MULTISCALAR MODELING OF POLYCENTRIC URBAN-REGIONAL SYSTEMS: ECONOMIC AGGLOMERATION, SCALE DEPENDENCY AND AGENT INTERACTIONS

by

Zhaoya Gong

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Geography and Urban Regional Analysis

Charlotte

2015

Approved by:

____
Dr. Jean-Claude Thill

____
Dr. Harrison Campbell

____
Dr. Wenwu Tang

____
Dr. Daniel Gutierrez

ABSTRACT

ZHAOYA GONG. Multiscalar modeling of polycentric urban-regional systems: economic agglomeration, scale dependency and agent interactions. (Under the direction of DR. JEAN-CLAUDE THILL)

This dissertation aims to study the causal relationship between the underlying processes of agglomeration economies and the formation of certain spatial structures at both intra- and inter-urban scales within the extent of megaregions. First, on the theoretical side, I develop a general framework to account for the interplay between market linkages and spatial costs across scales. The model system constructed from this framework allows us to study the evolution of intra- and inter-urban spatial structures in terms of monocentricity/polycentricity and agglomeration/dispersion. By examining the impacts of local spatial costs and interregional trade costs on the structural change at both scales, I find the interdependency of spatial structures across urban-regional scales. Second, I extend the theoretical models into a 2-D geographic model that can scale to real world applications. This geographic model is based on zonal geography connected by transportation networks. An agent-based approach is employed to model the discrete choice for locations and to approximate the equilibrium conditions as those in theoretical models. The simulations confirm that it is consistent with the theoretical models regarding the generated spatial structures. A demonstration application to the Carolinas megaregion is presented as a test bed for the geographic model. Three simulation scenarios presented provide insights to understand the observed pattern of urban-regional developments. In addition, high-performance computing technologies are leveraged to improve the computational performance of the geographic model.

DEDICATION

To my parents, Mr. Jinjun Gong and Mrs. Youli Li, and to my grandmother, Mrs. Ping Ou, for their endless affection, support and sacrifices.

To the loving memory of my late grandparents, whose roles in my life were, and remain, immense.

To my love, Tengteng, for her love, support and encouragement, and for sharing joys and frustrations with me throughout this challenging course of study.

ACKNOWLEDGMENTS

I would like express my deepest gratitude to my advisor, Dr. Jean-Claude Thill, who has guided me through this journey of Ph.D. study by introducing me to many new realms of knowledge, offering insights to my research questions, inspiring me to bridge disconnected ideas, generously supporting my study, and continuously encouraging me to fulfill my ultimate goal. He provided the best combination of freedom and directing, which gives rise to my pursuit of the current shape and essence of my dissertation. I extend my acknowledgments to my committee members, Drs. Harrison Campbell, Wenwu Tang and Daniel Gutierrez, for their insightful advice, comments and great patience. Particularly, my appreciation goes to Dr. Wenwu Tang for his invaluable guidance, warmhearted help, and unceasing assistance and support, after I was introduced into the domain of high-performance computing.

I am truly grateful to Dr. Jacques Thisse for offering an enlightening discussion about my dissertation proposal. Without his suggestions I would not have accomplished chapters 4 and 5 of my dissertation. I am sincerely indebted to Dr. Takatosh Tabuchi and Dr. Pierre-Philippe Combes for their generosity to share computer codes of their research, which allows me to learn from their work.

I acknowledge the financial support from the GASP grants and the equipment support from University Research Computing.

I also thank my colleagues and friends in the department for their help and encouragement.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 Cities, Systems of Cities, and Emerging Urban-Regions

2010 marks the first time in history that more than half the human population lives in urban areas and, in the next four decades, all the world's net population growth is expected to take place in urban areas (United Nations, 2011). As it manifests itself through variations in intensity of human settlements and land use, urbanization is probably the most extreme of all geographic inequalities. Historically, such inequalities have resulted from the geographic segregation of human population groups that have different political jurisdictions, religions, social standards, and ethnicities. In modern times, cities are usually considered the engines of economic development in that they hold a great share of the economy on a small share of land. For example, in 2000, 0.6% of the territory of the European Union is covered by its top 38 cities that accommodate about 25% of its population and about 30% of its GDP (Henderson and Thisse, 2004).

In reality, the urban landscape is more or less the combination of two natures. Uneven economic development may come from the natural features at different places, such as mineral resources, climate, rivers, and harbors, which are collectively called "first nature". That is the spatial heterogeneity of the surface of the earth, where places with exogenous uneven distribution of natural resources and amenities create comparative advantages that lead to specialization and trade. More important however is the "second nature", which is the economic mechanism emerging from human decisions and interactions to give rise to the spatial agglomeration of economic activities, dubbed agglomeration economies. On the other hand, urban agglomerations are subject to diseconomies such as crowding, congestion, pollution, crime and social segregation and consume a large amount of resources. The spatial agglomeration of economies may correspond to real world phenomena at different spatial scales that range from the North-South divide of world regions, large city-regions dominating their national economy at the country level (Seoul region in Korea, Paris region in France), single metropolis or cities playing a significant role at the regional level (New York and Tokyo), and large commercial districts or urban centers that frame the internal structure of an urban area (Manhattan in New York, Ginza in Tokyo). This is because the nature and balance of agglomeration (economy) and dispersion (diseconomy) forces at work to push and pull economic activities are different at different spatial scales. In other words, "it may be that the patterns that occur at different distance scales are influenced by different types of agglomeration economies, each based on interaction mechanisms with particular requirements for spatial proximity" (Anas et al., 1998, p. 1440).

Recently, global city-systems, or city-regions, have earned considerable economic prosperity in the context of globalization and of the shift to the advanced service economy. Examples include the "Blue Banana" (an area that stretches from London to northern Italy) in Europe, the Northeast megalopolis in the United States, the Yangtze River delta and Pearl River delta areas in China, and the Tokyo-Osaka corridor in Japan. These urban-regions usually span vast areas with diverse spatial forms, encompassing major cities, suburbs, exurbs and rural areas. It has been found that the majority of the projected population and employment growth along with future anticipated urbanization would occur in these urban-regions. To accommodate this need on the one hand, and to act on global competition on the other, regional planning and policy making has gained renewed interest at such broad scale to promote economic integration and reduce disparities within the region as well as to increase economic competitiveness and sustainability of the region as a whole.

Different concepts have been used to denote the new urbanization forms at this broad scale, for instance polycentric urban regions (Kloosterman and Musterd, 2001), global city-regions (Scott, 2001), mega-city regions (Hall and Pain, 2006), and megaregions (RPA, 2006). Though these concepts bear on specific definitions, in general they refer to an urban-regional system, or urban-regions, encompassing networks of cities and metropolitan areas that are linked with a multimodal transportation infrastructure, share ecosystems, topography and culture affinities, and have strong economic ties like overlapping commuting patterns, frequent business travel, common labor pool, and industrial supply chains, which together form a common region and a basis for shared policy consideration.

A core characteristic emphasized here is polycentricity. This construct not only exhibits a multi-nodal spatial organization in a morphological sense (space of places), but also stresses economic interactions between the linked cities or metropolitan areas in a functional sense (space of flows). At variance with the hierarchical structure of monocentric urban systems, this polycentric structure aims to promote alternative urban nuclei or metropolitan areas and their horizontal functional linkages (inter-urban level) and to increase economic diffusion from urban cores to peripheral areas vertically (intra-urban level) by overcoming local fragmentation, taking advantage of complementary economic factors, and encouraging greater collaboration. A higher level of agglomeration economies is expected via such network externalities.

On the contrary, a hierarchical structure of urban systems, advocated in traditional planning, emphasizes monocentricity and encourages economic concentration and complex functions in major cities (providing services and goods unidirectionally to small towns at lower levels of the hierarchy) that increasingly face tremendous urban costs as congestion, higher housing price, degradation of water and air quality, and inefficient use of natural resources. In contrast, polycentric linkages can provide greater flexibility to relieve the stresses of large cities and accommodate development and growth at newly established urban agglomerations. This process features the emergence of new metropolitan areas, cities and settlements along with the conversion of rural lands to urban land uses. Improvements in transportation and communication infrastructures are pivotal to enable these polycentric economic linkages. Represented by the increased mobility of workers, business travelers, goods, and information, these interactions on the other extreme may lead to urban sprawl that generates severe traffic congestion, air pollution and high energy consumption from longer commuting trips, occupies excessive agricultural lands, and intrudes on environmentally sensitive areas. As these aspects of an urban-regional system bring about increasing pressure from the rapid population growth and low density development, it is imperative to coordinate policies at this expanded geographic scale.

Megaregions, the conceptualization of urban-regions in the America 2050 initiative led by the Regional Plan Association (2006), are defined on five major categories of relationships: “environmental systems and topography, infrastructure systems, economic linkages, settlement patterns and land use, and shared culture and history”. A more cohesive and stronger megaregion will share a higher degree of these common characteristics. Borrowing this definition, I situate my research in the context of urban-regions and limit myself to the two intervening aspects of economic linkages, and settlement patterns and land use. In other words, it is my intention to study the causal relations between the underlying processes of agglomeration economies and the spatial structures of urban-regional development at both intra- and inter-urban scale within the extent of urban-regions.

1.2 Research Statement

To gain understanding of the urbanization process in emerging urban-regions, I call upon the “second nature”, agglomeration economies (functional processes), to explain the formation of urban agglomerations (spatial structures and land use patterns such as metropolises, cities, and urban centers) across scales. Broadly speaking, I draw upon the agent-based theories of urban and regional spatial structure, a schema falling into microeconomic theoretic approaches of the urban and regional economics theorization tradition categorized by Briassoulis (2000) in her comprehensive review book on land-use change analysis. In particular, I am interested in the spontaneous emergence of agglomerations when numerous decentralized economic agents (e.g., households and firms) make decisions on their locations to pursue their own interest. Such a mechanism of urbanization features a self-organizing process of interacting atomistic agents that collectively form an urban agglomeration as a complex system. This allows me to exclude the alternative mechanism for urban formation (developed in urban economics), a centralized institutional approach relying on local governments or developers (large agents) who create cities as their intentions, in the literature of systems of cities (Abdel-Rahman and Anas, 2004).

1.2.1 A Synthesis of Agglomeration and Dispersion Forces

Externalities are known to be essential in the agglomeration forces because they generate a self-reinforcing "snowball effect" in which an increasing number of consumers/workers and firms come together to enjoy and benefit from either a larger diversity of goods/services or a higher degree of specialization in labor. In other words, spatial agglomeration itself generates an advantageous economic circumstance that sustains further agglomeration. In economic terms, it is called increasing returns to scale (or increasing returns in short). Scitovsky (1954) distinguishes two types of externalities: technological externalities (spillovers) and pecuniary externalities. The former refers to the nonmarket interactions generating increasing returns external to households and firms that directly affect their utility and production due to the spatial proximity to each other. Thus, their influences decay with distance and are usually constricted by limited geographic areas (e.g., intra-urban districts or cities). Technological externalities are mostly viewed as black boxes through which complex micro-interactions (e.g., individual communications and knowledge spillovers) are modeled in an ad hoc way similar to the neighborhood effects in spatial models.

In contrast, pecuniary externalities, as the by-products of market interactions, relate to the increasing returns arising from market exchanges among firms and between firms and consumers/workers mediated by the price mechanism. They are rooted in the interplay between pricing decisions and location choices of firms that concentrate production in large markets (access to consumers and workers) while seeking to avoid competitions and trading off transport costs of products and input factors. Thus, increasing returns must exist internal to firms that would otherwise disperse to serve each local market. Due to the Spatial Impossibility Theorem (Starrett, 1978), this is only relevant when the market is imperfectly competitive. Recent advances in the new economic geography (NEG; Krugman, 1991) have employed the Chamberlinian models of monopolistic competition (Spence, 1976; Dixit and Stiglitz, 1977) to provide clearer origin for these pecuniary externalities. Firms' internal increasing returns lie in their product (inputs) differentiation which responds to consumers' (firms') preferences for variety (specialization). In addition, this approach enables a general equilibrium framework that is more problematic under oligopolistic competition setting as an alternative modeling approach for imperfectly competitive market. Finally, NEG models are also consistent with the atomistic agents approach adopted here rather than the large agents considered in the models of oligopolistic competition.

According to Marshall's threefold classification, the sources of externalities can be summarized as: 1) the backward (demand) and forward (supply) linkages due to specialized input providers in final and intermediate products markets; 2) the advantage of a large pool of labors with similar and specialized skills; 3) the creative activities due to communications, the exchange of information, and knowledge spillovers. Among these, NEG selectively focuses on the market linkages (the first of the list) as the centripetal (agglomeration) force because it remains difficult to model explicitly the micro-foundations of the others. On the other hand, the centrifugal (dispersion) force in NEG models rise both from the increasing competition between firms that are agglomerated and from the assumed spatial immobility of resources, such as land and labor. Specifically, these immobile resources are involved in the production of an agriculture sector spatially spread, thus they either generate demands or supply inputs that are dispersed. This convenient assumption has been criticized as arbitrary since in reality neither is labor immobile nor is city spaceless. Because urban agglomeration (manufacturing sector) takes no land and is treated as a dimensionless point in their abstraction of space (e.g., two-region, racetrack, and continuous line), NEG models always neglect agglomeration diseconomies such as urban costs (land rent, commuting costs) that are more relevant in developed economies. This centrifugal force is the emphasis of urban land use theory (Alonso, 1964; Mills, 1967) and systems of cities theory (Henderson, 1974) developed in urban economics. Thus, it is my intention to synthesize the centripetal force from the NEG and the centrifugal force from urban economics in seeking an in-depth explanation of urban agglomerations in urban-regions by taking into account the micro-foundations of agglomeration economies.

1.2.2 Interdependent Polycentricity across Scales

This section aims to disentangle the agglomeration and dispersion forces working at both intra-urban and inter-urban spatial scales and briefly discusses the theoretical models that incorporate them in the literature.

At the intra-urban level, households and firms seek to congregate because they need to interact on a daily or a short-run basis for various economic and social purposes such as commuting and shipping. For instance, in the service economy I observe clusters of stores selling similar goods and employment centers hosting different kinds of jobs. In such cases, the agglomeration forces are generated by market interactions among firms and consumers/workers for consumption of a variety of services, consumption goods and intermediate goods. On the other hand, the main dispersion force lies in the urban costs borne by workers and firms residing in large agglomerations. That is, workers compete for housing and bear commuting and shipping costs while firms compete for land and bear wage costs, and for both their land rents and transport costs increase with the size of urban agglomerations as the population expands. As a result, although stronger agglomeration economies exist for a larger city, growing urban costs may push jobs from its urban core either to its suburb or to other distant but smaller cities. However, it has been observed that large cities become polycentric with the formation of secondary business districts (SBDs). In SBDs, workers bear lower commuting costs and housing rents and firms pay lower wages and land rents while they both keep enjoying most of the benefits created by large urban agglomerations. Thus, I can expect that metropolises are able to maintain their attractiveness with a polycentric structure that reduces the average urban costs. In fact, this phenomenon is well recognized in most metropolitan areas of the United States.

Polycentric urban models considering the aforementioned internal urban forces have been mainly developed in urban economics. Two types exist in line with the atomistic agents approach. First, multicentric models pre-specify multiple centers/sub-centers and leave their formation unexplained (White, 1988; Helsley and Sullivan, 1991). In contrast, the second type aims to explain the existence of agglomerations. These non-monocentric models relax the assumption of monocentricity and endogenize the formation of centers/sub-centers by taking account of externalities explicitly (Ogawa and Fujita, 1980; Fujita and Ogawa, 1982; Fujita, 1988). Among the second type, models incorporating technological externalities have been demonstrated to manifest both monocentric and polycentric structures at multiple equilibria (Fujita and Ogawa, 1982; Fujita and Smith, 1990; Anas and Kim, 1996; Lucas and Rossi-Hansberg, 2002; Berliant and Wang, 2008). However, existing non-monocentric models with pecuniary externalities are only able to exhibit monocentric structures at the equilibrium (Fujita, 1988; Anas and Xu, 1999; Anas and Liu, 2007; Picard and Tabuchi, 2013). Although Fujita (1990) extends the formulation of Fujita (1988) to a two-sector model with intermediate goods, employment, and commuting, and conjectures that it would exhibit multiple equilibria and polycentric patterns, their systematic analysis is lacking, and has been left for future study. Except this, none of the other polycentric models with pecuniary externality has so far been realized formally to the best of our knowledge.

At the inter-urban level, what matters are the inter-industry linkages, the intercity trading of goods and services, and the labor migration across urban agglomerations through the product and labor markets in the long run. At this scale, pecuniary externalities play a significant role arising from the imperfectly competitive market where firms and consumers/workers participate and exchange for goods/services and economic factors (Fujita and Thisse, 2002). Such externalities lie at the heart of models of monopolistic competition that assume preference for variety (horizontally differentiated goods) on the demand side and increasing returns to scale on the supply side. In the process of urbanization, firms balance the advantages of being close to existing major cities to enjoy larger demand but fierce competition against those associated with less competition but smaller markets in the emerging cities. As differentiated varieties of different goods are traded at the inter-urban level, new cities would distribute themselves according to a network structure. Positive inter-urban trade costs are critical here in that, without considering it, cities would be treated like floating islands and their location is irrelevant. The intriguing question is whether there exists regularity in the inter-urban spatial organization and whether it follows a central place hierarchy (Christaller, 1933; Losch, 1940) or the polycentricity (as defined in section 1.1).

Inter-urban spatial structure models that deal with pecuniary externalities at this scale are notably NEG models of urban systems (Fujita and Krugman, 1995; Fujita and Mori, 1997; Mori, 1997; Fujita et al., 1999; Tabuchi and Thisse, 2011). By modeling the costly trade of goods and perfect mobility of workers as backward and forward linkages between urban agglomerations, this approach presents the evolutionary formation and spatial distribution of agglomerations along a continuous line space as population grows. The resulting urban systems, where each agglomeration as a clustering of industries is surrounded by agricultural hinterlands, suggest some regularities of the emerging inter-urban spatial structure. That is, it morphologically provides a reminiscence of and some justification for central place theory in terms of the size, relative location, and industrial composition of cities; in a sense of functional linkages it fosters bidirectional trade (of differentiated goods) between all cities, which represents horizontal relations among places, and thus exhibits functional polycentricity.

However, due to the inter-urban focus of NEG models, the internal spatial structure of cities is abstracted as spaceless points (e.g., industrial production does not consume land), and thus the induced urban costs are entirely ignored as a dispersion force. In contrast, systems of cities theory developed in urban economics emphasize the monocentric urban internal structure and the trade-off between commuting costs and land rents (a simplified setting for urban costs) while giving much less consideration to the inter-urban spatial structure by assuming costless trade between cities (Henderson, 1974; Anas, 2004). Some models have been proposed as the unification of NEG and the systems of cities approaches either with a standard two-region (Tabuchi, 1998; Ottaviano et al., 2002; Murata and Thisse, 2005; Anas and Xiong, 2005; Tabuchi and Thisse, 2006; Thisse, 2010; Gaigné et al., 2012) or an extended equidistant multi-region spatial configurations (Anas and Xiong, 2003; Anas, 2004; Tabuchi et al., 2005). However, they usually embed monocentric urban economics models core-periphery (CP) models of NEG with two or multiple regions. Due to the predefined symmetry (in terms of the spatial cost between any pair of regions) and discreteness of space, they are unable to represent the richness of spatial interaction in an asymmetric manner or distinguish at which spatial scale the polycentric structures emerge (e.g., either a new city or a new sub-center with a city). A more realistic approach would entail the study of the tension between inter-urban linkages and intra-urban costs and their impact on the spatial organization of agglomerations in a continuous and asymmetric space (e.g., a linear or racetrack economy), which is absent from the current literature.

The space-economy of urban-regions hinges on the interactions among agglomeration and dispersion forces across different scales. Specifically, the location of economic activities within and across urban agglomerations is the consequence of the interplay between various spatial frictions at different scales: commuting costs of workers and shipping and trading costs of firms at both the intra-urban and the inter-urban scales. In other words, these costs are associated within the following trade-off: concentrating workers and firms in a small number of large cities minimizes intercity trading costs but yields longer average distance for commuting and shipping; dispersing workers and firms across numerous small cities has the opposite effect. As demonstrated by models relying on urban costs as a centrifugal force, the economy involves initial dispersion, then agglomeration, and finally re-dispersion (a bell curve), as inter-urban transport costs keep falling. In contrast, when intercity transport costs are sufficiently low and start to foster re-dispersion, a decrease in urban costs (e.g., faster commuting) can sustain the agglomeration equilibrium. This can be achieved by advocating a polycentric structure internal to large cities. In other words, the polycentricity at the intra-urban scale will slow down the re-dispersion from the large cities to the small towns, lower the level of interactions among them, reduce the possibility of emergence of new agglomerations, and thus downgrade the degree of polycentricity at the inter-urban scale. Hall and Pain (2006) have found empirical evidences for this theoretical mechanism that polycentricity is scale-dependent, i.e., if polycentricity exists at one scale, there may be monocentricity at another.

To incorporate the mechanism that spatial structures (polycentricity vs. monocentricity) at different spatial scales are interdependent, a model needs to derive polycentric structures at both intra-urban and inter-urban scales. To the best of my knowledge, the only model suited to this purpose was developed by Cavailhès et al. (2007). Through a simple multiscalar theoretical model, their study suggests that the polycentric structure internal to a city (i.e., more subcenters) fosters the agglomeration of the city as a whole, which at the interregional level otherwise implies a monocentric structure (fewer large cities, more small towns). Their model enables a polycentric urban model that incorporates land rent, commuting costs, and communication costs between firms at the predetermined CBD and the induced SBD, while assuming that firms do not occupy land and thus all (sub)centers are dimensionless. However, this model assumes zero transport costs for goods consumed internally for simplicity and it does not allow trading between firms such that the input-output linkages are missing. Moreover, at the inter-urban level it still has a two-region discrete setup, which makes it impossible to test asymmetric transportation configurations between regions. Thus, it limits the capability to model the impacts of transport costs at different scales on the intra- and inter-urban spatial structures.

1.3 Research Objectives and Contributions

The general goal of this research is to study the causal relationships between the underlying processes of agglomeration economies and spatial structures in terms of polycentric and monocentric development at both intra- and inter-urban scales within the extent of a megaregion. Specifically, I develop an economic model of urban-regions that incorporates centripetal (backward and forward linkages) and centrifugal (urban costs) forces at different geographic scales taking advantage of pecuniary externalities. This model thus enables to test for the possible interdependency between spatial structures across scales in various policy contexts. This effort leads to the following two goals.

The first goal is to develop a theoretical framework that models the interplay between market linkages and spatial costs across scales. It accounts for the interdependency of spatial structures at intra- and inter-urban scales. Specifically, two objectives are in order for constructing the framework in line with the principles outlined above.

First, as the building block, to develop polycentric urban models that feature pecuniary externalities. Particularly, it extends the non-monocentric models developed in Fujita (1988, 1990) and incorporates monopolistic competition among producers and the preference for differentiated final and intermediate goods of consumers. Taking into account these demand and supply linkages, firms and consumers compete for land while trading off their costs for wages and transporting goods and labors, respectively. Firms tend to locate close to workers for good access to larger labor pools; workers seek to be near to firms for shorter commuting and better employment opportunities. It is expected that a polycentric structure may emerge endogenously when incorporating both final and intermediate goods as a reflection of developed economies prevailing in U.S. metropolitan areas.

Second, in full conformity with the generic framework, to build a hyper-model for systems of polycentric urban models that incorporates urban costs and the internal spatial structures of cities. Specifically, it is to embed a number of polycentric urban models into one hyper-model structure of urban systems (Fujita et al., 1997; Tabuchi and Thisse, 2011). For the sake of simplicity, only one differentiated sector is considered and the urban systems model is constructed on a racetrack space. In essence, the agglomeration force hinges on the land competition between all agents and the commuting and shipping costs borne by workers internal to cities. The aim is to examine, from a theoretical perspective per se, whether certain evolutionary paths of equilibria for intra- and inter-urban spatial structures would emerge with the change of transport technologies and the free migration of labor across cities.

Thus, this framework can be used to construct a multiscalar urban systems model that enables the derivation of polycentricity at each spatial scale by accounting for the interaction of economic linkages and spatial frictions across scales. Once implemented, the proposed model produces simulated urban systems, which allow us to examine the evolution of intra- and inter-urban spatial structures in terms of monocentricity/polycentricity and agglomeration/dispersion. Consequently, I am able to address issues such as whether the change of intra-urban spatial structure due to local factors (e.g., urban costs) will affect the spatial structure at the inter-urban scale, or whether the change of inter-urban spatial configuration due to global factors (e.g., trade costs) will exert influence on the internal urban structures. In sum, this framework contributes to the theoretical body of literature on the spatiotemporal organization of urban and regional economies in terms of how spatial scale makes differences on the crucial agglomeration and dispersion forces at work.

The second goal is to extend the theoretical model into a two-dimensional geographic model that can scale to real world applications. With this objective, I transform a stylized unidimensional spatial representation (linear or circular space) into a two-dimensional zonal geography connected by transportation networks. This more realistic depiction of space takes into account the fact that the accessibility to locations of households and firms varies across a network of places. An agent-based approach is employed to model the location choices made by households and firms. The resulting geographic model can better reflect the spatial structures of urban and regional development by coupling the first nature, a realistic heterogeneous space, with the space-economy. It is of great importance to empirically assess the theoretical framework by calibrating the geographic model against real world data. It also helps practical policy testing and scenario-based planning. For demonstration purposes, a real world application to the Carolinas region (including North and South Carolinas) is presented as a test bed for the geographic model. Given the complexity of the urban systems model itself as well as the geographic resolution and extent of simulations with changing parameters conducted in this modeling effort, massive computations are involved regarding to the hundreds of thousands of agents that participate in decision making, interactions, and location choices. In addition, model evaluation entails testing parameters varying over a wide range of values through a large number of simulations while model calibration requires a huge amount of data such as population, employment, commuting flows, and commodity flows. All these types of computational complexity must to be handled appropriately. This challenge can be tackled by leveraging high-performance computing (HPC) technologies to boost the computational performance of both theoretical and geographic models. The contributions of this effort are twofold. First, it contributes to the existing corpus of urban models by being the first to incorporate the economic agglomeration effect of NEG style into location choice modeling; second, it contributes to the computational aspects of existing large-scale urban modeling by enabling the support of HPC.

1.4 Dissertation Outline

The rest of this dissertation is organized as follows. Chapter 2 provides a thorough review of the literature on urban and regional modeling efforts made in different disciplines. Chronologically, three generations have been witnessed. Given this review, decisions are made on the appropriate approach to adopt for this study. Chapter 3 proposes the general multiscalar modeling framework tailored to the polycentric urban development in urban-regional systems. Chapter 4 details the construction of polycentric urban models by incorporating final goods and intermediate goods in an incremental way. Short-run equilibrium conditions for these intra-urban models are presented and the feasibility and stability of multiple equilibria are examined for a full understanding of model behaviors. Chapter 5 develops the multiscalar model of urban systems built on the polycentric urban models in chapter 4. It is followed by the determination of long-run equilibrium conditions and the dynamic adjustment process. With choice rules designed for detecting the transition between multiple long-run equilibria, simulations are conducted to study the evolutionary path of urban-regional development under exogenous change of transportation costs at both local and global scales. Chapter 6 proposes a geographic model that extends the theoretical model in chapter 5 with a realistic representation of space and an agent-based location choice approach for spatial equilibrium approximation. Simulations of the geographic model are performed to demonstrate its consistency with the theoretical models. Chapter 7 discusses how to leverage existing HPC technologies to tackle the computational complexity of the proposed theoretical and geographic models. Parallel models are designed and implemented in a powerful modeling environment to take advantage of heterogeneous HPC platforms. Chapter 8 applies the geographic model in a real world case study, the Carolina area, with discussions on the data availability and potential strategies to calibrate the model. A partial calibration procedure is employed with the limitation of data sources. Simulation results are presented under several scenarios to assess the applicability of the geographic model. Finally, this dissertation is concluded in Chapter 9 with discussions on future studies.

CHAPTER 2: LITERATURE REVIEW: URBAN AND REGIONAL MODELS

An urban and regional system is an open, dynamic and complex system that comprises various dimensions, scales, processes, actors and their interactions. Urbanization, manifested by the change of land use patterns, is a process of concentration of human population and activities. In order to better understand the intrinsic mechanisms of urbanization, urban modeling has become a multidisciplinary effort (from disciplines such as geography, planning, regional science, urban and regional economics, and environmental science) that intends to create scientific models to account for functions and processes that generate urban spatial structures at either intra-urban or inter-urban scales. At the intra-urban scale, the focus is on the internal structure of a city which itself is treated as a system, whereas at the inter-urban scale the external relations of cities are emphasized to explain the distribution of urban centers in a system of cities (or urban-regional system). Practically, to facilitate policy making for planning and sustainable development, these models are implemented as computer programs fed with empirical data to make predictions of future urban development patterns.

This scientific modeling approach is based on the perspective of social science positivism and on the paradigm of rationalism in planning that started from the early 1950s. The first generation of urban-regional models emerged in the 1950s and culminated in the 1960s. Following it, two other generations followed, which can be distinguished not only by the time frame of their development, but also by the theoretical, methodological, and operational dimensions they bear on. In the following review, I pay special attention to how models developed from different fields to address the causal relationships between the underlying economic processes (the human behavioral aspect) and the resulting urban spatial structures and land use patterns (the geographic aspect). This review is not meant to be comprehensive, but rather illustrative of the existing paradigms.

2.1 The First Generation

2.1.1 Origins of Urban Modeling

The first generation of large-scale urban models started to be developed in the 1950s for U.S. metropolitan areas in order to contribute to the paradigm of rational urban planning dominant in most western countries at the time. The concept of a metropolitan area is defined based on an urban core with a substantial population nucleus surrounded by adjacent areas with less population but having a high degree of economic and social integration with the urban core (US Census, 2010). This hierarchical-nodal structure, consistent with central place theory (CPT; Christaller, 1933; Losch, 1940), was regarded as the ideal spatial organization of functions and activities by planners (Low, 1975; Hall, 1997). CPT concerns a monocentric urban structure characterized by a hierarchy of central places, where their spatial locations relative to one another are related to their size as marketplaces and the level of the functions they provide (i.e., larger cities are surrounded by smaller towns and provide goods or services to smaller ones; Mulligan, 1984). In other words, the center is self-sufficient in that it provides the full range of goods/services, whereas central places at lower levels are dependent on central places at higher levels for the supply of goods. The study of the spatial organization of urban systems is rooted in location theory (von Thunen, 1826; Weber, 1909), which addresses the questions of what economic activities are located where and why, given an exogenously located marketplace. With the monocentricity assumption, these models by von Thunen and Weber focus on the competition for land among various economic activities (e.g., agricultural or industrial productions and goods) by trading off between land bid rents and transport costs to the central marketplace. Grounded on this principle, two main approaches have been developed in the traditions specific to different disciplines, namely a disaggregate approach and an aggregate approach. Both hinge on the joint determination of travel and location decisions.

2.1.2 Disaggregate versus Aggregate Approaches

Applied to an urban context, new urban economics models (NUE; Alonso, 1964; Muth, 1969; Mills, 1967) follow the tradition of von Thunen's agricultural land use model with a microeconomic foundation. These models assume the internal structure of a city as monocentric with a prespecified center of production activities (CBD) where all employment is concentrated. Workers optimize their residential locations by trading off commuting cost and land rent on a competitive land market. The equilibrium land use pattern is characterized by concentric rings of residential areas surrounding the CBD and a decreasing gradient of residential density with distance from the CBD. However, a crucial question remains: why would economic agglomeration occurs in the center? Specifically, what are the exact agglomeration forces and economic mechanisms that push all firms to the urban center? Furthermore, treating the geography as a featureless continuous space with a dimensionless center, the commuting cost (dependent on distance) to the center is the only determinant of land use pattern. In other words, these models fail to capture the heterogeneous landscape features in reality and ignore their effects on land use decisions (the Ricardian tradition). Therefore, their ability to explain spatially heterogeneous land use patterns at a disaggregate level is limited, although they model decision-making at an individual level. Due to the above reasons, this approach has been more theoretical, rather than empirically applied.

More in line with the traditions in regional science and planning, the spatial-interaction-based Lowry-Garin type models (Lowry, 1964; Garin, 1966) make use of a gravitational analogy in social physics. Models of this kind were first built to coordinate empirical land use and transport planning with the recognition of the "land-use and transport feedback cycle". This approach takes into account the spatial distribution of aggregate flows of population and employment across rather coarse geographic delineations represented by discrete zonal units connected by transportation networks. Thus, economic activities can be allocated and land use can be determined accordingly in each zone. Combined with macroeconomic models such as economic-base theory and input-output analysis, this strand of urban models can also account for the spatial distribution of macro-level flows represented by production and consumption between various economic sectors (e.g., manufacturing, retail), and they together constitute the aggregate approach of first generation urban models. The two strands together lay the basis for a school of operational urban models, collectively called integrated land-use and transport models (ILUT) that have been empirically calibrated with real world data and applied practically in planning to facilitate policy analysis.

2.1.3 Mismatch between Theories and Empirics of Urban Models

However, the era for the first generation of urban models ended in the mid-1970s because of the widening discrepancy between the models and the changing planning context. This mismatch can be attributed to several aspects. First, there was a lack of theoretical foundation that could reflect the reality of spatial organization of urban systems. Though the insights offered by CPT are fundamental and intuitive to the understanding of organization of spatial economies of cities, CPT has been largely descriptive in that it gives no explanation about why central places should emerge. In other words, the plausibility of a hierarchical structure has been suggested by Christaller, yet no microeconomic underpinnings have been developed to account for how such a hierarchy emerges from the interactions of households and firms making location decisions. Furthermore, with rapid suburbanization between 1950 and 1970 (Berry, 2002), the decentralization of population and economic activities radically transformed the spatial structure of U.S. cities such that their spatial organization became increasingly disconnected with the classic monocentric model represented by CPT. A better formal theoretical model is lacking to explain the new development of the suburbs and the formation of subcenters, which empirically deviates from the monocentric hierarchical models.

Second, urban models of the first generation, whether along the aggregate or disaggregate approach, are essentially static in that they assume that an equilibrium exists in the spatial structure at a cross section in time through a general equilibrium or a partial equilibrium mechanism in modeling. Lacking the capability to account for dynamics, these models are unable to explain and respond to changes in spatial structure. Last but not least, due to the lack of data at the disaggregate level in this period, only macroscopic models could be applied meaningfully in most practical cases. It raised the question for aggregate models, whether their model components and units of analysis are the appropriate representation of the key elements operating at different scales of urban systems. Oversimplification that sometimes results from aggregation may substantially eliminate spatial heterogeneity in the models and lead to inadequate richness of detail in the outcome to be useful for policy decision-making.

2.2 The Second Generation

2.2.1 Paradigm Shift

From the early 1970s through 1980s, the second generation of urban modeling emerged. During this time a paradigm shift was witnessed in planning, geography, urban studies and related social sciences. This shift is marked by the abandonment of the scientific approach, comprehensive plans in planning, and the end of the course of the quantitative revolution in the social sciences. It was partially due to the disconnection between theories and urban reality and the lack of practically useful urban models. Instead, spurred by ongoing urban decline and deindustrialization in this period, the emphasis in planning practice shifted to individual cities from wider urban-regional areas as planning units with a more routinely pragmatic and managerial approach (Batty, 1994). In urban studies, the behavioral approach took the role of positivist paradigm with a focus on the spatial behavior of households and firms at the micro level and how their decisions shape the urban systems at large (Batty, 1994). As a result, only a few operational urban models were developed in the U.S. during this period (Wegener, 1994). On the other hand, significant theoretical improvements in urban modeling still continued being achieved in academia. In addition to the theoretical developments, there were continuous efforts to refine existing applications of operational urban land-use and transportation models in terms of model calibration and disaggregation to fit available data (Batty, 1976).

2.2.2 Advances in Theoretical Modeling

The theoretical enhancements of urban models of the second generation emerged from three main aspects: optimization as a unifying approach, dynamics of urban growth, and alternative intra-urban and inter-urban structures.

Optimization as a Unifying Approach

First, there is a synthesis of urban theories that combines location-based activity and transportation modeling via the general framework of optimization (Wilson, 1967, 1970; Anas, 1983). This framework unifies the aggregate and the disaggregate approaches by making connections between spatial interaction models derived through entropy maximization, microeconomic land market models based on utility maximization, route searching through cost minimization, and discrete choice analysis of travel behavior underpinned by random utility maximization. Thus, it can provide a consistent representation of economic behavior through various forms of optimization and enables an operational approach to construct theoretically solid urban models.

A similar approach has been developed in the natural sciences tradition for land use/cover change (LUCC) studies (Irwin and Geoghegan, 2001; Verburg et al., 2004; Brown et al., 2006). These spatially explicit models replicate the pattern of land-use change and explore processes that lead to this pattern. To this end, land suitability for certain use is assumed to be driven by biophysical (e.g., soil and slope) and socio-economic (e.g., land price and accessibility) characters of a land parcel. This suitability concept can be interpreted on the behavioral ground that, based on a parcel's characteristics, the landowner makes the optimal land conversion decision that maximizes the expected utility given various land uses (van Schrojenstein and Lantman, 2011). This validates using a discrete choice framework. Taking advantage of remotely sensed data at finer spatial resolutions, LUCC models employ increasingly disaggregate grid cells or landscape units. However, their units of analysis, instead of being individual decision-makers, may pose problems such that the boundaries of individually owned land parcels are not in line with the boundaries of land cells. The economic interpretation of implicit landowners is "ad hoc" and it limits LUCC models' ability to explain underlying economic processes.

Dynamics of Urban Growth

Second, following Forrester's (1969) early attempt to introduce dynamics into urban systems theory, theoretical developments of disequilibrium models focused on the dynamic process of nonlinear growth of urban systems that can generate not only continuous change but also discontinuity and catastrophe. Harris and Wilson (1978) embed a spatial interaction model of retail centers in a dynamic framework that can give rise to nonlinearities and qualitative change when some parameter exceeds a critical value. Allen and Sanglier (1979, 1981) build a dynamic model of a central place system and show how it can generate bifurcations where new centers may emerge because of random fluctuations and grow along different paths otherwise. In particular, the growth of population and employment in their model is interdependently determined by accounting for both agglomeration economies and congestion diseconomies in an ad hoc way.

Specifically, existing employment/population attracts new employment/population, but eventually the capacity of places hits a ceiling. Their models have been calibrated and applied to a number of cities and geographies (Allen, 1997). Although these dynamic models may employ “ad hoc” specifications for considerations of the economic motivation of individuals, they lack microeconomic foundations as other non-economic models do. Furthermore, their dynamic behavior is backward- rather than forward-looking, that is individual decisions are dynamic if they consider future expected benefits and costs.

Alternative Intra- and Inter-Urban Structures

Third, due to the increasing incongruity between the monocentric hierarchical models and tremendous ongoing suburbanization in cities across the United States, increasing consideration has been directed to alternative theories of the spatial organization of urban systems that structurally emphasize the bi-directional and mutual functional linkages between the urban core and subcenters, or between large cities and small cities, or even between large metropolitan complexes (Pred, 1977). It can be represented by a polycentric network model that contrasts with the central place model of urban hierarchies where linkages are unidirectional and only from large cities to small ones (correlation between urban size and urban functions). Therefore, it allows relationships between cities to be not only competitive (as in urban hierarchies) but also cooperative. As the accumulation of studies on empirical identification of subcenters (Gordon et al., 1986; Richardson, 1988) in urban economics and regional science, formal theoretical models of polycentric organization of economic activities were developed at both intra-urban and inter-urban scales.

To account for polycentricity and suburbanization within a city, there have been multicentric models with multiple exogenously specified centers assumed (White, 1976; Sullivan, 1986) and non-monocentric models with endogenous formation of the city center and subcenters (Ogawa and Fujita, 1980; Fujita and Ogawa, 1982; Fujita, 1988). Both types of models were constructed through either general or partial equilibrium approaches. Although multicentric models allow economic activities to be analyzed under the spatial structure of multiple centers, why these centers exist at all remains unexplained. In contrast, by incorporating scale economies through non-market or market interactions, non-monocentric models provide a framework to endogenize interdependent location decisions made by households and firms and to determine their distribution jointly. It thus provides a theory of the spatial agglomeration of economic activities (monocentric or non-monocentric) of a city without a priori assumption.

On the other hand, the theory of systems of cities started to develop in urban economics (Henderson, 1972, 1974) at the inter-urban scale (e.g., a large metropolitan area encompassing multiple cities). Henderson assumes that each city has a monocentric internal structure and a finite size, which is based on Mills' (1967) work about city formation determined by the trade-off between the scale economies in production and the commuting cost borne by the workers. In particular, city developers or local governments play central roles in creating new cities to maximize their utilities. When costless trading is assumed, cities benefit from specialization in the production of certain goods, because commuting costs and land rents will start to increase when hosting the production of multiple goods within one single city. Once it allows different degrees of scale economies for the production of different goods, his model is able to describe how a hierarchy of cities by size and type emerges, but it has nothing to say about the city locations because cities are treated as "floating islands". In other words, the urban spatial structure described in CPT remains unexplained.

2.3 The Third Generation

From the early 1990s onward, urban system theory and modeling efforts have been advanced in three main streams, which feature the third and latest generation of urban-regional models.

2.3.1 Polycentric Urban Systems and Economic Agglomeration

The first stream, along the same line as Pred's polycentric network paradigm, features renewed interest in the spatial organization of urban systems by concentrating on polycentricity at larger spatial scales. It is marked by work on edge cities, urban networks, polycentric urban regions, and megaregions (Kloosterman and Musterd, 2001; Scott, 2001; Hall and Pain, 2006; RPA, 2006). Fueled by the advances in transportation, information and communication technologies and the growing importance of the service economy and globalization, historically and spatially independent cities and regions (urban regions or metropolitan areas) become increasingly interconnected through external cooperative linkages (e.g., complementarity between specialized cities due to localization economies and agglomeration of diversified cities due to urbanization economies), and tend to form an economic and social coalescence comprising a larger morphological and functional polycentric urban region, which exhibits many characteristics shared with Gottmann's Megalopolis. This polycentric network paradigm has been introduced into urban and regional planning to enhance territorial cohesion and regional economic strength and competitiveness, while less theoretical and empirical models have been developed to test the polycentric development policies.

Advances in spatial economics in the urban and regional context primarily come from two fields. The new economic geography (NEG) appeared as applications of trade theory to spatial economy in 1990s (Krugman, 1991; Fujita et al., 1999; Baldwin et al., 2003). It presents a unified theoretical approach to explain the emergence of various economic agglomerations in the geographic space at different scales (e.g., urban system, regional level, international level) by conceptualizing the balance between centripetal (aggregative) and centrifugal (dispersive) forces in the interplay among increasing returns, transport costs, and the mobility of economic factors. At the scale of urban systems, this approach enables the evolution of the spatial structure of city systems (Fujita and Krugman, 1995; Fujita and Mori, 1997; Fujita et al., 1999) and megalopolises (e.g., industrial belt; Mori, 1997) by combining general equilibrium with growth dynamics. It is noteworthy that these models yield multiple equilibria of spatial patterns and adopt self-organizing mechanisms to the selection of equilibrium.

Fujita and Krugman (1995) provided a microeconomic approach to endogenously determine the urban and agricultural land-use pattern and, for the first time, justified the emergence of a monocentric urban structure that is predetermined in the classical von Thunen model. Furthermore, Fujita, Krugman and Mori (1999) take a first step to show that as population grows, a fairly regular hierarchical city-system emerges in the space-economy, where cities at higher levels provide a larger group of goods/services than cities at the lower levels. In contrast to the unidirectional linkages in the central place model, two-way trade exists between cities; large and diversified cities (at high level) trade a larger variety of goods/services than small and specialized cities (at low level). The resulting urban networks are much more complex than with the hierarchical central places model in that "they combine both the hierarchy of various centers with the existence of networks of cities exchanging specialized goods and services", which is more consistent with the urban systems of modern space-economies described by Pred (1966, 1977). However, due to their theoretical considerations, these models usually employ a highly abstract and simplistic representation of the space (linear or circular space) and are far from being operational.

In the field of urban economics, models of systems of cities continue to be developed. The aims are to address three fundamental issues (Abdel-Rahman and Anas, 2004): 1) the number and size distribution of cities in the economy; 2) the variation of industrial composition among cities with different sizes and the efficiency of such variation; 3) the distribution of labor with various skills within and between cities in the system. Following the tradition of Henderson (1974, 1987, 1988) or the NUE approach, these models usually assume that cities have a monocentric internal structure and explicitly incorporate agglomeration economies including forces derived from market or non-market externalities (production side) or from local public goods (consumption side). Specific studies concern various issues such as no trade vs. trade with zero transport cost between cities, predefined vs. endogenous industrial composition of cities, and identical vs. heterogeneous labor.

