THREE ESSAYS ON SOCIAL DYNAMICS AND LAND-USE CHANGE: FRAMEWORK, MODEL AND ESTIMATOR

BY

SAKET SARRAF

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Regional Planning
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Professor Varkki G. Pallathucheril, Chair
Professor Kieran P. Donaghy, Director of Research
Professor Geoffrey J. D. Hewings
Associate Professor Brian Deal

Abstract

Simulations of future regional land-use change help planners and policymakers understand how scenarios with alternative public policy and investment choices will play out in the future, especially in conjunction with different economic and demographic trends. Models used to simulate land-use change are driven by physical, economic, infrastructure, environmental and geographical factors. The magnitude of growth is determined exogenously, independent of the existing land-use and socio-economic conditions. Even though significant relationship among income inequality, racial segregation, housing abandonment, intra-regional migration, school quality, amenities and urban growth have been established, the current generation of land-use change models does not explicitly model these social dynamics. This omission results in underestimation of new housing construction, and failure to account for urban decline, sprawl, intra-regional migration, regional imbalances and social externalities.

The research described in this dissertation seeks to address this omission and is presented as three essays. The first essay explores how the spatial pattern of socio-economic characteristics determines the magnitude and location of growth and is shaped by it. It presents a framework for modeling the relationship among regional inequalities, urban distress and growth. The framework is tested using data from the St. Louis region. The second essay uses the framework to model the impact of the social dynamics on location and magnitude of growth and decline in the region. The forecasts from this model, together with population and income forecasts from a regional economic model, are used to derive probabilities of development and decline at the Census tract level and to assess the total magnitude of growth endogenously. The last essay evaluates the performance of the full information Feasible Generalized Spatial Three Stages Least Square estimator used to estimate the model presented in the second essay. Using Monte Carlo experiments, the sensitivity of results to varying degrees of spatial dependences, choice of spatial weight matrix, sample size and variance covariance matrices is analyzed. These Monte Carlo simulations provide confidence in the results of the social dynamics model.

To my family

Acknowledgements

This dissertation would not have been possible without the support of those listed here and many more who I have inadvertently omitted.

My faculty advisors Varkki Pallathucheril, Kieran Donaghy, Geoffrey Hewings and Brian Deal for trusting and encouraging me to join the Ph.D. program, for their patience in allowing me to explore countless research paths, and their guidance during the course of this work leading to a meaningful conclusion.

Varkki, for his emphasis on structured argumentation, unlimited access to his time, careful technical edits of multiple versions, and for unwavering support for all issues, without which I would not have made it so far.

Kieran, for introducing me to the fields of modeling and philosophy of science, and for inculcating critical thinking.

Geoff, for constant encouragement and mentorship, and making me a part of the REAL laboratory.

Brian, for emphasis on real life problems and for unconditional support through all these years.

Varkki and Brian for allowing me to use LEAM as a test bed for this dissertation sustained and funding during my five years at UIUC which allowed me to completely focus on my work.

Colleagues and fellow researchers at LEAM laboratory, who through numerous discussion sessions shared their ideas and helped identify the many gaps in the land-use change models.

Bimal Patel for being the first person to inspire me towards a doctoral degree in Urban and Regional Planning, and never failing to ask me if I have finished.

Robert Olshansky for giving me the encouragement and support to finish this dissertation.

Arnab and Shilpa, whose hospitality and presence in the final days made it possible for me to focus on my work during the critical days. Murtaza, for proofreading at a very short notice. Jane Terry, for the welcoming smile, and prompt responses to numerous emails and queries. Jennifer, Mike and Dianne, for working out the crucial administrative support.

Numerous faculty members, staff and friends in the University of Illinois, for their love, warmth, support, and direction. Special thanks to Amita Sinha, Faranak Miraftab and Ken Salo, Lew and Susan Hopkins, Rajmohan and Usha Gandhi, Stacy Harwood, Varkki and Susan, and Zorica Nedović-Budić.

My wife Parul, for continuing to believe in me, and for her support to manage everything else whilst I was working on this dissertation, without which it would never see the light of day. And to my parents and sisters, who always trusted me with all my decisions and provided unconditional support all throughout.

Table of Contents

CHAPTER 1: INTRODUCTION 1
1. Problem Statement 1
2. Land-use Evaluation and Impact Assessment Model 3
3. Three Essays 4
4. Expected Contributions 6

CHAPTER 2: INCORPORATING SOCIAL DYNAMICS INTO MODELS OF LAND-USE CHANGE.. 8
1. Introduction 8
2. Models of Land-use Change..10
3. Model Forecasts and Observed Phenomena 12
4. Social Dynamics as a Driver of Land-use Change 16
5. Framework for Modeling Social Dynamics..17
6. Social Dynamics and Land-use Change in St. Louis 24
7. Conclusion 30

CHAPTER 3: IMPLEMENTING SOCIAL DYNAMICS AS A DRIVER OF RESIDENTIAL LAND-USE CHANGE.. 32
1. Introduction 32
2. Framework for Modeling Social Dynamics..35
3. Social Dynamics Model 39
4. Integration with Land-use Change Model 47
5. Conclusion 49

CHAPTER 4: PERFORMANCE OF THE FGS3SLS ESTIMATOR IN SMALL SAMPLES - A MONTE CARLO STUDY 52
1. Introduction 52
2. Model Structure 54
3. Monte Carlo Experiments 57
4. Results 62
5. Conclusion 70

CHAPTER 5: CONCLUSION 74

REFERENCES 79

APPENDIX A: EXPLORATORY DATA ANALYSIS.. 91

APPENDIX B: SOCIAL MODEL CODE AND DIAGNOSTICS.................................. 99

APPENDIX C: MONTE CARLO CODE 110

CHAPTER 1: INTRODUCTION

Simulations of future regional land-use play a central role in planning. These simulations are a critical part of scenarios analysis, which records not just the outcome but also how the future unfolds: the factors that drive the outcome and the process through which the factors are linked to changes over space and time. In developing and working with scenarios, planners are required to perform several types of analyses. They must understand and explain how land-use currently changes and use that knowledge to predict future change. To make better sense of a set of scenarios, they must compute metrics that allow comparisons among scenarios. These metrics measure the state of various social, economic, and environmental attributes. Models of land-use change can be designed to do much of the above.

1. Problem Statement

Models of land-use change have two components: one directed at estimating demand for new land; a second directed at changing land use across space to meet this demand. This sequential spatial allocation of demand based on exogenously determined population and employment projections ignores the impact of intra-regional migration, social inequalities and urban decline on growth projections. However, most applied models of land-use change do not explicitly account for these dynamics. This absence can possibly be explained by the lack of spatial data and methodological advancements in handling complex spatial relationships. The exclusion of social dynamics from land-use change model prevents planners and policy makers from fully understanding the impacts of various public policy and investment choices.

Neighborhood ecology models are often used to model the socio-economic dynamics in regions but seldom incorporate spatial spillover effects. They are used as an overlay to spatially distribute the demand for new housing (estimated exogenously) but never to update the demand in the context of land-use change modeling. These have resulted in a failure of many land-use change models to adequately capture the phenomena of sprawl, decline of the urban core and the disproportionately higher amount of new construction relative to the increase in number of households.

This dissertation brings together four aspects of land-use modeling that have received individual attention to some extent but have never been integrated into model (a) modeling the regional economy which provides macro controls for the amount of growth in the region in terms of net change in number of households, (b) predicting the amount of new housing units needed to accommodate this growth, (c) modeling the spatial distribution of new households and housing units, (d) modeling the impact of social dynamics and regional inequality on the location and magnitude of growth.

How are regional inequalities–such as urban distress, sprawl, and regional economic growth—interlinked and how does this interrelationship inform the magnitude and location of growth in the region as simulated by a land-use change model? This is the primary question that this dissertation attempts to answer. Modeling these interdependencies provides a richer interpretation of the underlying social, economic and spatial dynamics defining land-use change in many regions of the nation than currently available. In turn, this helps planners and policymakers make better informed decisions.

2. Land-use Evaluation and Impact Assessment Model

To examine these questions, I utilize the ongoing land-use modeling efforts in the Land-use Evaluation and Impact Assessment Model (Deal & Pallathucheril, 2008) concentrating on the St. Louis region. LEAM generates land use change projections in specified geographic regions for future time horizons. It uses the local causal mechanisms of change in a region at a small scale (30 meters by 30 meters) to project land use change at annual time increments over very large regions. A discrete-choice model controls whether land use in each grid cell is transformed from its present state to a new state (residential or industrial and commercial use) in a particular time step (Figure 1.1). This transformation is based on a development score computed in each time step for each cell based on a number of factors associated with the cell (e.g., proximity to cities, employment centers, roads, highways; slope; location within wetlands, floodplains) and characteristics of surrounding cells (e.g., degree of development, type of development). The impact of each factor on a cell's development score is calibrated based on current land-use patterns in the region. In a given time step, the regional demand for new development and the development score associated with the cell determine whether or not a cell is transformed. A cell that is available for development and has a high enough development score to successfully compete to satisfy the regional demand for new land is likely to be transformed.

The model results consist of a framework of maps, movies and GIS layers that describe emergent future land use patterns (Deal, 2001). Alternative scenarios can be tested and analyzed by varying drivers or constraints, forming the basis for using LEAM as a tool for deliberation and policy analysis. The resulting implications of the emergent development patterns (on the local environmental, social and economic systems) are then analyzed.

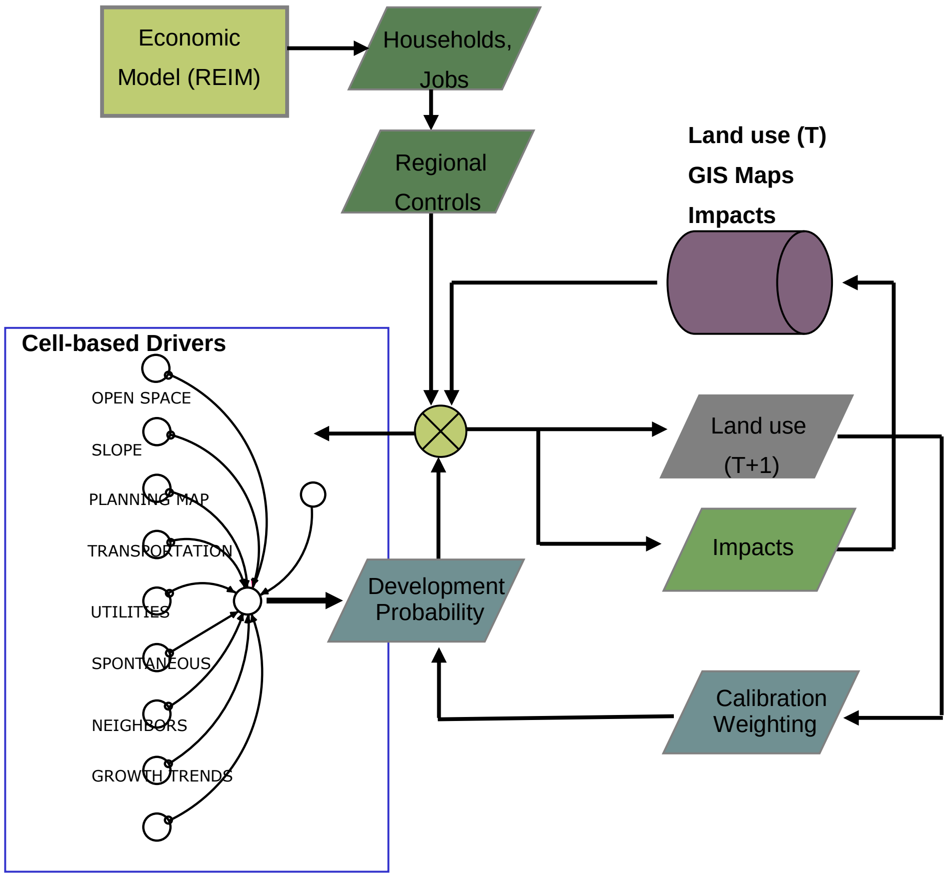


Figure 1.1: The Structure of Land-use Evolution and Impact Assessment Model (LEAM)

3. Three Essays

The research described in this dissertation address these issues in a three-essay format. It (a) creates a framework to explain the relationship between social dynamics and land-use change, (b) models the social dynamics to inform land-use change by explicitly acknowledging the spatio-temporal relationships, and (c) explores the robustness of the estimator used in the social model using Monte Carlo simulations, enabling some confidence in the model results.

The first essay explores how the spatial pattern of socio-economic characteristics determines the magnitude and location of growth and is also shaped by it. It presents a framework for modeling the relationship among regional inequalities, urban distress and growth. The framework is tested using data from the St. Louis region. These dynamics are too complex to be translated into a modeling framework and hence a reduced form approach is suggested where only certain variables relevant to modeling land-use change are further analyzed. Their inter-relationships are studied through exploratory spatial data analysis using data from St. Louis region. This exploratory exercise does not test individual hypotheses presented in the causal map, but assesses if the outcomes predicted by these dynamics corroborate with the observed reality.

The second essay uses the above framework to model the impact of the social dynamics on location and magnitude of growth and decline in the region. A reduced form of the causal map is modeled with a system of simultaneous equations to predict the location and magnitude of growth with the endogenous variables consisting of proxies for socio-economic fabric (income, vacancy, number of households), growth (new housing construction) and decline (housing abandonment). Given the spatial nature of urban processes, spatial dependences in endogenous variables as well as in the error structure are explicitly treated. A full information feasible generalized spatial three stages least square estimator (FGS3SLS) is used to estimate the model. The forecasts from this model, together with population and income forecasts from a regional economic model, are used to derive probabilities of development and decline at the Census tract level and to also assess the total magnitude of growth endogenously. It aims to better link the regional economy, social dynamics, and land-use change. These development probabilities are useful in multiple ways within a land-use change models. Together with the traditional drivers of land use change these probabilities can be used to more realistically calculate overall development probabilities.

The third and the final essay evaluate the performance of the full information feasible generalized spatial three stages least square (FGS3SLS) estimator (Kelejian & Prucha, 2004) used to estimate the model presented in the second essay. This estimator is consistent and asymptotically normal but its small sample properties are less known. In absence of very large samples as is the case in most applied works, it is difficult to interpret the results with confidence based on asymptotic results only. One alternative to employ in a situation such as this is to use finite sample approximations or asymptotic expansions. However, these approximations tend to be very complex, the results difficult to interpret and the computations very advanced. In contrast, the method of Monte Carlo replaces the skills needed in asymptotic approximations by relying on computational power of computers. Here, the properties of the parameters of interest are studied through a series of stochastic simulations and their statistics are analyzed (Davidson & MacKinnon, 1993). These Monte Carlo simulations provide confidence in the results of the social dynamics model.

4. Expected Contributions

The dissertation aims to provide answers to the question of 'how much to grow and where to grow' in the context of land-use change, highlighting the impact of social dynamics on land-use change and the interdependence between the two. Regional forecasting of housing demand based on macro-level models often ignores the micro-level social dynamics (e.g. intra-regional migration, urban decline and sprawl) that are an important driver for new housing construction; on the other hand, micro-level drivers alone cannot reflect the economy-wide impacts on housing demand. Modeling these interdependencies provides a richer interpretation of the underlying social, economic and spatial dynamics driving land-use change in regions than is currently available. Thus, it bridges the gap between regional economic models, social dynamics and land use change to provide planners and policymakers with a more substantial knowledge base on which to deliberate about the region and its future.

CHAPTER 2: INCORPORATING SOCIAL DYNAMICS INTO MODELS OF LAND-USE CHANGE

1. Introduction

Simulations of future regional land-use change help planners and policymakers understand how alternative public policy and investment choices will play out in the future, especially in conjunction with different economic and demographic trends (Pallathucheril & Deal, 2005). These simulations are a critical part of scenarios analysis, which records not just the outcome but also how the future unfolds: the factors that drive the outcome and the process through which the factors are linked to changes over space and time.

Scenarios play a different role in planning than was envisaged for alternatives in the rational comprehensive model of planning. Rather than choosing a particular outcome as the desired future, decision-makers refer to an entire set of scenarios that have been developed in advance, and the knowledge of the link between actions and consequences embedded in this set, to plan for the future and to make choices as they encounter particular circumstances in the future.

According to Steinitz (1990):

> A scenario-based approach to land-use planning offers several advantages. First, a scenario that describes the future using a multivariate approach is not only able to consider the implications of policy choices, but also the inter-relationships between possible actions. Second, a set of scenarios designed to bracket a set of alternative outcomes can include differing viewpoints. This can encourage a

> diverse set of opinions within the planning process. Finally, scenario futures are also valuable in helping to manage uncertainty and risk. (p. 136)

In developing and working with scenarios, planners are required to perform several types of analyses. They must understand and explain how land-use currently changes and use that knowledge to predict future change. To make better sense of a set of scenarios, they must compute metrics that allow comparisons among scenarios. These metrics measure the state of various social, economic, and environmental attributes.

Models of land-use change can be designed to do much of the above. *Explanations* of patterns of land-use change involve understanding various factors, their inter-relationships and the causal mechanisms that have brought about the perceived change directly or indirectly over a period. *Predictions* of change are based on the interaction between the agents and their environment or changes in the behavior of agents or the environment. Explanations and predictions have both spatial and temporal dimensions.

In addition to various economic factors, regional land-use change is shaped by social dynamics of income inequality, racial segregation, housing abandonment and intra-regional migration. These dynamics have been extensively studied in literature and the collective impact on urban change is well known. Sampson, Morenoff and Gannon-Rowley (2002) present a detailed literature review on this issue. However, few applied models of land use change explicitly account for these dynamics. Their absence from the land-use change models can possibly be explained with lack of spatial data and methodological advances in handling complex spatial relationships. This has led to gaps between model forecasts and observed patterns of land use change. How might social dynamics be incorporated in models of future land-use change?

This chapter presents a causal mechanism through which these dynamics play out and drive land-use change in the St. Louis region. These dynamics are too complex to be translated into a modeling framework and hence a reduced form approach is suggested where only certain variables relevant to modeling land-use change are further analyzed. Their inter-relationships are studied through exploratory spatial data analysis using data from St. Louis region. This exploratory exercise does not test individual hypotheses presented in the causal map, but assesses wheter the outcomes predicted by these dynamics corroborate the observed reality. This exercise suggests that these social dynamics can be modeled using the reduced set of variables.

This rest of the chapter is structured as follows. Section 2 provides an overview of the models of land-use change. Section 3 presents the gap between model forecasts and observed patterns of land-use change. These gaps are attributed to the failure to address the underlying social dynamics in the region. Section 4 creates a case for modeling social dynamics and reviews the literature on social dynamics, its incorporation in land-use change models and methodological advancements to explicitly handle spatial processes. A framework to model these dynamics using causal maps is presented in Section 5. Section 6 explores some of outcomes as observed in St. Louis region and finds them confirming to these social dynamics. The chapter concludes by re-emphasizing the role of social dynamics to inform the land-use change models along with a short discussion on unresolved issues.

2. Models of Land-use Change

The models of land use change can be broadly characterized into two categories: theoretically explicit models and simulation based models. In presence of theoretical knowledge, the behavior of the system is modeled using explicit mathematical equations to representing the behavior of the system (Batty, 1976). However, the land use change does not lend itself to an explicit theoretical framework often due to the absence of a strict set of assumptions required for a theoretical framework, lack of understanding of the land use changes processes, or lack of integrated theory to cover all aspects and interrelationship between involved processes. Consequently, they can only be used to perform a limited set of thought experiments focusing on small number of components in isolation of other dynamics occurring around them. Thus, the theoretical models have limited use to planners who need to explain land-use change, predict future, and perform impact assessment and scenario analysis across a multitude of issues at the same time. The complexity of real life problems faced by planners needs models which can help them answer relevant questions even at the cost of the analytical rigor of theoretical models.

The second category of models is the simulation model which is used when the complexity of the system does not lend itself to be modeled via direct analytic approach (Wilson, 1974; Batty, 1976), or when probabilistic determinants of change are used. Simulation based models are based on statistics, econometrics, spatial interactions, optimization, neural networks, genetic algorithms, etc. or some combination of these. Wegner (1986) defines integrated models as those where different components of the spatial system namely environment, society, economics etc. interact with each other and in relation with the land-use and its changes. The modular nature of some of the integrated models allows connecting with other models with relative ease expanding the scope of the processes they are trying to model. This connection entails either simultaneous or iterative use of one component's output as an input for other model and vice versa. Briassoulis (2000) provides a detailed argument in favor of integrated simulation models as a tool to predict land use change and perform scenario planning.

Some of the examples of integrated simulation models are CATLAS (Anas, 1982, 1983), IRPUD (Wegener, 1982, 1985, 1986a, 1986b, 1994), TRANUS (Barra, 1989, 2001), MEPLAN (Echenique et al., 1990), ITLUP (Putman, 1983, 1991), Cellular Automata (White & Engelen, 1994), CUFM (Landis, 1994, 1995), UrbanSim (Wadell, 2002), CLUE-CR (Veldcamp & Fresco, 1996), IMPEL (Rounsevell, 1999), and LEAM (Deal & Pallathucheril, 2008).

3. Model Forecasts and Observed Phenomena

Models that produce land-use simulations typically have two components: one directed at estimating demand for new housing; a second directed at changing land use across space to meet this demand. Wegener (1994), Briassoulis (2000), U.S. EPA (2000), Parker, Manson, Janssen, Hoffman, and Deadman (2003), Brail and Klosterman (2001), Agarwal, Green, Grove, Evans, and Schweik (2002) among others provide a detailed survey of models of land-use change. These models usually rely on exogenously generated control totals for growth, or use a regional economic sub-model to derive them. In either case, the projected demand is not affected by the spatial pattern of urban structure. The 'control totals' determine the amount of growth in the region as a function of net change in number of households. The new households are then located in space based on biophysical (land forms, natural resources, soil types, etc.), and socio economic drivers (amenities, employment centers, congestion, social, economic, political and institutional factors) of land-use change (Briassoulis, 2000).

The sequential method of spatial allocation of demand based on exogenously determined population and employment projections ignores the impact of intra-regional migration, social inequalities and urban decline on growth projections. This section analyzes some of the differences between model prediction and observed phenomena. The differences are generalized under issues with prediction of the magnitude of demand, location of growth, location of housing abandonment and the lack of interaction between these variables.

a) Predicting the magnitude of demand: In most models of land-use change, population change drives the amount of new construction in the region. The change in population is proportionately converted into the amount of new housing units with a correction for household size with some correction for vacancy. These housing units are then used as macro totals and distributed in space. In the case of St. Louis region, the population increased by approximately 96,000 between 1990 and 2000 leading to a corresponding increase of 66,000 households. Considering one housing unit per household and adjusting for vacancy which decreased by four thousand, one would expect an additional demand for about 62,000 new housing units. However, there were 144,000 new units constructed in the region and the housing stock grew only by 60,000. These figures (Table 2.1) in effect suggest that 84,000 units were abandoned, a phenomenon ignored in the current methodology.

	1990	2000	2000-1990
Population	2,444,102	2,540,138	96,036
No of Households	923,642	990,249	66,607
Vacant Units	81,279	76,975	-4,304
Expected New Construction			62,607
Housing Stock	1,006,012	1,066,358	60,346
Actual New Construction			144,818
Abandoned Units			84,472

Table 2.1: Changes in population and housing units in the St. Louis region between 1990 and 2000. Source: Bureau of Census, Geolytics database

Summarizing, a population increase of 96,000 led to new construction of 144,000 and abandonment of 84,000 housing units. Thus, we see that predicting the magnitude of growth directly from population change alone can be grossly misleading. The contribution of housing abandonment and spatial patterns of socio-economic characteristics have a greater role to play to determine the amount of new construction as shown later in the chapter.

b.) Predicting the location of growth: Contrary to the observed trends of decline in the core and growth in regional periphery, the traditional drivers of land-use change models like proximity to employment and amenities predict high degree of development in the urban core (Figure 2.1).

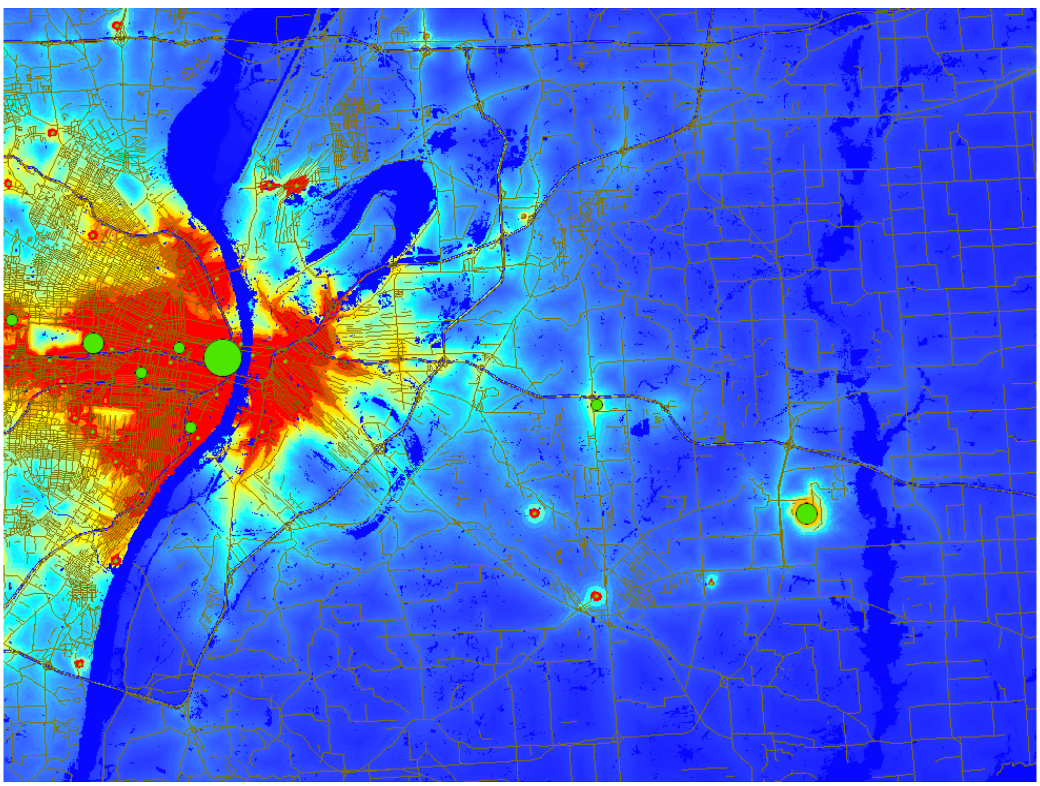


Figure 2.1: Development attractiveness computed from travel time to employment centers in St. Louis region. Areas marked in red are more attractive than those in blue due to shorter travel time to employment centers shown in circles.

c.) Predicting decline: Conventional methods of modeling land-use change either ignore or fail to capture urban decline as defined by vacancy, housing abandonment of existing housing stock together with lack of new construction.

d.) Lack of interaction between magnitude of growth, location of growth and the socio-economic fabric of the region: Land-use change in the St. Louis region shows a trend of movement further away from the core. Economic mobility and poor quality of life in the urban core creates demand for peripheral location of new housing units. The houses left vacant by households who can afford to move to the periphery in search of better quality of life are occupied by the not-so-wealthy living in the core who cannot afford to move. This process is commonly referred to as filtering and is discussed at length in Baer and Williamson (1988). This mobility chain is a continuous process independent of migration into the region that leaves behind even more abandoned units in the most distressed areas and spurs new housing construction in the peripheries. These abandoned units add to distress promoting further outward mobility and spread of urban distress. Thus, the amount of new housing construction is a function of housing abandonment which is directly related to the socio-economic fabric of the region. This implies that estimating magnitude and location of growth are interconnected and cannot be done independently or sequentially. Some models use housing prices as a proxy for social dynamics however, the housing prices are an outcome of social dynamics and are also affected by macro forces beyond the region. Cameron and Ian (2006) find that they "may not capture the full effects of environmental disamenities". However, the land-use change models have failed to address this phenomenon.

Spread of urban decline characterized by abandoned housing units, poverty, poor school quality, etc. from one area to neighboring areas has occurred in many regions. As a result there is a spatial mismatch between where the development actually occurs and where it should take place from the perspective of travel time, cost, accessibility, etc. There have been limited attempts to model these social dynamics in the context of land-use modeling.

4. Social Dynamics as a Driver of Land-use Change

Sampson et al. (2002) summarize results from 40 peer reviewed articles on the study of social processes in American cities. They find considerable socioeconomic inequality and racial segregation among neighborhoods. These studies found a strong correlation at neighborhood level between a number of social problems like crime, school dropout, well-being, etc. The results have been robust at different levels of geographical analysis like community areas and census tracts. The spatial concentrations of poor and affluent neighborhoods were found to be increasing over time. Given, the presence of such strong social forces in shaping land-use change, the need for modeling their dynamics for predicting land-use change is of immense importance.

Briassoulis (2000) reviewed evolution of social models and their relevance to land-use change modeling. She finds that the overall ability of social models to inform land-use change models is very limited. The reviewed models of social dynamics do not explicitly acknowledge the spatial and temporal dimension of social interactions. Even when they do so, they do not refer to the actual land-use or its changes over which the social agents interact with themselves or with their spatial environment. Further, some of these are very specifically focused on a particular socio-political and cultural setting and their applications to other regions would violate some of their basic assumptions. There has been some theoretical support to modeling efforts from the urban economic theory but its impact has been limited in terms of application to the spatial diversity of the social and cultural nature of land use change models that are relevant to planners.

There is an increasing trend of incorporating spatial process in social models. This has followed form methodological advancement in spatial statistics and econometrics to explicitly handle space. This was supported by increased availability of spatial data in public domain and available computing capabilities. Spatial externalities are now started playing a central role in the recent emergence of “spatial thinking” in the mainstream social sciences (Goodchild, Anselin, Appelbaum, & Harthorn, 2000). Akerlof (1997) shows the increase in use of models of social interaction between actors and environment. The neighborhood ecology models and its variants ranging from Case (1992) about neighborhood influence and technology change, Kelejain and Robinson (1992) about police expenditures at county level, Case, Rosen, and Hines (1993) on budget spillovers, Boarnet (1994) on population and employment growth within a metropolitan area. Spatial models of neighborhood effects have also started surfacing in demography and criminology including formal notions of spatial spillovers and dependences (Abbot, 1997; Sampson et al., 2002; Messner & Anselin, 2004).

Given the role of social dynamics in shaping land-use change, availability of spatial data and advancements in methodology to deal with spatial processes, it is now possible to incorporate social dynamics to inform land-use change. However, before doing so, it is important to create a framework that will inform this modeling exercise.

5. Framework for Modeling Social Dynamics

A report from Focus St. Louis (2001) attributes urban distress to intra-regional migration, housing segregation, economic and educational disparities. These social factors sustain urban distress in some areas (primarily the core) and promote growth in others (primarily in the periphery). Lee and Leigh (2007) find similar trends in Atlanta, Cleveland, Philadelphia, and Portland, using longitudinal census data from 1970 to 2000.

A framework to explain social dynamics in reference to its relation with land-use change based on the literature surveyed in the previous section is presented here. The arguments are first constructed at micro level of households explaining the wealth dynamics and housing segregation, and then expanded to explain the resultant spatio-temporal feature of socio-economic disparities at the regional level. Some of these ideas presented in this section were initiated in Sarraf and Bendor (2005).

Wealth dynamics

Carter, Schill, and Wachter (1998) find that proportion of families in a tract with less than average income is influenced by the income in the previous decade, condition of the housing stock and changes in the housing stock during the last decade. Using their study and findings from the Focus St. Louis (2001), Figure 2.2 develops a causal mechanism of wealth dynamics at household level. Wealthy households choose to live in neighborhoods that provide better quality of life in terms of infrastructure and access to opportunities. Better infrastructure and higher ownership rates leads to faster appreciation of property values. Higher property values mean higher resources for the school districts. Better schools together with wealthier parents provide better opportunities for higher education, leading to a higher income for the next generation. Vartanian (1999) also shows the influence of childhood neighborhood conditions on the economic well-being including employment prospects of adults. These also come with better economic opportunities in terms of transportation, and access to capital. Thus, the benefits of living in a wealthier neighborhood are high and households with high income tend to gravitate towards neighborhoods which are occupied by other wealthy households.

Households with sufficient economic resources start moving out of poor neighborhoods in search of better prospects. Continuous exodus from poor neighborhoods leads to decreasing tax revenues, poor maintenance of infrastructure, deteriorating school quality, etc. Neighborhood conditions start affecting school quality and dropout rates (Clark, 1992). This further accentuates out-migration of wealthier households and fall in property prices in these neighborhoods. The result is increased vacancy, abandoned units and disincentive for new construction. Loss of tax revenues leads the communities to adopt higher tax rates repelling further investment in the neighborhood.

Each household within a rich neighborhood has positive spillover over others and those in poor neighborhoods have negative spillovers. Collectively the households in rich neighborhoods enjoy better schools, better infrastructure and are able to steer housing market and regional and local policies in their favor. The average wealth of households in richer neighborhoods increases faster than if the households were homogenously dispersed in the region. This becomes further attractor for richer households to agglomerate

The resulting income segregation is further strengthened by home owner associations through exclusionary zoning laws to maintain the lower bound on average household income. Over time neighborhoods tend to get differentiated by income and over many years, clear segregation patterns emerge. However, the neighborhood boundaries are not rigid and sealed. The interactions among adjoining neighborhoods lead to externalities.

Relationship between neighborhood character and crime has been extensively studied (Savoie, 2008; Wallace, Wisener, & Collins, 2006). Neighborhoods that have higher vacancy and housing abandonment have higher crime or are perceived to have higher levels of crime. Spelman (1993) found that crime in blocks with abandoned houses were twice as high as similar ones without abandoned units. This has huge negative externality on surrounding neighborhoods and induce decline there. Just like individual households of certain income groups cluster, neighborhoods also starts clustering. A richer neighborhood may exert positive externality over the adjoining poorer neighborhoods or may be faced with negative externality. The strength and direction of this externality determines the fate of the surrounding neighborhoods. These dynamics eventually result in much stronger and sustained segregation in terms of the desirability for new construction of housing.

Racial Segregation

The impact of race on neighborhood dynamics is similar to that of impact of wealth and income. This is because the African-American households constitute a huge part of low income households. Figure 2.3 elaborates on the impact of race on neighborhood dynamics and regional land-use change. When the percentage of African American households reaches a certain threshold, it is perceived as a negative signal by other households. This perception of the so called 'tipping point' or thresholds (Quercia & Galster, 2000) leads to exodus of white households, further strengthening the signal. This exodus is dealt in detail in Harris (1997) titled The Flight of Whites: A Multilevel Analysis of why Whites Move. It leads to exodus, increased vacancy and fall in property values. The lower property price attracts other of lower income and or African American households. If the neighborhood is very old, and many household have moved out, the infrastructure starts deteriorating creating further disincentives to move in. This process is spatially diffusive. At the regional level, it creates segregated neighborhoods and the children who grow up in such neighborhood tend to have less racial tolerance. When they grow up, they tend to choose more segregated locations, further strengthening segregation. New migrants into the region are 'steered' by real estate agents into segregated neighborhoods. High degree of fragmentation of local governments facilitates this segregation through discriminatory local policies leading to intra-regional mobility and thereby creating sprawl.

