Using Neighborhood Indicators to Understand Inner city Markets

By

Mi Diao

Master of Architecture, Tsinghua University 2002
Bachelor of Architecture, Tsinghua University 1996

Submitted to the Department of Urban Studies and Planning
In Partial Fulfillment of the Requirements for the Degree of

Master in City Planning

At the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2006

Signature of Author

Department of Urban Studies and Planning
June 2006

Certified by

Professor Joseph Ferreira, Jr.
Department of Urban Studies and Planning
Thesis Supervisor

Accepted by

Professor Langley Keyes
Chair, MCP Committee
Department of Urban Studies and Planning

Using Neighborhood Indicators to Understand Inner city Markets

By

Mi Diao

Submitted to the Department of Urban Studies and Planning
On May 25, 2006 in Partial Fulfillment of the Requirements for
the Degree of Master in City Planning

Abstract

The economic distress of America's inner cities is one of the most pressing issues facing the nation. Many analysts have asserted the unmet retail demand in inner-city neighborhoods and the potential for translating this demand into investment. Tapping the unmet retail demand has been considered an important strategy to accelerate the economic development in inner cities.

The purpose of this study is to propose an analytical framework that can reveal the spatial patterns of retail markets and test whether and to what extent inner-city neighborhoods are actually 'under served'. With the help of Geographic Information Systems (GIS) and Relational Database Management System (RDBMS) tools, this study designs and calculates neighborhood indicators of the demand, supply and gaps in retail markets with census tract level socioeconomic data and parcel level business data. Based on the indicators, econometric models are developed to quantitatively estimate the 'pure' impact of an inner-city location on the local retail supply level.

The neighborhood indicator system is applied to the food store markets in the Boston Metropolitan Statistical Area (MSA). Econometric analysis shows that inner-city tracts have an annual food store retail sales level (in millions of dollars per square mile) that is significantly lower than non inner-city tracts in the Boston MSA, after controlling for other factors that may influence retail supply level.

The proposed analytical framework can be easily applied to other retail markets as well as other MSAs. The spatial patterns of retail markets revealed by the neighborhood indicators can be helpful for business owners to identify opportunities for future business expansion or recruitment.

Thesis Supervisor: Joseph Ferreira, Jr.
Title: Professor of Urban Planning and Operation Research

Thesis Reader: Xavier de Souza Briggs
Title: Associate Professor of Sociology and Urban Planning

Acknowledgments

I would like to thank the people who generously offered advice, encouragement, inspiration, patience, and friendship throughout my years at MIT. Without their support, the completion of this thesis would have been impossible.

My heartfelt thanks go to Professor Joseph Ferreira Jr., my advisor, for his invaluable and constant support and encouragement in leading this thesis to completion. Joe guided me through my life at MIT for three years. His insightful advice always helped me look into problems deeper and consider meaningful solutions. I am honored to continue my studies under his supervision at MIT.

I am grateful to Professor Xavier de Souza Briggs for his invaluable comments, advices and great patience. His insightful thoughts helped me to improve the quality of this thesis, and led me to think about the policy issues behind the technical explorations.

Thanks go to Brookings Institution's Urban Market Initiative (UMI) – this thesis is part of a project partially funded by UMI. Thanks also go to Holly St. Clare from the Metropolitan Area Planning Council (MAPC), who generously provided some important data used in this study.

Finally, special thanks go to my family – this thesis is dedicated to them. I am grateful to my parents, Diao Keyong and Huang Xiaohua, for their extraordinary love and support. I am thankful to my wife, Cai Zuhua, who has been enormously supportive in my academic life. I feel fortunate to have someone so special beside me now and in the future. Special thanks go to JinJin (Amy Diao), my beloved baby daughter, who has been my source of joy and happiness ever since her birth last year.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Problem Statement

What are inner cities? In the United States, the term "inner-city" is perhaps most widely understood to mean the poorer parts at the center of a major city, which commonly have large minority population, higher level of poverty, unemployment rate, crime, single-parent families, lower education attainment, and so forth.

The economic distress of America's inner cities is one of the most pressing issues facing the nation (Porter 1997). There are substantial researches and development programs focusing on the revitalization of dilapidated inner-city neighborhoods. Many studies have documented the unmet retail demand in inner cities and the potential for translating this demand into investment, which are sorely needed by the inner-city neighborhoods (Cotterill and Franklin 1995; Porter 1995, 1997; ICIC 1998; U.S. Department of Housing and Urban Development 1999; Weissbourd and Berry 1999; Pawasarat and Quinn 2001; Sabety and Carlson 2004; Seidman 2004a).

Despite these studies, inner cities are experiencing a significant shortage of investment and retail market activities. As a result, low-income urban residents typically lose possible employment opportunities, pay more for groceries, and spend more time traveling to distant supermarkets (MacDonald and Nelson 1991; Whelan et al 2002; Clifton 2004). This puzzle raises two questions: do inner-city neighborhoods actually have unmet retail demand? If inner cities have retail market potential indeed, why, then have retailers ignored them? A growing number of studies indicate that an information gap plays a crucial role in answering these two questions (U.S. Department of Housing and Urban Development 1999; Weissbourd and Berry 1999; Sabety and Carlson 2004). Although the broad-brush expenditure data reveal general inner-city market potential, businesses need more specific and reliable assessments of different communities' potential market strength to make investment decision. However, literature suggests that conventional market assessment methods may not accurately represent the market potential of inner-city neighborhoods. The aggregate purchasing power of dense urban areas is largely undervalued (Weissbourd and Berry 1999; Pawasarat and Quinn 2001; Chieffo et al 2004; Seidman 2004a). When such crucial information about inner-city markets is not available, not accurate, or not used by market actors, an information gap exists and the inner cities fall out of retail investment (Sabety and Carlson 2004).

This study seeks to fill some information gaps in this field by proposing a new analytical framework that can reveal the spatial patterns of retail markets and test whether and to what extent inner-city neighborhoods are actually 'under served'. With the help of Geographic Information Systems (GIS) and Relational Database Management System (RDBMS) tools, this study designs and calculates neighborhood indicators of the demand, supply and gaps in retail markets with census tract level socio-economic data and parcel level business data. Based on these indicators, econometric models are developed to quantitatively estimate the pure impact of an inner-city location on the local retail supply level.

To make more accurate analysis about retail markets, this study differentiates retail markets into categories based on the Standard Industrial Classification (SIC) code. The food stores (with two-digit SIC industry code '54') in the Boston Metropolitan Statistical Area (MSA) is selected as a case study. The proposed analytical framework can be easily applied to other retail categories as well as other MSAs.

Among the many definitions of inner city, a methodology proposed by the Initiative for a Competitive Inner City (ICIC) is applied in this study, which defines inner cities as census tracts that currently have 20% poverty rate or higher, or meet two of the following three criteria:

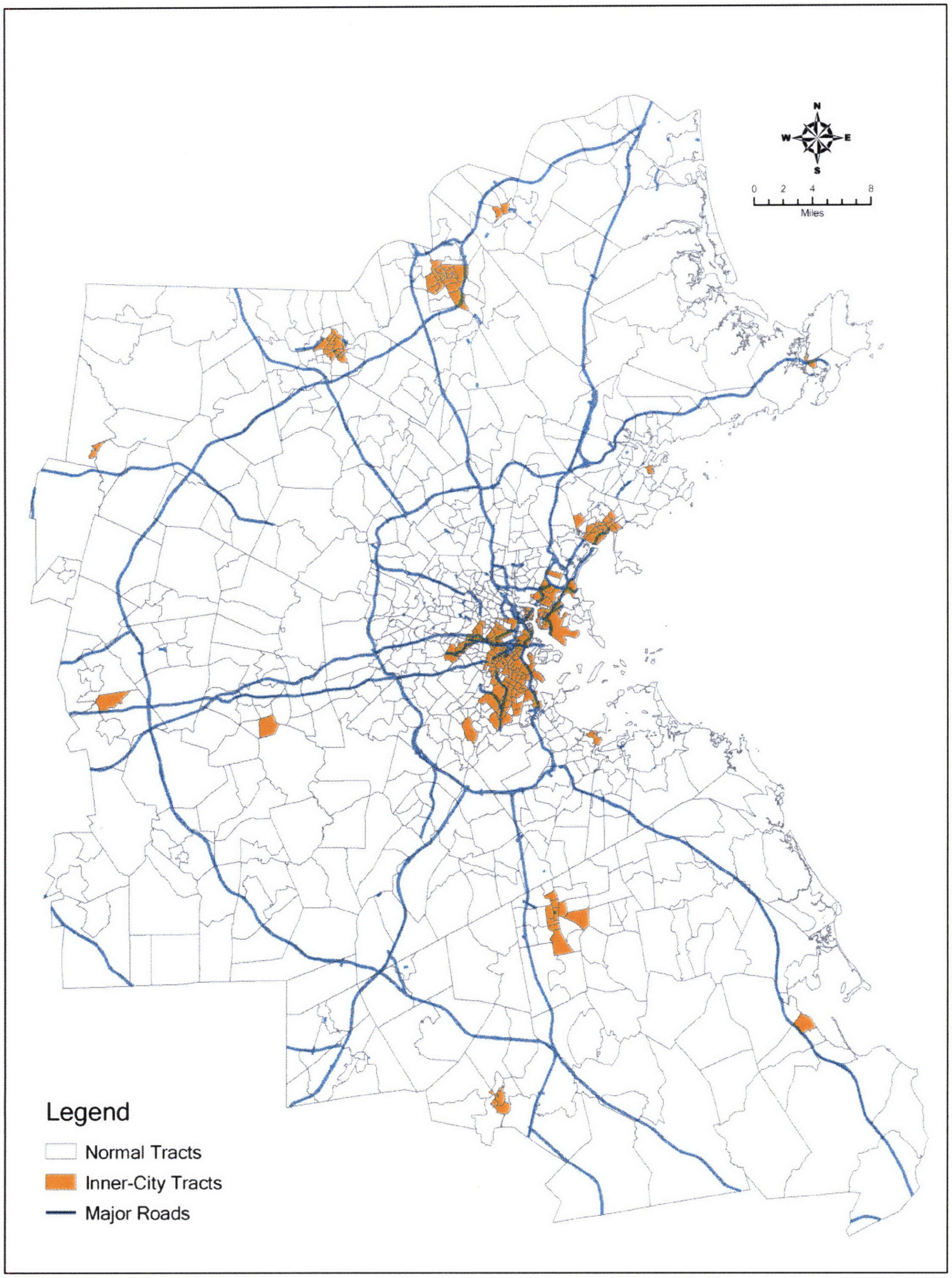
- poverty rate of 1.5 times or more that of their MSAs;
- median household income of 1/2 or less that of their MSAs;
- unemployment rate of 1.5 or more that of their MSAs.[1]

This definition not only catches tracts in absolute poverty status, but also the relatively poorer part of the MSA. Figure 1.1 shades census tracts that are classified as inner-city neighborhoods in the Boston MSA by the above criteria.

---

[1] www.icic.org

Figure 1-1: Inner-City Neighborhoods in the Boston MSA



1.2 Research Objectives and Hypothesis

The objectives of this thesis are the following:

1. to develop an analytical framework that quantitatively reveal the spatial patterns of retail markets.
2. to contribute to the inner city revitalization literature by statistically examining the existence and magnitude of the retail market potential in inner cites.
3. to explore new ways that information technologies like GIS and RDBMS tools can contribute to economic development study.
4. to explore a way to integrate and reinterpret spatially disaggregated raw datasets and generate useful indicators that can help residents to quantify and evaluate the social, economic, and environmental health of their communities, and provide decision support for various stakeholders.

To achieve these objectives, this study attempts to answer the following questions:

1. What are the spatial patterns of the food store retail markets in the Boston MSA?
2. How much of the variations in food store retail supply level of census tracts in the Boston MSA can be attributed to the inner-city location?

The research hypothesis is:

There is considerable unmet food store retail demand in the inner-city neighborhoods of the Boston MSA. An inner-city tract has a significantly lower annual food store sales (in millions of dollars per square mile) than non inner-city

tracts in the Boston MSA, after controlling for other factors that may influence retail supply level.

Findings of this study have implications for relevant stakeholders in the public, private and nonprofit sectors. It can help business owners to identify opportunities for future business expansion or recruitment. It may also initiate new thinking in public and nonprofit organizations about the strategies to revitalize inner-city neighborhoods and improve the quality of life of urban residents.

1.3 Thesis Structure

This thesis is organized into six chapters. Chapter two reviews the related literature to provide background information for the study.

Chapter three presents the analytical framework of the study. A new neighborhood indicator system for retail markets is proposed. Econometric models to test the research hypothesis are developed and possible variables are discussed. Major data sources and associated limitations are also presented.

Chapter four outlines how the neighborhood indicator system is structured, calculated, and visualized with GIS and RDBMS tools. This neighborhood indicator system is constructed for the food stores in the Boston MSA to uncover the spatial patterns of this market.

By incorporating indicators calculated in Chapter four, Chapter five uses econometric analysis to estimate the impact of an inner-city location on local retail supply. Regression results and related tests are discussed.

Chapter six gives a brief summary of research findings, discusses the limitations and challenge of the research, and suggests directions for future studies.

Chapter 2: Related Literature

This chapter reviews literature that provides background information for the study. Two important related fields are inner-city economic development strategies and retail market analysis.

2.1 Literature on Inner-City Economic Development Strategies

The distress of America's inner cities has brought serious social, economic and environmental problems to the urban core. Meanwhile, it has reduced the economic performance of the larger city and the region as well, which fail to reach their full potential. Substantial efforts have been made to revitalize the inner-city economy. Some focus on increasing human capital and meeting the basic human needs of disadvantaged populations, while others emphasize the power of markets. The next few sections outline the principles and practices of four types of important strategies.

2.1.1 Firm and Cluster Based Strategy

Firms play one important role in the economic redevelopment of inner cities. One well-known advocate of the firm-based strategy is the Initiative for a Competitive Inner-city (ICIC) founded by Harvard Business School professor Michael Porter. ICIC and Michael Porter emphasize the importance of market force and the central role of the private sector in the revitalization of inner cities, and suggest government and community-based organizations should shift their focuses from direct intervention to preparing and training the inner-city workforce and creating a favorable environment for business.

Porter (1997) summarizes the genuine competitive advantages of inner cities as:

1. Strategic location

Strategic location provides a competitive edge to inner-city neighborhoods, especially for logistics-sensitive and location-sensitive businesses, which can benefit from the proximity of inner cities to consumers, transportation infrastructure, and business clusters.

2. Integration with regional clusters

To achieve economic revitalization, inner cities should focus on developing clusters within inner cities, rather than isolated companies, and linking them better to those clusters in the surrounding economy.

3. Unmet local demand

Despite low average incomes, high population density of inner-city neighborhoods translates into a large local market with substantial purchasing power. This market is even more attractive considering the fact that there tend to be few competitors.

4. Human resources

Although inner-city populations present many workforce readiness challenges, inner-city residents can also be an attractive labor pool for businesses that rely on a loyal, modestly skilled workforce.

Porter (1997) argues that although inner-city neighborhoods have substantial assets, there are also many misperceptions and biases about inner cities and their opportunities. The redevelopment of inner cities will come only from recognizing the potential advantages of an inner-city location, while dealing with the present disadvantages of inner cities as business locations.

2.1.2 Community Development Based Strategy

Community development is an important redevelopment strategy employed by many nonprofit organizations. One such example is the Local Initiatives Support Corporation (LISC) - a national non-profit institution that directs financing and technical expertise to community development organizations. LISC/Chicago invested nearly $120 million in Chicago-based community development projects since its establishment in 1980, leveraging over $2 billion in total private and public investments[2].

LISP/Chicago aims to stimulate the comprehensive development of healthy, stable neighborhoods and improve the quality of life for Chicago residents. In order to achieve this goal, LISC/Chicago's main approaches include:

1. Provides technical expertise, information and training in the field of community development;

[2] Website of LISC/Chicago: http://www.lisc-chicago.org/about.php

2. Promotes policies, networks, relationships and resources that enhance neighborhood development;
3. Invests high-risk capital in a comprehensive program of neighborhood development, including but not limited to housing, commercial and industrial development, community facilities, open space, education and employment;
4. Invests in community development organizations, neighborhood institutions, public agencies and private sector companies;
5. Creates measurement instruments by which to assess the health and stability of the invested neighborhoods and to gauge the effectiveness of the investments.[3]

Rather than highlighting the power of market forces as Porter does, LISC's approach stresses social capital, and emphasizes community building through participation of community organizations, the technical guidance and resources from intermediaries like LISC, and the effective interaction between them.

2.1.3 Marketing Organization Strategy

Many marketing organizations with support from the governments rely on the Business Improvement District (BID) model and the Main Street model to facilitate the economic development of inner cities.

---

[3] These principles are summarized in LISC/Chicago's website: http://www.lisc-chicago.org

Hoyt (2005) describes the BID as publicly sanctioned – yet privately directed - organization that supplements public services to improve shared, geographically defined, outdoor public spaces. Moreover, such organizations subscribe to a self-help doctrine, whereby a compulsory self-taxing mechanism generates multi-year revenue. By providing services such as street and sidewalk maintenance, public safety officers, park and open space maintenance, marketing, capital improvements, and various development projects, BID model aims to increase the attractiveness of the downtown area, and improve its ability to compete with regional office parks and shopping malls.

'Main Street' programs function like BIDs, but they do not rely on a self-taxing funding mechanism. The 'Main Street' model combines activities in four intersecting areas:

1. Design and physical improvement to enhance the district's attractiveness;
2. Promotion and marketing to strengthen the district's image and attract more customers;
3. Economic restructuring to identify the district's economic potential, build on existing assets, and attract new business and capital;
4. Organizational development to create a strong volunteer-driven organization that engages all major concerned parties in planning and executing commercial district revitalization. [4]

Seidman (2004b) expands these principles to effectively address two key urban challenges: improving public safety and securing service from a fragmented city government.

---

[4] These principles are summarized in Dane, "Main Street success Stories", pp. 6-8

The Main Street Model gains a lot of popularity in practice. Since the introduction of the Main Street model more than twenty years ago, about 1,600 U.S. communities have used it to revitalize their downtowns or neighborhood commercial districts (Seidman 2004b).

2.1.4 Information Gap Based Strategy

Many studies assert that 'information gaps' constitute barriers for inner-city neighborhoods to regain economic development (HUD 1999; Weissbourd and Berry 1999; Sabety and Carlson 2004). Realizing that the current information gap biases the investment decision of market actors, Urban Markets Initiative (UMI)[5] aims to improve the quality of the information available on urban communities and use it to unleash the full power of those markets while connecting them to the economic mainstream.

UMI's approaches include: (1) identifying real information gaps that impede business, nonprofit and government investment; (2) crafting collaborative solutions to address those information gaps by leveraging technology; (3) educating businesses, non-profits and governments to create awareness, initiate adoption, spur new development of 'best practices', and identify future needs in the world of urban information[6].

---

[5] The Brookings Institution launched the Urban Markets Initiative (UMI) in September 2003 with support from Living Cities -- a non-profit, public-private partnership working to improve physically and economically distressed inner-city neighborhoods.
[6] UMI website: http://www.brookings.edu/metro/umi.htm

Echoing UMI's strong emphasis on using information to drive change, a large and growing number of agencies are actively engaged in efforts to quantify and evaluate the social, economic, and cultural health of communities. Increasingly, these efforts involve the use of city, regional, and national data, together with various GIS and statistical tools[7], to generate various neighborhood indicators[8]. While these efforts can be quite helpful, they are usually very labor-intensive and not easily replicated or sustained. For example, each update of the core datasets will ruin previous achievements; various agencies keep repeating similar works for different communities. The ideal model should have flexibility whereby users (NGOs and businesses) can easily tune the indicators to reflect their own beliefs and interests without needing to change or rework the underlying core datasets.

To address this issue, MIT and Metropolitan Area Planning Council (MAPC) have collaborated on one project 'Intelligent Middleware for Understanding Neighborhood Markets', with support from UMI. This project aims to prototype and test an intelligent middleware approach for sharing data within a metropolitan area in a manner that is likely to be more effective, scalable, and sustainable than the traditional approach (Ferreira 2004). By keeping official datasets at the backend 'untouched', and decomposing the data processing steps into reusable and tunable modules, the middleware approach make the data processing more efficient and sustainable, which

---

[7] An MIT Master of City Planning thesis written by Hideo Sakamoto and supervised by Prof. Ferreira experimented with such GIS models: "Socioeconomic Topography: Inner-city Economic Development and Geographic Information Systems," MIT MCP thesis, 1999, http://web.mit.edu/uis/theses/sakamoto.
[8] An example is the Nation Neighborhood Indicator Partnership (NNIP). The Urban Institute website for the NNIP is available at: http://www.urban.org/nnip

can substantially facilitate the construction, maintenance, development and sharing of neighborhood indicators to close the information gap.

2.2 Literature on Retail Market Analysis

Despite the many studies that suggest the existence of unmet market demand, the inner city is treated as a risky-investment for the majority of retailers, and thus continues to be underserved by mainstream commercial enterprises (Weissbourd and Berry 1999). A brief review on retail market analysis methods may shed light on this puzzle.

Retail market analysis plays an important role both in the business and planning domain. Chieffo et al (2004) indicate that business location decisions, whether they are made by in-house site selection staff or from suggestions of developers and brokers, are largely based on marketing firm data, which usually enter the location decision at an early stage. Seidman (2004a) describes market analysis as a planning tool that is increasingly used to inform economic development and commercial revitalization planning for urban neighborhoods.

However, a growing number of studies suggest that conventional market analysis methods may not accurately represent the market potential of inner-city neighborhoods. The aggregate purchasing power of urban areas is largely undervalued, which could make the inner-city neighborhood fall out of retail investment (Weissbourd and Berry 1999; Pawasarat and Quinn 2001; Chieffo, et al, 2004).