In a sense, the NUE and NEG approaches are complementary to each other. In NUE, cities have an internal spatial structure while the transport cost between cities is suppressed. In NEG, cities are dimensionless but inter-city trade is not costless. In other words, the two approaches focus on spatial frictions at different scales. The former emphasizes the role of commuting costs in the formation of cities, whereas the latter pays attention to inter-city transport cost. It has been suggested that the synthesis of the two would lead to a comprehensive theory of systems of cities (Tabuchi, 1998). On the institutional mechanism of city formation, the atomistic agent approach taken by NEG contrasts with NUE's large agent approach, where city developers or local governments are responsible for creating new cities by coordinating the actions of firms and workers. In NEG, the emergence of cities is the consequence of a myriad of individual decisions made by economic agents who are not planning to form a city as a priori but rather seeking their own benefits. That is, "a city appears as a complex system whose existence is the result of a self-organizing process" (Fujita and Thisse, 2002, p. 354).

2.3.2 Urban Complexity and Agent-based Models

Second, in the same vein as the early development of dynamic urban models but with a rather aggregate approach, the new disequilibrium modeling paradigm arising during this period views cities as open, complex, and self-organizing systems and aims to incorporate both temporal dynamics and spatial heterogeneity that characterize urban processes in a highly disaggregate and decentralized approach. Therefore, these models are advocated to accommodate the increasing availability of finer-resolution land use/cover data in space and time facilitated by the advent of new information technologies. This includes the penetration of information technologies such as personal computers into the entire modern society and the diffusion of geospatial technologies such as geographic information system (GIS) and remote sensing (RS) throughout urban planning practices in terms of data collecting, digital mapping and spatial database management. Along with the developments in complexity theory, this approach emphasizes a new bottom-up paradigm, rather than the top-down fashion of previous aggregate urban models, where the accumulation and aggregation of numerous localized decisions and decentralized interactions in the disaggregate spatial-temporal dimensions give rise to the evolution of macroscopic urban structure as an globally emergent property. With a focus on urban morphology, cellular automata (CA), as a typical urban model of this paradigm, has gained significant popularity due to its simplicity and explicit spatial characteristics to represent the city and region as a fine-scale grid, where urban development and land use change can be simulated as diffusion processes that are reflected as the iteratively changing state (land use type or development) of each cell on the grid governed by decision rules and neighborhood effects.

Interestingly a decision-making process with a dedicated human behavioral point of view is the central concept of this paradigm. Agent-based models (ABM) provide an explicit representation for individuals or group entities as agents/actors to simulate their decision-making behaviors, interactions and responses to the urban context and policy environment through various processes of change with different speed ranging from daily travel to relocation, from housing choice to real estate development. Among those dynamic and behavioral models, CA has been widely applied empirically to a variety of domains such as natural sciences, geography, urban studies, and LUCC studies, but only a few of them have become operational models in that, being primarily physical environment driven, they pay little attention to transportation and spatial economy such as land price and travel costs; their highly disaggregate and dynamic structure raises issues for model calibration, which makes them remain indicative rather than predictive; as a result they lack the capability to test policies and support practical planning.

2.3.3 Activity-based Approach and Microsimulation

Third, the continuation of the behavioral paradigm that has prevailed in planning and urban studies since the 1970s, which is reflected in accommodations for the changing policy environment in transportation planning and travel demand management during this time, serves as a major catalyst for the activity-based research on travel behavior in particular, and human behavior in general. The activity-based approach is intellectually rooted in activity analysis. This body of literature was singularly initiated by Hagerstrand (1970), Chapin (1974), and Fried et al. (1977) on the patterns of activity behavior, the constraints on and the social structural causes for activity participation under the space-time context. Along the same line, the tenet of this approach is that travel decisions are based on the demand for activity participation, and therefore the understanding of travel behavior depends on the understanding of the underlying activity behavior. As such, activity-based models aim to replace the traditional trip-based aggregate models that generate trips based on a spatial interaction framework with a highly disaggregate approach. This approach focuses on the formation of daily activity agendas, the scheduling of activity programs, and the choice process of associated decisions for participation performed by individuals and households at the micro-level, which constrains the spatial pattern of their activities and characterizes their travel behaviors.

As a typical realization specialized in transportation planning, the activity-based approach belongs to a broader concept named microsimulation which has close relationship with the CA/ABM approach. It allows the simulation of the decision-making process of and complex interactions between individual actors within an open system at the micro-level (disaggregate). Therefore, it enables to trace the evolution of the whole system over time at the macro-level (aggregate) by accounting for path dependence and stochastic elements. With the advances in computing power and increasing availability of disaggregate data, microsimulation has been introduced to land-use and transportation modeling to account for the dynamics and complexity of urban systems, exemplified by practical applications such as TRANSIMS (Nagel et al., 1999), UrbanSim (Waddell, 2002), and ILUTE (Miller, 2004). These models were intentionally developed as fully operational models that have targeted purposes for planning support, efficient computer programs for the implementation of model algorithms, clear specification for data requirements, well organized procedures for empirical calibration or even validation, and powerful capability for policy analysis. However, due to their enormous data requirements, they only have been practically applied to city contexts where data availability is not an issue.

CHAPTER 3: A GENERIC FRAMEWORK

The generic theoretical framework of multiscalar urban agglomerations (proposed as the first objective in Chapter 1) is illustrated in Figure 3.1. It also depicts how the theoretical framework guides the development of multiscalar urban-regional models with the real geography of land use and transportation and how the implementation of both theoretical and geographic models can be supported by high-performance computing technologies (the second objective).

This generic framework comprises two components, a polycentric urban model and an urban system model, corresponding to the intra-urban and inter-urban scales, respectively. In the polycentric urban model component, market linkages (pecuniary externalities) are the centripetal forces at work while the centrifugal forces are characterized by the market competition in products, land and transportation costs to overcome various spatial or transactional frictions when delivering either goods/services or production factors in different markets. The interplay between the two types of forces will allow the establishment of urban centers and subcenters, which amounts to a polycentric urban structure.

In this framework the behaviors of two basic types of economic agents, firms and households, are modeled in three different types of markets. In the commodity market, they represent the production and consumption sides and bear transaction and shopping costs, respectively. In the labor market, households supply labor to firms and bear commuting costs. In the land market, where location comes into play, both households and firms compete for the limited and immobile land resources to reside on and bear land rent costs. At this intra-urban scale, given a fixed population of households, short-run equilibria exist under conditions that all households achieve the highest and same utility level, all firms earn zero profit, and all markets clear their goods.

The urban systems model features similar centripetal forces as those prevailing at the intra-urban scale while inter-industry linkages become more prominent. The centrifugal forces at this inter-urban scale are not only derived from the aggregate combination of all types of firm and household costs internal to the urban model but also those rooted in the trade costs between urban models. At the inter-urban scale, long-run equilibria feature a series of comparative statics that involve dynamics arising from the inter-urban migration and exogenous population growth of households.

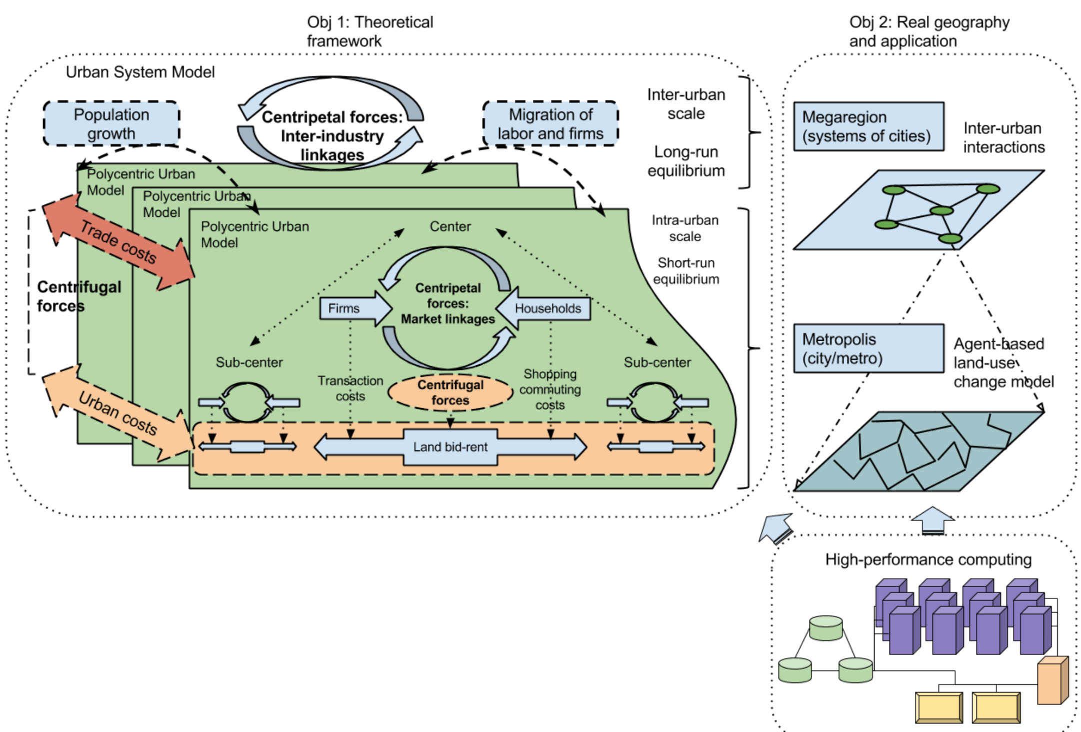


Figure 3.1: The generic framework

3.1 Increasing Returns and Monopolistic Competition Market

Based upon the above description, this framework characterizes the interactions between three fundamental principles of spatial economics: increasing returns (scale economies), transport costs of goods and services, and factor mobility, which is also the central idea of NEG. I will address the specification of these aspects separately in subsequent sections. Before that, a clear root and market form for the existence of increasing returns must be provided. Increasing returns arises from the location problem of a firm because economic activities are not perfectly divisible. Thus, each firm faces a trade-off between concentrating its production at a few places to minimize fixed costs and optimizing locations to minimize transport costs to its consumers and resources. The Spatial Impossibility Theorem (hereafter SIT, Starrett, 1978) states that without appealing to space heterogeneity (the first nature, comparative advantage models in neoclassical theory of trade) or technological externalities (non-market interaction due to spatial proximity), a perfectly competitive market equilibrium cannot exist with positive transportation costs for trade across locations. Indeed, if economic activities are perfectly divisible, a competitive equilibrium exists as exhibited in a general equilibrium model of a competitive economy such as the Arrow-Debreu (1954) model. This model assumes that consumers' preferences and consumption sets (including a relation of preference, a bundle of initial resources, and shares in firms' profits) and firms' production sets (containing production plans that describe input-output relations) are convex and commodities are differentiated by both their own properties and locations. Since convexity implying preferences for a combination of different commodities consumed in small quantity each rather than a single commodity consumed in large amount, consumers and firms will spread their consumption and production over every location in an homogeneous space for different commodities (goods or inputs) in order to avoid the positive transport costs. This situation features an equilibrium under which each location acts as an autarchy and contradicts the reality that a consumer usually lives in one house and a firm chooses a few places where to establish plants. In summary, a competitive model is not compatible with increasing returns arising from market interactions (pecuniary externalities). It points out that only imperfectly competitive markets are relevant here. Under imperfect markets, firms become price-makers instead of price-takers due to a certain level of market power. In this study, I intentionally focus on monopolistic competition rather than oligopolistic competition which is much problematic for the nonexistence of a general equilibrium (Gabszewicz and Thisse, 1992).

Tracing back to Chamberlin (1933), the monopolistic competition model emerges as a market structure determined by firms' internal increasing returns to production and consumers' heterogeneous tastes. On the supply side, firms are price-makers in that each of them produces a horizontally differentiated product (monopolistic), called a variety, due to its internal increasing returns. However, the number of firms is sufficiently large so that one firm's pricing level does not directly affect the decision of any of its competitors but only exerts indirect impacts on the global price index as a reflection of the aggregate behavior of its competitors in the market (competition). On the demand side, through aggregation, heterogeneous consumers differentiated by their tastes are collectively represented by a homogeneous type of consumers (representative agent) who exhibit preferences for varieties. Thus, their utilities increase with not only the quantity of each variety they consume but also with total number of varieties provided. In summary, The Chamberlinian model of monopolistic competition has four basic assumptions: 1) firms sell the varieties of a differentiated product of the same nature but they are not perfect substitutes; 2) every firm produces a single variety under increasing returns and chooses its price; 3) the number of the firms within the same industry is so large that each of them is negligible to the entire industry; 4) it is free entry and exit for the firms, so their profits are zero at the equilibrium.

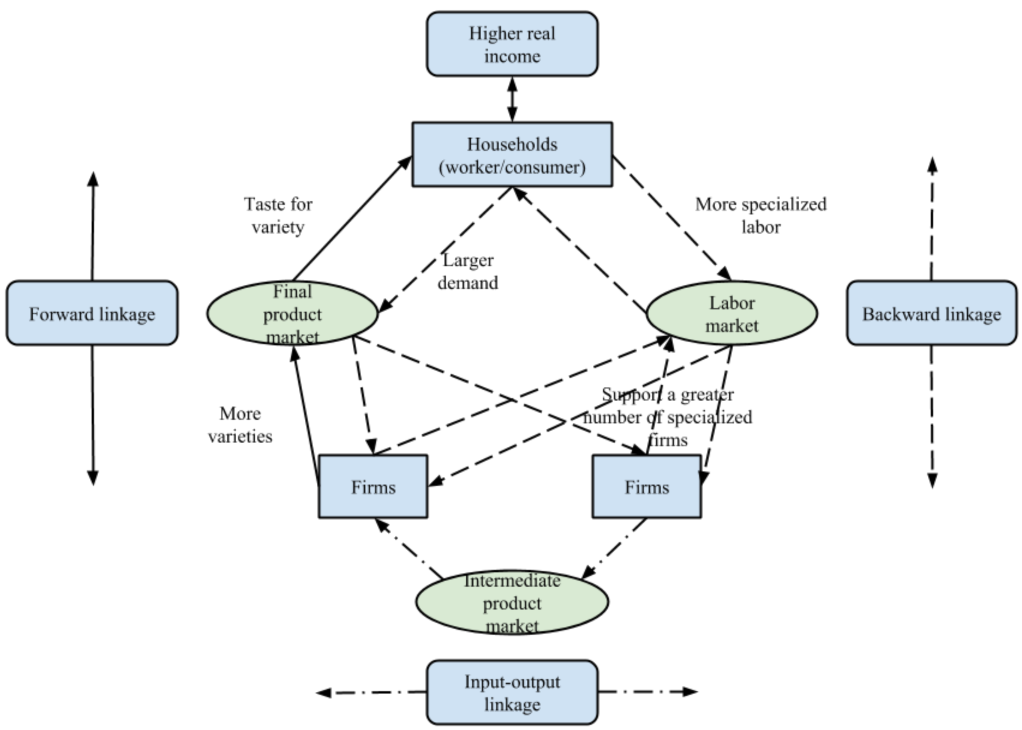


Figure 3.2: Circular causalities: forward and backward linkages

3.2 Circular Causalities: Forward and Backward Linkages

This section disentangles the agglomeration forces, forward and backward linkages, which is also the focus of NEG models. Specifically, two effects need to be distinguished. First, the market-access effect concerns the propensity of firms to concentrate their production close to the larger market rather than the smaller one in order to save on transport costs of goods. Second, the price index effect (or cost-of-living effect) refers to the impact of spatial concentration of firms on the level of price index (a generalized price indicator) in the local product market (or local cost of living). Figure 3.2 depicts the relationship between these two effects. Suppose a homogeneous space over which households and firms are dispersed without any agglomeration of cities. Let us assume that, for some reason, a household relocates from one point to another, which breaks the initial homogeneity and makes the local demand around it larger than anywhere else. Because of the market-access effect, this larger local market attracts more firms from other places. According to the 'home-market effect' of Krugman (1980), the number of firms relocated due to the local attraction of the larger market is more than proportional to the initial shift of demand. The reason is that more firms create more jobs, which employ more workers. This more-than-proportional increase of jobs will encourage further shift of demand via the relocation of households. This creates a reinforcing process called 'demand-linked circular causality' or 'backward linkages'.

On the other hand, once the initial shift of demand induces the shift of production (firm relocation due to market-access effect), the increase of firms brings in a larger variety of products in the local market because of monopolistic competition. Since a larger proportion of varieties are consumed in the bigger local market rather than in the smaller distant ones, the cost of consumption bears less transport burdens in total and results in a lower price index as an indicator of the cost of living (price index effect). Due to a lower cost of living reflected by a relatively higher real wage (given a nominal wage) in the local market, additional households will be further attracted to relocate here and, in turn, will induce even more shift of producing firms. At the same time, the increase of households locally will supply a larger range of labor to support much more specialized production which provides a greater amount of varieties. This creates another reinforcing process called 'cost-linked circular causality' or 'forward linkages'.

The key of the two circular causalities is the dual aspect of households. On the one hand, households are consumers of the products produced by firms. When households relocate, they move with their demand of consumption to the destination local market. On the other hand, households are workers who provide labor as a production factor to firms. When households migrate, they move with their supply of labor to the destination local market and spend their earnings there. It is this dual aspect of households that connects with firms on both the demand and supply sides in an enlarged local market that gives rise to the circular causation in location choices.

It is noteworthy that there are input-output linkages between firms of the same or different sectors, because firms sell and buy intermediate products. If the market structure of this intermediate sector is also characterized by monopolistic competition, firms of this sector produce horizontally differentiated intermediate products. Following the same logic of demand-linked and cost-linked circular causalities between households and final product sector, increasing firms in final product sector enlarges the demand for intermediate products (upstream supplier) via market-access effect; adding new firms in the intermediate product sector producing more varieties lowers the cost of production of final products (downstream customers) through price index effect.

3.3 Market Competition and Spatial Costs

This section discusses the sources of dispersion forces and how various spatial costs at different scales affect the relative strength of agglomeration and dispersion forces. In the product markets (either final or intermediate), what plays a major role is the market-crowding effect that the location choice of firms is driven by the severity of market competition. In other words, firms seek to locate in a market with fewer competitors. Specifically, after the initial relocation of a household, firms follow the demand shifting due to market-access effect. Increasing varieties brought by new firms joining the local market creates a price index effect. As more firms arrive, the given demand in the local market becomes increasingly fragmented over more varieties while firms maintain their prices unchanged. That is, the local market becomes more competitive, which can also be reflected by the lower price index indicator. Due to this fragmentation effect, each firm has a lower share of the total demand and thus a reduction of its sales and operating profit. Because of this, firms would have to pay lower wage to their workers in the local market than other places, which in turn drives households to leave the local market for some places else offering a higher wage. Other than a circular causality, this process features a self-correcting mechanism (or negative feedbacks) where the initial relocation of households tends to be discouraged in the cycle.

The market-crowding effect also exists in the land market. As increasing households and firms congregate to the same area and occupy land, competition in the land market becomes severe and raises the land rent in the local market given the limited supply of land as an immobile factor. Higher land rents inevitably increase the costs of households and firms and thus decrease their profits and utilities, respectively. This market-crowding effect will in turn propel households and firms to leave the local market and seek cheaper places because of the high ‘cost of living’ in the real estate aspect.

Spatial costs are the fundamental determinant of the strength of the dispersion forces. At the intra-urban scale, transport costs of products such as shopping costs (for households) and transaction costs (for firms) are prominent in the final and intermediate product markets. As these costs are reduced by transportation technology or accessibility improvements, the market-crowding effect in the local market diminishes. Because of the reduction in transport costs, the demand shared by the firms is no longer limited to the local market. In other words, a local market extends to a global market (e.g., the market in the urban center extends to one of the entire city). As a result, even if firms agglomerate in the local market, competition is no more localized and thus firms have fewer incentives to move away from the agglomeration since it cannot raise their revenues. In turn, households are also less intent to leave because firms elsewhere cannot pay them more. Following a similar logic, decrease in trade costs at the inter-urban scale has a same impact on the competition between cities and thus on the migration decisions of firms and households.

In contrast, the commodity in the land market, land itself, is immobile such that its transport is prohibitive. Therefore, the market-crowding effect is permanent in the land market. Furthermore, with the increase of accessibility the reduction in commuting costs will strengthen this dispersion force because households no longer need to stay close to their workplaces where firms locate. Similarly, firms no longer need to locate back-to-back because of the reduction in transaction costs (including transportation and communication costs) between them. In all, the alleviation of local competition in the land market becomes a stronger incentive through the improvement in transportation and communication means.

On the other hand, spatial costs also determine the strength of agglomeration forces. As spatial costs decrease, both market-access effects and price index effects become weaker since their essences lie in saving transport costs on either the production or the consumption when they agglomerate. Their saving matters less with the general reduction in spatial costs due to transportation improvements. In summary, the stable status (equilibrium) as either agglomeration or dispersion depends on the balance of the relative strength of these forces. As transportation and communication means keep improving, if the strength of agglomeration forces diminishes more rapidly than that of dispersion forces the dispersion equilibrium dominates; otherwise the agglomeration dominates.

CHAPTER 4: POLYCENTRIC URBAN MODELS AS BUILDING BLOCKS

The aim of this chapter is to construct theoretical polycentric urban models as the building blocks for the generic framework (Chapter 3). At the intra-urban scale, the proposed polycentric urban model extends the partial equilibrium non-monocentric models developed in Fujita (1988, 1990) into a general equilibrium model in three markets (land, products and labor). The extensions lie in two aspects: first, market linkages are incorporated for both final and intermediate goods; second, labor market linkages are modeled through commuting costs and patterns between residence and workplace locations. I approach it in an incremental manner. First, a polycentric model is formulated to endogenize the agglomeration by incorporating the final goods market linkages and spatial costs of transporting goods and commuting without considering market interactions between firms (Section 4.1). Second, I extend the model into a polycentric model with both final and intermediate goods (Section 4.2), that also captures the input-output linkages between firms. I then proceed in section 4.3 to study the equilibrium conditions for the proposed models and examine the existence of multiple equilibria and their feasibility and stability. Numerical results prove that our models are able to generate polycentric spatial structures at the intra-urban scale.

4.1 Polycentric Urban Model with Final Products

Non-monocentric models of urban land use do not assume an a priori location of firms and households. Their development through the pecuniary externality approach and the monopolistic competition formulation are pioneered by Fujita (1988), in which a simple specification with the entropy type model is used for an analytical solution. I use the standard Dixit-Stiglitz (1977) form in NEG that follows the four basic assumptions of Chamberlinian monopolistic competition. In addition, by incorporating employment and commuting (labor market), our model is of a general equilibrium rather than a partial equilibrium as the one by Fujita (1988). The specification of the proposed model is as follows.

Assume that the whole economy occurs along a linear space, where locations are represented by continuous and infinite space $X \equiv \mathbb{R}$. In the economy, there is a continuum of homogenous households with a size N, each of which provides one unit of labor and is free to choose any location. Let $n(y)$ be the household density at $y \in X$. First, I consider only one industry sector in this chapter. The industry consists of a continuum of firms with a size M, each of which produces a differentiated variety with the same production technology. Thus, the total number of varieties is also M. Let $m(x)$ be the firm density at $x \in X$. Second, land is a type of perfectly immobile good that can be consumed by both households and firms. The land quality is homogenous and the land density is equal to 1 at every location. Lastly, an unproduced homogenous good (e.g., agricultural sector) is assumed to be tradable without cost and taken as the numeraire.

4.1.1 Household Consumption

Every household is represented by a type of aggregate consumer with identical consumption preference and follows a quasi-linear utility with the CES (constant elasticity of substitution) preferences:

$$U = \alpha \ln \mathcal{M} + H, \alpha > 0 \quad (4.1.1.1)$$

$$\mathcal{M} = \left[\int_0^M q(i)^{\frac{\sigma-1}{\sigma}} di\right]^{\frac{\sigma}{\sigma-1}}, i \in [0, M]$$

where H is the quantity of the numeraire, and $\mathcal{M}$ the quantity of the composite good from the industry sector. $q(i)$ is the demand for variety i, and σ is the elasticity of substitution between any two varieties ($\sigma > 1$). A smaller σ indicates more differentiated varieties sold by firms such that they can soften competition at a higher degree. This formulation assumes the differentiated varieties affect the utility in a symmetric way. In addition, each household consumes a fixed amount of land S_h assuming $S_h = 1$ without loss of generality. Thus, the household density $n(y) = 1$ if $y \in X$ occupied by households.

Consider a household at location y, and let $q_i(y)$ represent the demand for variety *i* and $H(y)$ be the demand for the numeraire good by the household. Since variety *i* is produced by firm *i*, whose location is at $x(i)$, then it gives:

$$q_i(y) = q_i(y|x(i)) \quad (4.1.1.2)$$

where $q_i(y|x(i))$ presents a household's consumption of variety *i* from firm *i* at location $x(i)$. I assume that the cost to transport a variety *i* from firm location *x* to household location *y* takes Samuelson (1952)'s "iceberg" form: shipping one unit of variety *i* over a distance $|y - x|$ requires $\tau(y|x) = \tau^{|y-x|}$ units of numeraire, where $\tau > 1$. In equilibrium, all varieties produced by firms at the same location *x* must have the same (f.o.b.) price $p(x)$ due to their same production technology, transport costs and symmetric preferences over all varieties. Due to the strict concavity of the CES function, households optimize their choice of consumption bundle such that for every variety *i* produced at *x* it has:

$$q_i(y|x(i)) = q(y|x) \quad (4.1.1.3)$$

where $q(y|x)$ represents the consumption of the variety from each firm locating at *x* by the household at *y*. $H(y)$ Based on (4.1.1.1) – (4.1.1.3), the utility function of each household at location *y* becomes:

$$U(y) = \alpha \ln\left[\int_{x\in X} m(x)q(y|x)^{\frac{\sigma-1}{\sigma}}\, dx\right]^{\frac{\sigma}{\sigma-1}} + H(y) \quad (4.1.1.4)$$

where $m(x)$ is the density of firms at *x*. The budget constraint of a household residing at location *y* and working at x_w is given as:

$$\int_{x\in X} m(x)p(y|x)q(y|x)\, dx + R(y)S_h + t|y - x_w| + H(y) = W(x_w) \quad (4.1.1.5)$$

where $W(x_w)$ is the wage offered by firms at location x_w , $R(y)$ is the land rent prevailing at location *y*, $t > 0$ is the unit commuting cost, and $p(y|x)$ is the delivered price (including the f.o.b. price $p(x)$ and the transport cost) of the each variety supplied at location *x* for households at location *y*. According to iceberg transport technology, it follows:

$$p(y|x) = p(x)\tau(y|x) \quad (4.1.1.6)$$

Following Fujita and Thisse (2002, 6.3), I define a commuting function $J(y) = x_w$ to associate a residential location to a job location; thus it describes the commuting pattern. And it follows that a household residing at y must work at $J(y)$ to maximize its wage net of commuting cost:

$$W[J(y)] - t|y - J(y)| = \max_{x\in X}\{W(x) - t|y - x|\}, \quad y \in X \quad (4.1.1.7)$$

Then by maximizing (4.1.1.4) subject to (4.1.1.5) with respect to the choice of consumption bundle, the demand for a variety from location *x* by a household at *y* is as follows:

$$q(y|x) = \alpha \frac{p(y|x)^{-\sigma}}{\int_{x\in X} m(x)p(y|x)^{1-\sigma}\, dx} \quad (4.1.1.8)$$

Note that introducing new varieties will increase the denominator and thus lead to a decrease in the demand for the existing varieties if their prices remain constant. That is, the entry of a new firm fragments the total demand over more varieties, which is known as the market-crowding effect. Then, the indirect utility function is obtained as:

$$V(y) = \alpha\left[\ln\alpha - 1 - \ln\overline{P(y)}\right] + W(x_w) - t|y - x_w| - R(y) \quad (4.1.1.9)$$

where $\omega(x)$ is the real income of a household at location *x* and

$$\overline{P(y)} = \left[\int_{x\in X} m(x)p(y|x)^{1-\sigma}\,dx\right]^{\frac{1}{1-\sigma}} \quad (4.1.1.10)$$

is called the price index. Note that $\overline{P(y)}$ decreases with the entry of new varieties, which means that the price index becomes lower when the market becomes more competitive. At the same time, household utility rises as $\overline{P(y)}$ decreases, thus reflecting the utility getting higher with lower cost-of-living. This is known as the price index effect. Thus, (4.1.1.8) can be written as

$$q(y|x) = \frac{\alpha}{\overline{P(y)}}\left[\frac{p(y|x)}{\overline{P(y)}}\right]^{-\sigma} \quad (4.1.1.11)$$

Finally, the associated bid rent function of a household at *y* is defined as:

$$\Psi(y, U^*) = \alpha\left[\ln\alpha - 1 - \ln\overline{P(y)}\right] + W(x_w) - t|y - x_w| - U^* \quad (4.1.1.12)$$

which is the maximum rent per unit of land that a household can bid at location *y* while working at x_w and enjoying the equilibrium utility level of U^*.

4.1.2 Firm Production

The production in the industry is under increasing returns at the firm level. For a firm at location *x* to produce one unit of a variety, it needs a fixed requirement of $f > 0$ labor and a marginal requirement of $a > 0$ units of the numeraire. In addition, all firms consume a fixed amount of land S_f. Because of the assumption of monopolistic competition, a firm at location *x* chooses its f.o.b. price $p(x)$ so as to maximize its profit $\Pi(x)$ that is as follows:

$$\Pi(x) = \int_{y \in X} n(y)[p(y|x) - \tau(y|x)a]q(y|x)\,dy - W(x)f - R(x)S_f = \int_{y \in X}[p(x) - a]n(y)\tau(y|x)q(y|x)\,dy - W(x)f - R(x)S_f \quad (4.1.2.1a)$$

It can be verified that the price elasticity of the total demand for any variety is independent of the spatial distribution of the demand, and equals the price elasticity σ of each household's demand. Therefore, the marginal revenue is equal to the marginal cost for a firm at location *x*: $p(x)(1 - \sigma^{-1}) = a$. As a result, the optimal f.o.b. price a firm charges at location *x* is as follows:

$$p(x) = a\frac{\sigma}{\sigma - 1} = p^* \quad (4.1.2.2)$$

which indicates that each firm charges its f.o.b. price at a markup over the marginal cost a. Substituting (4.1.2.2) into (4.1.2.1), the profit of a firm at *x* becomes

$$\Pi(x) = [p(x) - a]Q(x) - W(x)f - R(x)S_f \quad (4.1.2.1b)$$

where $Q(x) = \int_{y \in X} n(y)\tau(y|x)q(y|x)\,dy$. Thus, at the equilibrium of the economy, according to the zero-profit condition (free entry and exit of firms), a firm at location *x* has its equilibrium production as

$$Q^*(x) = \frac{\sigma - 1}{a}\left[W(x)f + R(x)S_f\right] \quad (4.1.2.3)$$

which depends on the land rent and wage it pays at *x*.

Finally, the associated bid rent function of a firm at *x* is defined as

$$\Phi(x, \Pi^*) = \frac{1}{S_f}\left[\frac{a}{\sigma - 1}Q(x) - W(x)f - \Pi^*\right] \quad (4.1.2.4)$$

which represents the maximum rent a firm is willing to pay for a unit piece of land at location x while earning an equilibrium profit equal to Π^*.

4.2 Polycentric Urban Model with Both Final and Intermediate Products

The polycentric model here is built on the basis of the model proposed in section 4.1 by considering the intermediate product market that accounts for the input-output linkages between firms. In that sense, the technology for production needs to be reformulated while the consumption side remains the same as that in the previous model.

Firms' establishment requires a fixed amount of labor $f > 0$ and a fixed amount of land S_f. According to monopolistic competition, the production function takes a quasi-linear form, similar to consumers' utility, using both intermediate products and homogenous good as inputs. Under increasing returns, firms produce following two additional assumptions: 1) the elasticity of substitution is identical for both final and intermediate consumption; 2) each variety enters its own production. Thus, the production function of a firm at location *x* is as follows:

$$aQ(x) = \mu \ln I(x) + H, \ \mu > 0 \quad (4.2.1)$$

where $Q(x)$ is the output of production, H is a constant quantity of the numeraire, and *a* is the marginal requirement as before. $I(x)$ is a composite intermediate good of a CES-type:

$$I(x) = \left[\int_{x_I \in X} m(x_I) q_I(x|x_I)^{\frac{\sigma-1}{\sigma}} dx_I\right]^{\frac{\sigma}{\sigma-1}} \quad (4.2.2)$$

where $q_I(x|x_I)$ is the demand for the intermediate variety from each firm locating at x_I by a firm at *x*. Then, the cost function of this firm is given by:

$$C(x) = W(x)f + R(x)S_f + \int_{x_I \in X} m(x_I) p(x|x_I) q_I(x|x_I)\, dx_I + H \quad (4.2.3)$$

where $p(x|x_I)$ is the effective price (including the f.o.b. price $p(x)$ and transaction cost such as transport cost, communication cost and other fee incurred). Again, assuming iceberg transport technology, it follows:

$$p(x|x_I) = p(x_I)\,\tau(x|x_I) \quad (4.2.4)$$

Thus, by maximizing production $Q(x)$ in (4.2.1) subject to the cost constraint (4.2.3), the optimal demands for labor and intermediate products can be obtained respectively:

$$q_I(x|x_I) = \frac{\mu}{\overline{P(x)}}\left[\frac{p(x|x_I)}{\overline{P(x)}}\right]^{-\sigma} \quad (4.2.5)$$

where the price index at *x* where the firm locates is the same as (4.1.1.11) but using different notations: $\overline{P(x)} = \left[\int_{x_I \in X} m(x_I) p(x|x_I)^{1-\sigma}\, dx_I\right]^{\frac{1}{1-\sigma}}$. Inserting (4.2.5) into (4.2.1), the cost function can be obtained as a function of $Q(x)$:

$$C(x) = W(x)f + R(x)S_f + \mu - \mu\left[\ln\mu - \ln\overline{P(x)}\right] + aQ(x) \quad (4.2.6)$$

where a is the marginal production cost of a firm at location *x*. In this model, $C(x)$ depends not only on the wage bill $W(x)$ as in the model (4.1.2.1) but also on the price index $\overline{P(x)}$. Note that the decline in price index leads to a decrease in production costs, which allows the emergence of input-output linkages effect as a new centripetal force. Consequently, firms have an incentive to locate where hosting the largest varieties of intermediate inputs, as they can benefit from lower production costs.

Based on the same logic to derive (4.1.2.2), the optimal f.o.b. price charged by a firm at location *x* is the markup $\frac{\sigma}{\sigma-1}$ multiplied by the marginal production cost:

$$p(x) = a\frac{\sigma}{\sigma-1} = p^* \quad (4.2.7)$$

Then, the profit earned by a firm at *x* is given as

$$\Pi(x) = p^* Q(x) - C(x)$$

Therefore, applying zero-profit condition, its equilibrium production is obtained as

$$Q^*(x) = \frac{\sigma-1}{a}\left[W(x)f + R(x)S_f + \mu\left(1 - \ln\mu + \ln\overline{P(x)}\right)\right] \quad (4.2.8)$$

Finally, the associated bid rent function of a firm at *x* is as follows:

$$\Phi(x, \Pi^*) = \frac{1}{S_f}\left[\frac{a}{\sigma-1}Q(x) - W(x)f - \mu\left(1 - \ln\mu + \ln\overline{P(x)}\right) - \Pi^*\right] \quad (4.2.9)$$

which represents the maximum rent a firm is willing to pay for a unit piece of land at location x while earning an equilibrium profit equal to Π^*.

4.3 Short-run Equilibrium for Polycentric Urban Models

This section investigates the short-run equilibrium (or city equilibrium) conditions for the proposed polycentric urban models. The purpose is to study under what conditions polycentric and monocentric equilibria exist, respectively. I first describe in Section 4.3.1 the general conditions for each market equilibrium (land, labor, and products) and basic assumptions. Then, I examine the specific conditions that various spatial cost variables (commuting and goods transport) need to fulfill to allow the generation of monocentric or polycentric structures by the models. Given the existence of multiple equilibria, I further test the stability of each type of equilibria. I proceed with the last two tasks by treating the two models respectively in Sections 4.3.2 and 4.3.3.

4.3.1 Equilibrium Conditions

A short-run equilibrium is defined as an equilibrium configuration at the intra-urban scale such that all households achieve the same level of utility U^* by choosing their home locations and workplaces and the amount of final products they consume; all firms earn the same level of profits Π^* by choosing their locations, prices and wages to compete on in the land, product and labor market until no firm can profitably enter the market (zero-profit condition); all markets are cleared for all goods. At this urban equilibrium determined by the interplay of the households' and firms' bid rent for land, no household wants to change its residence and/or workplace, and no firm has an incentive to change its location. This equilibrium configuration is conditioned on a fixed population of households N within the urban context. Thus, the unknowns are: 1) the household distribution $n(y)$ and the firm distribution $m(x)$; 2) the wage rate $W(x)$; 3) the land rent $R(x)$; 4) the commuting function $J(y)$; 5) the equilibrium utility level U^* and profit level Π^*. In this section, I specify the equilibrium conditions discussed above.

1. Households' and firms' *population constraints*:

$$N = \int_{y \in X} n(y) dy \quad (4.3.1.1a)$$

$$M = \int_{x \in X} m(x) dx \quad (4.3.1.1b)$$

2. *Commuting equilibrium* is given by

$$W[J(y)] - t|y - J(y)| = \max_{x \in X} \{W(x) - t|y - x|\}, \quad y \in X \quad (4.3.1.2)$$

which determines the commuting destination $J(y)$ that maximizes the net wage for each potential residential location y. I assume that no *cross-commuting* occurs at the equilibrium condition (Fujita and Thisse, 2002, p. 189).

3. I assume that every household in the city has a job somewhere. Thus, given the fixed population of households, each of which provides one unit of labor, I have the *full employment condition*:

$$N = Mf \quad (4.3.1.3a)$$

and the *labor market equilibrium* is defined for each set of locations HL where households reside and all commute to the same firm location $J(\text{HL})$ it holds:

$$\int_{\text{HL}} n(y) dy = m(J(\text{HL})) f \quad (4.3.1.3b)$$

which ensures the equality of labor demand and supply under the commuting function J.

4. Assuming that the opportunity cost of land is $R_A = 0$, then *land market equilibrium* is given by

$$R(x) = max\{\Psi(x, U^*), \Phi(x, \Pi^* = 0), R_A\} \quad (4.3.1.4)$$

$$\Psi(x, U^*) = R(x) \quad if\ n(x) > 0$$

$$\Phi(x, \Pi^* = 0) = R(x) \quad if\ m(x) > 0$$

$$S_h n(x) + S_f m(x) = 1 \quad if\ R(x) > R_A$$

which ensure that each location is occupied by a household or a firm with the highest bid rent.

5. *Product market equilibrium* is given by equating the production $Q(x)$ of a variety by a firm at location *x* and the demand $D(x)$ for it from the entire economy

$$Q(x) = D(x) = \int_{y \in X} n(y)\tau(y|x)q(y|x)dy \quad (4.3.1.5a)$$

if only final products are considered (model in section 4.1), and

$$Q(x) = D(x) = \int_{y \in X} n(y)\tau(y|x)q(y|x)dy + \int_{y \in X} m(y)\tau(y|x)q_I(y|x)dy \quad (4.3.1.5b)$$

if both final and intermediate products are considered (model in section 4.2).

4.3.2 Polycentric City with Final Products

To study the equilibrium of the model developed in section 4.1, I first make several simplifications to ease the burden of our analysis. I assume that the fixed amount of land a firm consumes equals to zero ($S_f = 0$). The existence of a city requires a CBD located at the origin $0 \in X$ and may have SBDs located in the suburbs. To explain why the CBD exists in the first place, it would require the construction of non-monocentric models as in Fujita (1988) that would make the analysis much more involved without adding new insights to our results. Firms are free to locate in the CBD or to form SBDs in the suburbs. Households locate residential areas outside of CBD and SBDs and each consumes one unit of land. Because firms consume no land, both CBD and SBDs are dimensionless. I made this simplification because having firms to consume land and thus CBD and SBDs having spatial extension would make the analysis more cumbersome without changing the nature of our results. In addition, for analytical convenience, I focus on the basic symmetric setting where the CBD is surrounded by two SBDs. This symmetric spatial setting can be extended to have more SBDs (4, 6, 8, and so on), which would make the analysis much more involved while providing no further insight than the basic 2-SBDs setting in terms of allowing us to study the conditions for the existence of monocentric and polycentric equilibria. As a result, I determine the size of the CBD and the location and size of the two SBDs, which renders the internal structure of the city endogenous.

Let $\Psi^c(y)$ and $\Psi^s(y)$ be the bid rent at $y \in X$ of a household working in the CBD and the SBD. The land market equilibrium (4.3.1.4) becomes

$$R(y) = max\{\Psi^c(y), \Psi^s(y), 0\} \quad (4.3.2.1).$$

Let the right endpoint of the area formed by residents working in the CBD be l_c. Let l_s be the right endpoint of the residential area surrounding the SBD located at x_s, and l_s' the symmetrical residential endpoint on the left-hand side of the SBD. l_s is also the outer limit of the city. Because there are costs to ship the differentiated good, the two residential areas are adjacent ($l_c = l_s'$), and the total size of residential areas measure the size of the city l. Because each household occupies one unit of land, the total size of households $N = l$. Due to the symmetric setting, the left side of the CBD can be derived similarly. Therefore, the critical points are as follows:

$$l_c = \frac{\lambda}{2}N,\ x_s = \frac{1+\lambda}{4}N,\ l_s = \frac{1}{2}N \quad (4.3.2.2)$$

where λ is the share of households who commute to the CBD. An illustration of the land rent profile is provided for the right side of the space setting in Figure 4.1. Because of the no cross-commuting assumption (4.3.1.2), households in the residential areas surrounding the CBD and SBDs commute to their corresponding workplaces. And due to condition (4.3.1.3), the share of firms θ located in the CBD is equal to λ

$$\theta = \lambda \quad (4.3.2.3).$$

At equilibrium, the bid rent at l_s is 0 since the opportunity cost of land is 0; because of $l_c = l_s'$ the bid rents at l_c and l_s' are equal.

$$\Psi^s(l_s, U^*) = 0;$$

$$\Psi^c(l_c, U^*) = \Psi^s(l_s', U^*)$$

The latter implies that

$$\alpha\left[\ln\alpha - 1 - \ln\overline{P(l_c)}\right] + W_c - tl_c - U^*$$

$$= \alpha\left[\ln\alpha - 1 - \ln\overline{P(l_s')}\right] + W_s - t(x_s - l_s') - U^*$$

based on (4.1.1.12). Because of $l_c = l_s'$, (4.3.2.2) and (4.3.2.3), it can be reduced to

$$W_c - W_s = t\frac{3\lambda-1}{4}N = t\frac{3\theta-1}{4}N \quad (4.3.2.4)$$

where W_c and W_s are the wages in CBD and SBDs respectively. Thus, the difference of the wages prevailing in the CBD and SBDs compensates exactly the households for the difference of their corresponding commuting costs. When the size of the CBD is greater than the size of each SBD $\theta > 1/3$ (due to the assumption of two symmetric SBDs), the wage difference $W_c - W_s$ is positive. In addition, as the number of households grows, the wage difference increases. Because the average commuting cost increases with the size of the city, firms in the CBD must pay a higher wage to compensate workers' increasing commuting costs.

In equilibrium, since households at all locations achieve the same utility level U^*, I can apply (4.1.1.9) to obtain $V(0) = V(l_c) = V(x_s) = V(l_s) = U^*$. Combining with (4.1.1.5), it is readily verified that

$$H(y) = H_0 + \alpha \ln \frac{\overline{P(y)}}{\overline{P_c}}, \ 0 \leq y \leq l_s \quad (4.3.2.5)$$

where $H_0 = H(0)$ is the consumption of the numeraire good by a household at location 0 in equilibrium and $\overline{P_c} = \overline{P(0)}$ is the price index at the CBD. The bid rent functions then can be derived accordingly:

$$\begin{cases} \Psi^c(y) = t(l_c - y) + H(l_s) - H(y), & 0 \leq y \leq l_c \\ \Psi^s(y) = t(l_s - 2x_s + y) + H(l_s) - H(y), & l_c \leq y \leq x_s \\ \Psi^s(y) = t(l_s - y) + H(l_s) - H(y), & x_s \leq y \leq l_s \end{cases} \quad (4.3.2.6)$$

Inserting (4.3.2.6) back into (4.1.1.9) and combining with (4.3.2.4), the indirect utility function of households becomes

$$V(0 \leq y \leq l_c) = \alpha\left[\ln \alpha - 1 - \ln \overline{P(l_s)}\right] + W_c - tl_c = V(l_c \leq y \leq l_s) =$$

$$\alpha\left[\ln \alpha - 1 - \ln \overline{P(l_s)}\right] + W_s - t(l_s - x_s) \quad (4.3.2.7)$$

where $V(0 \leq y \leq l_c)$ represents the utility of a household who works in the CBD while $V(l_c \leq y \leq l_s)$ represents the utility of a household who works in a SBD. Note that their first parts in the two utilities are the same and that they are dependent on the constant α parameter and the price index at the city edge. As the price index at the city edge increases, equilibrium utility decreases. And the price index increases with the internal trade cost rate τ. Because price index can be interpreted as the "cost of living" measured by the inverse of consumer's accessibility to the whole range of varieties of products produced in both CBD and SBDs, the price index at the city edge represents the upper bound of the cost of living in the city due to the lowest accessibility to varieties. So, it is reasonable that customer utility degrades with higher cost of living. Then, the second part of the two utilities indicate that, at the equilibrium utility level, households who work in CBD and SBDs respectively trade off their wages earned from different work places with their commuting costs according to (4.3.2.4). The equilibrium utility increases with the wages earned and decreases with higher commuting cost rate t and larger city size N that determines l_c , x_s and l_s .