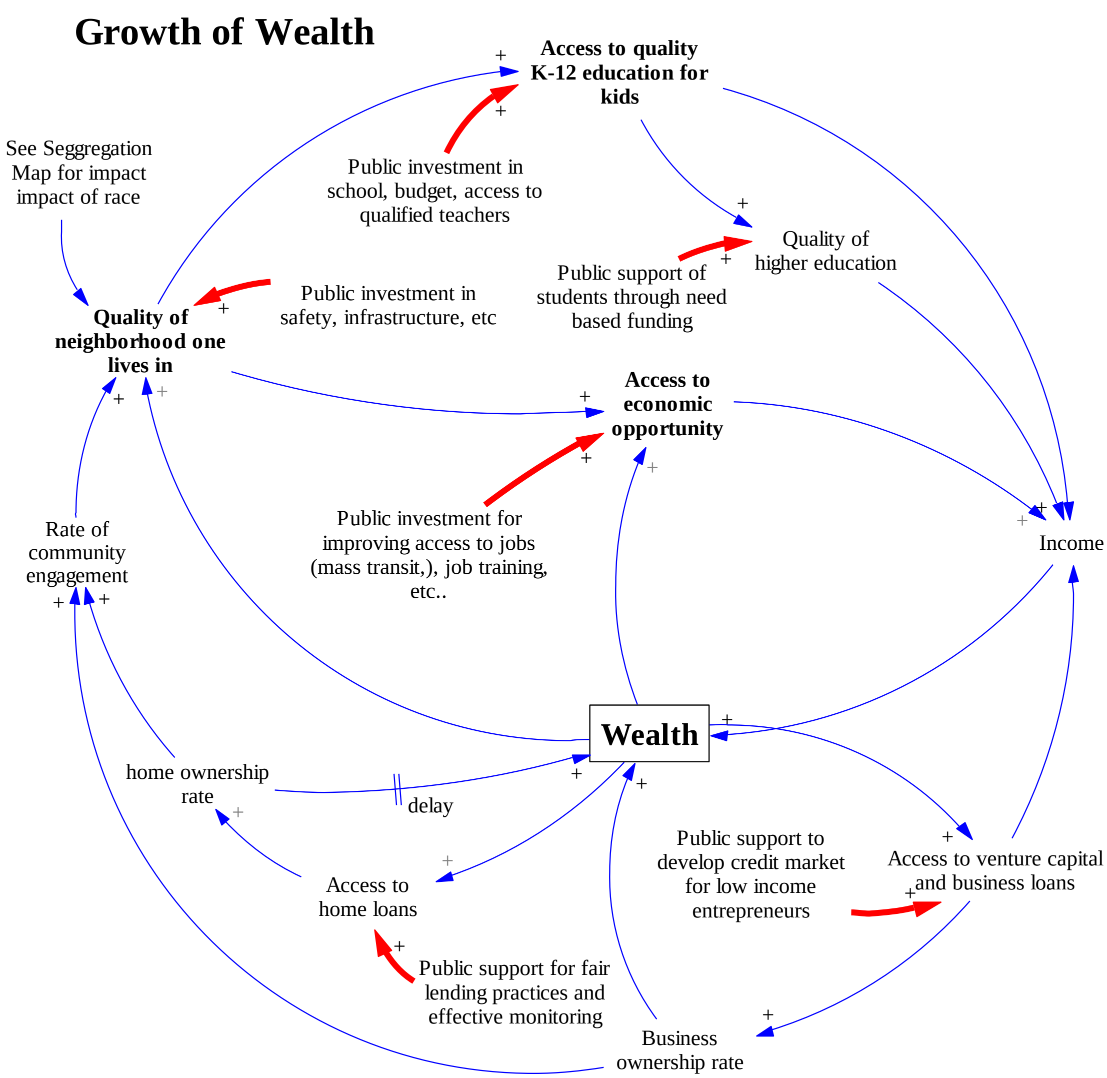


Figure 2.2: Wealth accumulation, temporal persistence and spatial clustering of poverty

Regional impacts

The interconnectedness between regional disparities in income, education and housing segregation based on race and income is shown in Figure 2.4. Economic disparities lead to housing segregation due to differential location decision making process. Housing segregation leads to education disparity characterized by access to schools, job oriented programs and ability to pursue higher education. Differential access to opportunities

Neighborhood Segregation

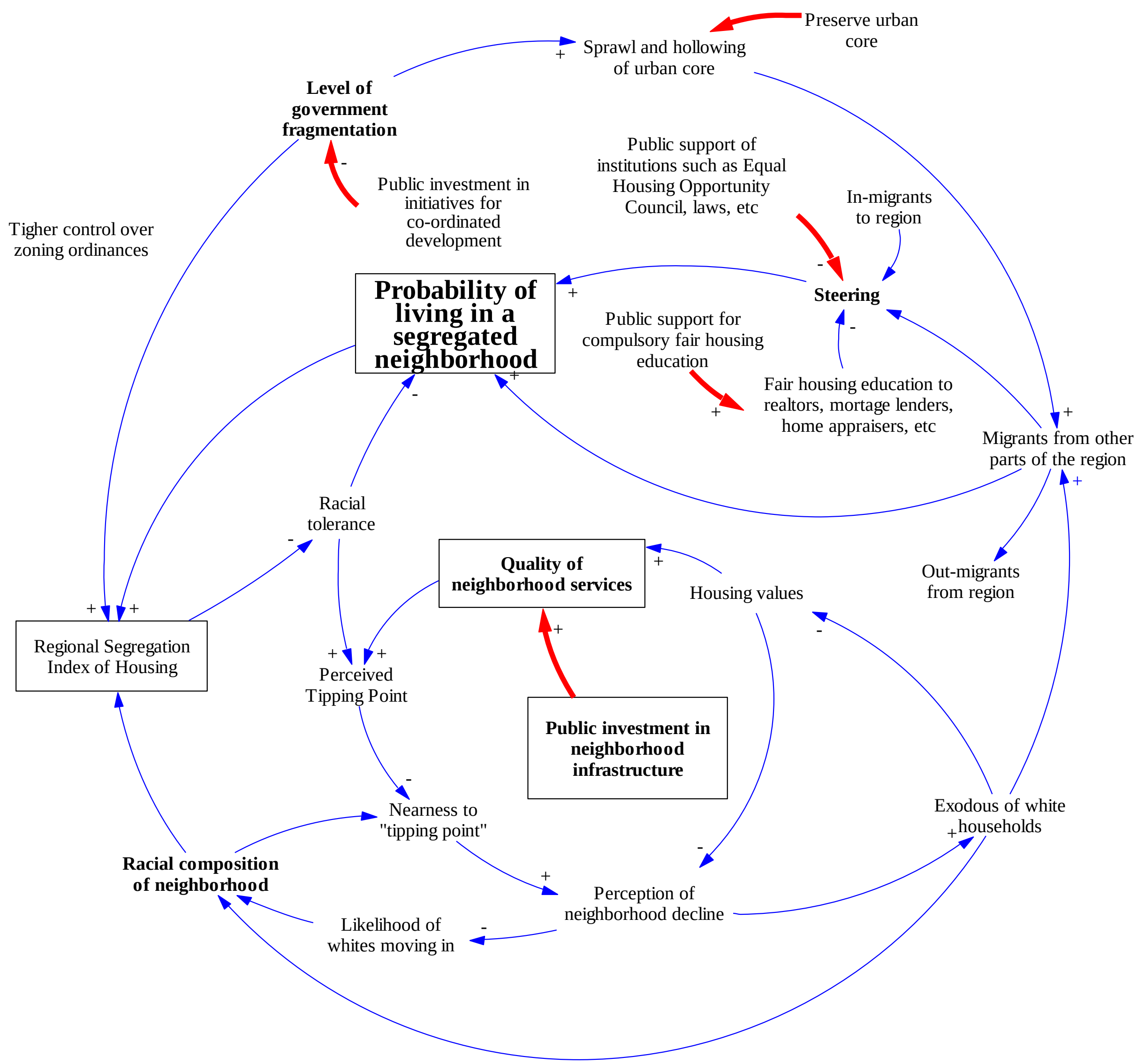


Figure 2.3: Neighborhood segregation, vacancy, urban decline and location of new construction

further enhances the economic disparity. All these together constitute the regional socio-economic disparities that affect the land-use change over time.

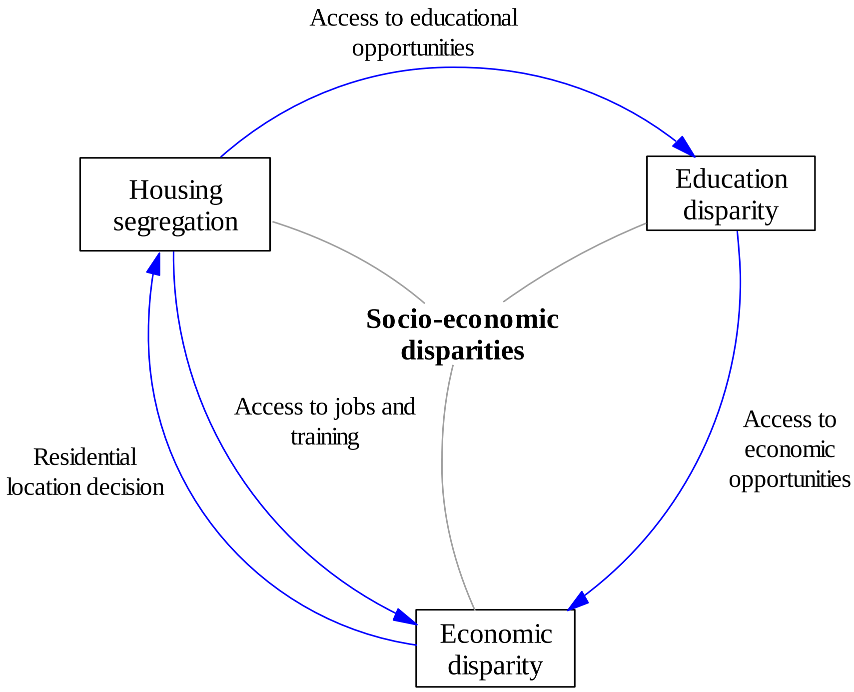


Figure 2.4: Interrelation between different regional inequalities

It is difficult to incorporate all the complexities of socio-economic dynamics into land-use change models. Only those components that can endogenously be determined within the model are useful predictor of future land-use and can be shaped by the evolving pattern of urban growth. This is important to complete the feedback loop between social dynamics and land-use change. There are other important variables e.g. racial composition, schools quality, etc. which are critical to explain social dynamics. However, their future values are difficult to predict in models that do not distinguish between residential land-use based on racial composition of households. The next section presents exploratory data analysis of the observed social dynamics in the St. Louis region between 1990 and 2000 for endogenously determined variables namely average household income, number of households, new construction and vacancy. The data analysis provides empirical support to the outcomes predicted by causal maps.

6. Social Dynamics and Land-use Change in St. Louis

The relationship between the key variables affecting land-use change is studied using statistical exploratory data analysis (EDA) and exploratory spatial data analysis (ESDA) without assuming *a priori* structure of relationship among them. It is important to not impose prior structure in absence of a strong existing theory to explore the nature of relationships. EDA includes the use of descriptive statistical analysis (Tukey, 1977) and dynamically linked statistical graphics to interact with different views of the data (Cleveland, 1993; Buja, Cook, & Swayne, 1996). ESDA consists of methods to describe and visualize spatial data including identification of outliers, spatial autocorrelation, spatial regimes and other forms of spatial heterogeneity (Messner et al., 1999; Anselin, 1998, 1999; Haining, 1990; Bailey & Gatrell, 1995). It includes but is not limited to univariate and bivariate global and local spatial autocorrelation analysis (Anselin, 1995, 1996), multivariate exploratory space-time analysis (Anselin, Syabri, & Kho, 2006) including identification of spatial cluster and spatial outliers.

The St. Louis region under study consists of five counties (Clinton, Jersey, Madison, Monroe and St. Clair) in the state of Illinois and five counties (City of St. Louis, St. Louis, St. Charles, Jefferson and Franklin) in the state of Missouri (Figure 2.5). The region approximately measures 120 miles in east west direction and about 90 miles in the north south direction. The Neighborhood Change Database (Tatian, 2002) provides census data reconfigured to make the aggregation consistent across 1990 and 2000 census boundaries and is available at the tract level. There are 515 tracts in the region, out of which three tracts were removed from the analysis. They had witnessed serious flood in early 1990's leading to large scale migration of households and their inclusion would have biased the analysis as these migrations were not a results of regional social dynamics. The analysis was carried out using Geoda (Anselin, Syabri, & Kho, 2006) and the R software for statistical analysis (R Development Core Team, 2004) with the spatial package spdep (Bivand et al., 2006).

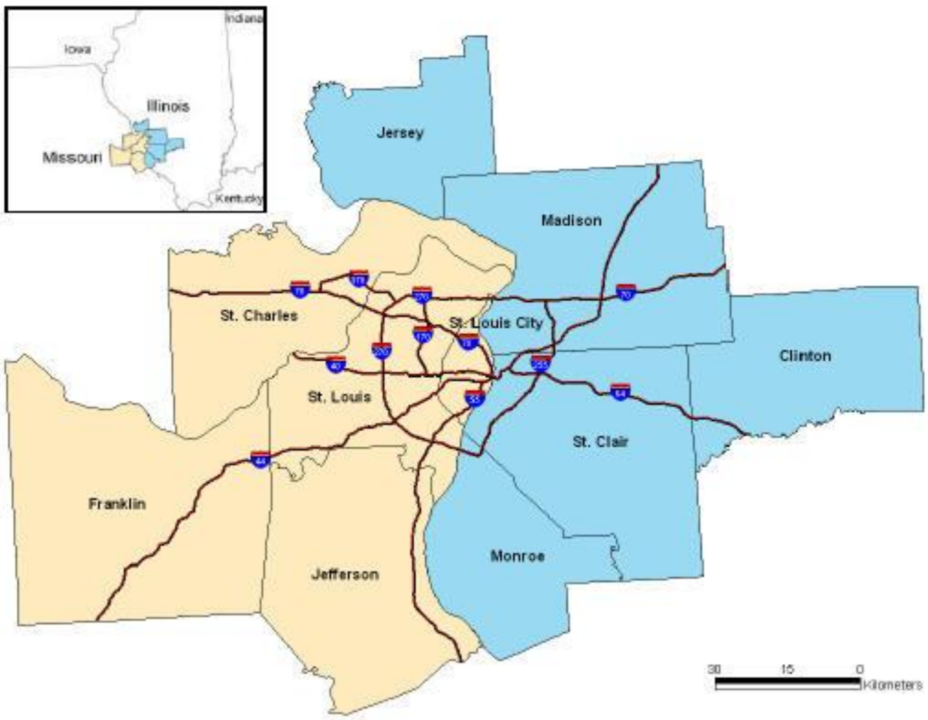


Figure 2.5: The ten counties in the St. Louis region

Income

The income inequality in the region is large and spatially persistent over time. The census tracts in the top 2% average household income group in 1990 continued to remain among the top 2% in 2000. Some tracts moved from the top quartile group in 1990 to second quartile in 2000 while some moved from the second quartile in 1990 to being among the top quartile in 20000. In the similar manner, some tracts have witnessed downward mobility from the third quartile in 1990 to fourth quartile in 2000 and vice versa. However, none of the tracts have dramatically moved between 1^{st} and 3^{rd} or 2^{nd} and 4^{th} quartile. This shows that there is both upward and downward mobility of average household income in tracts but there seems to be a lot of inertia with which the changes are happening, and that the transitions are smooth. The correlation between average household incomes in the two periods was 0.94. In absolute terms, the richer tracts saw maximum increase in their average household income, but in terms of percentage increase in income over 1990, the distribution was more even (Figure 2.6).

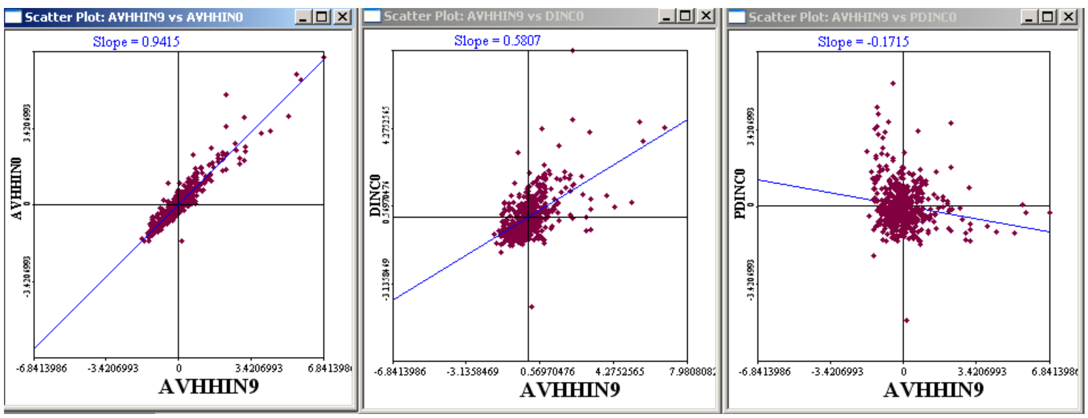


Figure 2.6. From left to right: Standardized scatter plots between Average household income (AvHHIn9) on the X axis and a) Average household income on 2000, b) Change of AvHHInc between 1990 and 2000, c) Percentage change in AvHHInc between 1990 and 2000 on Y axis.

Spatially, the richer and poor tracts witnessed more clustering with statistically significant spatial clusters expanding to include more neighboring tracts through a process of diffusion. The tracts with upwards mobility were mostly located in the regional periphery, while the tracts which witnessed downward mobility of income were located closer to the core. LISA (Local Indicators of Spatial Autocorrelation) maps (Figure 2.7) show a statistically significant cluster of tracts with high average income surrounded by other tracts with high income west of the city of St. Louis and in between interstates I-44 and I-70. Clusters of tracts with low household income surrounded by other tracts with low income are present in the tracts in the city of St. Louis and in the neighboring tracts in Illinois across the river. Both of these clusters have grown in geographical spread from 1970 to 2000. Excluding the richest tracts in year 2000, the spatial autocorrelation coefficient (Morans I) have remain very stable during this period.

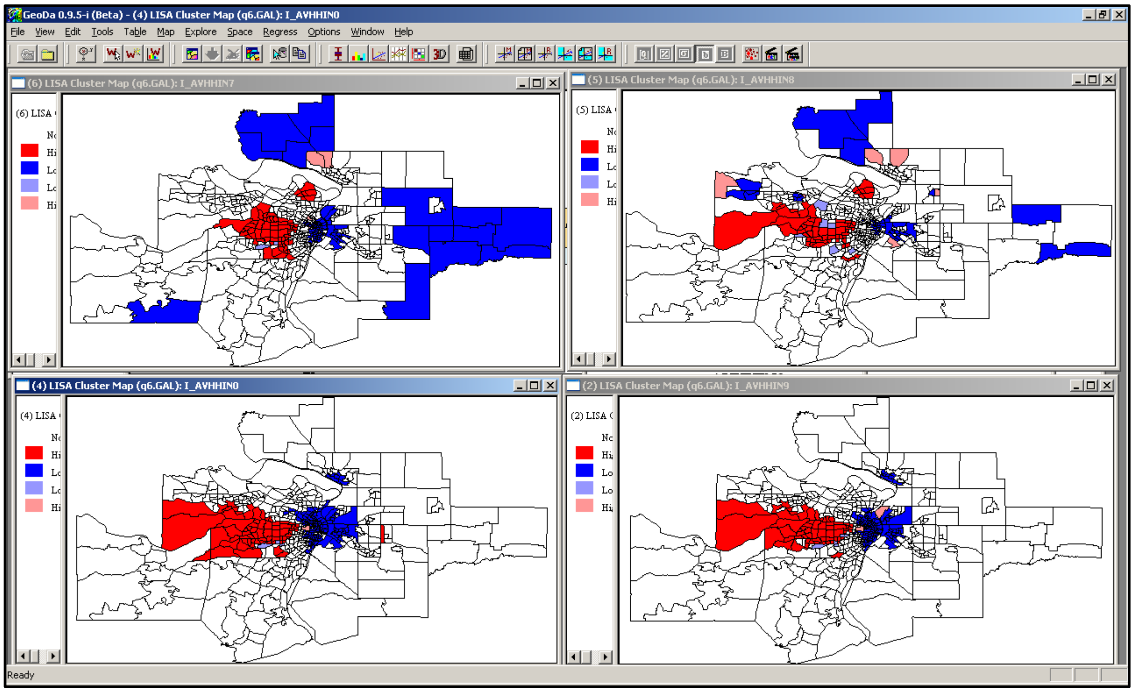


Figure 2.7: Tracts in red fill show statistically significant clusters with high average household income surrounded by other tracts with high income. Tracts in blue fill shows tracts with low average household income surrounded by other tracts with low income. Both these clusters have grown in geographical spread from 1970 to 2000 (clockwise from top, LISA maps for 1970, 1980, 1990 and 2000)

Households

Tracts that saw in-migration of households between 1990 and 2000 were located outside the regional core (Figure 2.8). Most of these tracts had witnessed gains during 1980 to 1990 as well. Tracts that lost households during 1980-1990 and 1990-2000 were concentrated in the regional core. This shows a long and steady process of decline of the urban core. Surprisingly, some tracts that gained household during 1980's lost many households during the 1990's and all of them were located North West of the regional core. Further, there is a spatial cluster of tract that gained households in 1980's and lost them 1990's returning to their 1980 level on the average. These were the tracts where the construction of new housing units was small in 1980's and much below average in 1990's. They did not have any other remarkable feature indicating such a decline in terms of income, vacancy or school quality.

Tracts gaining households in both decades witnessed maximum new construction. Households were moving into the regional periphery while the tracts where they came from, located primarily in the core saw housing abandonment, loss of households and no new construction. Further, all the new construction was occurring in the periphery where densities were low. Low density was important for households to move in, but not sufficient, i.e., most of the gainers were low density tracts, but not all low density tracts gained. Similarly, most out-migration was happening from tracts that had higher densities. The tracts that gained households were the tracts with higher average household income in both decades suggesting that people with higher income had greater mobility and attracted other similar households.

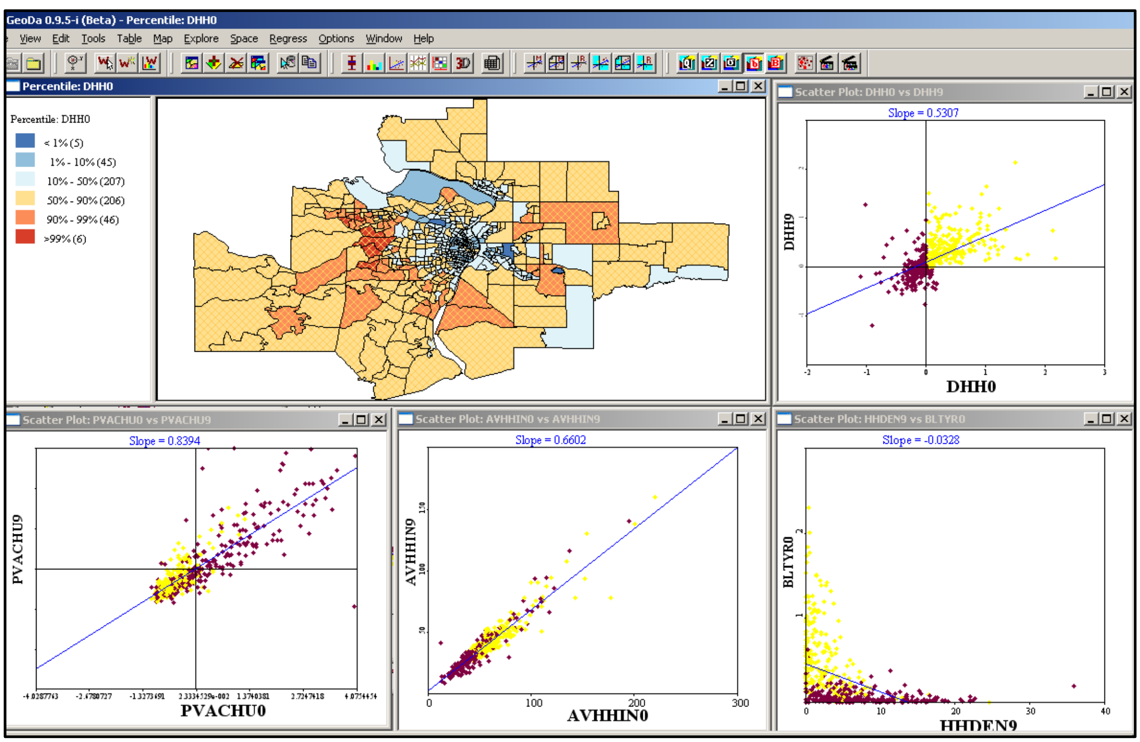


Figure 2.8: Tracts that gained households (highlighted in yellow) during the last two decades had very few vacancy, high average household income and low densities. There were all present in the regional periphery.

Vacancy

The distribution of vacancy among census tracts is skewed with only few tracts with high vacancy clustered in the regional core (Figure 2.9). These tracts also have low average household income, fewer new constructions, and are surrounded by other tracts with high vacancy. They have witnessed large out-migration over the years. Tracts with low vacancy in 1980's had most of new construction in 1990s, further bringing down their vacancy rates. They were relatively richer tracts. A small number of tracts that have moved from high vacancy to low vacancy between 1990 and 2000 witnessed more immigration. Tracts which had high vacancy over the years had higher housing abandonment and lower in-migration, and fewer new constructions.

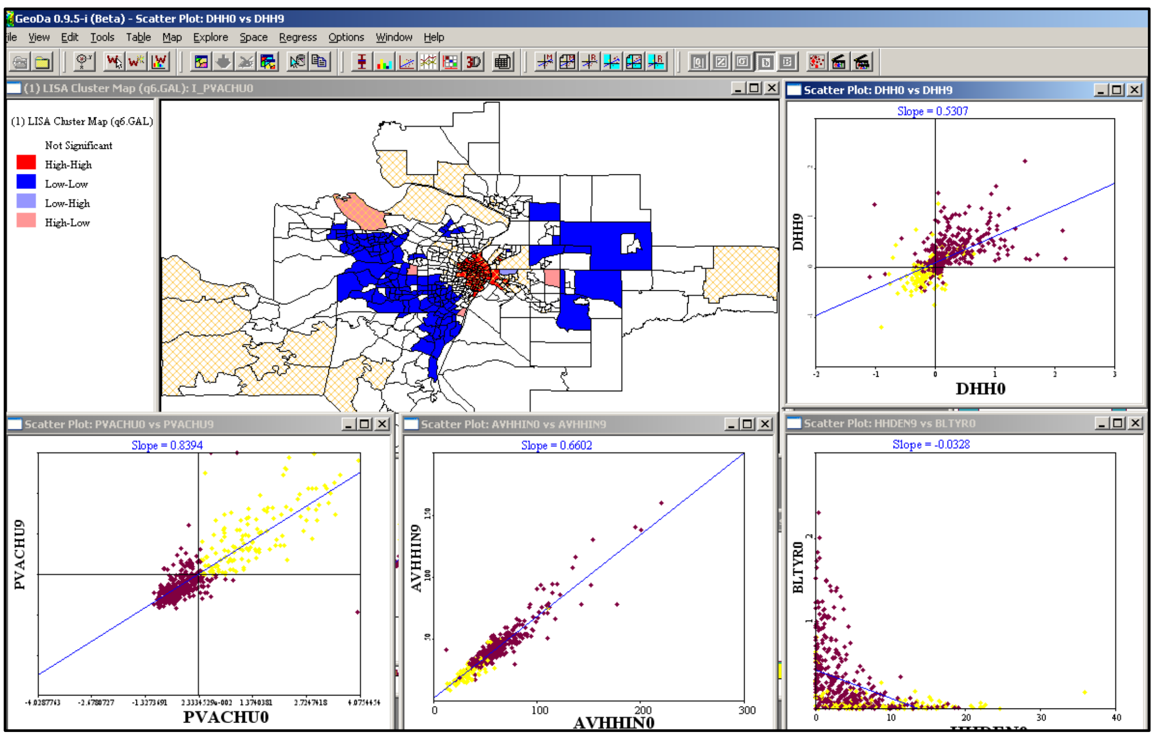


Figure 2.9: Clusters of high vacancy rates at the regional core, and that of low vacancy are located west of core. Places with high vacancy in both decades had very little new construction, were mainly poorer tracts and saw a lot of out migration of households. Bivariate LISA shows that vacancy and income were spatially clustered.

Additional graphs are provided in Appendix A.

7. Conclusion

By sequentially and independently estimating the amount of growth, the current generation of land-use change models fails to capture the full impact of social dynamics on additional new demand for housing. For example, between 1990 and 2000 in the St. Louis region, the number of households increased by 66,000 while 144,000 new housing units were constructed. This disproportionate growth in quantity of new housing is largely due to the social dynamics in the region and is not explained by economic models. Some of the other specific failures of models that fail to address social dynamics include not being able to account for intra-regional migration, hollowing out of urban core, increasing inequality and its associated impact on the total new housing construction in the region. Again in the case of St. Louis, the construction of 144,000 new units was accompanied with abandonment of 84,000 units, a phenomenon largely unrecognized in land-use change models.

As a result, the land-use change models in absence of social dynamics fail to adequately respond to redevelopment policy and its full impacts to curb urban sprawl. Though substantial progress has been made on modeling social dynamics with explicitly recognizing the role of space, the interaction between agents and environment continue to occur in abstract space rather than on a landscape on which human activities take place. Further, there is a lack of an integrated theory of interaction among social agents, their environment and the land-use where these interactions take place.

This chapter presents an integrated spatio-temporal framework to analyze the social dynamics in the region in order to better inform the models of land-use change. It develops a causal model as a first step towards creating a new driver for land-use change. The causal map provides a rich explanation of the various socio-economic attributes that has shaped the region in its current form. In the absence of an established theoretical relationship between urban distress and the social fabric, only a reduced form model structure based on the causal mechanisms seems feasible. It will be reduced in the sense that even though we know the causal linkages between variables but the exact functional form of these relationships are not known. This chapter provides descriptive statistics, maps, dynamic spatial linking and brushing techniques in interactive computing environment to perform exploratory spatial data analysis to support some of the relationships presented in the causal maps which can be used to inform the reduced form model.

There are a number of unresolved relationships that remain to be explored, race being one of the most important among them. It may be kept out of the purview of land-use change models for substantive and technical reasons. The racial composition of new households in the region is not known. To incorporate race, one will have to decompose households within a tract by racial groups and then separately model their residential choice behavior for each group. However, the residential choice behavior is not independently governed by race either. It depends on the age of the head of the householder, income group and household composition. We have very little theoretical understanding of dynamics at this level of detail and very little data to model such intricacies. The pay off to land-use change model, which is our objective, is also limited, compared to the efforts that would be needed to pursue such an endeavor. For details on race and its impact on urban structures, see papers by Brown and Chung (2006), Chen, Irwin, Jayaprakash, and Warren (2005), Charles (2003), Massey, and Denton (1993), Meyer (2000), Cutler, Glaeser, and Vigdor (1999), and Galster (1990, 1992).

CHAPTER 3: IMPLEMENTING SOCIAL DYNAMICS AS A DRIVER OF RESIDENTIAL LAND-USE CHANGE

1. Introduction

In addition to various economic factors, regional land-use change is shaped by social dynamics of income inequality, racial segregation, housing abandonment, intra-regional migration. These dynamics have been extensively studied in literature and the collective impact on urban change is well known. Sampson, Morenoff, and Gannon-Rowley (2002) present a details literature review on this issue.

The current generation of land-use change models does not explicitly model social dynamics and, hence, faces several shortcomings. (See Briassoulis [2000] for a detailed review of land-use change models). These shortcomings include (a) ignoring the impact of social dynamics (e.g. spatial distribution of income inequality) on location of new housing units, (b) ignoring the impact of housing abandonment, which increases the demand for new housing units, and only using population projections as a proxy for demand, (c) ignoring differences among types of housing units (abandoned, vacant and occupied) and only focusing on locating generic households , (d) ignoring the interdependence between magnitude of growth and location of housing units, and only carrying out these steps sequentially, and (e) ignoring the role of land-use as a driver of social dynamics . These omissions result in underestimation of new housing construction, and failure to account for sprawl, intra-regional migration, regional imbalances and social externalities. This prevents planners and policy makers from fully understanding the impacts of various public policy and investment choices.

These issues and other aspects of the problem are discussed in detail in Chapter 2 where a framework is proposed for modeling social dynamics as a driver of land-use change. This chapter describes an implementation of the proposed framework within the Land-use Evolution and impact Assessment Model (LEAM). (LEAM is briefly described in Chapter 1; for more details see Deal and Pallathucheril [2008].) As seen in Figure 3.1, LEAM uses a regional economic model to drive the demand for land over time. This demand for land is met by locating land-use change in space using a discrete choice framework that combines input from various spatial sub-models.

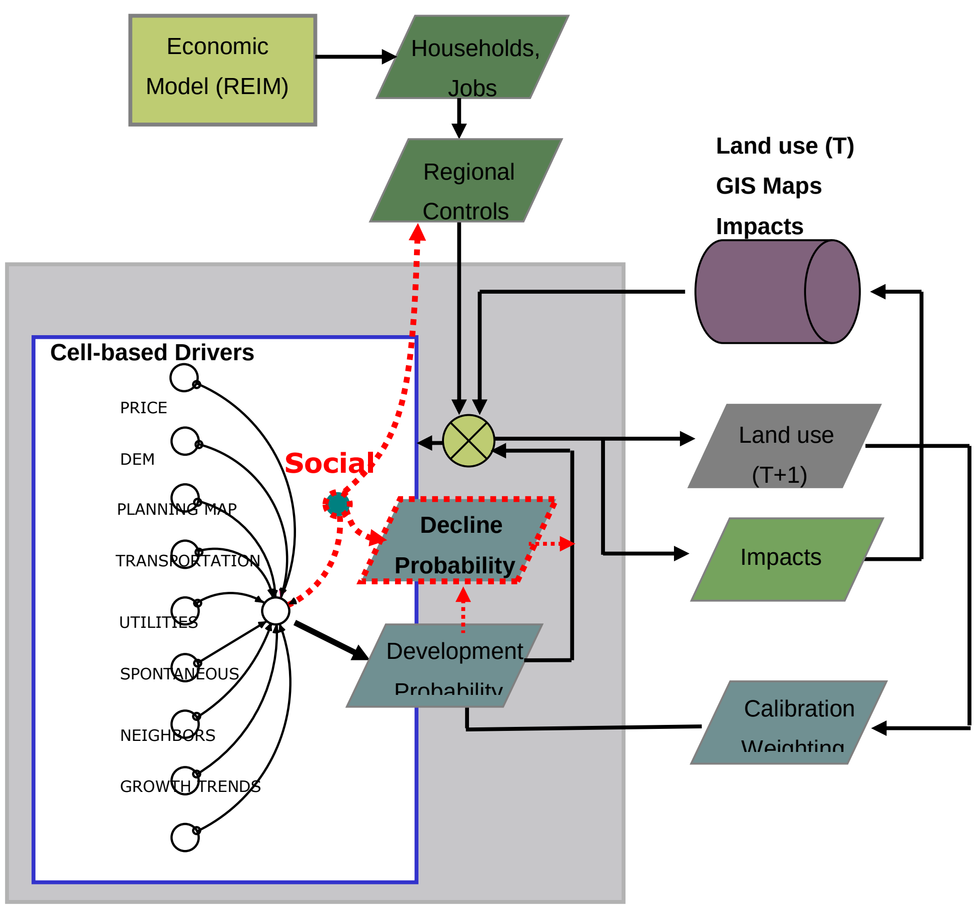


Figure 3.1: Structure of LEAM. Proposed links and components are shown in red dotted lines.