The traditional approach applied by large national marketing firms is clustering - classifying neighborhoods according to a template of neighborhood types developed by each firm. For example, leading marketing firm Claritas uses 'segmentation technology' to do a market analysis. Its mainstream product PRIZM divides the U.S. neighborhoods into 15 different groups[9] and 66 different segments. The 'Urban Cores' segments are characterized by 'relatively modest incomes, educations and rental apartments, ... One of the least affluent social groups... Among the group's preferences: TV news and daytime programming, Spanish and black radio, telephony services and pagers, cheap fast food and high-end department stores.' These qualitative classifications are often used to determine and rank the commercial viability of neighborhoods, and have substantial impacts on firms' location decision. The classification strategy is based on numerous variables but depends highly on household income. It does not adequately take into account density, the key competitive advantage presented by the inner city.

Another limitation of conventional market analyses is that they often use data that have biases with respect to low-income residents, which may lead to an inaccurate picture of inner-city markets. Pawasarat and Quinn (2001) point out that census data as a basis for conventional market analysis undercount the population and income of inner-city neighborhoods. Weissbourd and Berry (1999) indicate that 'unrecorded economy' is

---

[9] The fifteen groups include: Urban Uptown, Midtown Mix, Urban Cores, Elite Suburbs, The Affluentials, Middleburbs, Inner Suburbs, 2nd City Society, City Centers, Micro-City Blues, Landed Gentry, Country Comfort, Middle American, Rustic Living.

another source of discrepancy, which includes activities from nannies and tutors to home contractors and small businesses, and could reach 20% of the GNP.

In response to the perceived deficiency in conventional market analyses, alternative approaches are just now appearing, such as studies conducted by Initiative for a Competitive Inner City (ICIC), University of Wisconsin-Milwaukee Employment and Training Institute (ETI), and Social Compact. Typically, these studies employ density-sensitive indicators such as retail expenditure per square mile to measure the purchasing power of residents within a study area, and find substantial unmet retail demand in inner cities by comparing the demand and supply of a study area. For instance, a study by HUD (1999) compares retail sales and household purchasing power aggregated from tract level income and retail sales data for 48 inner cities with retail gaps, and finds a total shortfall of $8.7 billion. Another study by ICIC (1998) estimates that inner cities represent a market size of about $85 billion that is being consistently under serviced by a lack of suppliers within the neighborhoods. A market analysis model DrillDown developed by Social Compact builds on data from a spectrum of diverse sources like tax assessor, building permit, commercial credit companies, realtors, utility and police. Social Compact (2006) uses indicators generated with DrillDown to compare the residents' purchasing power and retail sales within two study areas in Santa Ana, and finds a $246 million market leakage. By addressing the defects of the traditional approaches, these alternative approaches provide more accurate pictures of the purchasing power, comparative advantage, and market potential of inner cities.

These alternative approaches are not without their flaws. First, they only consider the demand and supply situation within a study area itself, while the trade area of a store may well go beyond such geographic boundaries. These approaches could be misleading by omitting spillover effects. For example, a dense census tract close to the central business district (CBD) may have a very low retail store presence. But it does not necessarily mean that there is unmet retail demand in this tract, because the residents' retail demand can be conveniently met by the substantial retail supply in the nearby CBD area. Second, the definition of market opportunity is too narrow – concentrating mainly on demand of local residents. Of course local residents is a core market, but there are additional markets that could also provide market opportunities for the retail stores in a neighborhood. Seidman (2004a) indicates that there may be a 'visitor market' in which people live outside the commercial trade area but visit it regularly can be cultivated as a sustained market for business. For example, no-resident employees are an important source of retail demand, because many employees would like to grab some goods at stores close to their working place to save time and transportation cost.

To expand the studies in this field, this study will develop a new neighborhood indicator system for retail market analysis. The primary focus of this indicator system is to provide more complete pictures of retail markets by going beyond the geographic boundaries to catch spillover effects and covering broader sources of market opportunity. Furthermore, using the revised market indicators as variables in econometric models, the study will statistically test the spatial correlation between inner-city neighborhoods and under-served areas. To gain greater generality, this system maximizes the use of the existing secondary data. But it is ready to integrate local knowledge. Practitioners with rich primary data can easily tune the indicators to reflect neighborhood conditions.

Chapter 3: Methodology and Data

This chapter describes the methodology and data employed in this study. Section 3.1 proposes a new market-oriented neighborhood indicator system to uncover the spatial patterns of retail markets; Section 3.2 develops econometric models to estimate the impacts of an inner-city location on retail market supply level. Major data sources and limitations are discussed in Section 3.3.

3.1 A New Neighborhood Indicator System

When a business owner makes investment decisions, she may consider several basic market situations: what is the purchasing power of residents in the target community, how many competitors are doing business in this area, how much potential retail demand could be captured by the new store, and so forth. To meet the customized needs of market actors, the new neighborhood indicator system proposed in this study covers the demand, supply, and gaps of retail markets.

Retail demand refers to the purchasing power of consumers that could be captured by local retail businesses. Generally, it is determined by the demographic characteristics of the community. Based on different assumptions on consumers' shopping activities, this paper provides five methods to estimate people's purchasing power, which lead to five sets of indicators for a retail category:

1. Indicators calculated with the assumption that people spend a flat dollar amount annually in particular business categories;
2. Indicators calculated with the assumption that people spend a fixed proportion of their income in particular business categories;
3. Indicators calculated with Consumer Expenditure Survey (CES) estimation of household expenditure (in household income brackets) for each retail category;
4. Indicators calculated with the assumption that per capita expenditure is proportional to the average household expenditure in a tract;
5. Indicators calculated with the 'per capita expenditure' by 'per capita income' relation generated from the household expenditure by household income, based on the CES data.

Retail supply refers to the actual sales of retail stores that currently exist in the target community. The InfoUSA database used in this study contains the SIC code, location, sales and employment information of retail stores. The overall retail supply level of a community can be calculated by aggregating the sales of each store located in this community.

A comparison of demand and supply can help identify retail gaps (demand exceeds supply), which reflects the market potential in the neighborhood. If there appears to be a significant amount of unmet demand, it may be an opportunity for an existing business to expand or a new business to be developed. This gap also reflects the retail service level in this community. If a large portion of retail demand is locally provided, it means local residents can save transportation cost and enjoy a higher quality of life, assuming the prices across the MSA are similar.

Due to the spillover effects, market actors not only care about the supply-demand situation in the target community itself, but also the supply-demand situations in neighboring communities. To reflect the market situation in trade areas with different sizes, indicators are calculated at various levels, including the target community itself, and floating catchment areas within certain distances to the target community. In the proposed neighborhood indicator system, indicators at each aggregate level have the same structure, consisting of indicators on retail demand, retail supply, and the gap between demand and supply. These indicators can suggest whether the aggregation area is under-served or over-served in specific business categories.

Through this new neighborhood indicator system, the retail market situation can be quantitatively described. Further more, the spatial patterns of retail markets can be revealed vividly by shading the study area thematically with the value of these indicators.

GIS techniques and RDBMS tools are extensively used in this study to construct this neighborhood indicator system. Data are prepared, stored, processed and displayed in ArcMap and Oracle.

3.2 A Retail Sales Forecasting Model with Simple Distance Control

Due to spillover effects and the fact that multiple factors can influence retail supply level, the question onto whether or not a community is underserved in the retail markets cannot be answered by a simple comparison between the demand and supply of the target community. To address these issues, this study employs econometric models to test the hypothesis that inner-city neighborhoods are underserved in retail markets and quantitatively estimate the impacts of an inner-city location on local retail supply level.

Among the numerous factors that may affect the retail supply level, four factors are widely considered in previous research:

1. Retail demand of local residents;
2. Retail demand of residents from neighboring areas;
3. Retail demand of non-resident employees;
4. Location characteristics.

Apparently, retail demand of local residents is a substantial determinant of local retail supply level. People are less likely to travel long distance to serve their basic needs.

Consumers will not limit their shopping activities within the boundary of their residence areas, so retail demand of residents from neighboring areas can also influence the retail supply level of the target community. Meanwhile, this demand should be adjusted based on the retail supply level in the neighboring areas, considering competition effects.

Research has revealed that shopping trips may not be home based (Brown 1992). The consumer may grab some items from the stores near his working place to save time and transportation cost. Therefore, non-resident employees are another source of retail demand. This demand should also be adjusted for the retail leakage from residents of the target community who work outside the community, because these residents may also shop near their workplaces.

Masked by its long-time physical and socioeconomic distress, the retail demand of inner cities has not been fully recognized and tapped. Many studies suggest that an inner-city location substantially reduces the community's retail supply level. However, robust tests and estimations of such impacts are relatively scarce.

This study expands the previous research by using a multivariate linear regression model to quantitatively estimate the impacts of an inner-city location on local retail supply level. The dependant variable is the retail sales in each census tract. The explanatory variables in the model include retail demand from local residents, adjusted retail demand from neighboring residents (obtained by subtracting neighboring retail supply from neighboring retail demand), adjusted retail demand from non-resident employees (obtained by subtracting the number of workers living in this tract from the number of jobs in this tract), and an inner-city location dummy variable. The model can be expressed as:

$$Q_S = \alpha + \beta Q_D^l + \gamma Q_D^n + \delta Q_D^w + \lambda InnerCity + \varepsilon \qquad (3.1),$$

Where Q_s is local retail sales;

Q_D^l is retail demand from local residents;

Q_D^n is adjusted retail demand from neighboring residents;

Q_D^w is adjusted retail demand from non-resident employees;

InnerCity is a location dummy variable, which equals 1 for inner-city tracts, 0 otherwise;

ε is an error term;

$\alpha, \beta, \gamma, \delta, \lambda$ are coefficients to be estimated.

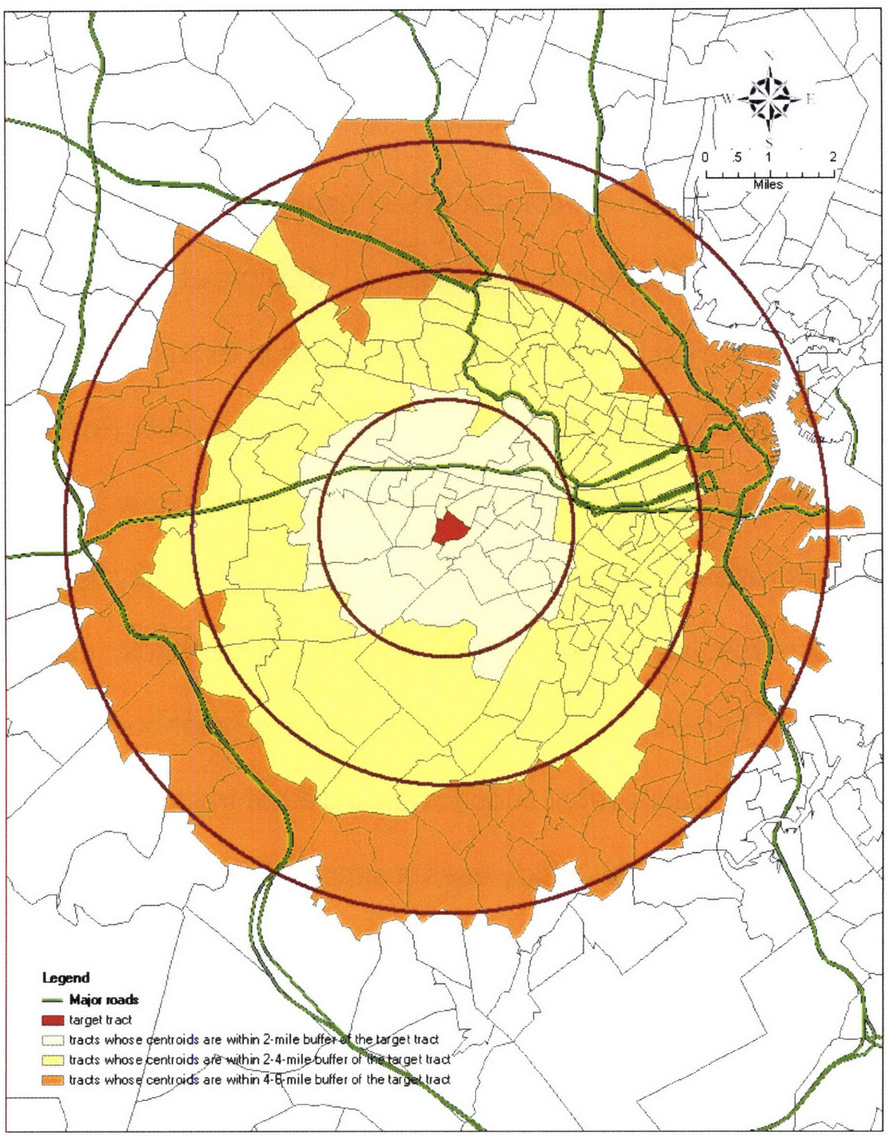
Q_s, Q_D^l, Q_D^n, and Q_D^w should be divided by the land area of their aggregation unit to avoid the size effect[10].

Space or distance has substantial influence on consumer's shopping activities. To control the impacts of the spatial (distance) factor, the classical retail gravity model is referred to in this study. The retail gravity model is extensively used in examining the spatial distribution of retail sales (Lakshmanan 1965; Eppli and Schilling 1996; Porojan 2000; Lee and Pace 2001). It draws an analogy with Newton's gravitational law to account for human shopping activities. In the retail gravity model, the possibility that a consumer shops at a store decreases as the distance between them increases, just as gravity diminishes with distance. Therefore, retail demand of a nearby neighboring tract has a greater influence on the target tract's retail supply than a distant tract, given the amount of the retail demand is the same. This study uses a simple method to control the impacts of distance. Consider a set of rings with different diameters around the centroid of the target tract. All tracts whose centroids within a ring are classified into one

[10] For example, a tract with a slightly higher volume of retail sales does not necessarily have a higher retail supply level than another tract, because its land area may be much bigger than that of the second tract.

group, and are assumed to have the same distance impact factor. Figure 3.1 illustrates this idea with an example. For the target tract (shaded in red in the map), three groups of tracts, whose distances to the target tract are less than 2 miles (target tract itself excluded), 2-4 miles and 4-6 miles respectively, are rendered in different colors, from light yellow to dark yellow.

Figure 3-1: Groups of Census Tracts Classified by Distance to the Target Tract



At this point, the regression model can be specified as below:

$$Q_S = \alpha + \beta Q_D^l + \gamma_1 Q_D^{n1} + \gamma_2 Q_D^{n2} + \gamma_3 Q_D^{n3} + \delta Q_D^w + \lambda InnerCity + \varepsilon \qquad (2.2),$$

where Q_D^{n1} is the adjusted retail demand of tracts in ring 1;
Q_D^{n2} is the adjusted retail demand of tracts in ring 2;
Q_D^{n3} is the adjusted retail demand of tracts in ring 3.

The distances to the target tract have the following relation: ring 1 < ring 2 <ring 3.

More (adjusted) retail demand means more market potential, so the coefficient of local demand, neighboring demand and employee demand is expected to be positive. It should be noted that the neighboring demand variables discussed in the section are a little different from the neighboring area indicators proposed in the preceding section. Their aggregation areas are tracts within different buffer rings of the target tract, while the indicators in the preceding section are calculated for circle buffer zones of the target tract. But the calculation methods/scripts are very similar.

The coefficient of the inner-city location dummy λ can be explained as the impacts of an inner-city location on local retail supply level. Negative value of λ is expected, according to the hypotheses that inner-city neighborhoods are actually underserved in the retail markets.

3.3 Data Sources and Limitations

Four major data sets are used in this study, including Census data, Consumer Expenditure Survey (CES) data, InfoUSA business data and Census Transportation

Planning Package (CTPP) data. The Census, CES, and CTPP data are used to calculate indicators from the demand side of the retail markets, while the InfoUSA data are used to calculate indicators from the supply side of the retail markets.

1. U.S. Census data, 2000

The census is the most accessible source for demographic data at different geographic levels, such as the county, zip code zone, census tract, and block group. The census is conducted every 10 years by the U.S. census bureau.

The latest version -- Census 2000, will be used in this study. It provides a detailed demographic and socioeconomic profile of residents, including household income and per capita income at the census tract level, which is an important determinant of the retail demand of communities.

2. U.S. Consumer Expenditure Survey data, 2000

The U.S. consumer survey program provides information on the buying habits of American consumers, including their expenditures, income, and consumer unit (families and single consumers)[11] characteristics. The data are collected in independent quarterly interview and weekly diary surveys of approximately 7,500 sample households (5,000 prior to 1999). Each survey has its own sample, and each collects data on household income and socioeconomic characteristics. The interview survey includes monthly out-of-pocket expenditures such as housing, apparel, transportation, health care, insurance,

[11] In this study, consumer unit is used as a surrogate of household.

and entertainment. The diary survey includes weekly expenditures of frequently purchased items such as food and beverages, tobacco, personal care products, and nonprescription drugs and supplies[12].

In this study, the 2000 CES data is used. CES contains data about household expenditure in a retail category by household income groups, which helps us to link residents' income with their actual purchasing power.

3. InfoUSA business data, 2003

InfoUSA, founded in 1972, is a leading compiler of several proprietary databases, which provide detailed information on majority of businesses and consumer households in the United States and Canada. These databases are compiled and updated from thousands of public sources such as yellow pages, white pages, newspapers, incorporation records, real estate deed transfers and various other sources.[13]

The InfoUSA dataset used in this paper covers over 268,000 businesses in Massachusetts for the year of 2003. It contains detailed information for business establishments in different categories, such as SIC code, sales (in sales category), employment (in number of employees), and location (in latitude and longitude), and so forth. In this study, InfoUSA data are used to estimate the retail supply level of census tracts. Since the sales data are provided in categories rather than accurate numbers,

---

[12] Consumer Expenditure Survey website: http://www.bls.gov/cex/home.htm
[13] http://www.infousa.com. As a regional agency, MAPC has access to the InfoUSA data and we use the data for this study through the collaboration of MIT and MAPC on the UMI project.

the employment data will be used as a proxy of retail supply level, assuming the constant productivity of employees across the MSA. To match the social-economic data at the census tract level, store level employment data are aggregated to census tract level with GIS tools based on each store's location.

4. Census Transportation Planning Package, 2000

Census Transportation Planning Package (CTPP) is a set of special tabulations from the decennial census designed for transportation planners. The data are tabulated from answers to the Census long form questionnaire, mailes to one in six U.S. households. CTPP provides information about jobs as well as about workers for small geographic areas. Part 1 of CTPP contains data for workers by place of residence, and Part 2 provides data for workers by place of work. Here the former measures workers residing in each census tract and the latter measures jobs located in each census tract.

CTPP 2000 data for Massachusetts is used for this study. The number of jobs (adjusted by number of workers in the target tract) works as a proxy for retail demand of non-resident employees, which can be calculated with the CTPP data.

Several limitations in the datasets may cause potential biases to the indicators and analyses, which need special attention.

1. Inherent drawbacks of national data

Data used in this study are mainly survey data at the national level, which have some inherent drawbacks. As Weissbourd and Berry (1999) point out, using U.S. Census reported income as a basis for expenditures is not accurate for low-income households that often expend more than their reported incomes due to the inclusion of government issued cash-equivalents. Census income estimates also fail to account for the 'unrecorded economy' – an economy estimated at $1 trillion annually which is comprised of predominately legal but unrecorded activities mostly cash transactions from nannies, tutors, and even small businesses that occur at disproportionate rates in the inner-cities. The CES data is limited because it is based on a national sample and provides no local area specific information. The integration of national data with state and local data, for instance, annual income tax data, home mortgage disclosure act data, local spending pattern survey data, will surely improve the quality of the analysis. However, due to the high collection cost and/or inaccessibility of such data, the Census and Consumer Expenditure Survey will still be the base of this study, while the proposed analytical framework can be easily adjusted to include local knowledge.

2. Time inconsistency

Another problem is the time inconsistency of datasets – Census data, CES data, and CTPP data are for the year 2000, while InfoUSA data is for the year 2003[14]. This inconsistency may cast doubts on the outcomes of the analysis. One possible argument is that retail markets adjust very slowly to changes in retail demand. Considering this

[14] The author cannot get access to the 2000 InfoUSA data, while the Census and CTPP data are not available for 2003.