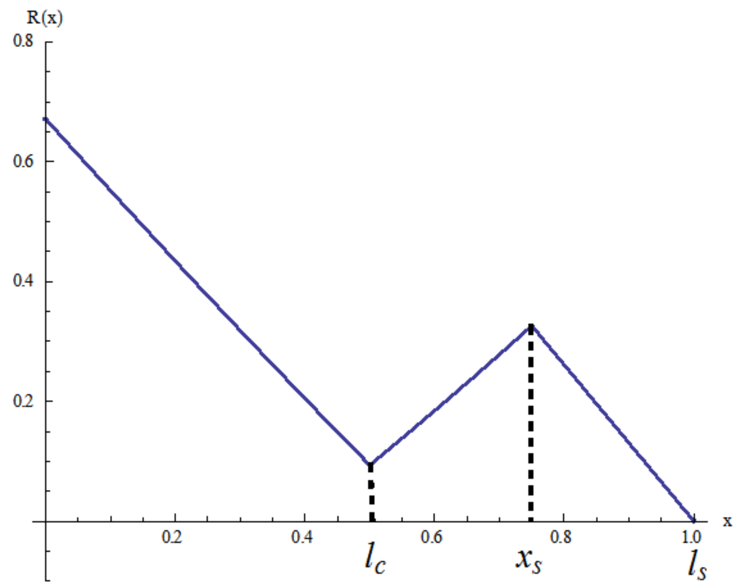


Figure 4.1: Land rent profile for the right side of an intra-urban space ($N = 2, t = 1, \phi = 0.4, \theta = 0.5$)

Through the equilibrium wages, I can study the equilibria of city structure represented by θ and their feasibility and stability in relation to the transport cost of products and the commuting cost. Given the zero-profit condition (4.1.2.3), the equilibrium wage can be expressed as $W^*(x) = Q^*(x)\frac{a}{f(\sigma-1)}$, with condition (4.3.1.5a) I have

$$W_c^* = \frac{2\alpha}{n\sigma}\int_0^{\frac{N}{2}} \frac{\phi(y|c)}{\phi(y|c)\theta+\phi(y|s)\frac{1-\theta}{2}+\phi(y|s')\frac{1-\theta}{2}} dy \quad (4.3.2.8a)$$

$$W_s^* = \frac{\alpha}{n\sigma}\int_{-\frac{N}{2}}^{\frac{N}{2}} \frac{\phi(y|s)}{\phi(y|c)\theta+\phi(y|s)\frac{1-\theta}{2}+\phi(y|s')\frac{1-\theta}{2}} dy \quad (4.3.2.8b)$$

where $\phi = \tau^{1-\sigma} \in (0,1)$ called the freeness of trade, $\phi(y|c) = \tau(y|c)^{1-\sigma} = \phi^{|y|}$, $\phi(y|s) = \phi^{\left|y-\frac{1+\theta}{4}N\right|}$, and $\phi(y|s') = \phi^{\left|y+\frac{1+\theta}{4}N\right|}$.

First, I look at what conditions need to be satisfied to make the monocentric structure ($\theta = 1$) feasible. If $\theta = 1$, $W_c = \frac{\alpha}{\sigma}$ and there are no firms in SBDs, so the wages in SBDs are not available. However, I can calculate the potential level of wages if there were firms in SBDs based on (4.3.2.4). Noting that the equilibrium profit in CBD is equal to 0, it will be attractive for firms to move from the CBD to SBDs if the potential profit in SBDs is greater than 0. In other words, monocentricity is not sustainable if (4.3.2.9) holds:

$$\Pi_s = [f(\phi) - g(t)]f > 0 \quad (4.3.2.9)$$

where $f(\phi) = \frac{\alpha}{N\sigma}\phi^{\frac{N}{2}}\left(\frac{N}{2} + \frac{1-\phi^{-N}}{2\ln\phi}\right)$ and $g(t) = \frac{\alpha}{\sigma} - \frac{N}{2}t$. Figure 4.2 shows the profit surplus $f(\phi)$ when $\alpha = 1, \sigma = 4, N = 1,2,3,4,5$. Because $\min_{0<\phi<1} f(\phi = e^{-\frac{2}{N}}) = \frac{\alpha}{\sigma}\left(\frac{e}{4} + \frac{1}{4e}\right)$, inequality (4.3.2.9) holds if $g(t) < \frac{\alpha}{\sigma}\left(\frac{e}{4} + \frac{1}{4e}\right)$, which can be reduced to

$$t > \frac{\alpha}{N\sigma}\left(2 - \frac{e}{2} - \frac{1}{2e}\right) \quad (4.3.2.10)$$

It means that when commuting cost is larger than a certain threshold, monocentricity is not a feasible equilibrium regardless of the value of transport cost for trading. In addition, this threshold becomes smaller as the city grows, which implies that a larger city requires a lower commuting cost to maintain its monocentric structure.

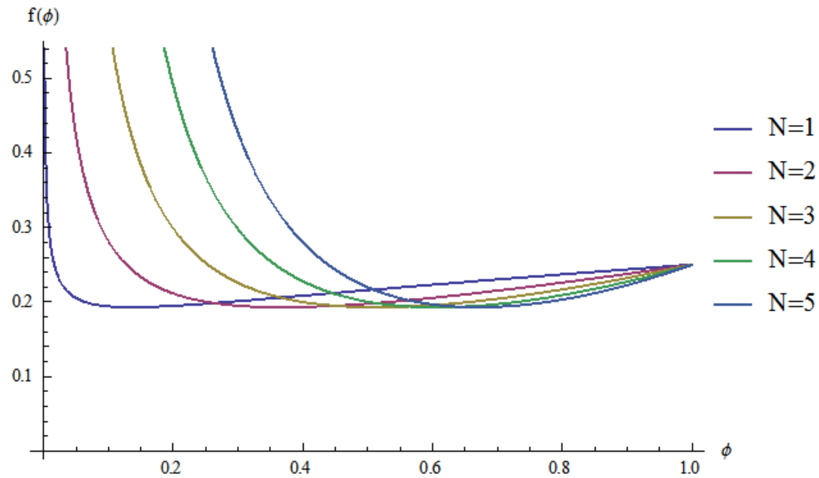


Figure 4.2: Function $f(\phi)$ when $\alpha = 1, \sigma = 4, N = 1,2,3,4,5$

However, $t \leq \frac{\alpha}{N\sigma}\left(2 - \frac{e}{2} - \frac{1}{2e}\right)$ is only a necessary but not sufficient condition to have a feasible monocentric equilibrium. Let $f(\phi) - g(t) = 0$; I can obtain t as a function of ϕ: $t = CT(\phi)$. Because $\max_{0<\phi<1} CT(\phi) = \frac{\alpha}{N\sigma}\left(2 - \frac{e}{2} - \frac{1}{2e}\right)$, $0 < t \leq \frac{\alpha}{N\sigma}\left(2 - \frac{e}{2} - \frac{1}{2e}\right)$. Correspondingly, I can obtain $LB < \phi < 1$ to meet this condition for t, where $\phi = LB$ is the smaller root of $CT(\phi) = 0$. Because of the nonlinearity of $CT(\phi)$, I cannot obtain an analytical solution. Figure 4.3 is a demonstration of the value for LB when $N = 1,2,3,4,5, \alpha = 1, \sigma = 4$. In summary, the city is monocentric if and only if

$t \leq CT(\phi)$ where $0 < t \leq \frac{\alpha}{N\sigma}\left(2 - \frac{e}{2} - \frac{1}{2e}\right)$ and $LB < \phi < 1$ (4.3.2.11)

which is the area under curve $CT(\phi)$ but above 0. When $N = 2, \alpha = 1, \sigma = 4$, $LB \approx 0.128$. Note that as the city grows, curves in Figure 4.3 are shifted to the left and lower commuting and trade costs are required to sustain the monocentric equilibrium.

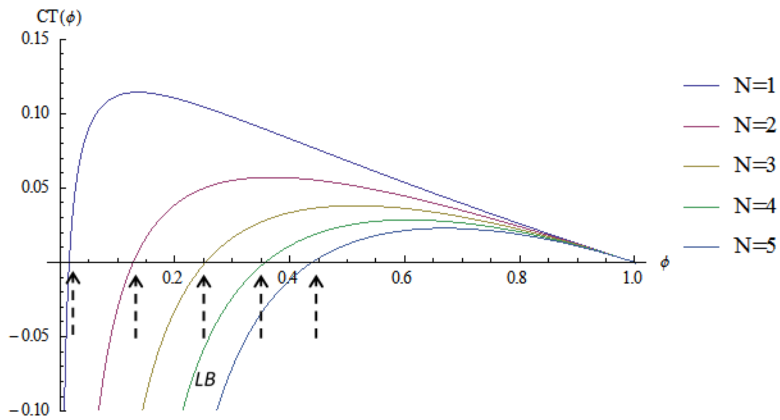


Figure 4.3: Function $CT(\phi)$ with $N = 1,2,3,4,5, \alpha = 1, \sigma = 4$ and corresponding LBs

To examine the stability of the equilibrium, I define a profit difference function between firms at CBD and SBDs

$$\Delta\Pi(\theta, \phi, t) = \Pi_c - \Pi_s = \left(W_c^* - W_s^* - t\frac{3\theta-1}{4}N\right)f \quad (4.3.2.12)$$

Then, I can examine $\frac{\partial\Delta\Pi(\theta,\phi,t)}{\partial\theta}$ under the condition (4.3.2.11) to determine the stability of monocentric equilibrium. Because the analytic solution is unable to obtain, I plot $\Delta\Pi(\theta, \phi, t)$ under the condition (4.3.2.11) when $N = 2, \alpha = 1, \sigma = 4$ as a demonstration. Figure 4.4 shows $\Delta\Pi\big(\theta, \phi, t = CT(\phi)\big)$ when $\phi = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$. Except for $\phi = 0.2$, all curves decrease all the way until $\theta = 1$ where $\Delta\Pi$ becomes zero as θ increases (Figure 4.5). I can expect that the sign of $\frac{\partial\Delta\Pi(\theta)}{\partial\theta}$ changes at some point between $0.2 < \phi < 0.3$. I can numerically solve this point as $\phi \approx 0.21$. Because the condition $t = CT(\phi)$, $\theta = 1$ is a feasible equilibrium. And because $\left.\frac{\partial\Delta\Pi(\theta)}{\partial\theta}\right|_{\theta=1} < 0$, $\theta = 1$ is a stable equilibrium when $\phi = 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$, while $\theta = 1$ is not a stable equilibrium for $\phi = 0.2$ due to $\left.\frac{\partial\Delta\Pi(\theta)}{\partial\theta}\right|_{\theta=1} > 0$. From the trend of these curves, I expect that $\left.\frac{\partial\Delta\Pi(\theta)}{\partial\theta}\right|_{\theta=1} < 0$ holds for $0.21 < \phi < 1$ where $\theta = 1$ is thus a stable equilibrium while $\frac{\partial \Delta\Pi(\theta)}{\partial \theta}\Big|_{\theta=1} > 0$ holds for $LB < \phi < 0.21$ ($LB \approx 0.12$) where $\theta = 1$ is not a stable equilibrium. Figure 4.6 shows $\Delta\Pi(\theta, \phi, t < CT(\phi))$ when $\phi = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$. Because all curves are greater than zero when $\theta \in [0, 1]$, $\theta = 1$ is the only feasible and stable equilibrium when $\phi = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$. Based on the trend of these curves, I expect the stability holds for $LB < \phi < 1$ where $LB \approx 0.128$ under the cond$CT(\phi)$.

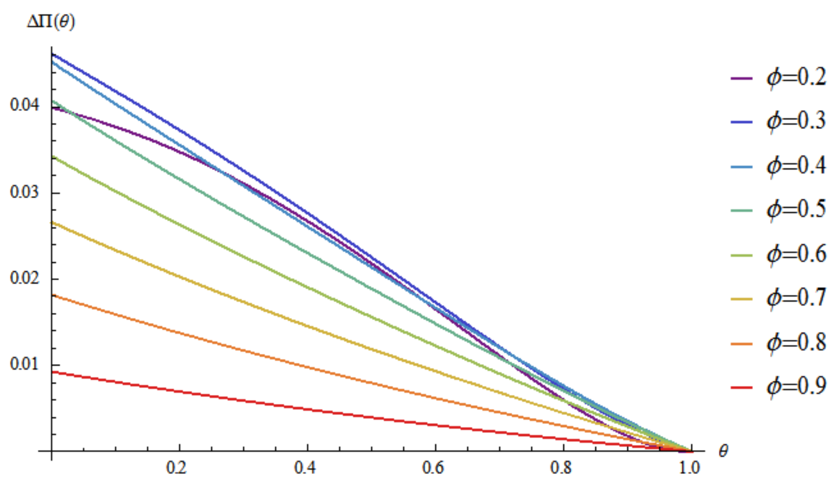


Figure 4.4: Stability of $\Delta\Pi(\theta, \phi, t = CT(\phi))$ for different ϕ when $N = 2, \alpha = 1, \sigma = 4$

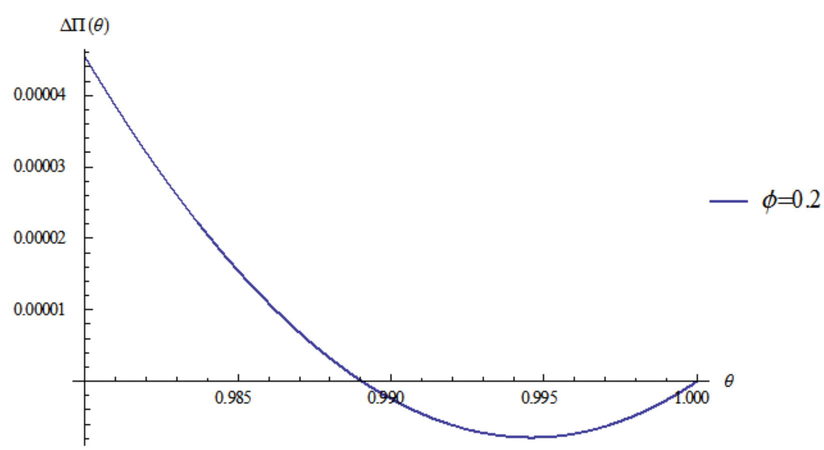


Figure 4.5: Zoom-in of $\Delta\Pi(\theta, \phi, t = CT(\phi))$ when $\phi = 0.2$ to check its stability

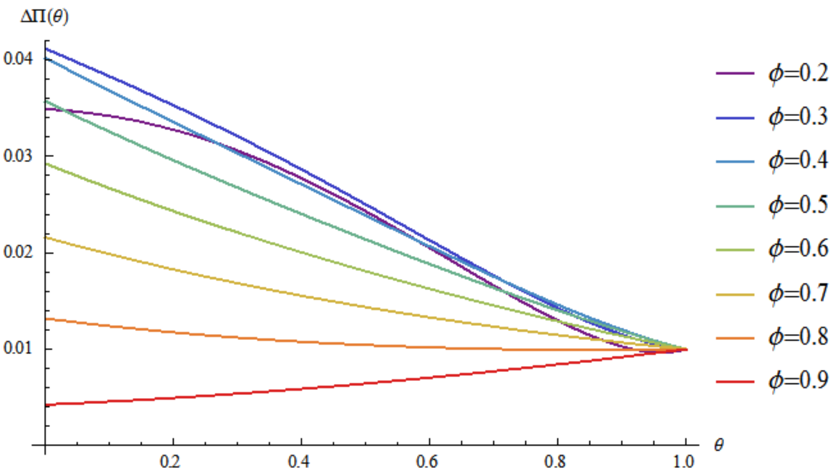


Figure 4.6: Stability of $\Delta\Pi\big(\theta, \phi, t < CT(\phi)\big)$ for different ϕ when $N = 2, \alpha = 1, \sigma = 4$

The second critical point is $\theta = 1/3$ when the CBD is the same size as two SBDs. Let $WF(\phi) = W_c^* - W_s^*$ given equations (4.3.2.8). Using equation (4.3.2.4) I know $W_c^* - W_s^* = 0$ when $= 1/3$. Thus, I can look at what conditions make $WF(\phi) = 0$ to examine whether $\theta = 1/3$ is a feasible equilibrium or not. $WF(\phi)$ does not have analytical solution, so I plot it when $\alpha = 1, \sigma = 4$ and $N = 1, 2, 3, 4, 5$ respectively (Figure 4.7). It illustrates that $WF(\phi) > 0$ always holds when $0 < \phi < 1$ regardless of the value of N (the proof is not provided here), which means that $\theta = 1/3$ is not a feasible equilibrium.

When $\theta < 1/3$, $W_c - W_s < 0$ based on equation (4.3.2.4). Similarly, I can examine what conditions make $WF(\phi) < 0$ by plotting $WF(\phi)$ when $\alpha = 1, \sigma = 4, N = 2$ and $\theta = 0.1, 0.2, 0.3$. Figure 4.8 demonstrates that when $\theta < 1/3$ $WF(\phi) > 0$ always holds in the range of $0 < \phi < 1$ (proof is not provided here). It means that $\theta < 1/3$ is not a feasible equilibrium in any condition.

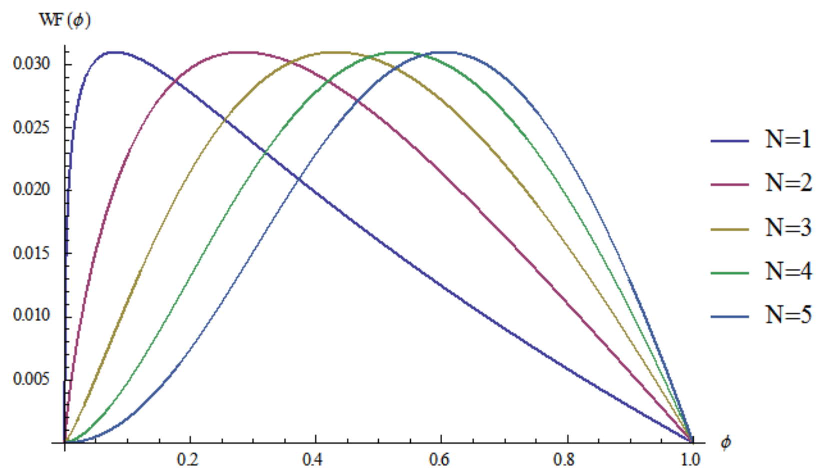


Figure 4.7: Function $WF(\phi)$ for $\theta = \frac{1}{3}, \alpha = 1, \sigma = 4, N = 1, 2, 3, 4, 5$

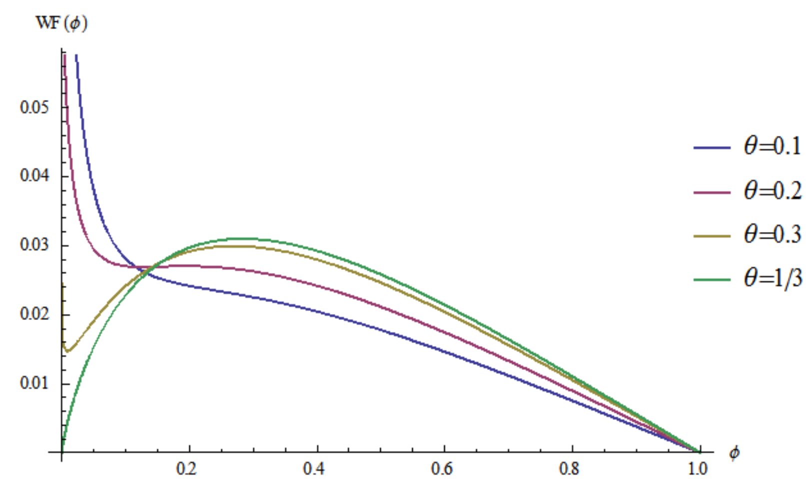


Figure 4.8: Function $WF(\phi)$ when $\theta < \frac{1}{3}, \alpha = 1, \sigma = 4, N = 2$

Finally, I look at the condition to fulfill $\frac{1}{3} < \theta < 1$. Insert $WF(\phi)$ into equation (4.3.2.4), then I can obtain t as a function of ϕ and θ: $t = GCT(\phi, \theta)$, which is a general form of $CT(\phi)$. That is, $CT(\phi) = GCT(\phi, \theta = 1)$. To demonstrate it, I plot $GCT(\phi, \theta)$ when $N = 2, \alpha = 1, \sigma = 4$ and $\theta = 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$ (Figure 4.9). It shows that when $\frac{1}{3} < \theta < 1$ $GCT(\phi, \theta)$ curves exhibit similar forms as that for $\theta = 1$. Thus, general forms of conditions for t, θ can be derived as I did for (4.3.2.11). As a result, for each

$\theta \in (\frac{1}{3}, 1)$ there exists a relationship between t and θ: $t = GCT(\phi, \theta)$ that makes θ a feasible equilibrium as a polycentric city if and only if

$$0 < t \leq \max_{0<\phi<1} GCT(\phi, \theta) \text{ and } LB(\theta) < \phi < 1 \quad (4.3.2.13)$$

where $\phi = LB(\theta) > 0$ is the smaller root of $GCT(\phi) = 0$. When $t > GCT(\phi, \theta)$, it can be verified that $\Pi_s > \Pi_c$ following the same logic used to derive (4.3.2.9). Therefore, θ becomes infeasible and a decrease of θ is followed as firms move from the CBD to SBDs until $\Pi_s = \Pi_c$ and $\theta' < \theta$ becomes a new feasible equilibrium. In contrast, if $t < GCT(\phi, \theta)$, then θ becomes infeasible because $\Pi_c > \Pi_s$. As a result, an increase of θ will be the trend as firms move from SBDs to the CBD until $\Pi_s = \Pi_c$ and $\theta' > \theta$ becomes a new feasible equilibrium. Given a fixed t, as trade cost deceases (ϕ increases) the city first becomes more concentrated in the CBD (θ increases) when trade cost is reduced to the intermediate level, then the city becomes more polycentric (θ decreases) with growing SBDs and a shrinking CBD when the trade cost reaches a sufficiently low level (Figure 4.10). This confirms the bell-shaped curve, spreading-agglomeration-spreading phenomena, found in the literature due to urban costs (Cavailhès et al., 2007).

I examine the stability of the equilibria when $\frac{1}{3} < \theta < 1$ under condition (4.3.2.13) using the same approach as that for $\theta = 1$. I plot $\Delta\Pi(\theta, \phi, t)$ under the condition (4.3.2.13) with $\phi = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ and $N = 2, \alpha = 1, \sigma = 4$ to illustrate $\frac{\partial \Delta\Pi(\theta, \phi, t)}{\partial \theta}$ when $t = GCT(\phi, \theta)$ and $\theta = 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ respectively (Figure 4.11-4.16). They all show $\frac{\partial \Delta\Pi(\theta)}{\partial \theta} < 0$, and thus enable their respective θ to become stable equilibria. Based on the trend reflected by these curves, I can expect the stability to hold for $\theta \in (\frac{1}{3}, 1)$ with their respective condition $LB(\theta) < \phi < 1$.

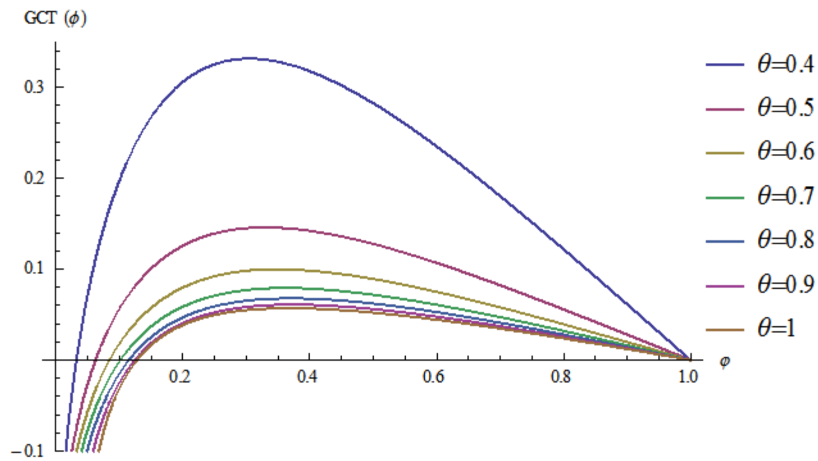


Figure 4.9: $GCT(\phi, \theta)$ when $N = 2, \alpha = 1, \sigma = 4$ and $\theta = 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$

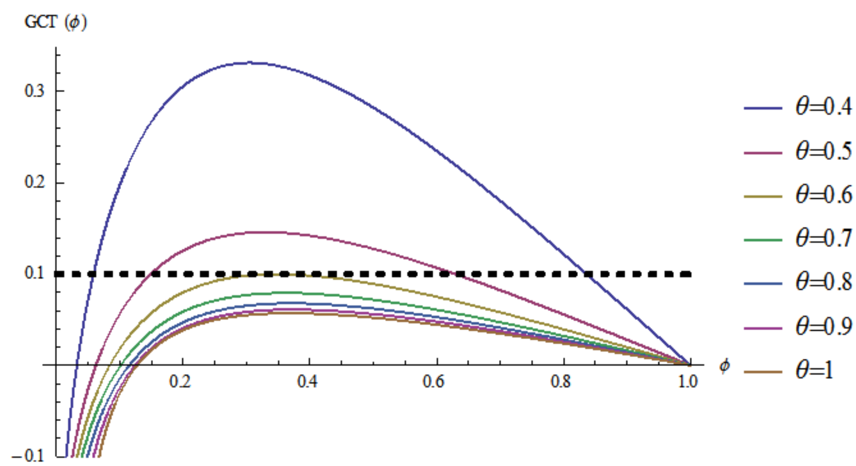


Figure 4.10: Change of feasible θ as ϕ increases given fixed t

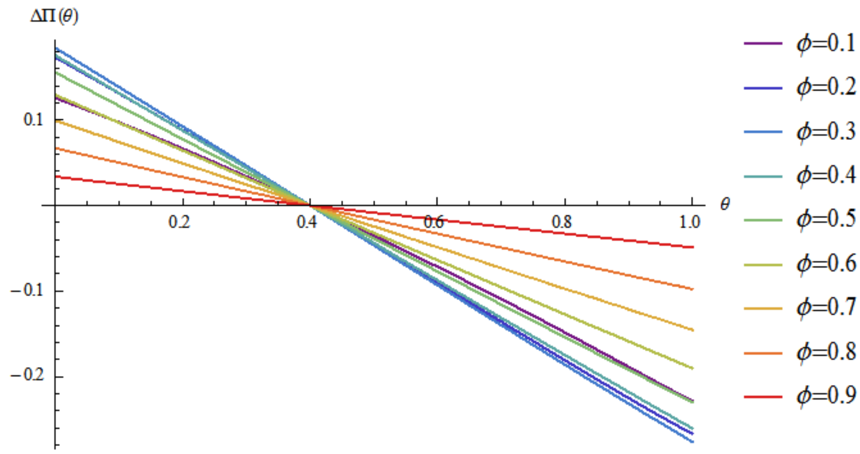


Figure 4.11: Stability of $\Delta\Pi\big(\theta, \phi, t = GCT(\phi, \theta = 0.4)\big)$ for different ϕ

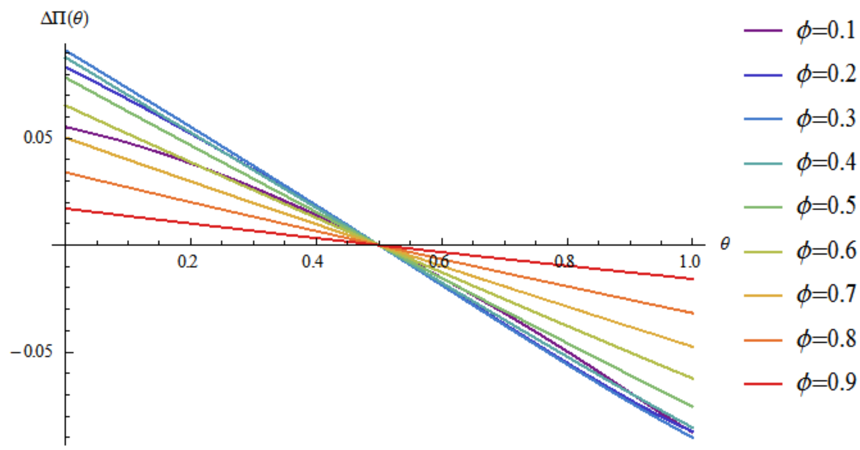


Figure 4.12: Stability of $\Delta\Pi\big(\theta, \phi, t = GCT(\phi, \theta = 0.5)\big)$ for different ϕ

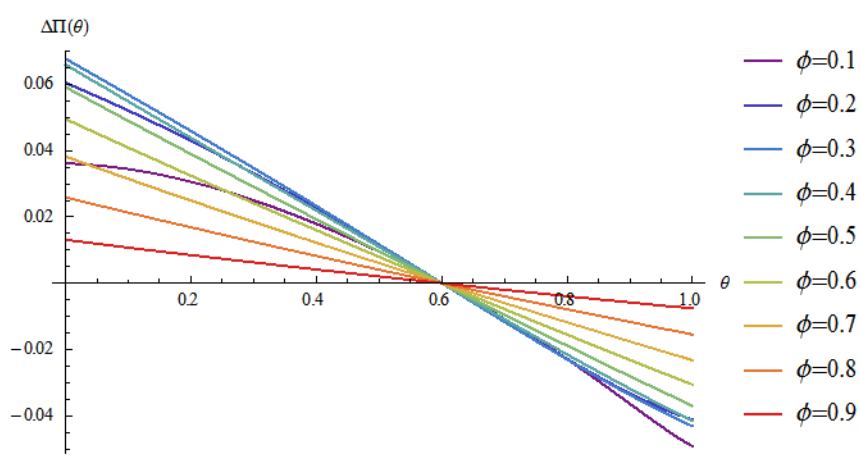


Figure 4.13: Stability of $\Delta\Pi\big(\theta, \phi, t = GCT(\phi, \theta = 0.6)\big)$ for different ϕ

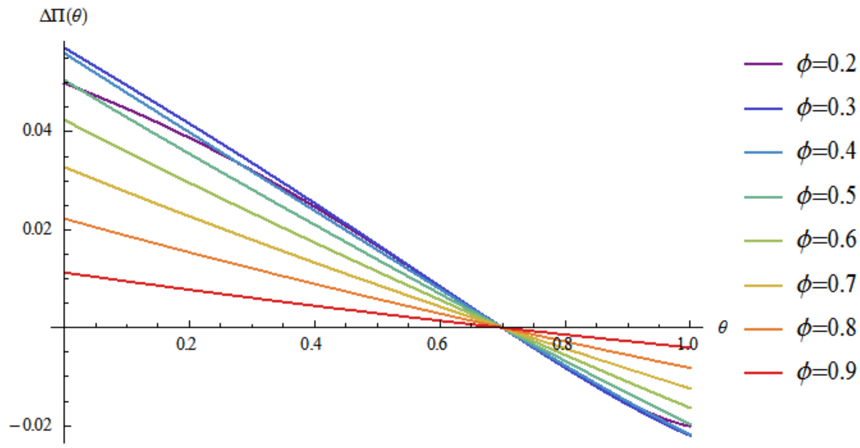


Figure 4.14: Stability of $\Delta\Pi\big(\theta, \phi, t = GCT(\phi, \theta = 0.7)\big)$ for different ϕ

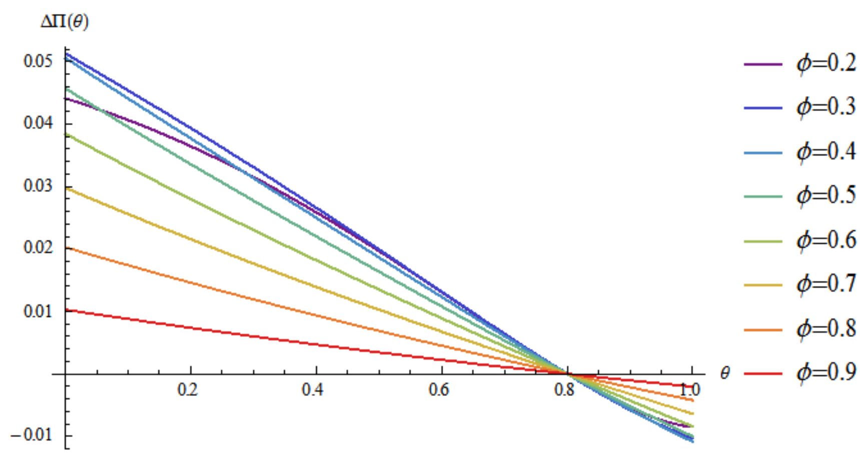


Figure 4.15: Stability of $\Delta\Pi\big(\theta, \phi, t = GCT(\phi, \theta = 0.8)\big)$ for different ϕ

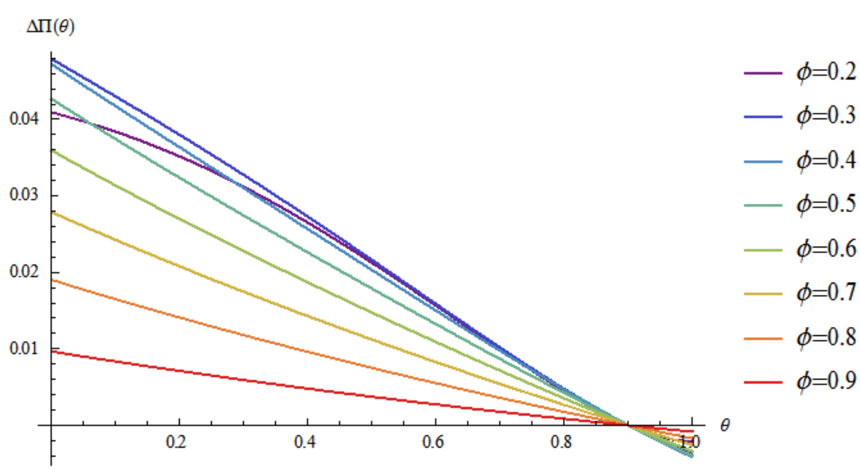


Figure 4.16: Stability of $\Delta\Pi\big(\theta, \phi, t = GCT(\phi, \theta = 0.9)\big)$ for different ϕ

4.3.3 Polycentric City with Both Final and Intermediate Products

I study the equilibrium of the model in section 4.2 following the same assumptions and space setting described in section 4.3.2. Through the equilibrium wages I can study how the city structure represented by θ relates to the transport cost of products and the commuting cost with the presence of intermediate products. Given the zero-profit condition (4.2.8), the equilibrium wage can be expressed as $W(x) = \frac{1}{f}\left[\frac{\alpha}{\sigma-1}Q(x)^* - \mu\left(1 - \ln\mu + \ln\overline{P(x)}\right)\right]$, with condition (4.3.1.5b) I have

$$W_c^* = \frac{\mu}{\sigma f}\left[\frac{\theta}{\theta+\phi(c|s)(1-\theta)} + \frac{\phi(s|c)(1-\theta)}{\phi(S|C)\theta+\left(\phi\left(s\middle|s'\right)+1\right)\frac{1-\theta}{2}}\right] +$$

$$\frac{2\alpha}{n\sigma}\int_0^{\frac{N}{2}}\frac{\phi(y|c)}{\phi(y|c)\theta+\phi(y|s)\frac{1-\theta}{2}+\phi(y|s')\frac{1-\theta}{2}}dy - \frac{\mu}{f}\left(1 - \ln\mu + \ln\overline{P_c}\right) \quad (4.3.3.1a)$$

$$W_s^* = \frac{\mu}{\sigma f}\left[\frac{\phi(c|s)\theta}{\theta+\phi(c|s)(1-\theta)} + \frac{(\phi(s'|s)+1)\frac{1-\theta}{2}}{\phi(S|C)\theta+\left(\phi\left(s\middle|s'\right)+1\right)\frac{1-\theta}{2}}\right] +$$

$$\frac{\alpha}{n\sigma}\int_{-\frac{N}{2}}^{\frac{N}{2}}\frac{\phi(y|s)}{\phi(y|c)\theta+\phi(y|s)\frac{1-\theta}{2}+\phi(y|s')\frac{1-\theta}{2}}dy - \frac{\mu}{f}\left(1 - \ln\mu + \ln\overline{P_s}\right) \quad (4.3.3.1b)$$

where $\phi(s|c) = \phi(c|s) = \phi^{\frac{1+\theta}{4}N}$, $\phi(s'|s) = \phi(s|s') = \phi^{\frac{1+\theta}{2}N}$, and $\overline{P_c} = \overline{P(0)}$ and $\overline{P_s} = \overline{P(\frac{1+\lambda}{4}N)}$ are the price indexes in CBD and SBD respectively.

First, I look at what conditions need to be satisfied for a feasible monocentric equilibrium ($\theta = 1$). Similarly to (4.3.2.11), monocentricity is feasible if the following condition holds: $\Pi_s = [f(\phi) - g(t)]f \le 0$, where $f(\phi) = W_s(\theta = 1)$ and $g(t) = W_c(\theta = 1) - \frac{N}{2}t$. When $f(\phi) - g(t) = 0$, I can again obtain t as a function of ϕ: $t = CT(\phi)$. Since I cannot have analytical solutions for $\max_{0<\phi<1} CT(\phi)$ and $\phi = LB$, numerical results are presented when $\mu = 2, \sigma = 4, f = 1; \alpha = 2, N = 4,5,6,7,8$ in Figure 4.17. It is easy to verify that the conditions follow a similar form as in (4.3.2.11). That is, the city is monocentric if and only if

$$t \leq CT(\phi) \text{ where } 0 < t \leq \max_{0<\phi<1} CT(\phi) \text{ and } LB < \phi < 1 \quad (4.3.3.2).$$

Otherwise, the monocentric structure is not sustained. Again, note that as the city grows lower commuting and trade costs are required to sustain the monocentric equilibrium.

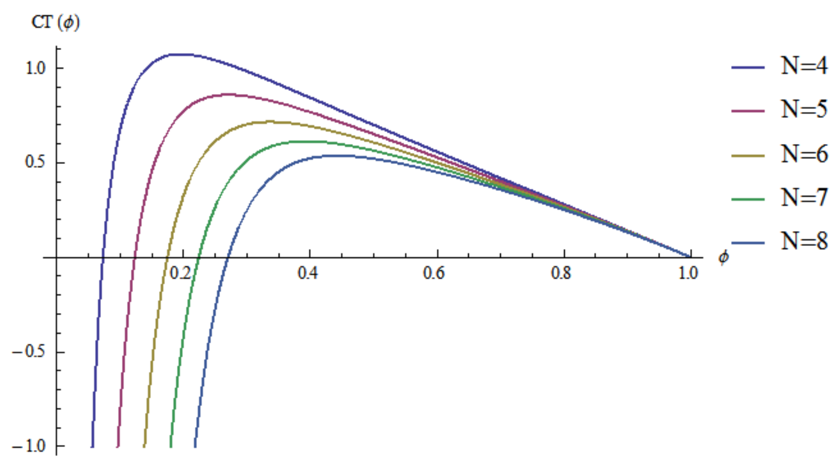


Figure 4.17: Function $CT(\phi)$ with $\mu = 2, \sigma = 4, f = 1; \alpha = 2, N = 4,5,6,7,8$ and corresponding LBs

To test its stability, I plot the profit differential function under this condition. Figure 4.18 shows $\Delta\Pi(\theta, \phi, t = CT(\phi))$ when $\mu = 2, \sigma = 4, f = 1, \alpha = 2, N = 4$ and $\phi = 0.01, 0.09, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$. Note that $\left.\frac{\partial\Delta\Pi(\theta)}{\partial\theta}\right|_{\theta=1} < 0$ does not hold for all cases. When $\phi = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6$ $\left.\frac{\partial\Delta\Pi(\theta)}{\partial\theta}\right|_{\theta=1} > 0$, while $\left.\frac{\partial\Delta\Pi(\theta)}{\partial\theta}\right|_{\theta=1} < 0$ holds for $\phi = 0.01, 0.09, 0.7, 0.8, 0.9$. I can expect that $\left.\frac{\partial\Delta\Pi(\theta)}{\partial\theta}\right|_{\theta=1}$ has its sign changed somewhere between $0.09 < \phi < 0.1$ and $0.6 < \phi < 0.7$ (refer to Table 4.1 for details). There is no way to solve the conditions analytically, but these points can be found numerically with given N, α and σ. And in a more general sense, though I cannot provide analytic solution to indicate what conditions for ϕ and t under (4.3.3.2) allow monocentricity to become a stable equilibrium, I am able to numerically test its stability given a set of specific values for ϕ and other parameters. Figure 4.19 shows $\Delta\Pi\left(\theta,\phi,t<CT(\phi)\right)$ when $\phi=0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,\mu=2,\sigma=4,f=1,\alpha=2,N=4$. Note that all curves have positive $\Delta\Pi$ at $\theta=1$, thus I can expect monocentricity is a stable equilibrium under the condition $t<CT(\phi)$ when $LB<\phi<1$.

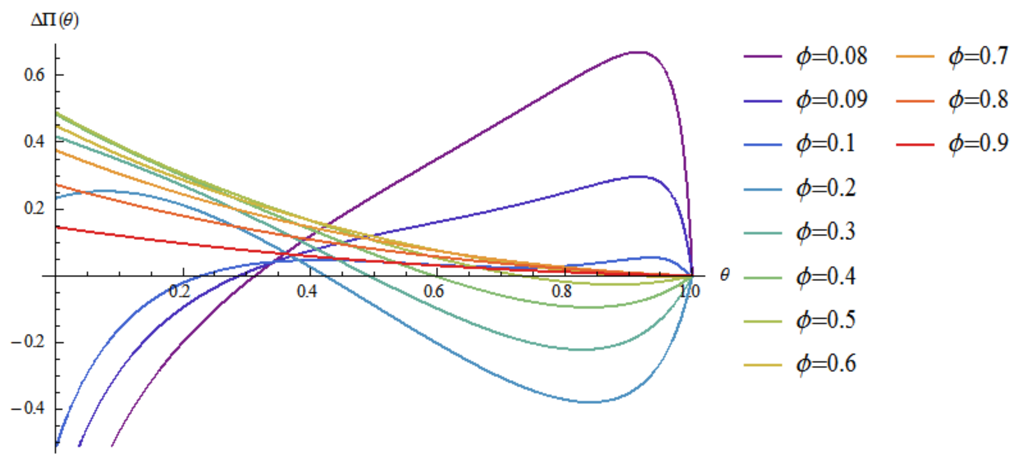


Figure 4.18: Stability of $\Delta\Pi\left(\theta,\phi,t=CT(\phi)\right)$ for different ϕ when $\mu=2,\sigma=4,f=1,\alpha=2,N=4$

Table 4.1: Stability analysis for $\Delta\Pi(\theta, \phi, t)$ with $\frac{\partial \Delta\Pi(\theta)}{\partial \theta}$

θ	**ϕ**											
	0.001	**0.01**	**0.09**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
1	N/F	N/F	-13.660	0.664	8.912	3.674	1.447	0.484	0.079	-0.061	-0.074	-0.037
0.9	-10.607	-7.431	-0.279	0.132	1.637	1.153	0.519	0.110	-0.089	-0.146	-0.123	-0.064
0.8	-1.821	-1.281	-0.021	0.062	0.425	0.285	0.032	-0.156	-0.243	-0.243	-0.187	-0.099
0.7	-0.572	-0.424	-0.115	-0.098	-0.053	-0.160	-0.298	-0.392	-0.416	-0.373	-0.278	-0.149
0.6	-0.244	-0.212	-0.215	-0.223	-0.340	-0.488	-0.613	-0.677	-0.664	-0.577	-0.428	-0.231
0.5	-0.118	-0.132	-0.315	-0.342	-0.635	-0.908	-1.105	-1.192	-1.154	-0.997	-0.739	-0.399
0.4	-0.043	-0.081	-0.590	-0.668	-1.510	-2.294	-2.854	-3.100	-3.009	-2.602	-1.928	-1.042
0.3	0.018	0.085	1.094	1.253	N/F	N/F	N/F	N/F	N/F	N/F	N/F	N/F
0.2	-0.057	0.649	0.681	0.681	0.947	1.202	1.408	N/F	N/F	N/F	N/F	N/F
0.1	-1.884	-0.074	1.451	1.406	0.937	0.757	0.744	0.760	0.737	N/F	N/F	N/F
0	-12,536.1	-71.814	4.962	4.294	1.249	0.536	0.388	0.387	0.395	0.366	0.288	N/F

Table 4.2: Stability analysis for $\Delta\Pi(\theta, \phi, t)$ with $\frac{\partial \Delta\Pi(\theta)}{\partial \theta}$ for $\theta = 0.3$

θ	**ϕ**									
	0.000001	**0.000002**	**0.000003**	**0.000004**	**0.000005**	**0.000006**	**0.000007**	**0.000008**	**0.000009**	**0.000010**
0.3	-0.003219	-0.002238	-0.001614	-0.001147	-0.000769	-0.000450	-0.000172	0.000074	0.000296	0.000499

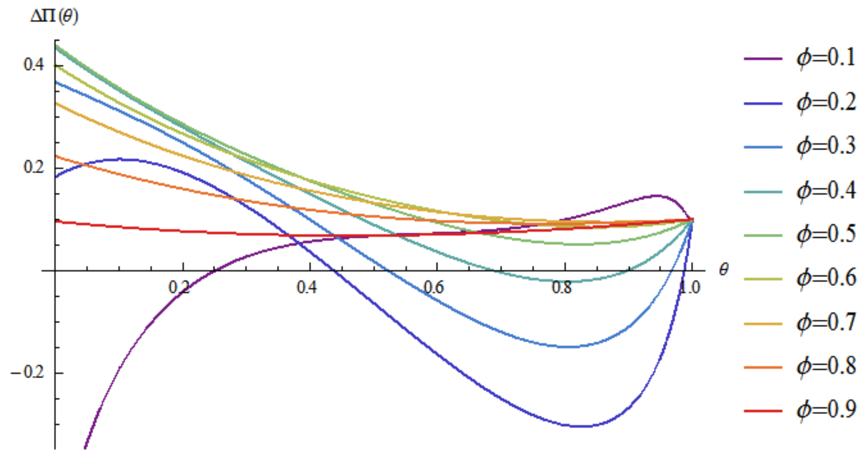


Figure 4.19: Stability of $\Delta\Pi\big(\theta, \phi, t < CT(\phi)\big)$ for different ϕ when $\mu = 2, \sigma = 4, f = 1, \alpha = 2, N = 4$

When $\theta = 1/3$, I use $WF(\phi) = W_c - W_s = 0$ to determine whether it is a sustainable equilibrium or not as in section 3.2.3.2. Due to not having analytical solutions, I plot $WF(\phi)$ when $\mu = 2, \sigma = 4, f = 1, \alpha = 2, N = 1,2,3,4,5,6,7,8,9,10$ to explore potential solutions with simulations (Figure 4.20). It illustrates that $WF(\phi) > 0$ always holds when $0 < \phi < 1$ regardless of the value of N (the proof is not provided here), which means that there is no condition to allow $\theta = 1/3$ as a feasible equilibrium.