Linkages introduced in the study described in this chapter are shown in red dotted lines. A new driver is introduced that implements social dynamics and has three effects: it alters development probability in a grid cell; it computes a probability of decline; and it modifies the regional demand for residential land use. Development probabilities for new construction, which are driven by social dynamics, are computed at the census tract level and assigned to all grid cells in the tract. Probabilities are separately computed for housing abandonment and vacancy. Regional demand for new construction of housing units reflects demand due to new households as well as (a) loss of housing stock due to abandonment and transformation of land-use from residential to commercial use and (b) intra-regional migration.

This study, therefore, aims to better link the regional economy, social dynamics, and land-use change. These development probabilities are useful in multiple ways within a land-use change models. Together with the traditional drivers of land use change these probabilities can be used to more realistically calculate overall development probabilities. As a result, they can alter the location choice of households, the attractiveness of a place for new construction, and capture the negative spatio-temporal externalities that lead to decline. They also provide feedback to the other sub-models. In the transportation model, for instance, they can be used to estimate travel demand more precisely by differentiating between abandoned, vacant and occupied units.

This chapter is organized as follows. Section 2 presents a framework for modeling social dynamics, the main components of social dynamics, and their expected behavior. Section 3 describes the model specification, estimation process and presents a discussion of the results. Section 4 provides a framework for forecasting decline and development probabilities and its integration with LEAM. Section 5 concludes the chapter by summarizing the key points

2. Framework for Modeling Social Dynamics

The broad structure of the model is presented in Figure 3.2. The regional social fabric—defined by variables like school quality, crime, income, race, etc. — interacts with regional land-use (new construction, occupied, vacant and abandoned housing units). The outcomes of these interactions are constrained by regional economic parameters and vice versa.

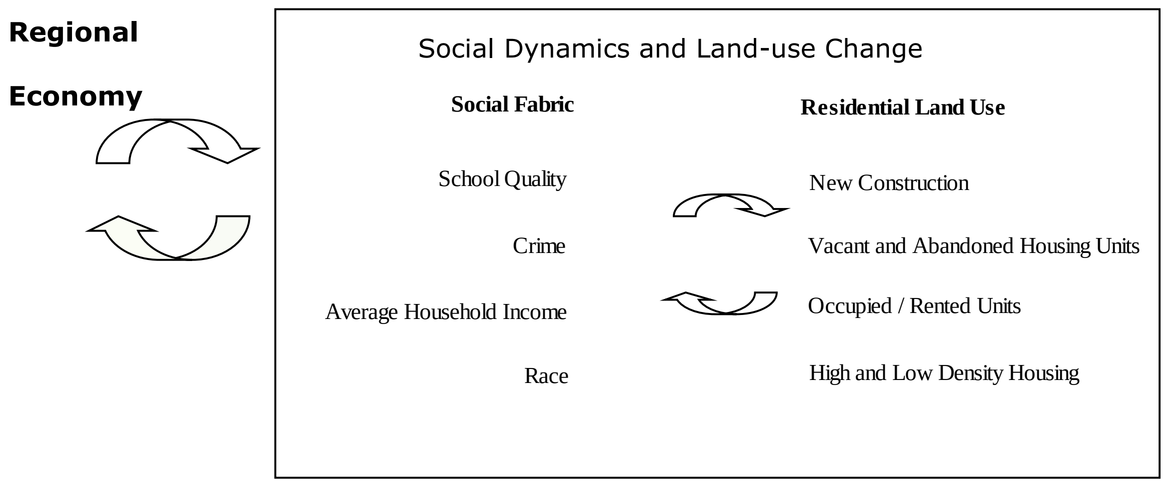


Figure 3.2: Broad Structure of the Social Dynamics Model

Figure 3.3 is a condensed causal map using discussions in Chapter 2 as a starting point. The encircled variables are those that are affected by spatial, temporal and spatio-temporal lags. Key variables, the values and spatial distribution of which are forecasted, include average household income, number of households, new construction, vacancy rate and housing abandonment. The main hypothesis is summarized as follows. Richer tracts continue to be rich over years and being surrounded by richer tracts has positive spillover effects. Tracts with more new construction normally see their average income increase. This is partly due to the fact that higher income household have more mobility. Tracts with high densities tend to particularly repel richer households and thus, their income levels decrease over time. There is sustained growth in lower density tracts. Changes in number of households in a tract are gradual and persistent like that of average household income. Households tend to move out of tracts with high housing abandonment, higher vacancies and poor schools, signaling these as indicators of distress. They also tend to move into tracts with more new construction, which occurs more in tracts with low vacancies, low housing abandonment and high income. This is because relatively poor households tend to move into houses vacated by richer households in the filtering process. Vacancy rates are lower in higher income tracts while higher in tracts with high vacancy and housing abandonment in the previous period.

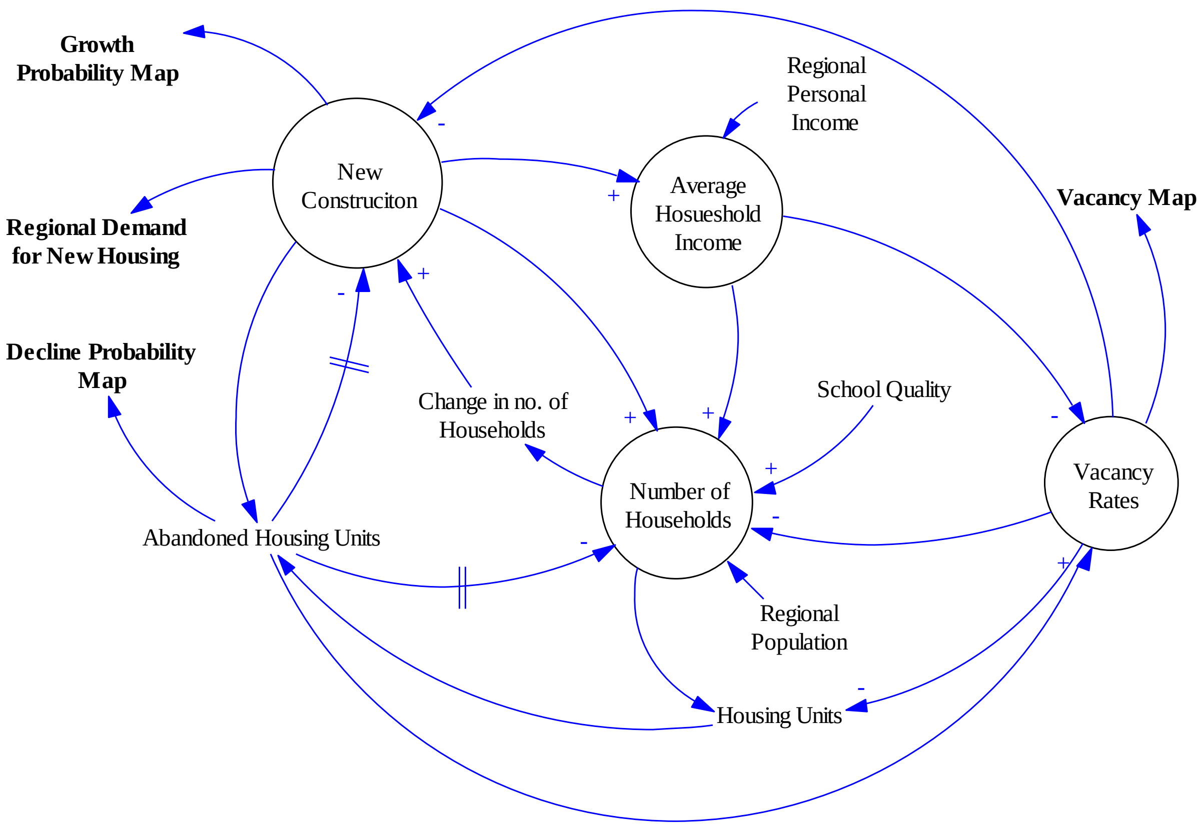


Figure 3.3: Mechanisms affecting location of new households and vacancy. Model outputs are highlighted in bold letters

Data on crime is not available at the tract level. It is either available at county level which is too large a geographic scope or at the level of places (as defined in census) which does not cover the entire region. However, the impact of crime can still be captured by spatially autocorrelated errors in the model specification, given that crime and its perception are spatially diffusive in nature.

Even though race is a very important driver for change in the neighborhood, it has not been included in this study for substantive and technical reasons. Land-use change models treat all households as homogenous and do not differentiate households by race. To incorporate race, one will have to decompose households within a tract by racial groups and then separately model the residential choice behavior for each group. But residential choice behavior is not independently governed by race. It also depends on the age of the head of the householder, income group and household composition. We have very little theoretical understanding of dynamics at this level of detail and very little data to model such intricacies. The racial composition of new households in a region is not known. But race is strongly correlated with income and some of its impact may be captured through income linkages with other social variables.

Data

Data used to estimate the model is taken from the Neighborhood Change Database for years 1980, 1990 and 2000 (Tatian, 2002) which provides geographically consistent census data at tract level for different time periods. There are a total of 518 tracts in the region, of which 3 were excluded from the analysis as they suffered from heavy floods in early 1990s leading to mass out-migration of households. Summary statistics of the data are presented in Table 3.1.

Endogenous Variables

	Average Household Income, 2000	Number of Households, 2000	New Construction, 1990-2000	Vacancy Ratio, 2000	Density, 2000	Change in # HH, 1990-2000
	AVHHIN0	**NUMHHS0**	**BLTYR0**	**PVACHU0**	**DENSITY0**	**DHH0**
Min.	12.840	0.025	0.000	0.000	0.010	-1.094
1st Qu.	39.280	1.211	0.026	0.035	1.320	-0.077
Median	50.310	1.851	0.114	0.056	3.944	0.041
Mean	54.790	1.923	0.281	0.084	4.974	0.129
3rd Qu.	63.560	2.594	0.394	0.104	7.022	0.252
Max.	219.950	5.082	2.296	0.389	36.245	2.181
	thousand $	thousand units	thousand units			thousand units

Exogenous Variables

	Average Household Income, 1990	Number of Households, 1990	New Construction, 1980-1990	Vacancy Ratio, 1990	Percent High School Dropout, 1990	Abandoned Units, 1990	Abandonment Ratio, 1990
	AVHHIN9	**NUMHHS9**	**BLTYR9**	**PVACHU9**	**HSDROP9**	**ABHU9**	**ABHUR9A**
Min.	8.130	0.018	0.000	0.000	0.000	0.000	0.000
1st Qu.	27.710	1.168	0.049	0.042	4.600	0.042	0.026
Median	35.430	1.733	0.222	0.058	9.680	0.092	0.069
Mean	38.070	1.793	0.323	0.085	11.820	0.123	0.094
3rd Qu.	45.070	2.393	0.480	0.105	16.190	0.163	0.128
Max.	158.870	5.017	2.409	0.359	67.960	1.202	0.660
	thousand $	thousand units	thousand units			thousand units	

Table 3.1: Summary Statistics

Housing abandonment is not reported in any database and thus is estimated in the model. It mostly consists of units that were part of the housing stock in the previous period and not reported in the current period. Some of these may have either been converted into other land-use or are left unclaimed. Housing abandonment is estimated as the difference between magnitude of new construction and change in housing stock. For example, if there is no new construction and the reported housing units in a tract decreased by 100 units, then we could assume that 100 units were abandoned. Similarly, if there are 500 new units constructed, while the housing stock increased by only 450, then we could assume that 50 units were abandoned. When units are abandoned in a tract, it creates negative externalities in the next period and creates disincentives for residents to continue living there.

3. Social Dynamics Model

Specification

The nature of the problem at hand demands explicit handling of the spatial structure of data and that of urban dynamics in two ways. First, the variables in a tract are spatially correlated with its neighbors. This is because the processes in social dynamics have spatial spillovers through externalities and diffusion, e.g., being surrounded by tracts with high housing abandonment makes it more likely for a tract to witness abandonment. Second, the disturbance structure in the specification for each endogenous variable can be spatially correlated in cross section and across equations. This may happen due to omitted variables like crime which is spatially diffusive in nature, or due to special circumstances in some tracts (like the presence of a brownfield) which may affect surrounding tracts. Further, the data collected at tract level (which is an artificially imposed political boundary) is different from the spatial geography at which urban dynamics occur, creating the problem of ecological fallacy and thus spatial dependence in error structure.

In the absence of an established theoretical relationship between urban distress and the social fabric, a reduced form model structure based on the causal mechanisms as described in Chapter 2 is used. It is reduced in the sense that even though the causal linkages between variables are known the exact functional form of these relationships is not known.

The social dynamics from Figure 3.3 is modeled using a system of m simultaneous equations with one equation for each endogenous variable and spatially correlated disturbance. Each of these equations is specified in an autoregressive framework with the spatial and temporal lags of endogenous variables including that of the dependent variable appearing on the right hand side of each equation. A typical equation (e.g., i^{th} variable in period *t*) in the model has the following structure.

$$
\begin{aligned}
Y_{i,t} &= \beta Y_{j \neq i,t} + \gamma W Y_{.,t} + \delta h(X_{k,.}) + u_{i,t} \\
u_{i,t} &= \rho W u_{i,t} + \varepsilon_{i,t}, \quad \operatorname{cov}(\varepsilon_{j=1..m,t}) = \Sigma
\end{aligned}
\qquad ..(1)
$$

Where, $Y_{i,t}$ is the i^{th} endogenous variable at time *t*. The first term $Y_{j \neq i,t}$ represents the impact of other endogenous variables that contemporaneously determine the behavior of the i^{th} dependent variable. For example, if we consider the average household income as the i^{th} variable, it is contemporaneously determined by the change in number of households or density in a tract. The term $W Y_{.,t}$ represents the spatial lag of Y's (the average value of the variable in the neighboring tracts with the neighborhood structure defined by matrix *W*) at time *t*. The spatial lag of endogenous variables in the current period accounts for the neighborhood effect of these variables in jointly determining the state of these variables. For example, income in a tract is dependent on the average income and vacancy in surrounding tracts in the current period. Here one should be cautious to interpret that other variables don't change the income of households living in a tracts but affect the migration decisions of households in certain income groups leading to changes in the average income.

$h(X_{k,.})$ is a function of exogenous variables including policy variables and other time invariant characteristics of the tract. It includes the spatio-temporal lag of variable, e.g. the average income of surrounding tracts in previous period. It also includes some function of the lagged value of dependent variables $g(Y_{.,t-1})$, e.g., income in the previous year or vacancy in the previous year. Functions *g* and *h* are assumed to be linear in absence of any theoretical reasons to choose otherwise or is guessed using exploratory data in ways that are consistent with the underlying causal mechanisms. β, γ, δ are vector of corresponding coefficients which are estimated using the feasible generalized spatial three stage least square (FGS3SLS) method based on the methods of moments estimator as proposed in Kelejian and Prucha (2004).

This system of equations can be seen as a spatial extension to 'neighborhood ecology models' in concept and an extension to the widely used single equation model of Cliff and Ord (1973, 1981) in terms of econometric specification. Each of the *i* equations in (1) can be compressed and written in matrix notation corresponding to Kelejian and Prucha (2004) as follow.

$$\mathbf{Y}_n = \mathbf{Y}_n\mathbf{B} + \mathbf{X}_n\mathbf{C} + \overline{\mathbf{Y}}_n\Lambda + \mathbf{U}_n \qquad ..(2)$$

with $\mathbf{Y}_n = (\mathbf{y}_{1,n}, \mathbf{y}_{2,n}.....\mathbf{y}_{m,n})$, $\mathbf{X}_n = (\mathbf{x}_{1,n}, \mathbf{x}_{2,n}.....\mathbf{x}_{k,n})$, $\mathbf{U}_n = (\mathbf{u}_{1,n}, \mathbf{u}_{2,n}.....\mathbf{u}_{m,n})$,

$\overline{\mathbf{Y}}_n = (\overline{\mathbf{y}}_{1,n}, \overline{\mathbf{y}}_{2,n}.....\overline{\mathbf{y}}_{m,n})$, $\overline{\mathbf{y}}_{m,n_n} = \mathbf{W}_n\mathbf{y}_{j,n}$ $\quad j = 1,..., m$

where, $\mathbf{y}_{j,n}$ is the $n\times 1$ vector of cross sectional observations on each of the dependent variables. Since the model is conditional on the realized value of $\mathbf{y}_{j,n}$ in the previous period, the temporally lagged endogenous variables are treated as given and forms a component of vector $\mathbf{x}_{j,n}$. $\mathbf{u}_{j,n}$ is the $n\times 1$ disturbance vector in the j^{th} equation, $\mathbf{W}_n$ is an $n \times n$ row standardized weights matrix of known constants, and $\mathbf{B}$, $\mathbf{C}$, and Λ are correspondingly defined parameter matrices of dimension $m\times m$, $k \times m$ and $m\times m$ respectively. The spillovers in endogenous variables and spatial correlation of disturbance terms are modeled as follow. The vector $\overline{\mathbf{y}}_{j,n} = \mathbf{W}_n.\mathbf{y}_{j,n}$ models the spatial spillover of endogenous variable, often referred to as the spatial lag of $\mathbf{y}_{j,n}$. Its r^{th} element represents the average of the values of y in the neighborhood of r. The error terms are assumed to be generated by the following spatially autoregressive process

$$\mathbf{U}_n = \overline{\mathbf{U}}_n \mathbf{R} + \mathbf{E}_n \qquad ..(3)$$

with $\mathbf{E}_n = (\varepsilon_{1,n}, \varepsilon_{2,n} .....\varepsilon_{m,n}), \quad \mathbf{R} = diag_{j=1}^{m}(\rho_j)$

$$\overline{\mathbf{U}}_n = (\overline{\mathbf{u}}_{1,n}, \overline{\mathbf{u}}_{2,n} .....\overline{\mathbf{u}}_{m,n}), \quad \overline{\mathbf{u}}_{j,n} = \mathbf{W}_n \mathbf{u}_{j,n}, \quad j = 1,..., m$$

where, $\varepsilon_{j,n}$ is the *n × 1* vector of disturbances, ρ_j is the spatial autoregressive parameter in the j^{th} equation. The vector $\overline{\mathbf{u}}_{1,n}$ is defined analogous to the spatial lag variable $\mathbf{y}_{j,n}$. The disturbances are not only assumed to be spatially correlated across space but also across equations.

Estimation

Conventional estimators like the maximum likelihood estimator are difficult to compute for the model defined in equations (2) and (3) as it involves calculation of Eigen values of large matrices in situations where the roots are not guaranteed to be real. Thus, for reasons of computational simplicity and without the foreseeable risk of losing any information, a generalized method of moments estimator proposed by Kelejian and Prucha (2004) and referred to as limited information Feasible Generalized Spatial Two Stage Least Squares procedure (FGS2SLS) and full information Feasible Generalized Spatial Three Stage Least Square (FGS3SLS) estimators are used. These estimators are asymptotically efficient and easier to compute relative to the maximum likelihood estimator. The consistent estimate of the spatial autocorrelation of the error terms in each equation is obtained by imposing three moment conditions on the error terms ε_j in each equation. However, this methodology does not give the standard error of estimate for the spatial dependence parameter of the error terms. Nevertheless, the estimates are consistent and asymptotically efficient (Kelejian & Prucha, 1999).

An important issue in estimation is the choice of the weight matrix which defines the neighborhood structure of tracts. It is usually assumed to be known *a priori*, or is based on ad hoc considerations. Two alternative definitions of weight matrices were considered, namely (a) a single weight matrix for the whole region, and (b) separate weight matrix for the east and west parts of the region as they belong to different states. These were computed from contiguity based relationships. The latter option was chosen as it does not allow for spatial spillovers across the river that divides the region in two parts and allows greater flexibility in modeling. A more generalized option to be considered may include distance between two spatial units or length of shared border (Cliff & Ord, 1973, 1981) but this option was not pursued in this study.

Results

The model described in equations (2) and (3) is estimated using the FGS2SLS and FGS3SLS estimators and the results are presented in Table 3.2. The parameter estimates were almost identical from the two estimates but, the latter had smaller standard errors for most of the coefficients. It also had a marginal improvement in goodness of fit. For estimation purposes, the minimum values of some variables were capped to 0.01 where an inverse of the variable was needed. Diagnostics for presence of spatial autocorrelation, spatial regimes and heteroskedasticity in residuals was performed for each equation (Anselin, 2002; Anselin & Bera, 1998). Some of the model diagnostic results are graphically presented in Appendix B. The model results are explained below.

Dependent	Average HH Inc	Exp Sign	GS2SLS	s.e.	t-stat	GS3SLS	s.e.	t-stat
Endogenous	**DENSITY0**	-	-0.117	0.083	-1.406	-0.114	0.082	-1.392
	BLTYR0	+	6.908	1.689	4.090	6.641	1.674	3.966
Exogenous	**AVHHIN9**	+	1.296	0.025	52.287	1.298	0.024	53.176
	1/ABHU9	+	0.036	0.023	1.514	0.047	0.023	2.006
Spatial Lags	**Dependent Variable**	+	0.055	0.021	2.646	0.052	0.021	2.520
	Error		0.090			0.090		
Squared correlation			0.900			0.900		

Dependent	Number of Households	Exp Sign	GS2SLS	s.e.	t-stat	GS3SLS	s.e.	t-stat
Endogenous	**BLTYR0**	+	0.894	0.042	21.532	0.918	0.039	23.290
Exogenous	**c**		-0.281	0.039	-7.128	-0.257	0.038	-6.847
	AVHHIN9	+	0.002	0.001	2.993	0.002	0.001	2.737
	NUMHHS9	+	0.973	0.013	73.150	0.975	0.012	78.177
	1/HSDROP9	+	0.003	0.001	1.772	0.002	0.001	1.640
Spatial Lags	**Dependent Variable**	+	0.053	0.020	2.682	0.043	0.018	2.358
	Error		0.050			0.050		
Squared correlation			0.959			0.961		

Dependent	New Construction	Exp Sign	GS2SLS	s.e.	t-stat	GS3SLS	s.e.	t-stat
Endogenous	**DHH0**	+	0.584	0.050	11.674	0.678	0.047	14.307
	PVACHU0	+	0.669	0.358	1.869	0.983	0.327	3.009
Exogenous	**BLTYR9**	+	0.320	0.029	11.110	0.278	0.027	10.169
	PVACHU9	-	-0.152	0.374	-0.408	-0.467	0.340	-1.376
Spatial Lags	**Dependent Variable**	+	0.225	0.036	6.192	0.216	0.034	6.414
	Error		-0.182			-0.182		
Squared correlation			0.844			0.850		

Dependent	Vacancy Rate	Exp Sign	GS2SLS	s.e.	t-stat	GS3SLS	s.e.	t-stat
Endogenous	**AVHHIN0**	-	0.0001	0.000	-2.617	-0.0002	0.000	-2.785
	DHH0	-	-0.0310	0.008	-3.723	-0.0272	0.008	-3.256
Exogenous	**1/AVHHIN9**	+	0.0003	0.000	1.668	0.0004	0.000	1.728
	1/BLTYR9	+	0.0002	0.000	2.282	0.0003	0.000	2.635
	ABHUR9	+	0.0765	0.027	2.850	0.0684	0.027	2.552
	PVACHU9	+	0.6068	0.047	12.995	0.6021	0.047	12.933
Spatial Lags	**Dependent Variable**	+	0.2569	0.041	6.220	0.2552	0.041	6.192
	Error		0.0500			0.0500		
Squared correlation			0.796			0.793		

Table 3.2: Model Estimation Results

Average household income increased in tracts with larger numbers of new housing units. Higher density was correlated with decreasing income, but its effects were statistically insignificant. Income in a tract was highly correlated with its past values and current values in neighboring tracts. This shows a strong spatio-temporal persistence and propagation of income patterns in the region. Housing abandonment was inversely related to income. Impact of vacancy and school quality were not significant and were excluded. This may be because these effects were indirectly accounted for through new construction and income in surrounding tracts and income in the previous period.

The number of households in tracts increased with new construction. Tracts with higher income in previous period saw more in-migration of households and tracts with better school performances attracted more households, though the effect of better schools was only marginally significant. Surprisingly, school quality did not achieve significance in any regression and this may be attributed to the poor quality of the school data. Like income, household dynamics in each tract too had significant spatio-temporal propagation and persistence.

New construction and in-migration of households occurred simultaneously in tracts. The new constructions were associated with the increase in vacancy in the current period probably reflecting the delay between construction and occupation. It was more in tracts which had higher new construction in the previous decade indicating that growing tracts continued to attract growth. New construction in neighboring tracts had positive spillovers on a tract making it attractive for development. However, the growth was deterred in tracts which had more vacancies in the previous period as they were considered less attractive for households to move in.

Vacancy rates were lower in tracts with higher income and larger in-migration. It increased with out-migration of households and led to conditions more prone for decline. It was higher in tracts which had lower income, lower construction, higher vacancy and higher housing abandonment in the previous decade, implying these to be major repellants of households. Vacancy in surrounding tracts also created vacancy possibly due to spatial autocorrelation among other factors of distress.

Housing abandonment was derived from these equations as the difference between magnitude of new construction and change in housing stock. The results show significant concentration of abandoned housing units in the urban core (Figure 3.5) especially in the tracts that had high vacancy in the previous period and those that witnessed maximum out migration. This result is heartening because even though the housing abandonment was not explicitly modeled, its spatio-temporal dynamics were captured in the reduced form model.

It seems reasonable that construction of new housing units in a tract would lead to increase in average income, attract more households and reduce housing abandonment in the same decade and leading to decrease vacancy, attraction of new construction in the next decade. This has strong policy implications for positive intervention in tracts facing decline.

Generating Macro Control Totals Endogenously

Once the model is estimated, it is used to predict the future values of endogenous variables at the tract level consistent with regional forecasts of number of households and average household income from the economic driver model. To achieve this, an objective function (Equation 4) is set up to minimize the deviation of the sum of forecasted values at tract level and the regional aggregate from economic model.

Currently equal weights are given to deviations of forecasted income and number of households from the predictions from the regional economic model.

$$\min_{HH,Inc} \left(\sum_n NHH_n - NHH_{econ}\right)^2 + \left(\sum_n (NHH_n . AvgHHInc_n) - Inc_{econ}\right)^2 \quad ..(4)$$

Subject to

$$\mathbf{Y}_n \leq \mathbf{Y}_n \hat{\mathbf{B}} + \mathbf{X}_n \hat{\mathbf{C}} + \overline{\mathbf{Y}}_n \hat{\Lambda} + \sigma_n$$

where, HH_{econ} and Inc_{econ} are the predictions for the total number of households and total personal income respectively from the regional economic model, NHH_n and $AvgHHInc_n$ represents the number of households and the average household income in tract *n*. The constraints are the set of equations estimated using FGS3SLS. However, instead of having each equation as strict equality, a margin of error in each equation is permitted within the standard error of regression obtained from the estimates of each equation. Further, non-negative constraints are imposed on number of new construction and other variables as deemed necessary. The solution of the minimization problem as stated in equation 4 then (a) endogenously determines the magnitude and location of new construction, housing abandonment, vacancy etc. consistent with the regional increase in household and income growth and (b) provides growth, housing abandonment and vacancy maps to inform the land-use change.

4. Integration with Land-use Change Model

Figure 3.4 shows the probability of development attractiveness resulting from conventional drivers of land-use change, and projected values of changes in number of households, housing abandonment and vacancy in each tract. Development attractiveness resulting from conventional drivers such as access to jobs shows that the regional core has a high probability of growth. However, the prediction of household behavior in response to housing abandonment and vacancy reveals its unfavorable position. The relative tradeoff between these two opposite forces would eventually determine the pattern of land-use change in the region.

The total demand for housing units and the regional income is usually derived from the regional economic model. Using the model described in the previous sections, the land-use and spatial distribution of income, vacancy, housing abandonment, etc. can now be used to determine the total demand for housing, along with the input from the regional economic model. This will account for new housing constructions needed to accommodate the new households as well additional demand generated through intra-regional migration, vacancy and housing abandonment. This proposed mechanism of integration with land-use change model is graphically depicted in Figure 3.5.

In the proposed framework, the single residential land-use category is disaggregated into occupied, vacant, abandoned and new housing units. This is because each of these categories has different effect on its neighbors in terms of attracting or repelling growth. New units may be attractors for growth while abandoned units repel growth. The spatial distribution of new construction, vacancy and housing abandonment are then used in conjunction with other drivers of land-use change like accessibility to job, amenities, environmental resources, geographical characteristics, etc. to determine the final location of new housing units and households in the region (Figure 3.5). The new land-use thus generated is used as an input for subsequent years and the cycle is repeated.

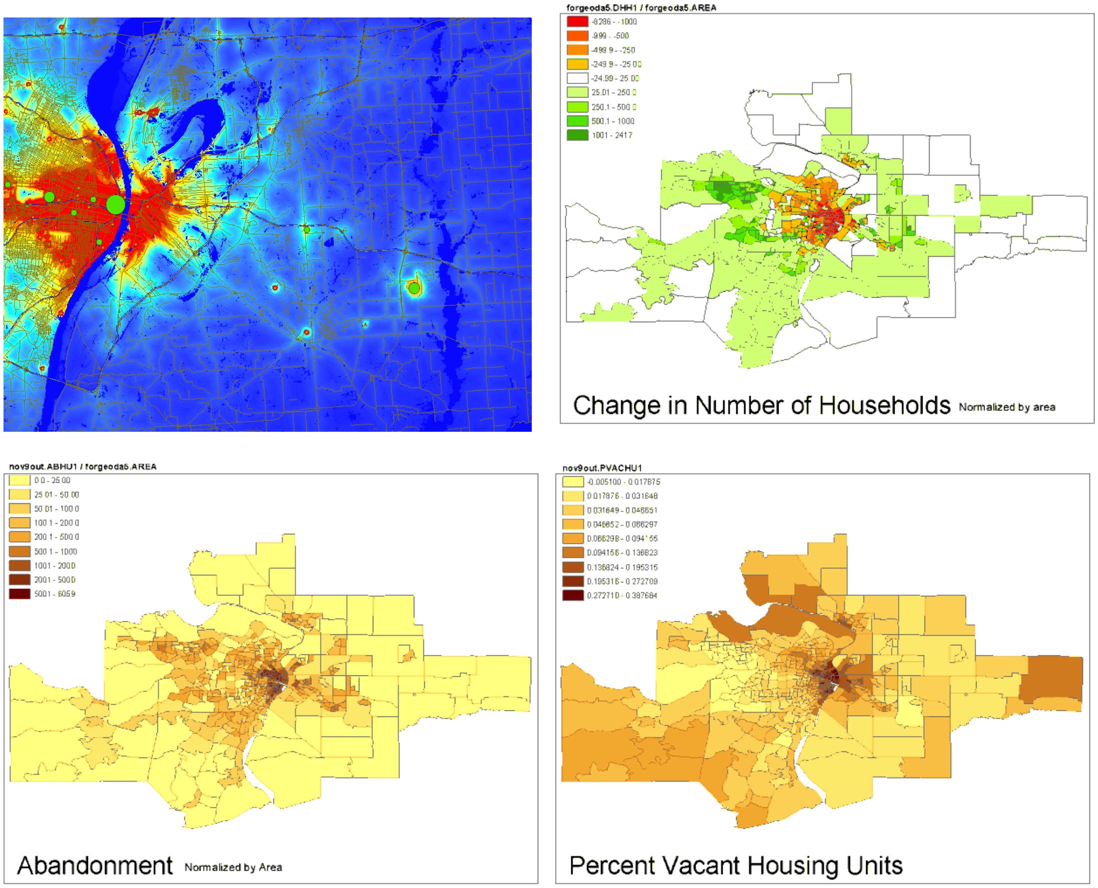


Figure 3.4: clockwise from top left. (a) Development attractiveness computed from travel time to employment centers in St. Louis region. Areas marked in red are more attractive than those in blue due to shorter travel time to employment centers shown in circles. (b) Projected change in number of households in 2010. Dark red tracts the biggest losers of households while dark green tracts are the biggest gainers. (c) Projected location of vacant housing units in 2010, mostly concentrated in the regional core. (d) Projected location of housing abandonment in 2010, mostly concentrated in the regional core

5. Conclusion

This chapter demonstrates how the impact of urban distress (as measured by school quality, income disparity, vacant and abandoned housing units) can be incorporated within the land-use change models to predict the location, type and magnitude of land-use change. Changes in land use in turn affect measures of urban distress. The enhanced model predicts magnitude and location of new housing units and the behavior of households in response to spatial variations in the conditions of the social fabric. The model structure is flexible and allows us to incorporate and evaluate the impacts of alternative scenarios. Some of the possible scenarios planning exercise using this enhance model include evaluating the impacts of improving the school quality, promoting construction of new housing units, providing targeted income generating activities, etc. in distress parts of the region.

