification of spatial models: A simulation study on weights matrices. *Papers in Regional Science* 88, no. 2: 389-408.

Tierney, Kathleen J., and Gary R. Webb. 2001. Business vulnerability to earthquakes and other disasters. available at: http://dspace.udel.edu:8080/dspace/handle/19716/729.

Ward E.J. 2008. A review and comparison of four commonly used Bayesian and maximum likelihood model selection tools. *Ecological Modelling* 211: 1-10.

Zacharias, John. 2001. Pedestrian behaviour and perception in urban walking environments. *Journal of Planning Literature* 16, no. 1: 3-18.

Zellner, Arnold. 1971. *An introduction to Bayesian inference in econometrics*. New York: John Wiley & Sons.

Zhou, B, and K.M. Kockelman. 2008. Neighbourhood impacts on land use change: A multinomial logit model of spatial relationships. *The Annals of Regional Science* 42, no.2: 321-40.

VITA

Xingjian LIU was born in Wuhan, China, on September 29, 1986, the son of Yaolin LIU and Yanfang LIU. After completing his Bachelor's study in Geographic Information System at Wuhan University, China in 2008, Xingjian came to the U.S., and entered the Master's program in Geography at Texas State-San Marcos. Xingjian has been a student of Geography for a long time, as his parents are professors of Geography in China.

Permanent address: School of Resources and Environment Science,
Wuhan University, Wuhan, China 430079

This thesis was typed by Xingjian Liu.