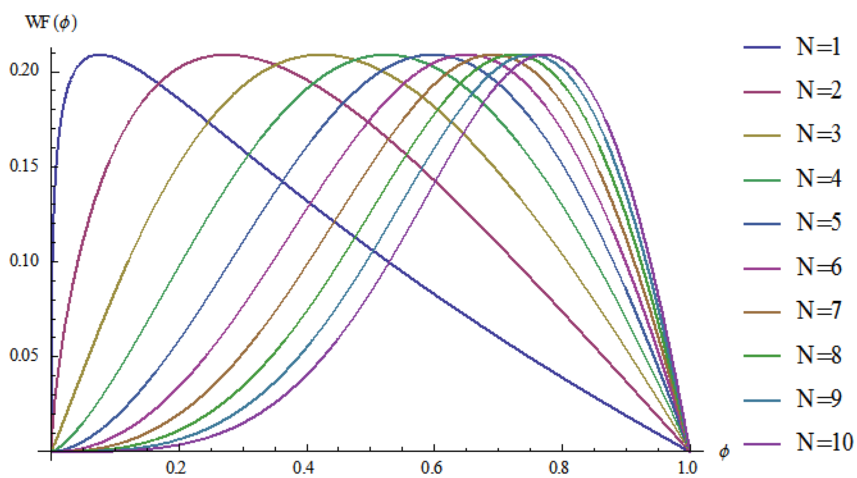


Figure 4.20: $WF(\phi)$ for $\mu = 2, \sigma = 4, f = 1, \alpha = 2, N = 1,2,3,4,5,6,7,8,9,10$

When $\frac{1}{3} < \theta < 1$, I can obtain t as a function of ϕ and θ: $t = GCT(\phi, \theta)$, which is a general form of $CT(\phi)$ as in section 4.3.2. Again, I plot the numerical results when $\mu = 2, \sigma = 4, f = 1, \alpha = 2$ and $\theta = 0.4,0.5,0.6,0.7,0.8,0.9,0.95,1$ to explore the form of $GCT(\phi, \theta)$ in order to decide the conditions (Figure 4.21). It shows that it always holds that $t \leq \max_{0<\phi<1} GCT(\phi, \theta)$ when $\frac{1}{3} < \theta < 1$ (the proof is not provided here for the existence of a single maximum for $GCT(\phi, \theta)$). In correspondence to the condition for $0 < t \leq \max_{0<\phi<1} GCT(\phi, \theta)$, ϕ must fulfill $LB < \phi < 1$ where $LB = \begin{cases} 0, if\, a\ smaller\ nonnegative\ root\ not\ exists\ for\ GCT(\phi) \\ a\ smaller\ nonnegative\ root\ for\ GCT(\phi), if\ exists \end{cases}$. In summary, for each $\theta \in (\frac{1}{3}, 1)$ there exists a relationship between t and θ: $t = GCT(\phi, \theta)$ that makes θ a feasible equilibrium of a polycentric city if and only if

$$0 < t \leq \max_{0<\phi<1} GCT(\phi, \theta) \text{ and } LB < \phi < 1 \quad (4.3.3.3)$$

When $t > GCT(\phi, \theta)$, it can be verified that $\Pi_s > \Pi_c$ following the same logic used to derive (4.3.2.9). Therefore, θ becomes infeasible and a decrease of θ follows as firms move from the CBD to SBDs until $\Pi_s = \Pi_c$ and $\theta' < \theta$ becomes a new feasible equilibrium. In contrast, if $t < GCT(\phi, \theta)$, then θ becomes infeasible because of $\Pi_c > \Pi_s$. As a result, an increase of θ will be the trend as firms move from SBDs to the CBD until $\Pi_s = \Pi_c$ and $\theta' > \theta$ becomes a new feasible equilibrium. Note that in Figure 4.21 the curves for $0.4 \leq \theta \leq 0.8$ follow a bell shape similar to those in Figure 4.10, which features a spreading-agglomeration-spreading process as well. However, the portion of $0.1 < \phi < 0.5$ for these curves reveals a fuzzy area of multiple equilibria, where the same combination of (ϕ, t) values can correspond to multiple θ values. It is interesting that for curves $0.9 \leq \theta \leq 1$ within this portion, lower commuting costs are needed to support less concentrated patterns, although the monocentricity is also feasible due to the multiplicity of equilibria. A possible explanation is that as a small portion of firms move to the SBDs, transport costs of goods to both households and firms increase because of this decentralization and need to be compensated by a reduction on their wage bills caused by the falling commuting costs of their workers.

Similarly as before, I use plotting and numerical methods to test the stability for $\theta \in (\frac{1}{3}, 1)$. I plotted $\Delta\Pi\big(\theta, \phi, GCT(\phi, \theta)\big)$under the condition (4.3.3.3) with $\phi = 0.01, 0.09, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, \mu = 2, \sigma = 4, f = 1, \alpha = 2, N = 4$ for feasible equilibria $\theta = 0.9, 0.8, 0.7, 0.6, 0.5, 0.4$ respectively (Figure 4.22-4.27). $\frac{\partial\Delta\Pi(\theta)}{\partial\theta}$ for each curve is compiled in Table 4.1 for reference. For $\theta = 0.9$ and 0.8, not all curves are stable at their θ. The sign change of $\frac{\partial\Delta\Pi(\theta)}{\partial\theta}$ occurs between $0.09 < \phi < 0.1$ and $0.5 < \phi < 0.6$ for $\theta = 0.9$ and between $0.09 < \phi < 0.1$ and $0.4 < \phi < 0.5$ for $\theta = 0.8$. For $\theta = 0.7, 0.6, 0.5, 0.4$, all cases plotted are stable ($\frac{\partial\Delta\Pi(\theta)}{\partial\theta} < 0$) at their feasible θ. Again, though I cannot provide analytic solution to indicate what conditions for ϕ and t constrained by (4.3.3.2) allow polycentricity reflected by θ to be a stable equilibrium, I am able to numerically test its stability given a set of specific values for ϕ and other parameters.

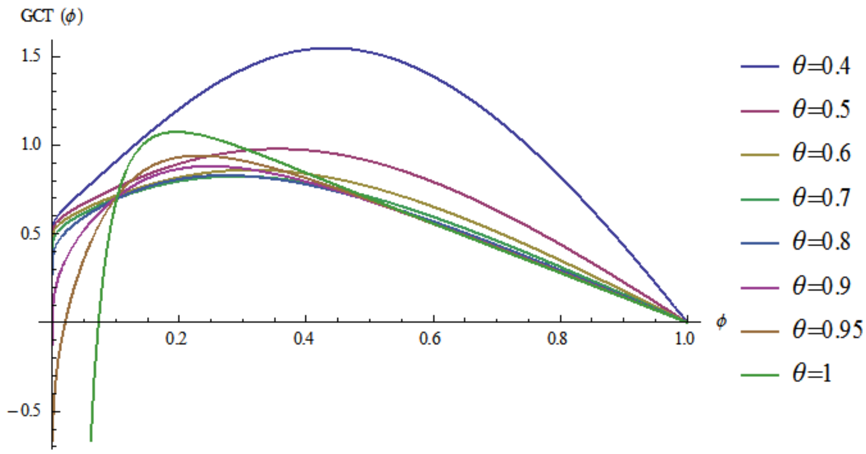


Figure 4.21: $GCT(\phi,\theta)$ when $\mu = 2, \sigma = 4, f = 1, \alpha = 2, N = 4$ and $\theta =$ 0.4,0.5,0.6,0.7,0.8,0.9,0.95,1

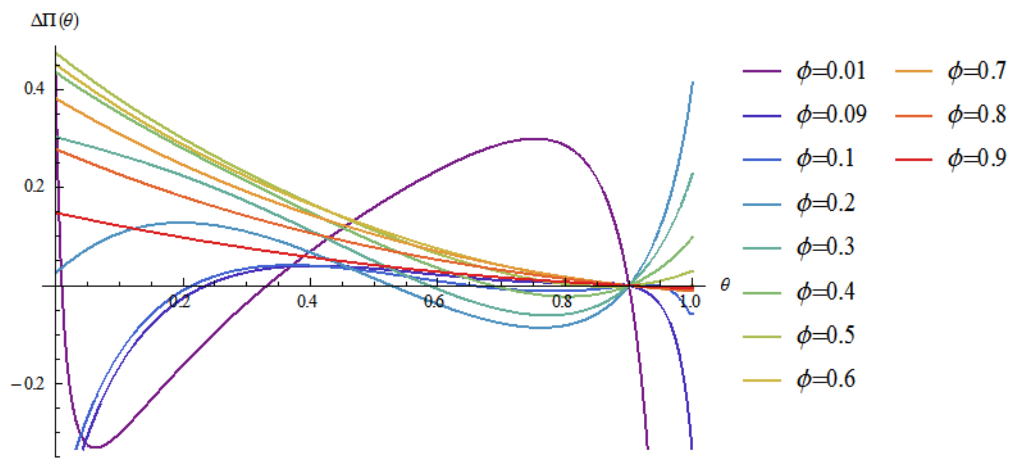


Figure 4.22: Stability of $\Delta\Pi\big(\theta, \phi, t = GCT(\phi, \theta = 0.9)\big)$ for different ϕ

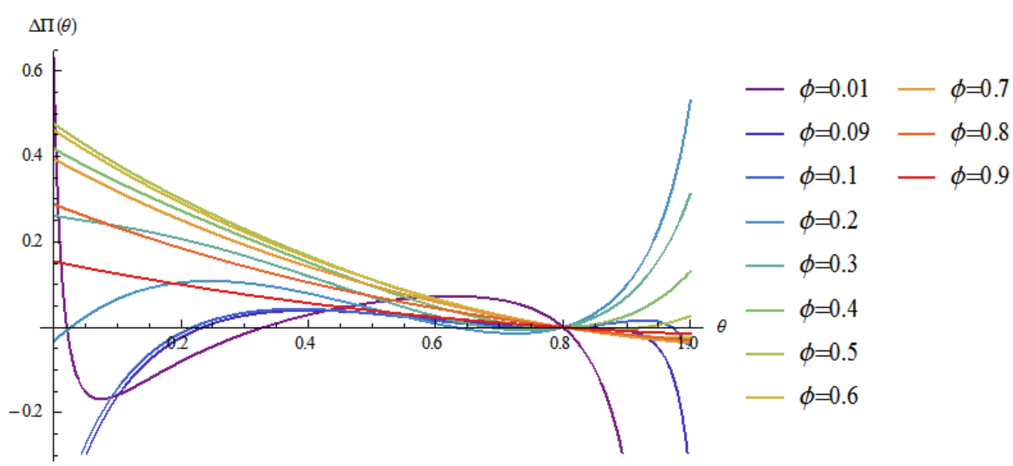


Figure 4.23: Stability of $\Delta\Pi\big(\theta, \phi, t = GCT(\phi, \theta = 0.8)\big)$ for different ϕ

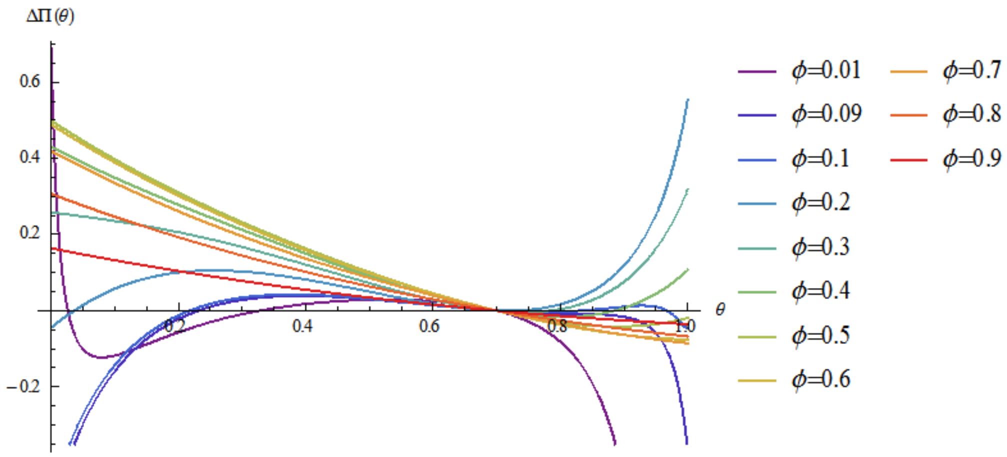


Figure 4.24: Stability of $\Delta\Pi\big(\theta, \phi, t = GCT(\phi, \theta = 0.7)\big)$ for different ϕ

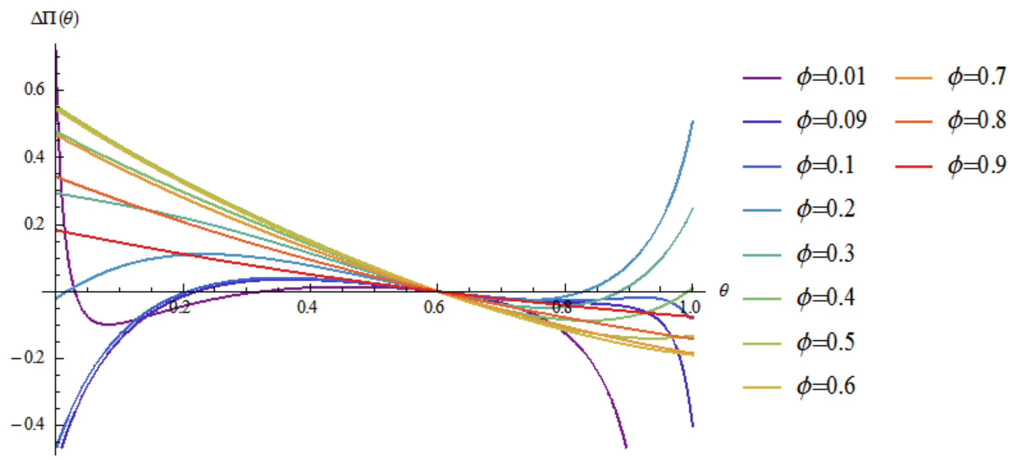


Figure 4.25: Stability of $\Delta\Pi\big(\theta, \phi, t = GCT(\phi, \theta = 0.6)\big)$ for different ϕ

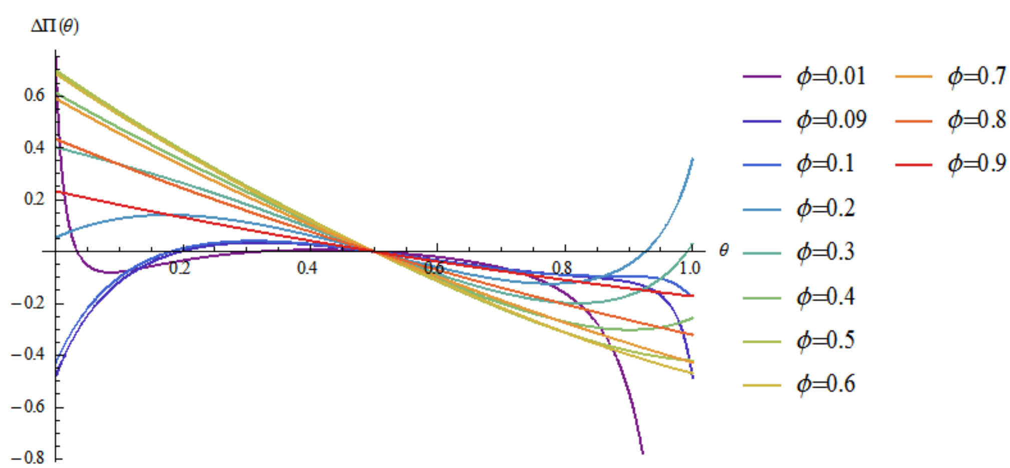


Figure 4.26: Stability of $\Delta\Pi\big(\theta, \phi, t = GCT(\phi, \theta = 0.5)\big)$ for different ϕ

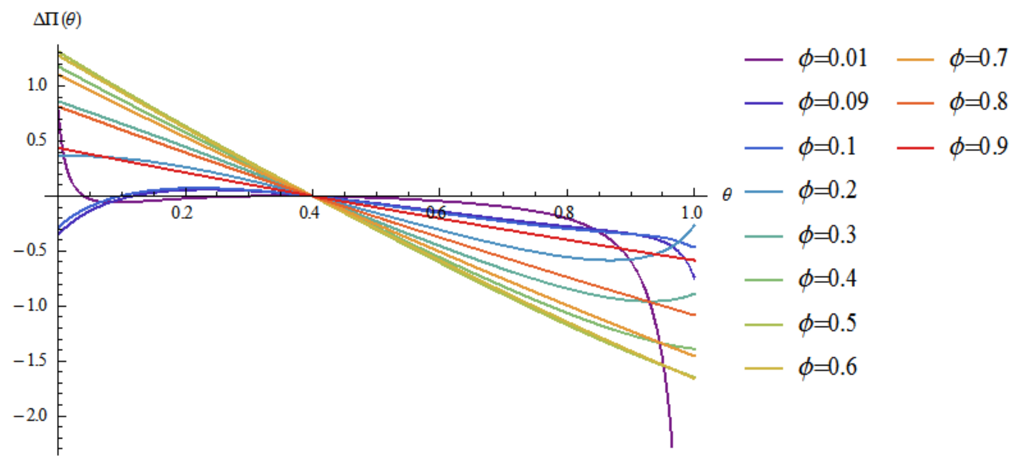


Figure 4.27: Stability of $\Delta\Pi\big(\theta, \phi, t = GCT(\phi, \theta = 0.4)\big)$ for different ϕ

When $0 < \theta < 1/3$, $W_c - W_s < 0$ based on equation (4.3.2.4). Plotting $WF(\phi)$ with cases of $\mu = 2, \sigma = 4, f = 1, \alpha = 2$ and $\theta = 0.1,0.2,0.3,1/3$ (Figure 4.28), I see that contrary to Figure 4.8, $WF(\phi)$ can be fulfilled by certain conditions when $0 < \theta < 1/3$. Again, using $t = GCT(\phi, \theta)$ I plot it when $\mu = 2, \sigma = 4, f = 1, \alpha = 2$ and $\theta = 0.01,0.05,0.1,0.2,0.3$ to explore these conditions (Figure 4.29). It also shows that $t \leq \max_{0<\phi<1} GCT(\phi, \theta)$ always holds when $0 < \theta < 1/3$. In correspondence to the condition for $0 < t \leq \max_{0<\phi<1} GCT(\phi, \theta)$, ϕ must fulfill $LB < \phi < UB$ where $LB = \begin{cases} 0, if\ root_1 \geq 0\ not\ exists \\ root_1 \geq 0, if\ exists \end{cases}$, $UB = \begin{cases} 1, if\ root_2 < 1\ not\ exists \\ root_2 < 1, if\ exists \end{cases}$ and $root_1 < root_2$ are both root of $GCT(\phi) = 0$. In summary, for each $\theta \in (0, \frac{1}{3})$ there exists a relationship between t and θ: $t = GCT(\phi, \theta)$ that makes θ an feasible equilibrium as a polycentric city if and only if

$$0 < t \leq \max_{0<\phi<1} GCT(\phi, \theta) \text{ and } LB < \phi < UB \quad (4.3.3.4)$$

When $t > GCT(\phi, \theta)$, $\Pi_s > \Pi_c$ follows. Then, θ become infeasible and θ decreases until $\Pi_s = \Pi_c$ and $\theta' < \theta$ becomes a new feasible equilibrium. When $t < GCT(\phi, \theta)$, θ

becomes infeasible because of $\Pi_c > \Pi_s$. θ will increase until $\Pi_s = \Pi_c$ and $\theta' > \theta$ becomes a new feasible equilibrium. Note that in Figure 4.29, decreasing trade costs support the decentralization of firms and the concentration of economy in SBDs (θ declines). However, higher commuting costs are needed to facilitate this process, because workers commute less distance if they find jobs in SBDs.

To test their stability, I plot $\Delta\Pi\big(\theta, \phi, GCT(\phi, \theta)\big)$ under condition (4.3.3.4) with $\mu = 2, \sigma = 4, f = 1, \alpha = 2, N = 4$ for feasible equilibria $\theta = 0.3, 0.2, 0.1$ respectively (Figure 4.30-4.33). Note that for $\theta = 0.2$ and 0.1(Figure 4.32 and 4.33 respectively), $\frac{\partial \Delta\Pi(\theta)}{\partial \theta}$ changes sign between $0.001 < \phi < 0.01$ and $0.01 < \phi < 0.09$ respectively (refer to Table 4.1). Thus, they are stable when trade cost τ reaches a very high level (ϕ is very low). For $\theta = 0.3$ (Figure 4.30 and 4.31), $\frac{\partial \Delta\Pi(\theta)}{\partial \theta}$ changes sign between $0.000007 < \phi < 0.000008$ (refer to Table 4.2 and Figure 4.31), which suggests that trade cost τ needs to reach an extremely high level for this equilibrium to become stable.

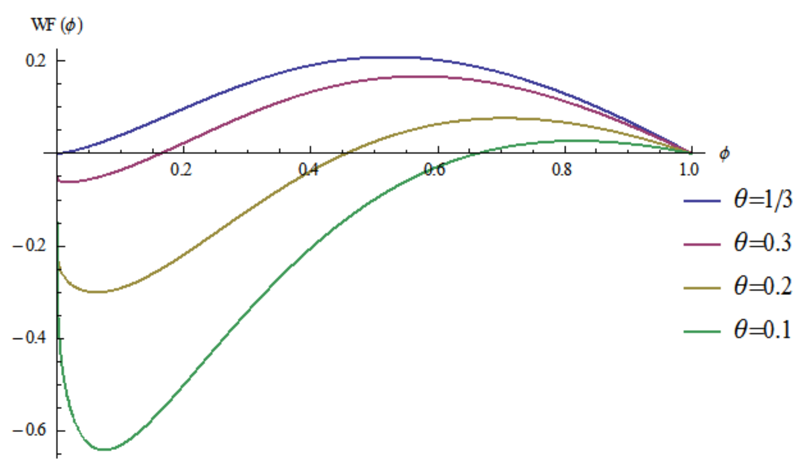


Figure 4.28: Function $WF(\phi)$ when $\theta < \frac{1}{3}, \mu = 2, \sigma = 4, f = 1, \alpha = 2, N = 4$

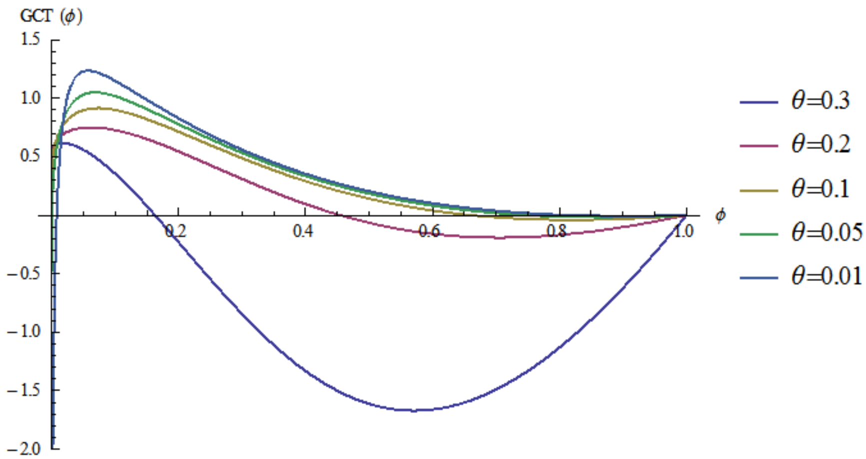


Figure 4.29: $GCT(\phi, \theta)$ when $\mu = 2, \sigma = 4, f = 1, \alpha = 2, N = 4$ and $\theta = 0.01, 0.05, 0.1, 0.2, 0.3$

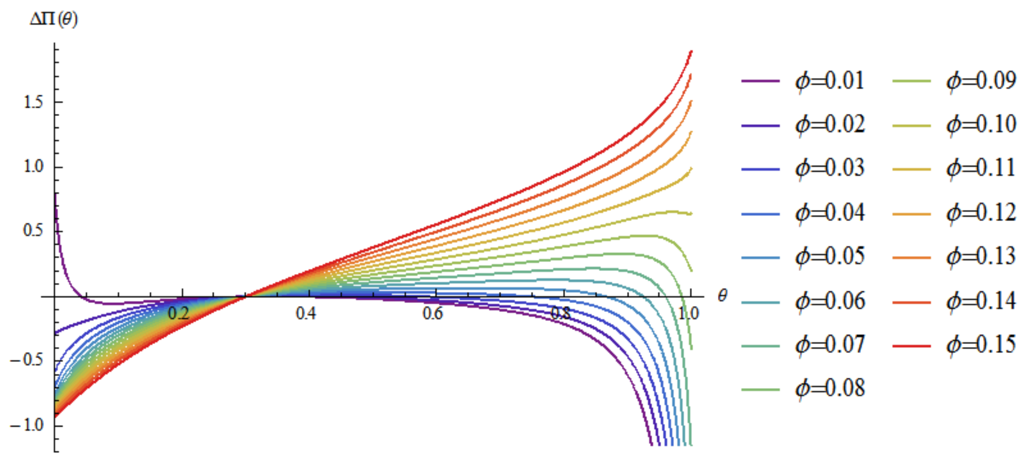


Figure 4.30: Stability of $\Delta\Pi\big(\theta, \phi, t = GCT(\phi, \theta = 0.3)\big)$ for different ϕ

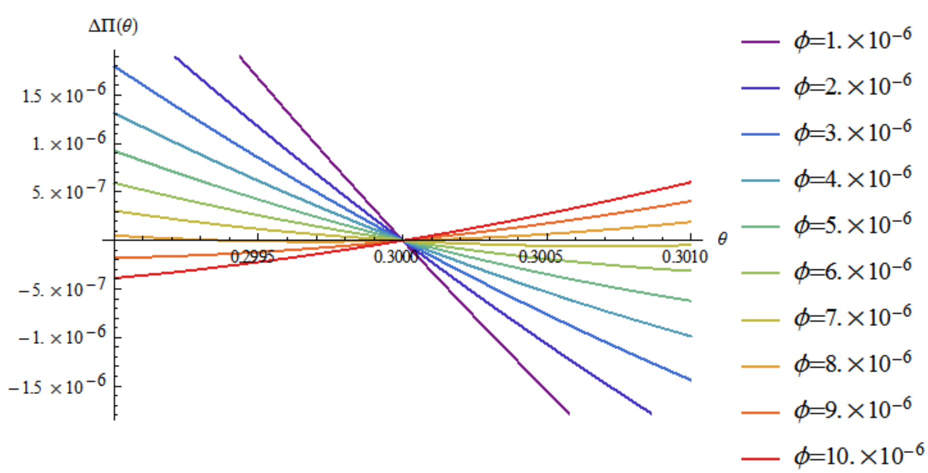


Figure 4.31: Stability of additional $\Delta\Pi\big(\theta, \phi, t = GCT(\phi, \theta = 0.3)\big)$ for small values of ϕ

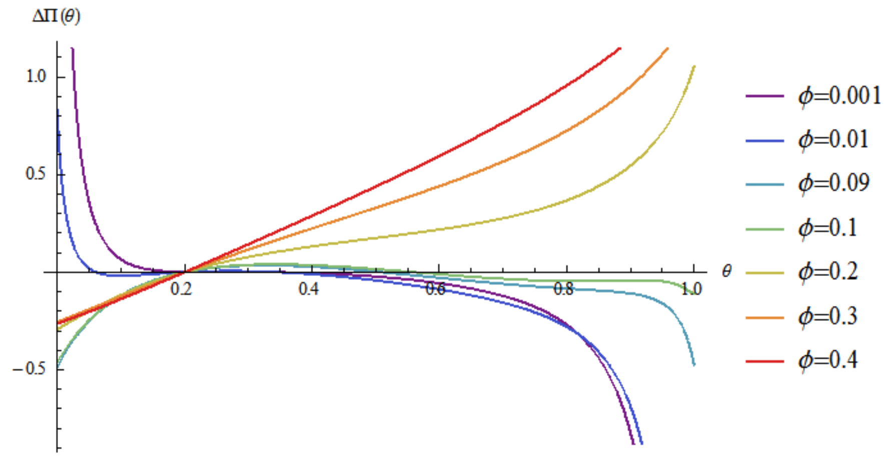


Figure 4.32: Stability of $\Delta\Pi\big(\theta, \phi, t = GCT(\phi, \theta = 0.2)\big)$ for different ϕ

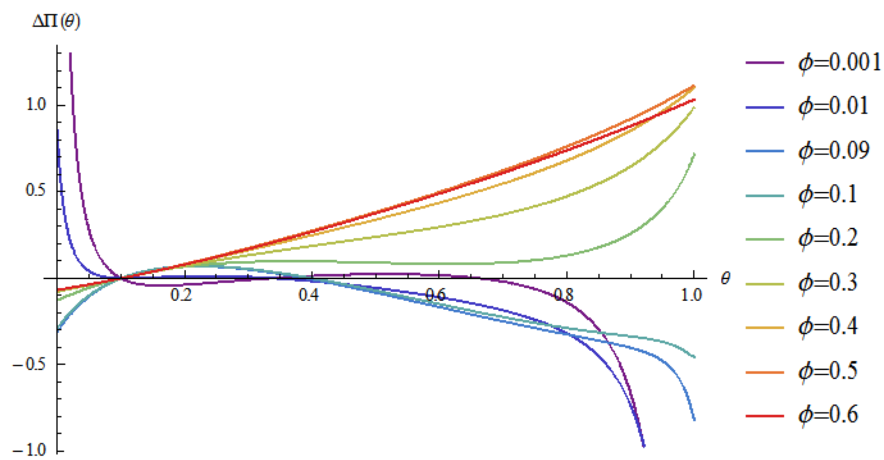


Figure 4.33: Stability of $\Delta\Pi\big(\theta, \phi, t = GCT(\phi, \theta = 0.1)\big)$ for different ϕ

Finally, when $\theta = 0$, the CBD disappears after its shrinkage and the city is still polycentric but it only has two symmetric SBDs. To examine the stability of this equilibrium, I follow the same logic when I look at the conditions for $\theta = 1$. I derive the potential level of wages and firms' profit if there were firms in the CBD. Noting that the equilibrium profits in SBDs are equal to 0, it will be attractive for firms to move from SBDs to the CBD if the potential profit in the CBD is greater than 0. In other words, the polycentricity with only two SBDs is not stable if the following condition holds:

$$\Pi_c = [f(\phi) - g(t)]f > 0 \quad (4.3.3.5)$$

where $f(\phi) = W_c(\theta = 0)$ and $g(t) = W_s(\theta = 0) - \frac{N}{4}t$. Let $f(\phi) - g(t) = 0$, I can obtain t as a function of ϕ: $t = GCT(\phi)$. Then (4.3.3.5) can be reduced to $t > GCT(\phi)$. Plotting $GCT(\phi)$ numerically in Figure 4.34 ($N = 2,3,4,5,6,7,8,9, \mu = 2, \sigma = 4, f = 1, \alpha = 2$), it shows a similar form as that for $0 < \theta < 1/3$. I find its condition follows a similar form as (4.3.3.4). That is, the city is in feasible polycentric form with only two SBDs if and only if

$$t \leq GCT(\phi) \text{ where } 0 < t \leq \max_{0<\phi<1} GCT(\phi) \text{ and } LB < \phi < UB. \quad (4.3.3.6)$$

Otherwise, this polycentric structure is infeasible. Note that as city size N grows lower maximum t and τ cost are needed to sustain the equilibrium's feasibility.

Under condition (4.3.3.6), I test the stability of feasible equilibria by plotting $\Delta\Pi(\theta, \phi, t)$. Figure 4.35 shows that $\Delta\Pi\big(\theta, \phi, t = GCT(\phi)\big)$ when $N = 2, \alpha = 1, \sigma = 4, \phi = 0.001, 0.01, 0.09, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$. Note that $\frac{\partial \Delta\Pi(\theta)}{\partial \theta}\Big|_{\theta=0}$ changes sign between $0.01 < \phi < 0.09$. Therefore, I can expect $\theta = 0$ is stable when $t = GCT(\phi)$ and ϕ is sufficiently low (trade cost is sufficiently high). Figure 4.36 shows $\Delta\Pi\big(\theta, \phi, t < GCT(\phi)\big)$ when $N = 2, \alpha = 1, \sigma = 4, \phi = 0.01, 0.09, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6$. Note that $\Delta\Pi(\theta = 0) < 0$ for all curves, suggesting the stability of the equilibrium.

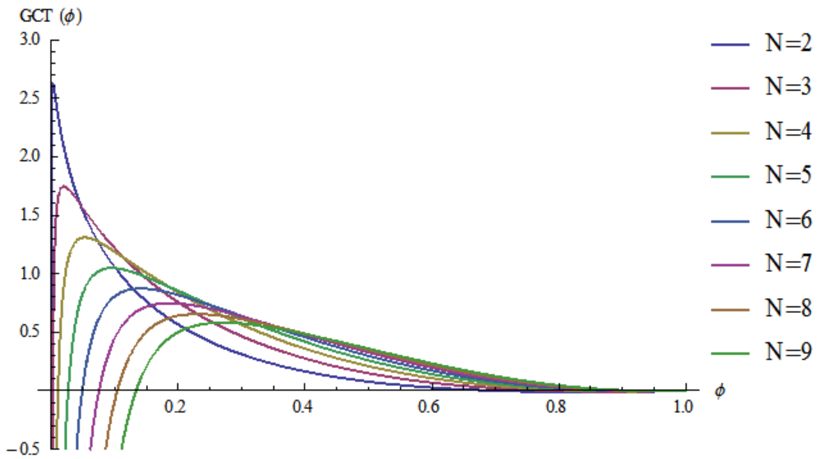


Figure 4.34: $GCT(\phi, \theta = 0)$ with corresponding LBs and UBs when $N = 2,3,4,5,6,7,8,9, \mu = 2, \sigma = 4, f = 1, \alpha = 2$

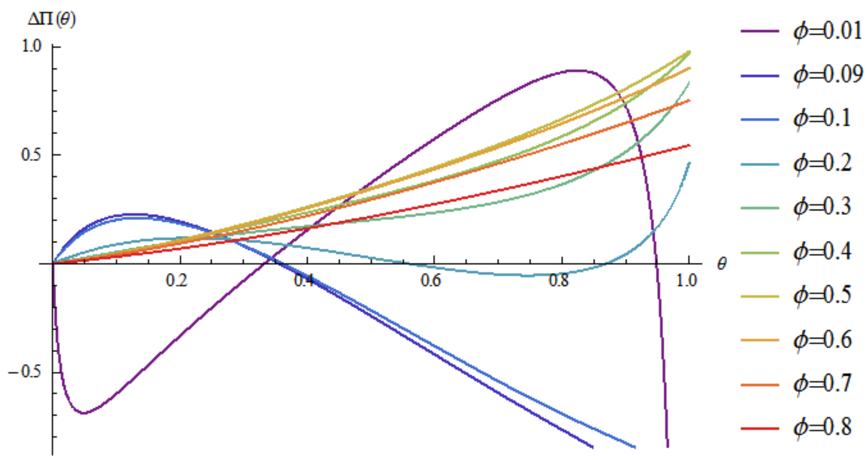


Figure 4.35: Stability of $\Delta\Pi\big(\theta, \phi, t = GCT(\phi, \theta = 0)\big)$ for different ϕ

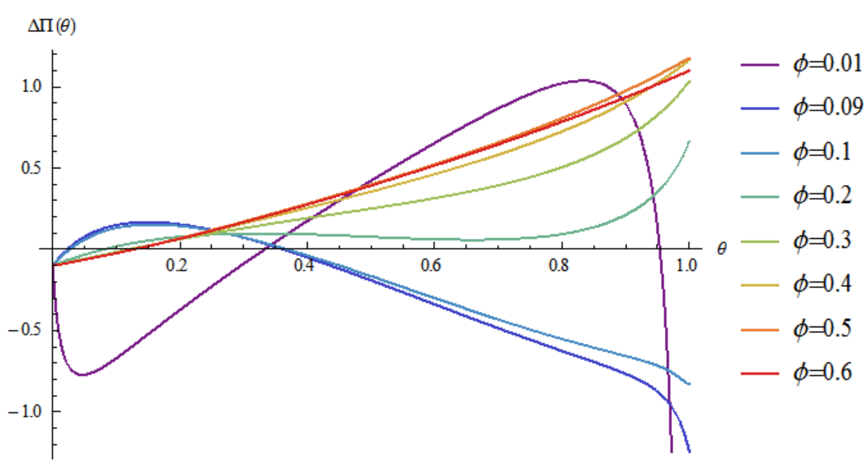


Figure 4.36: Stability of $\Delta\Pi\big(\theta, \phi, t < GCT(\phi, \theta = 0)\big)$ for different ϕ

4.3.4 Summary

In general, the study of equilibrium conditions for the proposed models proves our expectations that incorporating market linkages in final and intermediate goods with trade costs and labor market linkage with commuting costs enables the generation of both monocentric and polycentric structures from the models. Specifically, the main findings suggested by the numerical results in Sections 4.3.2 and 4.3.3 are threefold. First, as demonstrated by both polycentric urban models in general, increasing (decreasing) commuting costs strengthen (weaken) the dispersion forces and weaken (strengthen) the agglomeration forces. Regarding the trade cost, as it decreases, it first strengthens the agglomeration force of market-access, and then makes the dispersion forces of market-crowding dominant. This features a spreading-agglomeration-spreading process consistent with that found in literature. Second, the polycentric model with both final and intermediate goods incorporate input-output linkages between firms, and thus involve additional agglomeration forces compared to the model with only final goods. This added type of forces allows it to exhibit more complex behaviors, such as multiplicity of equilibria. Furthermore, it enables stronger decentralization and concentration of economy on SBDs which are absent in the model with only final goods. Third, because of the non-linearity of the proposed models, their equilibrium conditions and stability are not analytically tractable and proved. Instead, I seek to employ numerical methods to solve for their conditions and test their stability.

CHAPTER 5: MULTISCALAR POLYCENTRIC MODEL OF URBAN SYSTEMS

This chapter aims to construct a multiscalar polycentric urban systems model in full conformity with the generic framework (Chapter 3) in order to account for the interdependency of spatial structures across scales. Urban systems models of NEG type employ abstract space settings such as standard two-region settings, linear space (Fujita and Mori, 1997; Mori, 1997; Fujita et al., 1999) or a racetrack (the circumference of a circle) configuration (Krugman, 1993, 1996; Mossay, 2003; Picard and Tabuchi, 2010; Tabuchi and Thisse, 2011; Ikeda et al., 2012; Akamatsu et al., 2012). For the sake of simplicity, all existing models that unify NEG models with internal urban extensions adopt the standard but less realistic two-region setting. Except for the work by Cavailhès et al. (2007) that allows the emergence of polycentric urban structures, they all embed classical urban economic models with monocentricity assumptions. Though Cavailhès et al. also assume a given CBD, they endogenize the size and location of SBDs by modeling spatial costs within and between cities such as commuting costs, communication costs and trade costs. In this chapter, I develop the multiscalar urban systems model by nesting the polycentric urban models (Section 4.2) within racetrack urban systems characterized by inter-urban trading in final and intermediate goods markets. This racetrack configuration of urban systems allows any number of cities to be modeled, which provides a more general approach than the two-region setting. The choice of racetrack space lies in its symmetric structure (not symmetric distance as in multi-region CP model) that eliminates the spatial boundary effect in linear space (the two ends of the line). Then, I describe the long-run equilibrium conditions for the urban systems and specify the transition rules that help to single out evolutionary paths for the urban systems under structural changes. Finally, I study the possible evolutionary paths of long-run equilibria given the exogenous change of different types of spatial costs at both intra- and inter-urban scales and how these spatial costs affect the interdependency of spatial structures at different scales in terms of polycentricity/monocentricity.

5.1 Systems of Polycentric Urban Models

Assume a racetrack economy, where r regions are equally and sequentially spaced around the circumference of a circle (Figure 5.1), with region $i + 1$ next to region i, and with region r next to region 1. The land quality is homogenous and the land is evenly distributed in space with a density equal to 1. Transportation is only possible along the circumference where the distance between any two adjacent regions is one unit. Hence, the distance between any two arbitrary regions is the length of the shortest route along the circumference (e.g., the distance from region 1 to region 6 is 3 units). Based on this configuration, each city/region is modeled by a polycentric urban model (Section 4.2) with a spatial extension. The internal urban space is still represented by a line as in Chapter 4. The spatial size of a city is related to the population it accommodates and is independent to the racetrack space. In other word, I assume the distances between cities according to their locations on the racetrack are constant and much larger than the spatial size of cities. Hence, as cities grow their spatial size will not affect their distances to other cities. Cities interact with each other via trading goods subject to the trading cost. Assuming iceberg transport technology, shipping one unit of variety over a distance $|i -$

$j|$ requires $T(i|j) = T^{|i-j|}$ units of numeraire, where $|i - j|$ is the shortest distance along the circumference from region *j* to *i* and $T > 1$.