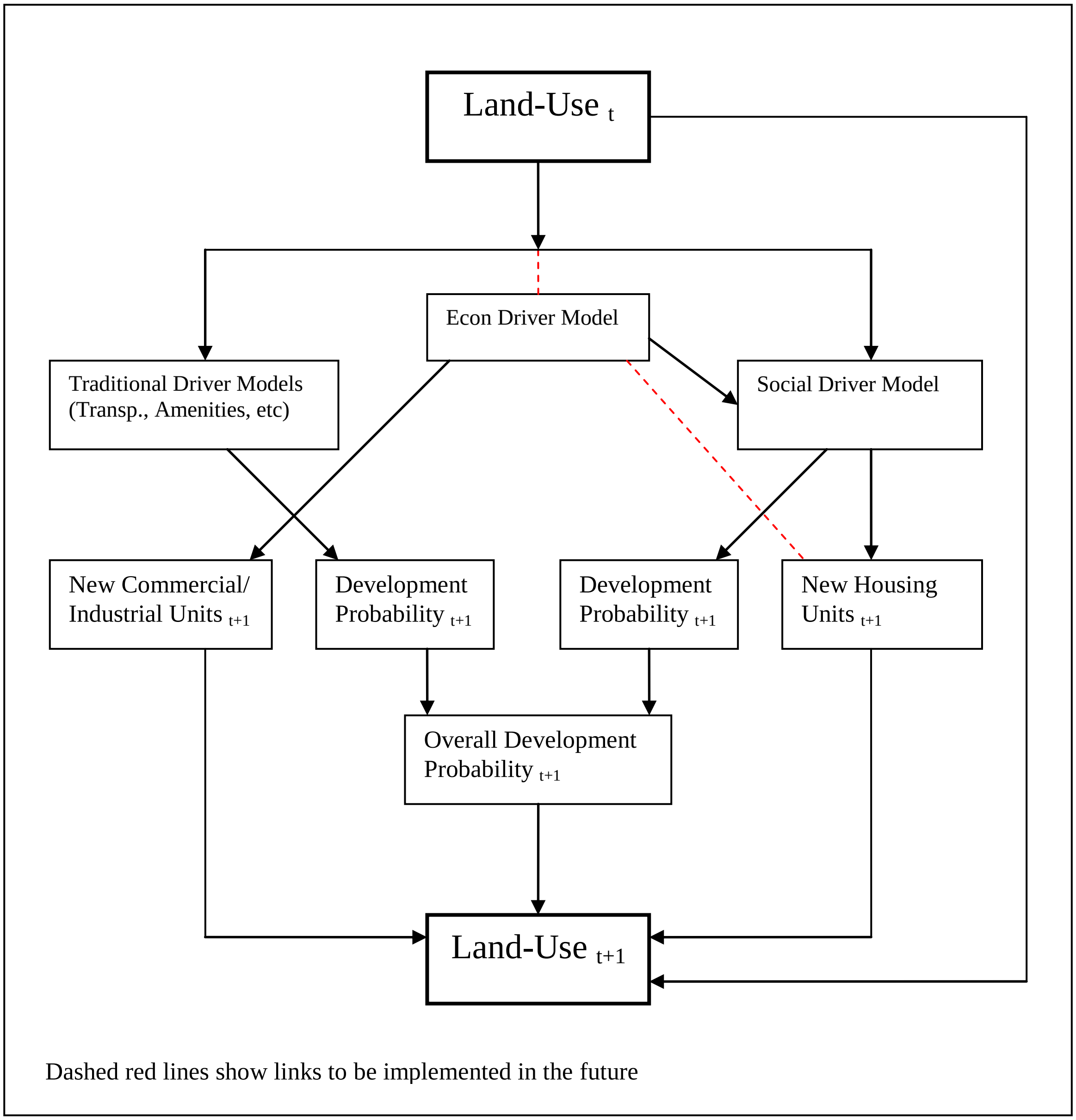


Figure 3.5: Proposed flow of information in an integrated model

An important point made in this chapter is that population and economic growth are not the only determinants of the number of new housing units constructed in a region. In region after region, a small increase in population has been associated with a large amount of new housing units through vacancy and abandonment of housing units. Sprawl, largely caused by declining urban core and intra-regional migration, leads to more residential growth in the periphery than is needed to accommodate the increase in the number of households. The amount of new residential development depends on the amount and intensity of urban distress in addition to conventional determinants like interest rates, economic growth, etc. Further, the location of these new households is not merely based on normative concepts of optimizing travel time to work or maintaining job-housing ratios but also on perceptions held by households of the social environment. This point has been largely ignored in the literature on land-use change and addressing it was the precise aim of this chapter.

CHAPTER 4: PERFORMANCE OF THE FGS3SLS ESTIMATOR IN SMALL SAMPLES - A MONTE CARLO STUDY

1. Introduction

The modeling of spatial processes has attained a mainstream position in social sciences (Goodchild, Anselin, Appelbaum, & Harthorn, 2000). In the simplest cases, the variables of interest are spatially correlated with their neighbors and with other variables. As we move from one variable to a system of variables, modeling the spatial interactions gets complex. The complexity further increases as the randomness too gets correlated spatially and across equations. Modeling the strength of spatial interactions and externalities requires the specification and estimation of spatial econometric models. However, the available estimators (Anselin, 1988; Case, 1991; Case, Rosen and Hines, 1993) lack methodological sophistication and computational simplicity to accurately estimate simultaneous systems with spatial autoregressive dependent variables and spatially interrelated cross sectional equations. They are often based on quasi maximum likelihood procedures and might not have feasible solutions in medium to large samples. Further, they are designed for single equation frameworks. See Kelejian and Prucha (1999) for a discussion on this issue.

To estimate models for such processes, Kelejian and Prucha (2004) proposed the limited information Feasible Generalized Spatial Two Stage (FGS2SLS) and full information Feasible Generalized Three Stage Estimators (FGS3SLS). These estimators are based on generalized methods of moments using approximation of optimal instruments, and thus are computationally simple. Kelejian and Prucha show that the estimators are consistent and asymptotically normal. Some of the applied examples of this estimator include

Ngeleza, Florax, and Masters (2006) to determine the geographical and institutional determinants of real income, Driffield (2006) for modeling spatial spillovers of foreign direct investment, Fishback, Horrace, and Kantor (2006) for modeling the impact of New Deal expenditures on mobility during the great depression.

It is important to understand how this estimator behaves in applied works given its relevance to estimate many of the complex spatial processes which were largely ignored so far. However, our understanding of this estimator is at best, rudimentary. The number of published works using this estimator is relatively few, and only its asymptotic properties have been established so far. In absence of very large samples as is the case in most applied works, it is difficult to interpret the results with confidence based on asymptotic results only.

One alternative to employ in a situation such as this is to use finite sample approximations or asymptotic expansions. However, these approximations tend to be very complex, the results difficult to interpret and the computations very advanced. Some early work on this topic is summarized in Philips (1983) and Rothenberg (1984). In contrast, the method of Monte Carlo replaces the skills needed in asymptotic approximations by relying on computational power of computers. Here, the properties of the parameters of interest are studied through a series of stochastic simulations and their statistics are analyzed (Davidson and MacKinnon, 1993).

This chapter investigates the performance of the FGS3SLS estimator for a system of simultaneous equations, with spatial autoregressive dependent variables and spatially autocorrelated error structures using Monte Carlo experiments. Performance is measured by its ability to estimate parameters of the model and sensitivity of the results to varying degree of spatial dependences, choice of spatial weight matrix, sample size and variance covariance matrices. The chapter concludes by emphasizing the need for further studies on the subject to increase our understanding of the estimator's behavior in applied work.

The rest of the chapter is organized as follows. Section 2 sets up the model used for the study. Section 3 briefly describes the estimator and section 4 describes the experiment design. The results of simulation exercise are presented in section 5. Section 6 summarizes the main findings and concludes the study with direction for future research.

2. Model Structure

The performance of the FGS3SLS estimator was tested using a model specification closely resembling the structure of the model used in Chapter 3 to analyze the regional social dynamics and its impacts on Land-use change. The model used here consists of a system of simultaneous equations with two endogenous variables, their spatial and temporal lags and two exogenous variables. The spatial lag of the dependent variable is treated as endogenous while the temporal lag is considered as predetermined, since the model is conditioned on past values of dependent variable. The disturbances are assumed to be correlated across space and across different equations. This form of model allows (a) capturing spatial processes like diffusion across space, (b) address problems of ecological fallacy or presence of some local conditions leading to spatially correlated error structure, and (c) correlation between two spatial processes. Further, the specification allows forecasting of the value of dependent variables conditional on its own past values, and other exogenous variables after accounting for the underlying spatial processes.

Let y_1 represent percent abandoned housing units in a census tract and y_2 represent net in-migration of households. Equation 1 states that percent abandoned units y_1 depend on - magnitude of net in-migration of households (y_2) and percent housing abandonment in neighboring tracts (Wy_1) in the current period; percent housing abandonment in the previous period (x_1); distance from interstate (x_3); and a random component (u_1). Simultaneously, the net in-migration of households is endogenous and depends on percent of units abandoned since higher housing abandonment tends to repel more households from the region. According to the equation (2), the magnitude of net in migration of households (y_2) in a tract depends on - percent housing abandonment (y_1) and net in migration of households in the neighboring tracts (Wy_2) in the current period; lagged values of net in-migration (x_2); the condition of infrastructure (x_4); and a random component u_2 . Thus, housing abandonment and net in migration of households are jointly determined. Note that x_3 is treated as fixed over time while x_4 is time dependent but still exogenous.

$$y_1 + \gamma_1 y_2 + \lambda_1 W_1 y_1 + \beta_1 x_1 + \beta_3 x_3 = u_1 \quad ..(1)$$
$$\gamma_2 y_1 + y_2 + \lambda_2 W_2 y_2 + \beta_2 x_2 + \beta_4 x_4 = u_2 \quad ..(2)$$

where, $y_i, (i = 1, 2)$ represents the endogenous variables we are interested to forecast. $W_i y_i$'s are the spatially lagged dependent variables with the spatial lag parameter λ_i . W_i is the row standardized weight matrix of known constants describing the neighborhood structure of observations. x_1 and x_2 are the temporally lagged values of dependent variable y_1 and y_2 respectively, x_3 and x_4 are the exogenously determined variables whose values either remain fixed throughout the simulation or are known *a priori*. u_1 and u_2 represent the stochastic component of the model whose behavior is elaborated below.

The disturbance vectors u_1 and u_2 in equations (1) and (2) are assumed to be correlated across space and across equations. The spatial geography at which the social dynamics are occurring is different from the administrative geography of census tracts. The aggregation of data at tract level leads to correlation of disturbances across tracts. Further any randomness affecting housing abandonment and change in number of households may be correlated. Thus, the current specification allows for randomness that is also correlated across equations.

$$u_1 = \rho_1 W_3 u_1 + \varepsilon_1 \text{ ..(3)}$$
$$u_2 = \rho_2 W_4 u_2 + \varepsilon_2 \text{ ..(4)}$$
$$with \quad \Sigma = Cov(\varepsilon_1, \varepsilon_2) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} \text{ ..(5)}$$

Equations (3) and (4) characterize the correlation across space where $W_3 u_1$ and $W_4 u_2$ are average value of error terms in the neighboring locations, and ρ_1 and ρ_2 depict the degree of spatial correlation of the error terms. ε_1 and ε_2 are non-spatially correlated disturbances but are correlated across equations with the variance covariance matrix Σ (equation 5). This completes specification of the hypothetical model.

Generalized form

For brevity, the model system represented in equations (1) to (5) can be rewritten in matrix notation as

$$\begin{pmatrix} I_n+\lambda_1.W_1 & \gamma_1.I_n \\ \gamma_2.I_n & I_n+\lambda_2.W_2 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}+\begin{pmatrix} \beta_1.I_n & 0.I_n \\ 0.I_n & \beta_2.I_n \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}+\begin{pmatrix} \beta_3.I_n & 0.I_n \\ 0.I_n & \beta_4.I_n \end{pmatrix}\begin{pmatrix} x_3 \\ x_4 \end{pmatrix}$$

$$=\begin{pmatrix} (1-\rho_1 W_3)^{-1}.\varepsilon_1 \\ (1-\rho_2 W_4)^{-1}.\varepsilon_2 \end{pmatrix} \quad ..(6)$$

$$or, \quad B.Y+T_a.X_a+T_b.X_b=U \quad ..(7)$$

where, $Y=\left(y_1,y_2\right)$ is a vector of endogenous variables, $X_a=\left(x_1,x_2\right)$ is a vector of temporally lagged endogenous variable *Y*, $X_b=\left(x_3,x_4\right)$ is a vector of exogenous variables and I_n is an identity matrix of dimension *n*. *B*, T_a and T_b represents the coefficient associated with these variables in equation (6). $U=\left(u_1,u_2\right)$ represent the vector of disturbance terms. The estimator is described in Appendix C.

3. Monte Carlo Experiments

With the model structure in place, designing the Monte Carlo experiment consists of three additional parts, namely defining the parameter settings; generating the spatio-temporal array of synthetic data for different variables consistent with the underlying spatial process; and designing the simulations to reduce errors due to randomization and analysis of alternative scenarios. Each of these steps is elaborated below.

Parameter settings

This section assigns values to the parameters used in the model specified in equations (1) through (5) including the values of all the coefficients, the variance covariance matrix of disturbance terms, the weight matrix and the spatial dependence parameters.

The parameter settings for the model are defined as follow

$\gamma_1 = -0.3, \qquad \gamma_2 = 0.7, \qquad \beta_1 = 2.0, \qquad \beta_2 = 2.5 \qquad \beta_3 = 2.5, \qquad \beta_4 = -2.0$

The following two alternative forms of the variance covariance matrix Σ are used:

$$\Sigma_1 = \begin{pmatrix} 900 & 450 \\ 450 & 900 \end{pmatrix} \quad \text{and} \quad \Sigma_2 = \begin{pmatrix} 3000 & -2500 \\ -2500 & 4000 \end{pmatrix}$$

In the first case, $\sigma_1 = \sigma_2$ and the correlation between error terms $corr(\varepsilon_1, \varepsilon_2) = \sigma_{12} / \sigma_1\sigma_2$ is 0.50. In the second case, $\sigma_1 \neq \sigma_2$ and the correlation between error terms is -0.72. This corresponds to an average R^2 value of roughly 0.75 and 0.6 respectively, where R^2 is defined as the average squared correlation coefficient (Carter & Nagar, 1977) between y_i and the mean value of y_i as explained by the model in different experiments. The parameters for spatial lag of dependent variable and for spatial autocorrelation in error terms {λ_i, ρ_i} include all possible combinations from the set {-0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8} in different experiments for each choice of Σ. For clarity of the exposition, I assume a common neighborhood structure $W (= W_1 = W_2 = W_3 = W_4)$, $\rho_1 = \rho_2$ and $\lambda_1 = \lambda_2$. It should be noted that in applications, this is not the case. Weight matrices for different variables will take different specifications depending on the nature of spatial processes that influence them. However, there is no loss of generality by using the same weight matrix for different process for the Monte Carlo experiments.

I consider three samples sizes of 100, 250 and 500 observations each. For each sample size, two different weight matrices are considered. The specification of *W* closely follows the weigh matrix described in Kelejian and Prucha (1999) and Das, Kelejian, and Prucha (2003). These matrices differ in the degree of sparseness. For the first specification, a hypothetical circular world is considered where each observation (y_i and u_i) is related to exactly one neighbor immediately before it and one neighbor immediately after it. Thus, the i^{th} row of *W* has non-zero entities only in *i-1* and *i+1* column, for each *i = 2, 3…. (n-1)*. For the first row, the non-zero elements are in the 2^{nd} and n^{th} column while for the last row, the non-zero elements are in the $n\text{-}1^{\text{th}}$ and 1^{st} column. Further, the matrix *W* is row standardized such that sum of elements in each row =1. This matrix is termed as "one ahead and one behind". The second matrix is analogously defined as "three ahead and three behind" where each observation is related to exactly three other observations behind and ahead of it. Thus, the average number of neighbors in the first matrix is 2 while in the second matrix is 6. Kelejian and Prucha reports that the results from hypothetical weight matrices and "real world" weight matrices are similar. I conjecture similar implications in this case.

Generating Synthetic Data

A dataset which is a realization of the spatial process under study is needed for purpose of estimation. It should be a generated from interdependencies between variables, random components and the spatial interactions between them as specified in the model structure. For each scenario, a different dataset is generated influenced by the parameter settings, nature and strength of spatial dependence, variance-covariance structure and sample size. It is ensured that the variation in results of different scenarios only reflects the changes in the scenarios rather than the randomness in the data generation process to make comparison possible. The data generation process consists of two parts namely generating the values of the disturbance terms and that for the regression variables.

Generating the values of disturbance terms

The process of generating spatially correlated random components starts with random draws from independently and identically distributed normal random variables $\nu_i, (i = 1, 2)$

with zero mean and unit variance. These are then transformed to reduced form disturbances ε_i that are correlated across equations with zero mean and variance covariance matrix Σ using the following transformation

$E = V\Sigma_*$ where, $E = (\varepsilon_1, \varepsilon_2)$, $V = (\nu_1, \nu_2)$

and Σ_* is the m x m lower triangular matrix such that $\Sigma_*^{'}\Sigma_* = \Sigma$. The disturbance terms u_i's in the model are then obtained by using the transformation $u_i = (I - \rho W)^{-1}\varepsilon_i$ resulting in randomness that is correlated across space as well as across equations.

Generating the values of regression variables

The starting values for a large number of time series on two exogenous variables $X_{b,t} = \left(x_3, x_4\right)$ are independently drawn from a normal distribution with zero mean and unit variance. x_3 is treated fixed over time while x_4 is assumed to grow at a rate of one percent in every period. To avoid the sensitivity of results to exogenous variables, they are generated using the same set of random realizations in every experiment.

Values of Y_t are generated conditional on $X_{a,t}$ and $X_{b,t}$ using reduced form autoregressive data generation process described as follows. Re-writing equation (7) with time subscript, substituting $X_{a,t} = Y_{t-1}$ and taking expectations, we get

$B.Y_t + T_a.Y_{t-1} + T_b.X_{b,t} = 0$ ..(8) or,

$Y_t = -(B)^{-1}.(T_a.Y_{t-1} + T_b.X_{b,t})$ ..(9)

True values for Y_t are generated using equation (9) for each period starting from initial values of $Y_{t=0}$ from normal random variables, and exogenously generated values for the variable $X_{b,t}$. This process is iterated several times to ensure that the pre-determined variable Y_{t-1} is generated using the same underlying spatial process as Y_t. The observed value of Y_t is subsequently obtained by perturbing its true values with disturbances $U = (u_1, u_2)$ whose values were generated in the previous section.

$$Y_{t,observed} = Y_t \pm U_t \quad ..(10)$$

The results from Monte Carlo simulations are at best random. In order to obtain sufficiently accurate results, a large number of repetitions is required. The errors due to the number of repetitions were reduced by use of antithetic variates. Thus, in equation (10) both positive and negative error terms are used to generate the observed values of *Y*.

Simulation design

Random samples are drawn from a specified distribution, and a set of data consistent with the model is generated. It is then used to estimate model parameters using the FGS2SLS and FGS3SLS estimator. This process is repeated several times. The estimates are then averaged to obtain the expected values of parameters of interest. The whole process is repeated for varying degrees of spatial dependences, sample size and the neighborhood structure to analyze the performance of FGS3SLS estimator under different conditions to analyze the sensitivity of the results to the data generation process. The complete code for the experiments is written in the statistical package R (R Development Core Team, 2005).

4. Results

Monte Carlo simulation using above parameters and synthetically generated data is performed for all combinations of the weight matrix *W*, sample size *n* and the spatial lag parameter λ. 500 random samples of errors are generated for each set of *n*, ρ, the neighborhood matrix *W* and the covariance matrix Σ. Each vector of errors is used twice (as thetic and anti-thetic variates) resulting in 1000 repetitions for each experiment. This setup yields a total of 2 values for Σ, 2 for *W*, 9 for λ_i, 9 for ρ_i and 3 for *n* resulting in 972 experiments with 1000 repetitions for each experiment.

The performance of the Feasible Generalized Spatial Three Stage Least Square estimator (FGS3SLS) was found to be superior to the Feasible Generalized Spatial Two Stage Least Square estimator (FGS2SLS) which in turn was found to be superior to the ordinary two stage least square estimator under varying conditions. Table 4.1 demonstrates the gains of using FGS3SLS for the parameter settings described in this chapter. The estimates from FGS3SLS have smaller bias and variance compared to the FGS2SLS estimator. The gains of using the former are more when the spatial correlation in disturbance terms is high, the spatial lag parameter has low absolute value, the sample size is small and the neighborhood structure is less dense.

Given the overall superiority of the FGS3SLS estimator under different conditions, I will only focus on the properties of FGS3SLS in the subsequent analysis. The simulations permit analysis of the impact of sample size, neighborhood density, variance-covariance structure of disturbances and the strength of spatial dependence on parameter estimates obtained using this estimator.

Impact of sample size on parameter estimates

In this section, I analyze the impact of sample size on parameter estimates using root mean square errors (RMSE) as a measure of performance for the FGS3SLS estimator. An attempt is made to isolate the interaction effects of sample size with neighborhood density (average number of neighbors), variance covariance matrix of error structures, degree of spatial dependences in endogenous variables and spatial autocorrelation in errors. For brevity of presentation, I choose one value of ρ and show the impact of varying sample size on RMSE of $\hat{\gamma}_2$ for different values of λ. Similarly, I choose one value of λ and show the impact of varying sample size on RMSE of $\hat{\gamma}_2$ for different values of ρ. The exercise is repeated for the two variance-covariance matrices (Figure 4.1).

Increasing the sample size from 100 to 250 observations had a huge impact on the RMSE of a parameter estimates irrespective of other control variables like neighborhood density or the variance covariance matrix. However, the gains in increasing the sample size from 250 to 500 were marginal except at extreme values of spatial dependence parameters λ and ρ. A large sample size improves the performance much more when the spatial lag parameter of the dependent variable is small, the spatial autocorrelation in errors is high, the number of neighbors is large and the variance covariance structure of error are large.

Impact of the average number of neighbors specified in the weight matrix W

The choice of neighborhood structure as defined by W is often decided *a priori* using exploratory data analysis or is based on the goodness of fit criteria. This is because the data generation process is not known in practice and the theory behind selection of the weight matrix is weak.

According to the simulations, the impact of neighborhood density on RMSE of parameter estimates depends on the strength of spatial dependences (ρ , λ) as shown in Figure 4.2. For all parameter estimates except that of ρ , increasing the average number of neighbors increased the RMSE noticeably at the following two combinations of spatial dependence parameters – (a) Extreme negative values of ρ and high positive λ , and (b) Small absolute values of λ and high positive ρ . However, at small absolute values of λ and extreme negative values of ρ , the RMSE actually decreased. The estimates of ρ conditional on *W* behaved slightly differently. Increasing density marginally increased the RMSE at small ρ irrespective of λ while drastically decreased at extreme negative value of ρ (except at high positive λ).

The experiments with different number of average neighbors revealed that as the structure becomes denser, the bias in parameter estimates increases many folds. The effect is more pronounced as the spatial autocorrelation in the dependent variable and error structure increases. An increase in the sample size consistently and greatly reduces the bias due to increase in neighborhood density. Thus, in a large sample the increase in bias due to denser neighborhood structure is marginal. The result for estimates of γ_2 for different values of sample size and degree of spatial dependences are shown in Figure 4.3. Estimates of other model parameters behaved in similar fashion.

Simulations suggest that the choice of neighborhood structure should not only involve goodness of fit criteria but also concern for increased bias in parameter estimates due to denser neighborhood structure.

λ_i	ρ_i	λ_1	γ_1	β_1	β_3	λ_1	γ_1	β_1	β_3
FGS3SLS		Bias				Variance			
-0.8	-0.8			-0.001	-0.006				0.001
-0.8	-0.4			-0.001	0.002				0.002
-0.8	0.0			-0.003	0.023				0.002
-0.8	0.2		0.001	-0.003	0.039				0.003
-0.8	0.6	-0.001	0.001	-0.008	0.106				0.009
-0.4	-0.8	-0.004	0.002	-0.014	0.249				0.033
-0.4	-0.4	-0.005	0.003	-0.016	0.301				0.047
-0.4	0.0	-0.006	0.004	-0.019	0.360				0.049
-0.4	0.2	-0.005	0.004	-0.019	0.380				0.043
-0.4	0.6	-0.009	0.007	-0.031	0.652			0.001	0.107
0.0	-0.8	-0.002	0.011	-0.009	0.328				0.108
0.0	-0.4	-0.001	0.012	-0.008	0.384				0.142
0.0	0.0	-0.001	0.010	-0.005	0.386				0.157
0.0	0.2	-0.003	0.010	-0.005	0.390				0.120
0.0	0.6	-0.019	0.013	-0.001	0.624	0.002			0.199
0.2	-0.8	-0.001	0.010	-0.006	0.232	0.001			0.152
0.2	-0.4	-0.001	0.007	-0.006	0.177	0.001			0.133
0.2	0.0	-0.001	0.006	-0.005	0.143	0.001			0.105
0.2	0.2	-0.005	0.005	-0.002	0.131	0.001			0.099
0.2	0.6	-0.029	0.004	0.012	0.098	0.004		0.001	0.089
0.6	-0.8	-0.005	0.003	-0.001	0.028	0.001			0.077
0.6	-0.4	-0.005	0.002		0.013				0.083
0.6	0.0	-0.003	0.001		-0.021				0.059
0.6	0.2	-0.005	0.001	0.002	-0.037				0.051
0.6	0.6	-0.013		0.010	-0.187	0.001			0.063
Gains over FGS2SLS		Reduciton in Absolute Bias				Reduction in Variance			
-0.8	-0.8			0.001	0.010				
-0.8	-0.4		0.001	0.002	0.027				
-0.8	0.0	0.001	0.001	0.003	0.042				0.001
-0.8	0.2	0.001		0.005	0.055				0.002
-0.8	0.6	0.002	0.002	0.010	0.118				0.010
-0.4	-0.8	0.007	0.003	0.018	0.250				0.053
-0.4	-0.4	0.007	0.003	0.019	0.264				0.068
-0.4	0.0	0.007	0.003	0.019	0.286				0.063
-0.4	0.2	0.009	0.004	0.019	0.299				0.075
-0.4	0.6	0.012	0.006	0.023	0.394				0.208
0.0	-0.8	0.002	0.014	-0.002	0.521				0.130
0.0	-0.4	0.003	0.014	-0.004	0.523				0.145
0.0	0.0	0.005	0.016	-0.004	0.582				0.126
0.0	0.2	0.008	0.015	-0.004	0.553				0.160
0.0	0.6	0.029	0.017	0.013	0.659				0.197
0.2	-0.8	-0.001	0.015	-0.005	0.536				0.201
0.2	-0.4	0.005	0.013	-0.005	0.489				0.129
0.2	0.0	0.014	0.011		0.513				0.095
0.2	0.2	0.017	0.010	0.008	0.508				0.101
0.2	0.6	0.049	0.011	0.031	0.643				0.116
0.6	-0.8	-0.003	0.005	0.004	0.258			0.001	0.027
0.6	-0.4		0.004	0.002	0.236	0.001			0.008
0.6	0.0	0.003	0.003		0.191				0.010
0.6	0.2	0.004	0.003		0.168				0.006
0.6	0.6	0.017	0.002	0.011	-0.057			0.001	0.002

Table 4.1: FGS3SLS Bias and Variances for n=250, Σ_2*, W=6,* $\gamma_1 = -0.3$*,* $\beta_1 = 2.0$*,* $\beta_3 = 2.5$

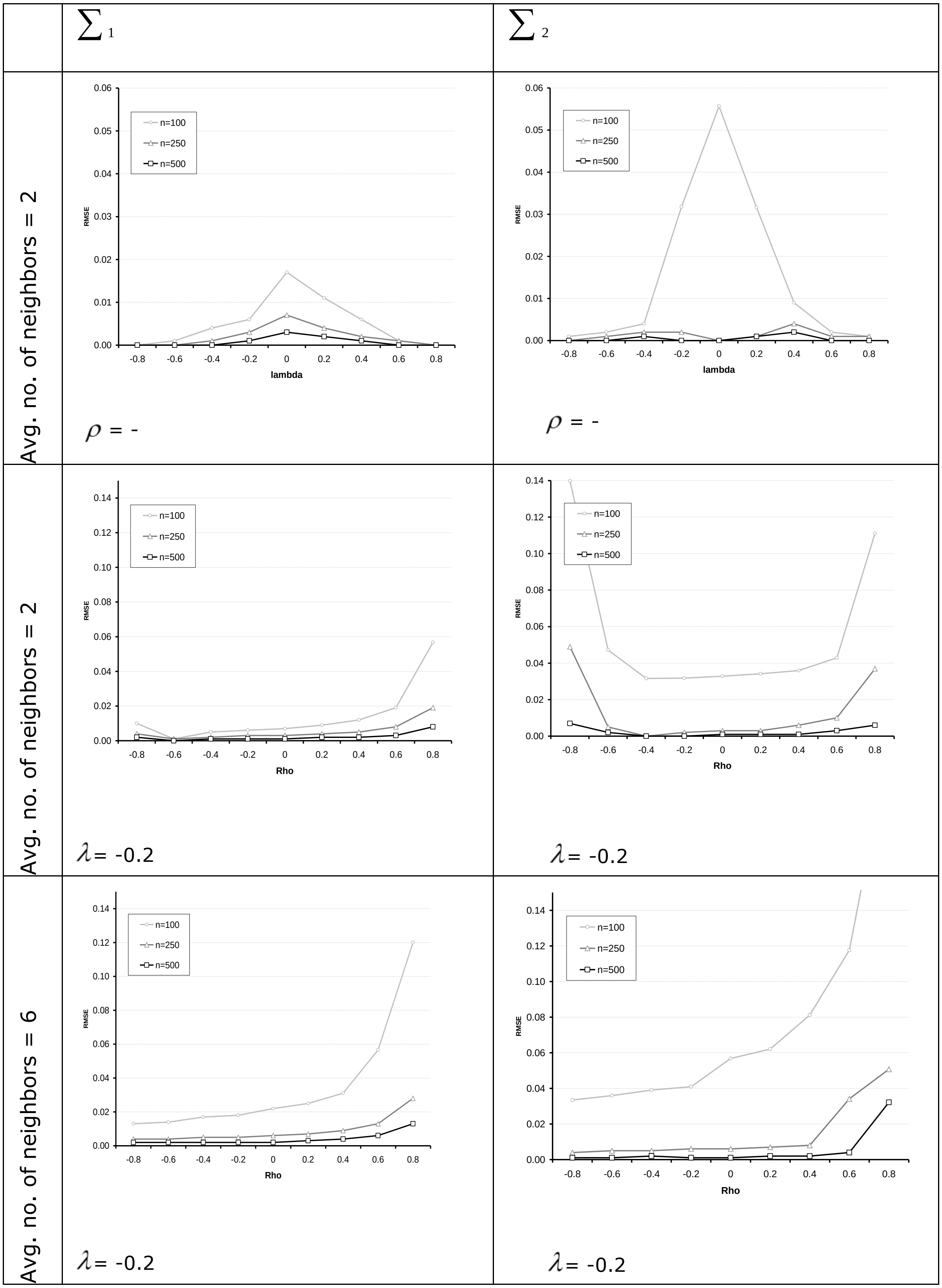


Figure 4.1: Impact of sample size on RMSE of γ_2

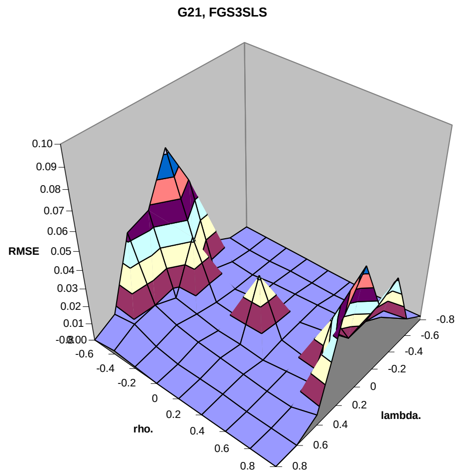
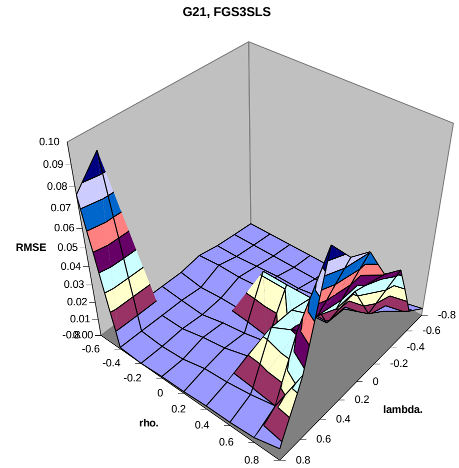
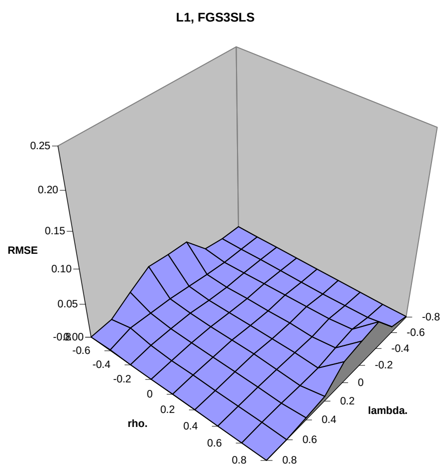
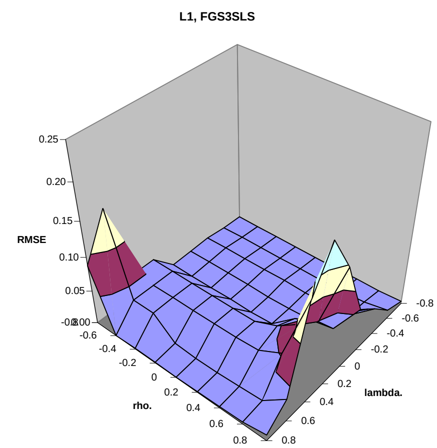
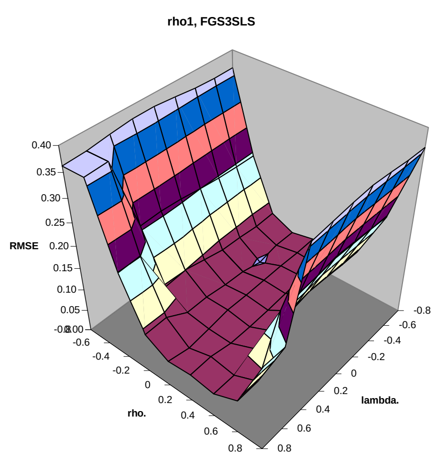
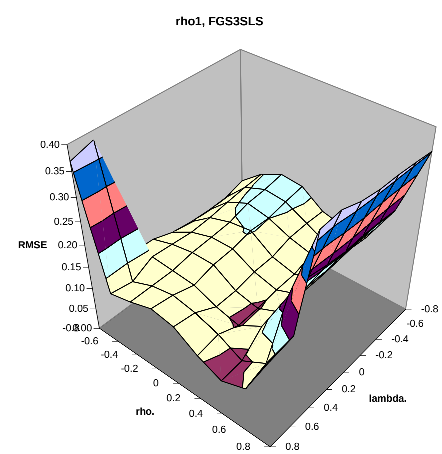
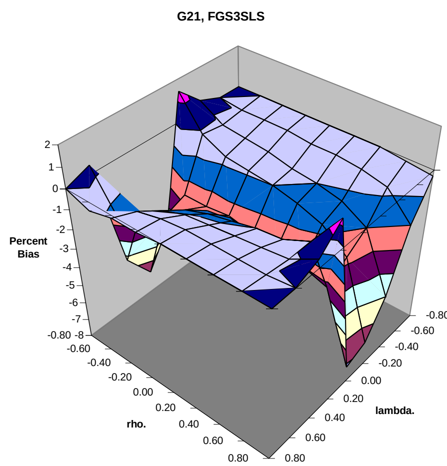
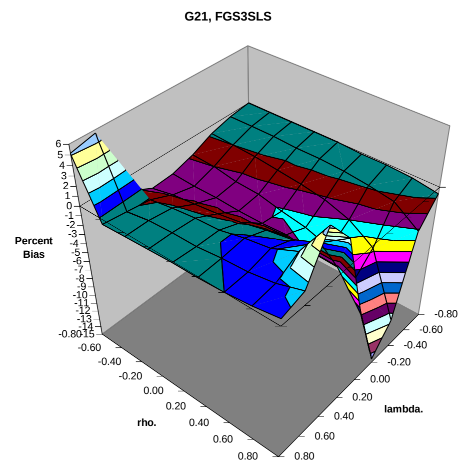
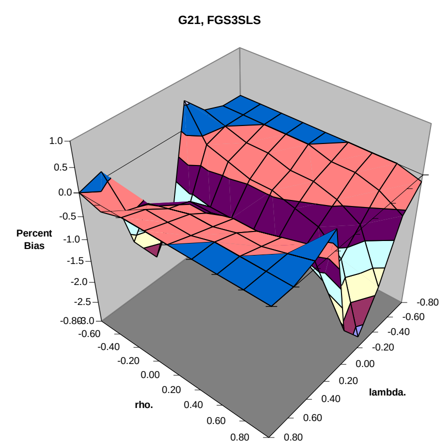
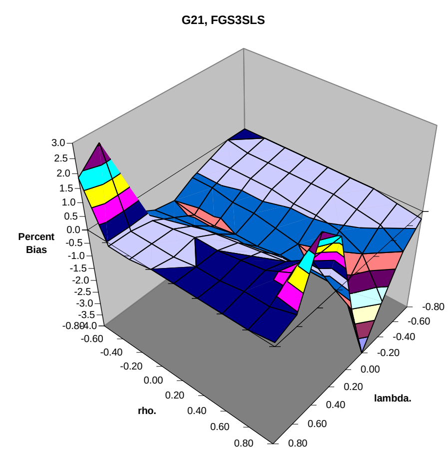
<table>
<tr><th></th><th>Average number of neighbors=2</th><th>Average number of neighbors=6</th></tr>
<tr><td>γ_2</td><td>
</td><td>
</td></tr>
<tr><td>λ</td><td>
</td><td>
</td></tr>
<tr><td>ρ</td><td>
</td><td>
</td></tr>
</table>

Figure 4.2: Impact of average number of neighbors on RMSE for Σ_2 and n = 250

<table>
<tr><td></td><td>Average number of neighbors=2</td><td>Average number of neighbors=6</td></tr>
<tr><td>N=

100</td><td>
</td><td>
</td></tr>
<tr><td>N=

250</td><td>
</td><td>
</td></tr>
<tr><td>N=

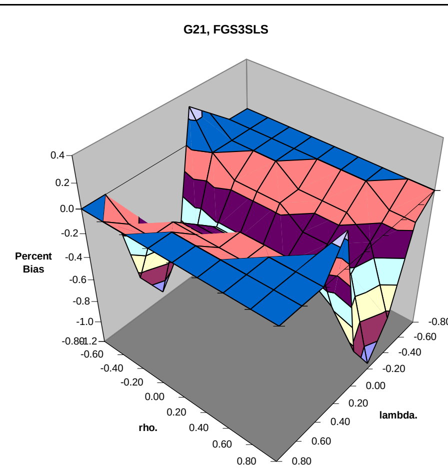
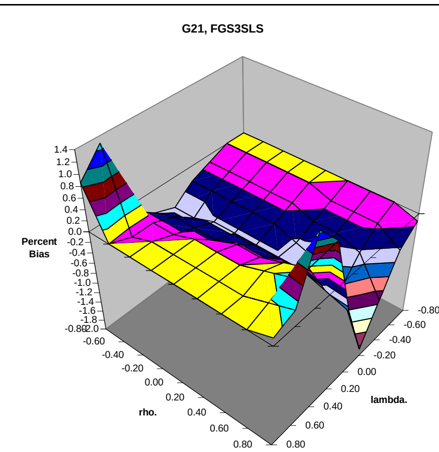
500</td><td>
</td><td>
</td></tr>
</table>

Figure 4.3: Impact of average no. of neighbors on percentage bias of $\gamma_2(=0.7)$ *for* Σ_1

Estimates of λ and ρ

The bias in the estimate of the spatial autocorrelation parameter in error terms ρ was analyzed under different sample sizes, variance-covariance structure and weight matrices conditional on different values of the spatial lag parameter λ. Similar analysis was done for the estimates of the spatial lag parameter λ conditional on ρ (Figure 4.4). The estimator does not provide a direct way to calculate the variance of ρ and therefore, it was derived computationally. One point of caution is that the estimation of ρ requires an optimization procedure where the objective function may not be well defined and is susceptible to choice of starting parameters. This was not found to be the case in our experiments as the results were stable with respect to the choice of starting parameters. However, it is a concern to be borne in mind while using the estimator.