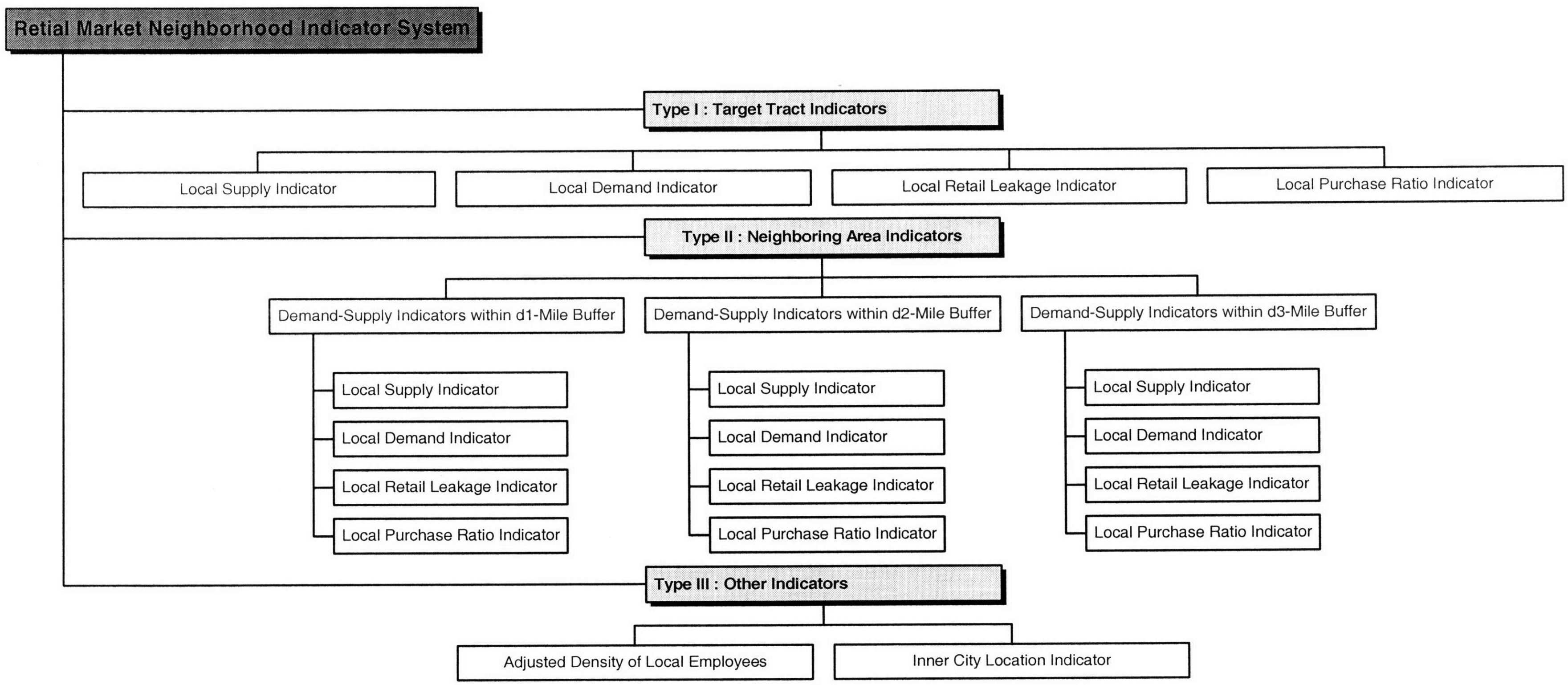
time lag, the retail supply level of 2003 (measured by the level of retail employment) may actually be a preferable choice compared with 2000 in order to reflect the retail market demand of 2000.

Chapter 4: Indicator Generation and Visualization

This chapter covers the definition, structure, calculation and visualization of the neighborhood indicator system proposed in Chapter 3. This system consists of three types of indicators: target tract indicators, neighboring area indicators, and other indicators. Figure 4.1 shows the structure of the retail market neighborhood indicator system applied in this study.

This chapter proceeds as follows: Section 4.1 provides five methods to calculate the target tract indicators based on different assumptions about people's purchasing power; Section 4.2 covers the neighboring area indicators; other indicators are discussed in Section 4.3; Section 4.4 summarize the differences of the five methods; finally, in Section 4.5 the spatial patterns of food store markets in the Boston MSA are illustrated by visualizing the corresponding neighborhood indicators. In Section 4.1 – 4.3, general calculation methods, which can be applied to any retail categories as well as any analysis units, are first presented. Then as an example, these methods are applied to the food store markets in the Boston MSA at the census tract level. The reasons for selecting the Boston MSA as the study area is that Boston has a typical inner-city area and it is the only MSA that all the necessary data are available for the author. Food stores is chosen as the retail category under investigation because it has fundamental influences on people's quality of life and the urban grocery store gap has been documented by many studies (Cotterill and Franklin 1995; Donohue 1997).

Figure 4-1: Structure of Retail Market Neighborhood Indicator System



4.1 Indicator Generation I: Target Tract Indicators

Target tract indicators reflect the retail market situations in the tract under investigation itself, including local supply indicator, local demand indicator, local retail leakage indicator, and local purchase ratio indicator. This section provides five calculation methods. The first method is presented in detail, covering the calculation process of all indicators, while the presentation of the other four methods focuses on their differences with Method 1. Table 4-1 provides a brief summary of the five methods.

Table 4-1: Brief Summary of Five Calculation Methods

No.	Name	Brief Discription
1	Flat dollar per capita	Calculated with the assumption that people spend flat dollar amount annually in a retail category
2	Flat percent of income	Calculated with the assumption that people spend a fixed proportion of their income in a retail category
3	Per household CES income adjustment	Calculated with CES* estimation of household expenditure (in household income brackets) for each retail category
4	Per capita CES income adjustment	Calculated with the assumption that per capita expenditure is proportional to the average household expenditure by income in a tract
5	Household size adjusted per capita CES income adjustment	Calculated with per capita expenditure by per capita income relation generated from the household expenditure by household income relation provided by the CES* data

* Consumer Expenditure Survey

Procedural Language/SQL (PL/SQL) is the major tool to conduct the calculation. Part of the scripts used in this chapter is listed in Appendix A.

4.1.1 Method One (Flat Dollar Per Capita)

Compared with the other four methods, Method 1 is the most straightforward way to calculate the retail market neighborhood indicators. It assumes that people spend flat dollar amounts annually in a retail category.

1. Local Supply Indicator: S_m^i

Local supply indicator S_m^i refers to the actual sales of stores in retail category *i* that currently exist in tract *m*. It can be calculated from the InfoUSA database or any other business database that contains information about the retail category, location (in terms of latitude and longitude coordinates or street addresses), sales, and employment of stores.

The InfoUSA database used in this study does not provide exact sales data (sales data are provided in categories rather than accurate numbers), employment data are used as a proxy, which can be transformed to sales using a statewide sales to employment ratio assuming constant productivity of employees across the state. The following are the steps to estimate tract *m*'s retail sales using store level employment data:

Step 1: Create a table containing all the stores in retail category *i* from the InfoUSA database. Use food stores as an example. All the stores whose two-digit SIC industry code equals '54' are selected, which include retail stores primarily engaged in selling food for home preparation and consumption.

Step 2: Aggregate the employment of retail category *i* in tract *m*. First, create a point layer representing all the stores in retail category *i* with their latitude and longitude coordinates[15] using ArcMap[16]. Figure 4-3 shows such a map for food stores in the Boston MSA. Next, use the spatial join function of ArcMap to join the Massachusetts 2000 census tract layer[17] to store layer and add the associated census tract ID to each store. Then, sum up the overall employment for tract *m* by associated tract ID of each store. This procedure can be accomplished with ArcMap or any database management software like ORACLE. Note that in the InfoUSA database, 596 out of 4192 food stores in the Boston MSA only have an employee number range (1-4 employees or 5-9 employees) rather than exact numbers. For these stores, the mean of the associated range is used as the employment of the store.

Step 3: Calculate the sales to employment ratio for retail category *i* in the state. The economic census by the U.S. Census Bureau provides data about the overall employment and sales by business category and state. The sales to employment ratio can be calculated by dividing the overall sales by the overall employment of retail category *i* in the state. For example, in 1997 the total number of paid employees of food stores in Massachusetts is 93,579, the total sales of food stores in Massachusetts is about 10,835.4 million dollars, so the sales to employment ratio of Massachusetts is 115,789 dollars per employee.

---

[15] All the stores in the InfoUSA database have latitude and Longitude coordinates.
[16] ArcMap is the central application in ArcGIS Desktop -- a collection of software products developed by ESRI, which is used to create, import, edit, query, map, analyze, and publish geographic information with standard desktop computers.
[17] This layer contains boundaries of all census tracts in Massachusetts in 2000. It can be downloaded from MASSGIS website: http://www.mass.gov/mgis/

Step 4: Transform the employment of retail category *i* in tract *m* to sales by multiplying the total employment of tract *i* by the sales to employment ratio of the state, assuming constant productivity of food store employees across the state.

Step 5: Normalize the retail supply by dividing it by tract m's land area to avoid the size effect.

Tract '25025000401' is a typical inner-city tract in the Brighton neighborhood of Boston, which is used as an example to show the calculation process throughout this chapter. Some basic socioeconomic data about this tract are listed in Table 4.2. Table 4-3 shows the detail calculation process for tract '25025000401'.

Table 4-2: Socioeconomic Data of Census Tract '25025000401'

Land Area (sq.mi.)	Population	No. of Households	Poverty Rate	Med. Household Income ($)	Unemployment Rate	PerCapita Income ($)	No.of Food Stores	Food Store Employment
0.17	5796	2946	0.217	29491	0.022	22588	3	8.5

Source: U.S. Census 2000

Table 4-3: Example of Food Store Retail Supply Calculation with Method 1

	Procedures	Methods	Units	Values
A	Calculate the total employment of food stores in the tract	Aggregate store employment data with GIS tools	person	8.5
B	Estimate the sales to employment ratio for the State	B=Total sales in the State/total employment in the State	thousand $ / person	115.79
C	Calculate the local retail sales for the tract	C=A*B	thousand $	984.21
D	Normalize the retail sales by dividing the land area of the tract	D=C/land area	thousand $ / square mile	5,789.45

Source: Calculated by the author.

2. Local Demand Indicator: D_m^i

Local demand indicator D_m^i refers to the purchasing power of consumers in tract *m* that could be captured by local stores in retail category *i*. By assuming people spend flat dollars in retail category *i*, Method 1 estimates the local demand in tract *m* as the product of the population in this tract and the per capita expenditure in the MSA, as is shown in Equation (4.1)

$$D_m^i = \frac{E_m^i * POP_m}{A_m}, \ E_m^i = E_s^i \qquad (4.1),$$

where E_m^i is the per capita expenditure in retail category *i* of tract *m*; E_s^i is the per capita expenditure in retail category *i* of the MSA; POP_m is the population of tract *m*; A_m is the land area of tract *m*.

The following are the detailed steps to calculate the local demand indicator for retail category *i* in tract *m*:

<u>Step 1</u>: Calculate the MSA wide per capita expenditure in retail category *i* by dividing the MSA wide sales in retail category *i* by the population of the MSA. Here the retail sales for the MSA are used as a surrogate for the total consumer expenditure in the MSA. The underlying assumption is that the aggregate consumer demand at the MSA level is fairly well captured by the MSA itself. In the Boston MSA, the total number of food store employees is 60,516. By multiplying total number of employees by the statewide sales to employment ratio (115,789 dollars per employee), the estimated annual food store sales of the Boston MSA is 7,007.09 million dollars. The population of

the Boston MSA in 2000 is 4,306,692. Hence the MSA wide per capita expenditure is 1,627 dollars.

<u>Step 2</u>: Calculate the retail demand in tract *m* by multiplying the population in tract *m* by the MSA wide per capita expenditure in retail category *i*.

<u>Step 3</u>: Normalize the retail demand by dividing it by tract m's land area.

3. Local Retail Leakage Indicator: RLI_m^i

Local Retail Leakage Indicator of retail category *i* in tract *m* (RLI_i^m) represents whether and to what extent the tract is losing potential sales to other tracts. It can be calculated as the margin between local demand indicator and local supply indicator, as is shown in the equation below:

$$RLI_m^i = D_m^i - S_m^i \qquad (4.2).$$

RLI_m^i represents the volume of the retail gap.

4. Local Purchase Ratio Indicator: LPI_m^i

Local Purchase Ratio Indicator LPI_m^i is the ratio of tract *m*'s supply to its demand in retail category *i*. This ratio can be understood as an indicator of the degree at which local residents spent their retail dollars locally. A value less than 1 means this community is losing retail dollars to other communities. A value greater than 1 means this community is attracting retail dollars from other communities. The calculation formula is:

$$LPI_m^i = S_m^i / D_m^i . \qquad (4.3).$$

Obviously, there are some shortcomings associated with Method 1. Resident' expenditure is influenced by income level. Normally, a resident with high income will spend more than a low-income resident in a retail category. Method 1 tends to overestimate the demand of the poor, while underestimate the demand of the rich. Therefore, the retail gap in dilapidated inner cities is likely to be exaggerated.

4.1.2 Method Two (Flat Percent of Income)

Compared with Method 1, Method 2 uses a different approach to estimate the local retail demand, while their methods to calculate local supply indicator, local retail leakage indicator, and local purchase ratio indicator are just the same. So this section only presents the calculation of local retail demand with Method 2.

Rather than assuming people spend flat dollar amount annually in a retail category as in Method 1, Method 2 assumes people spend a fixed proportion of their income in a retail category. Therefore, a tract's per capita expenditure is proportional to its per capita income and can be calculated from the MSA wide per capita expenditure. The local demand indicator can be then calculated with the following equation:

$$D_m^i = \frac{E_m^i * POP_m}{A_m}, \ E_m^i = E_s^i * \frac{INC_m}{INC_s} \qquad (4.4),$$

where E_m^i is the per capita expenditure in retail category *i* of tract *m*; E_s^i is the per capita expenditure in retail category *i* of the MSA; POP_m is the population of tract *m*; A_m is the land area of tract *m*; INC_m is the per capita income of tract *m*; INC_s is the per capita income of the MSA.

The following are the detailed steps to calculate the local demand indicator for retail category *i in* tract *m* with Method 2:

Step 1: Calculate MSA wide per capita expenditure in category *i* by dividing the MSA wide sales in category *i* by the population of the state. As is presented in the preceding section about Method 1, the per capita expenditure in food stores of the Boston MSA is about 1,627 dollars.

Step2: Calculate the adjustment factor, which is the ratio of per capita income in tract *m* to per capita income in the MSA.

Step 3: Calculate the per capita expenditure of tract *m*. As is shown in Equation 4.4, the per capita expenditure of a tract can be calculated from the per capita expenditure of the MSA, adjusting the difference in their per capita incomes.

Step 4: Calculate retail demand of tract *m* by multiplying the population and the per capita expenditure of tract *m*.

Step 5: Normalize the retail demand by dividing it by tract m's land area.

Table 4-4 shows the food store demand calculation process for a tract in the Boston MSA (with tract ID '25025000401') with Method 2.

Table 4-4: Example of Food Store Retail Demand Calculation with Method 2

	Procedures	Methods	Units	Values
A	Calculate the MSA wide per capita food store expenditure	A=Total food store expenditure in the state/State population	thousand $ / person	1.627
B	Estimate the adjustment factor of the tract	B=Per capita income of the tract/per capita income of the MSA	N.A.	0.802
C	Calculate the per capita expenditure of the tract	C=A*B	thousand $ / person	1.304
D	Estimate the overall expenditure of the tract	D=C*Population	thousand $	7,559
F	Normalize the retail sales by dividing the land area of the tract	F=D/Land area	thousand $ / square mile	44,462

Source: Calculated by the author

There are also some drawbacks associated with Method 2. In real life, though consumer's expenditure is positively related to income, the relationship cannot be simplified as a linear relationship as in Equation 4.4. It can be expected that as the income increase, the marginal effect of an additional unit of income will decrease. Method 2 tends to overestimate the purchasing power of the rich, while underestimate the purchasing power of the poor.

4.1.3 Method Three (Per Household CES Income Adjustment)

The Consumer Expenditure Survey (CES) by the Bureau of Labor Statistics provides detailed information about the spending characteristics of consumer units. In this study, consumer unit is used as a replacement of household[18]. Method 3 uses the summation

[18] According to the glossary of CES, a consumer units is defined as members of a household related by blood, marriage, adoption, or other legal arrangement; a single person living alone or sharing a household with others but who is financially independent; or two or more person living together who share responsibility for at least 2 out of 3 major type of expense – food, housing, and other expense.

of household expenditure in a retail category to estimate the local retail demand, assuming total expenditure is equal to the total household expenditure in the target tract. This method can be expressed with Equation 4.5

$$D_m^i = \frac{\sum_j HOUEXP_{mj}^i}{A_m} \qquad (4.5),$$

where j is summed over the households living in tract *m*; $HOUEXP_{mj}^i$ is expenditure of household *j* in retail category *i*; A_m is the land area of tract *m*.

The following are the procedures to calculate the local demand indicator for retail category *i* in tract *m*:

Step 1: Calculate the number of households in each household income group in tract *m* using the 2000 Census data. Note that the dividing points of household income group in Census data are slightly different from that in CES data. So some transformation work is needed.

Step 2: Calculate the retail expenditure for each household income group in tract *m*, which equals the product of number of households and the average household expenditure of that group. Table 4.5 is part of the 2000 CES report, showing the annual food expenditure by household income group. In this study, food at home category is used as a surrogate for the food store expenditure.

Step 3: Estimate the total retail expenditure of tract *m*. To get this figure, sum up the retail expenditure of all household income groups in tract *m*.

Step 4: Normalize the retail demand by dividing it by tract m's land area.

Table 4-5: Annual Household Food Expenditure by Household Income, U.S., 2000

Item	Complete reporting of income									
	Total complete reporting	Less than $5,000	$5,000 to $9,999	$10,000 to $14,999	$15,000 to $19,999	$20,000 to $29,999	$30,000 to $39,000	$40,000 to $49,000	$50,000 to $69,000	$70,000 and over
Number of consumer units (thousand)	81,454	3,627	7,183	8,037	6,677	12,039	9,477	7,653	11,337	15,424
Consumer unit characteristics:										
Income before taxes ($)	44,649	1,980	7,638	12,316	17,319	24,527	34,422	44,201	58,561	112,586
Average number in consumer unit:										
Persons	2.5	1.8	1.7	2.0	2.2	2.4	2.5	2.6	2.9	3.2
Average annual expenditure	40,238	17,946	15,703	21,199	24,331	29,852	35,609	42,323	49,245	75,964
Food ($)	5,435	2,627	2,462	2,984	3,743	4,507	5,118	6,228	6,557	8,665
Food at home ($)	3,154	1,603	1,723	2,108	2,556	2,921	2,995	3,552	3,605	4,483
Food away from home ($)	2,280	1,024	738	876	1,187	1,586	2,122	2,676	2,952	4,182

Source: Consumer Expenditure Survey Data, Bureau of Labor Statistics

Table 4-6: Example of Food Store Retail Demand Calculation with Method 3

Household Income Branket	Number of Households	Food Store Expenditure Per Household (thousand $ / household)	Food Store Retail Demand (thousand $)
Less than $10,000	775	1.663	1,289
$10,000 to $14,999	315	2.108	664
$15,000 to $19,999	141	2.556	360
$20,000 to $29,999	224	2.921	654
$30,000 to $39,999	174	2.995	521
$40,000 to $49,999	275	3.552	977
$50,000 to $69,000	423	3.605	1,525
$70,000 and over	619	4.483	2,775
Total	2946		8,765
Normalized by Land Area (thousand $ / sqmi)			51,561

Source: Calculated by the Author based on 2000 CES data and 2000 Census data

Table 4-6 shows how a tract's food store retail demand can be calculated with Census and CES data. A tract in the Boston MSA (with tract ID '25025000401') is used as an example.

It should be noted that the estimation of local retail supply in Methods 3 and 5 is different from that in Methods 1, 2, and 4. Rather than assuming statewide constant employee productivity as in Methods 1,2, and 4, Methods 3 and 5 assumes that the productivity is a constant at the MSA level[19]. The advantage of this approach is that it allows for the difference of employee productivity between the state and the MSA.

The following procedures to estimate the local retail supply of retail category *i* in tract *m* are used in Method 3 and Method 5.

Step 1: Calculate the total retail sales of the MSA. According to the assumption that retail demand and supply are balanced at the MSA level, the total retail sales in the MSA equal the total retail expenditure in the MSA, which is the sum of each tract's retail demand calculated above.

Step 2: Estimate the sales to employment ratio for the MSA. This figure equals the total retail sales divided by the total employment of the MSA. The total employment data can be aggregated from the store level employment data in the InfoUSA database.

---

[19] InfoUSA database does not contain exact store sales data. The retail supply level can only be estimated from the store employment data by assuming constant employee productivity. To calculate the productivity, both the sales and number of employees at a curtain aggregation level should be known. These data are available at the state level from the Economic Census, but not the MSA level. In methods 1,2 and 4, the calculation starts from the supply side, then goes to the demand side. So we can only assume constant employee productivity at the state level. While in method 3 and 5, the calculation starts from the demand side. The total retail demand at the MSA level can be calculated directly with the CES data. By assuming retail demand and supply are balanced at the MSA level, the MSA wide average sales to employment ratio (productivity) can be obtained.

Step 3: Calculate the total employment of retail category *i* in tract *m*, using the method presented in Method 1.

Step 4: Calculate the local retail sales for tract *m*, which is the product of total employment of retail category *i* in tract *m* and the sales to employment ratio of the MSA.

Step 5: Normalize the retail sales by dividing it by tract m's land area.

Table 4-7 shows the calculation process for a tract in the Boston MSA (with tract ID '25025000401') with Method 3.

Table 4-7: Example of Food Store Retail Supply Calculation with Method 3

	Procedures	Methods	Units	Values
A	Calculate the total retail sales of the MSA	A=Total expenditure of the MSA	thousand $	5,818,020
B	Calculate the total employment of food stores in the MSA	Aggregate store exployment data with GIS tools	person	60,516
C	Estimate the sales to employment ratio for the MSA	C=A/B	thousand $ / person	96.140
D	Calculate the total employment of food stores in the tract	Aggregate store employment data with GIS tools	person	9
E	Calculate the local retail sales for the tract	E=C*D	thousand $	817
F	Normalize the retail sales by dividing the land area of the tract	F=E/land area	thousand $ / square mile	4,807

Source: Calculated by the author.

The local retail leakage indicator and local purchase ratio indicator can be calculated using Equation 4.1 and Equation 4.2, as is presented in Section 4.1.1.

The advantages of Method 3 reside in: (1) rather than make assumptions about the expenditure patterns of consumers as in Method 1 and 2, Method 3 uses 'real' household expenditure data to do the calculation, which may better reflect the purchasing power of consumers; (2) Method 3 uses the 'true' distribution of household income in the target tract. While in the other four methods, per capita income is used as an indicator of the wealthy level of the community, which embodies the critical assumption that residents within a given zone are fairly homogeneous. This assumption may cause aggregation bias in understanding the highly heterogeneous socio-economic activities in modern cities. For instance, an individual with an extremely high income will have a big impact on the average income of the study tract; however, the total food expenditure of the study tract will not be significantly different.