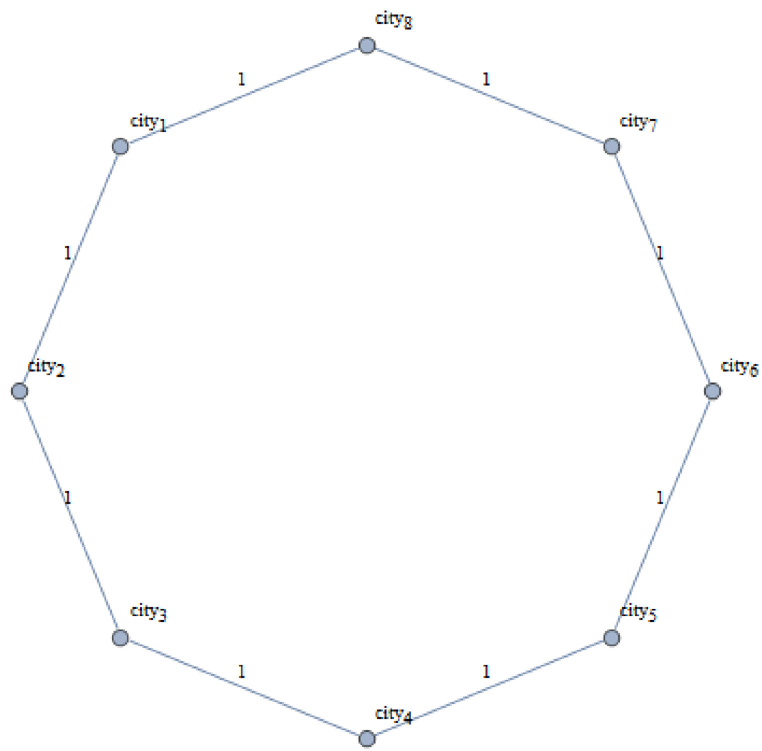


Figure 5.1: A racetrack configuration for a system of 8 cities

5.1.1 Household Consumption

In an urban system, due to the preference for varieties, households not only consume locally produced varieties but also those from remote regions through trading. Thus, the utility function of a household at location *y* of region *i* is given as:

$$U_i(y) = \alpha \ln \left[\sum_j \int_{x \in X}^{M_j} m_j(x) q_{ji}(y|x)^{\frac{\sigma-1}{\sigma}} dx \right]^{\frac{\sigma}{\sigma-1}} + H_i(y) \quad (5.1.1.1)$$

where the demand of a household at location *y* in region *i* for varieties produced at location *x* in region *j* is $q_{ji}(y|x) = \begin{cases} q_j(x)\tau(y|x)^{-1}, & i = j \\ q_j(x)T(i|j)^{-1}, & i \neq j \end{cases}$, which is differentiated by the spatial costs borne within or between urban regions. Accordingly, the budget constraint of a household residing at location *y* and working at $J_i(y)$ (the commuting function at region *i*) is as follows:

$$\sum_j \int_{x\in X}^{M_j} m_j(x)p_{ji}(y|x)q_{ji}(y|x)\,dx + R_i(y)S_h + H_i(y) = W_i[J_i(y)] - t|y - J_i(y)| \quad (5.1.1.2)$$

where the delivered price from the production location *x* of region *j* to the consumption location *y* of region *i* depends on the spatial costs borne within or between urban regions due to the iceberg transport technology:

$$p_{ji}(y|x) = \begin{cases} p_j(x)\tau(y|x), & i = j \\ p_j(x)T(i|j), & i \neq j \end{cases} \quad (5.1.1.3)$$

Similarly to what was discussed before (Equation 4.2.5), by maximizing (5.1.1.1) subject to (5.1.1.2) with respect to the choice of consumption bundle, demand for a variety from location *x* of region *j* by a household at location *y* of region *i* is as follows:

$$q_{ji}(y|x) = \frac{\alpha}{\overline{P_i(y)}}\left[\frac{p_{ji}(y|x)}{\overline{P_i(y)}}\right]^{-\sigma} \quad (5.1.1.4)$$

where the price index at location *y* of region *i* is defined as:

$$\overline{P_i(y)} = \left[\sum_j \int_{x\in X}^{M_j} m_j(x)p_{ji}(y|x)^{1-\sigma}\,dx\right]^{\frac{1}{1-\sigma}} \quad (5.1.1.5)$$

Inserting (5.1.1.5) back into (5.1.1.1), I finally obtain the indirect utility function for a household at location *y* of region *i*

$$V_i(y) = \alpha\left[\ln\alpha - 1 - \ln\overline{P_i(y)}\right] + W_j[J_i(y)] - t|y - J_i(y)| - R_i(y) \quad (5.1.1.6)$$

and its associated bid rent function

$$\Psi_i(y, U_i^*) = \alpha\left[\ln\alpha - 1 - \ln\overline{P_i(y)}\right] + W_j[J_i(y)] - t|y - J_i(y)| - U_i^* \quad (5.1.1.7).$$

5.1.2 Firm Production

The production of firms must also take into account the trade between regions. Based on the production technology with both final and intermediate goods proposed in section 3.2.2, I expand it to account for the costly trading of both final and intermediate goods between each pair of urban regions. Hence, the production function of a firm at location *x* in region *i* is as follows:

$$aQ_i(x) = \mu \ln I_i(x) + H,\ \mu > 0 \quad (5.1.2.1)$$

where $Q_i(x)$ is the output of production, H is a constant quantity of the numeraire, and *a* is the marginal requirements. $I_i(x)$ is a composite intermediate good taking varieties across all regions: $I_i(x) = \left[\sum_j \int_{x_I \in X}^{M_j} m_j(x_I) q_{I,ji}(x|x_I)^{\frac{\sigma-1}{\sigma}} dx_I\right]^{\frac{\sigma}{\sigma-1}}$,

where $q_{I,ji}(x|x_I)$ is the demand for the intermediate variety from each firm at x_I of region *j* by a firm at *x* of region *i*. Since every firm also consume a fixed amount of labor $f > 0$ and a fixed amount of land S_f, I have the cost function of a firm at location *x* in region *j* defined as:

$$C_i(x) = W_i(x)f + R_i(x)S_f + \sum_j \int_{x_I \in X}^{M_j} m_j(x_I) p_{ji}(x|x_I) q_{I,ji}(x|x_I) dx_I + H \quad (5.1.2.2)$$

where the effective price from location *x* of region *j* to location *y* of region *i* is subject to spatial costs either within or between urban regions:

$$p_{ji}(x|x_I) = \begin{cases} p_j(x_I)\tau(x|x_I),\ i = j \\ p_j(x_I)T(i|j),\ i \neq j \end{cases} \quad (5.1.2.3)$$

By maximizing production $Q_i(x)$ in (5.1.2.1) subject to the cost constraint (5.1.2.2), the optimal demands for labor and intermediate products by a firm at location *x* of region *i* can be obtained respectively:

$$q_{I,ji}(x|x_I) = \frac{\mu}{\overline{P_i(x)}}\left[\frac{p_{ji}(x|x_I)}{\overline{P_i(x)}}\right]^{-\sigma} \quad (5.1.2.4)$$

where the price index at *x* of region *i* is in the same form as (5.1.1.5):

$\overline{P_i(x)} = \left[\sum_j \int_{x_I \in X}^{M_j} m_j(x_I) p_{ji}(x|x_I)^{1-\sigma}\, dx_I\right]^{\frac{1}{1-\sigma}}$. Then, following the same procedure as in section 3.2.2, I can rewrite cost function:

$$C_i(x) = W_i(x)f + R_i(x)S_f + \mu - \mu\left[\ln\mu - \ln\overline{P_i(x)}\right] + aQ_i(x) \quad (5.1.2.5)$$

where for a firm at location *x* of region *i* its marginal production cost is a. Then, the optimal f.o.b. price charged by a firm at location *x* of region *i* is the markup $\frac{\sigma}{\sigma-1}$ multiplied by the marginal production cost:

$$p_i(x) = a\frac{\sigma}{\sigma-1} = p^* \quad (5.1.2.6)$$

After applying the zero-profit condition to the profit function of a firm at location *x* of region *i*

$$\Pi_i(x) = p^*Q_i(x) - C_i(x) = 0 \quad (5.1.2.7),$$

the equilibrium production of the firm can be obtained as:

$$Q_i^*(x) = \frac{\sigma-1}{a}\left[W_i(x)f + R_i(x)S_f + \mu\left(1 - \ln\mu + \ln\overline{P_i(x)}\right)\right] \quad (5.1.2.8)$$

Finally, I can obtain its associated bid rent function as follows:

$$\Phi_i(x,\Pi^*) = \frac{1}{S_f}\left[\frac{a}{\sigma-1}Q_i^*(x) - W_i(x)f - \mu\left(1 - \ln\mu + \ln\overline{P_i(x)}\right) - \Pi^*\right] \quad (5.1.2.9)$$

5.1.3 Short-run Equilibrium

In the model of urban systems, the short-run equilibrium for each region follows the same definition as in section 4.3.1. That is, given a fixed population of households N_i for region *i*, conditions (1-4) apply to *population constraints* (4.3.1.1), *Commuting equilibrium* (4.3.1.2), *labor market equilibrium* (4.3.1.3), and *land market equilibrium*

(4.3.1.4) respectively at the intra-urban scale. However, due to the inter-regional trading at the inter-urban scale, condition 5 must be specified differently for product market equilibrium, because the demand for consumption comes not only from inside the region but also from outside the region due to inter-regional trading. I have a modified condition 5 as:

Product market equilibrium is given by equating the production $Q_i(x)$ of a variety by a firm at location *x* of region *i* and the demand $D_i(x)$ for it from both the intra-urban level and the inter-urban level:

$$Q_i(x) = D_i(x) = \left\{\int_{y\in X}\left[n_i(y)q_{ii}(y|x)\tau(y|x) + m_i(y)q_{I,ii}(y|x)\tau(y|x)\right]dy\right\} +$$

$$\left\{\sum_{j\neq i} T(i|j)\int_{y\in X}\left[n_j(y)q_{ij}(y|x) + m_j(y)q_{I,ij}(y|x)\right]dy\right\} \quad (5.1.3.1)$$

where the first item on the right-hand side is the internal demand and the second item is the external demand.

5.2 Long-run Equilibrium and Adjustment Dynamics

In the long run, labor is assumed to be mobile inter-regionally. As a result, the population of households in each urban-region is not fixed anymore. Specifically, households respond to the differences in the utility level and choose to migrate to regions with a higher one, which in turn affects the utility level at both origin and destination regions, until the utility levels in all regions are equal (then there is no incentive to migrate). Mathematically, given the total population N in the urban system economy, it is to find $\left(N_i^*, \widehat{U}\right)$ to satisfy:

$$\begin{cases} \left(U_i^* - \widehat{U}\right)N_i^* = 0, \quad N_i^* \geq 0, \quad U_i^* \leq \widehat{U}, \quad (i = 1,\cdots,r) \\ \sum_{i=1}^{r} N_i^* = N \end{cases} \quad (5.2.1)$$

where U_i^* is the short-run equilibrium at region i, $\widehat{U}$ is the highest and equilibrium utility level prevailing in the whole system in the long run, and N_i^* is the long-run equilibrium population in region *i* under utility level $\widehat{U}$.

In addition, the exogenous change to transport costs (e.g., inter-urban and intra-urban trade costs and commuting cost) can be assumed as a long-run process in the economy in order to examine the evolution of spatial agglomeration over time. I assume here that the endogenous migration process takes place very fast compared with the exogenous change to transport costs such that the adjustment can be accomplished over 'fictitious time'. In other words, starting from an equilibrium state, whenever the a transport cost changes a certain amount it temporarily stops and allows the migration adjustment to settle the economy into a new equilibrium state; then it repeats these two steps. In this sense, it essentially exhibits a series of comparative statics of long-run equilibria regarding to the exogenous change to transport costs. Also note that a long-run equilibrium is based on a sequence of short-run equilibria in that the migration adjustment process equalizes the utility level between regions (5.2.1).

To account for the possible emergence of new spatial configurations (growth/born and shrinkage/death of centers and cities), it is worth examining the stability of long-run equilibria. In other words, at a given time $\tilde{\mathbf{t}}$, the spatial distribution of population in the urban system may experience a small random fluctuation from equilibrium. This equilibrium is (locally) stable if the perturbed population distribution recovers to the original one. Otherwise, it phases into another stable equilibrium because the original one is not stable. The dynamic adjustment (migration dynamics) adopted to reach long-run equilibrium is a type of replicator dynamics, which is from evolutionary game theory and has been extensively applied in the context of NEG (Krugman, 1991; Fujita and Mori, 1997; Fujita, Krugman, and Mori, 1999).

Specifically, I consider that an urban system at a long-run equilibrium contains total population $N(\tilde{\mathbf{t}})$ and $r(\tilde{\mathbf{t}})$ urban regions at a given time $\tilde{\mathbf{t}}$. Let $\{1,2,\cdots,i,\cdots,r(\tilde{\mathbf{t}})\} \equiv \mathfrak{R}(\tilde{\mathbf{t}})$ be the set of exiting urban regions, $\left(x_1, x_2, \cdots, x_i, \cdots, x_{r(\tilde{\mathbf{t}})}\right) \equiv \boldsymbol{\mathcal{X}}(\tilde{\mathbf{t}})$ be their location, and $\left(N_1^*(\tilde{\mathbf{t}}), N_2^*(\tilde{\mathbf{t}}), \cdots, N_i^*(\tilde{\mathbf{t}}), \cdots, N_{r(\tilde{\mathbf{t}})}^*(\tilde{\mathbf{t}})\right) \equiv \boldsymbol{\mathcal{N}}^*(\tilde{\mathbf{t}})$ be the corresponding population distribution of the urban system. Then, the system follows the migration adjustment process:

$$\dot{N}_i = \eta N_i[U_i^* - \bar{U}^*]/N(\tilde{\mathbf{t}}) \quad for\ i \in \mathfrak{R} \quad (5.2.2)$$

where η (positive constant) is the adjustment rate, U_i^* is the short-run equilibrium utility level prevailing in region *i*, and $\bar{U}^*$ is the average short-run equilibrium utility level: $\bar{U}^* = \sum_{i=1}^{r(\mathbf{t})} N_i\, U_i^*/N(\tilde{\mathbf{t}})$.

5.3 Evolutionary Path for Long-run Equilibrium

It has been shown in section 4.3.3 that there may exist multiple equilibria regarding the internal urban structure in the short run given fixed transport costs. And I know that the adjustment dynamics applies to a series of short-run equilibria before reaching a long-run equilibrium. As a result, there are a large number of possible paths for the urban system as it evolves towards a long-run equilibrium. Therefore, in order to single out a unique evolutionary path, I need a set of rules to guide path choice when multiple equilibria encountered.

Rule 1 (stability): The chosen short-run equilibrium must be a stable one following the stability definition in section 4.3.2.

Rule 2 (monocentricity first): In the evolutionary process, I assume that city growing follows a natural process, in which one employment center (CBD) first forms surrounded by residential areas and then as the size of the city increases tremendously it allows the formation of sub-centers (SBD) as a way to relieve congestion. Thus, at the beginning of the process, monocentricity has the priority to be chosen as long as it is stable given the set of transport costs. Note that monocentricity here does not necessarily mean that only CBD exists ($\theta = 1$). It could refer to the situation that among a set of polycentric equilibria ($0 < \theta < 1$) the largest θ has the priority to be chosen.

Rule 3 (continuity): I assume urban inertia to prevent catastrophic change of urban structures. That is, as long as population changes gradually, the internal structure of a city (represented by θ) should change continuously so that it follows a smooth evolutionary path. Technically, this means that the path should move to the nearest stable equilibrium θ' from the previous one θ.

Rule 4 (transition): When it becomes impossible to maintain the continuity of the evolutionary path, an adjacent stable equilibrium should be chosen based on the following rules. If there is only one stable equilibrium available, it will be chosen. If there are more than one stable equilibria (denote this set as s), three situations will be considered regarding the transition from θ to θ':

1) If $\theta < \min_{i \in s} \theta_i$, $\theta' = \min_{i \in s} \theta_i$;

2) If $\theta > \max_{i \in s} \theta_i$, $\theta' = \max_{i \in s} \theta_i$;

3) If $\theta_j < \theta < \theta_i$ and $j, i \in s$:

If $\Delta\Pi(\theta) > 0$, $\theta' = \theta_i$;

If $\Delta\Pi(\theta) < 0$, $\theta' = \theta_j$.

5.4 Interplay between Spatial Costs at Intra- and Inter-urban Scales

In this section, I study the impact of changes in spatial costs, the transport costs of goods and activities, on the spatial structures at intra- and inter-urban scales. In other words, I examine how the interplay between spatial costs across scales causes the interdependency of spatial structures in terms of polycentricity/monocentricity across scales. Considering the historical change of transport costs, I assume they follow a decreasing process and study how this process affects the space-economy. Transportation technology or accessibility at the intra-urban scale affects the trading and commuting costs of households and the transaction costs of firms. Inter-urban trade costs are also affected by the change of transportation and communication means. I explore long-run equilibria through numerical simulations based on the rules described in the previous section. The stability of equilibria is tested using the method depicted before (section 5.2). Without losing generality, I employ a racetrack with 8 regions as the inter-urban geography setting. Due to the limit amount of simulations, I cannot cover the whole range for parameters. However, I use adaptive interval strategies (finer interval for equilibria shifting area) to extract major types of equilibria. Each equilibrium can be characterized by two vectors, $\boldsymbol{\theta}(\theta_i^*)$ and $\boldsymbol{\lambda}(\lambda_i^*)$ ($i \in 1,2,3,4,5,6,7,8$), where the former indicates the firm share in CBD (intra-urban structure) and the later represents the population share of the region (inter-urban structure). G is the set $\lambda_i^* > 0$ and $|G|$ is the number of region in the set. There exist five major types summarized as follows:

a) a single monocentric urban region ($\theta_i^* = 1, \lambda_i^* = 1, \lambda_{j\neq i}^* = 0, |G| = 1$), denoted $(M, 0)$;

b) two or more monocentric urban regions ($\theta_i^* = 1, 0 < \lambda_i^* < 1, |G| > 1$), denoted (M, M);

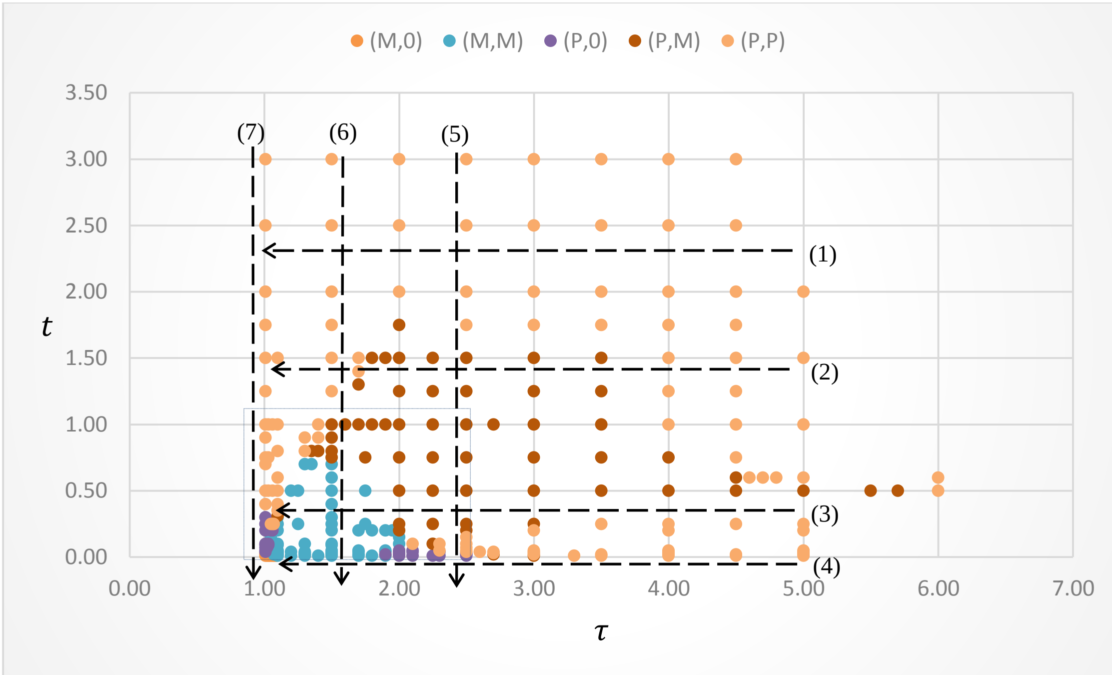
c) a single polycentric urban region ($\theta_i^* < 1, \lambda_i^* = 1, \lambda_{j\neq i}^* = 0, |G| = 1$), denoted $(P, 0)$;

d) two or more polycentric urban regions ($\theta_i^* < 1, 0 < \lambda_i^* < 1, |G| > 1$), denoted (P, P);

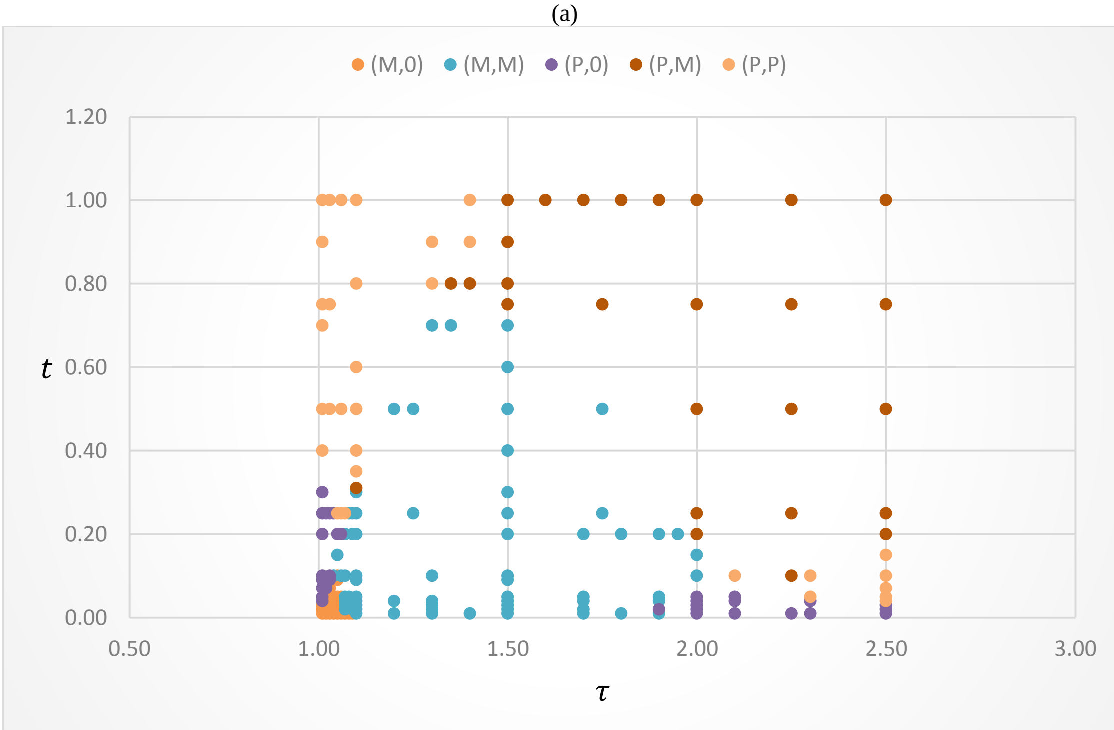
e) mix of monocentric and polycentric urban regions ($\theta_i^* \leq 1, 0 < \lambda_i^* < 1, |G| > 1$), denoted (P, M).

5.4.1 How Intra-urban Spatial Costs Affect Local and Global Spatial Structures

I first focus on the decreasing commuting cost (t) and the intra-urban trade cost (τ) within urban agglomerations and examine their effects on the spatial structures internal to and between urban regions. The five equilibrium types are illustrated in Figure 5.1 based on results assembled from simulations with the parameter setting as: $N = 32, \sigma = 4, f = 1, \mu = 2, \alpha = 2, T = 256$. It shows that the (τ, t) plane is roughly divided by the partitions dominated by certain equilibria, but the partitions are not clear-cut due to the multiplicity of equilibria.



(a)



(b)

Figure 5.2: (a) Equilibrium states in (τ, t) plane from simulations with $N = 32, \sigma = 4, f = 1, \mu = 2, \alpha = 2, T = 256$; (b) zoom-in of the dash box regon in (a)

The Impacts of Intra-urban Trade Costs

Along the τ axis from high to low value in Figure 5.2, four types of paths can be distinguished. First, when t is very high (e.g, $t = 2$), the evolution of equilibria follows a path:

$$(P, P) \rightarrow (P, P) \quad \text{Case (1)},$$

where a polycentric structure prevails. Because the commuting cost is so large that monocentricity is not feasible (too expensive for all households to commute to the single CBD). Externally, high commuting cost alleviates urban cost to the level that agglomeration within a single urban region cannot be afforded, which leads to a dispersion pattern between urban regions. When the internal trade cost τ is very high, each urban region is formed by one CBD and two SBDs with the same size. As τ decreases to middle level, centralization of activities occurs in the CBD because firms trend to concentrate in the major center to enjoy larger home-market effect. As τ continues to decrease to very low level, a re-decentralization to SBDs occurs because firms in SBDs can still access a larger market due to reduced τ but also save on wage bills and land rent. This process resembles the spreading-agglomeration-spreading phenomena in the discussion regarding short-run equilibrium.

Second, when t is at the intermediate level (e.g., $t = 1.5$), the evolution of equilibria follows the path:

$$(P, P) \rightarrow (P, M) \rightarrow (P, P) \quad \text{Case (2)}.$$

When internal trade cost is high, decentralization makes polycentricity the equilibrium urban structure with similar sized CBD and SBDs. High internal trade cost also prevents agglomeration at the inter-urban scale, which is why the dispersion pattern prevails with

8 similar sized urban regions to keep the size of each urban regions as low as possible. When τ decreases to the intermediate level, agglomeration between urban regions is triggered by market-access effect as large urban region grows at the expense of the small regions. In the meanwhile, as small regions decrease in size, the intermediate level of commuting cost makes monocentricity possible for these regions where CBDs regain their dominance and become the only center in the regions. As τ keeps decreasing, decentralization to SBDs occurs in small monocentric urban regions to take advantage of the low internal trade cost. However, for large urban regions decentralization continues since commuting costs are not sufficient low to support monocentricity. At the inter-urban scale, the rise of wages induced by the saving on τ for households in large urban regions is less sufficient to cover their expenses in commuting compared to that for households in small urban regions. As a result, small urban regions re-gain population from large urban regions as a process of dispersion between regions. Finally, when τ falls to a very low level, re-decentralization occurs in all regions no matter their size. The higher commuting cost can be better covered by the gain of sufficiently reduced τ in wages for households in large urban regions. Together with further decreased price index due to very low τ, higher real wages level makes large regions more attractive and lead to agglomerations towards large regions at the inter-urban scale.

Third, when t takes low value (e.g, $t \in [0.2, 0.5]$), the evolution of equilibria follows a path:

$$(P, P) \rightarrow (P, M) \rightarrow (M, M) \rightarrow (P, P) \rightarrow (P, 0) \quad \text{Case (3).}$$

When internal trade cost is very high, t is not sufficient low relative to τ to afford monocentric structure for urban regions. And activities are dispersed across regions to avoid forming large agglomerations so that internal trade and commuting cost can be kept at lower level. As τ decreases, t become sufficient low relative to τ to allow monocentric structure to occur in smaller regions as that in the second case. As τ keeps decreasing, more regions are able to form monocentric structure and this centralization process continues until τ reach the level that can afford all regions to become monocentric. As τ further decreases to low level, the effect starts causing agglomerations of fewer large monocentric urban regions at the inter-urban scale because sufficiently reduced trade cost and low commuting cost allow large agglomerations to be monocentric. As before, once τ reaches very low level, it triggers the re-decentralization within each urban regions. And because the polycentric structure can sustain larger population while maintaining the access to large market, further agglomeration between urban regions also prevails. This process continues until all population agglomerates in a single polycentric region when τ reaches an extremely low value.

Fourth, when t is extremely low (e.g., $t = 0.01$), the evolution of equilibria follows a path:

$$(P,P) \rightarrow \begin{cases} (P,M) \rightarrow (P,0) \rightarrow (M,0) & 4.1 \\ (M,M) & 4.2 \end{cases} \quad \text{Case (4).}$$

When internal trade cost is very high, polycentricity is the preferred structure in order to save on urban costs and to reduce the upper bound of price index. However, at the inter-urban scale, population tends to agglomerate in few urban regions instead of spreading across space because urban costs are sufficiently reduced by the extremely low commuting cost. As τ decreases, low commuting cost starts to affect the urban internal structure making monocentricity possible for small regions for the same reason as in case 3, which results in a few agglomerations as a mix of large polycentric regions and small monocentric regions. As τ continues decreasing to the intermediate level, because of the asymmetric distribution of population between these regions, further agglomeration proceeds until forming a fully agglomerated polycentric region. When τ reaches very low level, together with very low commuting cost it makes monocentricity sustainable in one fully agglomerated region. Therefore, a single mega-city with a huge CBD emerges. Note that there exists a bifurcating path (case 4.2) where (P, P) move towards (M, M) as τ decrease from high to low. This is because sufficiently reduced τ enable low commuting cost to take effect on the centralization of activities within urban regions, which otherwise can be less effective to promote agglomerations between regions due to the well-balanced symmetric pattern. Taking which path depends on the degree of symmetricity of the equilibrium (P, P) at the bifurcation point, which further depends on the initial distribution of population. When the initial population distribution is fully dispersed and symmetric, the path goes with case 4.2; otherwise, it goes with 4.1. This well demonstrates the path dependency in the evolution process of urban systems.

The Impacts of Commuting Costs

Regarding commuting costs falling from high to low value, three types of paths can be identified along the t axis. First, when the internal trade cost rate τ is high (e.g., $\tau = 2\ or\ 2.5$), the evolution of equilibria follows a path:

$$(P, P) \rightarrow (P, M) \rightarrow \begin{cases} (P, P) \rightarrow (P, 0) & 5.1 \\ (M, M) & 5.2 \end{cases} \quad \text{Case (5).}$$

When t is high, large τ leads to polycentric urban equilibrium. They also contribute to high urban costs and price index (consumption costs) that lead to dispersion at the inter-urban scale. Therefore, (P, P) equilibrium exhibits a near symmetric pattern. As t decreases, falling urban costs create potential for agglomeration across regions. And the perturbation of population distribution caused by regular migration may take advantage of this potential to enlarge the relative imbalance of activities between large and small regions. This imbalance can trigger further agglomeration from small regions to large ones. As small regions shrink in size, lower commuting cost makes monocentricity stable equilibrium. Thus, a mix of polycentric large urban regions and monocentric small one emerge as equilibrium. As t keeps decreasing, the path reaches a bifurcation point towards two different paths (case 5.1 and 5.2). Case 5.1 features further agglomeration in that falling urban costs accelerate the growth of large regions by attracting population from small ones until all small monocentric urban regions vanish. Therefore, it exhibits an asymmetric partial agglomerated (P, P) equilibrium different from the initial one. When t falls at the low level, agglomeration between large polycentric regions continues until a single fully agglomerated polycentric region emerges $(P, 0)$. The second path (case 5.2) is characterized by a centralization process of internal urban structure while re-dispersion prevails from larger regions to small ones. The centralization caused by falling commuting cost allows monocentricity affordable in larger regions and leads to a symmetric configuration of monocentric urban regions. Then, the question comes in what condition the path chooses cases 5.1 or 5.2. Simulation testing reveals that, in equilibrium (P, M), when the number of large polycentric regions is less than half of the total regions, it tends to bifurcate to case 5.1. Otherwise, the path more probably bifurcates to case 5.2. This reflects that the degree of symmetricity of population distribution in previous phase of equilibrium influences the outcome of the bifurcating paths for the following phase, which again demonstrates the path dependency of the evolving urban systems.

Second, when τ is at the intermediate level (e.g., $\tau = 1.5$), the evolution of equilibria follows a path:

$$(P,P) \rightarrow (P,P)/(P,M) \rightarrow (M,M) \text{ Case (6).}$$

Seemingly similar as case 5.2, this path proceeds with a more progressive transition and it does not involve partial and fully agglomeration and bifurcation. When t is high, activities are decentralized within regions and dispersed across regions for the same reason as that in case 5. As t decreases, reduced urban costs take effects on both centralization forces within regions and agglomeration forces between regions. The effects are reinforced by the intermediate level of internal trade cost, which leads to the transition from symmetric (P,P) to $(P,P)/(P,M)$ involving both partially agglomerated equilibrium of few large polycentric regions and dispensed equilibrium of large polycentric regions together with small monocentric regions. The former results from stronger external agglomeration forces, while the latter is caused by stronger internal centralization forces. Once t falls to a very low level, it makes monocentricity affordable for all regions no matter their size manifesting as dispersed (symmetric) or partially agglomerated (asymmetric) equilibria of monocentric urban regions. Note that a fully agglomerated monocentric region is not equilibrium because internal trade cost is not sufficient low.

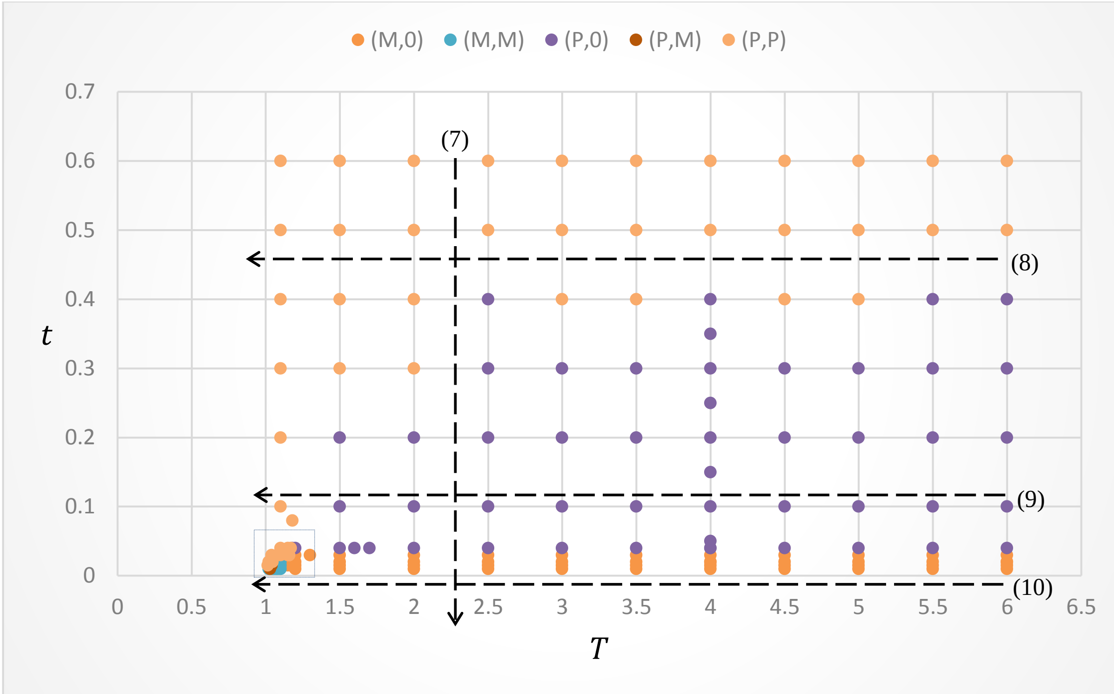
Third, when τ is at the very low level (e.g., $\tau = 1.01$), the evolution of equilibria follows a path:

$$(P,P) \rightarrow (P,0) \rightarrow (M,0) \text{ Case (7).}$$

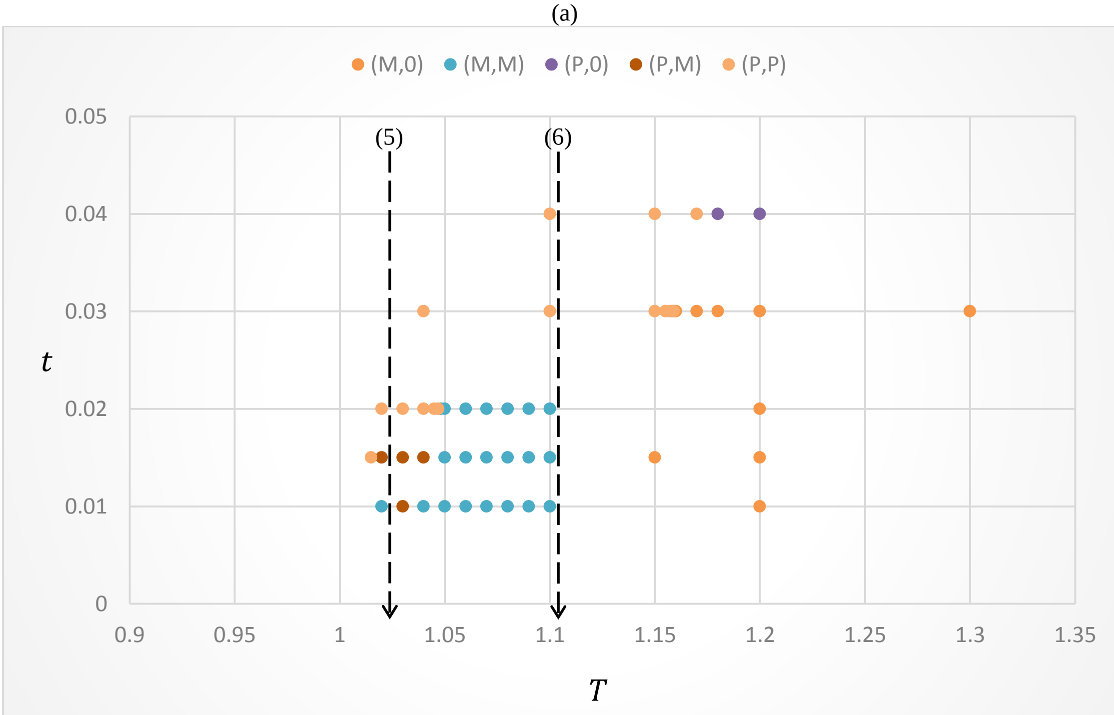
When the commuting cost is high, urban costs are so large that activities are decentralized within regions and dispersed between regions. As t decreases to the intermediate level, with reduced urban costs very low internal trade cost starts to strengthen the agglomeration forces between regions compared to relative high inter-urban trade cost. This leads to one fully agglomerated polycentric region since the commuting cost is not sufficient low to sustain a monocentric one. Once t falls to an extremely low level, centralization becomes the dominant trend, and as a result, a single fully agglomerated monocentric urban region emerges.

5.4.2 How Inter-urban Spatial Costs Affect Local and Global Spatial Structures

In this section, I examine the interaction between the commuting cost (t) and the inter-urban trade cost (T) and their effects on both the external and internal spatial structure of urban regions. Given that the intra-urban trade cost (τ) are constant, the five equilibrium types are all covered and demonstrated in Figure 5.3. The results are from simulations with the parameter setting as: $N = 32, \sigma = 4, f = 1, \mu = 2, \alpha = 2, \tau = 1.01$. Again, the (T, t) plane shows rough partitions dominated by certain equilibria, but the partitions are not clear-cut due to the existence of multiplicity of equilibria.



(a)



(b)

Figure 5.3: (a) Equilibrium states in (T, t) plane from simulations with $N = 32, \sigma = 4, f = 1, \mu = 2, \alpha = 2, \tau = 1.01$; (b) zoom-in of the dash box regon in (a)

The Impacts of Inter-urban Trade Costs

Along the T axis from high to low value in Figure 5.3, three types of paths can be identified. First, when t takes high values (e.g., $t > 0.4$), the evolution of equilibria follows a path:

$$(P, P) \rightarrow (P, P) \text{ Case (8),}$$

where polycentric equilibrium prevails because monocentricity cannot be sustained with high commuting costs and inexpensive internal trade. And when T is high, it creates potential for agglomeration at the inter-urban scale. However, full agglomeration into a single urban region is not possible with large urban costs due to the high commuting cost, which leads to a partial agglomerated equilibrium with few polycentric regions. As T falls to the low level, the internal polycentricity remains while dispersion becomes a stronger force as inter-regional trading is promoted by cheaper costs. This process continues until full dispersion emerges.

Second, when t is at the intermediate level (e.g., $t \in [0.04, 0.4]$), the evolution of equilibria follows a path:

$$(P, 0) \rightarrow (P, P) \text{ Case (9),}$$

where the polycentricity within regions prevails because the commuting cost is not low enough to sustain a monocentric structure, while the very low internal trade cost tends to encourage decentralization within regions. When T is large compared to τ, agglomeration becomes the dominant forces between regions and is able to form a single polycentric region supported by lower commuting costs comparing to case 7. As T decreases, agglomeration is replaced by dispersion at the inter-urban scale, which eventually leads to a fully dispersed equilibrium with polycentric regions.

The third type involves a group of paths that follow the same mechanism, when t takes very low value (e.g., $t \in [0.01, 0.03]$). The evolution paths of equilibria are as follows:

$$\begin{cases} (M,0) \rightarrow (M,M) \rightarrow (P,M) \rightarrow (P,P) & 10.1 \\ (M,0) \rightarrow (M,M) \rightarrow (P,P) & 10.2 \\ (M,0) \rightarrow (P,P) & 10.3 \end{cases} \quad \text{Case (10).}$$

Due to very high inter-urban trade cost, activities agglomerate in one region to avoid inter-regional trading. And because of the extremely low commuting cost, monocentricity is stable equilibrium. All three paths follow a common trend characterized by dispersion across regions and decentralization within regions. From case 10.1 to 10.3, t takes larger values. I start with case 10.1. Once T decreases, in order to take advantage of the reduced trade cost between regions, activities disperse to the nearest region and form the second monocentric urban region with a similar size as the origin. This leads to a partial dispersed equilibrium (M, M). As T decreases further, dispersion continues with new urban regions emerging at the expense of two large monocentric regions. However, as the new regions grow from small size, internal decentralization starts to take effects because low commuting and internal trade costs together hit the threshold of stable polycentricity for their size. Since larger regions require lower internal trade costs and higher commuting costs than those for the smaller regions to trigger decentralization, internal trade costs sufficient low (or commuting costs sufficient high) for small regions to become polycentric cannot break the monocentricity equilibrium for larger regions. It is why the equilibrium of two large monocentric regions and two small polycentric ones become stable before it evolves to four regions with the same size. As T continues to decrease to the low level, inter-regional trading is promoted by further dispersion from large regions to small ones. This trend continues until four regions reach the same size which is small enough to allow the polycentricity enabled for these regions. In case 10.2, because commuting cost is larger, the break point of monocentricity can be reached faster by skipping the equilibrium (P, M) when the dispersion from (M, M) to (P, P) forming more regions with a smaller size. Case 10.3 skips more intermediate equilibrium phases with even higher commuting cost rate for the same reason.

The Impacts of Commuting Costs

I look at again the impact of commuting costs with respect to the change of inter-urban trade costs. Three types of paths can be distinguished along the t axis from high to low values. First, when T is large relative to the internal trade cost τ (e.g., $T \in [1.5, 6]$), the evolution of equilibria follows a path that is the same as case 7:

$$(P, P) \to (P, 0) \to (M, 0).$$

Case 7 is the path when τ is very low relative to T, which is an equivalent condition to the one here. When t is high, large urban costs lead to both decentralization within regions and dispersion between regions. As t falls, reduced urban costs promote agglomeration across regions and ultimately lead to a fully agglomerated polycentric urban region. As t continues to decrease to a very low level, together with very low τ it enables monocentricity within a fully agglomerated region. In sum, falling commuting costs first strengthen the agglomeration force externally then the centralization force internally. It takes this order because a very low internal trade cost rate is given.

Second, when T takes intermediate values (e.g., $T = 1.1$), the evolution of equilibria follows a path that is similar as case 6:

$$(P, P) \to (M, M).$$

Case 6 has a condition of τ at the intermediate level, it is equivalent to the condition here according to the relative difference between τ and T. The initial equilibrium of full dispersion with polycentric regions results from the large urban costs due to high commuting costs. With falling t, agglomeration forces take effects first, and then it is followed by centralization forces as the same order as the previous case. At the inter-urban scale, a partial agglomeration pattern emerges with regions forming monocentric structures when t decreases to the low level.

Third, when T is at the low level (e.g., $T = 1.02$), which is relatively larger than but close to the low value of τ, the evolution of equilibria follows a path that is the same as case 5.2:

$$(P, P) \rightarrow (P, M) \rightarrow (M, M).$$

Case 5.2 is conditioned on the high internal trade cost level with high level of T, which is equivalent to the condition here according to the relativity between τ and T. The initial dispersed polycentric urban regions emerge as stable equilibrium for the same reason as before. As t decreases, agglomeration prevails and creates imbalance of size between regions. During the process of certain regions growing at expense of others, polycentric regions with larger size become sustainable for monocentricity enabled by the set of very low for internal trade cost and modest low commuting cost. This leads to a mix of larger monocentric regions and smaller polycentric regions as stable equilibrium. Once t falls to extremely low level, monocentricity prevails within regions with partial dispersion across regions.

5.4.3 Summary

The simulation results unveil three main findings. First, I build a multiscalar urban systems model to take into account the interdependency of spatial structure across scales by incorporating urban costs (land rent, commuting), intra-urban trade costs and inter-urban trade cost. This has been illustrated by the simulation results. Second, specifically the evolutionary paths of equilibrium transition regarding intra- and inter-urban spatial structures are characterized by 10 major cases extracted from the simulation results. These equilibrium transitions indicating the change and interdependency of spatial structures (in terms of polycentricity/monocentricity) at different scales are discovered under the exogenous change of intra- and inter-urban spatial costs, respectively. Hence, by examining the impact of these spatial costs on local and global organization of the space-economy, I have successfully established the causal relations between the interplay of market linkages and spatial costs across scales and the interdependency of spatial structures across scales. Third, the interdependency of spatial structures across scales proved by our model highlights the importance to policy making in the context of urban regions. For example, our results (case 3) confirm the idea that the polycentric structure fosters further agglomeration. Its policy implication suggests advocating the development of secondary business centers.