Estimates of λ were very robust to varying degree of spatial dependences over most of the (ρ , λ) space. As the neighborhood density increases, there is an increase in the bias and is mostly independent of ρ it is conditioned upon. The estimator performs well at low and moderate degree of spatial dependencies in endogenous variables except when there is a simultaneous presence of a very high spatial dependence in randomness. Surprisingly, higher bias in the parameter estimate of λ was accompanied by higher variances, signifying the poor performance of the estimator in such conditions.

The bias and variance of ρ was largely independent of the values of λ it was conditioned upon except at very high values of λ. The bias increased very rapidly when its true parameter value increased from -0.8 to +0.8. However, unlike λ, there was a clear bias variance trade off in the estimates of ρ.

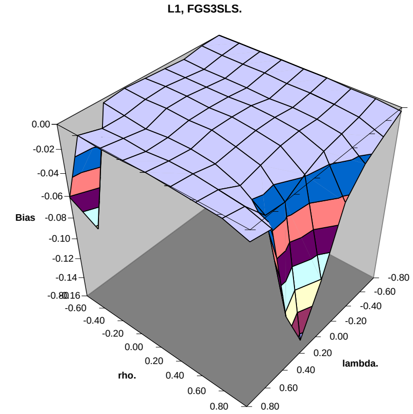
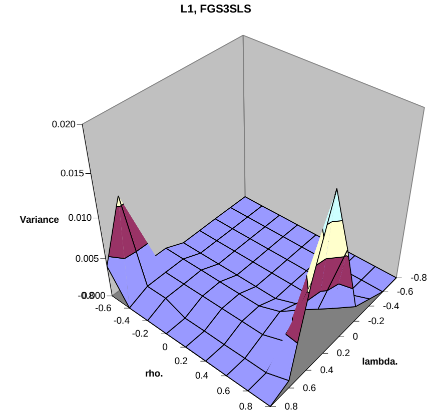
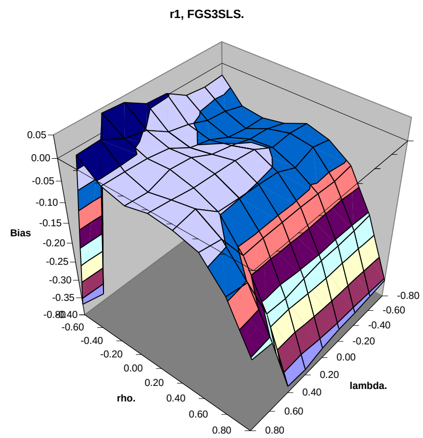
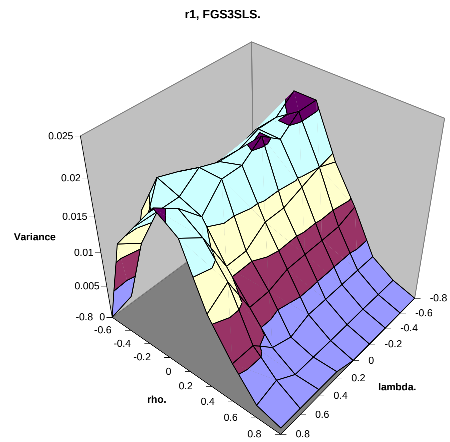
<table>
<tr><td></td><td>Bias</td><td>Variance</td></tr>
<tr><td>λ</td><td>
</td><td>
</td></tr>
<tr><td>ρ</td><td>
</td><td>
</td></tr>
</table>

Figure 4.4: Estimates of λ and ρ, at n=250 for Σ_2, average number of neighbors = 6

5. Conclusion

In this chapter, I analyzed the small sample properties of the limited information Feasible Generalized Spatial Two Stage Least Squares (FGS2SLS) and the full information Feasible Generalized Spatial Three Stage Least Squares (FGS3SLS) estimator for a system of simultaneous equations with spatial dependence in error terms and in the dependent variable. Given relatively few published applications of this estimator and lack of theoretical understanding about its behavior in small samples, this chapter provides a starting point for analyzing the behavior of this estimator. A Monte Carlo framework was used to explore the impacts of sample size, neighborhood structure, variance co-variance matrix and varying degree of spatial dependence parameters on the estimators' performance.

The FGS3SLS estimator performed better than the FGS2SLS estimator in terms of smaller bias and lower variance. The gains of using the former are higher when the spatial correlation in disturbance terms is high, the spatial lag parameter has low absolute value, the sample size is small and the neighborhood structure is dense. Given superiority of the FGS3SLS estimator over the FGS2SLS in the simulations described in this chapter, the detailed study of the impacts of sample size, neighborhood structure, variance-covariance matrix and degree of spatial dependence on estimator's behavior was made limited to the FGS3SLS estimator.

The performance of the FGS3SLS estimator drastically increased when the sample size was increased from 100 to 250 observations. Increasing the sample size to 500 observations yielded only marginal gains. Gains with increasing sample size are more significant when the heterogeneity is high, the spatial lag parameter of the dependent variable is small, the spatial autocorrelation in errors is high, the number of neighbors is large and the variance covariance structure of error is large. The performance of the estimator was found to be sensitive to the value of the spatial dependence parameters. It was worse at low values of spatial lag parameter in dependent variable (λ) and at extreme high values of the spatial dependence in the error structure (ρ). Thus, the estimator pays a premium in terms of bias and variance when the spatial lag is small but has huge gains as the spatial lag increases. The estimator for λ performed well at low and moderate degree of spatial dependencies in endogenous variables except when there is a simultaneous presence of a very high spatial dependence in randomness. Spatial structures with higher average number of neighbors led to higher bias and variances in the estimates. The effect is more pronounced as the spatial autocorrelation in the dependent variable and error structure increases. In large samples the increase in bias due to denser neighborhood structure is marginal. The results presented here are sensitive to the model specification, choice of the data generation process, distribution of the exogenous variable, etc. However, the results are useful in comparative exercise to assess the relative changes in performance under different conditions and should not be taken as an absolute measure of performance.

Understanding the impacts of the sample size, varying degree of spatial dependencies, neighborhood structure and the error structure on the forecasted value is essential. However, given its magnitude of the work in analyzing the forecasts of a spatial data and comparing with that of its true values, it is not discussed in this chapter.

Additional research is needed in order to make this estimator a commonplace in applied work. It is computationally intensive and there is no software or standard code to implement this estimator. Efforts in this direction are very much warranted. A useful extension would be to analyze the impact of increasing model complexity and choice of instruments on the performance of the estimator. Further, the estimates of ρ are obtained from an optimization routine, where the objective function may have multiple optimums. In such cases, the parameter estimate of ρ may be susceptible to the choice of starting values and various techniques may be indeed to insure that a global optimum is reached. This makes the task more computationally demanding. Work is needed to theoretically corroborate the findings of this chapter in a generalized framework. Over the last five decades, we have learnt a lot about the properties of the three stage least squares estimator in terms of impacts of misspecification, nonlinearity, multicollinearity, etc., many of which have been studied through Monte Carlo simulations. A parallel series of literature needs to be developed for the Feasible Generalized Three Stage Least Square estimator.

CHAPTER 5: CONCLUSION

Population change and economic growth are not the only determinants of the number of new housing units constructed in a region. A small increase in population has been associated with a large amount of new housing units through vacancy and abandonment of housing units. Sprawl, largely caused by declining urban core and intra-regional migration, leads to more residential growth in the periphery than is needed to accommodate the increase in the number of households. Further, the location of these new households is not merely based on normative concepts of optimizing travel time to work or maintaining job-housing ratios but also on perceptions of the social environment held by households.

This dissertation explores answers to the question of 'how much to grow and where to grow' in the context of land-use change, highlighting the impact of social dynamics on land-use change and the interdependence between the two. Regional forecasting of housing demand based on macro-level models often ignores the micro-level social dynamics (e.g. intra-regional migration, urban decline and sprawl) that are an important driver for new housing construction; on the other hand, micro-level drivers alone cannot reflect the economy-wide impacts on housing demand. Modeling these interdependencies provides a richer interpretation of the underlying social, economic and spatial dynamics driving land-use change in regions than is currently available. Thus, it bridges the gap between regional economic models, social dynamics and land use change to provide planners and policymakers with a more substantial knowledge base on which to deliberate about the region and its future.

The research described in this dissertation was presented as three essays. The first essay explored how the spatial pattern of socio-economic characteristics determined the magnitude and location of growth and is also shaped by it. It presented a framework for modeling the relationship among regional inequalities, urban distress and growth. The framework was tested using data from the St. Louis region. The second essay used this framework to model the impact of the social dynamics on location and magnitude of growth and decline in the region. The forecasts from this model, together with population and income forecasts from a regional economic model, were used to derive probabilities of development and decline at the Census tract level and to also assess the total magnitude of growth endogenously. The final essay evaluated the performance of the full information feasible generalized spatial three stages least square (FGS3SLS) estimator used to estimate the model presented in the second essay. Using Monte Carlo experiments, the sensitivity of results to varying degrees of spatial dependences, choice of spatial weight matrix, sample size and variance covariance matrices is analyzed. These Monte Carlo simulations provided confidence in the results of the social dynamics model.

Limitations and Improvements

There are several limitations to the research presented in this dissertation. Some of these, along with suggestions for addressing them, are included below.

(1) The research uses a reduced-form of model. While this is useful for making predictions, the true dynamics for scenario analysis can only be captured by a full-form model. Inferences about causation and correlation must be made with considerable caution while using the proposed model in the dissertation. Nevertheless, since the model is based on a detailed causal map, it remains useful for forecasting and scenario analysis. Crime, age of infrastructure, land value, etc. have been omitted, largely because of lack

of data. There is a high degree of collinearity among income, race and school quality. These two issues necessitate the need for a more in-depth study at a micro level.

(2) Relationships between the variables used in this analysis are assumed to be linear in absence of any theoretical guidance on the functional form to be used. These relationships were further assumed to be constant throughout their distribution but in reality may drastically differ at different points in the distribution. For example, households living in poor income tracts would likely respond more strongly to marginal changes in school quality than those living in richer income tracts. Incorporating this information inside a system of simultaneous equations framework is a methodological challenge. It has been tried in multivariate single-equation setup (Koenker, & Bassett, 1978) and its extensions to system of equations with spatial externalities will be useful in our context.

(3)The proposed model uses population forecast from the regional economic model to drive the amount of land-use change rather than changes in number of households or household size. Between 1990 and 2000, the contribution of population to increase in number of households was same as that of decrease in household size. Decreases in household size and changes in family structure are critical factors affecting land-use change. The issue becomes complex as the household size varies with income, Census tract and race. Methodological improvements in modeling this will be critical.

(4) Explicit modeling of the effect of race is methodologically and substantively problematic. It may be kept out of the purview of land-use change models for substantive and technical reasons. The racial composition of new households in the region is not known. To incorporate race, one will have to decompose households within a tract by racial groups and then separately model their residential choice behavior for

each group. However, the residential choice behavior is not independently governed by race either. It depends on the age of the head of the householder, income group and household composition. Alternatively, a cohort component model for population broken by race could be used, but the issue of converting population to number of households by race would still remain. We have very little theoretical understanding of dynamics at this level of detail and very little data to model such intricacies.

(5) The social model was estimated using the FGS3SLS estimator, whose small sample properties are not well understood. Understanding the impacts of sample size, varying degree of spatial dependencies, neighborhood structure and the error structure on the forecasted values of the social model is essential. However, given its magnitude of the work in analyzing the forecasts of a spatial data and comparing with that of its true values, it is not discussed in this dissertation. The choice of a spatial structure for neighborhood spillover matrices, which are assumed a priori, must be given special attention. A further desired methodological extension for FGS3SLS estimator is to theoretically establish the finite sample properties under low level of regularity conditions given its immense usefulness in varied contexts.

Extensions

This dissertation provides a starting step aiming towards a symbiotic relationship between methods of regional science (Isard et al., 1998) and the planning support systems literature (Brail & Klosterman, 2001, Geertman & Stillwell, 2009). The former has focused largely on macro-level issues while the later has concerned itself with micro-level details in the context of land-use change models, assuming the macro-level changes to be exogenously provided. The two can be integrated by expanding the regional econometric input-output model (Israilevich, Hewings, Sonis, Schindler, 1997) and including in it a Social Accounting Matrix with disaggregated households and labor supply. The behavior of these households (income, expenditures, level of skill, etc.), their number and location in the region will be jointly determined, with equilibrium established by intra-regional migration. This, at the minimum, would supply much needed information to better integrate socio-economic dynamics, urban decline, regional housing supply and demand and economic growth at regional and micro levels. A more difficult but eventually necessary step from the perspective of policy analysis would be to model the impact of micro-level policy decisions (land-use control and transformation, local government fragmentation, etc.) that eventually affects the macro-level competitiveness of the region and vice versa (Figure 5.1). This is critical from a planner's perspective.

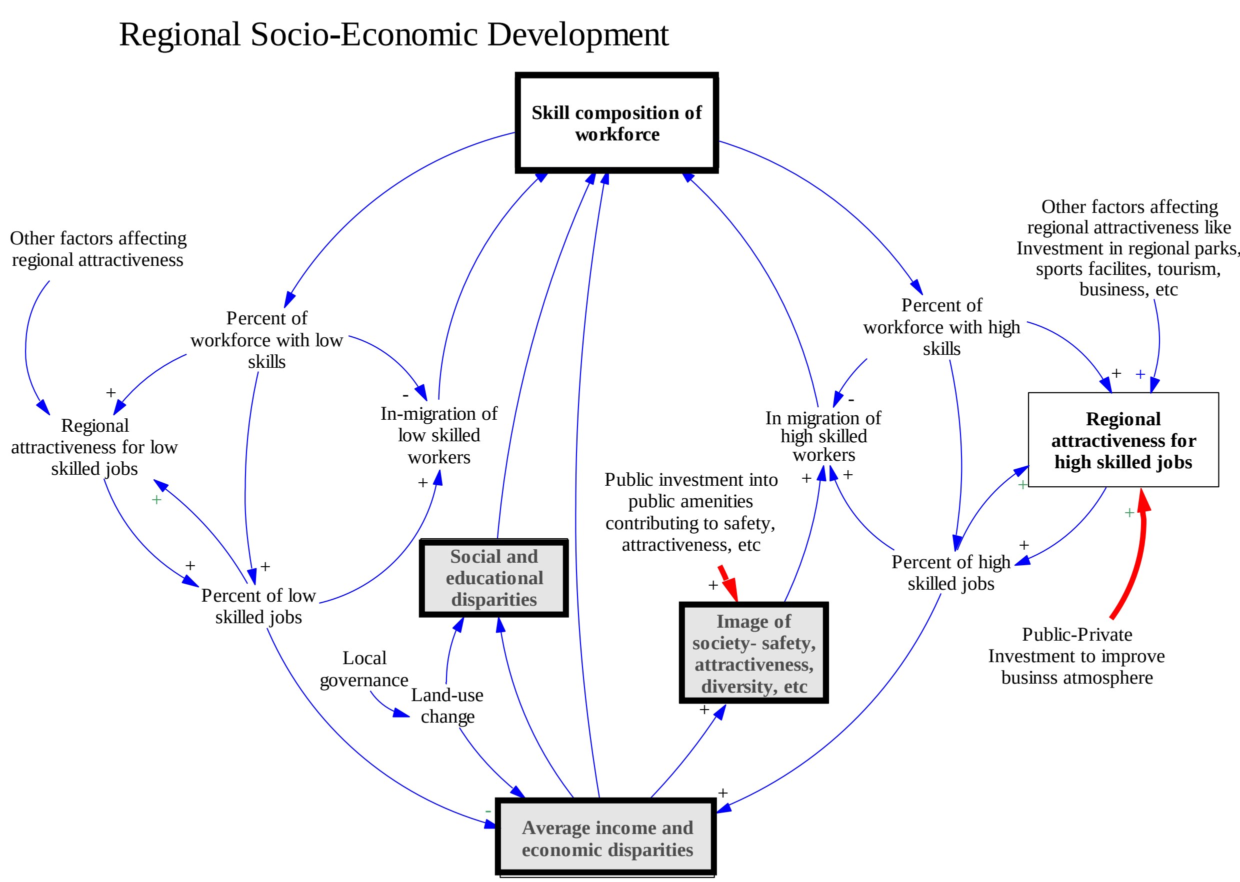


Figure 5.1: Relationship between socio-economic structure and the regional economy through changing availability and skill composition of work force and household consumption by income groups over time.

REFERENCES

<Chapter 1>

Davidson, R., & MacKinnon, J. G. (1993). *Estimation and inference in econometrics*. New York, NY: Oxford University Press.

Deal, B. (2001). Ecological urban dynamics: The convergence of spatial modeling and sustainability. *The Journal of Building Research and Information, 29(5),* 381-393.

Deal, B., & Pallathucheril, V. G. (2008). Simulating regional futures: The land-use evolution and impact model (LEAM). In R. K. Brail (Ed.), *Planning Support Systems for Cities and Regions (*pp. 61–8). Cambridge, MA: Lincoln Institute of Land Policy.

Kelejian, H. H., & Prucha, I. R. (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations, *Journal of Econometrics, 118*, 27-50.

<Chapter 2>

Abbot, A. (1997). Of time and space: The contemporary relevance of the Chicago school. *Social Forces, 75,* 1149–1182.

Agarwal, C., Green, G. M., Grove, J. M., Evans, T. P., & Schweik, C. M. (2002). *A review and assessment of land-use change models: Dynamics of space, time, and human choice.* TR NE-297, US Department of Agriculture, Forest Service, Northeast Research Station, Newton Square, PA.

Akerlof, G. A. (1997). Social distance and social decisions. *Econometrica*, *65(5),* 1005-1027.

Anas, A. (1982). *Residential location markets and urban transportation economic theory, econometrics and policy analysis with discrete choice models*. Oxford: Academic Press.

Anas, A. (1983). *The Chicago area transportation and land use analysis system*. Urban and Regional Planning Program, Department of Civil Engineering, Northeastern University.

Anselin, L. (1995). Local indicators of spatial association — LISA. *Geographical Analysis, 27,* 93–115.

Anselin, L. (1996). The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In M. Fischer, H. Scholten, & D. Unwin (Eds.), *Spatial analytical perspectives on GIS in environmental and socio-economic sciences* (pp. 111–125). London: Taylor and Francis.

Anselin, L. (1998). Exploratory spatial data analysis in a geocomputational environment. In P. A. Longley, S. M. Brooks, R. McDonnell, & B. Macmillan (Eds.), *Geocomputation, a primer* (pp. 77–94). New York, NY: Wiley.

Anselin, L. (1999). Interactive techniques and exploratory spatial data analysis. In P. A. Longley, M. F. Goodchild, D. J. Maguire, & D. W. Rhind (Eds.), *Geographical information systems: Principles, techniques, management and applications* (pp. 251–264). New York, NY: Wiley.

Anselin, L., Syabri, I., & Kho, Y. (2006). GeoDa: An introduction to spatial data analysis. *Geographical Analysis, 38(1),* 5–22.

Baer, W. C., & Williamson, C. B. (1988). The filtering of households and housing units. *Journal of Planning Literature, 3,* 127-152.

Bailey, T. C., & Gatrell, A. C. (1995). *Interactive spatial data analysis*. New York, NY: John Wiley and Sons.

Barra, T de la. (1989). *Integrated land use and transport modeling*. Oxford: Cambridge University Press.

Barra, T de la. (2001). Integrated land use and transport modeling: The TRANUS experience. In R. K. Brail & R. E. Klosterman (Eds.), *Planning support systems:*

Integrating geographic information systems, models and visualization tools (pp. 129-156). Redlands, CA: ESRI Press.

Batty, M. (1976). *Urban modeling: Algorithms, calibrations, predictions*. Cambridge, MA: Cambridge University Press.

Bivand, R. with contributions by Luc Anselin, Olaf Berke, Andrew Bernat, Marilia Carvalho, Yongwan Chun, Carsten Dormann, Stéphane Dray, Rein Halbersma, Nicholas Lewin-Koh, Hisaji Ono, Pedro Peres-Neto, Michael Tiefelsdorf and Danlin Yu. (2006). *spdep: Spatial dependence: weighting schemes, statistics and models*. R package version 0.3-25.

Boarnet, M. G. (1994). An empirical model of intermetropolitan population and employment growth. *Papers in Regional Science, 73,* 135–152.

Brail, R. K., & Klosterman, R. E. (Eds.). (2001). *Planning support systems: Integrating geographic information systems, models, and visualization tools.* Redlands, CA: ESRI Press.

Briassoulis, H. (2000). Analysis of land use change: Theoretical and modeling approaches. In S. Loveridge (Ed.), *The web book of regional science*. Morgantown, WV: West Virginia University.

Brown, L. A., & Chung, S. (2006). Spatial segregation, segregation indices and the geographical perspective. *Population Space and Place, 12,* 125-143.

Buja, A., Cook, D., & Swayne, D.F. (1996). Interactive high-dimensional data visualization. *Journal of Computational and Graphical Statistics, 5,* 78-99.

Cameron, T. A., & McConnaha, I. (2006). *Evidence of environmental migration: Housing values alone may not capture the full effects of local environmental disamenities*. University of Oregon Economics Department Working Papers, 2010-7. University of Oregon Economics Department.

Carter, W. H., Schill, M. H., & Watcher, S. M. (1998). Polarisation, public housing and racial minorities in US cities. *Urban Studies, 35(10),* 1889-911.

Case, A. C. (1992). Neighborhood influence and technological change. *Regional Science and Urban Economics, 22(3),* 491-508.

Case, A. C., Rosen, H. S., & Hines, J. R. (1993). Budget spillovers and fiscal policy independence: Evidence from the states. *Journal of Public Economics, 52,* 285–307.

Charles, C. Z. (2003). The dynamics of racial residential segregation. *Annual Review of Sociology, 29,* 167-207.

Chen, K., Irwin, E. G., Jayaprakash, C., & Warren, K. (2005). The emergence of racial segregation in an agent-based model of residential location: The role of competing preferences. *Computational and Mathematical Organization Theory, 11(4),* 333-338.

Clark, R. (1992). *Neighborhood effects on dropping out of school among teenage boys*. Discussion Paper PSC-DSC-UI-13, Urban Institute, Washington, DC.

Cleveland, W. S. (1993). *Visualizing data*. Summit, NJ: Hobart Press.

Cutler, D. M., Glaeser, E. L., & Vigdor, J. L. (1999). The rise and decline of the American ghetto. *Journal of Political Economy, 107(3),* 455-506.

Deal, B., & Pallathucheril, V. G. (2007). Developing and using scenarios. In L. D. Hopkins & M. A. Zapata (Eds.), *Engaging the future: Forecasts, scenarios, plans, and projects* (pp. 221-241). Cambridge, MA: Lincoln Institute of Land Policy.

Deal, B., & Pallathucheril, V. G. (2008). Simulating regional futures: The land-use evolution and impact model (LEAM). In R. K. Brail (Ed.), *Planning support systems for cities and regions (*pp. 61–8). Cambridge, MA: Lincoln Institute of Land Policy.

Echenique, M. H., Anthony, D. J., Flowerdew, J., Hunt, D., Mayo, T.R., Skidmore, I.J., & Simmonds, D.C. (1990). The MEPLAN models of Bilbao, Leeds and Dortmund. *Transport Reviews, 10,* 309-322.

Focus St. Louis. (2001). *Racial inequality in the St. Louis region: A community call to action*. FOCUS St. Louis, St. Louis, MO. Available from http://www.focusstl.org/ProgramsInitiatives/InfluencingPolicy/RacialEqualitySocialJustice/ACommunityCalltoAction.aspx.

Galster, G. C. (1990). Racial steering in housing markets during the 1980s: A review of the audit evidence. *Journal of Planning Education and Research, 9,* 165–75.

Galster, G. C. (1992). Research on discrimination in housing and mortgage markets: Assessment and future directions. *Housing Policy Debate, 3(2),* 639–83.

Goodchild, M. F., Anselin, L., Appelbaum, R. P., & Harthorn, B. H. (2000). Toward spatially integrated social science. *International Regional Science Review, 23(2),* 139–59.

Haining, R. (1990). *Spatial data analysis in the social and environmental sciences*. Cambridge, MA: Cambridge University Press.

Harris, D. R. (1997). *The flight of whites: A multilevel analysis of why whites move.* Population Studies Center Report no. 97-386. University of Michigan.

Kelejian, H. H., & Robinson, D. P. (1992). Spatial autocorrelation: A new computationally simple test with an application to per capita county police expenditures. *Regional Science and Urban Economics, 22,* 317–333.

Landis, J. (1994). The California urban futures model: A new generation of metropolitan simulation models. *Environment and Planning B, 21,* 399-420.

Landis, J. (1995). Imagining land use futures: Applying the California urban futures model. *Journal of the American Planning Association, 61(4),* 438-457.

Lee, S., & Leigh, N. G. (2007). Intrametropolitan spatial differentiation and decline of inner-ring suburbs: A comparison of four U.S. metropolitan areas. *Journal of Planning Education and Research, 27,* 146-164.

Massey, D. S., & Denton, N. A. (1993). *American apartheid: Segregation and the making of the underclass*. Cambridge, MA: Harvard University Press.

Messner, S. F., & Anselin, L. (2004). Spatial analyses of homicide with areal data. In M. Goodchild & D. Janelle (Eds*.), Spatially integrated social science* (pp. 127-144). New York, NY: Oxford University Press.

Messner, S. F., Anselin, L., Baller, R. D., Hawkins, D. F., Deane, G., & Tolnay. S. E. (1999). The spatial patterning of county homicide rates: An application of exploratory spatial data analysis. *Journal of Quantitative Criminology, 15(4),* 423-450.

Meyer, S. G. (2000). *As long as they don't live next door: Segregation and racial conflict in American neighborhoods.* New York, NY: Rowman and Littlefield.

Pallathucheril, V., & Deal, B. (2005). Modeling to link forecasts, scenarios, and plans. Proceedings from *Symposium on Envisioning and Creating Urban Futures: Using Forecasts, Scenarios, and Plans.* Cambridge, MA: Lincoln Land Institute

Parker, D. C., Manson, S. M., Janssen, M. A., Hoffman, M. J., & Deadman, P. (2003). Multi-agent systems for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers, 93,* 314-337.

Putman, S. H. (1983). *Integrated urban models*. London: Pion.

Putman, S. H. (1991). *Integrated urban models 2: New research and applications of optimization and dynamics*. London: Pion.

Quercia, R. G., & Galster, G. C. (2000). Threshold effects and neighborhood change. *Journal of Planning Education and Research, 20, 146-162.*

R Development Core Team. (2004). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, Available from http://www.R-project.org.

Rounsevell, M. D. A. (Ed.). (1999). *Spatial modeling of the response and adaptation of soils and land use systems to climate change – An integrated model to predict European land use (IMPEL).* Research report for the European commission, framework IV programme, environment and climate. Contract Nos. ENV4-CT95-0114 and IC20-CT96-0013.

Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing neighborhood effect: Social processes and new directions in research. *Annual Review of Sociology, 28,* 443-478.

Sarraf, S., & Bendor, T. (2005). *Effective public investment in the St. Louis region: More for your money*. Gateway Blueprint Model Workshop. St. Louis, MO: Unpublished.

Savoie, J. (2008). *Neighbourhood characteristics and the distribution of crime: Edmonton, Halifax and Thunder Bay*. Crime and Justice Research Paper Series. Statistics Canada Catalogue no. 85-561-MIE No. 10. Ottawa.

Spelman, W. (1993). Abandoned buildings: Magnets for crime? *Journal of Criminal Justice, 21(5),* 481-495.

Steinitz, C. (1990). A framework for theory applicable to the education of landscape architects (and other environmental design professionals). *Landscape Journal*, 9(2), 136-143.

Tatian, P. A. (2002). *CensusCD neighborhood change data base (NCDB): Data users' guide short form release*. Washington, DC: The Urban Institute.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison–Wesley.

US EPA. (2000). *Projecting land-use change: a summary of models for assessing the effects of community growth and change on land-use patterns.* EPA/600/R-00/098, US Environmental Protection Agency, Office of Research and Development. Cincinnati, OH.

Vartanian, T. P. (1999). Adolescent neighborhood effects on labor market and economic outcomes. *Social Service Review, 73(2),* 142-167.

Veldkamp, A., & Fresco, L.O. (1996). CLUE-CR: An integrated multi-scale model to simulate land use change scenarios in Costa Rica*. Ecological Modelling, 91(1-3),* 231-248.

Waddell, P. (2002). UrbanSim: Modeling urban development for land use, transportation and environmental planning. *Journal of the American Planning Association, 68,* 297–314.

Wallace, M., Wisener, M., & Collins, K. (2006). *Neighbourhood characteristics and the distribution of crime in Regina*. Crime and Justice Research Paper Series. Statistics Canada Catalogue no. 85-561-MIE. No. 8. Ottawa.

Wegener, M. (1982). Modeling urban decline: A multi-level economic-demographic model for the Dortmund region. *International Regional Science Review, 7(2),* 217-241.

Wegener, M. (1985). The Dortmund housing market model: A Monte Carlo simulation of a regional housing market. In K. Stahk (Ed.), *Microeconomic models of housing markets,* Lecture notes in economic and mathematical systems 239 (pp. 144-191). Berlin: Springer Verlag.

Wegener, M. (1986a). Transport network equilibrium and regional deconcentration. *Environment and Planning A, 18,* 437-456.

Wegener, M. (1986b). Integrated forecasting models of urban and regional systems. In . P. W. J. Batey & M. Madden (Eds.), *Integrated analysis of regional systems* (pp. 9-24). London: Pion.

Wegener, M. (1994). Operational urban models: State-of-the-art. *Journal of the American Planning Association, 60(1),* 17-29.

White, R., & Engelen, G. (1994). Cellular dynamics and GIS: Modeling spatial complexity. *Geographical Systems, 1(3),* 237-253.

Wilson, A. G. (1974). *Urban and regional models in geography and planning*. New York, NY: John Wiley.

<Chapter 3>

Anselin, L. (2002). Under the Hood: Issues in the specification and interpretation of spatial regression models. *Agricultural Economics, 27(3),* 247-67.

Anselin, L., & Bera, A. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah & D. E. Giles (Eds.), *Handbook of applied economic statistics* (pp. 237– 289). New York, NY: Marcel Dekker.