The drawbacks of Method 3 include: (1) this method uses total household expenditure as a proxy of the total expenditure in the tract. Individuals that are not classified as households are excluded from the calculation. For example, 3 out of 894 tracts in the Boston MSA do not have household presence according to 2000 Census. The retail demand in these tracts would be totally ignored if Method 3 is used; (2) the Consumer Expenditure Survey are U.S. based. The difference between the purchasing power of average U.S. consumers and consumers in the Boston MSA is another source of biases.

4.1.4 Method Four (Per Capita CES Income Adjustment)[20]

Method 4 can be seen as a revised version of Method 2. Method 2 assumes a tract's per capita expenditure in a certain retail category is proportional to its per capita income. Method 4 assumes a tract's per capita expenditure is proportional to its average household expenditure by income group. Thus the per capita expenditure of tract m can be calculated from the per capita expenditure of the MSA accounting for the differences in their average household expenditure. The complete formulas to calculate tract m's retail demand are:

$$D_m^i = \frac{E_m^i * POP_m}{A_m}, \; E_m^i = E_s^i * \frac{AHS_m}{AHS_s} \quad (4.6),$$

where E_m^i is the per capita expenditure of retail category *i* in tract *m*; E_s^i is the MSA wide per capita expenditure of retail category *i*; POP_m is the population of tract *m*; A_m is the land area of tract *m*; AHS_m is the average household expenditure in tract *m*; and AHS_s is the MSA wide average household expenditure.

The following are the steps to calculate the local demand indicator using method 4:

Step 1: Calculate the MSA wide per capita expenditure in retail category *i*, using the method presented in Section 4.1.1 (Method 1). The per capita food store expenditure in the Boston MSA is 1,627 dollars per person

[20] Some market analysis studies use similar method to evaluate retail store demand, for example the online market analysis workbook by University of Wisconsin-Extension Center for Community Economic Development http://www.uwex.edu/ces/cced/dma/9.html

Step 2: Calculate the total household expenditure of retail category *i* in tract *m* as well as in the MSA, using the method presented in Section 4.1.3 (Method 3) and CES data.

Step 3: Calculate the average household expenditure of retail category *i* in tract *m* as well as in the MSA. This figure can be obtained by dividing total household expenditure in tract *m* (the MSA) with the total number of households in tract *m* (the MSA).

Step 4: Estimate the ratio of average household expenditure in tract *m* to the average household expenditure in the MSA.

Step 5: Calculate the per capita expenditure of tract *m* by multiplying the MSA wide per capita expenditure by the ratio obtained in step 4.

Step 6: Estimate the overall expenditure of tract *m*. This figure is the product of per capita expenditure and the population of tract *m*.

Step 7: Normalize the retail expenditure by dividing it by tract m's land area.

Table 4-8 shows the local demand indicator calculation for a tract in the Boston MSA with tract ID '25025000401'. Other indicators can be calculated with the same approach as is presented in section 4.1.1 (Method 1).

One the one hand, Method 4 improves Method 2 by using a more reasonable assumption about a tract's per capita expenditure. On the other hand, it also has two shortcomings: (1) it uses the per capita income as the base of calculation. As an average value, per capita income cannot reflect the true income distribution of the study tract, which may cause aggregation bias; (2) the CES data used in Method 4 come from a national sample survey, which ignores the difference between the U.S. and the MSA. A more detailed discussion can be found in Section 4.1.3.

Table 4-8: Example of Food Store Retail Demand Calculation with Method 4

	Procedures	Methods	Units	Values
A	Calculate the MSA wide per capita food store expenditure	A=Total food store expenditure in the MSA/MSA population	thousand $ / person	1.627
B	Calculate the total household expenditure of food stores in the MSA	B=Summation of expenditure of each household income group in the MSA	thoudand $	5,818,020
C	Calculate the total household expenditure of food stores in the tract	C=Summation of expenditure of each household income group in the tract	thousand $	8,766
D	Calculate the average household expenditure of food stores in the MSA	D=B/Number of households in the MSA	thousand $ / household	3.538
E	Calculate the average household expenditure of food stores in the tract	E=C/Number of households in the tract	thousand $ / household	2.975
F	Estimate the adjustment factor of the tract	F=E/D	N.A.	0.841
G	Calculate the per capita expenditure of the tract	G=A*F	thousand $ / person	1.368
H	Estimate the overall expenditure of the tract	H=G*Population	thousand $	7,930
I	Normalize the retail sales by dividing the land area of the tract	I=H/Land area	thousand $ / square mile	46,646

Source: Calculated by the author

4.1.5 Method Five (Household Size Adjusted Per Capita CES Income Adjustment)

Method 5 uses per capita expenditure estimated from Consumer Expenditure Survey's household expenditure data to calculate the local demand indicator. The local supply indicator can be calculated with the same approach as is presented in Section 4.1.3 (Method 3). Therefore, this section only discusses the calculation method for local demand indicator.

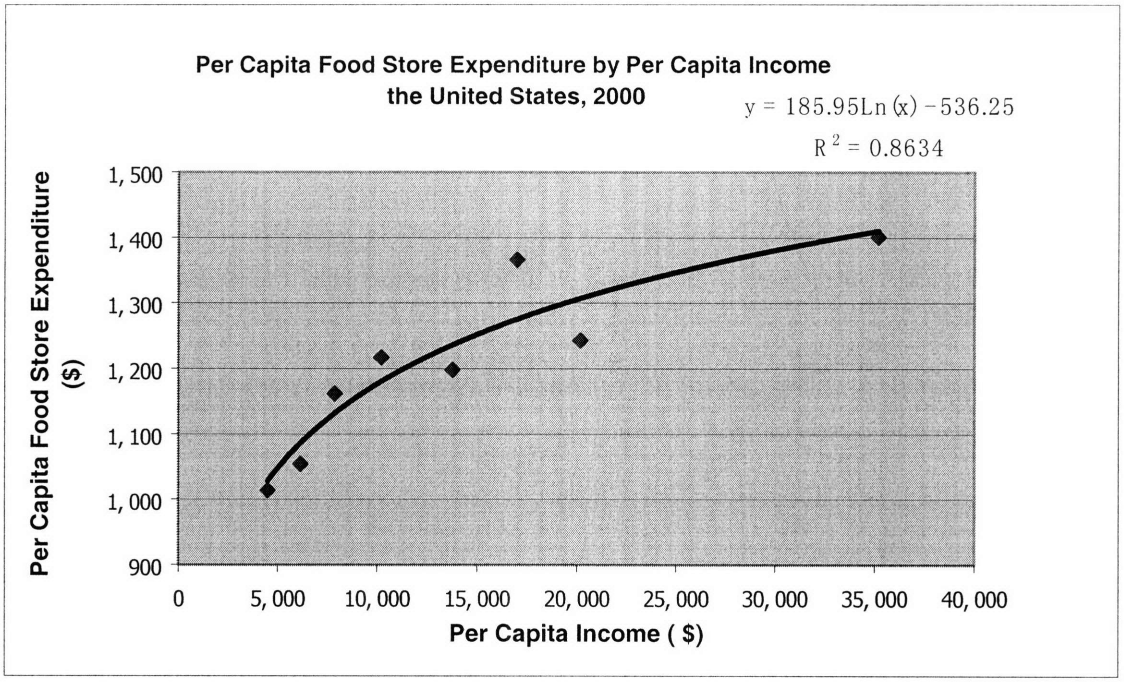
CES data contain information about household size and household expenditure by household income groups. The per capita expenditure and per capita income for each household income group can then be calculated, as is shown in Table 4-9. Using the data in Table 4-9, Figure 4-2 plots per capita food store expenditure by per capita income for the United State in 2000.

Table 4-9: Annual Per Capita Food Expenditure by Per Capita Income, U.S., 2000

	Complete reporting of income							
	$5,000 to $9,999	$10,000 to $14,999	$15,000 to $19,999	$20,000 to $29,999	$30,000 to $39,000	$40,000 to $49,000	$50,000 to $69,000	$70,000 and over
Consumer unit income before taxes	7,638	12,316	17,319	24,527	34,422	44,201	58,561	112,586
Average number in consumer unit	1.7	2.0	2.2	2.4	2.5	2.6	2.9	3.2
Food at home	1,723	2,108	2,556	2,921	2,995	3,552	3,605	4,483
Per capita food at home	1,014	1,054	1,162	1,217	1,198	1,366	1,243	1,401
Per capita income	4,493	6,158	7,872	10,220	13,769	17,000	20,193	35,183

Source: calculated by the author based on 2000 Consumer Expenditure Survey data from the Bureau of Labor Statistics

Figure 4-2: Per Capita Food Store Expenditure by Per Capita Income, U.S., 2000



Source: calculated by the author based on 2000 Consumer Expenditure Survey data from the Bureau of Labor Statistics

The relationship between per capita food store expenditure and per capita income can be presented by a log function (with an R^2 0.863):

$$E_m^i = 185.95 * \ln(INC_m) - 536.25 \qquad (4.7),$$

where E_m^i is the per capita food store expenditure, and INC_m is the per capita income of tract *m*. Taking derivatives on both sides of Equation (4.5), it can be represented by the relationship that $\frac{dE_m^i}{d(INC_m)} = \frac{185.95}{INC_m}$. An interpretation is that as the per capita income grows, the per capita food store expenditure will also increase, while the marginal increase of per capita expenditure will decline. Equation 4.7 is consistent with the common expectation of the relationship between income and expenditure.

Based on Equation 4.7, the complete formulas used to calculate local demand are

$$D_m^i = \frac{E_m^i * POP_m}{A_m}, \; E_m^i = 185.95 * \ln(INC_m) - 536.25 \tag{4.8}$$

The per capita income of tract *m* can be obtained from Census data. Using Equation 4.7, the per capita expenditure in retail category *i* of tract *m* can be calculated from the per capita income in tract *m*. Then the total retail expenditure, i.e., the total retail demand of retail category *i* in tract *m* is the product of per capita expenditure and population in tract *m*.

This method also has two drawbacks: (1) CES is based on a national sample and provides no local area data. The spending pattern of the Boston MSA is very likely different from that of the U.S.; (2) method 5 uses per capita income of a tract to represent the wealthy level of that tract, which may cause aggregation bias as discussed in section 4.1.3.

4.2 Indicator Generation II: Neighboring Area Indicators

Socioeconomic activities like shopping do not adhere to the standard boundaries for aggregating numbers like census tract, block groups, and towns. Sometimes, the target tract indicators themselves could be misleading because they do not incorporate demand and supply situation outside an area of residence. For example, a dense census tract close to the central business district (CBD) may have a very low retail store presence. But it does not necessarily mean that there is unmet retail demand in this tract, because the residents' retail demand can be conveniently met by the substantial retail supply in the nearby CBD area.

Besides retail demand from local residents, retail demand from residents of neighboring areas is also an important source of target community's market potential. Neighboring area indicators consist of several subsets of indicators. Each of them has the same structure as the target tract indicators, including local demand indicator, local supply indicator, local retail leakage indicator, and local purchase ratio indicator. The only difference is that each subset of indicators is calculated for a floating catchment area within a curtain distance from the target tract, rather than limit the calculation to the target tract itself. They describe the retail market situations in larger trade areas.

The neighboring area indicators are calculated based on the target tract indicators. It means that for a set of target tract indicators calculated using one of the five methods presented in Section 4.1, there is a set of neighboring area indicators associated with it. The equations to calculate neighboring area indicators are listed below:

$$S_m^{id} = \frac{\sum_n (S_n^i * A_n)}{\sum_n A_n} \tag{4.9}$$

$$D_m^{id} = \frac{\sum_n (D_n^i * A_n)}{\sum_n A_n} \tag{4.10}$$

$$RLI_m^{id} = D_m^{id} - S_m^{id} \tag{4.11}$$

$$LPI_m^{id} = S_m^{id} / D_m^{id} \tag{4.12}$$

where *d* is a threshold value for distance from a tract's centroid to target tract m's centroid; *n* is the tracts whose centroids are within distance *d* from the centroid of target tract *m*; A_n is the land area of tract *n*; S_n^i and D_n^i are the local supply indicator and local demand indicator of tract *n*; S_m^{id}, D_m^{id}, RLI_m^{id}, and LPI_m^{id} are the local supply indicator, local demand indicator, local retail leakage indicator and local purchase ratio indicator of tracts whose centroids are within distance *d* to the centroid of target tract *m* respectively.

Based on the characteristics of the retail category and study area under research, one or a set of values of distance *d* can be selected. In this study, for the food stores in the Boston MSA, three subsets of neighboring area indicators are calculated. They are the indicators for tracts within 2-mile, 4-mile, and 6-mile buffer to the target tract. The reason to choose 6 miles as the upper bound is that the 6-mile buffer zone of large food stores (with employment greater than 150) can cover almost all census tracts in the Boston MSA, i.e., almost every tract has at least one large food store within 6-mile distance. Therefore, It is reasonable to assume that a consumer will not travel beyond the 6-mile boundary to shop elsewhere.

PL/SQL, Oracle's procedural extension of SQL[21], is extensively used in the calculation. PL/SQL bridges the gap between database technology and procedural programming languages. With the help of PL/SQL, calculation work that used to be cumbersome is a lot easier. The PL/SQL scripts used in this part are listed in the Appendix. These scripts

[21] Many other RDBMS have SQL extensions like PL/SQL.

can be transformed to reusable and tunable modules, and easily applied to other retail categories, other MSAs, as well as aggregation areas with different sizes.

4.3 Indicator Generation III: Other indicators

Besides target tract indicators and neighboring area indicators, other elements considered in this study include the adjusted density of local employees and the inner-city location indicator.

4.3.1 Adjusted Density of local Employees

Employees may shop at stores close to their working place to save time and transportation cost. Adjusted density of local employees is a proxy of retail demand from non-resident employees, assuming all the employees across the MSA are identical consumers. The adjusted density of local employees W_m is calculated with the following formula[22]:

$$W_m = (JOB_m - WORKER_m) / A_m \qquad (4.13),$$

where JOB_m is the total number of jobs in tract m, $WORKER_m$ is the total number of workers living in tract m, and A_m is the land area of tract m. The total number of jobs and total number of workers in a tract are calculated from the 2000 CTPP data[23].

---

[22] The reason to use jobs minus workers rather than jobs in Equation 4.13 is that workers living in tract m while working outside m may also shop near their working place.

[23] The census tract level numbers of jobs and workers of the Boston MSA used in this study are calculated by Jiawen Yang from the Department of Urban Studies and Planning at MIT for a project called 'The Effectiveness of Job-Housing Balance as a Congestion Relief Strategy', and funded by the US Department of Transportation through their Region One (New England) University Transportation Center (UTC) research.

4.3.2 Inner-city Location Indicator

Inner-city census tracts can be selected based on their socioeconomic characteristics. Based on ICIC's definition, inner cities are tracts that have 20% poverty rate or higher or meet two of the following three criteria:

- poverty rate of 1.5 times or more that of their MSAs;
- median household income of 1/2 or less that of their MSAs;
- unemployment rate of 1.5 or more that of their MSAs.

The threshold values for the last three criteria in the Boston MSA are presented in Table 4-10.

Table 4-10: Threshold Values for the Definition of Inner City in the Boston MSA

Variable	Value of Boston-Worcester-Lawrence CMSA	Threshhold Value
Poverty Rate	0.086	0.128
Median Household Income	52,471	26,236
Unimployment Rate	0.042	0.064

Source: Calculated by the author based on US 2000 census data.

Among the 894 census tracts in the Boston MSA, 165 tracts are classified as inner-city neighborhoods. They mainly concentrate in the central part of Boston and traditional manufacturing centers like Lawrence and Lowell, as is shown in Figure 1-1.

4.4 Summary of Indicator Generation

Table 4-11 provides a summary of some key indicators calculated with the five methods for a tract with tract ID '25025000401'. Table 4-12 presents the comparison of the MSA average of these indicators using different calculation methods.

As is shown in Table 4-11 and Table 4-12, retail markets indicators are sensitive to the assumptions about consumers' purchasing patterns. The standard deviations of target tract demand and target tract supply indicator are around 10% of their mean values, which is reasonable. Since retail gap indicators measure the difference between estimated supply and demand, they are more volatile, especially for the buffer where the gap is relatively small since it is aggregated over many tracts. When practitioners refer to a market analysis report, they should always check the assumptions of the reports and decide whether or not the assumptions fit the neighborhood condition, and in what direction the bias, if any, might be.

Table 4-11: Comparison of Indicators Calculated with Alternative Methods for Census Tract '25025000401'

Indicators	Method 1	Method 2	Method 3	Method 4	Method 5	Mean	Standard Deviation
Retail Demand in the Target Tract (thousand $/sqmi)	55,471.1	44,462.1	51,562.7	46,651.0	45,274.8	48,684.3	4,688.6
Retail Supply in the Target Tract (thousand $/sqmi)	5,789.5	5,789.5	4,807.0	5,789.5	4,820.9	5,399.2	534.3
Retail Gap in the Target Tract (thousand $/sqmi)	49,681.7	38,672.6	46,755.7	40,861.6	40,453.9	43,285.1	4,693.8
Retail Gap in Tracts within 2-Mile Buffer (thousand $/sqmi)	4,872.5	8,165.0	5,632.5	4,544.1	4,393.1	5,521.4	1,553.2
Retail Gap in Tracts within 4-Mile Buffer (thousand $/sqmi)	-376.8	2,042.6	798.7	-1,000.7	-140.0	264.8	1,185.6
Retail Gap in Tracts within 6-Mile Buffer (thousand $/sqmi)	355.1	882.4	750.0	-263.9	272.8	399.3	451.2

Source: Calculated by the author

Table 4-12: Comparison of the MSA Average of Indicators Calculated with Alternative Methods

Indicators	Method 1	Method 2	Method 3	Method 4	Method 5	Mean	Standard Deviation
Retail Demand in the Target Tract (thousand $/sqmi)	14,106.2	12,449.9	11,797.3	13,084.2	11,457.6	12,579.1	1,057.5
Retail Supply in the Target Tract (thousand $/sqmi)	12,970.7	12,970.7	10,805.9	13,014.4	10,800.6	12,112.5	1,195.3
Retail Gap in the Target Tract (thousand $/sqmi)	1,135.5	-520.8	991.4	69.9	657.0	466.6	687.6
Retail Gap in Tracts within 2-Mile Buffer (thousand $/sqmi)	-558.6	-1,428.1	-497.6	-1,184.0	-622.1	-858.1	420.3
Retail Gap in Tracts within 4-Mile Buffer (thousand $/sqmi)	-160.8	-589.6	-69.6	-583.5	-220.3	-324.8	244.9
Retail Gap in Tracts within 6-Mile Buffer (thousand $/sqmi)	-19.6	-263.9	74.2	-297.9	-69.7	-115.4	160.2

Source: Calculated by the author

4.5 Spatial Patterns of Food Store Retail Markets in the Boston MSA

As a direct application of the retail markets neighborhood indicator systems, this section reveals the spatial patterns of the food store markets in the Boston MSA by shading the neighborhood indicators thematically using GIS tools. All the indicators mapped in this section are calculated based on Method 5.

Figure 4-5 and 4-6 are maps of food store demand and food store supply of the target tract respectively. For the convenience of comparison, they use the same quantile classification. These maps reveal an interesting difference in the spatial distributions of food store demand and supply: substantial purchasing power is concentrated in the core of the MSA, while the food store supply is much more dispersed.

One possible reason for the observed spatial pattern of food store demand is that in the urban core, the impact of higher population density on aggregate purchasing power exceeds the impact of lower income level. As is shown in Table 4-13, by average the per capita expenditure of inner-city tracts is about 0.91 times that of non inner-city tracts, while the average population density of the inner-city tracts is about 3.25 times that of non inner-city tracts. Overall, the inner-city group has an average food store expenditure 2.96 times that of the non inner-city group.

Table 4-13: Food Store Expenditure Characteristics by Census Tract Group

Census Tract Group	Average Population Density (number of people/sq. mi.)	Average Per Capita Expenditure (thousand $/person*year)	Average Total Expenditure Density (thousand $/year*sq. mi.)
Inner-City Group	19,919.8	1.247	24,925.1
Non-Inner-City Group	6,123.9	1.373	8,409.4

Source: Calculated by the author

The spatial distribution of food store supply shows two patterns. One apparent pattern is that many tracts with high food store supply level are located near the transportation corridors. This pattern is understandable because these stores can take advantage of their locations to reduce consumer's travel costs, and attract more business from neighboring areas. For example, 67 out of 95 food stores with an employee number greater than 150 are located within the one-mile buffer of a major highway, as is shown in Figure 4-4. Another pattern is that within the urban core, food store supply of some tracts is very high, while that of another group of tracts is extremely low. The former group of tracts is usually located in the CBD area with huge number of employees and visitors, while the latter group is typically poor residential tracts. This pattern suggested that demand from non-resident employees and visitors may be important factors that can influence a tract's food store supply level.

Figure 4-7 is a map of the food store retail gaps in the target tract. In this map, tracts are divided into two groups. One group of tracts with positive retail gaps (demand exceeds supply) is shaded with warm colors (yellow to brown) and quantile classification method within the group. Another group with negative retail gaps (supply exceeds demand) is shaded with cold colors (light blue to dark blue) and quantile classification method within the group. Figure 4-7 suggests that census tracts in the urban core are polarized in terms of their food store retail gaps. Tracts within the CBD area usually have extremely high negative retail gaps, while other tracts show high positive retail gaps. This finding is consistent with the previous analysis for the spatial patterns of retail supply and demand.