CHAPTER 6: GEOGRAPHIC MODELS OF POLYCENTRIC URBAN SYSTEMS

The proposed theoretical models in Chapters 4 and 5, following the convention of economic modeling, rely on a very abstract conceptualization of the geographic space, i.e., linear or circular space. This simplification indeed helps to single out the "second nature" forces at work that endogenously determine the location of spatial agglomerations; it may however reduce the empirical validity. The second objective of this research is thus to extend the proposed theoretical models with a true geographic dimension that is characterized by 1) a two-dimensional space and 2) heterogeneous and asymmetric places. In contrast to the aggregate style of the theoretical economic models (i.e., representative agents), the geographic model proposed here is of a flavor of disaggregation (e.g., heterogeneity with agents' location choices). In terms of accounting for system dynamics, the former commonly adopts differential models with a continuous representation of time, while the latter usually employs difference equations with a discrete one. Though there have been arguments that differential models are better suited to describe aggregate processes in regional economic systems (Thill and Wheeler, 1995), the discrete representation of time is more appropriate in this study given the disaggregated nature of our model.

6.1 Discrete Space and Time

First, for internal urban space (e.g., a city), I consider a two-dimensional discrete space where there exists a finite number of locations. The distances between these locations are nonhomogeneous. Each location has a capacity of land units that may be developed, in terms of being occupied by households or firms, or otherwise be undeveloped. Here, I still assume land quality is homogenous across locations. Consistent with the simplification made to the theoretical models (Section 4.3.2), I further assume fixed land consumption by each household and each firm such that $S_h = 1, S_f = 0$. This discrete space is best represented by a weighted graph with variable size of vertices (Figure 6.1). Each vertex represents a location with a capacity indicated by its size. The edge between two locations is a generalization of transportation networks between them. The weight of an edge represents the distance or travel cost via the generalized transportation link between two locations. If there is no edge (direct link) between two locations, distance (or travel cost) between them can be calculated by finding indirect links that comprise the shortest path between the two locations. Second, at the inter-urban scale, cities represented by individual subgraphs comprise the whole urban-regions by connecting to each other via higher level transportation networks such as interstate highways (Figure 6.2). The networks between two cities can be again generalized as a link between two subgraphs. For simplification, I assume that all locations in one city have the same distance or travel cost to all locations in another city. In other words, the distance between cities is independent from cities' internal extent. If cities (represented by a subgraph) are viewed as vertices at the inter-urban scale, the whole urban-regions can be still represented by a weighted graph with city vertices having variable size. In this sense, the entire space structure can be easily understood as a hierarchical weighted graph. Given this discrete space setting, recall that continuous urban models (Chapter 4) have an internal urban space based on a continuum of locations. The discrete space adopted here necessitates the transformation of models based on continuous space into models based on discrete space. Besides, this geographic model is subject to all other assumptions made to the specification of the theoretical model.

The behavior of households and firms is modeled in the context of discrete time intervals represented by model iterations. Within one iteration, households choose residential locations and workplace while firms choose business locations. It may take any number of iterations to achieve a short-run equilibrium in each urban region or city (conditions for short-run equilibrium are defined in Section 4.3). The inequality of utility levels achieved at short-run equilibria in different urban regions produces inter-regional migration dynamics as an adjustment process that entails consecutive short-run equilibria in each region, and eventually leads to a long-run equilibrium where utility levels are all equal in every region (defined in Section 5.2). Thus, it may take one or more short-run periods to reach a long-run equilibrium. Because of the explicit discrete treatment of time, this modeling approach is inherently dynamic and necessitates the transformation of the way directly calculating equilibrium solutions in continuous urban models into the way iteratively converging to equilibrium conditions through discrete model algorithms.

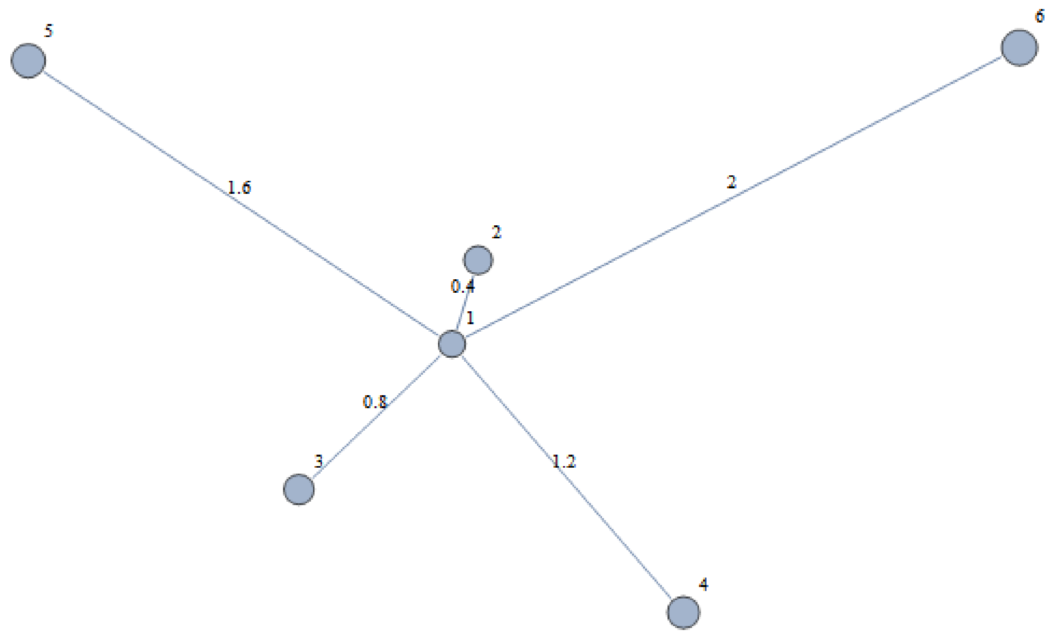


Figure 6.1: Graph representation of the discrete space for urban models

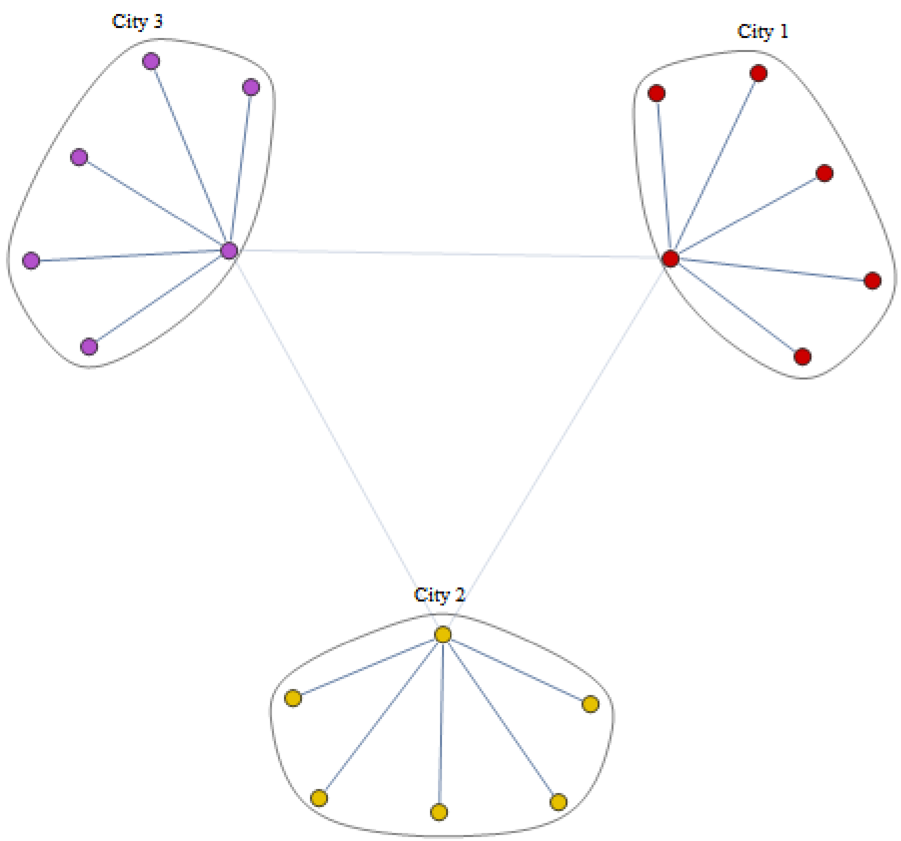


Figure 6.2: Graph representation of the discrete space for urban systems models

6.2 Agent-based Location Choice

For location choice problems, achievement of general (Walrasian) equilibrium requires the simultaneous adjustment of all locations and prices, meaning that perfect information for all agents must be assumed, while in reality bounded rationality of agents with imperfect information is usually the case. Consequently, the former entails global interaction, while the latter leads to local interaction. Due to this limitation, an alternative approach to market equilibrium modeling for location choice needs to be employed in compliance with the dynamic framework of the proposed model.

In order to better represent and study complex economic systems, the application of agent-based modeling to economics and market interactions has formed a domain of research on Agent-based Computational Economics (ACE; Tesfatsion, 2006). According to Tesfatsion (2006, p6), ACE is "the computational study of economic processes modeled as dynamic systems of interacting agents". That is, viewing economic systems as complex adaptive systems that hold two fundamental properties: 1) comprising interacting units, an economic system as a whole cannot be understood by simply aggregating the behaviors of its parts, individual economic agents; 2) system-wide emergent properties arise unexpectedly from the interactions of individual agents. Specifically, in contrast to the Walrasian auctioneer mechanism in a traditional economic model that ensures market clearing in a top-down manner, autonomous agents in ACE explicitly enable decentralized strategic interactions in various forms of economic processes (e.g., production, pricing, trade) giving rise to macroscopic phenomena that may manifest a globally stable equilibrium or only exhibit the dynamics of the system playing out over time. In other words, instead of focusing on the equilibrium state of an economic system, the essence of ACE is to observe and examine if certain forms of equilibria develop over time through computational simulations. As a result, the acquisition of the entire process of system dynamics aims at a holistic understanding of the phase transition of a system, i.e., all possible equilibria, tipping point, and catastrophic change. The agent-based approach I adopt here is in line with the fundamental idea of ACE.

Given the discrete setting of space and time, an agent-based approach is employed to model location choices made by individual households and firms to maximize their utility and profits respectively in the city equilibrium (short-run). Households choose their residential location and workplace with respect to given land rents and wages. In each workplace, the prevailing wages are determined by a bidding process in which firms compete for workers by offering them higher wages until no firm can profitably enter the market. Firms choose their business locations given the prevailing wages and local labor pool of workers in different workplaces. While the product market still concerns continuous choices of bundles of goods, both land and labor markets deal with discrete choices due to the discreteness of locations and the disaggregation of agents. Both types of discrete choices can be modeled via a bid-auction approach. In the land market, a bid-rent model is employed to determine the location choice for residence and the rent simultaneously based on the theoretical models in Chapter 4. In the labor market, the workplace choice problem is modeled by an equivalent approach where firms bid for workers under zero-profit condition. In such a process, the workplace and business location decisions and the wages can be determined simultaneously.

6.2.1 Residential Choice

Since Alonso (1964)'s bid-rent function, the urban land market has been recognized as an auction market, where under a given utility level households bid their willingness-to-pay (WTP) for a land parcel that is allocated to the highest bidder. Regarding the dual characteristics of land being a commodity (exchanged via price mechanism) and completely immobile (uniquely associated with a location), this approach simultaneously determines the two aspects, that is the land rent and the location choice. The bid-rent function can be derived from a classical problem of constrained consumer utility maximization (a general form of equations 4.1.1.4 and 4.1.1.5):

$$\max_{q,i} U(q, z_i) \quad (6.2.1)$$

$$s.t. pq + R_i \leq W$$

where consumer utility depends on a vector of continuous goods q and a location i, described by a set of attributes z_i. Its maximization is subject to the spending on q with price p and land rent R_i for location i constrained by the available income W. If I assume homogenous locations as in models of Chapter 4, z_i is the same for all locations and can be dropped from (6.2.1) (eliminate first nature).

The maximization problem becomes unconstrained by rewriting it as an indirect utility function:

$$\max_{i} V(p, W - R_i, z_i) \quad (6.2.2)$$

Thus, the bid-rent function can be derived by inverting the indirect utility in the rent variable conditional on a fixed level of maximum utility:

$$\Psi_{hi} = W_h - V_h^{-1}(p, U^*, z_i) \quad (6.2.3)$$

which can be understood as the maximum rent a consumer h is willing to pay for a location i, while maintaining a fixed level utility level U^*(Fujita, 1989). Martinez (2000) finds that it is possible to decompose (6.2.3) into components related to the consumer and the location respectively by assuming a quasi-linear utility structure for $U(q, z_i)$ in (6.2.1):

$$\Psi_{hi} = B_h^D(U^*) + B_h^L(z_i) \quad (6.2.4)$$

where $B_h^D(U^*)$ is related to the decision maker and $B_h^L(z_i)$ is related to the location. If only considering homogeneous locations, $B_h^L(z_i)$ can be dropped in (6.2.4) and the consumer's bid-rent only depends on the income and the component related to the consumer utility.

6.2.2 Simultaneous Workplace, Industry and Firm Location Choice

In this section, I consider the household workplace choices and firm location choices under the context multiple industries. That is, a household jointly chooses a workplace and an occupation in one industry, while a firm jointly chooses a business location and an industry in which it does business. I employ a bid-auction approach to model these decision-making processes by agents.

Given its residential location, a utility-maximizing household jointly chooses its workplace and industry with respect to the equilibrium wages for industries in workplaces net of the commuting cost (denoted as net wage $\widetilde{W}_s^k$) from its residential location to workplaces. From this perspective, a household at location s needs to examine the relative difference of the net wage for each industry from every location c and the average wage between industries at location s, where the worker does not need to commute. This difference of net wages can be expressed as:

$$\Delta\widetilde{W}_{s,c}^{k} = \widetilde{W}_{c}^{k} - \overline{W}_{s} = \left[W_{c}^{k} - t \cdot l_{s,c} \cdot F(s,c)\right] - \overline{W}_{s} \quad (6.2.5)$$

where $\widetilde{W}_{c}^{k}$ is the net wage for industry k at workplace c, W_{c}^{k} and W_{s}^{k} are the wages for industry k at workplace c and s respectively, $l_{s,c}$ is the distance between s and c, $\overline{W}_{s} = \sum_{k} W_{s}^{k}/K$ is the average wage at workplace s, K is the number of industries, and $F(s,c)$ is the commuting flow from s to workplace c. Here, I assume that the commuting costs depend on the commuting distance and the volume of commuting flow that is used as a proxy of the congestion level. Treating $\Delta\widetilde{W}_{s,c}^{k}$ as bids from firms belonging to different industries at different workplaces, a household evaluates the set of bids and makes a decision to maximize its utility. This process can be modeled by a discrete choice model. In city equilibrium (short-run), households at location s commuting to different workplaces and working in different industries have the same utility level, which means their net wages $\widetilde{W}_{c}^{k}$ are equal:

$$W_{c}^{k} - \overline{W}_{s} = t \cdot l_{s,c} \cdot F(s,c) \quad (6.2.6)$$

where $W_{s}^{k} = W_{s}^{k\prime},\ k, k' \in K \,\&\, k \neq k'$ when $c = s$.

From a firm's perspective, it jointly chooses its business location and the industry it belongs to with respect to the profit potential of the industry in that location. The household workplace-industry choices and the firm location-industry choices are linked by the bidding process in which firms compete for workers by offering them higher wages until earning zero profit. This is why firms in fact choose their business location and industry with the highest profit potential rather than the absolute profit level. In this sense, for a firm at location s to determine where to relocate and what industry to enter, it needs to examine the relative difference of profit potential for each industry at every location c and the average profit level between industries at its current location s. Following (4.3.2.12) and using (6.2.6), it can be expressed as:

$$\Delta\Pi_{s,c}^{k} = \Pi_{c}^{k} - \overline{\Pi}_{s} = \left[W_{c}^{k^*} - \overline{W}_{s}^{*} - t \cdot l_{s,c} \cdot F(s,c)\right] f \quad (6.2.7)$$

where $W_{c}^{k^*}$ is the wage under zero-profit condition for industry k at location c, $\overline{W}_{s}^{*} = \sum_{k} W_{s}^{*}/K$ is the average wage under zero-profit condition between industries at location s, and f is the fixed labor requirement which can be set to $f = 1$ without loss of generality. By examining $\Delta\Pi_{s,c}^{k}$ for the set of location-industry combinations, a firm determines where to relocate and what industry to enter. This process can also be modeled by a discrete choice model. Note that equation (6.2.5) is equal to equation (6.2.7) if zero-profit condition wages are used instead. This indicates that the household workplace-industry choices and the firm location-industry choices are a simultaneous process given the model setting.

6.2.3 An Equilibrium Model of Agent-based Location Choice

I propose an agent-based approach to modeling the dynamics of transition between consecutive intra-urban (short-run) equilibria, that gives rise to an inter-urban (long-run) equilibrium of location decisions made by disaggregate agents, and the evolution of a polycentric urban system through a series of long-run equilibria. This approach takes into account the agglomeration and dispersion forces taking place in market interactions (Chapter 3) and considers location decisions of agents in land and labor markets with a realistic geographic dimension. Instead of solving equilibrium problems, it aims to accomplish these through iterative simulations carried out by an agent-based model that results in outcome states converging to the equilibrium solutions obtained from the theoretical models (Chapters 4 and 5).

The proposed approach is based on the following assumptions:

- Once a household decides to relocate within an urban region, it potentially presents a chance for it to change its workplace under the assumption of zero cost of relocating and job searching. Therefore, a relocating household simultaneously joins the bidder pool of the local land market and the labor pool of the local labor market and participates in the actions of both markets in order to maximize its utility.
- Firms relocate to locations with higher market potential in order to maximize profits. The product market equilibrium is implied in that firms exploit profits by fulfilling the demand with production.
- The supply of land is exogenously determined and does not necessarily meet demand perfectly.
- Firms are free to enter and exit until all workers are employed and all firms earn zero profit.
- Land and labor are transacted in auctions. All active agents (households and firms look for vacant land for residence and firms look for workers without jobs) are potential bidder for all units (land or labor) available for auction in the corresponding markets.
- Agents do not have access to perfect information on the conditions of the markets (land or labor); they can only infer them from the prices (rent or wage) they observe in the previous time. Before bidding they adjust their expectations according to this information.

- Auctions take place simultaneously; the highest bidder gets the commodity. Prices are determined as the expected maximum bid. If an agent is the best bidder for more than one commodity in the market, it chooses the one that provides maximum utility/profit surplus.

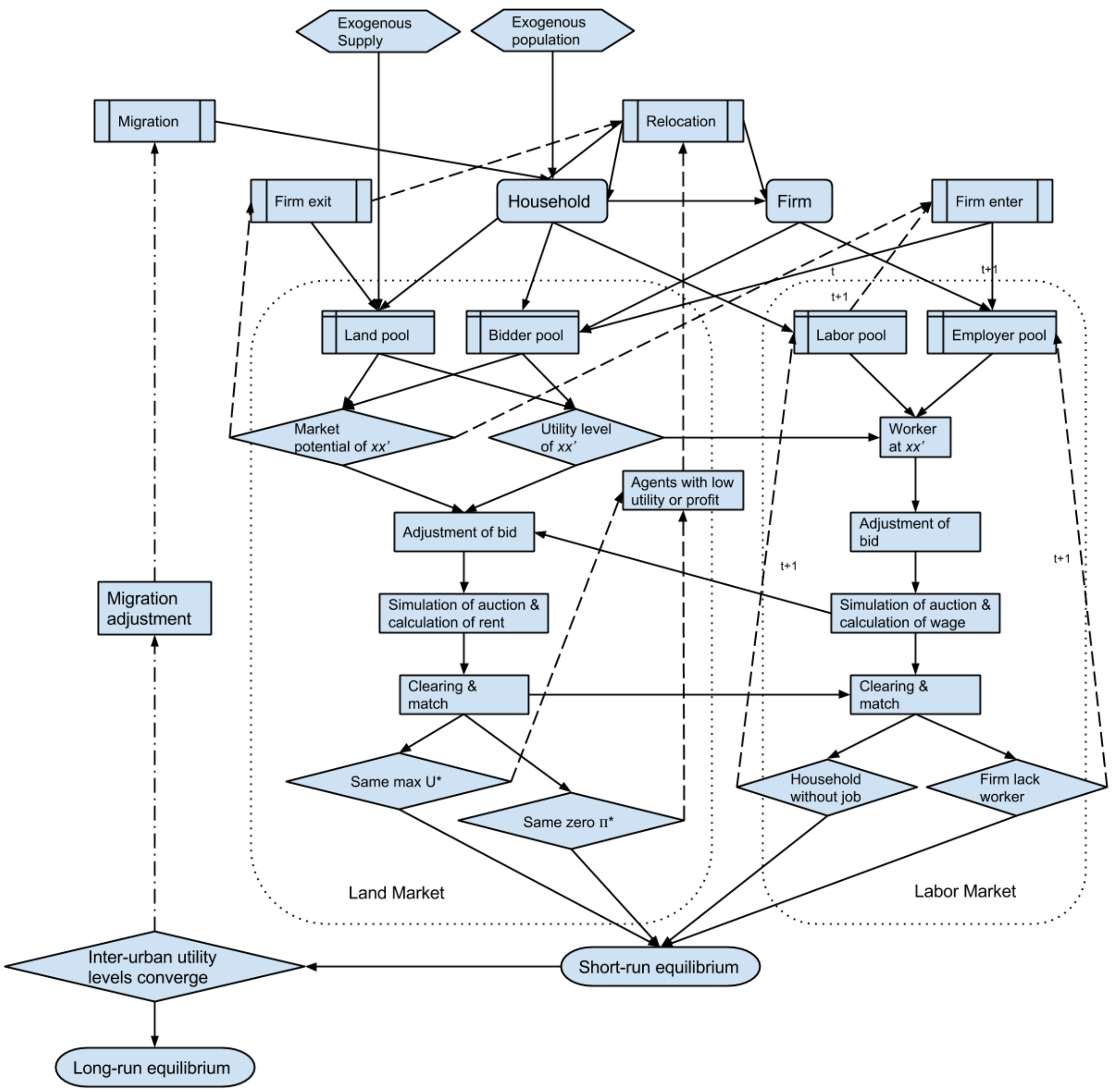


Figure 6.3: Algorithm of agent-based location choice with bid-action land and labor markets

Figure 6.3 describes the algorithmic procedures of the proposed agent-based model. The land supply and the population are exogenous to the model. Clearings of land, labor, and product markets are interdependent in the process towards a short-run equilibrium at the intra-urban scale. Within each urban region, households and firms with a low level of utility or profit prompt relocation to improve their own situation. Relocating agents vacate their residence land parcels, which become available in the local land pool. Relocating households join both the bidder pool of local land market and the labor pool of local labor market in order to optimize their residential location and workplace at the same time. Due to quitting of relocating households from their current employers, these firms become short of workers and thus join the employer pool of the local labor market as well. Relocating firms join the bidder pool of the local land market and seek locations with a higher level of market potential. However, households who work for these firms are assumed to stay put and continue being employed in the current period, even though the relocation of firms may consequently increase their commuting costs and cause them to relocate in the future. On the other hand, firms freely enter and exit the local product market in response to the market potential. Upon their entry, firms join the bidder pool of the local land market immediately but only join the employer pool of the local labor market in the following time period. Once firms exit, their workers become unemployed and thus join the local labor pool in order to find another job.

For each land parcel in their choice sets, relocating agents adjust their WTP for the rent based on observed prices previously and offer bids. For a household, its adjustment of bid rent is conditioned on the outcome of its auction in the labor market. That is, the maximum wage net of commuting cost to the winner firm out of the bidders in that auction. After all bidders adjust their bids for every land parcel in their choice sets, auctions determine the best bidder for each auctioned land parcel and its expected price.

Finally, by resolving conflicts such as a best bidder for multiple locations or for no location, a clearing mechanism of the land market matches land parcels with the winner households, which further invokes the clearing mechanism of the labor market that matches households with winner firms. This sub-process repeats until no household can improve its utility level compared to others' and no firm earns a non-zero profit, where a short-run equilibrium is reached at an intra-urban level. Once each urban region achieves its own short-run equilibrium, a migration adjustment is applied at the inter-urban scale such that households respond to the inter-regional difference of utility level (short-run equilibrium) and migrate to urban regions with higher utility level. Within each urban region, immigrant households look for residential locations and workplaces at the same time acting similarly as relocating households. It is here where the sub-process follows to reach a short-run equilibrium at each urban region. This super-process repeats until all regions have the same utility level, where a long-run equilibrium is reached across regions.

6.3 Demonstrations of Geographic Models

In this section, I demonstrate the constructed geographic model in a two-dimensional space specified in Section 6.1. The geographic model is able to accommodate any number of sectors. Here, I only use two sectors for demonstration purpose. First, I simulate a single urban model to confirm that it behaves consistently with the theoretical model in Section 4.2. Second, I illustrate the urban systems model comprising 8 urban models.

6.3.1 Two-dimensional Two-sector Urban Model

I simulate a two-dimensional two-sector urban model with both final and intermediate goods on the space setting given by Figure 6.1. Evidently, location 1 is the central location. The distances between locations in this space are detailed in Table 6.1. Two sectors with distinct degrees of differentiation are used to represent sectors with more differentiated goods and less differential goods, respectively (e.g., service sector and manufacturing sector). The sector with more differentiated goods has lower σ value and less competition between goods. I assume that sector 1 has more differentiated goods ($\sigma_1 = 4$) while sector 2 has less differentiated goods ($\sigma_2 = 8$). The demonstration tests one scenario that features the exogenous change of intra-urban trade costs. As the trade costs (τ) decline, the city equilibrium (short-run) is examined in terms of the change of the share of each industry at each location. The simulations are configured with $\alpha = 2, \mu = 2, f = 1, n = 4, t = 1$.

The simulation results are compiled in Tables 6.2 and 6.3. Table 6.2 shows that the concentration of sector 1 firms at location 1 first increases as trade costs decrease, but it start falling when τ reaches around 1.49 and the decrease continues until the lowest level of τ. Note that when $\tau = 1.49$, there is a full centralization of sector 1 at location 1, while when $\tau = 1.03$ a decentralization prevails with firms moving out of location 1 to other locations. This clearly depicts a spreading-agglomeration-spreading process, which is consistent with our theoretical findings. Compared to sector 1 firms, sector 2 firms react to the change of trade costs slowly. Table 6.3 shows that when τ goes from 4 to 2.15, the share of sector 2 only slight changes. Then, it shows a jump from 3% to 22% when τ increases from 1.49 to 1.26. This is actually a smoother change once the finer intervals are further tested (Table 6.4). It simply reflects the threshold where sector 2 starts reacting to changes. This share increase continues until $\tau = 1.13$, where sector 2 firms fully concentrate at the central location 1. After it hits the peak, sector 2 share starts falling when τ goes to 1.03. This change follows the decentralization-centralization-decentralization process similar to that for sector 1. However, this process is triggered for sector 2 until τ decreases to a lower level than that for sector 1. In other words, sector 1 firms are more sensitive to trade cost change than sector 2 firms. As trade costs decrease a small amount, sector 1 firms are able to take advantage of this change by moving to locations with larger local markets (higher concentration of firms and households) because they face less competition by selling more differentiated goods. In the meanwhile, they can easily supply consumers at other locations because trade costs are lower. In contrast, sector 2 firms selling less differentiated goods remain decentralized to avoid tough competition. When trade costs decrease to lower level, sector 2 firms start reacting to this change by moving to locations having good market access. However, these lower level costs are sufficient to trigger the re-dispersion of sector 1 firms in order to enjoy lower costs and competition and at the same time have access to the large market from remote locations. This explains why an interesting case occurs as $\tau = 1.03$ where sector 1 is dispersed while sector 2 concentrates at the central location.

Table 6.1 Distance matrix for the locations in the space setting in Figure 6.1

Location	**1**	**2**	**3**	**4**	**5**	**6**
1	0	0.4	0.8	1.2	1.6	2
2	0.4	0	1.2	1.6	2	2.4
3	0.8	1.2	0	2	2.4	2.8
4	1.2	1.6	2	0	2.8	3.2
5	1.6	2	2.4	2.8	0	3.6
6	2	2.4	2.8	3.2	3.6	0

Table 6.2 The share of sector 1 at all locations in equilibrium given changing intra-urban trade costs

τ	**Location**					
	1	**2**	**3**	**4**	**5**	**6**
4.00	0.715258	0	0.004316	0.032311	0.038723	0.040979
3.00	0.736808	0	0	0.019557	0.033943	0.038674
2.15	0.781213	0	0	0	0.011925	0.027762
1.49	0.770551	0	0	0	0	0
1.26	0.650295	0.101664	0	0	0	0
1.13	0.479971	0.237507	0.022178	0	0	0
1.03	0.155113	0.248961	0.162529	0.101227	0.055383	0.019877

Table 6.3: The share of sector 2 at all locations in equilibrium given changing intra-urban trade costs

τ	Location					
	1	**2**	**3**	**4**	**5**	**6**
4.00	0.072171	0.027151	0.015831	0.017133	0.017795	0.018331
3.00	0.068619	0.032301	0.018562	0.016454	0.017263	0.017819
2.15	0.058664	0.04537	0.02447	0.017889	0.015991	0.016716
1.49	0.030247	0.101529	0.031255	0.025321	0.021945	0.019153
1.26	0.221443	0	0	0.000379	0.009966	0.016253
1.13	0.260344	0	0	0	0	0
1.03	0.256911	0	0	0	0	0

Table 6.4: The share of sector 2 when τ changes between values of 1.49 and 1.26

τ	Location					
	1	**2**	**3**	**4**	**5**	**6**
1.4	0.031867	0.112689	0.02851	0.024886	0.022907	0.020734
1.37	0.046907	0.106016	0.024623	0.023266	0.022562	0.020979
1.35	0.06697	0.096125	0.019839	0.021086	0.021822	0.020913
1.34	0.089882	0.080818	0.01549	0.018994	0.021025	0.020698
1.33	0.128216	0.050336	0.009532	0.015516	0.019707	0.020239
1.327	0.162032	0.016725	0.005874	0.013719	0.018911	0.019978
1.32	0.188971	0	0.001419	0.010411	0.017122	0.019429
1.30	0.195698	0	0	0.008584	0.016043	0.019026
1.28	0.207692	0	0	0.004513	0.01352	0.017887

6.3.2 Two-dimensional Two-sector Urban Systems Model

This section presents the two-dimensional two-sector urban systems constructed by assembling 8 urban models that are same as the one demonstrated in Section 6.3.1. For the simplicity of testing, the symmetric racetrack space setting is adopted at the inter-urban scale. The two-dimensional internal urban space is embedded in each region on the racetrack as in Section 5.1. The distances between regions are detailed in Table 6.5. The same assumption is made about the two sectors as in Section 6.3.1. This illustration tests the same scenario as before in terms of the exogenous change. Here, I examine the impact of decreasing intra-urban trade costs on the inter-urban spatial structures in long-run equilibrium and check the conformity of the geographic model to the theoretical model. The simulations are configured with $\alpha = 2, \mu = 2, f = 1, n = 32, t = 1, T = 4$. Here, I assume very high inter-urban trade costs and modestly low commuting costs.

Table 6.5 Distance matrix for the regions in the racetrack space setting in Figure 5.1

Region	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0	4	8	12	16	12	8	4
2	4	0	4	8	12	16	12	8
3	8	4	0	4	8	12	16	12
4	12	8	4	0	4	8	12	16
5	16	12	8	4	0	4	8	12
6	12	16	12	8	4	0	4	8
7	8	12	16	12	8	4	0	4
8	4	8	12	16	12	8	4	0

Figure 6.4 shows the simulated results of population distributions across 8 regions in long-run equilibrium given the exogenous change of intra-urban trade costs. The values for τ are the same as those used in Section 6.3.1 for consistency. When the internal and external trade costs are both high ($\tau = 4,3$), spatial separation limits the concentration of firms at any location at the intra-urban scale and also prevents dispersion at the inter-urban scale. Based on the definition specified in Section 5.4, as τ decreases from 2.15 to 1.03, the long-run equilibria for sector 1 and 2 follow the evolutionary paths respectively as follows:

$$(P,P) \rightarrow (M,M) \rightarrow (P,P) \rightarrow (P,P) \rightarrow (P,0) \text{ Sector 1}$$

$$(P,P) \rightarrow (P,P) \rightarrow (P,P) \rightarrow (P,M) \rightarrow (M,0) \text{ Sector 2}$$

As τ decreases, the spatial structures at the inter-urban scale first become dispersed ($\tau = 1.49$), then become agglomerated again when internal trade costs are reduced to the low level ($\tau = 1.03$). Recall that in Section 6.3.1, falling τ causes the internal urban structure to follow a decentralization-centralization-decentralization process. When firms are concentrated at fewer locations in a large city to take advantage of lower trade costs, competition and urban costs become heavy burdens and eventually lead to the spreading of firms to smaller cities. With decreasing τ, this trend continues until a highly dispersed pattern becomes the long-run equilibrium ($\tau = 1.49$), where the concentration of sector 1 firms at the central location forms a monocentric pattern (M,M). As the internal trade costs keep falling to a low level, decentralization of sector 1 firms leads to a more polycentric structure that let sector 1 firms in SBDs can enjoy low urban costs and keep good access to large market in the city. And, because their less differentiated goods, sector 2 firms just start reacting to the reduced internal trade costs by concentrating at the central locations of all the smaller regions (P, M). The benefit of polycentricity motivates more firms to agglomerate at fewer large cities, which eventually causes a monocentric structure at the inter-urban scale ($\tau = 1.03$) with a fully agglomerated region. In contrast, sector 2 firms concentrate at the central location of this region exhibiting a monocentric structure at the intra-urban scale because they react the change slowly and only reach the stage of centralization in the decentralization-centralization-decentralization process. Therefore, the evolutionary paths of sector 1 and 2 correspond to the Case 3 and 4.1 for the theoretical model in Section 5.4.1. So far, the results from the geographic model have confirmed the interdependency of spatial structures across scales unveiled by our theoretical models, and thus prove the conformity of the geographic model to theoretical models.

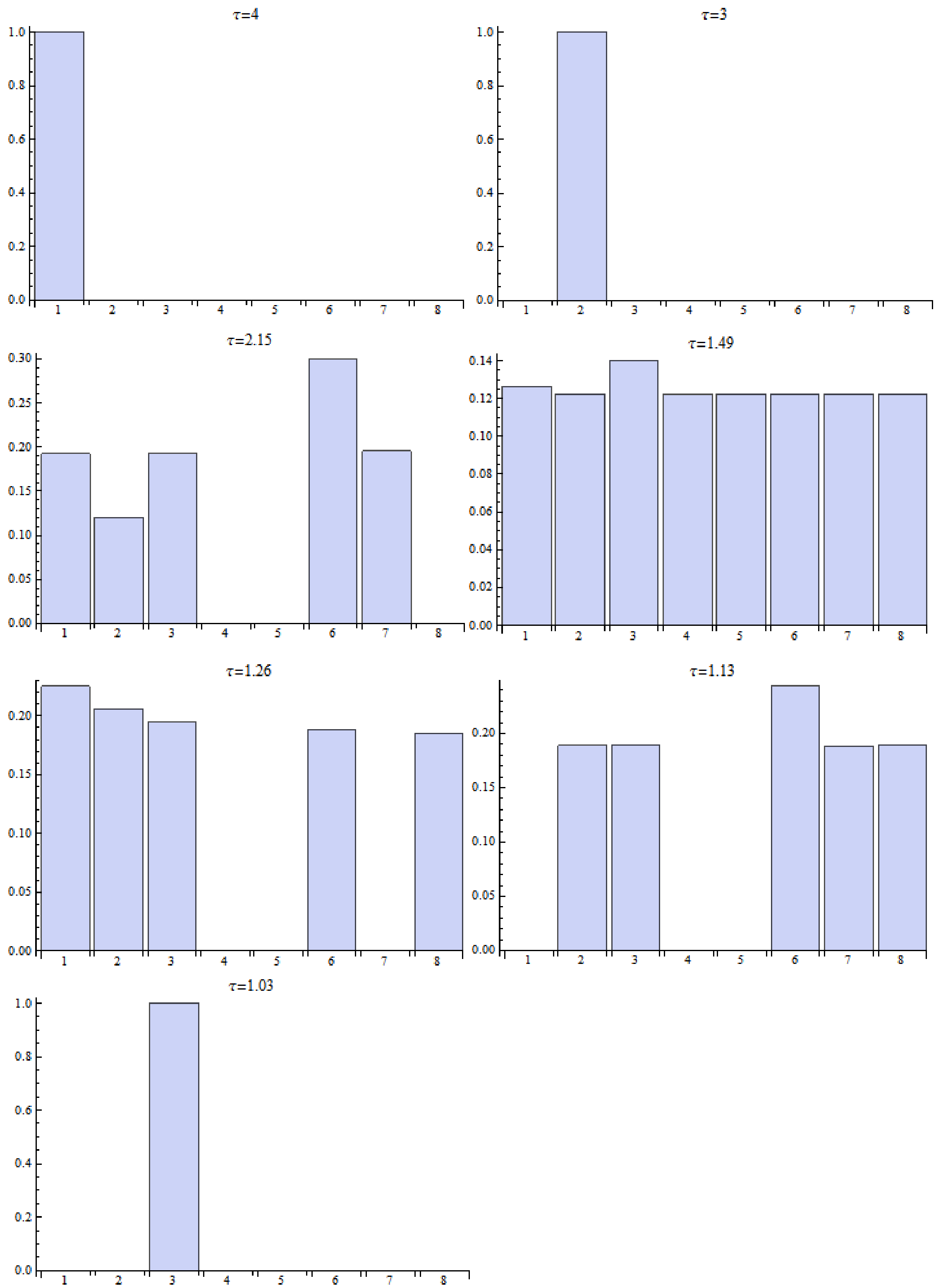


Figure 6.4: Equilibrium population distributions by share across 8 regions for changing intra-urban trade costs

CHAPTER 7: HIGH-PERFORMANCE MODELS OF URBAN SYSTEMS

The design and implementation of the proposed theoretical and geographic models concern the efficiency of computing when applied to large-scale problems, such as a megaregion containing multiple metropolitan areas and each metropolitan area contains many cities and small towns. The adopted strategy is to employ high-performance computing (HPC) technologies to tackle the computational issues. This section is centered on the design and implementation of HPC-enabled models of urban-regional systems. I first review the literature on applications of HPC to urban spatial modeling, and then discuss the paradigms of parallel computing and strategies for model implementation.

7.1 HPC in Urban Spatial Modeling

High-performance computing, in a layman expression, refers to the practice of increasing computing power through advanced hardware and/or software (e.g., supercomputers or parallel computers) that offer much higher performance than regular computers or workstations. Early explorations of applying HPC technologies to urban spatial modeling date back to the 1980s (Harris, 1985; Openshaw, 1987). Making use of HPC, parallel processing was proposed to accommodate the increasing complexity of spatial analysis and modeling (SAM) in terms of the volume of data at fine spatial and temporal resolutions and of more sophisticated algorithms and models (Armstrong, 2000). Openshaw and Turton (2000) identifies a list of opportunities that would lead towards a computational human geography:

- To speedup existing compute-bound tasks in order to engage large-scale experimentation and simulation of complex human and physical systems and real-time geospatial analysis.
- To improve the quality of results by using compute-intensive methods to reduce the number of assumptions and shortcuts forced by computational restraints.
- To permit larger databases to be analyzed or to obtain better results by being able to process finer resolution data.
- To develop completely new and novel approaches based on computational technologies such as computational intelligence methods.

Along these lines, several special journal issues stand out as landmark contributions. A 1996 issue of the International Journal of Geographic Information Science (IJGIS) initiated a focus on parallelization of existing computationally intensive geospatial operations (Ding and Densham, 1996; Clematis et al., 1996). Later, a 2003 issue of Parallel Computing extended this line of research to parallel spatial algorithms and data structures (Clematis et al., 2003; Wang and Armstrong, 2003). As cyberinfrastructure emerged as a new paradigm to harness the power of data and computation sciences, two special issues (a 2009 issue of IJGIS and a 2010 issue of Computers, Environment and Urban Systems) on geospatial cyberinfrastructure are in order aiming at elevating geospatial sciences to the next level with the support of HPC as one of its critical components (Yang and Raskin, 2009; Yang et al., 2010). Among those issues, significant enhancements have been demonstrated by specific applications employing HPC such as conducting computationally intensive geospatial analysis methods and large-scale forecasting of dust storms (Xie et al., 2010; Wang and Liu, 2009).

Riding this tide, recent research has been targeted at taking full advantage of the HPC resources available encompassing the development and application of parallel algorithms (Wang and Armstrong, 2009; Li et al., 2010; Guan et al., 2011; Yin et al., 2011; Widener et al., 2012) and parallel libraries (e.g., pRPL, Repast HPC and EcoLab, see Guan and Clarke, 2010; Repast Team, 2012; Standish, 2012) for SAM. The latest contribution is featured by a 2013 issue of IJGIS (Wang, 2013; Wang et al., 2013) that focuses on the development of a new generation of cyberinfrastructure-based GIS (CyberGIS) as the synthesis of advanced cyberinfrastructure, GIScience, and SAM (Wang, 2010). Expanding the frontiers of CyberGIS, this issue highlights establishing integrated and scalable geospatial software ecosystems with the pursuit of scalable methods, algorithms, and tools that can harness heterogeneous HPC resources, platforms, and paradigms (message passing vs. shared memory) (Shook et al., 2013; Tang, 2013; Zhang and You, 2013; Zhao et al., 2013).

Specific to the domain of applications in urban and regional modeling, HPC has been leveraged to support modeling endeavors that fall into three categories. Firstly, in the approach of integrated land-use and transport models, parallel computing is utilized to help solve general equilibria or fixed point problems that require a lot of iterations of numerical approximation. Specifically, this involves matrix balancing and estimation in spatial interaction, input-output, and spatial regression models (Turton and Openshaw, 1998; Davy and Essah, 1998; Wong et al., 2001) and spatial network and location optimization in path search, traffic assignment, and location allocation problems (Birkin et al., 1995; Wisten and Smith, 1997; Hribar et al., 2001; Lanthier et al., 2003; Smith et al., 2010). These studies revealed how effective HPC is to accelerate existing models so that they can be applied to a finer spatial detail or resolution on the largest available databases and thus provide improved levels of solution, accuracy, and representation.

Secondly, in the computational approach including CA, ABM and various computational intelligence methods (e.g., genetic algorithm and neural networks), the complex and dynamic urban geographic phenomena under simulation necessitate the support of HPC because these models enable the incorporation of heterogeneous factors and processes at multiple spatiotemporal scales and their decentralized micro-level interactions that give rise to macro-level structures or regularities, and thus prompt massive computational demands (Dattile et al., 2003; Guan, 2008; Guan and Clarke, 2010; Li et al., 2010; Tang et al., 2011; Gong et al., 2012a; Gong et al., 2012b; Meentemeyer et al., 2013; Porta et al., 2013; Pijanowski et al., 2014). Furthermore, due to their intrinsic mechanism of concurrency and parallelism (e.g., decentralization of cells/agents and interactions, evaluation of individuals in natural selection, and distributed processing of interconnected neurons), these computational models are inherently suitable for parallel computing (Wong et al., 2001; Tang and Bennett, 2011; Gong et al., 2012b). Especially, the calibration of these models involving estimating a large number of combinations of parameters and their simulations entailing a considerable number of iterations all justify the utilization of HPC, which in turn enables to gain unprecedented insights into the complexity and dynamics of urban regional systems and manifest the opportunities to discover new theories (Meentemeyer et al., 2013; Pijanowski et al., 2014; Tang and Jia, 2014; Gong et al., 2014).

Finally, the approach of microsimulation and activity-based modeling, increasingly popular as a comprehensive decision support system for practical urban and transportation planning, is notorious for the level of details of the data it requires (parcel level spatial resolution and individual travel activities) and the heavy computing load it relies on (individual level location choice and vehicle level traffic simulation). General modeling frameworks of this approach such as UrbanSim and TRANSIMS, without exception, resort to HPC for effective and efficient problem solving once applied to real world planning projects (Awaludin and Chen, 2007; Nagel and Rickert, 2001; Richert and Nagel, 2001; Cetin et al., 2002). Notably, operational models implemented via this approach and applied real world planning practices include Oregon Statewide models (ODOT, 2001, 2010) that combine macro- and micro-level simulations of statewide land-use, economy, and transport systems and Chicago Metro Evacuation Planning (TRACC, 2011) that adapt TRANSIMS' normal traffic forecasting capability to dynamic evacuation scenarios, which all take advantage of HPC clusters in order to achieve extraordinary performance.