Briassoulis, H. (2000). Analysis of land use change: Theoretical and modeling approaches. In S. Loveridge (Ed.), *The web book of regional science*. Morgantown, WV: West Virginia University.

Cliff, A., & Ord, J. (1973). *Spatial autocorrelation*. London: Pion.

Cliff, A., & Ord, J. (1981). *Spatial processes, models and applications*. London: Pion.

Deal, B. (2001). Ecological urban dynamics: The convergence of spatial modeling and sustainability. *The Journal of Building Research and Information, 29(5),* 381-393.

Deal, B. (2003). *Sustainable land-use planning: The integration of process and technology*. (Doctoral dissertation). Department of Urban and Regional Planning. University of Illinois, Urbana Champaign.

Deal, B., & Pallathucheril, V. G. (2008). Simulating regional futures: The land-use evolution and impact model (LEAM). In R K Brail (Ed.), *Planning support systems for cities and regions* (pp 61–8). Cambridge, MA: Lincoln Institute of Land Policy.

Kelejian, H. H., & Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, *40,* 509-533.

Kelejian, H. H., & Prucha, I. R., (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations, *Journal of Econometrics, 118*, 27-50.

R Development Core Team. (2004). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, Available from http://www.R-project.org.

Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing neighborhood effect: Social processes and new directions in research. *Annual Review of Sociology, 28,* 443-478.

Tatian, P. A. (2002). *CensusCD neighborhood change data base (NCDB): Data users' guide short form release*. Washington, DC: The Urban Institute.

<Chapter 4>

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Boston, MA: Kluwer Academic Publishers.

Carter, R. A. L., & Nagar, A. L. (1997). Coefficients of correlation for simultaneous equation systems. *Journal of Econometrics, 6,* 39–50.

Case, A. (1991). Spatial patterns in household demand. *Econometrica, 59,* 953–966.

Case, A., Rosen, H. S., & Hines, J. R. (1993). Budget spillovers and fiscal policy independence: Evidence from the states. *Journal of Public Economics, 52,* 285–307.

Cliff, A., & Ord, J. (1973). *Spatial autocorrelation*. London: Pion.

Cliff, A., & Ord, J. (1981). *Spatial processes, models and applications.* London: Pion.

Das, D., Kelejian, H. H., & Prucha, I. R. (2003). Small sample properties of estimators of spatial autoregressive models with autoregressive disturbances. *Papers in Regional Science, 82*, 1-26.

Davidson, R., & MacKinnon, J. G. (1993). *Estimation and inference in econometrics*. New York, NY: Oxford University Press.

Driffield, N. (2006). On the search for spillovers from foreign direct investment (FDI) with spatial dependency. *Regional Studies, 40(1),* 107-119.

Fishback, P. V., Horrace, W. C., & Kantor, S. (2006). The impact of New Deal expenditures on mobility during the great depression. *Explorations in Economic History, 43(2),* 179-222.

Goodchild, M., Anselin, L., Appelbaum, R., & Harthorn, B. (2000). Toward spatially integrated social science. *International Regional Science Review, 23,* 139–159.

Kelejian, H. H. & Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review, 40,* 509-533.

Kelejian, H. H., Prucha, I. R. (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations. *Journal of Econometrics, 118,* 27-50.

Ngeleza, G. K., Florax, R. J. G. M., & Masters, W. A. (2006). *Geographic and institutional determinants of real income: A spatio-temporal simultaneous equation approach.* Working Papers 06-15, Purdue University, College of Agriculture, Department of Agricultural Economics.

Phillips, P. C. P. (1983). Exact small sample theory in the simultaneous equations model. In Z. Griliches & M. Intriligator (Eds.), *Handbook of Econometrics, Vol. 1* (pp. 441-516). New York, NY: North Holland.

R Development Core Team. (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, Available from http://www.R-project.org.

Rothenberg, T. J. (1984). Hypothesis testing in linear models when the error covariance matrix is nonscalar. *Econometrica, 52,* 827–842.

<Chapter 5>

Brail, R. K., & Klosterman, R. E. (Eds.). (2001). Planning support systems: *Integrating geographic information systems, models, and visualization tools*. Redlands, CA: ESRI Press.

Geertman, S., & Stillwell, J. (Eds.). (2009). *Planning support systems: New methods and best practice (Advances in spatial science)*. New York, NY: Springer Publishers.

Isard, W., Azis, I., Drennan, M., Miller, R., Saltzman, S., & Thorbecke, E. (Eds.). (1998). *Methods of interregional and regional analysis.* U.K.: Aldershot.

Israilevich, P. R., Hewings, G. J. D., Sonis, M., & Schindler, G. R. (1997). Forecasting structural change with a regional econometric input-output model. *Journal of Regional Science, 37,* 565–90.

Koenker, R. W. and Bassett, G. Jr. (1978), Regression quantiles. *Econometrica*, 46, 33–50.

APPENDIX A: EXPLORATORY DATA ANALYSIS

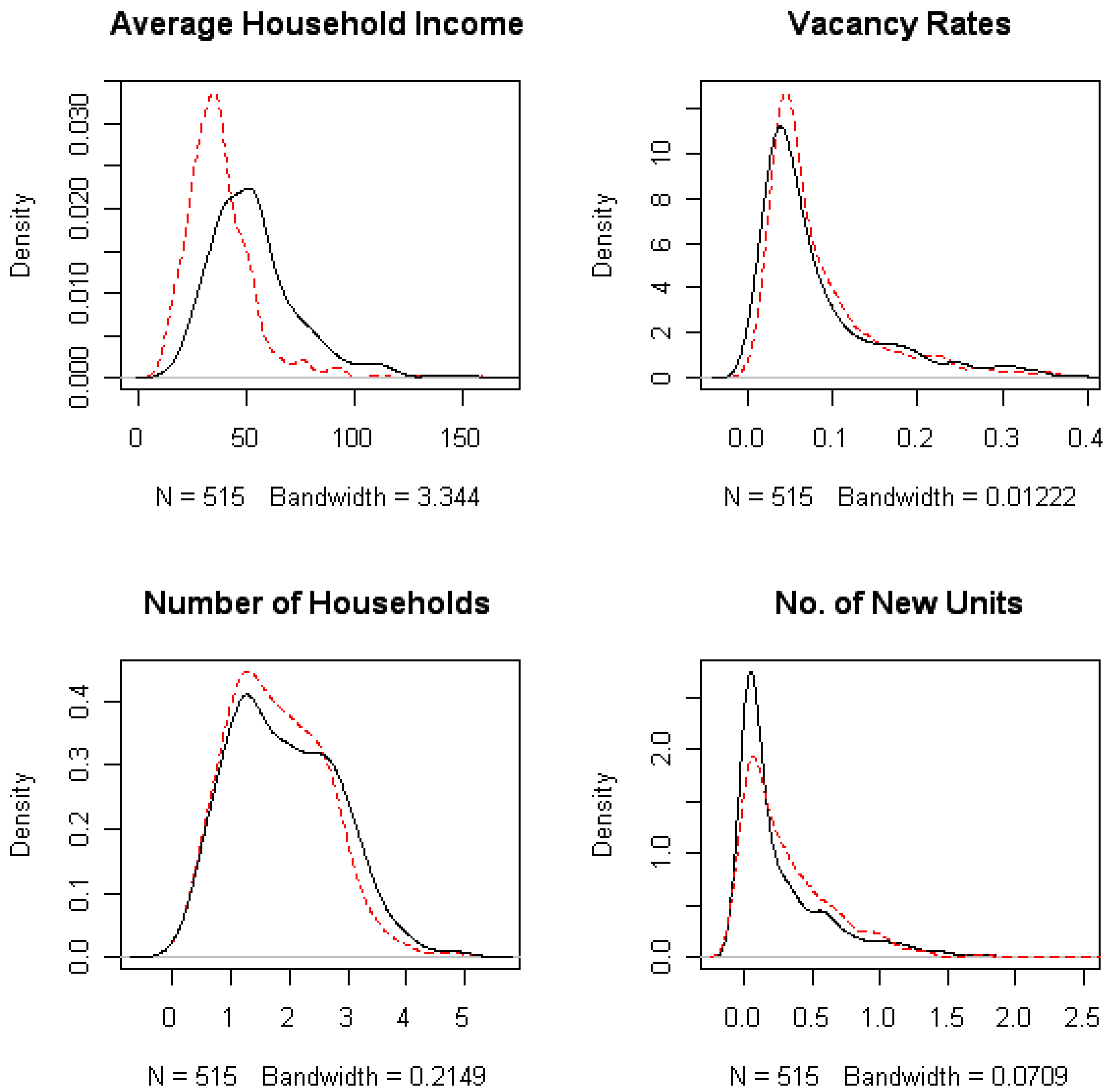


Figure A.1: Density distribution of key variables in 1990 (black solid line) and 2000 (red dotted lines)

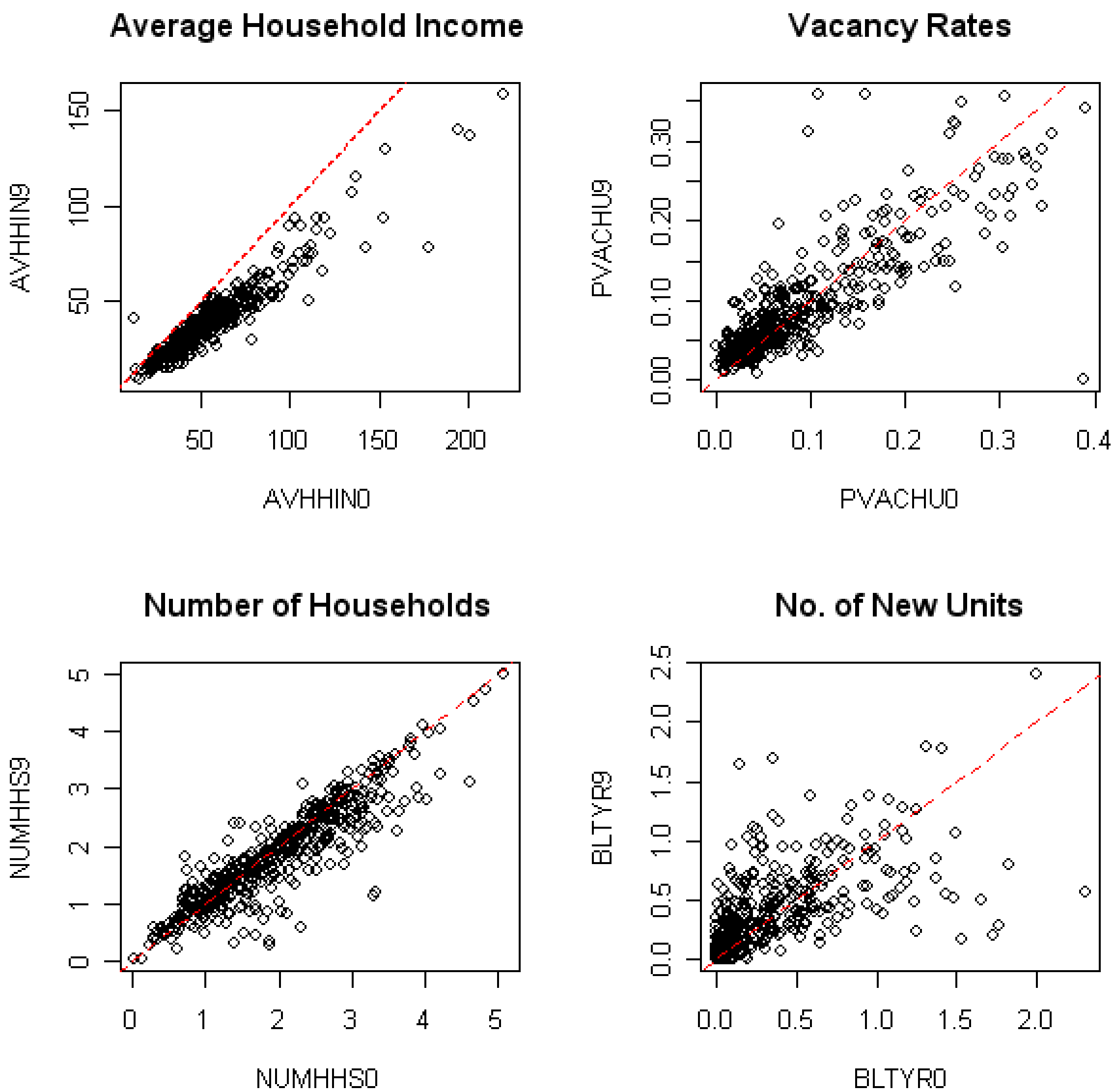


Figure A.2: Changes in key variables between 1990 and 2000

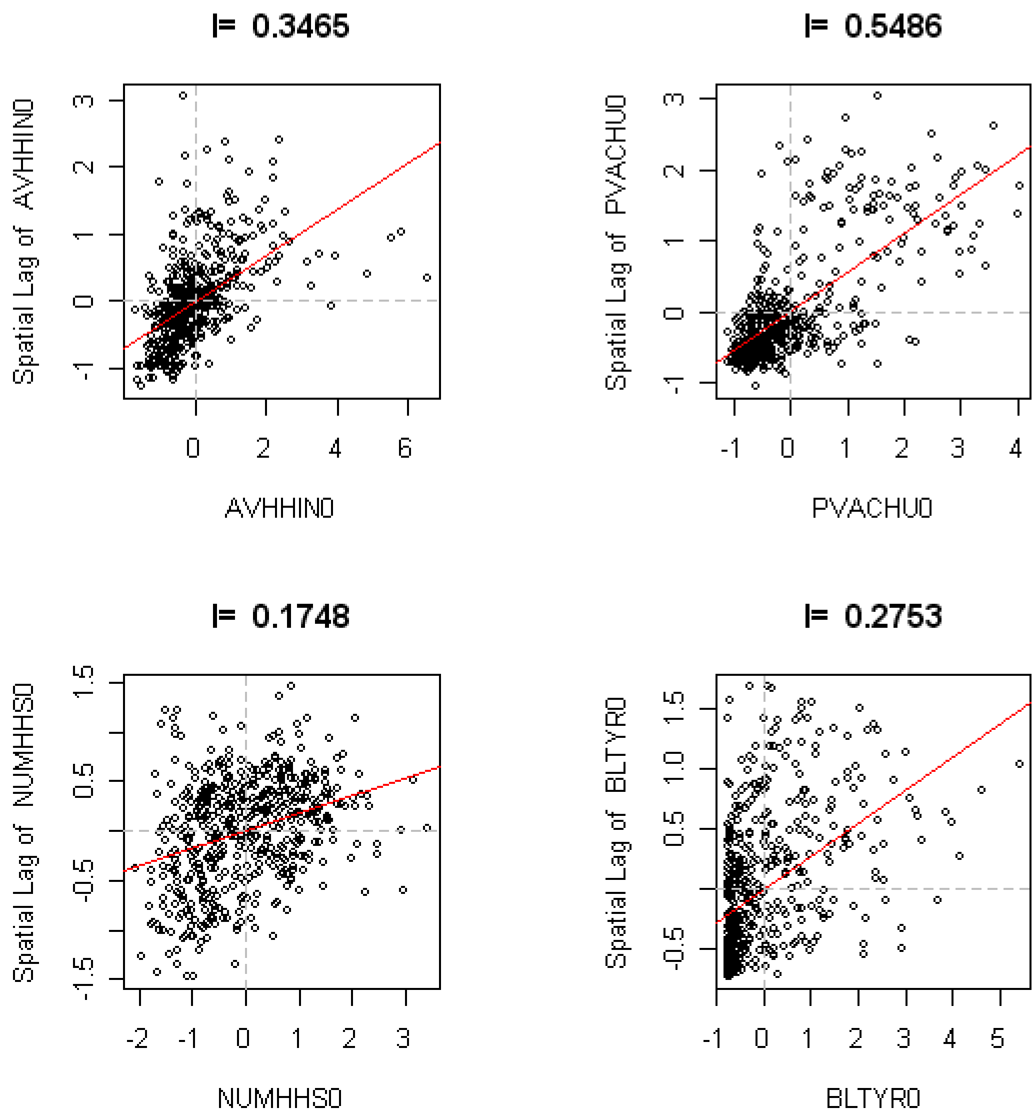


Figure A.3: Spatial correlation of key variables in 2000 with its value in neighboring tracts in 2000

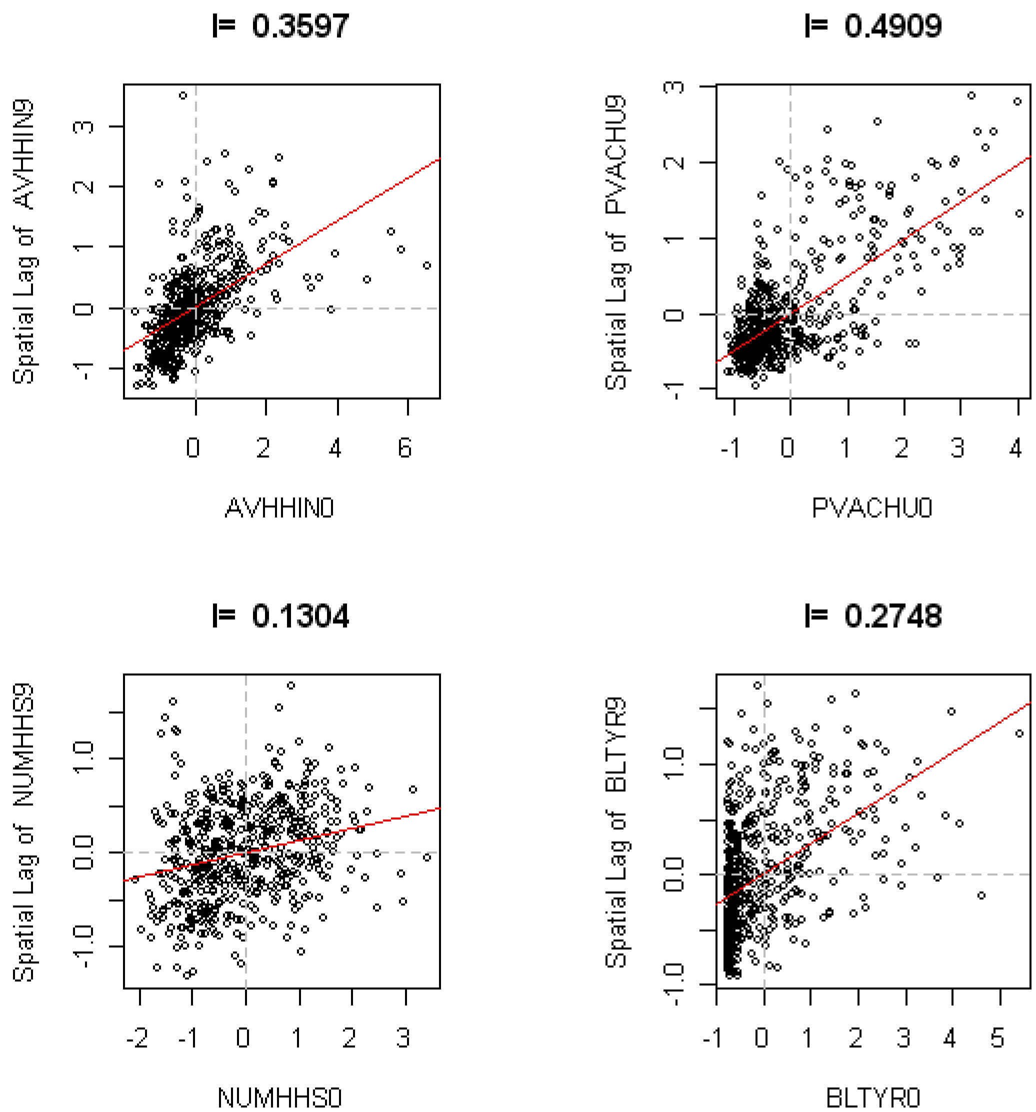


Figure A.4: Spatial correlation of key variables in 2000 with its value in neighboring tracts in 1990

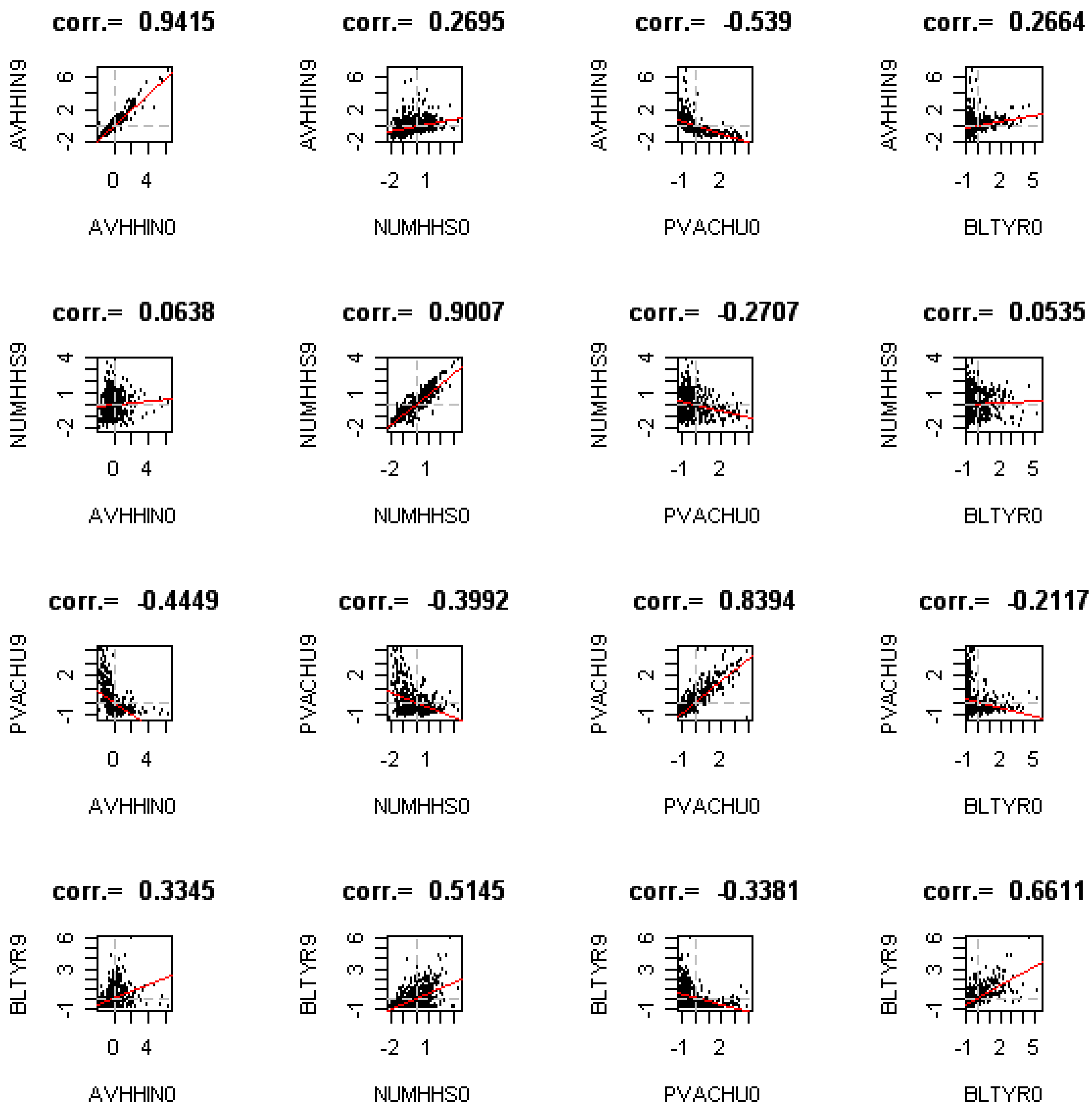


Figure A.5: Correlation matrix and interrelationship of key variables in 1990 and 2000

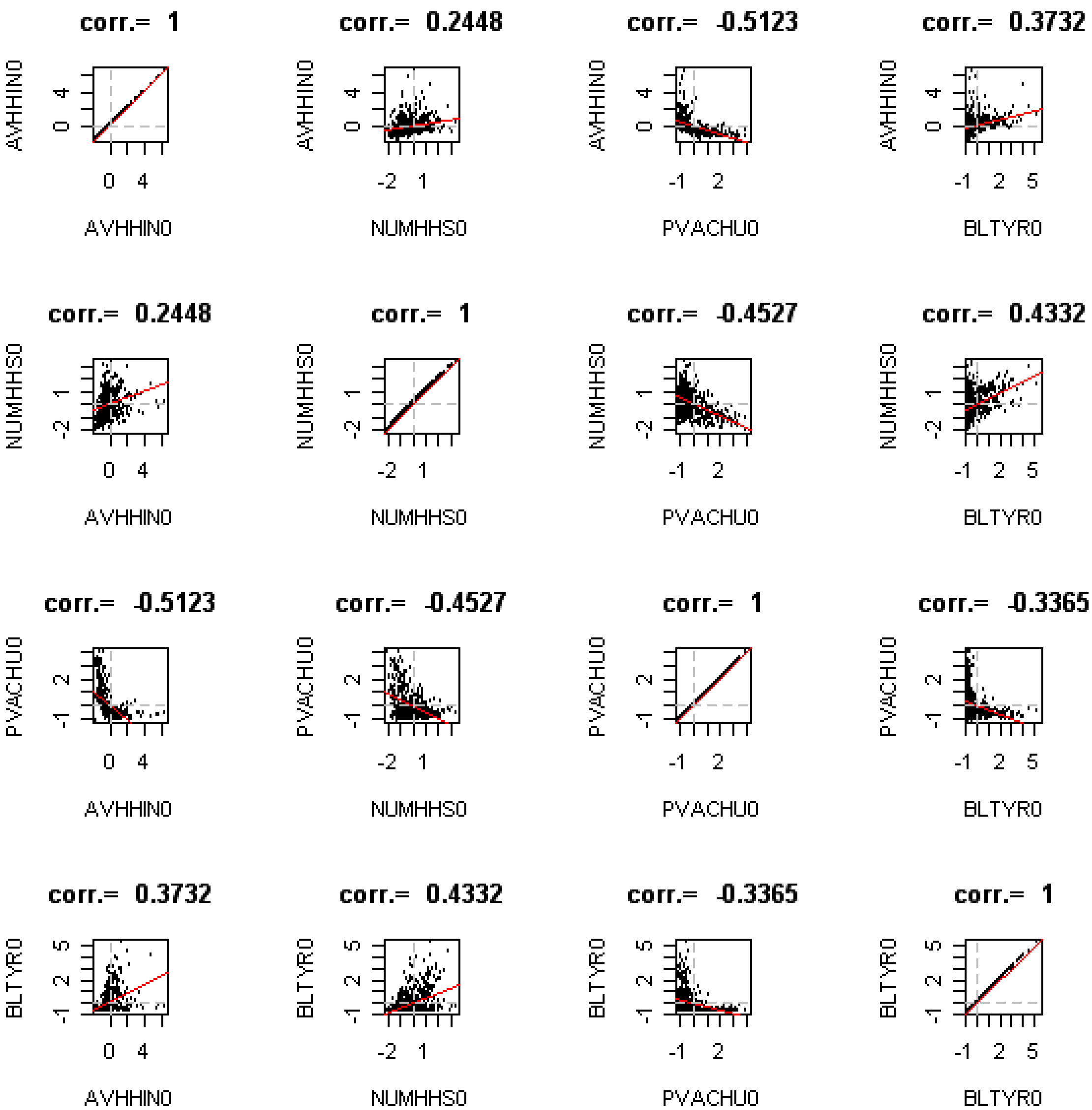


Figure A.6: Correlation matrix and interrelationship of key variables in 1990 and 2000

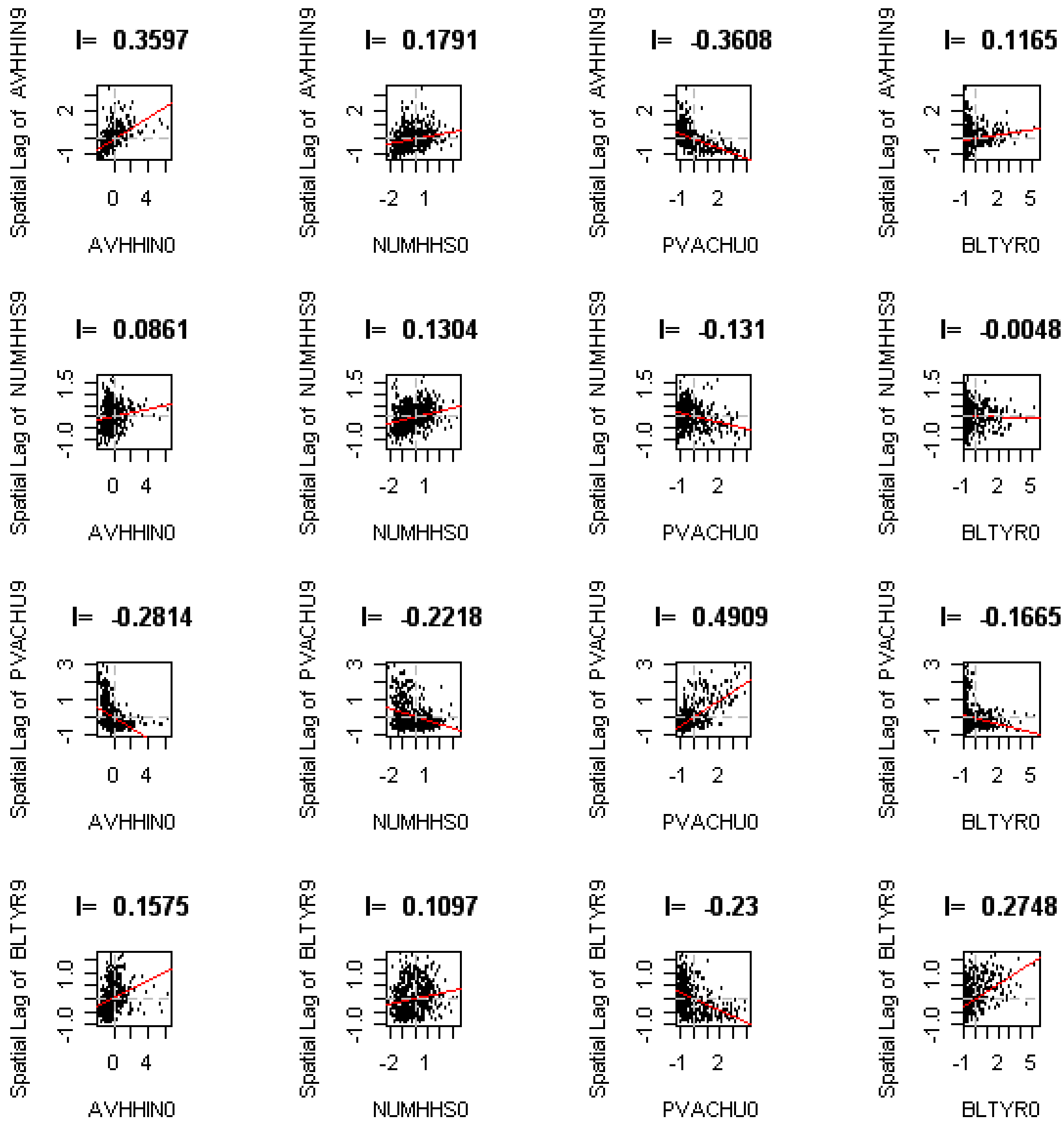


Figure A.7: Spatial correlation matrix and interrelationship of key variables in 2000 with their neighbors in 1990

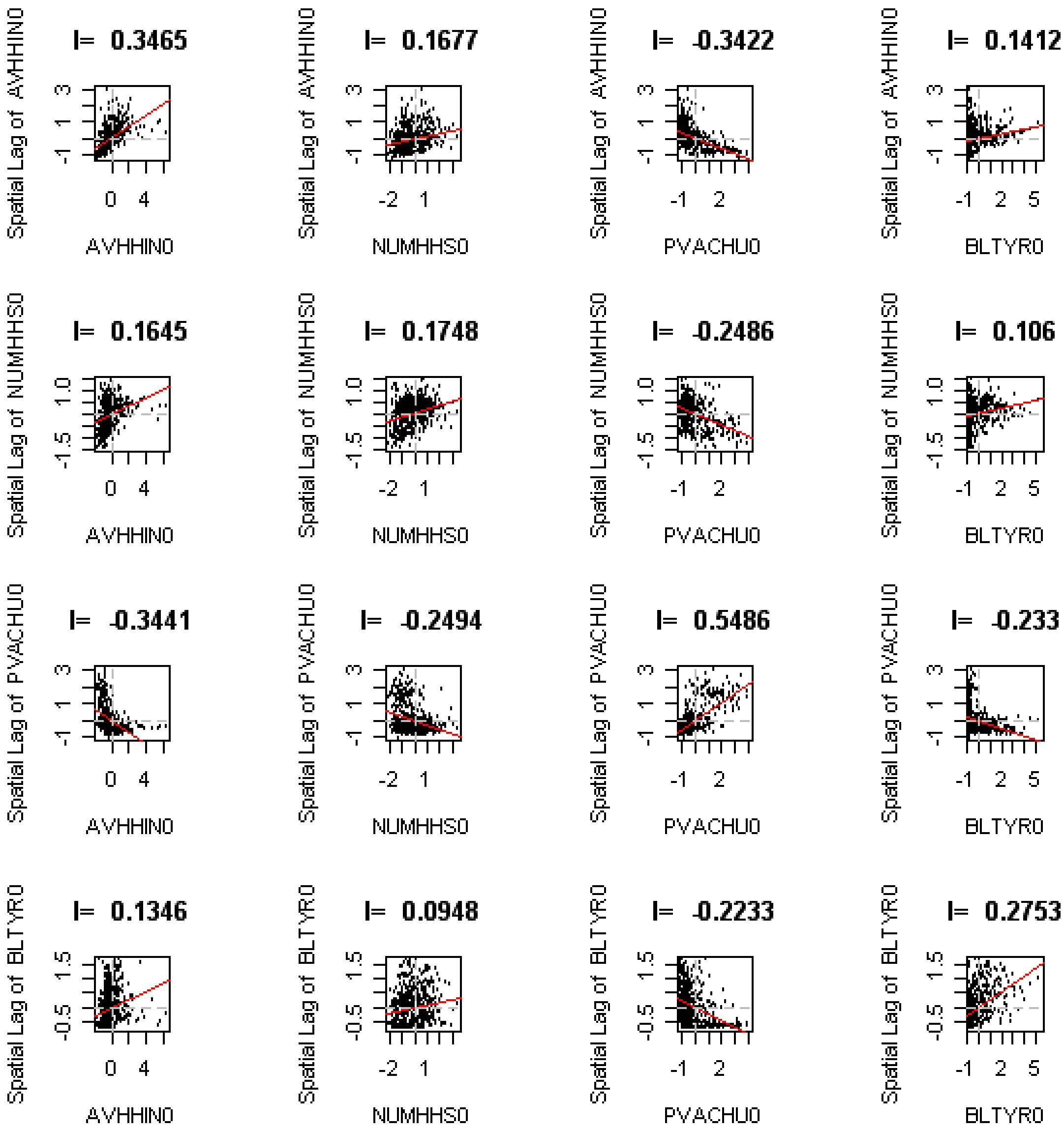


Figure A.8: Spatial correlation matrix and interrelationship of key variables in 2000 with their neighbors in 2000