Figure 4-8, 4-9 and 4-10 shades the food store retail gaps in trade areas with various sizes – within 2 mile, 4 miles and 6 miles buffers zone of the target tract respectively. For the convenience of comparison, they use the same color scheme and dividing points as Figure 4-7. Comparing these maps with Figure 4-7, it can be found that the larger the trade area's size, the smoother the retail gap surfaces in the maps and the smaller the magnitude of the retail gaps. It suggests that the food store demand and supply are more likely to be balanced as the size of the study area increases.

Though the maps discussed above illustrate the spatial patterns of the food store retail markets vividly, to test the spatial correlation of inner-city neighborhoods and underserved areas, some quantitative analyses are necessary. There are multiple factors that may affect the retail sale level of the target tract such as retail demand from neighboring areas and non-resident employees. The location factor is only one of them. To show the 'pure' impact of the location factor, other factors need to be controlled. In Chapter 5, regression analyses are implemented to extract the 'pure' impact of an inner-city location on local retail supply level.

Figure 4-3: Food Stores in the Boston MSA

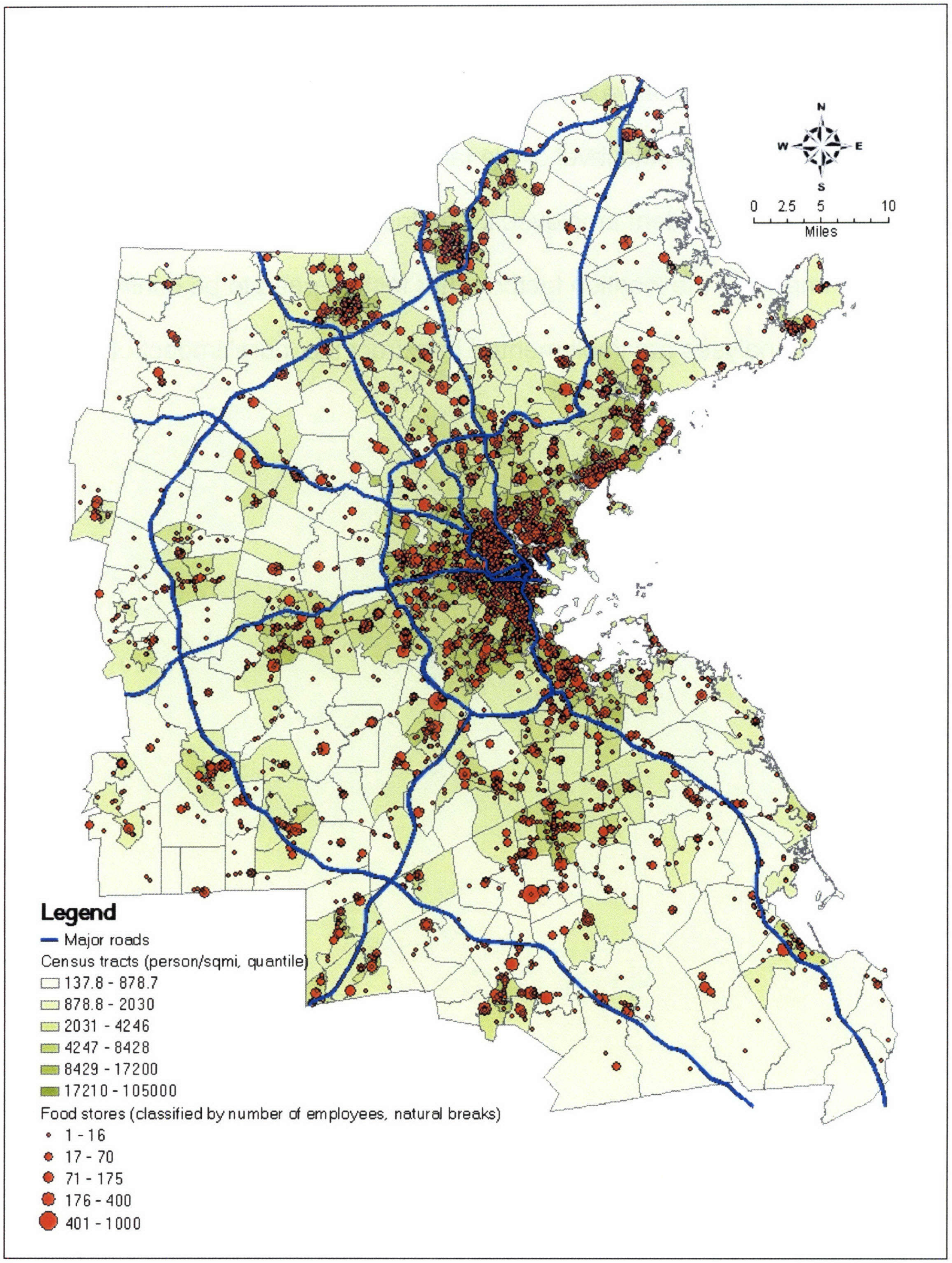


Figure 4-4: Large Food Stores (with 150 Employees or More) in the Boston MSA

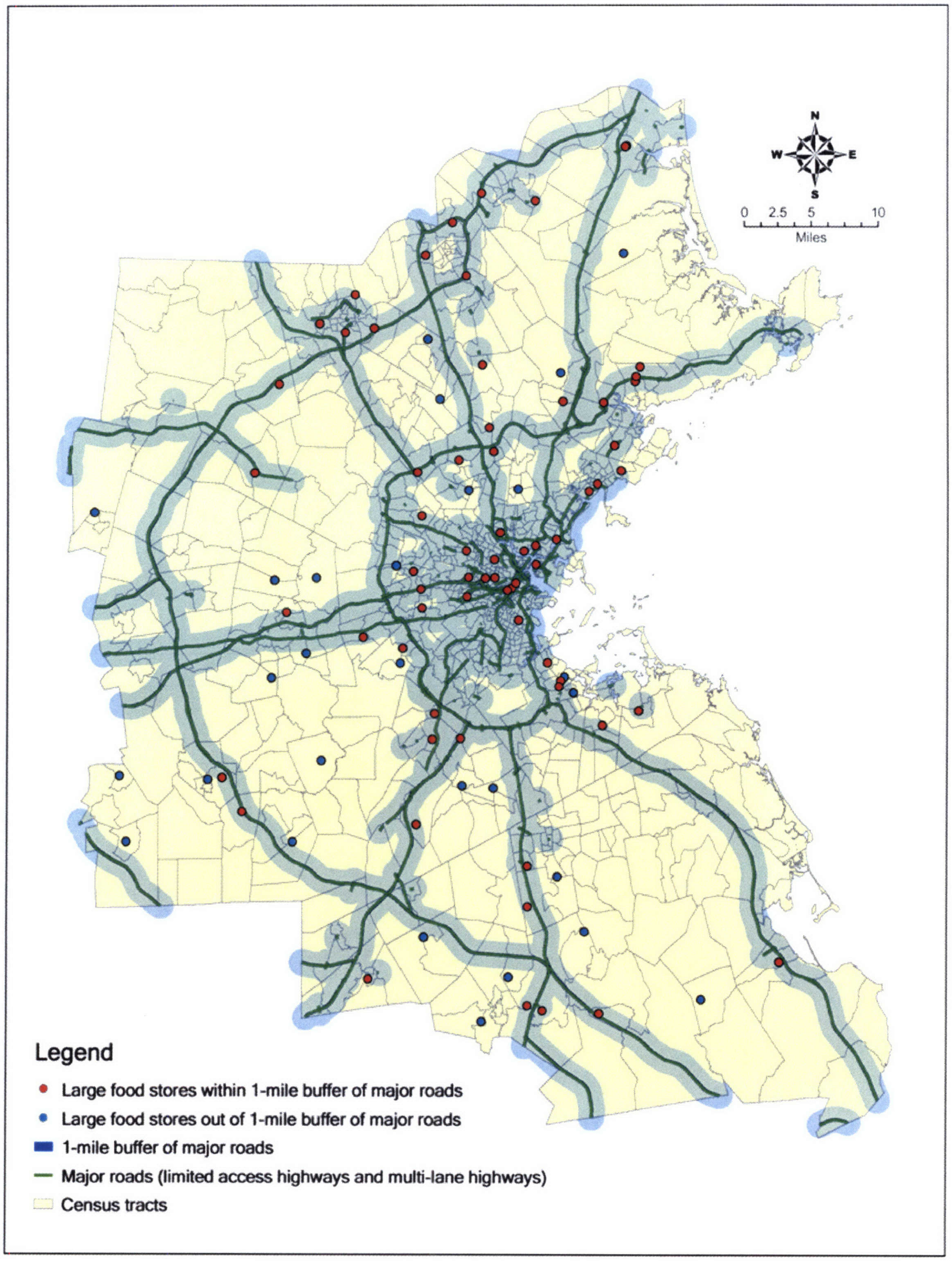
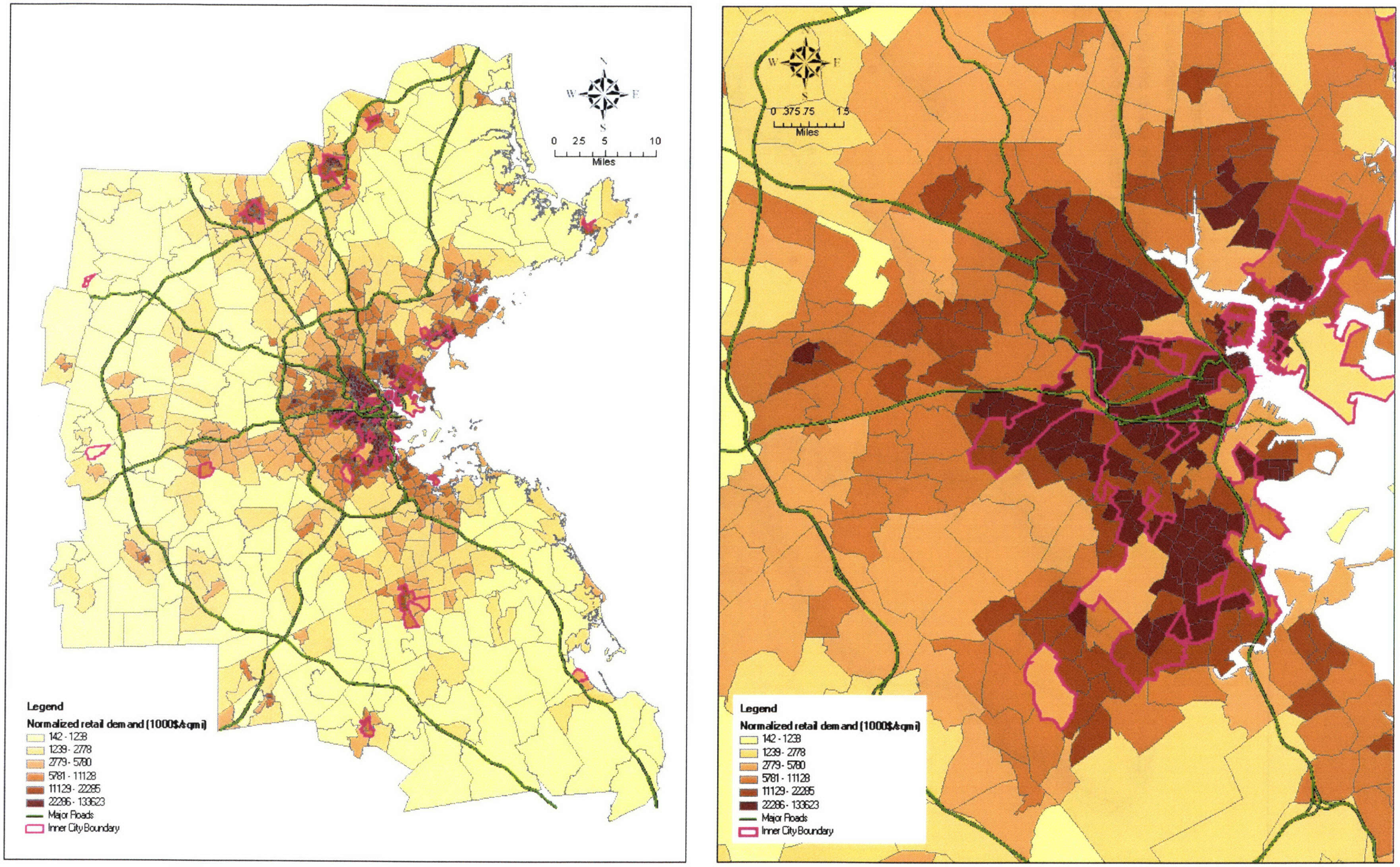
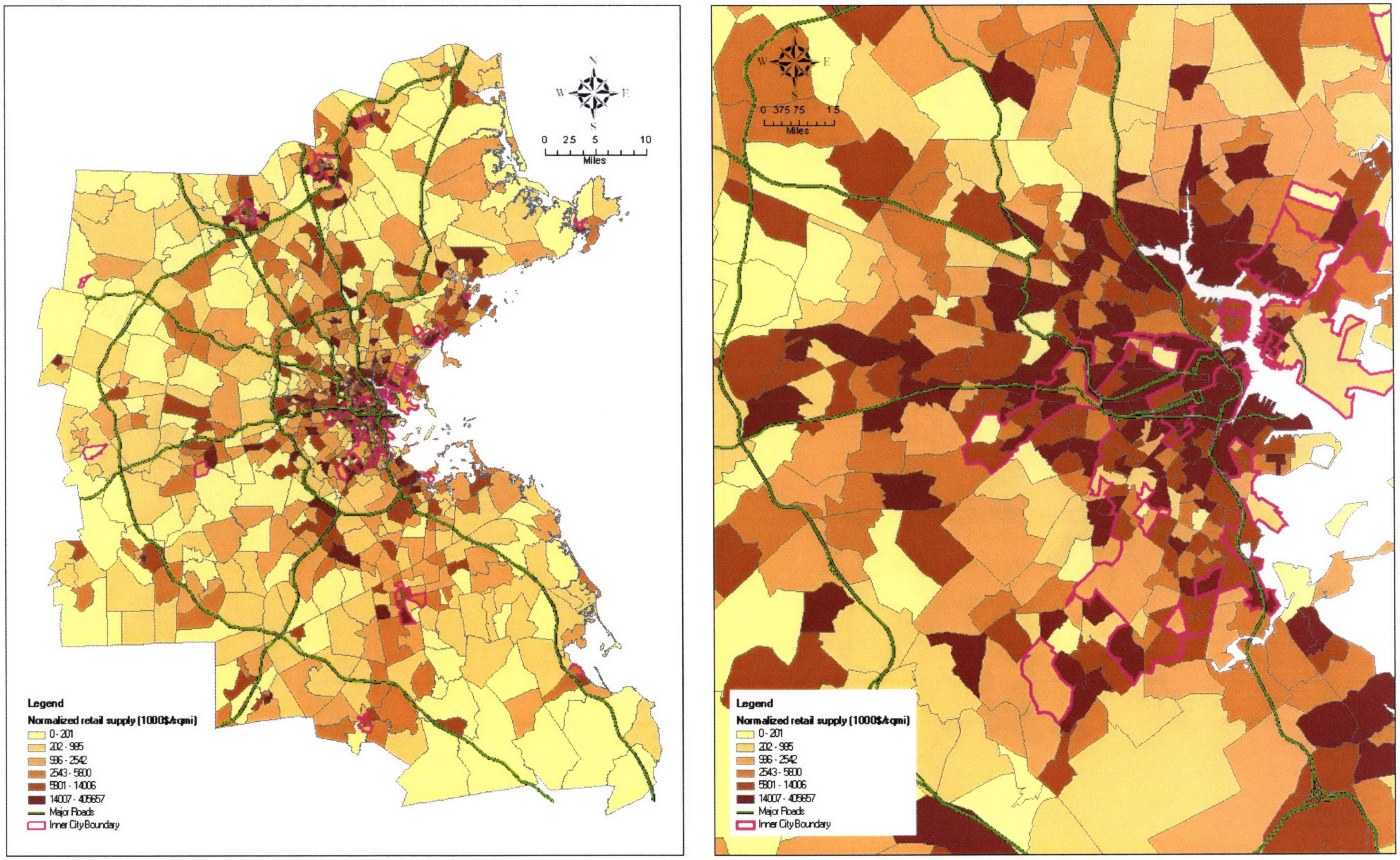


Figure 4-5: Food Store Demand in the Target Tract



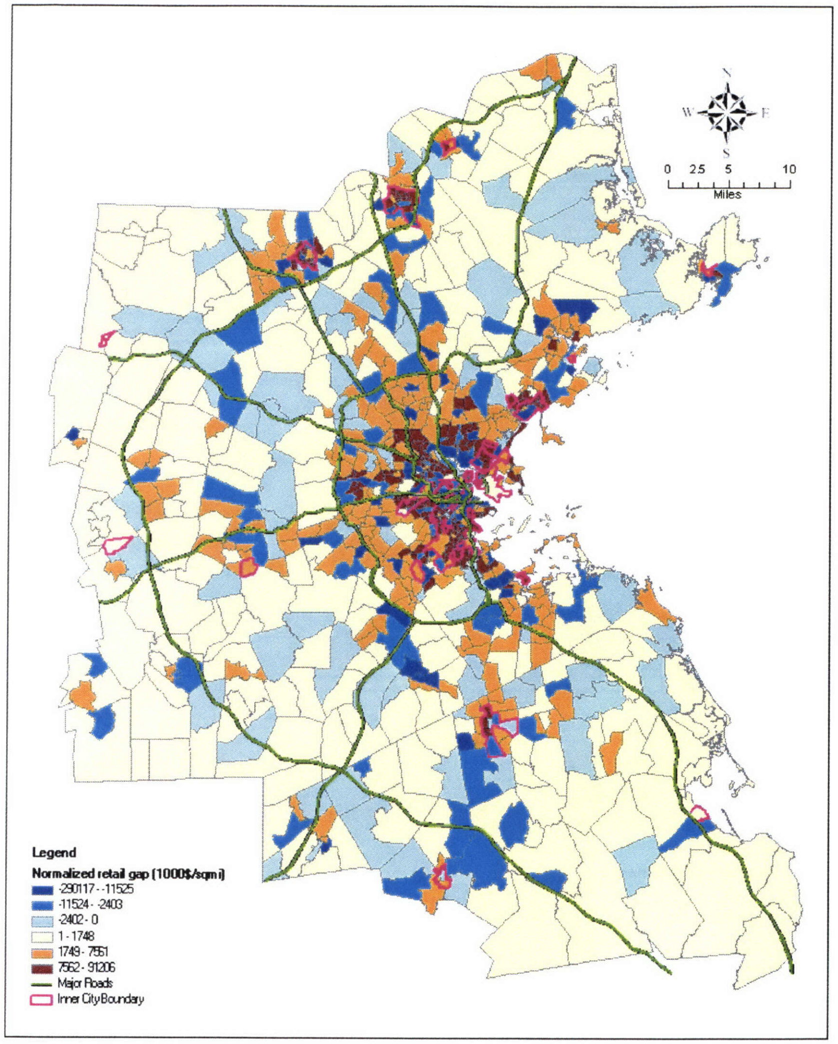
Note: Indicators are calculated based on Method 5 (household size adjusted per capita CES income adjustment).

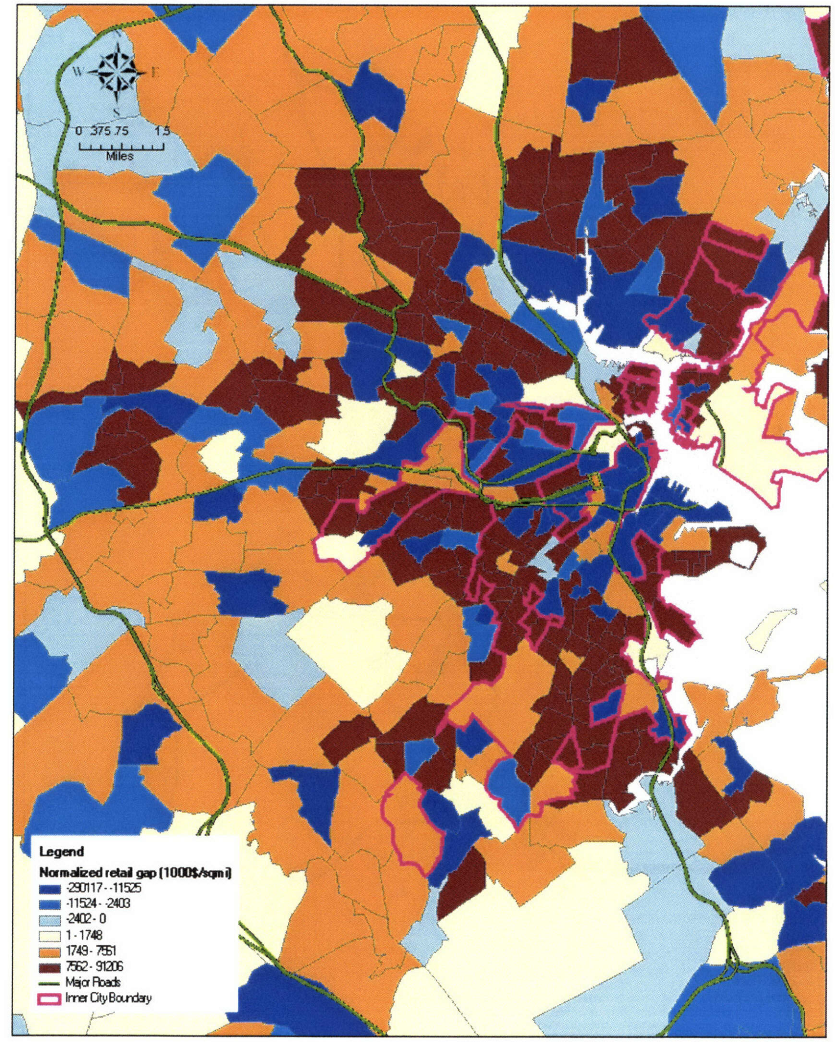
Figure 4-6: Food Store Supply in the Target Tract



Note: Indicators are calculated based on Method 5 (household size adjusted per capita CES income adjustment).