7.2 Computational Complexity in Models of Urban Systems

This section disentangles the computational complexity involved in computing the proposed theoretical and geographic models. By identifying the computational issues, performance concerns motivate the design and implementation of the proposed models to leverage the power of HPC in order to tackle these issues. The computational complexity of the proposed theoretical and geographic models lies in two aspects: computational intensity and data intensity.

The first aspect, computational intensity, relates to the nature of the proposed models as nonlinear systems with equilibrium searching at multiple levels. Regarding the theoretical urban models, their nonlinearity and one-dimensional continuous space setting require short-run equilibrium be calculated through numerical approximations which usually cost more time by running algorithmic programs, since the analytical solution does not exist. Furthermore, the stability of every short-run equilibrium needs to be tested numerically by calculating the derivative of certain functions. Due to the multiplicity of equilibria, the model must obtain all the stable equilibria and select an appropriate one based on the path choice rules. Once short-run equilibria are determined for all regions, it just completes one iteration towards the long-run equilibrium. If the testing shows that the long-run equilibrium conditions are not satisfied, the model will go to the next iteration with adjusted populations for each region through migration dynamics. This process continues until a long-run equilibrium is achieved. This is not an end because the stability of a long-run equilibrium also needs to be tested. The common approach is through perturbation of the initial distributions of households and firms across locations. Usually 5 to 10 instances of random perturbation runs are needed to justify the stability of a long-run equilibrium. Each perturbation run follows the same process as described above. This is how the models involve stochasticity to search for regularities.

The proposed geographic model follows a similar computational process except that it involves a two-dimensional discrete space setting for urban models. This space setting features added dimensions of complexity in that heterogeneous locations are defined by non-uniform sizes and more complex distance relations that are only allowed in two-dimensional space. As a result, it renders a large variation of the densities of firms and households across locations and a more complex commuting and trading patterns. To obtain short-run equilibrium with this space setting, instead of using standard numerical methods to solve nonlinear problems, an agent-based approach was proposed (in Section 6.2) and implemented as an iterative algorithm to approximate the equilibrium conditions. With all the added complexity, it usually takes more time for the geographic model to converge to a short-run equilibrium, and thus even more time to reach a long-run equilibrium.

Due to the nonlinearity of the model, model parameters cannot be examined analytically. Numerical simulation becomes the only way to study the model behaviors under different parameter settings. For example, Section 5.4 investigates the effects of changing spatial costs at different scales on the spatial equilibrium of urban systems. Each point represented by a set of parameter values in Figures 5.2 and 5.3 indicates a stable long-run equilibrium, which takes the entire process I have discussed above to obtain. Both figures must include more than 300 points in order to have clear pictures about the evolution of spatial structures of urban systems under the change of certain spatial costs. Computing stable equilibrium for more than 100 points can already take more than a week on a standard PC workstation. Examining the infinite combinations of these parameters is simply a mission impossible.

The second aspect, data intensity, relates to the amount of data involved to model all the entities in the proposed framework of urban systems. In general, the data intensity increases with the geographic units such as number of locations or number of regions.

Therefore, larger extent of the study area or finer granularity of the geographic unit can easily intensify the computational complexity. Moreover, with an increasing number of geographic units, the interactions between them, such as commuting flows and trade flows, grow exponentially, which imposes even higher intensity on data processing. In addition, data intensity also relates to the model complexity. For example, when multiple industries are considered, the agent types for firms and households is not only differentiated by locations but also by industries. This added dimension will cause more than proportional increase of the data intensity because there are interactions between firms of different industries. These types of data intensity will be exemplified when calibrating the geographic model against empirical data in a real world application to the Carolinas Region (Chapter 8).

7.3 Parallel Models of Urban Systems

This section investigates how to leverage the existing distinct HPC platforms to tackle the identified computational and data intensities involved in the proposed models of urban systems. Two existing paradigms for parallel computing are first discussed in terms of their advantages and disadvantages. Then, an implementation environment is proposed to develop the parallel models of urban systems.

It has become standard to use the paradigm of message-passing on platforms of cluster computing, grid computing, and cloud computing, which become less expensive computational resources to access. The advent of parallel computing in personal computers is now opening new avenues for parallel SAM on platforms such as multicore CPUs and many-core graphic processing units (GPUs, Owens et al., 2008) in compliance with the shared-memory paradigm.

In the message-passing paradigm, computing elements (e.g., individual computers) have their local memory space and exchange data through sending and receiving data streams packaged as messages over interconnected networks (Wilkinson and Allen, 2005). Thus, communications must be coordinated to reduce the costs of accessing remote memories during data exchange, which may significantly complicate the parallel programing. However, due to its explicit consideration of the communication process, the message-passing paradigm is highly flexible and portable to a range of parallel platforms such as vector supercomputers, computer clusters, and grid computing systems. In contrast to distributed memory systems in message-passing paradigm, the fundamental principle of shared-memory systems is that multiple processors or cores are organized in a way that multiple processing units access a common memory space simultaneously (Wilkinson and Allen, 2005). Such architecture supports thread-level parallelism to boost computational performance when physical limits curtail further clock rate increases of single CPU. Multicore architectures are based on a coarse-grained shared-memory paradigm aiming to exploit parallelism through coordination among multiple concurrent processing threads within a single program. Compared to many-core shared-memory architectures, such as GPUs which support a large number of fine-grained light-weight threads (millions), they use a small number of threads, each of which has much more powerful computational capability. In practical applications, the message-passing and shared-memory paradigms can be combined to maximize the exploitation of different types of parallelisms on heterogeneous parallel platforms (Kranz et al., 1993).

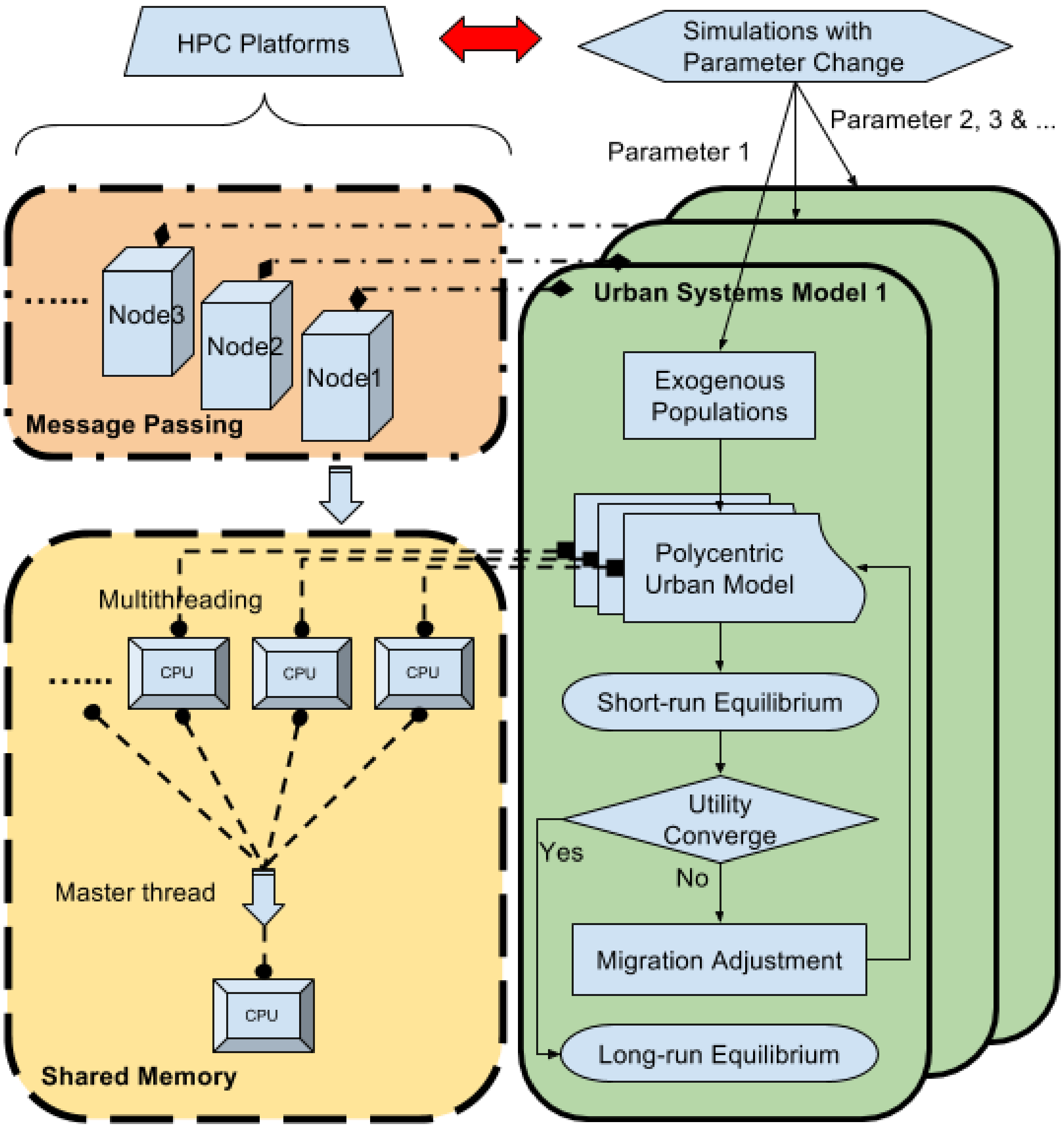


Figure 7.1: Parallel algorithms of urban systems model supported by both shared-memory and message-passing platforms

The parallel model of urban systems is designed to leverage both shared-memory and message-passing platforms. The parallelization strategies, depicted in Figure 7.1, are applied to both theoretical and geographic models, since they follow the same set of general procedures. Practically, the design of the parallel model aims to adapt to the general architecture of cluster computing, ranging from small-scale cluster computers to supercomputers, which can be characterized by a hierarchical structure of the combination of shared-memory and message-passing platforms. Because a computer cluster constitutes a massage-passing platform by implementing the massage-passing interface (MPI) protocol between a large number of interconnected computers via high-speed networks, among which each individual computer itself is a shared-memory platform. Given that the shared-memory paradigm is more efficient to handle communications between operations than the message-passing paradigm, the design of the parallel model follows a main principle that model operations with high dependency are assigned to shared-memory platform while relatively independent model runs (or simulations with changing parameters) are handled by the message-passing platform. Therefore, the computational and data intensity internal to the model are shared by CPUs and/or cores within a computer, while numerous simulations with different sets of parameters are processed in parallel by different computers in the cluster. Specifically, on a shared-memory platform, each urban model in the urban systems model is processed by a CPU/core and the interactions between urban models are through inter-thread communications by accessing the shared memory space among CPUs/cores. MPI is used currently to allocate simulations of models with different parameters to different computers. If an urban systems model is too large to be accommodated by a single computer, MPI can be leveraged to decompose the urban system model into smaller sub-domains, each of which can be adequately handled by an individual computer. This extension will be part of future studies.

The proposed theoretical and geographic models are implemented in WOLFRAM MATHEMATICA, which is a widely used symbolic and numerical modeling system with built-in ABM capability. More important, MATHEMATICA integrates both message-passing and shared-memory paradigms in a seamless HPC environment. Thus, it provides parallel computing support using both the threading-based multicore and the MPI-based computer cluster platforms. The resulting implementation of the parallel models speed up the model performance by several magnitudes. Specifically, the CPU time for 100 model runs has been reduced from more than a week to 2 hours by using a computer clusters PYTHON in University Research Computing at UNC Charlotte.

CHAPTER 8: APPLICATION TO THE CAROLINAS REGION

This chapter aims at assessing the applicability of the proposed geographic model (chapter 6) with a case study. The Carolinas region (Figure 8.1), a real world example of urban-regions, is presented as a test bed for the application. To serve this purpose, a calibration procedure is explored to investigate the extent to which the proposed model can be calibrated given the limitations on data availability. By calibrating the model with data collected from the study area, the structural parameters of the model can be set with meaningful values, and thus the model outputs have a more solid foundation in reality. As a result, the geographic model can be assessed in terms of its capability to reproduce the observed spatial structures of the inter- and intra-urban agglomerations and how these spatial structures causally relate the economic behaviors of agents when a certain heterogeneity is assumed in space such as transportation networks and population distribution. This superior explanatory power, carried out by the simulations of the geographic model, is of great importance to spatial policy making and scenario evaluation guiding the devising of policy instruments that can potentially direct urban developments.

8.1 Study Area

The study area (within red color boundary in Figure 8.1) is represented by a zonal geography connected by transportation networks. Treated as urban-regions (a megaregion), the study area includes the majority of the Carolinas region consisting of two U.S. states: North Carolina (NC) and South Carolina (SC), one county from Tennessee (TN) and six counties and three cities from Virginia (VA). These adjustments are made because I use the configurations of BEA economic areas from the Bureau of Economic Analysis (BEA) as the first-level geographic entities in the study area. I follow this geography because BEA economic areas defines the relevant regional market in terms of the economic relations between the metropolitan areas and surrounding counties (Johnson and Kort, 2004). Thus, each BEA area contains at least one core Metropolitan Statistical Areas (MSA, shaded areas in Figure 7.1) and adjacent areas, which in a general sense corresponds to a metropolis at the inter-urban scale. Each metropolis comprises a collection of second-level geographic entities such as counties that constitute the intra-urban scale. A location is defined by the finest geographic unit; its choice depends on data availability and on the computational concern. Here, for the sake of simplicity, I use counties as the unit for location (the finest geography in Figure 8.1). These geographic entities at different spatial scales represent a hierarchical zonal geography. The unit here is heterogeneous in terms of size and shape, which prompts the consideration of a variable amount of land consumption at different locations. Due to the assumption of fixed land consumption by a household, this heterogeneity in land consumption leads to an uneven distribution of populations across locations. Given these settings, the agent-based location choice approach developed in section 6.3 can be properly applied to the heterogeneous zonal geography here.

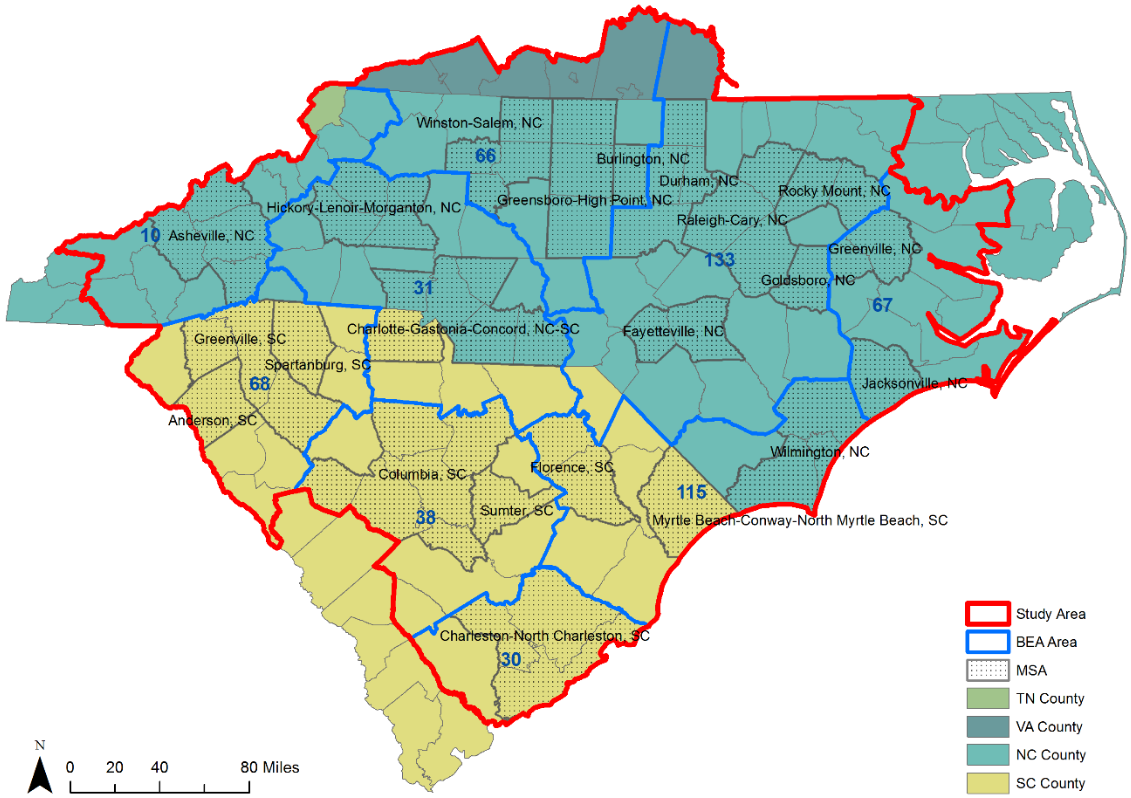


Figure 8.1: Study area (the majority of Carolinas region and some adjacent counties)

8.2 Data Availability and Assumptions

The data I collect for the study area over the recent years are at the county level (Table 8.1). For convenience, data for 2012 were used for model calibration, because it is the latest year in which data from all sources are available. I now discuss the characteristics of these data sets and the assumptions based on which they will be applied.

Population and employment data are from the data company IMPLAN, who provides economic input-output data and models. Employment data is available for 5 super sectors of the economy, which are detailed in relation to 2-digit naics code in Table 8.2. Employment will be used as a proxy for firms in the model. To ease the burden of simulation, the 5 super sectors are further reduced to one general sector. To simplify the model structure, only employed population is modeled. Given that, I further assume that both population that goes outside of the study area for employment and the employees that come from outside of the study area are not considered by the model.

Commuting patterns between counties are accounted for by the residence county to workplace county commuting flows from 5-year American Community Survey (ACS) for 2009 to 2013. Since this is an estimated data set, in order to reduce the risk of bias, the percentage of flows to workplace counties relative to their residence counties are derived and used instead of the absolute volume of flows. This can be directly applied to population and employment data without creating inconsistency issues between data from different sources. In addition, the flows that go and come outside of the study area are eliminated to be consistent with population and employment data.

To consider the real geography of transportation networks in transport costs, instead of using simple Euclidian distances, I collected the travel time by driving over existing road networks between counties from Google Maps Distance Matrix API. Here, the travel costs should be interpreted as the average costs between counties because each county (as the finest geographic unit in this application) is treated as a location, and thus the geography of a county is represented by its centroid according to our graph-based space setting (described in Section 6.1). This configuration is for the sake of simplicity in demonstration. More accurate measures of transport costs can be applied when model applications are based on finer geographic units (such as Traffic Analysis Zones).

Table 8.1: Data sources

Variable	Data	Spatial Resolution	Temporal Resolution	Source
Household	Population	County	2007-2012	IMPLAN social accounting matrix
Firm	Employment by 5 sectors	County	2007-2012	IMPLAN social accounting matrix
Commuting flows	Commuting pattern	County-to-County	2009-2013	Census 5-Year ACS
Transport/ commuting costs	Travel time	County-to-County	Present	Google Maps service

Table 8.2 Super economic sectors for employment

Sector	Description (NAICS codes)
1	Natural Resources, Construction, Utilities (11,13, 21-23)
2	Manufacturing (31-33)
3	Trade, Transportation, Post Office (42, 44-45, 48-49)
4	Communication, Information, Finance, Real Estate, Management of Companies and Business Services (51-56)
5	Other Services (61-62, 71-72, 81, 92)

8.3 A Partial Calibration Approach

This section proposes a partial calibration procedure of the geographic model in order to assess its capability to reproduce the observed inter- and intra-urban spatial structures. Calibrating the geographic model requires the identification of key equations and variables. In the literature, the wage equation lies at the heart of empirical studies of NEG in order to establish a spatial wage structure gradient from urban centers to suburbs (Bosker and Garretsen, 2007; Brakman et al., 2006; Knaap, 2006; Hanson, 2005; Redding and Venables, 2004). The reasons to rely on a wage equation rather than a traditional equilibrium price function are that: 1) on the theoretical side, labor migrates in response to levels of regional real wage differentials; 2) on the practical side, data on wages are more readily available than prices of consumption goods. In this section, I first discuss major calibration strategies that have been reported in the NEG literature regarding the way to estimate the wage equation and their data needs. Then, I explore a partial calibration procedure that can be applied with limited data.

The wage function is derived from the equilibrium condition for product market that equates supply and demand (5.1.3.1). That is, let the demand $D_i(x)$ for a variety produced by a firm at discrete location *x* of region *i* equal its equilibrium production $Q_i^*(x)$ (5.1.2.8), $Q_i^*(x) = D_i(x)$. Then, solving for the equilibrium wage at location x of region i as in (4.3.3.1) I can obtain the wage function:

$$W_i(x) = \frac{1}{\sigma \cdot f} \sum_y \frac{\tau(y|x)^{1-\sigma}}{PIS_i(y)} [\alpha \cdot n_i(y) + \mu \cdot m_i(y)] + \frac{1}{\sigma \cdot f} \sum_{j \neq i} \sum_y \frac{T(j|i)^{1-\sigma}}{PIS_j(y)} \left[\alpha \cdot n_j(y) + \mu \cdot m_j(y)\right] + \frac{\mu}{f} \left[\ln \mu - 1 - \ln \frac{\sigma}{\sigma - 1} - \frac{1}{1-\sigma} \ln PIS_i(x)\right] \quad (8.3.1)$$

and

$$PIS_i(x) = \sum_y m_i(y)\,\tau(y|x)^{1-\sigma} + \sum_{j\neq i}\sum_y m_j(y)\,T(i|j)^{1-\sigma} \quad (8.3.2)$$

$PIS_i(x)$ can be interpreted as the inverse of price index, and $[\alpha \cdot n_i(y) + \mu \cdot m_i(y)]$ is the expenditure at location y of region i. Rearranging (8.3.1) gives the following non-linear equation that can be estimated:

$$W_i(x) = \beta_0 + \beta_1 \left[\sum_y \frac{\tau(y|x)^{1-\sigma}}{PIS_i(y)} n_i(y) + \sum_{j\neq i}\sum_y \frac{T(j|i)^{1-\sigma}}{PIS_j(y)} n_j(y)\right] +$$

$$\beta_2 \left[\sum_y \frac{\tau(y|x)^{1-\sigma}}{PIS_i(y)} m_i(y) + \sum_{j\neq i}\sum_y \frac{T(j|i)^{1-\sigma}}{PIS_j(y)} m_j(y)\right] + \beta_3 \ln PIS_i(x) \quad (8.3.3)$$

$$\beta_0 = \frac{\mu}{f}(\ln \mu - 1 - \ln \frac{\sigma}{\sigma - 1}),\ \beta_1 = \frac{\alpha}{\sigma \cdot f},\ \beta_2 = \frac{\mu}{\sigma \cdot f},\ \beta_3 = \frac{\mu}{f(\sigma - 1)}$$

where β's are the estimated parameters from which in principle the structural NEG parameters can be inferred (Redding and Venables, 2004; Hanson, 2005).

There are two main strategies to estimate the wage equation (8.3.1) in the empirical NEG literature (Bosker and Garretsen, 2007). The first strategy is to estimate the non-linear wage equation directly (one step), which is introduced by Hanson (2005). To accomplish this, a trade cost function must be assumed to deal with the lack of directly measurable trade costs. Hence, the parameters in trade cost function are jointly estimated with the structural NEG parameters in the wage equation. Moreover, to deal with the common issue with the lack of data on regional price indices $\overline{P_i(x)}$ (or $PIS_i(x)$ here), this strategy usually approximates the price index of a region via an average of the wage levels in that region and the nearby economic centers. However, as the geographic units considered grow, the increasing number of simultaneous non-linear equations to be estimated poses a huge difficulty in terms of computation and finding viable solutions. This imposes a limitation on applications involving a large number of geographic units such as the urban systems model developed here.

The second strategy is a two-step procedure to estimate the wage equation with linear models by making use of inter-regional trade flow data, which is pioneered by Redding and Venables (2004). The first step involves the estimation of market and supplier capacity and inter-regional trade costs based on the trade flow data. It follows a spatial interaction model (gravity type) of trade flows from production sites (location *x* of region *i*) to consumption sites (location *y* of region *j*):

$$TF_{ij}(y|x) = m_i(x)p_i(x)\left[n_j(y)q_{ij}(y|x) + m_j(x)q_{I,ij}(y|x)\right] = m_i(x) \cdot \mathrm{T} \cdot \frac{\alpha \cdot n_j(y) + \mu \cdot m_j(y)}{PIS_j(y)}$$

where $\mathrm{T} = \begin{cases} \tau(y|x)^{1-\sigma}, & i = j \\ T(i|j)^{1-\sigma}, & i \neq j \end{cases}$ and $m_i(x)$ refers to supply capacity indicating firm density at the production site while $\frac{\alpha \cdot n_j(y) + \mu \cdot m_j(y)}{PIS_j(y)}$ refers to market capacity that reflects the magnitude of consumer expenditure deflated by the price index at the consumption site. Taking logs on both side of the above equation and replacing market and supply capacity by an importer and exporter dummy respectively, i.e. $sc_i(x) = m_i(x)$ and $mc_j(y) = \frac{\alpha \cdot n_j(y) + \mu \cdot m_j(y)}{PIS_j(y)}$, I obtain the following equation that can be estimated:

$$\ln TF_{ij}(y|x) = \ln sc_i(x) + \ln \mathrm{T} + \ln mc_j(y) + \varepsilon \quad (8.3.4)$$

where ε is an i.i.d. lognormal disturbance term.

The second step involves the construction of the so-called *real market access* (RMA) and *supplier access* (SA) based on the estimated market and supply capacity and the trade costs in the first step. In the wage equation, RMA refers to the first two terms in (8.3.1):

$$\sum_y \frac{\tau(y|x)^{1-\sigma}}{PIS_i(y)}[\alpha \cdot n_i(y) + \mu \cdot m_i(y)] + \sum_{j\neq i}\sum_y \frac{T(j|i)^{1-\sigma}}{PIS_j(y)}\left[\alpha \cdot n_j(y) + \mu \cdot m_j(y)\right] =$$

$$\sum_j \sum_y \frac{\alpha \cdot n_j(y) + \mu \cdot m_j(y)}{PIS_j(y)} \mathrm{T} = \sum_j \sum_y mc_j(y)\, \mathrm{T} = RMA_i(x)$$

which is a cost weighted sum of market capacities of all consumption sites. SA refers to $PIS_j(y)$ (8.3.2) that is the cost weighted sum of supply capacities of all production sites:

$$PIS_j(y) = \sum_i \sum_x m_i(x)\, \mathrm{T} = \sum_i \sum_x sc_i(x)\, \mathrm{T} = SA_j(y)$$

Therefore, the constructed RMA and SA can be used to estimate the wage equation in the following form:

$$W_i(x) = \beta_0 + \beta_4 RMA_i(x) + \beta_3 \ln SA_i(x) \quad (8.3.5)$$

where $\beta_4 = \frac{1}{\sigma \cdot f}$. In sum, this estimation strategy is more robust compared to the first one, however adopting this strategy is entirely conditioned on the availability of trade flow data.

Given the complexity of estimating many non-linear equations simultaneously using the first strategy, I prefer to adopt the second strategy. However, it was not possible to secure access to the inter-county trade flow data for the study region. Therefore, I ended up only having limited data sources available for model calibration. Given all the limitations, I turn to explore a calibration approach that can be applied in a data scarce situation. The data sets discussed in Section 8.2 are used for calibrating the corresponding variables. Other endogenous variables will be computed during simulations but not calibrated against any empirical data. These variables include land rent, wages, and price index. To calibrate only a portion of the whole set of endogenous variables, a partial calibration procedure needs to be developed. First of all, I take several parameters as fixed constants, such as $\alpha, \mu, f,$ and σ. Therefore, their values are provided in the simulations. Second, our focus is on estimating values for the parameters related to three spatial costs: $t, \tau,$ and T. The estimation process is to search for appropriate values for these parameters through simulations in order to let the model generate similar patterns given by the data used for model calibration. Specifically, a nonlinear least-squares method is employed to minimize the sum of squared difference between the observed population distribution and model outputs:

$$S = \sum_{i=1} r_i^2 \quad (8.3.6)$$

where $r_i = o_i - p(X_i, t, \tau, T)$, a residual, is the difference between the observed value o_i and the model output $p(X_i, t, \tau, T)$, which is a function of other observed variables X_i and adjustable parameters $t, \tau,$ and T. Since this simulation-based approach is a type of brute-force search, it is very time-consuming and thus prevents us from searching exhaustively the parameter space. Given the complexity of our model, it is only feasible to search for a very limited range of values for parameter estimation. Hence, an optimal or accepted solution is not guaranteed.

I take the following parameters as constants and fix their values during simulations: $\alpha = 2, \mu = 2, f = 1,$ and $\sigma = 4$ and only estimate the values for $t, \tau,$ and T. Because the simulation-based brute-force search is used, I cannot exhaust the entire space for parameter combinations. Consequently, I only take several portions from the parameter space as target for searching, where I think the solutions may exist. The fitness of resulting distributions from calibrated simulations can be determined by the corresponding least squares that they produces. The lower the least squares are the better the output distributions fit the observed data. That is, when the values of least squares are closer to zero, more accurate estimates (searched values for adjustable parameters) can be obtained for these parameters.

Model outputs include population distribution across BEA areas and populations for counties within each BEA area. Because workers only find jobs in the BEA areas where they live and because it is assumed that there is a constant relation between population and employment (4.3.1.3a), the population distribution across BEA areas also indicates the employment distribution. The generated county population distributions reflect the internal spatial structures of BEA areas. Within each BEA area, population growth and decline in counties are constrained at the same rate, while rates for different BEA areas may differ. This constraint is on purpose to fit the observed population distribution between counties within each BEA area. On the other hand, it is not constrained on the capacity of land for each county, which means hypothetically a full agglomeration into one single region may happen. Similarly, there is no constraint on the migration volume between BEA areas, although it could be done by calibrating against real migration flows. Consequently, it allows full migration from one BEA area to others due to the migration dynamics adopted in the model (specified in Section 5.2). This may create empty regions that do not host any population. Since I use a fixed number of spatial units (BEA areas and counties), the geography is constant in the model. The population size of each county and BEA area is endogenous, which allows the emergence of new centers or new metropolises. However, the model does not predict the formation of new geography, since it does not create counties or BEA areas.

8.4 Simulation Results and Discussions

The simulated population distributions for BEA areas are compared to the observed distribution for 2012 (Figure 8.3) in order to examine the differences in spatial structures they represent. The comparisons result in three categories of simulated distributions with respect to their agglomeration (or dispersion) structures comparing to that of the observed distribution. For example, an output distribution may appear more agglomerated (aggressive) or more dispersed (conservative) than the observed distribution. And the output distributions may also exhibit a balance between the above two extremes but are still different from the observed distribution. The three scenarios of simulated outputs are labeled as aggressive, conservative and balanced accordingly for convenience. In each scenario, the set of estimated parameter values that achieve the best fitness are detailed in Table 8.3 with the minimized least squares associated with each. The simulated distributions based on these estimates are presented in Figure 8.2 compared against the observed population distribution at BEA areas for 2012. Regions in Figure 8.2 are indicated by their BEA area codes. Table 8.4 lists these codes and the major city in the corresponding BEA areas for the purpose of easy identification.

Table 8.3 Estimated parameter values for different calibration scenarios

Scenario	**Conservative**	**Balanced**	**Aggressive**
t	2.51189E-8	3.16228E-8	3.98107E-8
τ	1.70998	1.88207	2.15443
T	0.99999	0.99995	0.99901
Least squares	0.03663	0.12519	0.21486

Table 8.4 BEA economic areas and their corresponding major cities

BEA Code	Center city
10	Asheville
30	Charleston
31	Charlotte
38	Columbia
66	Greensboro
67	Jacksonville
68	Greenville
115	Myrtle Beach
133	Raleigh

In Figure 8.2, it seems that the balanced scenario is the closest to the observed distribution. However, the conservative scenario actually is the best fit among others based on the minimized least squares in Table 8.3. This may be because the balanced scenario captures one (31: Charlotte) of the two largest agglomerations (Figure 8.5) as almost the same magnitude as in observed data in terms of population share, while the conservative scenario captures more the relative order of the distribution. This also explains why the conservative scenario achieves the smallest least squares. Without predicting any prominent large agglomeration, the conservative scenario generates a rather dispersed pattern of urban development across the study area, with just a few modest agglomerations (Figure 8.2). According to the spatial distributions of population in Figure 8.4, population leaves the three largest regions (133: Raleigh, 31: Charlotte, 66: Greensboro) and move to the smaller regions. Raleigh region has the largest loss, which drops its rank of population share dramatically. In general, population partially agglomerates in the western areas, since Charlotte and Greensboro regions still keep their relative ranks of population share even though they experience population loss. Columbia (38) and Asheville (10) regions have the largest population gains, which change their ranks of population share. Greenville region also gains population and becomes one of the three largest agglomerations in the conservative scenario. Note that all the center cities keep their relative dominance within their regions thanks to the constraint that is imposed on the simulations.

The balanced scenario generates three large agglomerations with some small ones, but only one of the large regions is correctly captured. Both the balanced and aggressive scenarios predict the entire disappearance of the Raleigh region. Figure 8.5 shows a clear concentration of population in the western regions. The greatest population gains come from the three regions in the southwest (10: Asheville, 68: Greenville, and 38: Columbia), which also indicates that the predictions for them carry the largest residuals. Apparently, they are spatially clustered surrounding the best predicted area, the Charlotte region, which with the Columbia and Greenville regions together becomes the core of the population concentration in the west. The agglomeration surrounding the Charlotte region benefits from the transportation facilities connecting them (see the interstates networks in these areas in Figure 8.3). Looking at the internal structures, center cities in the Greensboro region lose their dominance and the same for the Myrtle Beach region (115). Center city in Charlotte keeps its dominance, while center cities in the Asheville, Greenville and Columbia regions all grow with more concentrations. Interestingly, secondary centers emerge close to the dominant centers in the Greenville and Columbia regions, while Charlotte region maintains its relatively monocentricity-oriented structure that a single dominant center with many small towns.

The aggressive scenario generates three fully agglomerated regions with all other regions gone. The three regions are 31 (Charlotte), 38 (Columbia) and 68 (Greenville), which draw all populations from other areas and keep their relative rank of population as in the observed data. One reason could be that they are neighboring areas constituting a fair symmetric triangular topology that increases the accessibility between them (Figure 8.3). Since the model weights on the impact of spatial costs, it is not surprising that these areas are singled out. In terms of internal spatial structures, all primary and secondary centers grow with larger concentrations and all primary centers maintain their dominance. Charlotte region still keeps its monocentricity-oriented structure with a single largest center. This can find explanations in Charlotte's base economy of financial and business services that have more international or domestic connections rather than local economic linkages. It also reflects the fact that most surrounding counties serve as Charlotte's bedroom counties.

For all three scenarios, it is unexpected that the Raleigh region does not emerge as a meaningful urban agglomeration in the study area. One possible explanation may be that the extent of this area is much larger than others, which would cause very high urban costs. In this sense, if the current Raleigh region were split into two smaller regions, the Raleigh region could have emerged as a major agglomeration. Second, compared to Charlotte region, a competing agglomeration, Raleigh region has few supporting regions surrounding it. And internally, Raleigh is the single dominant center in the region without other secondary supporting centers, which indicates weak internal economic linkages.

The third reason could be found again in the geography of transportation networks serving this region. Comparing to the connections between Charlotte, Columbia, and Greenville regions via interstates, Raleigh region is not well connected with other regions with both I85 and I95 passing by and connecting Richmond, VA and Washington D.C. in the far north. The above considerations also highlight the boundary effects on the spatial agglomerations created by the particular geographic regions (BEA areas) used in this application. Imagine that if the study area extends further southwest to include Atlanta region, Charlotte region may lose its agglomeration to Atlanta region. Similarly, if the D.C. region is included, a different pattern of spatial agglomeration may exhibit as well. On the other hand, note that county is a quite large geographic unit since, for example, the Mecklenburg County itself includes the city of Charlotte that may has multiple centers. Therefore, if smaller geographic units, such as Census block group, are used and correspondingly smaller regions are configured, it can be expected that the model may generate quite different internal and external spatial structures in the study area. These geographic dependencies and boundary effects warrant future studies. In addition, other possible explanations may relate to noneconomic factors that cannot be captured by this model such as hosting the state capital in Raleigh.

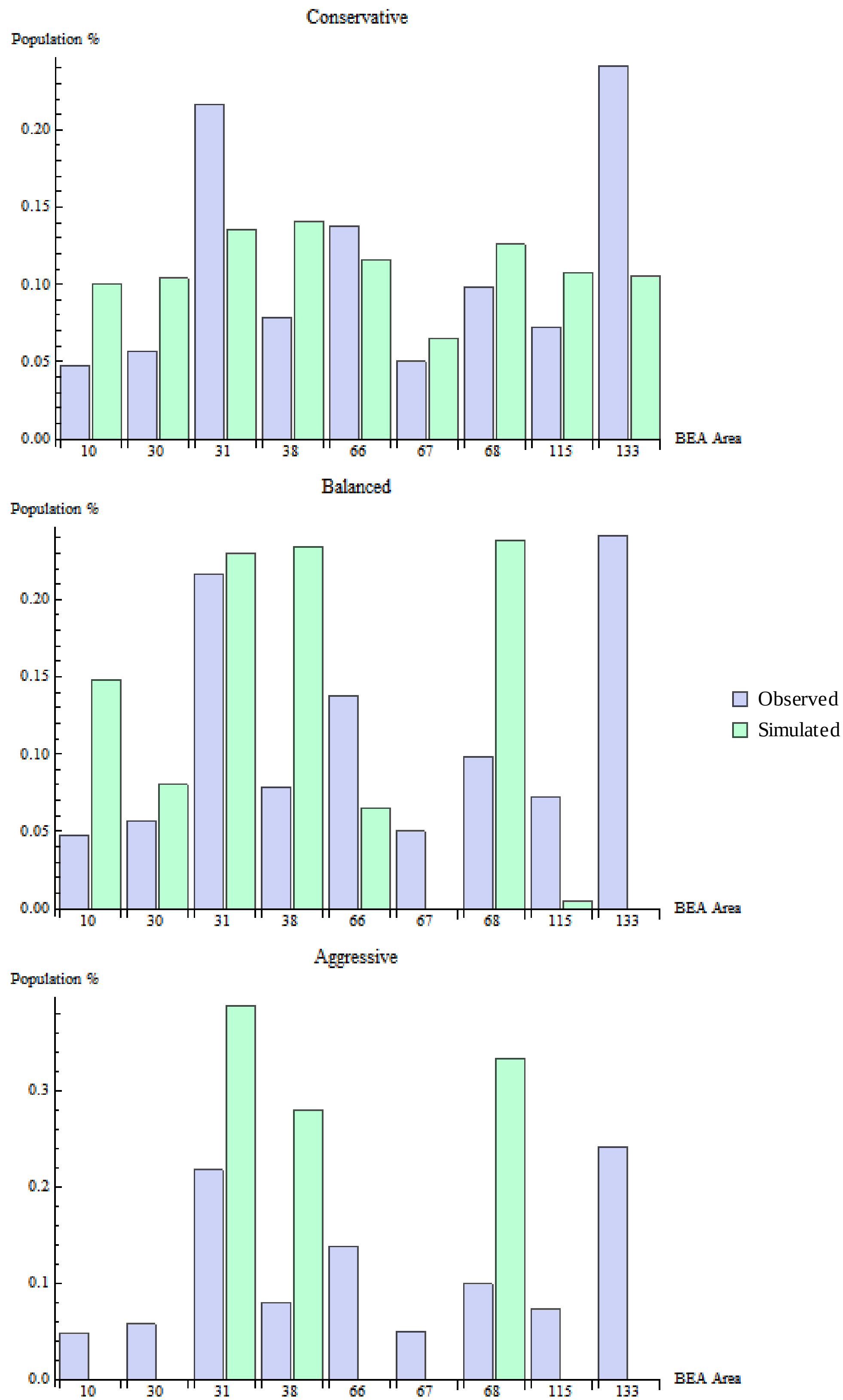


Figure 8.2: Simulated outputs by scenario compared against observed population distribution