APPENDIX B: SOCIAL MODEL CODE AND DIAGNOSTICS

```
library(quantreg)
library(foreign)
library(boot)
library(MASS)
library(spdep)
library(nlme)
rm(list=ls(all=TRUE))
set.seed(100)
par(mfcol=c(2,2))
q <- seq(0,250,0.01)

#MYR0 median year built
#MHV median Housing value
#PUH percent urban
#POHU percent occupied, percent vacant
#PNWHH percent non white HH
#PRHU percent rented
#TAX total tax/ total value of owner occupied

school1 <- data.frame(read.csv("school_processed.csv"))
school2 <- data.frame(read.csv("school_processed1.csv"))
cen2k <- data.frame(read.csv("cen2k_processed.csv"))
tax <- data.frame(read.csv("tax.csv"))
school <- merge(school1, school2)
cen <- merge(school,cen2k)
all <- merge(cen,tax)

#kelejian and Prucha 2004
gFit <- function(vpredict, vobserved){
        pseudoR2 <- cor(vpredict,vobserved)^2
        pseudoR2
}

read.geoda <- function(file,row.names=NULL) {
read.csv(file=file,header=TRUE,skip=1,row.names=row.names) }
stdata <- read.geoda("forgeoda5.txt")
stdata2 <- merge(stdata,all)

polgal <- read.gal("q6.GAL", override.id=TRUE)
colqueen <- nb2listw(polgal)
nmat <- nb2mat(polgal)
attach(stdata2)

#DHU0 <- TOTHSUN0-TOTHSUN9
#DHU9 <- TOTHSUN9-TOTHSUN8
BLTYR0 <- ifelse(BLTYR0-DHU0<0,DHU0,BLTYR0)
BLTYR9 <- ifelse(BLTYR9-DHU9<0,DHU9,BLTYR9)
ABHU0 <- BLTYR0-DHU0
ABHU9 <- BLTYR9-DHU9
ABHUR0A <- ifelse(TOTHSUN9==0,0,ABHU0/TOTHSUN9)
ABHUR9A <- ifelse(TOTHSUN8==0,0,ABHU9/TOTHSUN8)
```

```
n <- 515 # sample size
m <- 4  # number of equations
k <- 7  # number of exogenous variables
W <- nmat

PVACHU9D1 <- ifelse(PVACHU9>0.15, 1, 0)
PVACHU9D2 <- ifelse(PVACHU9<0.10, 1, 0)
ABHUR0A <- ifelse(ABHUR0A>0.66, 0.66, ABHUR0A)
ABHUR9A <- ifelse(ABHUR9A>0.66, 0.66, ABHUR9A)
BLTYR9Dm <- ifelse(BLTYR9<quantile(BLTYR9,0.05),1,0)
BLTYR9Dma <- ifelse(BLTYR9<quantile(BLTYR9,0.1),1,0)
BLTYR9Dmb <- ifelse(BLTYR9>quantile(BLTYR9,0.97),1,0)
PVACHU9Dm <- ifelse(PVACHU9>quantile(PVACHU9,0.9),1,0)

y1 <- (AVHHIN0)
y2 <- (NUMHHS0)
y2a <-(NUMHHS0-NUMHHS9)
y2b <- NUMHHS0/(AREA*1000)
y3 <- BLTYR0
y4 <- (PVACHU0)
Y <- cbind(y1,y2,y3,y4)
wy1 <- W %*% y1; wy2<-W %*% y2; wy3<-W %*% y3; wy4<-W %*% y4

x1 <- (AVHHIN9)
x12 <- (AVHHIN9^2)
Wx1 <- W%*%x1
x13 <- 1000/AVHHIN9
x2 <-(NUMHHS9)
x2b <- (NUMHHS9/(AREA*1000))
x3 <- (BLTYR9)
x3a <- 1/(ifelse(x3<0.01,0.01,x3))
x3b <- BLTYR9Dm
x3c <- BLTYR9Dma
x3d <- BLTYR9Dmb
x4 <- (PVACHU9)
x4a <- x4^2
x4b <- PVACHU9Dm
Wx4 <- W%*%x4
x5 <- HSDROP9
x5a <- HSDROP0
x5b <- 1/(ifelse(x5<0.01,0.01,x5))
x6 <- (AREA)
x7 <- ABHU9  #(ABHUR9A*NUMHHS8)
x7a <- ABHUR9A
#x7b<-ABHUR9A^2
x7b <- 1/(ifelse(x7<0.01,0.01,x7))
x8 <- x1*x3
x9 <- TAX0
x10 <- PVACHU9D1
x10a <- PVACHU9D2
#x11<-1/(ifelse(x4<0.01,0.01,x4))

X <- cbind(x1,  x2,x12, x2b, x3,Wx4,x7, x6)
y1n <- "AVHHIN0"
y2n <- "NUMHHS0"
y2an <- "DHH0"
```

```
y2bn <- "DENSITY0"
y3n <- "BLTYR0"
y4n <- "PVACHU0"

x1n <- "AVHHIN9"
Wx1n <- "WAVHHIN9"
x12n <- "AVHHIN9sq"
x13n <- "1/AVHHIN9"
x2n <- "NUMHHS9"
x2bn <- "DENSITY9"
x3n <- "BLTYR9"
x3an <- "1/BLTYR9"
x3bn <- "BLTYR9DMo"
x3cn <- "BLTYR9DMless"
x3dn <- "BLTYR9DMmore"
x4n <- "PVACHU9"
x4an <- "PVACHU9sq"
x4bn <- "PVACHU9Dm"
Wx4n <- "WPVACHU9"
x5n <- "HSDROP9"
x5an <- "HSDROP"
x5bn <- "1/HSDROP9"
x6n <- "AREA"
x7n <- "ABHU9A"
x7an <- "ABHUR9A"
x7bn <- "1/ABHU9A"
x8n <- "INCVAC9"
x9n <- "TAX0"
x10n <- "HighVacancyDummy1"
x10an <- "HealthyVacancy9Dummy"
#x11n <-"1/PVACHU9"

Z1 <- cbind(  y2b,   y3,                        x1,          x7b,     wy1 )
Z2 <- cbind(    y3,                    1, x1, x2,  x5b,             wy2 )
Z3 <- cbind( y2a,y4,                                x3,  x4, x7,       wy3 )
Z4 <- cbind(  y1, y2a,                           x13, x3a, x4,   wy4 )

dZ1 <- ncol(Z1); dZ2 <- ncol(Z2); dZ3 <- ncol(Z3); dZ4 <- ncol(Z4)

colnames(Z1) <- list(  y2bn,    y3n,              x1n,x7bn,         "wy1" )
colnames(Z2) <- list(  y3n,                   "c",   x1n, x2n,  x5bn,        "wy2" )
colnames(Z3) <- list(  y2an, y4n,                           x3n,  x4n, x7n,        "wy3" )
colnames(Z4) <- list(  y1n, y2an,                     x13n, x3an, x4n,     "wy4" )

H <- cbind(X, W %*% X, W %*% (W %*% X))
P <- H %*% solve(crossprod(H,H), t(H))

delta1_til <- solve( crossprod(P %*% Z1, Z1), crossprod( P %*% Z1 , y1))
delta2_til <- solve( crossprod(P %*% Z2, Z2), crossprod( P %*% Z2 , y2))
delta3_til <- solve( crossprod(P %*% Z3, Z3), crossprod( P %*% Z3 , y3))
delta4_til <- solve( crossprod(P %*% Z4, Z4), crossprod( P %*% Z4 , y4))

delta_til <- rbind(delta1_til, delta2_til, delta3_til, delta4_til)
colnames(delta_til) <- list("delta_til")
delta1_til <- delta_til[1:dZ1,]
delta2_til <- delta_til[(dZ1+1):(dZ1+dZ2),]
```

```
delta3_til <- delta_til[(dZ1+dZ2+1):(dZ1+dZ2+dZ3),]
delta4_til <- delta_til[(dZ1+dZ2+dZ3+1):(dZ1+dZ2+dZ3+dZ4),]

u1_til <- y1- Z1 %*% delta1_til
u2_til <- y2- Z2 %*% delta2_til
u3_til <- y3- Z3 %*% delta3_til
u4_til <- y4- Z4 %*% delta4_til

#############
# iterate from here
#############
wu1_til <- W %*% u1_til; wu2_til<- W %*% u2_til; wu3_til<- W %*% u3_til; wu4_til<- W %*% u4_til
wwu1_til <- W %*% wu1_til; wwu2_til<- W %*% wu2_til; wwu3_til<- W %*% wu3_til; wwu4_til<- W
%*% wu4_til
G1 <- array(0,c(3,3)); G2 <- array(0,c(3,3)); G3 <- array(0,c(3,3)); G4 <- array(0,c(3,3))

G1[1,] <- (1/n)*cbind( 2* crossprod(u1_til, wu1_til) ,  -crossprod(wu1_til, wu1_til),   n)
G1[2,] <- (1/n)*cbind( 2* crossprod(wwu1_til, wu1_til), -crossprod(wwu1_til, wwu1_til),
sum(diag(crossprod(W,W))))
G1[3,] <- (1/n)*cbind( crossprod(u1_til, wwu1_til) + crossprod(wu1_til, wu1_til), crossprod(wu1_til,
wwu1_til), 0)
g1 <- (1/n)*t( cbind( crossprod(u1_til, u1_til) , crossprod(wu1_til, wu1_til) , crossprod(u1_til, wu1_til) ))

G2[1,] <- (1/n)*cbind( 2* crossprod(u2_til, wu2_til) ,  -crossprod(wu2_til, wu2_til),   n)
G2[2,] <- (1/n)*cbind( 2* crossprod(wwu2_til, wu2_til), -crossprod(wwu2_til, wwu2_til),
sum(diag(crossprod(W,W))))
G2[3,] <- (1/n)*cbind( crossprod(u2_til, wwu2_til) + crossprod(wu2_til, wu2_til), crossprod(wu2_til,
wwu2_til), 0)
g2 <- (1/n)*t( cbind( crossprod(u2_til, u2_til) , crossprod(wu2_til, wu2_til) , crossprod(u2_til, wu2_til) ))

G3[1,] <- (1/n)*cbind( 2* crossprod(u3_til, wu3_til) ,  -crossprod(wu3_til, wu3_til),   n)
G3[2,] <- (1/n)*cbind( 2* crossprod(wwu3_til, wu3_til), -crossprod(wwu3_til, wwu3_til),
sum(diag(crossprod(W,W))))
G3[3,] <- (1/n)*cbind( crossprod(u3_til, wwu3_til) + crossprod(wu3_til, wu3_til), crossprod(wu3_til,
wwu3_til), 0)
g3 <- (1/n)*t( cbind( crossprod(u3_til, u3_til) , crossprod(wu3_til, wu3_til) , crossprod(u3_til, wu3_til) ))

G4[1,] <- (1/n)*cbind( 2* crossprod(u4_til, wu4_til) ,  -crossprod(wu4_til, wu4_til),   n)
G4[2,] <- (1/n)*cbind( 2* crossprod(wwu4_til, wu4_til), -crossprod(wwu4_til, wwu4_til),
sum(diag(crossprod(W,W))))
G4[3,] <- (1/n)*cbind( crossprod(u4_til, wwu4_til) + crossprod(wu4_til, wu4_til), crossprod(wu4_til,
wwu4_til), 0)
g4 <- (1/n)*t( cbind( crossprod(u4_til, u4_til) , crossprod(wu4_til, wu4_til) , crossprod(u4_til, wu4_til) ))

f1 <- function(alpha1){
        crossprod(g1-G1 %*% t(cbind(alpha1[1], alpha1[1]^2, alpha1[2])), g1-G1 %*%
t(cbind(alpha1[1], alpha1[1]^2, alpha1[2])))  }
rho1_til <- nlm(f1, c(1,1))$estimate[1]
s1_til <- nlm(f1, c(1,1))$estimate[2]

f2 <- function(alpha2){
        crossprod(g2-G2 %*% t(cbind(alpha2[1], alpha2[1]^2, alpha2[2])), g2-G2 %*%
t(cbind(alpha2[1], alpha2[1]^2, alpha2[2]))) }
rho2_til <- nlm(f2, c(1,1))$estimate[1]
s2_til <- nlm(f2, c(1,1))$estimate[2]
```

```
f3 <- function(alpha3){
	crossprod(g3-G3 %*% t(cbind(alpha3[1], alpha3[1]^2, alpha3[2])), g3-G3 %*%
t(cbind(alpha3[1], alpha3[1]^2, alpha3[2]))) }
rho3_til <- nlm(f3, c(1,1))$estimate[1]
s3_til <- nlm(f3, c(1,1))$estimate[2]

f4 <- function(alpha4){
	crossprod(g4-G4 %*% t(cbind(alpha4[1], alpha4[1]^2, alpha4[2])), g4-G4 %*%
t(cbind(alpha4[1], alpha4[1]^2, alpha4[2]))) }
rho4_til <- nlm(f4, c(0.05,0.05))$estimate[1]
s4_til <- nlm(f4, c(0.05,0.05))$estimate[2]

y1star <- y1 - rho1_til* W %*% y1; y2star<- y2 - rho2_til* W %*% y2; y3star<- y3 - rho3_til* W %*% y3;
y4star <- y4 - rho4_til* W %*% y4
Z1star <- Z1 - rho1_til* W %*% Z1; Z2star<- Z2 - rho2_til* W %*% Z2; Z3star<- Z3 - rho3_til* W %*%
Z3; Z4star <- Z4 - rho4_til* W %*% Z4

delta1_hat_2S <- solve ( crossprod(P %*% Z1star, Z1star), crossprod(P %*% Z1star, y1star))
delta2_hat_2S <- solve ( crossprod(P %*% Z2star, Z2star), crossprod(P %*% Z2star, y2star))
delta3_hat_2S <- solve ( crossprod(P %*% Z3star, Z3star), crossprod(P %*% Z3star, y3star))
delta4_hat_2S <- solve ( crossprod(P %*% Z4star, Z4star), crossprod(P %*% Z4star, y4star))

y1_hat_2S <- Z1 %*% delta1_hat_2S
y2_hat_2S <- Z2 %*% delta2_hat_2S
y3_hat_2S <- Z3 %*% delta3_hat_2S
y4_hat_2S <- Z4 %*% delta4_hat_2S

e1_til <- y1star-Z1star %*% delta1_hat_2S
e2_til <- y2star-Z2star %*% delta2_hat_2S
e3_til <- y3star-Z3star %*% delta3_hat_2S
e4_til <- y4star-Z4star %*% delta4_hat_2S

plot(y1, y1_hat_2S); plot(y2, y2_hat_2S); plot(y3, y3_hat_2S); plot(y4, y4_hat_2S)

vcov1_2S <- s1_til* solve( crossprod(P %*% Z1star, P %*% Z1star), diag(1,dZ1) )
vcov2_2S <- s2_til* solve( crossprod(P %*% Z2star, P %*% Z2star), diag(1,dZ2) )
vcov3_2S <- s3_til* solve( crossprod(P %*% Z3star, P %*% Z3star), diag(1,dZ3) )
vcov4_2S <- s4_til* solve( crossprod(P %*% Z4star, P %*% Z4star), diag(1,dZ4) )
var1_2S <- diag(vcov1_2S); var2_2S<-diag(vcov2_2S); var3_2S<-diag(vcov3_2S); var4_2S<-
diag(vcov4_2S)

delta_hat_2S <- rbind(delta1_hat_2S,delta2_hat_2S,delta3_hat_2S,delta4_hat_2S)
var_2S <- c(var1_2S, var2_2S, var3_2S, var4_2S)
GS2SLS <- cbind(delta_hat_2S, sqrt(var_2S), delta_hat_2S/sqrt(var_2S))
colnames(GS2SLS) <- list("2SLS coeff", "2SLS s.e.", "2SLS t value")

delta1_2S <- GS2SLS[1:dZ1,]
delta2_2S <- GS2SLS[(dZ1+1):(dZ1+dZ2),]
delta3_2S <- GS2SLS[(dZ1+dZ2+1):(dZ1+dZ2+dZ3),]
delta4_2S <- GS2SLS[(dZ1+dZ2+dZ3+1):(dZ1+dZ2+dZ3+dZ4),]

#GS3SLS
#############
SIGMAm <- array(0,c(m,m))
SIGMAm[1,] <- (1/n)*rbind(crossprod(e1_til,e1_til), crossprod(e1_til,e2_til), crossprod(e1_til,e3_til),
crossprod(e1_til,e4_til))
```

```
SIGMAm[2,] <- (1/n)*rbind(crossprod(e2_til,e1_til), crossprod(e2_til,e2_til), crossprod(e2_til,e3_til),
crossprod(e2_til,e4_til))
SIGMAm[3,] <- (1/n)*rbind(crossprod(e3_til,e1_til), crossprod(e3_til,e2_til), crossprod(e3_til,e3_til),
crossprod(e3_til,e4_til))
SIGMAm[4,] <- (1/n)*rbind(crossprod(e4_til,e1_til), crossprod(e4_til,e2_til), crossprod(e4_til,e3_til),
crossprod(e4_til,e4_til))

In <- diag(1,n)
Im <- diag(1,m)
ystar <- rbind(y1star,y2star, y3star, y4star)
PZstar <- rbind(  cbind( P %*% Z1star, 0*Z2star, 0*Z3star, 0*Z4star), cbind(  0*Z1star, P %*% Z2star,
0*Z3star, 0*Z4star),cbind(  0*Z1star, 0*Z2star, P %*% Z3star, 0*Z4star),cbind(  0*Z1star, 0*Z2star,
0*Z3star, P %*% Z4star))

delta_hat_3S<- solve( crossprod( PZstar,kronecker(solve(SIGMAm,Im),In)) %*% PZstar, crossprod(
PZstar,kronecker(solve(SIGMAm,Im),In)) %*% ystar )
vcov_3S<-solve( crossprod(PZstar,kronecker(solve(SIGMAm,Im),In)) %*% PZstar,
diag(1,dZ1+dZ2+dZ3+dZ4) )
var_3S<-diag(vcov_3S)

#3SLS coeff, s.e., t stat, p-value
GS3SLS<-cbind(delta_hat_3S, sqrt(var_3S), delta_hat_3S/sqrt(var_3S))
colnames(GS3SLS)<-list("3SLS coeff", "3SLS s.e.", "3SLS t value")

delta1_3S<-GS3SLS[1:dZ1,]
delta2_3S<-GS3SLS[(dZ1+1):(dZ1+dZ2),]
delta3_3S<-GS3SLS[(dZ1+dZ2+1):(dZ1+dZ2+dZ3),]
delta4_3S<-GS3SLS[(dZ1+dZ2+dZ3+1):(dZ1+dZ2+dZ3+dZ4),]

c(y1n, y2n, y3n, y4n)
cbind(delta1_til, delta1_2S, delta1_3S)
cbind(delta2_til, delta2_2S, delta2_3S)
cbind(delta3_til, delta3_2S, delta3_3S)
cbind(delta4_til, delta4_2S, delta4_3S)

c(rho1_til, rho2_til, rho3_til, rho4_til)

plot(y1, Z1 %*% delta1_3S[,1]); lines(q,q)
plot(y2, Z2 %*% delta2_3S[,1]); lines(q,q)
plot(y3, Z3 %*% delta3_3S[,1]); lines(q,q)
plot(y4, Z4 %*% delta4_3S[,1]); lines(q,q)

c(gFit(y1, Z1 %*% delta1_2S[,1]), gFit(y1, Z1 %*% delta1_3S[,1]))
c(gFit(y2, Z2 %*% delta2_2S[,1]), gFit(y2, Z2 %*% delta2_3S[,1]))
c(gFit(y3, Z3 %*% delta3_2S[,1]), gFit(y3, Z3 %*% delta3_3S[,1]))
c(gFit(y4, Z4 %*% delta4_2S[,1]), gFit(y4, Z4 %*% delta4_3S[,1]))

# the block below to be used if iteration is desired
u1_til<- y1- Z1 %*% delta1_3S[,1]
u2_til<- y2- Z2 %*% delta2_3S[,1]
u3_til<- y3- Z3 %*% delta3_3S[,1]
u4_til<- y4- Z4 %*% delta4_3S[,1]

#plot(BLTYR0, y1)
#plot(BLTYR0, y2)
#plot(BLTYR0, y4)
```

```
#plot(BLTYR0, BLTYR9)

ev1<-1/sd(u1_til)
ev2<-1/sd(u2_til)
ev3<-1/sd(u3_til)
ev4<-1/sd(u4_til)

############################################################
#         solving simultaneously same period

#current variables
y1c<-AVHHIN9
y1ac   <- ifelse(AVHHIN9<0.01,0.01, AVHHIN9)
y2c    <- NUMHHS9
y3c    <- BLTYR9
y3ac   <- ifelse(BLTYR9<0.01,0.01, BLTYR9)
y4c    <- PVACHU9
y5ac   <- x7          # ifelse(ABHU9<0.01,0.01, ABHU9)
#ABHUR9A <- ifelse(ABHUR9A<0.01,0.01, ABHUR9A)
y5bc  <- ifelse(x7<0.01,0.01, x7)
exo1c  <- ifelse(x5<0.01,0.01, x5)
#exo2c <- y2c + VACHU9  #housing stock

#forecast variables
i<-515
y1p <- rep(0,i)
y2p <- rep(0,i)
y3p <- rep(0,i)
y4p <- rep(0,i)

A_y1  <- cbind(    diag(1,i,i) - 0.05849686*nmat ,    diag(+0.08768054/(AREA*1000)),       diag( -
6.62444475 ,i,i)           ,  diag(0,i,i)          )
A_y2  <- cbind(    diag(0,i,i),                    diag(1,i,i)- 0.033185720*nmat ,      diag( -0.918842610 ,i,i)
,  diag(0,i,i)          )

A_y3  <- cbind(    diag(0,i,i),                    diag(-0.8634428,i,i) ,              diag(1,i,i) -0.1704842 *nmat
                ,  diag(-1.0559266 ,i,i))
A_y4  <- cbind(    diag(0.0001354090,i,i),           diag(0.0255989581,i,i) ,            diag(0  ,i,i)
,  diag(1,i,i) -0.2478047025*nmat )
A     <- rbind(A_y1, A_y2, A_y3, A_y4)

B_y1 <- (               1.30052053  * y1c                                                    + 0.01171603
* (1/y5bc)   )
B_y2 <- ( -0.270314182  + 0.001332000 * y1c      + 0.995502776* y2c
+ 0.002871906 * (1/exo1c)  )

B_y3 <- (                              - 0.8634428  * y2c     + 0.1553021     * y3c      -0.5127563    * y4c
+ 0.2077432   * y5ac       )
B_y4 <- (               0.0005965211* (1/y1ac) + 0.0255989581* y2c    + 0.0002307763  * (1/y3ac)  +
0.6098514338* y4c                         )
B    <- c( B_y1, B_y2, B_y3, B_y4 )

bigPredict<- solve(A,B)
y1p  <- bigPredict[1:i]
y2p  <- bigPredict[(i+1):(2*i)]
y3p  <- bigPredict[(2*i+1):(3*i)]
```

```
y4p  <- bigPredict[(3*i+1):(4*i)]

y5ap <- y3p - (  y2p /(1-y4p) - y2c /(1-y4c)   )
y5bp <- y5ap / (y2c /(1-y4c))  # abandonment / previous decade housing stock

gFit(y1p, AVHHIN0)
gFit(y2p, NUMHHS0)
gFit(y3p, BLTYR0)
gFit(y4p, PVACHU0)
gFit(y5ap, ABHUR0A)

plot(y1p,AVHHIN0); lines(q,q)
plot(y2p,NUMHHS0); lines(q,q)
plot(y3p,BLTYR0); lines(q,q)
plot(y4p,PVACHU0); lines(q,q)
#plot(y5ap,y5ac); lines(q,q)
#plot(y5bp,y5bc); lines(q,q)

library(quadprog)
BB_y1<-ifelse(B_y1<100,B_y1,100)
BB_y2<-B_y2-5
BB_y3<-B_y3-1
BB_y4<-B_y4-0.5
BB    <- c( BB_y1, BB_y2, BB_y3, BB_y4 )

#c<-t(cbind(t(AVHHIN9*3),t(NUMHHS9*1.2),t(BLTYR9),t(PVACHU9)))
c<-t(cbind(t(rep(200,i)),t(rep(1,i)),t(rep(0.2,i)),t(rep(0.0,i))) )
plot(A%*%c-BB, ylim=c(-3,3))
summary(A%*%c-BB)

d<-cbind(t(rep(0,i)),t(rep(1,i)),t(rep(0,i)),t(rep(0,i)))
D1<-cbind(diag(0,i),diag(0,i),diag(0,i),diag(0,i))
D2<-cbind(diag(1,i),diag(0,i),diag(0,i),diag(0,i))
D<-rbind(D1, D2, D1, D1)
fQp<-function(b){
crossprod((A%*%b-B), diag(c(rep(ev1,i),rep(ev2,i),rep(ev3,i),rep(ev4,i)))  )%*%(A%*%b-B)
}

#(d%*%b-990.249)^2*(crossprod(b,D)%*%b-57167.96)^2*
BP<-constrOptim(c, fQp, NULL, ui=A,ci=BB)
y1pp  <- BP$par[1:i]
y2pp  <- BP$par[(i+1):(2*i)]
y3pp  <- BP$par[(2*i+1):(3*i)]
y4pp  <- BP$par[(3*i+1):(4*i)]
y5app <- y3pp - (  y2pp /(1-y4pp) - y2c /(1-y4c)   )
y5bpp <- y5app / (y2c /(1-y4c))  # abandonment / previous decade housing stock

gFit(y1pp, AVHHIN0)
gFit(y2pp, NUMHHS0)
gFit(y3pp, BLTYR0)
gFit(y4pp, PVACHU0)
gFit(y5app, ABHUR0A)

plot(y1pp,AVHHIN0); lines(q,q)
plot(y2pp,NUMHHS0); lines(q,q)
plot(y3pp,BLTYR0); lines(q,q)
```

```
plot(y4pp,PVACHU0); lines(q,q)
#plot(y5app,y5ac); lines(q,q)
#plot(y5bpp,y5bc); lines(q,q)

############################################################
#          solving simultaneously next period

#current variables
y1c <-  ifelse(AVHHIN0<0.01,0.01, AVHHIN0)
y2c <-  ifelse(NUMHHS0<0.01,1, NUMHHS0)
y3c <-  ifelse(BLTYR0<0.01,0.01, BLTYR0)
y4c <-  ifelse(PVACHU0<0.01,0.01, PVACHU0)
y5ac <- ifelse(ABHU0<0.01,1, ABHU0)
y5bc <- ifelse(ABHUR0A<0.01,0.01, ABHUR0A)
exo1c <- ifelse(HSDROP0<0.01,0.01, HSDROP0)
exo2c    <- y2c + VACHU0  #housing stock

regFit<-gFit(AVHHIN0-eq_AVHHIN0$residuals, AVHHIN0)
plot(AVHHIN0,AVHHIN0-eq_AVHHIN0$residuals, ylab="predicted", col="red", main=round(regFit,4))
points(AVHHIN0,AVHHIN0p, pch="+")
lines(x,y)

regFit<-gFit(NUMHHS0-eq_NUMHHS0$residuals, NUMHHS0)
plot(NUMHHS0,NUMHHS0-eq_NUMHHS0$residuals, ylab="predicted", col="red",
main=round(regFit,4))
points(NUMHHS0,NUMHHS0p, pch="+")
lines(x,y)

regFit<-gFit(BLTYR0-eq_BLTYR0$residuals, BLTYR0)
plot(BLTYR0,BLTYR0-eq_BLTYR0$residuals, ylab="predicted", col="red", main=round(regFit,4))
points(BLTYR0,BLTYR0p, pch="+")
lines(x,y)

regFit<-gFit(PVACHU0-eq_PVACHU0$residuals, PVACHU0)
plot(PVACHU0,PVACHU0-eq_PVACHU0$residuals, ylab="predicted", col="red",
main=round(regFit,4))
points(PVACHU0,PVACHU0p, pch="+")
lines(x,y)
```

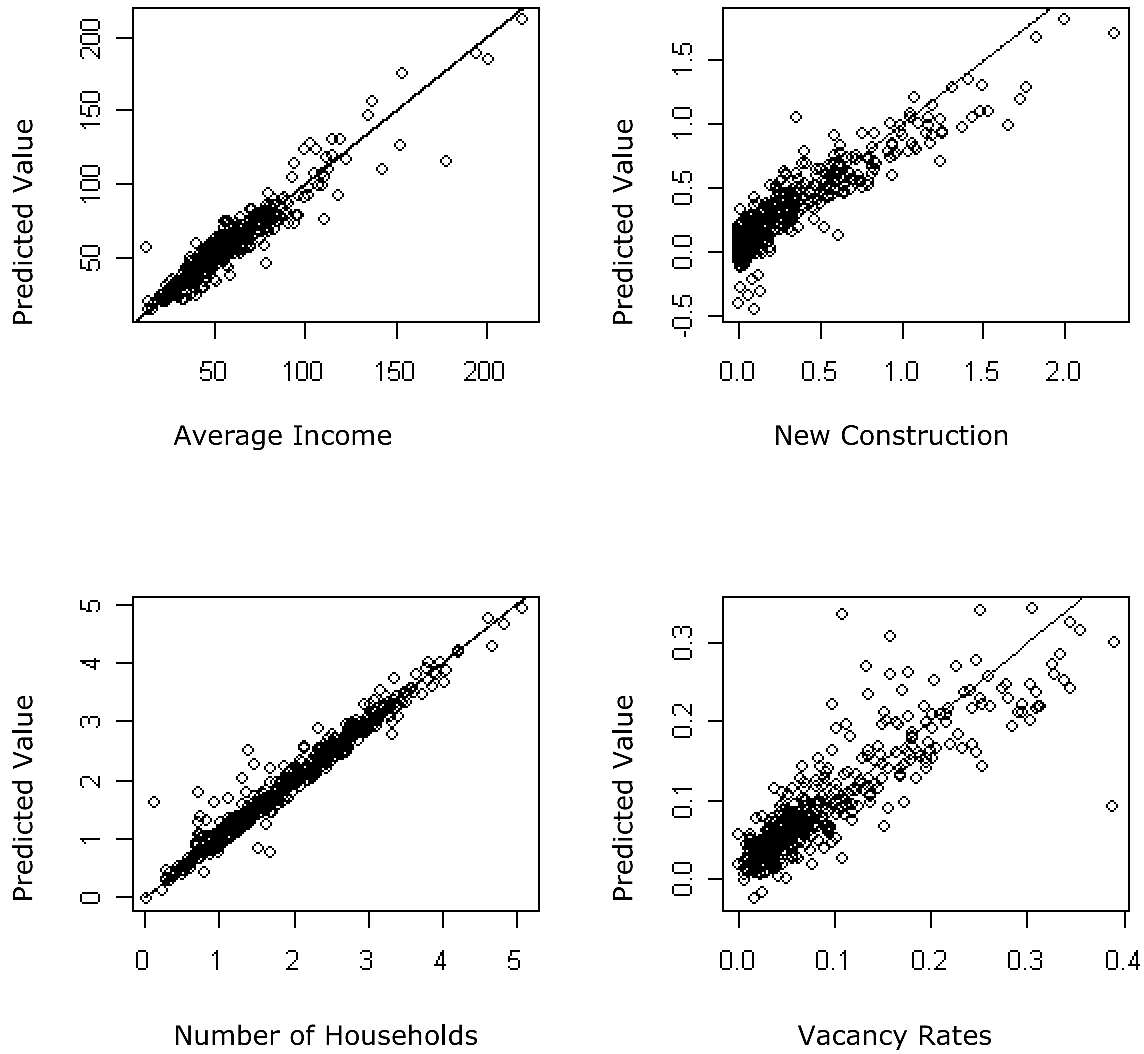


Figure B.1: Model diagnostics

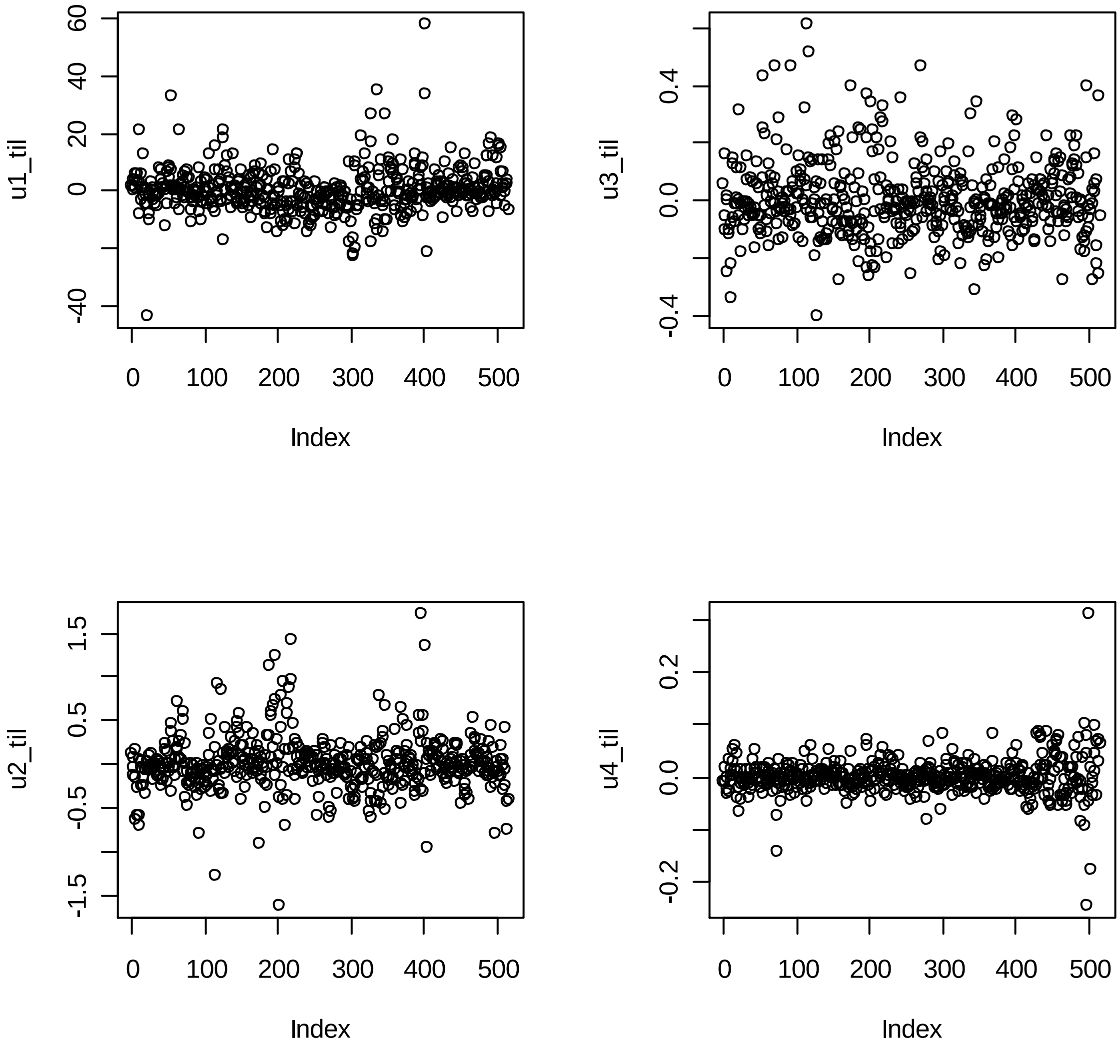


Figure B.2: Error diagnostics

APPENDIX C: MONTE CARLO CODE

The FGS3SLS Estimator

This appendix first summarizes the FGS2SLS and FGS3SLS estimator as derived in Kelejian and Prucha (2004) followed by a sample R code used to perform the Monte Carlo simulations. The system of equations considered in this chapter can be seen as a spatial extension to 'neighborhood ecology models' in concept and an extension to the widely used single equation model of Cliff and Ord (1973, 1981) in terms of econometric specification.

$$\mathbf{Y}_n = \mathbf{Y}_n\mathbf{B} + \mathbf{X}_n\mathbf{C} + \overline{\mathbf{Y}}_n\Lambda + \mathbf{U}_n \tag{C.1}$$

with $\mathbf{Y}_n = (\mathbf{y}_{1,n}, \mathbf{y}_{2,n}.....\mathbf{y}_{m,n})$, $\mathbf{X}_n = (\mathbf{x}_{1,n}, \mathbf{x}_{2,n}.....\mathbf{x}_{k,n})$, $\mathbf{U}_n = (\mathbf{u}_{1,n}, \mathbf{u}_{2,n}.....\mathbf{u}_{m,n})$,

$\overline{\mathbf{Y}}_n = (\overline{\mathbf{y}}_{1,n}, \overline{\mathbf{y}}_{2,n}.....\overline{\mathbf{y}}_{m,n})$, $\overline{\mathbf{y}}_{m,n_n} = \mathbf{W}_n\mathbf{y}_{j,n}$ $\quad j = 1,..., m$

where $\mathbf{y}_{j,n}$ is the *n x 1* vector of cross sectional observations on each of the dependent variables. Since the model is conditional on the realized value of $\mathbf{y}_{j,n}$ in the previous period, the temporally lagged endogenous variables are treated as given and forms a component of vector $\mathbf{x}_{j,n}$. $\mathbf{u}_{j,n}$ is the *n* × *1* disturbance vector in the j^{th} equation, $\mathbf{W}_n$ is an *n* × *n* row standardized weights matrix of known constants, and $\mathbf{B}$, $\mathbf{C}$, and Λ are correspondingly defined parameter matrices of dimension *m*×*m, k* ×*m* and *m*×*m* respectively. The spillovers in endogenous variables and spatial correlation of disturbance terms are modeled as follow. The vector $\overline{\mathbf{y}}_{j,n} = \mathbf{W}_n.\mathbf{y}_{j,n}$ models the spatial spillover of endogenous variable, often referred to as the spatial lag of $\mathbf{y}_{j,n}$. Its r^{th} element represents the average of the values of y in the neighborhood of r. The error terms are assumed to be generated by the following spatially autoregressive process

$$\mathbf{U}_n = \overline{\mathbf{U}}_n \mathbf{R} + \mathbf{E}_n \tag{C.2}$$

with $\mathbf{E}_n = (\varepsilon_{1,n}, \varepsilon_{2,n}.....\varepsilon_{m,n}), \quad \mathbf{R} = diag_{j=1}^{m}(\rho_j)$

$\overline{\mathbf{U}}_n = (\overline{\mathbf{u}}_{1,n}, \overline{\mathbf{u}}_{2,n}.....\overline{\mathbf{u}}_{m,n}), \quad \overline{\mathbf{u}}_{j,n} = \mathbf{W}_n \mathbf{u}_{j,n}, \quad j = 1,..., m$

where $\varepsilon_{j,n}$ is the *n × 1* vector of disturbances, ρ_j is the spatial autoregressive parameter in the j^{th} equation. The vector $\overline{\mathbf{u}}_{1,n}$ is defined analogous to the spatial lag variable $\mathbf{y}_{j,n}$. The disturbances are not only assumed to be spatially correlated across space but also across equations.