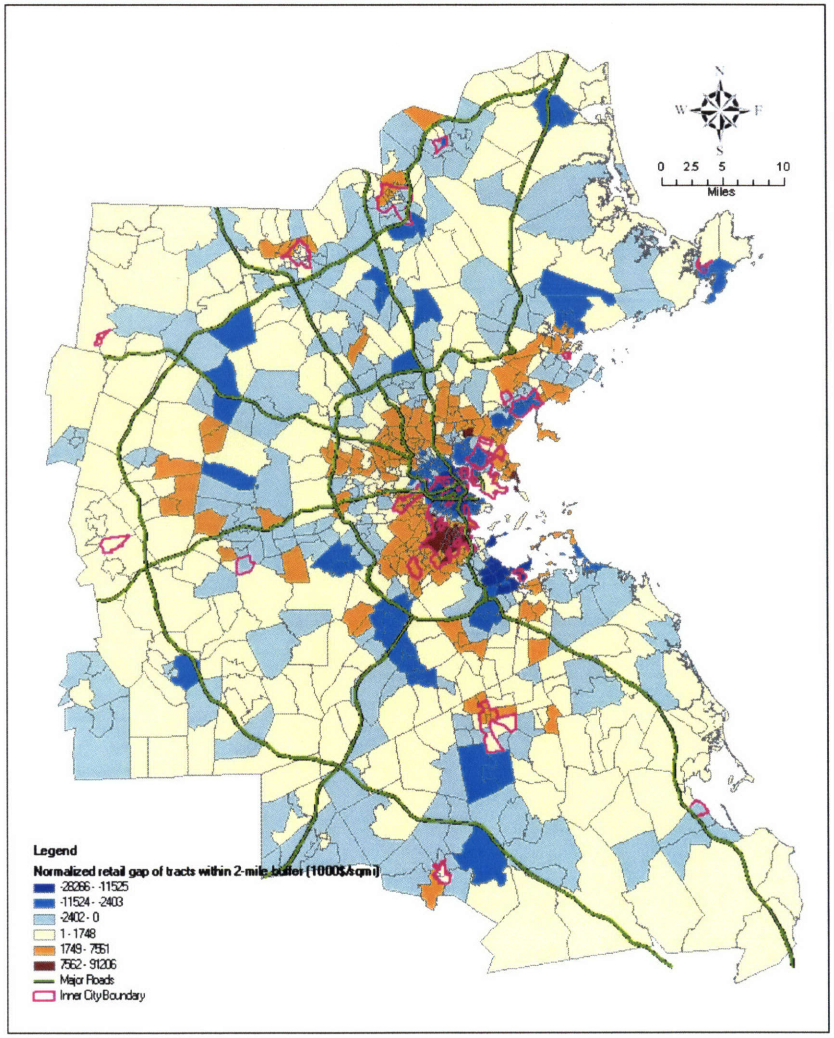
Figure 4-7: Food Store Retail Gap in the Target Tract

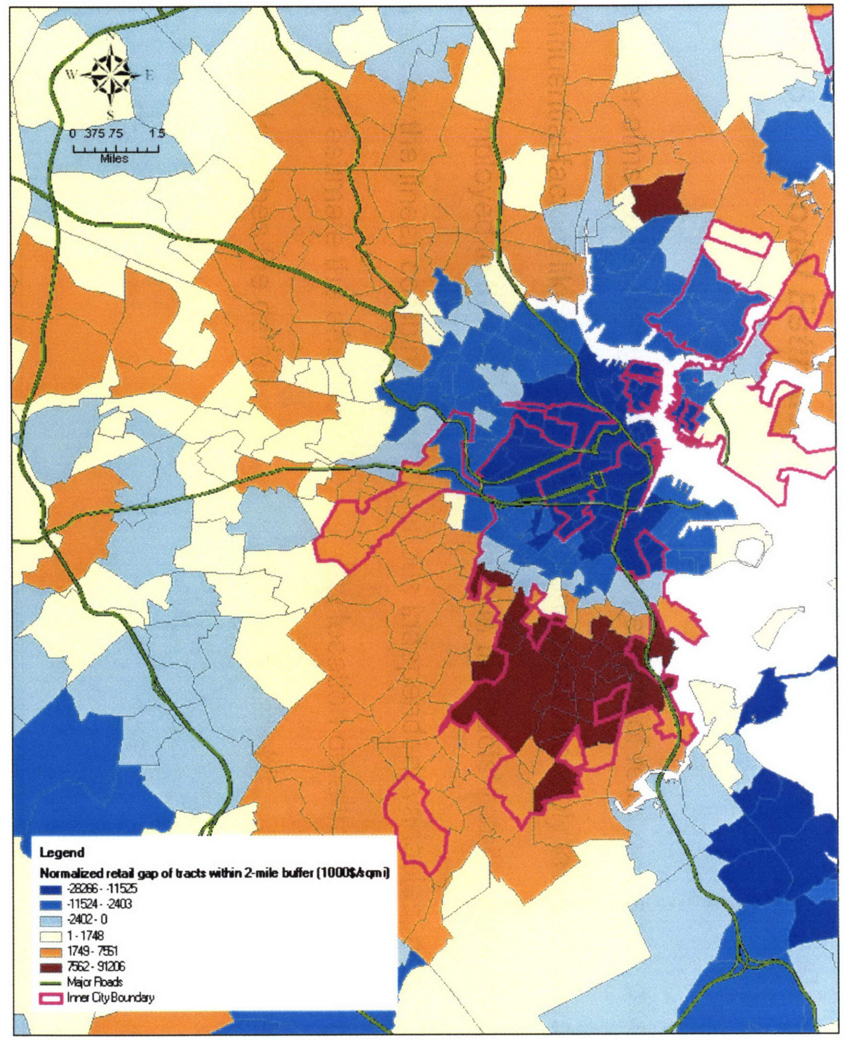




Note: Indicators are calculated based on Method 5 (household size adjusted per capita CES income adjustment). Positive retail gap means demand exceeds supply. Negative retail gap means supply exceeds demand.

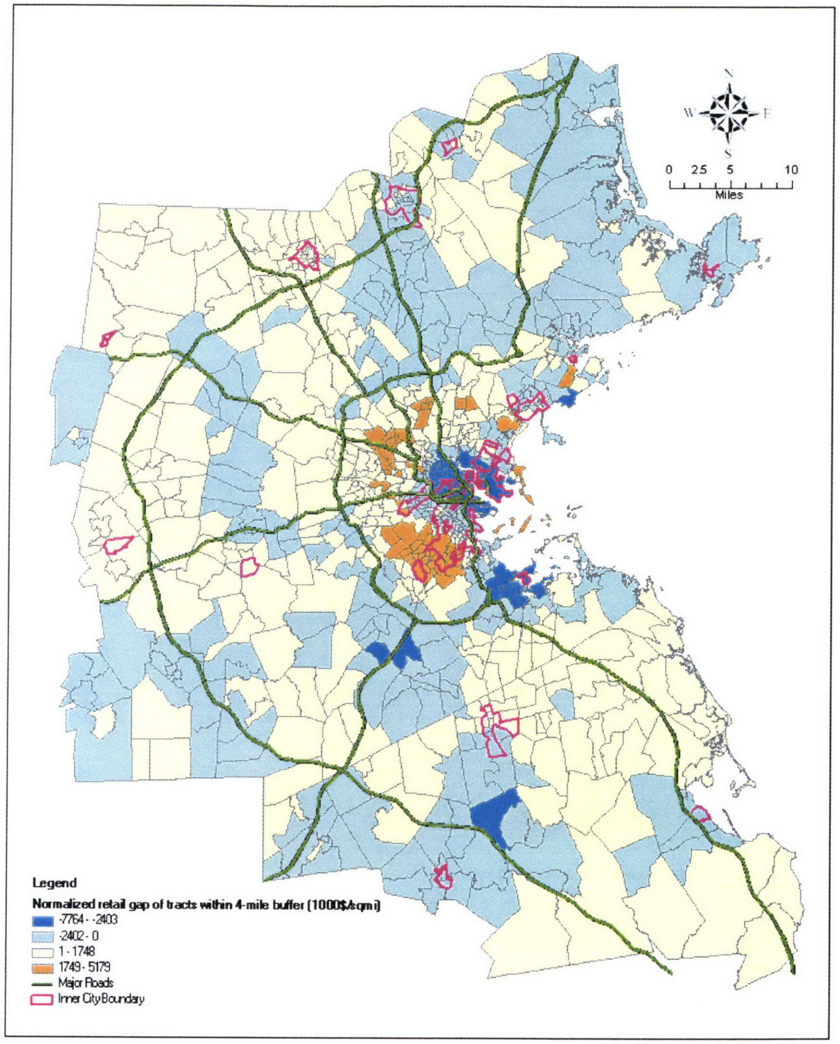
Figure 4-8: Food Store Retail Gap within 2-Mile Buffer Zone of the Target Tract

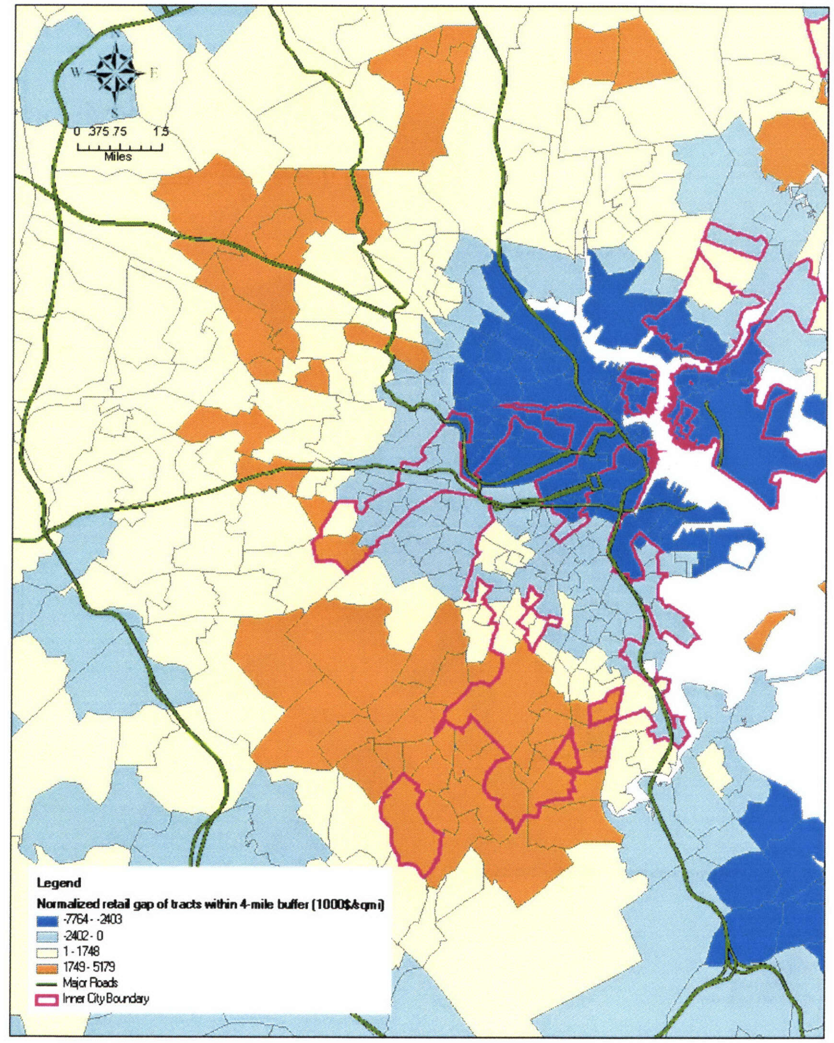




Note: Indicators are calculated based on Method 5 (household size adjusted per capita CES income adjustment). Positive retail gap means demand exceeds supply. Negative retail gap means supply exceeds demand.

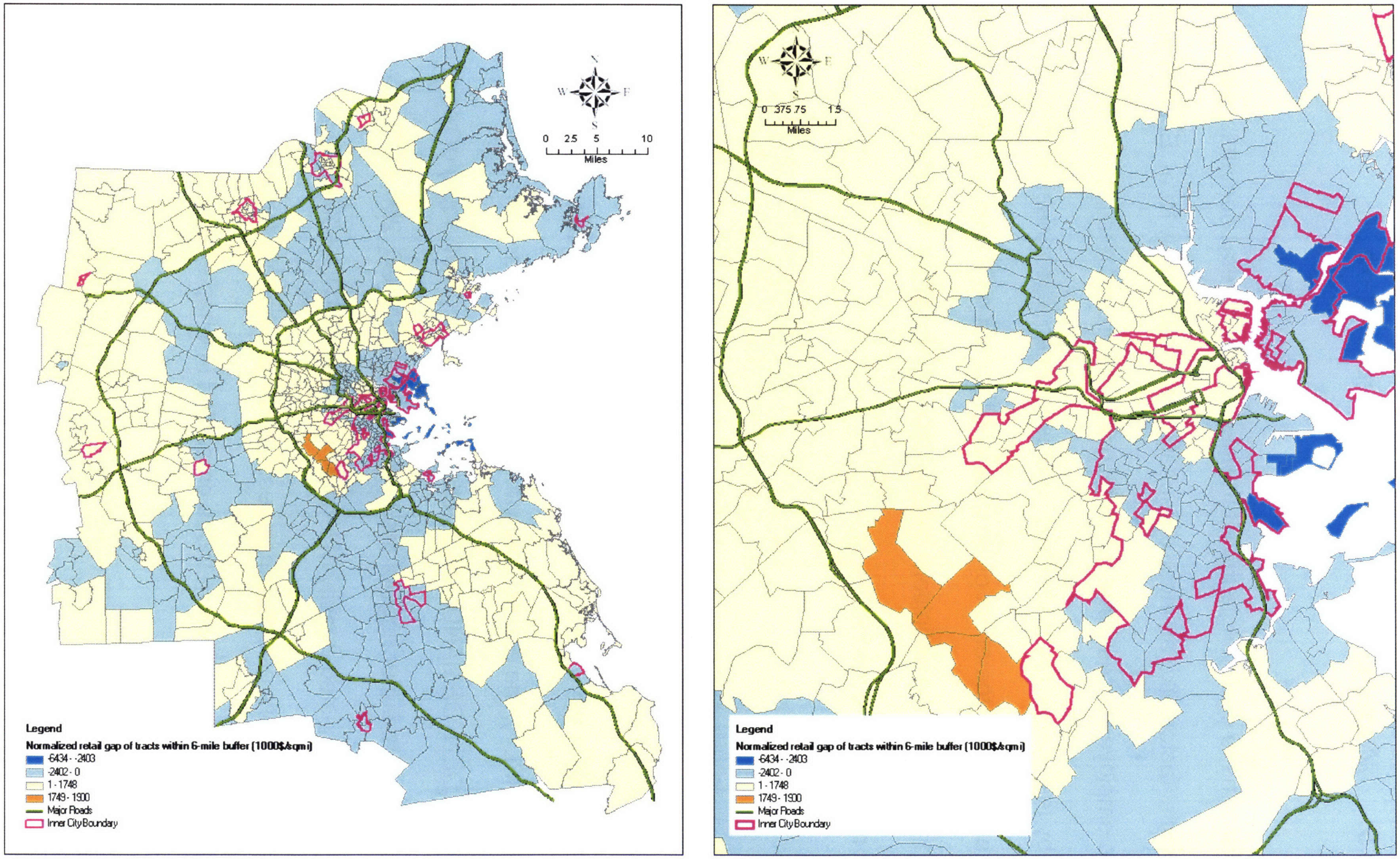
Figure 4-9: Food Store Retail Gap within 4-Mile Buffer Zone of the Target Tract





Note: Indicators are calculated based on Method 5 (household size adjusted per capita CES income adjustment). Positive retail gap means demand exceeds supply. Negative retail gap means supply exceeds demand.

Figure 4-10: Food Store Retail Gap within 6-Mile Buffer Zone of the Target Tract



Note: Indicators are calculated based on Method 5 (household size adjusted per capita CES income adjustment). Positive retail gap means demand exceeds supply. Negative retail gap means supply exceeds demand.

Chapter 5: Model Estimation and Interpretation

This chapter aims to disentangle the intertwined effects of retail demands, distance and other influential factors like an inner-city location on retail supply level. Econometric models are employed to test whether variations in the tracts' retail supply level can be explained by the linear combinations of various independent variables, and quantitatively estimate the effect of an inner-city location on a tract's retail supply level. Section 5.1 describes the definitions and statistics of variables. Section 5.2 presents the estimation results of the model.

5.1 Variable Descriptions

Based on the model specification presented in Chapter 3, this section describes the variables included in the model. The dependent variable is the retail supply in the target tract (DEN_SUPPLY); the explanatory variables considered are retail demand of target tract residents (DEN_DEMAND), adjusted retail demand (retail gaps) of residents from tracts within different buffer rings of the target tracts (DEN_G0_2MI, DEN_G2_4MI, and DEN_G4_6MI)[24], adjusted density of local employees (DEN_WORKER, as a proxy of retail demand from non-resident employees), and the inner-city location dummy variable (INNERCITY). Among the five methods presented in Chapter 4 to calculate the

[24] This study considers the impact of three groups of neighboring tracts on the target tract, whose centroids fall in 0-2, 2-4, 4-6-mile buffer rings of the centroid of the target tract respectively. The reason to choose 6 miles as the upper bound is that the 6-mile buffer zone of large food stores (with employment greater than 150) can cover almost all census tracts in the Boston MSA, i.e., almost every tract has at least one large food store within 6-mile distance. Therefore, It is reasonable to assume that a consumer will not travel beyond the 6-mile boundary to shop elsewhere.

variables, Method 5 (household size adjusted per capita CES income adjustment) are applied in this chapter, which is based on the per capita expenditure by per capita income relation generated from the household expenditure by household income relation provided by the CES data. Table 5-1 reports the definition and descriptive statistics of variables included in the model estimation.

Table 5-1: Descriptive Statistics for Variables (All Tracts, Method 5)

Variable	Description	Mean	Std. Dev.	Max	Min
Dependant Variable					
DEN_SUPPLY	Retail Supply in the target tract (k$/sqmi)	10,801	29,905	405,657	0
Independent Variable: Retail Demand of Local Residents					
DEN_DEMAND	Retail demand from residents in the target tract (k$/sqmi)	11,458	14,699	133,624	142
Independent Variable: Retail Demand of Residents of Neighboring Areas					
DEN_G0_2MI	Adjusted retail demand (retail gap) of tracts whose centroids fall in 2 mile buffer of the target tract's centroid (k$/sqmi) *	-699	5,556	10,124	-28,770
DEN_G2_4MI	Adjusted retail demand from tracts whose centroids fall in 2-4 mile buffer of centroid of the target tract (k$/sqmi)	-129	1,939	5,976	-16,863
DEN_G4_6MI	Adjusted retail demand from tracts whose centroids fall in 4-6 mile buffer of centroid of the target tract (k$/sqmi)	61	2,000	6,188	-10,167
Independent Variable: Retail Demand of Non-Resident Employees					
D_WORKER	Adjusted density of local employees (person/sqmi)	312	14,617	209,683	-55,732
Independent Variable: Location Characteristics					
INNERCITY	Inner city location dummy	0.18	0.39	1	0
Number of Observations: 894					

* The target tract itself is excluded from calculation.

Note: Variables are calcuated with method 5 (household size adjusted per capita CES income adjustment).

It should be noted that in Table 5-1, adjusted demand means retail gap (obtained by subtracting retail supply from retail demand) divided by land area of the aggregation area; adjusted density of local employees is the margin between employees working in the tract and workers living in the tract divided by the land area of the tract. In this research, densities of retail demand or supply (measured by thousand dollars per square mile) and density of local employees (measured by person per square mile) are actually used rather than the volume of retail demand or supply (measured by thousand dollars) and local employees (measured by person) to avoid the size effect of tracts. For example, a tract with a slightly higher volume of retail sales does not necessarily have a higher retail supply level than another tract, because its land area may be much bigger than that of the second tract. For the convenience of narration, 'retail demand' and 'retail supply' are still used hereafter, though they are actually normalized by the land area of the aggregation unit.

The statistics indicate that the mean value of retail supply (DEN_SUPPLY) in the Boston MSA is 10.8 million dollar per square mile with a standard deviation 29.9 million dollar per square mile, indicating the large variation across tracts. The average retail demand (DEN_DEMAND) is about 11.5 million dollars per square mile, while the range of this variable is also large with a standard deviation of 14.7 million dollars per square mile, a maximum 134 million dollars per square mile, and a minimum 0.142 million dollars per square mile. The average of the adjusted retail demand in tracts within 2-mile buffer ring of the target tract (DEN_G0_2MI, target tract itself excluded) is –0.700 million dollars per square mile, while this demand could reach as high as 10.1 million dollars per square mile, and as low as –28.8 million dollars per square mile. The mean of the adjusted retail demand in tracts within 2-4-mile buffer ring (DEN_G2_4MI) is –0.129 million dollars per square mile, while that of tracts within 4-6-mile buffer ring (DEN_G4_6MI) is -0.061 million dollars per square mile. The adjusted density of local employees (DEN_WORKER) also has a high variation. The maximum value can reach 209,683 people per square mile, while the minimum value is as low as -55,732 people per square mile. The mean value is 312 employees per square mile. The inner-city location dummy variable equals 1 if a tract belongs to inner-city neighborhoods; otherwise a value 0 is assigned. Overall, 18.5 percent of the tracts fall in the category of inner-city neighborhoods, while 81.5 percent of the tracts are non inner-city tracts.