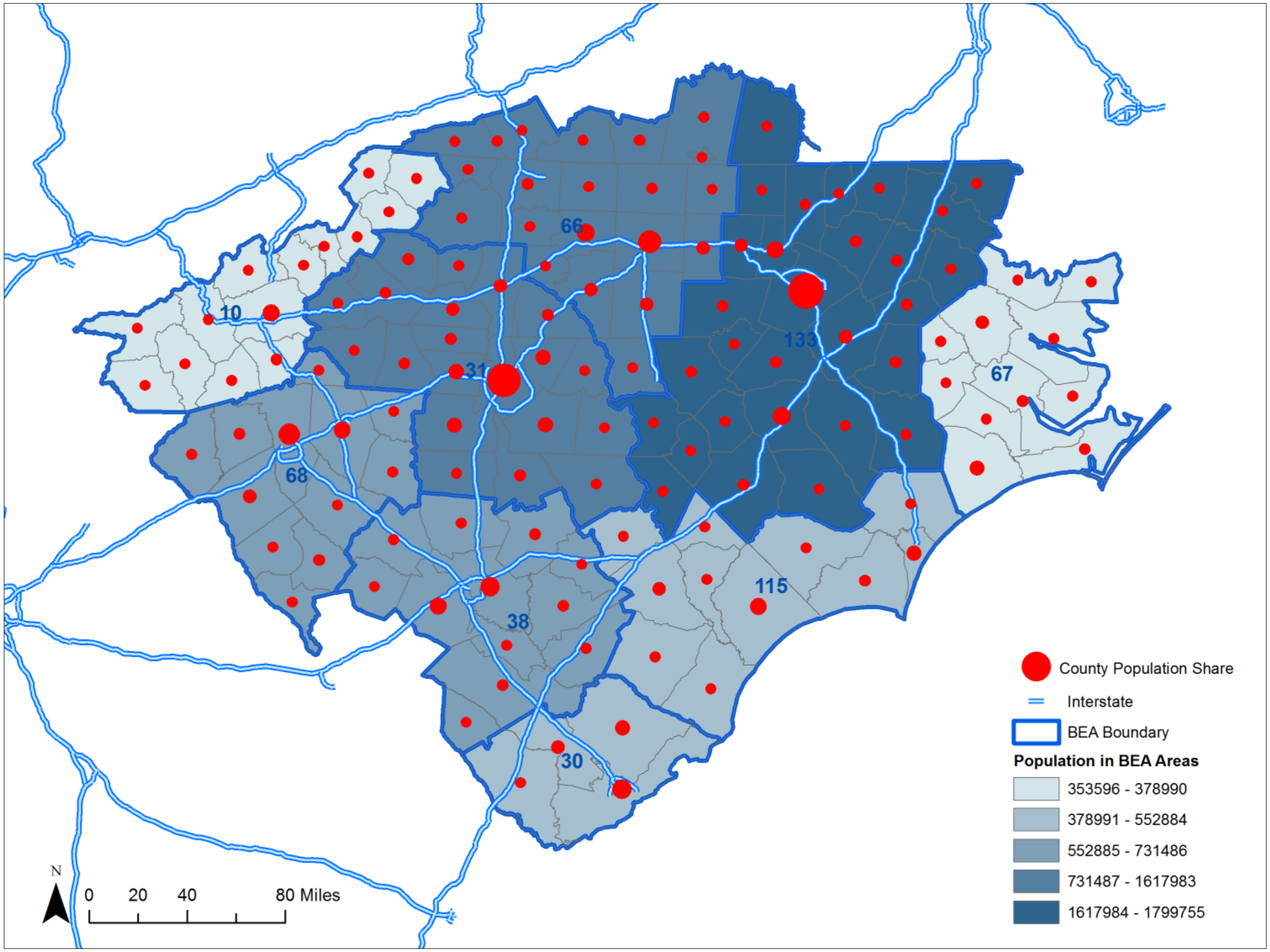


Figure 8.3: Observed population distribution in the study area for 2012

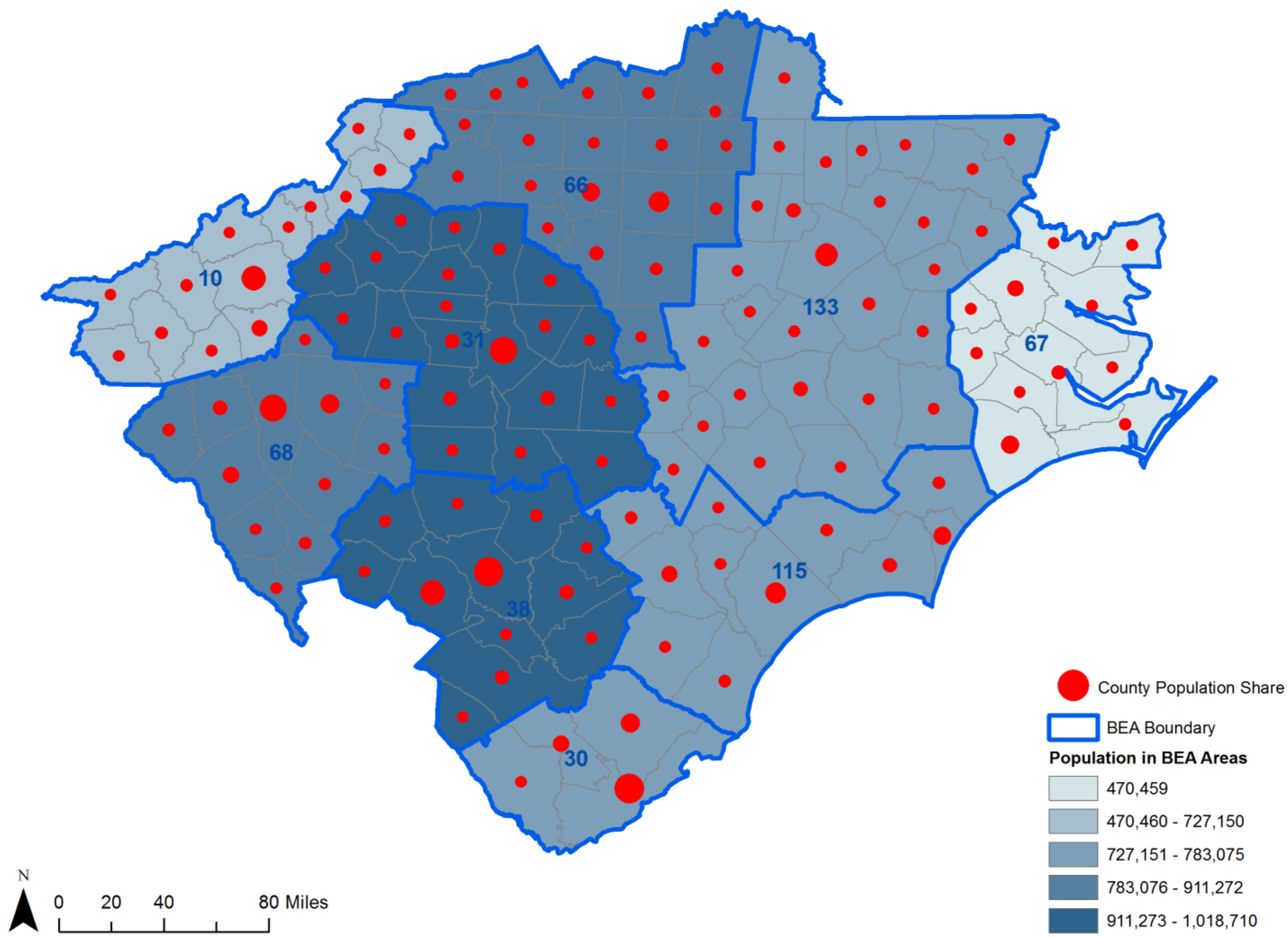


Figure 8.4: Simulated population distribution for the conservative scenario

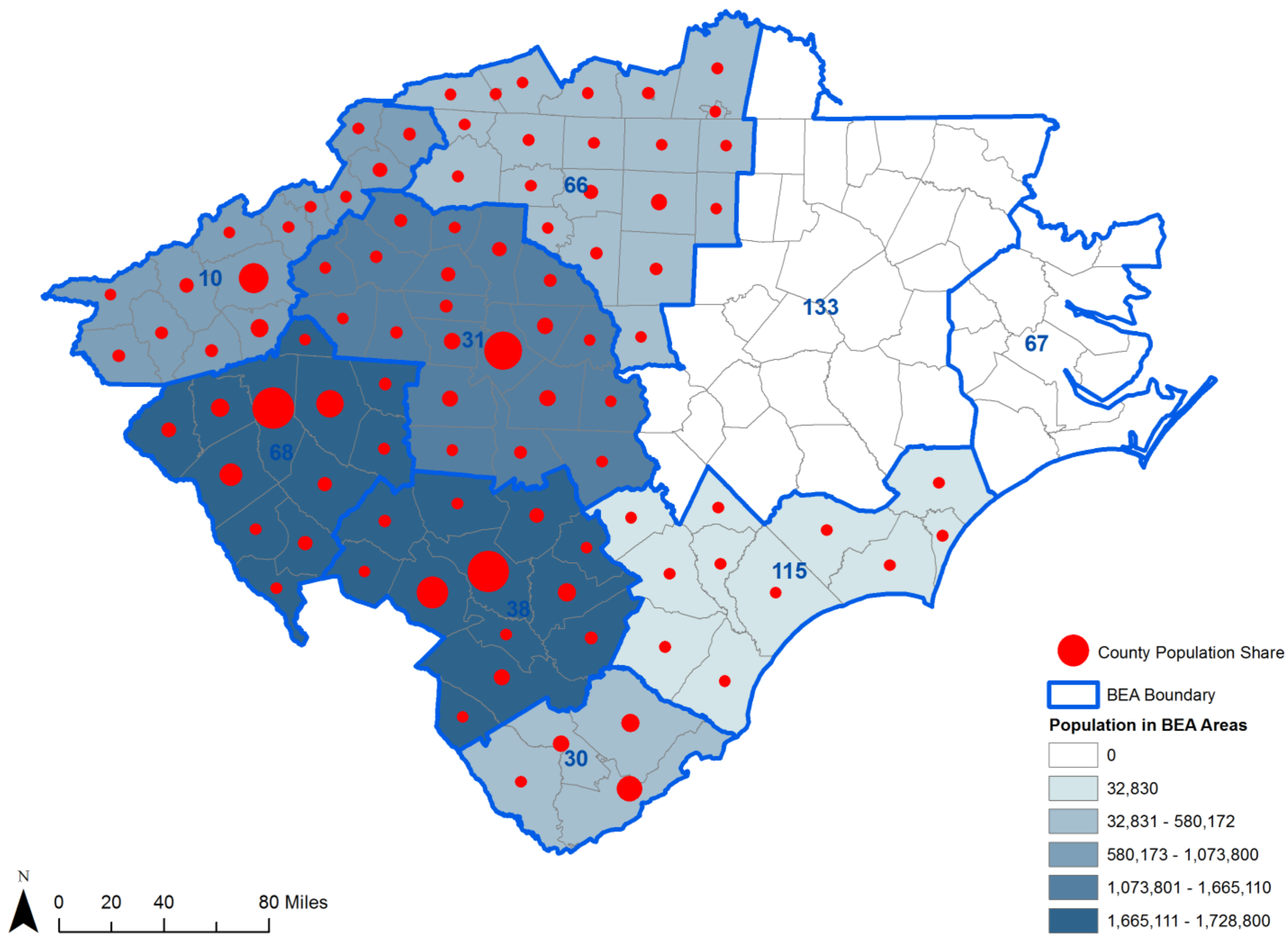


Figure 8.5: Simulated population distribution for the balanced scenario

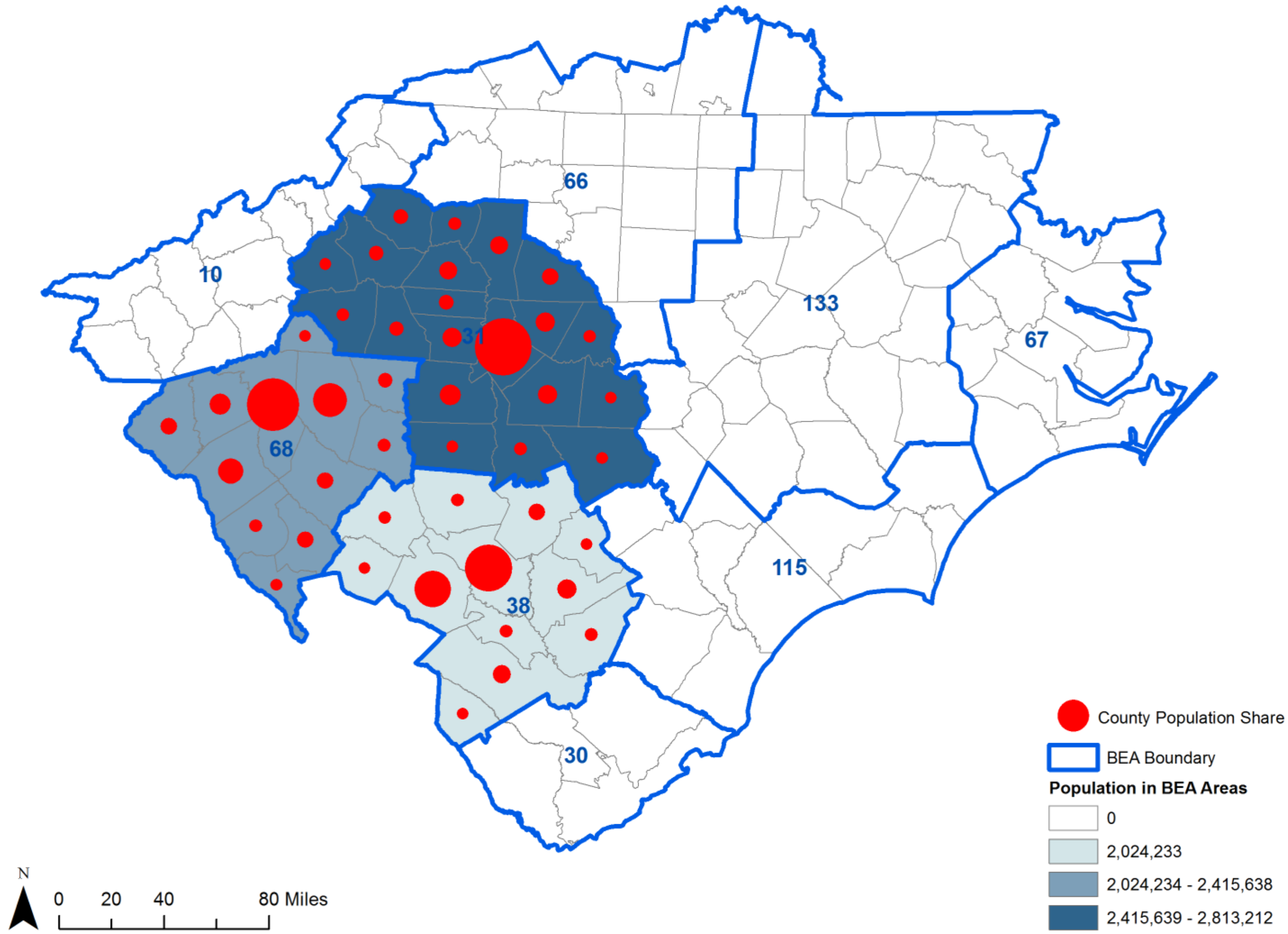


Figure 8.6: Simulated population distribution for the aggressive scenario

It is an interesting perspective to look at these scenarios in the order of conservative, balanced and aggressive as a series. It is clear to have a picture of agglomeration, first to the west, then to the three fully agglomerated regions in the aggressive scenario. If the three scenarios can be treated as the snapshots in an agglomeration process at the inter-region scale with the observed distribution as the starting point and with the spatial costs ($t, \tau,$ and T) changes given by each scenario (specified in Table 8.3), it can help to better understand why the agglomeration occurs and why it occurs that way. The whole process is affected by the change of three types of spatial costs: increasing commuting costs, increasing intra-urban trade costs and decreasing inter-urban trade costs (from conservative to aggressive scenario in Table 8.3). Once the local spatial costs (commuting and internal trade) start to increase, larger regions will bear more cost increase because of the different size of regions. The increasing urban costs push population and firms out of the large regions in size, such as the Charlotte region and the Raleigh region, to nearby smaller regions having lower urban costs but good access to large markets, which are in the south and west. And the decreasing interregional trade costs reinforce this process by allowing cheaper trading with remote places. It explains the fairly dispersed distribution in the conservative scenario. As the costs internal to regions continue increasing, the largest region (Raleigh) in size become the first having its entire population lost. Having this large market close by gone, population and firms in the Jacksonville region on the east side face a high disutility by being remote to other regions of large markets. This geographic disadvantage makes its population also leave this region for other places. And their destination are still those smaller regions having good access to large markets. This is why I see two empty regions to the east (133: Raleigh, 67: Jacksonville) and highly concentrated regions (68: Greenville, 38: Columbia) to the southwest. At this time, the concentrations at the three region cluster (Charlotte, Greenville and Columbia) constitute a huge market attracting all populations in the surrounding area. On the other hand, to enjoy this largest market and to compensate on the ever increasing urban costs, populations and firms in the surrounding regions move to the cluster of three regions to minimize the interregional trade costs as much as possible since the three regions are adjacent to each other and have good transportation facilities. This results in the outcome in the aggressive scenario.

Therefore, the partially calibrated models for each scenario actually give the possible long-run equilibria simulated with the observed data as the initial condition and with the influence of spatial cost factors as model parameters. In this sense, instead of searching for the best fit scenario, these scenarios can be linked in a meaningful way to help to understand where the observed situation stands towards the possible long-run equilibria of urban regional development.

CHAPTER 9: CONCLUSIONS

The objective of this dissertation is twofold. On the theoretical side, the goal is to establish a multiscalar modeling framework of economic agglomerations that accounts for the interdependency of spatial structures at intra- and inter-urban scales within the extent of a megaregion. On the methodological side, the goal is to extend the proposed theoretical model into a two-dimensional geographic model that accounts for the heterogeneity of space and can be computational scalable to be applied to real world applications. I summarize the mains insights of this research as follows.

To approach the first objective, I propose a theoretical framework that takes account of market linkages in terms of agglomeration and dispersion forces, various types of spatial costs and the interplay between them across scales. The main idea is to develop a general framework by integrating theories from New Economic Geography and Urban Economics. The resulting theoretical models are characterized by general equilibrium, pecuniary externalities of agglomeration economies, urban costs, scale-dependent transport costs of goods, and endogenous spatial structures (in terms of polycentricity and monocentricity) at both intra- and inter-urban scales. This framework highlights the cross-scale interactions of various forces and factors, thus it enables the construction of models at both intra- and inter-urban scales.

Chapter 4 develops a polycentric urban model that extends existing models in two aspects: 1) incorporating market linkages for both final and intermediate goods; 2) modeling labor market linkages with commuting costs between residence and workplace locations. Numerical results have proved our expectation that these extensions allow the model to generate both monocentric and polycentric structures under specific circumstances. The main findings revealed by the study of short-run equilibrium are threefold. First, commuting costs and internal trade costs exhibit different behaviors when they affect agglomeration and dispersion forces of market linkages. Commuting costs strengthen (weaken) the dispersion forces and weaken (strengthen) the agglomeration forces as they increase (decrease). Internal trade costs first strengthen the agglomeration force, and then reinforce the dispersion forces, as they decrease. This features a bell-shaped curve of spreading-agglomeration-spreading process that has been only found in the literature of NEG models. This is the first to build this mechanism into urban economics models. Second, by incorporating input-output linkages between firms into an urban model I found that it exhibits complex behaviors characterized by multiplicity of equilibria and stronger decentralization forces to secondary employment centers, which are absent in existing polycentric urban models. Third, I leverage numerical simulations to tackle the intractability of non-linear models. Although model behaviors cannot be analytically proved in general, specific conditions and their stability can be numerically tested with given parameters.

Chapter 5 builds the multiscalar urban systems model to account for the interdependency of spatial structures across scales. It is achieved by nesting a set of polycentric urban models (Section 4.2) within racetrack urban systems characterized by inter-urban trading in final and intermediate goods markets. The racetrack configuration of space allows a more general approach to model any number of cities than existing two-region models. Together, these two features make our model the first of this type. The interdependency of spatial structures across scales has been illustrated by examining the impact of the exogenous change of intra- and inter-urban spatial costs, respectively. Specifically, the evolutionary paths of equilibrium transition pertaining to the cross-scale interdependency of spatial structures are characterized by 10 cases extracted from the simulation results. These findings allow us to safely conclude that causal relations have been successfully established by the model between the interplay of market linkages and spatial costs across scales and the interdependency of spatial structures across scales. This conclusion fulfills our first objective that highlights the importance of scale dependency in understanding the formation of certain spatial structures at any spatial scale. In a word, scale matters in the spatiotemporal organization of urban and regional economies. This can have significant practical policy implications in that advocating a spatial structure at one scale may affect the spatial structure at another. For example, development of secondary business centers in a city may foster further agglomeration to the city from other regions.

Chapter 6 extends the theoretical models into a geographic model with real geography characterized by a two-dimensional discrete space and heterogeneous locations. This is achieved by utilizing a graph representation of geographic locations and distance. To transform the conventional continuous space setting in economic modeling into a graph-based discrete space setting, an agent-based discrete choice model is employed to determine the residence locations and workplaces for households and the business locations for firms. It further generalizes the model to be able to accommodate any number of industries. This approach iteratively approximates the equilibrium conditions that are equivalent to those in theoretical models. The geographic model has been tested in an artificial two-dimensional space with two industries. The simulation results confirm the main findings revealed in the theoretical models and prove their consistency. This demonstration highlights the importance of the geographic model in terms of carrying over the model mechanisms from the theoretical models and enabling their applications in empirical circumstances for validity assessment. In addition, the geographic model also contributes to the urban modeling literature by incorporating the economic agglomeration effect of NEG style into agent-based location choice.

Chapter 7 implements the parallel models of urban systems by leveraging heterogeneous HPC platforms to tackle the computational complexity involved in the modeling efforts. The parallel models take full advantage of a powerful, robust and flexible platform (MATHEMATICA) with native capability to support for parallel computing on heterogeneous HPC platform. Therefore, this work readily contributes to the computational aspect of the existing large-scale urban and economic modeling.

Chapter 8 assesses the applicability of the geographic model developed in Chapter 6 with a real world case study. To serve this purpose, a partial calibration procedure is used to handle the limitations in data availability. A simple demonstration is presented for the Carolinas megaregion. Three calibration scenarios are discussed regarding their fitness to observed data. In addition, I offer a perspective to link these scenarios in a meaningful way to reflect the long-run equilibria the observed urbanization pattern may approach and in turn to provide insights into existing urban regional developments. Since the model is not fully calibrated, its prediction power with regard to planning scenarios is limited. Its usefulness in policy making and scenario testing would require careful calibration on quality and complete data.

There exist limitations in this research work. However, limitations may lead to interesting future works. I discuss several of them here and provide some perspectives. First, regarding the theoretical model there are three points in order. 1) Rural areas related to agriculture sector and farmers are not considered in this model. Since rural-to-urban migration and the trading of agriculture product are so important in many developing countries, these factors cannot be missed. The unproduced public goods in current model specification can be quite easily modified to incorporate agriculture sector and farmers. And farmers can be treated as low skill workers in the labor market. 2) Numerical simulation has advantages in solving non-linear problems, but traversing the entire parameter space is often impossible to have a full knowledge of the model behavior. This is why economists prefer tractable models. Given the complexity of our models, more parameter space needs to be explored in the future. 3) The interdependency of spatial structures across scales suggested by our models is the opposite side of Zipf's law that indicates scale independence and similar fractal dimensions across scales. Since Zipf's law has been tested on a number of urban systems, it is worth our attention in future to study why and in what circumstances it contradicts our findings.

Second, regarding our geographic model, two aspects are discussed here. 1) The geographic model currently employs a graph-based discrete space in two dimensions. The accommodation of two-dimensional space can be extended to a continuous representation such as a cellular environment. This extension would allow our model to be applied to a wide range of applications involving environmental components. 2) I have severe limitations in model calibration. Data availability is certainly a problem. Given the availability of data, model specification may also be a concern because the model is calibrated on such a vast region where spatial heterogeneity may cause structural parameters to vary across space. This would require calibrating models for each smaller and homogenous area. In addition, an arguable issue that if an equilibrium model can be calibrated on data that may not reflect the equilibrium state. These points are worth our further investigation in future studies.

REFERENCES

Abdel-Rahman, H. M., and A. Anas. 2004. Theories of systems of cities. In *Handbook of Regional and Urban Economics*, eds. J. V. Henderson and J.-F. Thisse, 2293–2339. Amsterdam: Elsevier B.V.

Akamatsu, T., Y. Takayama, and K. Ikeda. 2012. Spatial discounting, Fourier, and racetrack economy: a recipe for the analysis of spatial agglomeration models. *Journal of Economic Dynamics and Control* 36 (11):1729–1759.

Allen, P. M., and M. Sanglier. 1981. A Dynamic Model of a Central Place System - II. *Geographical Analysis* 13 (2):149–164.

Allen, P. M., and M. Sanglier. 1979. A Dynumic Model of Growth in a Central Place System. *Geographical Analysis* 11 (3):256–272.

Alonso, W. 1964. *Location and Land Use*. Cambridge, MA and London, England: Harvard University Press.

Anas, A. 1983. Discrete choice theory, information theory and the multinomial logit and gravity models. *Transportation Research Part B: Methodological* 17 (I):13–23.

———. 2004. Vanishing cities: what does the new economic geography imply about the efficiency of urbanization? *Journal of Economic Geography* 4 (2):181–199.

Anas, A., R. Arnott, and K. Small. 1998. Urban spatial structure. *Journal of Economic Literature* 36 (3):1426–1464.

Anas, A., and I. Kim. 1996. General equilibrium models of polycentric urban land use with endogenous congestion and job agglomeration. *Journal of Urban Economics* 40 (2):232–256.

Anas, A., and K. Xiong. 2003. Intercity trade and the industrial diversification of cities. *Journal of Urban Economics* 54 (2):258–276.

Anas, A., and K. Xiong. 2005. The formation and growth of specialized cities: efficiency without developers or Malthusian traps. *Regional Science and Urban Economics* 35 (4):445–470.

Anas, A., and R. Xu. 1999. Congestion, land use, and job dispersion: a general equilibrium model. *Journal of Urban Economics* 45 (3):451–473.

Anas, A., and L. Yu. 2007. A regional economy, land use, and transportation model (RELU-TRAN): formulation, algorithm design, and testing. *Journal of Regional Science* 47 (3):41p.

Armstrong, M. P. 2000. Geography and computational science. *Annals of the Association of American Geographers* 90 (1):146–156.

Arrow, K. J., and G. Debreu. 1954. Existence of an equilibrium for a competitive economy. *Econometrica* 22 (3):265.

Awaludin, A., and D. Chen. 2007. *UrbanSim Parallel Programming Capstone Paper*.

Baldwin, R., R. Forslid, P. Martin, G. Ottaviano, et al. 2003. *Economic Geography and Public Policy*. Princeton, NJ: Princeton University Press.

Batty, M. 1994. A chronicle of scientific planning: the Anglo-American modeling experience. *Journal of the American Planning Association* 60 (1):7–16.

———. 1976. *Urban Modelling: Algorithms, Calibrations, Predictions*. New York: Cambridge University Press.

Berliant, M., and P. Wang. 2008. Urban growth and subcenter formation: a trolley ride from the Staples Center to Disneyland and the Rose Bowl. *Journal of Urban Economics* 63 (2):679–693.

Berry, B. J. L. 2002. Paradigm lost. *Urban Geography* 23 (5):441–445.

Birkin, M., M. Clarke, and F. George. 1995. The use of parallel computers to solve nonlinear spatial optimization problems - an application to network planning. *Environment and Planning A* 27 (7):1049–1068.

Bosker, M., and H. Garretsen. 2007. Trade costs, market access and economic geography: why the empirical specification of trade costs matters. *CEPR Discussion Paper 2071* :1–41.

Brakman, S., H. Garretsen, and M. Schramm. 2006. Putting new economic geography to the test: free-ness of trade and agglomeration in the EU regions. *Regional Science and Urban Economics* 36 (5):613–635.

Briassoulis, H. 2000. Analysis of land use change: theoretical and modeling approaches. http://www.rri.wvu.edu/WebBook/Briassoulis/contents.htm (last accessed 2 January 2013).

Brown, D. G. et al. 2006. Land-use and land-cover change. In *Land-Use and Land-Cover Change,* Global Change - The IGBP Series., eds. E. F. Lambin and H. Geist, 117–135. Berlin, Heidelberg: Springer Berlin Heidelberg.

Cavailhès, J., C. Gaigné, T. Tabuchi, and J. F. Thisse. 2007. Trade and the structure of cities. *Journal of Urban Economics* 62 (3):383–404.

Cetin, N., K. Nagel, B. Raney, and A. Voellmy. 2002. Large-scale multi-agent transportation simulations. *Computer Physics Communications* 147 (1-2):559–564.

Chamberlin, E. H. 1969. *The Theory of Monopolistic Competition*. Cambridge, MA: Harvard University Press.

Chapin, F. S. 1974. *Human Activity Patterns in the City*. New York, NY: Wiley.

Christaller, W. 1966. *Central Places in Southern Germany*. Jena, Germany: Fischer.

Clematis, A., B. Falcidieno, and M. Spagnuolo. 1996. Parallel processing on heterogeneous networks for GIS applications. *International Journal of Geographical Information Systems* 10 (6):747–767.

Clematis, A., M. Mineter, and R. Marciano. 2003. High performance computing with geographical data. *Parallel Computing* 29 (10):1275–1279.

Dattilo, G., and G. Spezzano. 2003. Simulation of a cellular landslide model with CAMELOT on high performance computers. *Parallel Computing* 29 (10):1403–1418.

Davy, J., and W. Essah. 1998. Generating parallel applications of spatial interaction models. In *Euro-Par'98 Parallel Processing,* 136–145.

Ding, Y. Y. Y., and P. J. P. Densham. 1996. Spatial strategies for parallel spatial modelling. *International Journal of Geographical Information Systems* 10 (6):669–698.

Dixit, A. K., and J. E. Stiglitz. 1977. Monopolistic competition and optimum product diversity. *American Economic Review* 67 (3):297–308.

Forrester, J. . 1969. *Urban Dynamics*. Cambridge, MA: The MIT Press.

Fried, M., J. Havens, and M. Thall. 1977. *Travel Behavior-a Synthesized Theory*. Washington.

Fujita, M. 1988. A monopolistic competition model of spatial agglomeration. *Regional Science and Urban Economics* 18 (1):87–124.

———. 1990. Spatial interactions and agglomeration in urban economics. In *New Frontiers in Regional Science,* eds. M. Chatterji and R. E. Kuenne, 184–221. New York: New York University Press.

Fujita, M. 1989. *Urban Economic Theory*. Cambridge: Cambridge University Press.

Fujita, M., and P. Krugman. 1995. When is the economy monocentric?: von Thünen and Chamberlin unified. *Regional Science and Urban Economics* 25 (4):505–528.

Fujita, M., P. Krugman, and T. Mori. 1999. On the evolution of hierarchical urban systems. *European Economic Review* 43 (2):209–251.

Fujita, M., P. Krugman, and A. J. Venables. 1999. *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge, MA: MIT Press.

Fujita, M., and T. Mori. 1997. Structural stability and evolution of urban systems. *Regional Science and Urban Economics* 27 (4-5):399–442.

Fujita, M., and H. Ogawa. 1982. Multiple equilibria and structural transition of non-monocentric urban configurations. *Regional Science and Urban Economics* 12 (2):161–196.

Fujita, M., and T. E. Smith. 1990. Additive-interaction models of spatial agglomeration. *Journal of Regional Science* 30 (1):51–74.

Fujita, M., and J.-F. Thisse. 2002. *Economics of Agglomeration*. Cambridge: Cambridge University Press.

Gabszewicz, J. J., and J.-F. Thisse. 1992. Chapter 9 Location. In *Handbook of Game Theory with Economic Applications*, eds. R. J. Aumann and S. Hart, 281–304. Amsterdam: North-Holland.

Gaigné, C., S. Riou, and J.-F. Thisse. 2012. Are compact cities environmentally friendly? *Journal of Urban Economics* 72 (2–3):123–136.

Garin, R. A. 1966. Research note: a matrix formulation of the Lowry model for intra-metropolitan activity allocation. *Journal of the American Institute of Planners* 32 (6):361–364.

Gong, Z., W. Tang, D. A. Bennett, and J.-C. Thill. 2013. Parallel agent-based simulation of individual-level spatial interactions within a multicore computing environment. *International Journal of Geographical Information Science* 27 (6):1152–1170.

Gong, Z., W. Tang, and J. Thill. 2012. Parallelization of ensemble neural networks for spatial land-use modeling. In *Proceedings of the 5th International Workshop on Location-Based Social Networks - LBSN '12*, 48. New York, New York, USA: ACM Press.

Gong, Z., J.-C. Thill, and W. Liu. 2015. ART-P-MAP neural networks modeling of land-use change: accounting for spatial heterogeneity and uncertainty. *Geographical Analysis* 47 (4):376–409.

Gordon, P., H. W. Richardson, and H. L. Wong. 1986. The distribution of population and employment in a polycentric city: the case of Los Angeles. *Environment and Planning A* 18 (2):161–173.

Gottmann, J. 1957. Megalopolis or the urbanization of the northeastern seaboard. *Economic Geography* 33 (3):189.

Guan, Q. 2008. Parallel algorithms for geographic processing. *ProQuest*.

Guan, Q., and K. C. Clarke. 2010. A general-purpose parallel raster processing programming library test application using a geographic cellular automata model. *International Journal of Geographical Information Science* 24 (5):695–722.

Guan, Q., P. C. Kyriakidis, and M. F. Goodchild. 2011. A parallel computing approach to fast geostatistical areal interpolation. *International Journal of Geographical Information Science* 25 (8):1241–1267.

H. Hanson, G. 2005. Market potential, increasing returns and geographic concentration. *Journal of International Economics* 67 (1):1–24.

Hägerstrand, T. 1970. What about people in Regional Science? *Papers of the Regional Science Association* 24 (1):6–21.

Hall, P. 1997. Modelling the post-industrial city. *Futures* 29 (4-5):311–322.

Hall, P., and K. Pain. 2006. *The Polycentric Metropolis*. London: Routledge.

Harris, B. 1985. Some notes on parallel computing: with special reference to transportation and land-use modeling. *Environment and Planning A* 17 (9):1275–1278.

Harris, B., and A. G. Wilson. 1978. Equilibrium values and dynamics of attractiveness terms in production-constrained spatial-interaction models. *Environment and Planning A* 10 (4):371 – 388.

Helsley, R. W., and A. M. Sullivan. 1991. Urban subcenter formation. *Regional Science and Urban Economics* 21 (2):255–275.

Henderson, J. 1991. *Urban Development: Theory, Fact, and Illusion*. New York, NY: Oxford University Press.

———. 1974. The sizes and types of cities. *American Economic Review* 64 (4):640–656.

Henderson, J. V. 1987. General equilibrium modeling of systems of cities. In *Handbook of Regional and Urban Economics*, ed. E. S.Mills, 927–956. Amsterdam: North-Holland.

Henderson, J. V. 1972. Hierarchy models of city size: an economic evaluation. *Journal of Regional Science* 12 (3):435–441.

Henderson, V., and J. Thisse. 2004. *Handbook of Regional and Urban Economics: Cities and Geography* eds. J. V. Henderson and J.-F. Thisse. Amsterdam: Elsevier B.V.

Hribar, M. R. M., V. E. V. Taylor, and D. D. E. DE Boyce. 2001. Implementing parallel shortest path for parallel transportation applications. *Parallel Computing* 27 (12):1537–1568.

Ikeda, K., T. Akamatsu, and T. Kono. 2012. Spatial period-doubling agglomeration of a core-periphery model with a system of cities. *Journal of Economic Dynamics and Control* 36 (5):754–778.

Irwin, E. G., and J. Geoghegan. 2001. Theory, data, methods: developing spatially explicit economic models of land use change. *Agriculture, Ecosystems and Environment* 85 (1–3):7–23.

Johnson, B. K. P., and J. R. Kort. 2004. *2004 Redefinition of the BEA Economic Areas*. Bureau of Economic Analysis.

Kloosterman, R., and S. Musterd. 2001. The polycentric urban region: towards a research agenda. *Urban Studies* 38 (4):623–633.

Knaap, T. 2006. Trade, location, and wages in the United States. *Regional Science and Urban Economics* 36 (5):595–612.

Kranz, D., K. Johnson, A. Agarwal, J. Kubiatowicz, et al. 1993. Integrating message-passing and shared-memory. In *Proceedings of the Fourth ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming - PPOPP '93*, 54–63. New York, New York, USA: ACM Press.

Krugman, P. 1991. Increasing returns and economic geography. *Journal of Political Economy* 99 (3):483–499.

———. 1993. On the number and location of cities. *European Economic Review* 37 (2-3):293–298.

Krugman, P. 1980. Scale economies, product differentiation, and the pattern of trade. *The American Economic Review* 70 (5):950–959.

Krugman, P. 1996. *The Self-Organizing Economy*. Cambridge, MA: Blackwell Publishers.

Lanthier, M., D. Nussbaum, and J. g-R. iger Sack. 2003. Parallel implementation of geometric shortest path algorithms. *Parallel Computing* 29 (10):1445–1479.

Li, X., X. Zhang, A. Yeh, and X. Liu. 2010. Parallel cellular automata for large-scale urban simulation using load-balancing techniques. *International Journal of Geographical Information Science* 24 (6):803–820.

Low, N. 1975. Centrism and the provision of services in residential areas. *Urban Studies* 12 (2):177–191.

Lowry, I. S. 1964. *A Model of Metropolis*. Santa Monica, CA: RAND Corporation.

Lucas, R. E., and E. Rossi-Hansberg. 2002. On the internal structure of cities. *Econometrica* 70 (4):1445–1476.

Martinez, F. 2000. Towards a land-use and transport interaction framework. In *Handbook of Transport Modelling*, eds. D. Hensher and K. Button, 145. New York, NY: Elsevier Science.

Meentemeyer, R. K. et al. 2013. FUTURES: multilevel simulations of emerging urban–rural landscape structure using a stochastic patch-growing algorithm. *Annals of the Association of American Geographers* 103 (4):785–807.

Miller, E. J., J. Douglas Hunt, J. E. Abraham, and P. A. Salvini. 2004. Microsimulating urban systems. *Computers, Environment and Urban Systems* 28 (1-2):9–44.

Mills, E. S. 1967. An aggregative model of resource allocation in a metropolitan area. *The American Economic Review* 57 (2):197–210.

Mori, T. 1997. A modeling of megalopolis formation: the maturing of city systems. *Journal of Urban Economics* 42 (1):133–157.

Mossay, P. 2003. Increasing returns and heterogeneity in a spatial economy. *Regional Science and Urban Economics* 33 (4):419–444.

Mulligan, G. F. 1984. Agglomeration and central place theory: a review of the literature. *International Regional Science Review* 9 (1):1–42.

Murata, Y., and J.-F. Thisse. 2005. A simple model of economic geography à la Helpman–Tabuchi. *Journal of Urban Economics* 58 (1):137–155.

Muth, R. F. 1969. *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*. Chicago: University of Chicago Press.

Nagel, K., R. Beckman, and C. Barrett. 1999. TRANSIMS for urban planning. In *6th International Conference on Computers in Urban Planning and Urban Management*. Venice, Italy.

Nagel, K., and M. Rickert. 2001. Parallel implementation of the TRANSIMS micro-simulation. *Parallel Computing* 27 (12):1611–1639.

NetLogoTeam. 2014. NetLogo. http://ccl.northwestern.edu/netlogo/.

O'Neill, P. 2003. Global city-regions: trends, theory, policy. *Area* 35 (3):326–327.

ODOT. 2010. *Oregon Statewide Integrated Model (SWIM2) Model Description*.

———. 2001. *Transportation and Land Use Model Integration Program : Overview of the First Generation Models*.

Ogawa, H., and M. Fujitat. 1980. Equilibrium land use patterns in a nonmonocentric city. *Journal of Regional Science* 20 (4):455–475.

Openshaw, S. 1987. Research policy and review 17. Some applications of supercomputes in urban and regional analysis and modelling. *Environment and Planning A* 19 (7):853–860.

Openshaw, S., and I. Turton. 2000. *High Performance Computing and the Art of Parallel Programming: An Introduction for Geographers, Social Scientists and Engineers*. New York, NY: Routledge.

Ottaviano, G., T. Tabuchi, and J. F. Thisse. 2002. Agglomeration and trade revisited. *International Economic Review* 43 (2):409–435.

Owens, J. D. et al. 2008. GPU computing. *Proceedings of the IEEE* 96 (5):879–899.

Picard, P. M., and T. Tabuchi. 2013. On microfoundations of the city. *Journal of Economic Theory* 148 (6):2561–2582.

———. 2008. Self-organized agglomerations and transport costs. *Economic Theory* 42 (3):565–589.

Pijanowski, B. C. et al. 2014. A big data urban growth simulation at a national scale: configuring the GIS and neural network based Land Transformation Model to run in a High Performance Computing (HPC) environment. *Environmental Modelling and Software* 51:250–268.

Porta, J. et al. 2013. High performance genetic algorithm for land use planning. *Computers, Environment and Urban Systems* 37 (1):45–58.

Pred, A. 1966. *The Spatial Dynamics of U.S. Urban-Industrial Growth*. Cambridge, MA: The MIT Press.

Pred, A. R. 1977. *City Systems in Advanced Economies*. London: Hutchinson.

Redding, S., and A. J. Venables. 2004. Economic geography and international inequality. *Journal of International Economics* 62 (1):53–82.

Regional Plan Association. 2006. *America 2050: A prospectus*. New York: Regional Plan Association.

RepastTeam. 2014. Repast for High Performance Computing. http://repast.sourceforge.net/.

Richardson, H. W. 1988. Monocentric vs. policentric models: the future of urban economics in regional science. *The Annals of Regional Science* 22 (2):1–12.

Rickert, M., and K. Nagel. 2001. Dynamic traffic assignment on parallel computers in TRANSIMS. *Future Generation Computer Systems* 17 (5):637–648.

van Schrojenstein Lantman, J., P. Verburg, A. Bregt, and S. Geertman. 2011. *Land-Use Modelling in Planning Practice* eds. E. Koomen and J. Borsboom-van Beurden. Dordrecht: Springer Netherlands.

Scitovsky, T. 1954. Two concepts of external economies. *The Journal of Political Economy* 62 (2):143–151.

Shook, E., S. Wang, and W. Tang. 2013. A communication-aware framework for parallel spatially explicit agent-based models. *International Journal of Geographical Information Science* 27 (11):2160–2181.

Smith, T. R., G. Peng, and P. Gahinet. 1989. Asynchronous, iterative, and parallel procedures for solving the weighted-region least cost path problem. *Geographical Analysis* 21 (2):147–166.

Spence, M. 1976. Product selection, fixed costs, and monopolistic competition. *The Review of Economic Studies* 43 (2):217–235.

Standish, R. K. 2014. EcoLab. http://ecolab.sourceforge.net.

Starrett, D. A. 1974. Principles of optimal location in a large homogeneous area. *Journal of Economic Theory* 9 (4):418–448.

Sullivan, A. M. 1986. A general equilibrium model with agglomerative economies and decentralized employment. *Journal of Urban Economics* 20 (1):55–74.

Tabuchi, T. 1998. Urban agglomeration and dispersion: a synthesis of Alonso and Krugman. *Journal of Urban Economics* 44 (3):333–351.

Tabuchi, T., and J.-F. Thisse. 2011. A new economic geography model of central places. *Journal of Urban Economics* 69 (2):240–252.

———. 2006. Regional specialization, urban hierarchy, and commuting costs. *International Economic Review* 47 (4):1295–1317.

Tabuchi, T., J.-F. Thisse, and D.-Z. Zeng. 2005. On the number and size of cities. *Journal of Economic Geography* 5 (4):423–448.

Tang, W. 2013. Parallel construction of large circular cartograms using graphics processing units. *International Journal of Geographical Information Science* 27 (11):2182–2206.

Tang, W., and D. a. Bennett. 2012. Parallel agent-based modeling of spatial opinion diffusion accelerated using graphics processing units. *Ecological Modelling* 229 (19):108–118.

Tang, W., D. A. Bennett, and S. Wang. 2011. A parallel agent-based model of land use opinions. *Journal of Land Use Science* 6 (2-3):121–135.

Tang, W., and M. Jia. 2014. Global sensitivity analysis of a large agent-based model of spatial opinion exchange: a heterogeneous multi-GPU acceleration approach. *Annals of the Association of American Geographers* 104 (3):485–509.

Tesfatsion, L. 2006. Chapter 16 Agent-based computational economics: a constructive approach to economic theory. In *Handbook of Computational Economics*, eds. L. Tesfatsion and K. L. Judd, 831–880. New York, NY: Elsevier B.V.

Thill, J., and A. Wheeler. 1995. On chaos, continuous-time and discrete-time models of spatial system dynamics. *Geographical Systems* 2:121–130.

Thisse, J.-F. 2010. Toward a unified theory of economic geography and urban economics. *Journal of Regional Science* 50 (1):281–296.

von Thünen, J. H. 1966. *Isolated State: An English Edition of Der Isolierte Staat*. New York: Pergamom Press.

TRACC. 2011. Advanced Evacuation Modeling Using TRANSIMS. (September):2011. https://www.tracc.anl.gov/index.php/evacuation-planning.

United Nations. 2011. *Population Distribution, Urbanization, Internal Migration and Development: An International Perspective*.

US Census Bureau. 2010. 2010 Geographic Terms and Concepts - Core Based Statistical Areas and Related Statistical Areas. http://www.census.gov/geo/reference/gtc/gtc_cbsa.html.

Verburg, P. H., P. P. Schot, M. J. Dijst, and A. Veldkamp. 2004. Land use change modelling: current practice and research priorities. *GeoJournal* 61 (4):309–324.

Waddell, P. 2002. UrbanSim: modeling urban development for land use, transportation, and environmental planning. *Journal of the American Planning Association* 68 (3):297–314.

Wang, S. 2010. A cyberGIS framework for the synthesis of cyberinfrastructure, GIS, and spatial analysis. *Annals of the Association of American Geographers* 100 (3):535–557.

———. 2013. CyberGIS software: a synthetic review and integration roadmap. *International Journal of Geographical Information Science* 27 (11):2122–2145.

———. 2013. CyberGIS: blueprint for integrated and scalable geospatial software ecosystems. *International Journal of Geographical Information Science* 27 (11):2119–2121.

Wang, S., and M. P. Armstrong. 2003. A quadtree approach to domain decomposition for spatial interpolation in Grid computing environments. *Parallel Computing* 29 (10):1481–1504.

Wang, S., and M. P. Armstrong. 2009. A theoretical approach to the use of cyberinfrastructure in geographical analysis. *International Journal of Geographical Information Science* 23 (2):169–193.

Wang, S., and Y. Liu. 2009. TeraGrid GIScience gateway: bridging cyberinfrastructure and GIScience. *International Journal of Geographical Information Science* 23 (5):631–656.

Weber, A., and C. Friedrich. 1962. *Theory of the Location of Industries*. Chicago: University of Chicago Press.

Wegener, M. 1994. Operational Urban Models State of the Art. *Journal of the American Planning Association* 60 (1):17–29.

White, M. J. 1976. Firm suburbanization and urban subcenters. *Journal of Urban Economics* 3 (4):323–343.

———. 1988. Location choice and commuting behavior in cities with decentralized employment. *Journal of Urban Economics* 24 (2):129–152.

Widener, M. J., N. C. Crago, and J. Aldstadt. 2012. Developing a parallel computational implementation of AMOEBA. *International Journal of Geographical Information Science* 26 (9):1707–1723.

Wilkinson, B., and M. Allen. 2004. *Parallel Programming: Techniques and Applications Using Networked Workstations and Parallel Computers (2nd Edition)*. Prentice-Hall, Inc.

Wilson, A. 2010. Entropy in urban and regional modelling: retrospect and prospect. *Geographical Analysis* 42 (4):364–394.

Wilson, A. G. 1967. A statistical theory of spatial distribution models. *Transportation Research* 1 (3):253–269.

Wisten, M. B., and M. J. Smith. 1997. Distributed computation of dynamic traffic equilibria. *Transportation Research Part C: Emerging Technologies* 5 (2):77–93.

Wong, S. C., C. K. Wong, and C. O. Tong. 2001. A parallelized genetic algorithm for the calibration of Lowry model. *Parallel Computing* 27 (12):1523–1536.

Xie, J., C. Yang, B. Zhou, and Q. Huang. 2010. High-performance computing for the simulation of dust storms. *Computers, Environment and Urban Systems* 34 (4):278–290.

Yang, C., R. Raskin, M. Goodchild, and M. Gahegan. 2010. Geospatial cyberinfrastructure: past, present and future. *Computers, Environment and Urban Systems* 34 (4):264–277.

Yang, C., and R. Raskin. 2009. Introduction to distributed geographic information processing research. *International Journal of Geographical Information Science* 23 (5):553–560.

Yin, L. et al. 2012. A framework of integrating GIS and parallel computing for spatial control problems – a case study of wildfire control. *International Journal of Geographical Information Science* 26 (4):621–641.

Zhang, J., and S. You. 2013. High-performance quadtree constructions on large-scale geospatial rasters using GPGPU parallel primitives. *International Journal of Geographical Information Science* 27 (11):2207–2226.

Zhao, Y., A. Padmanabhan, and S. Wang. 2013. A parallel computing approach to viewshed analysis of large terrain data using graphics processing units. *International Journal of Geographical Information Science* 27 (2):363–384.