The j^{th} equation of (1) after imposing exclusion restrictions may be expressed as

$$\mathbf{y}_{j,n} = \mathbf{Y}_{j,n}\beta'_j + \mathbf{X}_{j,n}\gamma'_j + \overline{\mathbf{Y}}_{j,n}\lambda'_j + \mathbf{u}_{j,n} \tag{C.3}$$

with $\mathbf{u}_{j,n} = \rho \mathbf{W}_n \mathbf{u}_{j,n} + \varepsilon_{j,n}$

and β_j, γ_j and λ_j are the coefficients associated with the endogenous, exogenous and spatially dependent endogenous variables respectively,

Rewriting (A.3) as

$$\mathbf{y}_{j,n} = \mathbf{Z}_{j,n}\delta_j + \mathbf{u}_{j,n} \tag{C.4}$$

with $\mathbf{u}_{j,n} = \rho \mathbf{W}_n \mathbf{u}_{j,n} + \varepsilon_{j,n}$

where $\mathbf{Z}_{j,n} = (\mathbf{Y}_{j,n}, \mathbf{X}_{j,n}, \overline{\mathbf{Y}}_{j,n})$ and $\delta_j = (\beta'_j, \gamma'_j, \lambda'_j)$

To estimate (A.4), let the instrument matrix $\mathbf{H}_n$ contain at least the linearly independent columns of $(\mathbf{X}_n, \mathbf{W}_n\mathbf{X}_n)$ and follow the regularity conditions as discussed in Kelejian and Prucha (2004). Defining

$\tilde{\mathbf{Z}}_{j,n} = \mathbf{P}_H \mathbf{Z}_{j,n}$ where $\mathbf{P}_H = \mathbf{H}_n(\mathbf{H}'_n\mathbf{H}_n)^{-1}\mathbf{H}'_n$, the initial 2SLS estimator of δ_j is then given by

$$\tilde{\delta}_{j,n} = (\tilde{\mathbf{Z}}'_{j,n}\mathbf{Z}_{j,n})\tilde{\mathbf{Z}}_{j,n}\mathbf{y}_{j,n}\,. \tag{C.5}$$

Next, define $\mathbf{y}^*_{j,n}(\rho) = \mathbf{y}_{j,n} - \rho\mathbf{W}_n\mathbf{y}_{j,n}$, $\mathbf{Z}^*_{j,n}(\rho) = \mathbf{Z}_{j,n} - \rho\mathbf{W}_n\mathbf{Z}_{j,n}$ and applying the Cochrane-Ocrutt-type transformation to (4), we get

$$\mathbf{y}^*_{j,n}(\rho_j) = \mathbf{Z}^*_{j,n}(\rho_j)\delta_j + \varepsilon_{j,n} \tag{C.6}$$

Replacing ρ_j with its consistent estimator $\tilde{\rho}_j$ estimated using generalized moments approach outlined in Kelejian and Prucha (1999), Feasible Generalized Spatial 2SLS (or FGS2SLS) can then be estimated as

$$\hat{\delta}^F_{j,n} = [\hat{\mathbf{Z}}^*_{j,n}(\tilde{\rho}_{j,n})'\hat{\mathbf{Z}}^*_{j,n}(\tilde{\rho}_{j,n})]^{-1}\hat{\mathbf{Z}}^*_{j,n}(\tilde{\rho}_{j,n})'\mathbf{y}^*_{j,n}(\tilde{\rho}_{j,n}) \tag{C.7}$$

where $\hat{\mathbf{Z}}^*_{j,n}(\tilde{\rho}_{j,n}) = \mathbf{P}_H\hat{\mathbf{Z}}^*_{j,n}(\tilde{\rho}_{j,n})$. The small sample distribution can be approximated by

$$\hat{\delta}^F_{j,n} \sim N(\delta, \tilde{\sigma}_{jj}[\hat{\mathbf{Z}}^*_{j,n}(\tilde{\rho}_{j,n})'\hat{\mathbf{Z}}^*_{j,n}(\tilde{\rho}_{j,n})]^{-1}) \tag{C.8}$$

with $\tilde{\sigma}_{jj}$ being a consistent estimator of σ_{jj} estimated using the generalized moments used above.

The GS2SLS estimator accounts for potential spatial correlation but does not utilize the full information by ignoring potential cross equation correlation in the error terms. The full information feasible generalized spatial 3SLS (FGS3LS) is then given by

$$\check{\delta}_n^F = [\hat{\mathbf{Z}}_n^*(\tilde{\rho}_n)'(\hat{\Sigma}_n^{-1} \otimes \mathbf{I}_n)\hat{\mathbf{Z}}_n^*(\tilde{\rho}_n)]^{-1}\hat{\mathbf{Z}}_n^*(\tilde{\rho}_n)'(\hat{\Sigma}_n^{-1} \otimes \mathbf{I}_n)\mathbf{y}_n^*(\tilde{\rho}_n) \tag{C.9}$$

Its small sample distribution can be approximated by

$$\hat{\delta}_n^F \sim N(\delta, [\hat{\mathbf{Z}}_n^*(\tilde{\rho}_n)'(\hat{\Sigma}_n^{-1} \otimes \mathbf{I}_n)\hat{\mathbf{Z}}_n^*(\tilde{\rho}_n)]^{-1}) \tag{C.10}$$

Sample Code for Monte Carlo Simulations in R

```
rm(list=ls(all=TRUE))
set.seed(100)
par(mfcol=c(2,2))
irow<-0

#initializing arrays to store final restuls
#array size : 9 rho x 9 lambda x 3 n =243 rows, 8 coeff x 3 methods + 3 identifiers for ni, li, ri= 29
columns
beta_summary_mean_f      <- array(0,c(243,29))
beta_summary_median_f    <- array(0,c(243,29))
beta_summary_bias1_f     <- array(0,c(243,29))    # mean bias
beta_summary_bias2_f     <- array(0,c(243,29))    # median bias
beta_summary_varp_f      <- array(0,c(243,29))    #variance pooled
beta_summary_vart_f      <- array(0,c(243,29))    #variance thetic antithetic
beta_summary_rmse_f      <- array(0,c(243,29))    #rmse
beta_summary_rmse_mod_f  <- array(0,c(243,29))    #sqrt(bias^2+(IQ/1.35)^2)
endo_summary             <- array(0,c(243,35))        #mean,bias, etc (6) + variance +corr,  of 2 endo
vars with 2 methods S2SLS and S3SLS plus 3 identifiers

#hlist is the columns heading for above arrays
hlist <- list("n", "lambda",
"rho","2sls_l1","2sls_g12","2sls_b1","2sls_b3","2sls_l2","2sls_g21","2sls_b2","2sls_b4","s2sls_l1","s2sl
s_g12","s2sls_b1","s2sls_b3","s2sls_l2","s2sls_g21","s2sls_b2","s2sls_b4","s3sls_l1","s3sls_g12","s3s
ls_b1","s3sls_b3","s3sls_l2","s3sls_g21","s3sls_b2","s3sls_b4","s_r1","s_r2")
colnames(beta_summary_mean_f)      <- hlist
colnames(beta_summary_median_f)    <- hlist
colnames(beta_summary_bias1_f)     <- hlist
colnames(beta_summary_bias2_f)     <- hlist
colnames(beta_summary_varp_f)      <- hlist
colnames(beta_summary_vart_f)      <- hlist
colnames(beta_summary_rmse_f)      <- hlist
colnames(beta_summary_rmse_mod_f)  <- hlist
#colnames(endo_summary) <- list("n", "lambda", "rho","y1_bias_S2SLS", "y2_bias_S2SLS",
"y1_bias_S3SLS", "y2_bias_S3SLS", "y1_var_S2SLS", "y2_var_S2SLS", "y1_var_S3SLS",
"y2_bias_S3SLS")
colnames(endo_summary)  <- list("n", "lambda", "rho",
"y1_min_2s","y1_1q_2s","y1_med_2s","y1_mean_2s","y1_2q_2s","y1_max_2s","y1_var_2s",
"y1_cor_2s", "y2_min_2s","y2_1q_2s","y2_med_2s","y2_mean_2s","y2_2q_2s","y2_max_2s",
"y2_var_2s", "y2_cor_2s",
"y1_min_3s","y1_1q_3s","y1_med_3s","y1_mean_3s","y1_2q_3s","y1_max_3s",
"y1_var_3s","y1_cor_3s",
"y2_min_3s","y2_1q_3s","y2_med_3s","y2_mean_3s","y2_2q_3s","y2_max_3s","y2_var_3s","y2_cor_
3s")

########################
# model parameters
########################
m   <- 2           # number of equations
rep <- 500         # number of repetitions * 2, one each for thetic and antithetic variates
Im  <- diag(1,m)

# variance covariance matrix
omega1 <- array(0,c(2,2))
omega2 <- array(0,c(2,2))
```

```
#gives R squared of 0.75 approximately
omega1[1,] <-c ( 900, 450)
omega1[2,] <-c ( 450, 900)
#r square of 0.60 approximately
omega2[1,] <- c(  3000,-2500)
omega2[2,] <- c( -2500, 4000)

omega  <- omega2
sigma1 <- sqrt(omega[1,1])
sigma2 <- sqrt(omega[2,2])

# equation parameters
g11 <-  1;    g12 <- -0.3
g21 <-  0.7;  g22 <-   1
b1 <-   2.0;   b3 <- 2.5;
b2 <-   2.5;   b4 <- -2.0

#####################
# genrating exogenous variables master set from which to sample for different n's
# x1 and x2 are lagged endogenous variables
# x3 and x4 are time dependent
#####################
X1a <- array(0,c(20,1000)); X2a <- array(0,c(20,1000))
X3a <- array(0,c(20,1000)); X4a <- array(0,c(20,1000))
X1a[1,] <- rnorm(1000, 0, 1); X2a[1,] <- rnorm(1000, 0, 1)
X3a[1,] <- rnorm(1000, 0, 1); X4a[1,] <- rnorm(1000, 0, 1)

nl      <- c(250)                                    # 100, 500, 1000 sample size
lambdal <- c(-0.8,-0.6,-0.4,-0.2,0,0.2,0.4,0.6,0.8)    #-0.8,-0.6,-0.4,-0.2,0,0.2,0.4,0.6,0.8
rhol    <- c(-0.8,-0.6,-0.4,-0.2,0,0.2,0.4,0.6,0.8)    #-0.8,-0.6,-0.4,-0.2,0,0.2,0.4,0.6,0.8

#______________________________loop
for(ni in 1:length(nl)){
        n  <- nl[ni]
        In <- diag(1,n)
        Zn <- diag(0,n)
        Y_fcast<- array(0, c(81,(6*n))) # 9rho x 9 lanbda, 3 pairs of values of y

        ####################
        # weights
        ####################
        #creating a standardized circular matrix, one ahead and one back
        # W1 <- diag(0,n)       #lattice
        # for(i in 1:n){
        #          W1[i,ifelse(i-1 <1 ,n-i+1,i-1)]<-1/2
        #          W1[i,ifelse(i+1 >n ,n-i+1,i+1)]<-1/2}

        #creating a standardized circular matrix, 3 ahead and 3 back
        W3 <- diag(0,n)       #lattice
        for(i in 1:n){
                W3[i,ifelse(i-3 < 1,n+i-3,i-3)] <- 1/6
                W3[i,ifelse(i-2 < 1,n+i-2,i-2)] <- 1/6
                W3[i,ifelse(i-1 < 1,n+i-1,i-1)] <- 1/6
                W3[i,ifelse(i+1 > n,n-i+1,i+1)] <- 1/6
                W3[i,ifelse(i+2 > n,n-i+2,i+2)] <- 1/6
```

```
		W3[i,ifelse(i+3 > n,n-i+3,i+3)] <- 1/6
	}
	W<-W3
	#______________________________loop
	for(li in 1:length(lambdal)){
		l1 <- lambdal[li]; l2 <- lambdal[li]
		######################
		# BY + TX = U
		# Y <- Inv(B)*(-T*X + U)
		######################
		#matrix B of coeff on dependent variables
		B <- array(0,c(2*n,2*n))
		B[1:n,]		<- c( (In*g11+l1*W), In*g12 )
		B[(n+1):(2*n),]	<- c( In*g21, (In*g22+l2*W) )
		T <- array(0,c(2*n,4*n))
		T[1:n,]		<- c( In*b1, Zn, In*b3, Zn)
		T[(n+1):(2*n),]	<- c( Zn, In*b2, Zn, In*b4)

		######################
		# Generating X1 and X2
		######################
		for(t in 2:8){
			Xt <- as.vector(cbind(X1a[t-1,1:n],X2a[t-1,1:n], X3a[t-1,1:n], X4a[t-1,1:n] ))
			tt <- - solve(B,T%*%Xt)
			X1a[t,1:n] <- tt[1:n]
			X2a[t,1:n] <- tt[(n+1):(2*n)]
			X3a[t,]	<- X3a[t-1,]*1.01
			X4a[t,]	<- X4a[t-1,]*1.01
		}
		#the t=(8-1) value is used as true value for forecast year. t=(8-2) value is used to
estimate the coefficients
		x1 <- X1a[t-2,1:n]; x2 <- X2a[t-2,1:n]; x3 <- X3a[t-2,1:n]; x4 <- X4a[t-2,1:n]
		X  <- as.vector(cbind(x1,x2,x3,x4))
		#______________________________loop
		for(ri in 1:length(rhol)){
			r1<-rhol[ri]; r2<-rhol[ri]
			beta_exp<-array(0,c(rep,52))
			#______________________________loop
			for(repi in 1:rep){
				print(c(ni, li, ri, repi))
				#####################
				# generating errors
				#####################
				v1 <- rnorm(n,0,1); v2 <- rnorm(n,0,1);
				V  <- cbind(v1,v2)
				omega_star <- chol(omega)
				E  <- V%*%omega_star
				e1 <- E[,1]; e2 <- E[,2]
				#U <-array(0,5*n)
				u1 <- solve( In-r1*W, e1 ); u2 <- solve( In-r2*W, e2 )
				U  <- c(u1,u2)

				################
				# Generating Y using thetic and anti-thetic variates..
				################
```

```
#____________________________loop
for (thetic in 0:1){
        Y   <- - solve(B, T%*%X + (-1)^thetic * U)
        y1  <- Y[1:n]; y2 <- Y[(n+1):(2*n)]
        wy1 <- W %*% y1; wy2 <- W %*% y2
        Xm  <- cbind(x1,x2,x3,x4)
        #y=Z.delta+U
        Z1 <- cbind(-wy1, -y2, -x1, -x3)
        Z2 <- cbind(-wy2, -y1, -x2, -x4 )
        #colnames(Z1) <- list("l1", "g12", "b1", "b3")
        #colnames(Z2) <- list("l2", "g21", "b2", "b4")
        dZ1 <- ncol(Z1); dZ2 <- ncol(Z2)

        #####################
        #2SLS
        #####################
        H<-cbind(Xm, W%*%Xm, W%*% (W%*%Xm),
W%*%(W%*% (W%*%Xm)))
        P<-H %*% solve(crossprod(H,H),t(H))

        delta1_til <- solve( crossprod(P %*% Z1, Z1), crossprod( P
%*% Z1 , y1))
        delta2_til <- solve( crossprod(P %*% Z2, Z2), crossprod( P
%*% Z2 , y2))
        delta_til  <- rbind(delta1_til, delta2_til)
        TSLS       <- delta_til
        colnames(TSLS) <- list("2SLS")
        u1_til   <- y1- Z1 %*% delta1_til
        u2_til   <- y2- Z2 %*% delta2_til
        wu1_til  <- W %*% u1_til ; wu2_til  <- W %*% u2_til
        wwu1_til <- W %*% wu1_til; wwu2_til <- W %*% wu2_til
        G1<-array(0,c(3,3)); G2<-array(0,c(3,3))
        G1[1,] <- (1/n)*cbind( 2* crossprod(u1_til, wu1_til) ,  -
crossprod(wu1_til, wu1_til),   n)
        G1[2,] <- (1/n)*cbind( 2* crossprod(wwu1_til, wu1_til), -
crossprod(wwu1_til, wwu1_til), sum(diag(crossprod(W,W))))
        G1[3,] <- (1/n)*cbind( crossprod(u1_til, wwu1_til) +
crossprod(wu1_til, wu1_til), crossprod(wu1_til, wwu1_til), 0)
        g1     <- (1/n)*t( cbind( crossprod(u1_til, u1_til) ,
crossprod(wu1_til, wu1_til) , crossprod(u1_til, wu1_til) ))
        G2[1,] <- (1/n)*cbind( 2* crossprod(u2_til, wu2_til) ,  -
crossprod(wu2_til, wu2_til),   n)
        G2[2,] <- (1/n)*cbind( 2* crossprod(wwu2_til, wu2_til), -
crossprod(wwu2_til, wwu2_til), sum(diag(crossprod(W,W))))
        G2[3,] <- (1/n)*cbind( crossprod(u2_til, wwu2_til) +
crossprod(wu2_til, wu2_til), crossprod(wu2_til, wwu2_til), 0)
        g2     <- (1/n)*t( cbind( crossprod(u2_til, u2_til) ,
crossprod(wu2_til, wu2_til) , crossprod(u2_til, wu2_til) ))

        f1 <- function(alpha1){
        crossprod(g1-G1 %*% t(cbind(alpha1[1], alpha1[1]^2,
alpha1[2])), g1-G1 %*% t(cbind(alpha1[1], alpha1[1]^2, alpha1[2])))  }
        minObj1  <- nlm(f1, c(1,1))
        rho1_til <- minObj1$estimate[1]
        s1_til   <- minObj1$estimate[2]
```

```
f2 <- function(alpha2){
crossprod(g2-G2 %*% t(cbind(alpha2[1], alpha2[1]^2,
alpha2[2])), g2-G2 %*% t(cbind(alpha2[1], alpha2[1]^2, alpha2[2]))) }
minObj2  <- nlm(f2, c(1,1))
rho2_til <- minObj2$estimate[1]
s2_til   <- minObj2$estimate[2]
y1star   <- y1 - rho1_til* W %*% y1; y2star <- y2 - rho2_til* W
%*% y2;
Z1star   <- Z1 - rho1_til* W %*% Z1; Z2star <- Z2 - rho2_til*
W %*% Z2;


####################
# FGS2SLS
####################
delta1_hat_2S <- solve ( crossprod(P %*% Z1star, Z1star),
crossprod(P %*% Z1star, y1star))
delta2_hat_2S <- solve ( crossprod(P %*% Z2star, Z2star),
crossprod(P %*% Z2star, y2star))
delta_hat_2S  <- rbind(delta1_hat_2S, delta2_hat_2S)
y1_hat_2S <- Z1 %*% delta1_hat_2S
y2_hat_2S <- Z2 %*% delta2_hat_2S
e1_til <- y1star-Z1star %*% delta1_hat_2S
e2_til <- y2star-Z2star %*% delta2_hat_2S
# vcov1_2S <- s1_til* solve( crossprod(P %*% Z1star, P %*%
Z1star), diag(1,dZ1) )
# vcov2_2S <- s2_til* solve( crossprod(P %*% Z2star, P %*%
Z2star), diag(1,dZ2) )
# var1_2S  <- diag(vcov1_2S); var2_2S<-diag(vcov2_2S)
# var_2S <- c(var1_2S, var2_2S)
GS2SLS <- delta_hat_2S
# colnames(GS2SLS) <- list("FGS2SLS")

#############################
# FGS3SLS
#############################
SIGMAm <- array(0,c(m,m))
SIGMAm[1,] <- (1/n)*rbind(crossprod(e1_til,e1_til),
crossprod(e1_til,e2_til))
SIGMAm[2,] <- (1/n)*rbind(crossprod(e2_til,e1_til),
crossprod(e2_til,e2_til))
ystar  <- rbind(y1star,y2star)
PZstar <- rbind(  cbind( P %*% Z1star, 0*Z2star), cbind(
0*Z1star, P %*% Z2star))
delta_hat_3S <- solve( crossprod(
PZstar,kronecker(solve(SIGMAm,Im),In)) %*% PZstar, crossprod(
PZstar,kronecker(solve(SIGMAm,Im),In)) %*% ystar )
# vcov_3S <- solve(
crossprod(PZstar,kronecker(solve(SIGMAm,Im),In)) %*% PZstar, diag(1,dZ1+dZ2) )
# var_3S  <- diag(vcov_3S)
GS3SLS <- delta_hat_3S
# colnames(GS3SLS) <- list("FGS3SLS")
beta_exp[repi,(26*thetic+1):(26+26*thetic)]<-round(c(TSLS,
GS2SLS, GS3SLS, rho1_til, rho2_til),4)
#print(c(repi,(26*thetic+1)))
} #end thetic
```

```
} #end repetitions


#########################
# summary of coefficients
#########################
beta_exp_pooled     <- rbind(beta_exp[,1:26],beta_exp[,27:52])
beta_summary_mean   <- round(apply(beta_exp_pooled,2,mean),3)
beta_summary_median <- round(apply(beta_exp_pooled,2,median),3)
beta_summary_varp   <- round(apply(beta_exp_pooled,2,var),3)
beta_summary_bias1  <- beta_summary_mean   -
c(l1,g12,b1,b3,l2,g21,b2,b4,l1,g12,b1,b3,l2,g21,b2,b4,l1,g12,b1,b3,l2,g21,b2,b4,r1,r2)
beta_summary_bias2  <- beta_summary_median -
c(l1,g12,b1,b3,l2,g21,b2,b4,l1,g12,b1,b3,l2,g21,b2,b4,l1,g12,b1,b3,l2,g21,b2,b4,r1,r2)
beta_summary_IQ     <- round(apply(beta_exp_pooled,2,IQR),3)

#calculating variance for thetic antithectic estimates
beta_summary_vart <- array(0,26)
for(vi in 1:26){
        beta_summary_vart[vi]  <-  round( (1/4)*(   var(beta_exp[,vi]) +
var(beta_exp[,26+vi]) + 2*cov(beta_exp[,vi],beta_exp[,26+vi])   ),3)
}

beta_summary_rmse     <- round(sqrt(beta_summary_bias1^2 +
beta_summary_vart^2 )   ,3)
beta_summary_rmse_mod <- round(sqrt(beta_summary_bias2^2 +
(beta_summary_IQ/1.35)^2),3)

############################
#populating the final tables
##############################
irow<-irow+1
beta_summary_mean_f[irow,]       <- c(nl[ni], lambdal[li], rhol[ri],
beta_summary_mean)
beta_summary_median_f[irow,]     <- c(nl[ni], lambdal[li], rhol[ri],
beta_summary_median)
beta_summary_bias1_f[irow,]      <- c(nl[ni], lambdal[li], rhol[ri],
beta_summary_bias1)
beta_summary_bias2_f[irow,]      <- c(nl[ni], lambdal[li], rhol[ri],
beta_summary_bias2)
beta_summary_varp_f[irow,]       <- c(nl[ni], lambdal[li], rhol[ri],
beta_summary_varp)
beta_summary_vart_f[irow,]       <- c(nl[ni], lambdal[li], rhol[ri],
beta_summary_vart)
beta_summary_rmse_f[irow,]       <- c(nl[ni], lambdal[li], rhol[ri],
beta_summary_rmse)
beta_summary_rmse_mod_f[irow,]   <- c(nl[ni], lambdal[li], rhol[ri],
beta_summary_rmse_mod)
print(irow)

##############################
#Calculating forecasted endogenous variables
###############################
# true value of forecasted endogenous variable (Y's at t=7) = (x1,x2 at t=8)
y1_ftrue <- X1a[8,1:n]; y2_ftrue<-X2a[8,1:n]
```

```
#exogenous variables
Xtfcast <- as.vector(cbind(X1a[7,1:n],X2a[7,1:n], X3a[7,1:n], X4a[7,1:n] ))

y1_fcast <- array(0,c(n,2))
y2_fcast <- array(0,c(n,2))

# calculating endogenous variables by two methods
# f=1 I gives forecast by FGS2SLS, f=2 gives forecast by FGS3SLS
for(f in 1:2){
	l1_est  <- beta_summary_median_f[irow,12]*(2-f)+
beta_summary_median_f[irow,20]*(f-1)
	g12_est <- beta_summary_median_f[irow,13]*(2-f)+
beta_summary_median_f[irow,21]*(f-1)
	b1_est  <- beta_summary_median_f[irow,14]*(2-f)+
beta_summary_median_f[irow,22]*(f-1)
	b3_est  <- beta_summary_median_f[irow,15]*(2-f)+
beta_summary_median_f[irow,23]*(f-1)
	l2_est <- beta_summary_median_f[irow,16]*(2-f)+
beta_summary_median_f[irow,24]*(f-1)
	g21_est<- beta_summary_median_f[irow,17]*(2-f)+
beta_summary_median_f[irow,25]*(f-1)
	b2_est <- beta_summary_median_f[irow,18]*(2-f)+
beta_summary_median_f[irow,26]*(f-1)
	b4_est <- beta_summary_median_f[irow,19]*(2-f)+
beta_summary_median_f[irow,27]*(f-1)

	#expected value of coeffs
	B_est <- array(0,c(2*n,2*n))
	B_est[1:n,]           <- c( (In*g11+l1_est*W), In*g12_est )
	B_est[(n+1):(2*n),]   <- c( In*g21_est, (In*g22+l2_est*W) )
	T_est<-array(0,c(2*n,4*n))
	T_est[1:n,]           <- c( In*b1_est, Zn, In*b3_est, Zn)
	T_est[(n+1):(2*n),]   <- c( Zn, In*b2_est, Zn, In*b4_est)

	#forecasted y
	Ytfcast  <- - solve(B_est,T_est %*% Xtfcast)
	y1_fcast[,f] <- Ytfcast[1:n]
	y2_fcast[,f] <- Ytfcast[(n+1):(2*n)]
} #end f
Y_fcast[irow,] <- c(y1_ftrue, y1_fcast, y2_ftrue, y2_fcast)
endo_summary[irow,]   <- c(nl[ni], lambdal[li], rhol[ri], summary(y1_fcast[,1] -
y1_ftrue), var(y1_fcast[,1] - y1_ftrue), cor(y1_fcast[,1], y1_ftrue), summary(y2_fcast[,1] - y2_ftrue),
var(y2_fcast[,1] - y2_ftrue), cor(y2_fcast[,1] , y2_ftrue), summary(y1_fcast[,2] - y1_ftrue),
var(y1_fcast[,2] - y1_ftrue), cor(y1_fcast[,2] , y1_ftrue), summary(y2_fcast[,2] - y2_ftrue),
var(y2_fcast[,2] - y2_ftrue), cor(y2_fcast[,2] , y2_ftrue))
#endo_var    <- c(  var(y1_fcast[,2] - y1_ftrue), var(y1_fcast[,2] - y1_ftrue))
#endo[irow,] <- c(endo_bias, endo_var)
print(irow)
print(c(irow,"a"))
		} #end rho
	} #end lambda
} #end n

#beta_summary_mean_f[1:50,]
#beta_summary_median_f[1:50,]
#beta_summary_bias1_f[1:50,]
```

```
#beta_summary_bias2_f[1:50,]
#beta_summary_varp_f[1:50,]
#beta_summary_vart_f[1:50,]
##beta_summary_rmse_f[1:50,]
##beta_summary_rmse_mod_f[1:50,]
#endo_summary[1:50,]

beta_mean <- paste(tempfile(,"/home/sarraf/"), ".csv", sep = "")
write.csv(beta_summary_mean_f,  "/home/sarraf/beta_mean.csv")
beta_median <- paste(tempfile(,"/home/sarraf/"), ".csv", sep = "")
write.csv(beta_summary_median_f,"/home/sarraf/beta_median.csv")
beta_bias1 <- paste(tempfile(,"/home/sarraf/"), ".csv", sep = "")
write.csv(beta_summary_bias1_f,"/home/sarraf/beta_bias1.csv")
beta_bias2 <- paste(tempfile(,"/home/sarraf/"), ".csv", sep = "")
write.csv(beta_summary_bias2_f,"/home/sarraf/beta_bias2.csv")
beta_varp <- paste(tempfile(,"/home/sarraf/"), ".csv", sep = "")
write.csv(beta_summary_varp_f,"/home/sarraf/beta_varp.csv")
beta_vart <- paste(tempfile(,"/home/sarraf/"), ".csv", sep = "")
write.csv(beta_summary_vart_f,"/home/sarraf/Wbeta_vart.csv")
endo_sum <- paste(tempfile(,"/home/sarraf/"), ".csv", sep = "")
write.csv(endo_summary,"/home/sarraf/endo_sum.csv")
endo_fcast <- paste(tempfile(,"/home/sarraf/"), ".csv", sep = "")
write.csv(t(Y_fcast),  "/home/sarraf/endo_fcast.csv")
beta_rmse <- paste(tempfile(,"/home/sarraf/"), ".csv", sep = "")
write.csv(beta_summary_rmse_f,  "/home/sarraf/beta_rmse.csv")
beta_rmse_mod <- paste(tempfile(,"/home/sarraf/"), ".csv", sep = "")
write.csv(beta_summary_rmse_mod_f,  "/home/sarraf/beta_rmse_mod.csv")
#W3OM15k_endo_fcast <- paste(tempfile(,"/home3/sarraf/"), ".dbf", sep = "")
#write.dbf(Y_fcast,  "/home3/sarraf/W3OM15k_endo_fcast.dbf")
#var(y1-X1a[6,1:n])
#summary(y1-X1a[6,1:n])
#W3OM1_beta_mean <- paste(tempfile(), ".dbf", sep = "")
#write.dbf(data.frame(beta_summary_mean_f),W3OM1_beta_mean)
#W3OM1_beta_median <- paste(tempfile(), ".dbf", sep = "")
#write.dbf(data.frame(beta_summary_median_f),W3OM1_beta_median)
#W3OM1_beta_bias1 <- paste(tempfile(), ".dbf", sep = "")
#write.dbf(data.frame(beta_summary_bias1_f),W3OM1_beta_bias1)
#W3OM1_beta_bias2 <- paste(tempfile(), ".dbf", sep = "")
#write.dbf(data.frame(beta_summary_bias2_f),W3OM1_beta_bias2)
#W3OM1_beta_varp <- paste(tempfile(), ".dbf", sep = "")
#write.dbf(data.frame(beta_summary_varp_f),W3OM1_beta_varp)
#W3OM1_beta_vart <- paste(tempfile(), ".dbf", sep = "")
#write.dbf(data.frame(beta_summary_vart_f),W3OM1_beta_vart)
#W3OM1_endo <- paste(tempfile(), ".dbf", sep = "")
#write.dbf(data.frame(endo_summary),W3OM1_endo)
```