Table 5-2 and Table 5-3 provide the variable statistics for inner-city tracts and non inner-city tracts respectively. The average supply and demand of the non inner-inner city group are almost balanced; while the inner-city group's average demand is about 1.25 time its average supply. Comparing the two groups, although the average supply of inner-city group doubles that of the inner-city group, the average demand of inner-city group is almost 3 times that of non inner-city group. This finding is consistent with the argument that when normalized by land area, inner-city tracts have larger aggregate purchasing power than the non inner-city tracts, and they tend to be underserved in the retail markets.

Table 5-2: Descriptive Statistics for Variables (Inner-City Tracts)

Variable	Description	Mean	S.D	Max	Min
Dependant Variable					
DEN_SUPPLY	Retail Supply in the target tract (k$/sqmi)	19,824	33,738	238,334	0
Independent Variable: Retail Demand of Local Residents					
DEN_DEMAND	Retail demand from residents in the target tract (k$/sqmi)	24,925	16,980	133,624	142
Independent Variable: Retail Demand of Residents of Neighboring Areas					
DEN_G0_2MI	Adjusted retail demand (retail gap) of tracts whose centroids fall in 2 mile buffer of the target tract's centroid (k$/sqmi)	-1,351	7,847	9,892	-21,771
DEN_G2_4MI	Adjusted retail demand (retail gap)from tracts whose centroids fall in 2-4 mile buffer ring of centroid of the target tract (k$/sqmi)	-633	2,504	5,976	-9,064
DEN_G4_6MI	Adjusted retail demand (retail gap) from tracts whose centroids fall in 4-6 mile buffer ring of centroid of the target tract (k$/sqmi)	211	2,490	5,241	-7,329
Independent Variable: Retail Demand of Non-Resident Employees					
D_WORKER	Adjusted density of local employees (person/sqmi)	3,505	28,261	209,683	-32,765
Number of Observations: 165					

Note: Varibles are calculated with method 5 (household income adjusted per capita CES income adjustment)

Table 5-3: Descriptive Statistics for Variables (Non Inner-City Tracts)

Variable	Description	Mean	S.D.	Max	Min
Dependant Variable					
DEN_SUPPLY	Retail Supply in the target tract (k$/sqmi)	8,758	28,600	405,657	0
Independent Variable: Retail Demand of Local Residents					
DEN_DEMAND	Retail demand from residents in the target tract (k$/sqmi)	8,409	12,234	131,050	166
Independent Variable: Retail Demand of Residents of Neighboring Areas					
DEN_G0_2MI	Adjusted retail demand (retail gap) of tracts whose centroids fall in 2 mile buffer of the target tract's centroid (k$/sqmi) *	-551	4,886	10,124	-28,770
DEN_G2_4MI	Adjusted retail demand (retail gap) from tracts whose centroids fall in 2-4 mile buffer of centroid of the target tract (k$/sqmi)	-15	1,769	5,676	-16,863
DEN_G4_6MI	Adjusted retail demand (retail gap) from tracts whose centroids fall in 4-6 mile buffer of centroid of the target tract (k$/sqmi)	27	1,872	6,188	-10,167
Independent Variable: Retail Demand of Non-Resident Employees					
D_WORKER	Adjusted density of local employees (person/sqmi)	-411	8,906	138,926	-55,732
Number of Observations: 729					

Note: Varibles are calculated with method 5 (household income adjusted per capita CES income adjustment)

5.2 Results Interpretation

Using the explanatory variables discussed in Section 5.1, this section develops multivariate linear regression models to estimate the effects of explanatory variables. The dependent variable in the models is the food store retail supply (DEN_SUPPLY).

5.2.1 Ordinary Least Square Estimation

Ordinary least square (OLS) regression is run first. The estimation results are presented in Table 5-4.

Table 5-4: Estimation Results of OLS

Variables	Coefficients	Std. Error	Beta	t	Sig.
(Constant)	-1,889.410	997.201	.	-1.89	0.058
DEN_DEMAND	1.323	0.065	0.651	20.45	0.000
DEN_G0_2MI	0.484	0.189	0.090	2.56	0.011
DEN_G2_4MI	0.598	0.442	0.039	1.35	0.176
DEN_G4_6MI	2.036	0.508	0.136	4.01	0.000
DEN_WORKER	0.633	0.057	0.309	11.08	0.000
INNERCITY	-12,889.490	2,283.021	-0.167	-5.65	0.000
Number of observations	894				
Adjusted R^2	0.41				
F	102.58				
Sig.	0.000				

Dependent Variable: DEN_SUPPLY

There are occasions in econometric modeling when the assumption of homoscedasticity is unreasonable. With the presence of heteroscedasticity, OLS estimators are still unbiased, consistent, but they are not efficient. In addition, the OLS estimated variance of the estimated coefficients would be biased estimator of the true variance. Therefore, the statistical inference given by OLS is invalid.

The White test is used in this study to test the assumption of homoscedasticity. The test statistic given by STATA[25] is 584.17. By comparing the test statistic with the critical value of χ^2 distribution with 26 degree of freedom, the null hypothesis of homoscedasticity can be rejected at the 0.05 level of significance, which means the standard errors and t-statistics in the OLS estimation are invalid.

The spatial spillover effects may lead to the spatial autocorrelation of error terms. Under this circumstance, the OLS estimators lose efficiency. Though the model proposed in this study captures the spillover effects by including neighboring area variables in the model specification, further test is needed to ensure that the no additional spatial autocorrelation remains unaccounted for.

The Moran'I test is used to test the existence of spatial autocorrelation. Two alternative spatial weights matrices are applied in the tests: the rook contiguity based spatial weights and the distance band spatial weights. Rook contiguity uses common

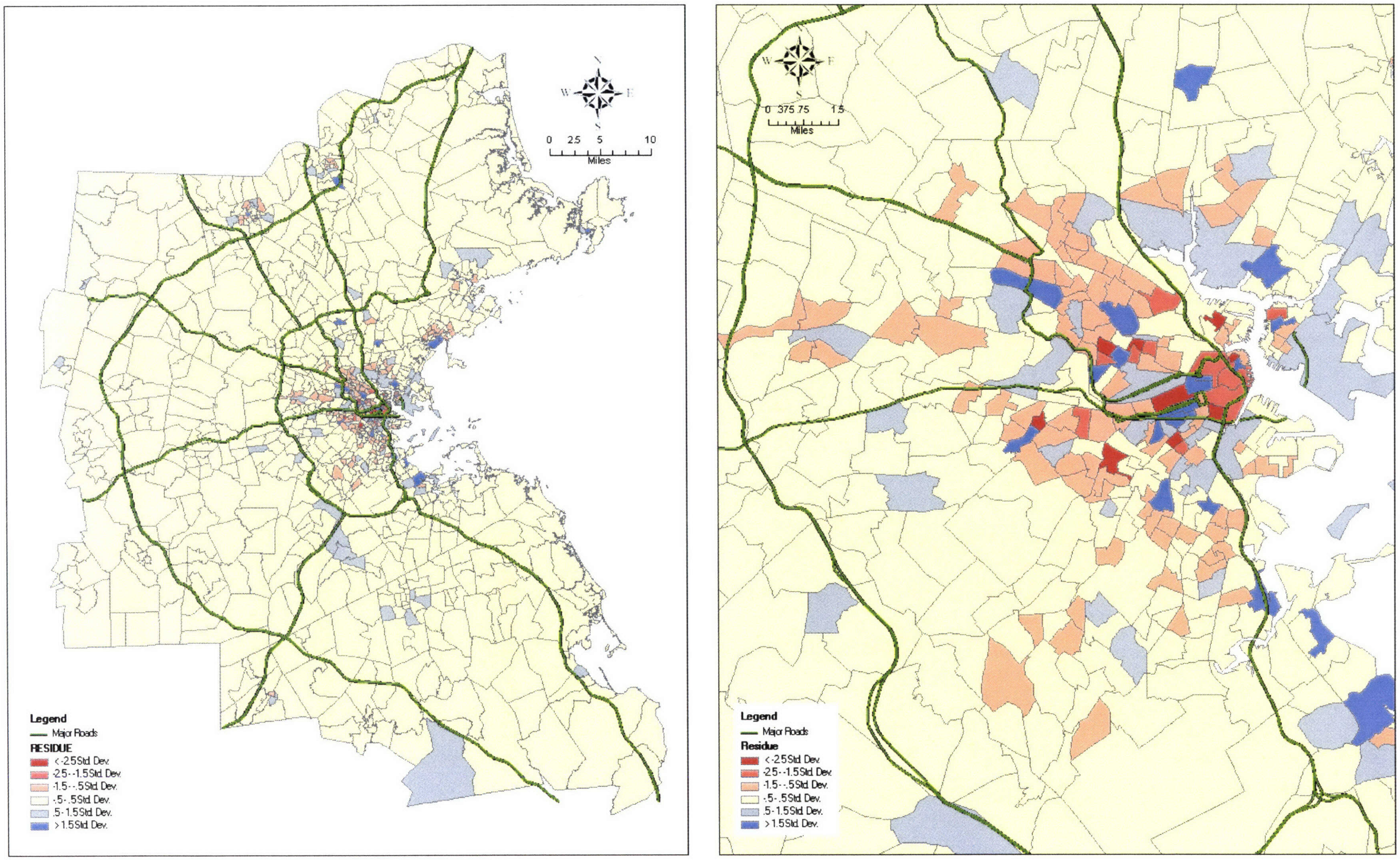
[25] STATA is a statistical package developed by Stata Corp LP.

boundaries to define 'neighbors'. The Moran's I value given by GEODA[26] using rook contiguity weights is 1.85, which is insignificant at the 0.05 level (probability = 0.065). Distance band weights define 'neighbors' by Euclidean distances. In this case, the Moran's I value 1.16 is also insignificant (probability = 0.247). The results of Moran's I tests suggest that the spatial spillover effects have been well captured by the neighboring area variables in the model, and there is no spatial autocorrelation in the error terms.

Figure 5-1 plots residues of the model estimation. The map uses standard deviation classification. It shows that there is no obvious autocorrelation in the spatial distribution of the residues. This result is consistent with the Moran's I tests.

[26] GEODA is a software tool for geodata analysis developed by Luc Anselin at the University of Illinois, Urbana-Champaign.

Figure 5-1: Residue Plot of Model Estimation



5.2.2 OLS Estimation with Robust Standard Errors

To deal with the heteroscedasticity problem, White-robust standard error method is applied to calculate the standard errors of the parameters. Hal White proposed this method for obtaining consistent estimates of variance and covariance of OLS estimates, which provide valid statistical tests for large samples.

Table 5-5 shows the regression results with robust standard errors. The robust standard errors are generally larger than the OLS estimates of standard errors, which lead to the smaller magnitudes of t-statistics than those presented in Table 5-4. Nonetheless, the same coefficients remain statistically significant at the 0.05 level.

Table 5-5: Estimation Results of OLS with Robust Standard Error (Method 5)

Variables	Coefficients	Robust Std. Error	Beta	t	Sig.
(Constant)	-1,889.410	1,621.733	.	-1.170	0.244
DEN_DEMAND	1.323	0.254	0.651	5.220	0.000
DEN_G0_2MI	0.484	0.205	0.090	2.360	0.018
DEN_G2_4MI	0.598	0.495	0.039	1.210	0.228
DEN_G4_6MI	2.036	0.537	0.136	3.790	0.000
DEN_WORKER	0.633	0.219	0.309	2.890	0.004
INNERCITY	-12,889.490	5,094.925	-0.167	-2.530	0.012
Number of observations	894				
Adjusted R^2	0.41				
F	10.73				
Sig.	0.000				

Dependent Variable: DEN_SUPPLY

The adjusted R-squared value 0.41 is acceptable for such a cross-sectional regression analysis with variable estimates that have large standard deviation. The F-statistic value 10.73 is significant at the 0.01 significance level, which can reject the null hypothesis that none of the explanatory variables helps explain the variation of the independent variable.

The estimated coefficient of local food store demand (DEN_DEMAND) is both positive and significant at the 0.01 significance level, indicating that higher local demand can attract more food stores to the target tract.

The coefficient of adjusted retail demand of tracts within 2-mile buffer ring (target tract itself excluded) DEN_G0_2MI is positive and significant at the 0.05 significance level. The magnitude of its effect is less than that of retail demand of target tract itself (DEN_DEMAND). This result is not surprising, given that people living in a close tract may well come over to the target tract and buy some food, but their propensity to shop at their own tracts is higher, all else equal. Overall, tracts with a higher adjusted retail demand of tracts within 2-mile buffer ring are more likely to have higher retail supply level, but the magnitude and significance level of its effect is much smaller than those of target tract retail demand effect.

The coefficient of adjusted retail demand of tracts within 2-4-mile buffer ring (DEN_G2_4MI) is also positive, but is insignificant at the 0.05 significance level. A possible reason is that the distance of 2-4 miles is too far for people to shop by walking but not far enough for people to arrive at a large regional mall. In real life, people would prefer to either purchase food from a nearby relatively small store to save time and transportation cost, or drive a long distance to bigger stores with better services and lower prices. The stores falling between are less likely to get patronage.

The coefficient of adjusted retail demand of tracts within 4-6-mile buffer ring (DEN_G4_6MI) is both positive and significant at the 0.01 level. Compared with the effect of target tract retail demand, the effect of adjusted demand of tracts with 4-6-mile buffer ring has a higher magnitude, but a lower t-statistic value. This result suggests that retail demand from tracts whose residents can shop at the target tract by car also play an important role in the target tract's retail supply level, which is consistent with the discussion for the coefficient of adjusted retail demand of tracts within 2-4-mile buffer ring.

The coefficient of adjusted employee density (DEN_WORKER) is positive and significant at the 0.01 significance level, reflecting that retail supply is significantly increased for tracts with larger number of employees. This result is consistent with the model expectation: employees may purchase food around their working place and bring back home to save time and transportation cost, hence more employees mean more demand and can generate higher retail supply level.

The coefficient of the inner-city location dummy is negative and significant at the 0.05 level, which justifies the research hypothesis. An inner-city location can reduce the retail supply level of the tract, and the magnitude of this effect is estimated to be around 12.89 million dollars per square mile annually, i.e., a non inner-city tract will have a annual food store sales 12.89 million dollars per square mile higher than an inner-city tract, all else factors equal. It should be noted that this figure cannot be directly compared with the retail demand of the target tract in order to get a underserved portion, because it is the underserved part with respect to the full potential of retail supply – the supply potential when the neighboring area demand and non-resident employee demand are taken into account. This figure may be more meaningful for retailers when they think about shifting from suburbs to inner cities, because what matters to make such decision is the difference between the two regions, not the retail gap of inner city itself.

Standardized coefficients describe the relative importance of the explanatory variables. For example, the standardized coefficient of local food store demand (DEN_DEMAND) is 0.651, which means that an increase of 1 standard deviation in the local food store demand will lead to a increase of 0.651 standard deviation in the local food store supply. According to their standardized coefficients, local food store demand is the most important determinant of local food store supply level; the second most important factor is adjusted density of local employees; the inner-city location dummy ranks the third with a value -0.167, which means that the inner-city location factor is more important than the neighboring area demand indicators in explaining the retail supply variation across the MSA.

5.3 Sensitivity Analysis

This section tests the sensitivity of the proposed model by using variables calculated with alternative methods, and by excluding large food stores from the analysis.

5.3.1 Sensitivity Analysis with Variables Calculated with Alternative Methods

The estimation results discussed in the preceding section are based on variables calculated with Method 5 (household income adjusted per capita CES income adjustment) as is presented in Section 4.1.5. To test the sensitivity of the analysis, the model are re-estimated with variables calculated based on Method 4 (per capita CES income adjustment) as is presented in Section 4.1.4. Table 5-6 reports the new descriptive statistics of variables.

Table 5-7 compares the estimated results of the model using variables calculated with alternative methods. All the important conclusions in the preceding section still hold with the estimates based on variables calculated with Method 4.The estimated coefficients are quite stable -- the coefficients based on Method 4 have similar values and similar significance levels as the coefficients based on Method 5. The magnitude of the inner-city effect is almost the same as the original estimation, and the significance level increases slightly.

Table 5-6: Descriptive Statistics for Variables (Method 4)

Variable	Description	Mean	Std. Dev.	Max	Min
Dependant Variable					
DEN_SUPPLY	Retail Supply in the target tract (k$/sqmi)	13,014	35,967	487,161	0
Independent Variable: Retail Demand of Local Residents					
DEN_DEMAND	Retail demand from residents in the target tract (k$/sqmi)	13,084	16,182	144,140	124
Independent Variable: Retail Demand of Residents of Neighboring Areas					
DEN_G0_2MI	Adjusted retail demand of tracts whose centroids fall in 2 mile buffer of the target tract's centroid (k$/sqmi) *	-1,285	7,069	11,037	-36,520
DEN_G2_4MI	Adjusted retail demand from tracts whose centroids fall in 2-4 mile buffer of centroid of the target tract (k$/sqmi)	-433	2,503	6,194	-20,397
DEN_G4_6MI	Adjusted retail demand from tracts whose centroids fall in 4-6 mile buffer of centroid of the target tract (k$/sqmi)	-59	2,441	6,976	-11,224
Independent Variable: Retail Demand of Local Employees					
D_WORKER	Adjusted density of local employees (person/sqmi)	312	14,641	209,683	-55,732
Independent Variable: Location Characteristics					
INNERCITY	Inner city location dummy	0.19	0.39	1	0
Number of Observations: 891					

* For the adjusted retail demand from tracts whose centroids fall in 2 mile buffer of the target tract's centroid, the target tract itself is excluded.

Note: Variables are calcualted with method 4 (per capita CES income adjustment).

Table 5-7: Estimation Results of Models Using Variables Calculated with Alternative Methods

	Variables Calculated with Method 4 (Per Capita CES Income Adjustment)			Variables Calculated with Method 5 (Household Size Adjusted Per Capita CES Income Adjustment)		
Variables	**Coefficients**	**t-stat.**	**Sig.**	**Coefficients**	**t-stat.**	**Sig.**
(Constant)	-2,840.133	-1.39	0.166	-1,889.410	-1.17	0.244
DEN_DEMAND	1.453	5.17	0.000	1.323	5.22	0.000
DEN_G0_2MI	0.447	2.24	0.025	0.484	2.36	0.018
DEN_G2_4MI	0.668	1.22	0.224	0.598	1.21	0.228
DEN_G4_6MI	2.139	3.83	0.000	2.036	3.79	0.000
DEN_WORKER	0.775	2.94	0.003	0.633	2.89	0.004
INNERCITY	-13,018.980	-2.29	0.022	-12,889.490	-2.53	0.012
Number of observations	891			894		
Adjusted R^2	0.40			0.41		
F	10.35			10.73		
Sig.	0.000			0.000		

Dependent Variable: DEN_SUPPLY
Note: Models are estimated using OLS with robust standard error method.

5.3.2 Sensitivity Analysis with Large Food Stores Excluded

In this section, the model is re-estimated with variables calculated using the same method (household size adjusted per capita CES income adjustment) as in the original estimation, but rather than include all food stores in the Boston MSA, 95 large food stores with 150 employees or more are excluded. The spatial locations of these large stores are plotted in Figure 4.4. Those stores represent about 1/3 of the food store supply in the Boston MSA. To make the food store demand and supply balanced at the MSA level, food store demand of each tract are reduced by the same ratio by which the total retail supply decrease, assuming that each tract spends a fix proportion of its food expenditure in large stores. This assumption is somewhat unreasonable in reality, because it can be expected that inner-city tracts would spend a lower proportion of their food expenditure in large stores compared with suburban tracts, due to the low automobile ownership rate in inner cities. Hence, the assumption is conservative in that it will tend to underestimate any inner-city effect.

Table 5-8 reports the new descriptive statistics of variables. Table 5-9 makes comparison between the new estimation results and the results with all food stores included.

Table 5-8: Descriptive Statistics for Variables (Large Stores Excluded)

Variable	Description	Mean	Std. Dev.	Max	Min
Dependant Variable					
DEN_SUPPLY	Retail Supply in the target tract (k$/sqmi)	9,910	28,202	487,161	0
Independent Variable: Retail Demand of Local Residents					
DEN_DEMAND	Retail demand from residents in the target tract (k$/sqmi)	11,797	17,233	203,068	88
Independent Variable: Retail Demand of Residents of Neighboring Areas					
DEN_G0_2MI	Adjusted retail demand of tracts whose centroids fall in 2 mile buffer of the target tract's centroid (k$/sqmi) *	713	3,758	8,361	-17,048
DEN_G2_4MI	Adjusted retail demand from tracts whose centroids fall in 2-4 mile buffer of centroid of the target tract (k$/sqmi)	937	1,606	5,877	-20,683
DEN_G4_6MI	Adjusted retail demand from tracts whose centroids fall in 4-6 mile buffer of centroid of the target tract (k$/sqmi)	1,030	1,411	6,489	-3,175
Independent Variable: Retail Demand of Local Employees					
D_WORKER	Adjusted density of local employees (person/sqmi)	312	14,641	209,683	-55,732
Independent Variable: Location Characteristics					
INNERCITY	Inner city location dummy	0.19	0.39	1	0
Number of Observations: 891					

* For the adjusted retail demand from tracts whose centroids fall in 2 mile buffer of the target tract's centroid, the target tract itself is excluded.

Table 5-9: Estimation Results of Model Using All Stores and Model Excluding Large Stores

	Large Food Store Exluded			All Food Stores Included		
Variables	Coefficients	t-stat.	Sig.	Coefficients	t-stat.	Sig.
(Constant)	-2,508.807	-1.62	0.106	-1,889.410	-1.17	0.244
DEN_DEMAND	1.748	4.35	0.000	1.323	5.22	0.000
DEN_G0_2MI	0.330	1.29	0.197	0.484	2.36	0.018
DEN_G2_4MI	1.241	1.42	0.157	0.598	1.21	0.228
DEN_G4_6MI	2.370	2.43	0.015	2.036	3.79	0.000
DEN_WORKER	0.363	4.38	0.000	0.633	2.89	0.004
INNERCITY	-8,620.053	-2.13	0.034	-12,889.490	-2.53	0.012
Number of observations	894			894		
Adjusted R^2	0.42			0.41		
F	22.24			10.73		
Sig.	0.000			0.000		

Dependent Variable: DEN_SUPPLY
Note: Models are estimated using OLS with robust standard error method

The estimated coefficient of retail demand of target tract (DEN_DEMAND) is still positive and significant at the 0.01 level. But the magnitude of the coefficient increases from 1.323 to 1.748, suggesting that as large food stores are excluded from the analysis, the relative importance of local food stores goes up.

The coefficient of the adjusted retail demand for the 0-2-mile buffer ring (DEN_G0_2MI) becomes smaller and insignificant. This result is not surprising - as large food stores are excluded, remaining expenditure are more likely to be close at hand. The estimated coefficient of the adjusted retail demand for the 2-4-mile buffer ring and 4-6-mile buffer ring is somewhat counter-intuitive. The normal expectation is that as large stores are excluded, consumers are less likely to drive a long distance to shop, thus the effect of the 2-4-mile and 4-6-mile buffer ring demands should diminish. However, the magnitudes of both coefficients increase. The coefficient of the 2-4-mile buffer ring demand is still insignificant at the 0.05 level. The coefficient of the 4-6-mile buffer ring demand is significant at the 0.05 level with a lower t-statistic value.

The estimated coefficient of adjusted employee density (DEN_WORKER) is still positive and significant, but the magnitude decreases. One possible reason is that many excluded large food stores are located near job centers. The exclusion of these stores makes the effect of adjusted employee density shrink.

The argument that an inner-city location will reduce the supply level of the target tract still holds, as is indicated by the negative and significant (at the 0.05 level) coefficient of INNERCITY location dummy. The magnitude of the estimated coefficient decreased from 12.89 to 8.62 million dollars per square mile. This decrease makes sense. There are more large stores located in suburban area than in the inner city. As they are excluded, the gap between the inner city and the suburban is expected to shrink.

Chapter 6: Conclusion

Tapping the unmet retail demand in the inner-city neighborhoods has long been considered as an important strategy to accelerate the economic revitalization of inner cities. This thesis proposes an analytical framework that can reveal the spatial patterns of retail markets and test whether and to what extent inner-city neighborhoods are 'underserved'. With the help of Geographic Information Systems (GIS) and Relational Database Management System (RDBMS) tools, this study designs and calculates neighborhood indicators of the demand, supply and gaps in retail markets with census tract level socio-economic data and parcel level business data. Based on these indicators, econometric models are developed to quantitatively estimate the pure impact of an inner-city location on the local retail supply level.

This concluding chapter first summarizes findings in the preceding chapters, then points out the limitations and challenges of the current study, and finally suggests future research directions.

6.1 Summary of Research Findings

The analysis of this study is broadly categorized into two parts. The first part aims to illustrate retail markets with a new neighborhood indicator system (Chapter 4); the second part examines the impact of an inner-city location on local retail supply level (Chapter 5).

In Chapter 4, a new neighborhood indicator system on retail markets is proposed. These indicators cover both the demand and supply sides of the retail markets, and incorporate market situations not only within an area of residence, but also in zones beyond that area to catch spillover effects. The calculation formulas and methods under different assumptions about consumers' purchasing patterns are presented in detail.

Chapter 4 also reveals the spatial patterns of food store retail markets in the Boston MSA with thematic maps of retail market indicators. These maps unveil a striking difference in the spatial distributions of food store demand and supply: substantial purchasing power is concentrated in the core of the MSA, while the distribution of food store supply is far more dispersed – many census tracts with high food store supply level are located along transportation corridors. This pattern suggests the potential existence of unmet retail demand in the urban core. However in the map of retail gaps, the urban core shows a dichotomous structure. The CBD area has an enormous negative retail gap (supply exceeds demand), while the poor residential neighborhoods have a very high positive retail gap (demand exceeds supply). One possible reason of the dichotomy is that retail demand from employees and visitors that cluster in the CBD area, is a significant determinant of retail supply level, and have considerable attractions for retailers. The spatial distribution pattern of food store markets changes as the size of aggregation area increases from the target tract itself, to 2-mile buffer zone, 4-mile buffer zone and 6-mile buffer zone of the target tract. As the size increases, the retail gap surfaces in the maps become smoother and smoother, and the magnitude of the gaps decreases. It implies that the food store demand and supply are more likely to be balanced when the size of the study area increases.

Chapter 5 incorporates the retail market neighborhood indicators developed in Chapter 4 into regression models to estimate the effect of an inner-city location on retail supply level. The dependent variable is the retail supply level. Explanatory variables include target tract retail demand, adjusted retail demand of tracts within 2-mile buffer ring, 2-4-mile buffer ring, and 4-6-mile buffer ring, adjusted employee density, and the inner-city location dummy.

The empirical results suggest that an inner-city location can significantly reduce local retail supply level. On average, inner-city tracts have an annual food store retail sales 12.89 million dollars per square mile lower than non inner-city tracts in the Boston MSA, after controlling for other factors that may influence retail supply level, such as demand of local residents, neighboring residents, and non-resident employees.

Some calculation and comparisons can help understand the meaning of the estimation results. By the model construction, the 12.89 million per square mile is the unrealized supply that can be attributed to inner-city location when the full food store retail potential of the target tract are considered, including target tract demand, neighboring demand and non-resident employee demand of the target tract. The potential supply density of an inner-city tract is the predicted value of the model using all the demand variables of the tract but assuming this tract is a non inner-city tract. The potential supply of all inner-city tracts can then be computed by multiplying the potential supply density of each inner-city tract by its land area and then summing up over inner-city tracts. The figure is 1097.67 million dollars. This estimated potential supply can be used as a benchmark to calculate the 'underserved' portion of the potential supply. The total 'true' supply of inner-city tracts is 779.82 million dollars, which is about 0.71 times of the potential supply. Hence the 'underserved' portion is about 29 percent of the potential supply.

6.2 Limitations and Challenges

There are several major limitations and challenges facing the current study.

The neighborhood indicators about retail markets are sensitive to assumptions about consumer's purchasing activities. As is shown in Section 4.4, this sensitivity could lead to the large variations of the indicators calculated based on different assumptions. To deal with this challenge, the proposed neighborhood indicator system adopts a flexible structure, which makes it easy to substitute alternative user-defined measures of supply and demand. Various tools can help integrate such user-specific local knowledge into the proposed system, such as the model builder extension in ARCMAP, web GIS service, and the 'intelligent middleware' approach proposed by Ferreira (2004).

As is discussed in Section 3.3, data used in this study are mainly national level data, which have some inherent drawbacks, such as the underestimation of inner city residents' purchasing power, and the lack of local area specific information. The integration of city, community, or parcel level data will surely improve the quality of the analysis. The proposed framework provides a general platform for practitioners, and is ready to be adjusted according to their local knowledge.

There are also several limitations in the econometric models. The first is that this model only has a very simple distance control. All tracts falling in a distance range are collapsed into one category and assumed to have the same distance impact. The second limitation is that constant returns to scale are assumed in this model. However, according to the Lakshmanan and Hansen retail gravity model, the possibility that consumers shop at a destination is decided not only by the distance, but also by the attractiveness of the destination (normally measured by size). This economy of scale has not yet been captured in the analysis.

The location decision of retail businesses is a complex process. Besides the retail demand factor, other factors such as crime rate, transit access, and trucking access, also plan important roles. This study mainly focuses on the demand supply analysis. A comprehensive study that take into account more influential factors may shed more light on store's location decision, and provide more support for various stakeholders.

Nonetheless, this study expands the previous studies by incorporating neighboring demand and non-resident employee demand into econometric models and systemically controlling for the distance factors. It provides more accurate pictures for inner-city markets and is useful as a starting point for future discussion of business opportunities.

6.3 Future Research Directions

There are three major potential extensions of the current study: (1) to refine the methodology to improve the empirical results; (2) to study the application of research results in practical settings, like business location decision-making process and community economic development planning; (3) to explore the associated data sharing and integration issues from the data infrastructure point of view.

1. Refinement of current methodology

As is discussed in the preceding section, the methodology used in the current study can be further improved. (1) The analytic framework can be further refined to explore the effects of factors like land rents, crime, and transportation accessibility on retail supply level, and see whether these factors can explain some of the inner-city location effects. (2) Due to the limitation of data, this study provides a snapshot for the year 2000. One study that would be important to examine in the future is the comparisons between results of various time points. This comparison can reveal the evolution patterns of the food store markets in the Boston MSA. (3) Since the current analytical framework can be readily applied to further research, the empirical analysis can be extended to include more study areas as well as more retail categories. There are considerable variations among the MSAs, and the nature of the retail markets varies from category to category. The comparison of various MSAs and retail categories can improve the understanding of inner-city retail markets. (4) There are substantial studies on the spatial location of economic facilities in the operations research literature, which employ complicated algorithms and models to predict and analyze firm's location. The current models in this thesis can be improved by integrating studies in operations research field to generate more refined models.

2. Interviews with retailers and community development corporations (CDC)

This thesis is mainly a quantitative study for the retail markets. Issues related to policy design and implementation like how the retailers can make use of these indicators to make location decisions, how the CDC can take advantage of these indicators to accelerate community development are still not explored. Therefore a qualitative research could be a good complementary study. Interviews with retailers and CDC officials can shed light on these issues.

3. Data integration and system sustainability

As is discussed in the previous sections, the integration of community or parcel level data can greatly improve the understanding of neighborhood markets. However, the collection of such data could be expensive and time-consuming, and the integration of more disaggregate datasets could be very labor-intensive and not easily replicated or sustained. (Ferreira, 2004). How to find a more effective, scalable, and sustainable approach for analyzing and sharing data is an important direction for the future research. Ferreira (2004) proposes an 'intelligent middleware' approach, which takes advantage of modern information and communication technologies to package the data processing, analysis and interpretation steps into reusable and tunable data intermediaries. It can help streamline data-sharing efforts while greatly enhance their likelihood of empowering grassroots planning. The inner-city retail markets example studied in this thesis could be a meaningful example to explore the application of the "middleware approach" in actual planning settings.

To summarize, further research should generate more in-depth insights into the nature of inner-city retail markets and should provide useful data for making effective programs that can help inner-city neighborhoods to accelerate economic development.

References

Brown, S. 1992. "The Wheel of Retail Gravitation?" *Environment and Planning A*, 24, 1409-1429.

Chieffo, Sara, et al. 2004. "Pittsburgh Purchasing Power Profiles," System Synthesis Project 2004, School of Public Policy and Management, Carnegie Mellon University.

Clifton, Kelly J. 2004. "Mobility Strategies and Food Shopping for Low-Income Families," *Journal of Planning Education and Research*, 23, 402-413.

Cotterill, Ronald W. and Andrew W. Franklin. 1995. "The Urban Grocery Store Gap," Food Marketing Policy Center Issue Paper No.8, Storrs: University of Connecticut.

Donohue, R.M. 1997. "Abandonment and Revitalization of Central City Retailing: The Case of Grocery Stores," Unpublished Doctoral Dissertation, University of Michigan, Ann Arbor.

Eppli, Mark J. and James D. Schilling. 1996. "How Critical Is a Good Location to a Regional Shopping Center?" *The Journal of Real Estate Research*, 12,3, 459-467.

Ferreira, Joseph, Jr. 2004. "Intelligent Middleware for Understanding Neighborhood Markets," Pilot Project Supported by the Brookings Institution through Urban Market Initiative Program (http://www.brookings.edu/metro/umi.htm)

Lakshmanan, T. R. and Walter G. Hansen. 1965. "A Retail Market Potential Model," *Journal of the American Institute of Planner*, 13, 134-143.

LISC/Chicago website: http://www.lisc-chicago.org

Lee, Ming-Long and R. Kelly Pace. 2001. "Spatial Distribution of Retail Sales," *The Journal of Real Estate Finance and Economics*, 31, 1, 53-69.

Hoyt, Lorlene. (2005). "The Business Improvement District: An Internationally Diffused Approach to Revitalization." Washington, D.C.: International Downtown Association, 1-65.

Initiative for a Competitive Inner-city. 1998. "The Business Case for Pursuing Retail Opportunities in the Inner-city," Boston: Author.

MacDonald, J.M., and P.E. Nelson. 1991. "Do the poor still pay more? Food price variations in large metropolitan areas," Journal of Urban Economics, 30, 344-359.

Pawasarat, John and Lois M. Quinn. 2001. "Exposing Urban Legends: The Real Purchasing Power of Central City Neighborhoods," Discussion Paper Prepared for The Brookings Institution Center on Urban and Metropolitan Policy.

Porojan A. 2000. "Trade Flows and Spatial Effects: the Gravity Model Revisited," *Open Economic Review*, 12, 265-280.

Porter, Michael E. 1995. "The Competitive Advantage of Inner-city," *Harvard Business Review*, May-June 1995, 55-73.

Porter, Michael E. 1997. "New Strategies for Inner-City Economic Development," *Economic Development Quarterly*, 11, 1, February 1997, 11-27.

Sabety, J. Pari and Virginia L. Carlson. 2004. "Using Information to Drive Change: New Ways to Move Urban Markets", Urban Markets Initiative, Metropolitan Policy Program, The Brookings Institution.

Seidman, Karl F. 2004a. "New Approach to Urban Market Analysis," Unpublished Paper.

Seidman, Karl F. 2004b. "Revitalize Commerce for American Cities, a Practitioner' Guide to Urban Main Street Programs," Fannie Mae Foundation.

Social Compact. 2006. "Santa Ana Neighborhood Market DrillDown Profile".

Social Compact Website: http://www.socialcompact.org/market.htm

University of Wisconsin-Milwaukee Employment and Training Institute Website: http://www.uwm.edu/Dept/ETI

U.S. Department of Housing and Urban Development. 1999. "New Market: The Untapped Retail Buying Power in American's Inner Cities," Washington DC: Author.

Weissbourd, Robert and Christopher Berry. 1999. "The Market Potential of Inner-City Neighborhoods: Filling the Information Gap," The Brookings Institution Center on Urban and Metropolitan Policy, March 1999.

Whelan, Amanda, Neil Wrigley, et al. 2002. "Life in a 'food desert'," *Urban Studies*, 39, 11, 2083-2100.

Appendix: PL/SQL Scripts Used to Calculate Neighboring Area Indicators

Purpose:

These PL/SQL scripts are used to calculate neighboring area indicators based on target tract indicators and prepare variables used in the econometric analysis.

```
------------------------------------------------------------------------------------------------------------------------

DROP TABLE BOS_TR00_54CA;

CREATE TABLE BOS_TR00_54CA AS
SELECT
CT_ID,TOTAL_POP,DRY_SQMI,TOT_HSHLD,INC_CAPITA,DEMAND_CA,DEN_DE_CA,BUSI
_NUM,EMPLOYMENT,SUPPLY SUPPLY_CA,DEN_SUPPLY
DEN_SUP_CA,GAP_CA,DEN_GAP_CA
FROM MSA_TR00_54CA;

------------------------------------------------------------------------------------

--Define a catchment area as composed of
--tracts within 2 mile of the target tract by air distance

alter table bos_tr00_54ca add
(
s2mile_ca number(38,8),
den_s2mile_ca number(38,8),
d2mile_ca number(38,8),
den_d2mile_ca number(38,8)
);

declare
  cursor geo_cursor is
   select * from ctpp.boston_centroid_00;
  geo_val geo_cursor%ROWTYPE;
  tempval1 number;
  tempval2 number;
  tempval3 number;
  tempval4 number;


Begin
  open geo_cursor;
  loop
```

```
  fetch geo_cursor into geo_val;
  exit when geo_cursor%NOTFOUND;

    select sum(b.supply_ca) into tempval1
    from ctpp.boston_centroid_00 t, bos_tr00_54ca b
    where t.tract=b.ct_id and
    sdo_within_distance (t.shape, geo_val.shape, 'distance=3219')='TRUE';

    select sum(b.supply_ca)/sum(b.dry_sqmi) into tempval2
    from ctpp.boston_centroid_00 t, bos_tr00_54ca b
    where t.tract=b.ct_id and
    sdo_within_distance (t.shape, geo_val.shape, 'distance=3219')='TRUE';

    select sum(b.demand_ca) into tempval3
    from ctpp.boston_centroid_00 t, bos_tr00_54ca b
    where t.tract=b.ct_id and
    sdo_within_distance (t.shape, geo_val.shape, 'distance=3219')='TRUE';

    select sum(b.demand_ca)/sum(b.dry_sqmi) into tempval4
    from ctpp.boston_centroid_00 t, bos_tr00_54ca b
    where t.tract=b.ct_id and
    sdo_within_distance (t.shape, geo_val.shape, 'distance=3219')='TRUE';

    update bos_tr00_54ca set s2mile_ca=tempval1 where ct_id=geo_val.tract;
    update bos_tr00_54ca set den_s2mile_ca=tempval2 where ct_id=geo_val.tract;
    update bos_tr00_54ca set d2mile_ca=tempval3 where ct_id=geo_val.tract;
    update bos_tr00_54ca set den_d2mile_ca=tempval4 where ct_id=geo_val.tract;

 end loop;
 close geo_cursor;
end;
.
run;
commit;

-------------------------------------------------------------------------------

--Define a catchment area as composed of
--tracts within 4 mile of the target tract by air distance

alter table bos_tr00_54ca add
(
s4mile_ca number(38,8),
den_s4mile_ca number(38,8),
d4mile_ca number(38,8),
den_d4mile_ca number(38,8)
);

declare
 cursor geo_cursor is
```

```
 select * from ctpp.boston_centroid_00;
geo_val geo_cursor%ROWTYPE;
tempval1 number;
tempval2 number;
tempval3 number;
tempval4 number;

Begin
 open geo_cursor;
 loop
  fetch geo_cursor into geo_val;
  exit when geo_cursor%NOTFOUND;

    select sum(b.supply_ca) into tempval1
    from ctpp.boston_centroid_00 t, bos_tr00_54ca b
    where t.tract=b.ct_id and
    sdo_within_distance (t.shape, geo_val.shape, 'distance=6437')='TRUE';

    select sum(b.supply_ca)/sum(b.dry_sqmi) into tempval2
    from ctpp.boston_centroid_00 t, bos_tr00_54ca b
    where t.tract=b.ct_id and
    sdo_within_distance (t.shape, geo_val.shape, 'distance=6437')='TRUE';

    select sum(b.demand_ca) into tempval3
    from ctpp.boston_centroid_00 t, bos_tr00_54ca b
    where t.tract=b.ct_id and
    sdo_within_distance (t.shape, geo_val.shape, 'distance=6437')='TRUE';

    select sum(b.demand_ca)/sum(b.dry_sqmi) into tempval4
    from ctpp.boston_centroid_00 t, bos_tr00_54ca b
    where t.tract=b.ct_id and
    sdo_within_distance (t.shape, geo_val.shape, 'distance=6437')='TRUE';

    update bos_tr00_54ca set s4mile_ca=tempval1 where ct_id=geo_val.tract;
    update bos_tr00_54ca set den_s4mile_ca=tempval2 where ct_id=geo_val.tract;
    update bos_tr00_54ca set d4mile_ca=tempval3 where ct_id=geo_val.tract;
    update bos_tr00_54ca set den_d4mile_ca=tempval4 where ct_id=geo_val.tract;

 end loop;
 close geo_cursor;
end;
.
run;
commit;

--------------------------------------------------------------------------------------

--Define a catchment area as composed of
--tracts within 6 mile of the target tract by air distance
```

```
alter table bos_tr00_54ca add
(
s6mile_ca number(38,8),
den_s6mile_ca number(38,8),
d6mile_ca number(38,8),
den_d6mile_ca number(38,8)
);

declare
  cursor geo_cursor is
   select * from ctpp.boston_centroid_00;
  geo_val geo_cursor%ROWTYPE;
  tempval1 number;
  tempval2 number;
  tempval3 number;
  tempval4 number;

Begin
  open geo_cursor;
  loop
    fetch geo_cursor into geo_val;
    exit when geo_cursor%NOTFOUND;

      select sum(b.supply_ca) into tempval1
      from ctpp.boston_centroid_00 t, bos_tr00_54ca b
      where t.tract=b.ct_id and
      sdo_within_distance (t.shape, geo_val.shape, 'distance=9656')='TRUE';

      select sum(b.supply_ca)/sum(b.dry_sqmi) into tempval2
      from ctpp.boston_centroid_00 t, bos_tr00_54ca b
      where t.tract=b.ct_id and
      sdo_within_distance (t.shape, geo_val.shape, 'distance=9656')='TRUE';

      select sum(b.demand_ca) into tempval3
      from ctpp.boston_centroid_00 t, bos_tr00_54ca b
      where t.tract=b.ct_id and
      sdo_within_distance (t.shape, geo_val.shape, 'distance=9656')='TRUE';

      select sum(b.demand_ca)/sum(b.dry_sqmi) into tempval4
      from ctpp.boston_centroid_00 t, bos_tr00_54ca b
      where t.tract=b.ct_id and
      sdo_within_distance (t.shape, geo_val.shape, 'distance=9656')='TRUE';

      update bos_tr00_54ca set s6mile_ca=tempval1 where ct_id=geo_val.tract;
      update bos_tr00_54ca set den_s6mile_ca=tempval2 where ct_id=geo_val.tract;
      update bos_tr00_54ca set d6mile_ca=tempval3 where ct_id=geo_val.tract;
      update bos_tr00_54ca set den_d6mile_ca=tempval4 where ct_id=geo_val.tract;

   end loop;
   close geo_cursor;
```

```
end;
.
run;
commit;

--------------------------------------------------------------------------------------

--Calculate the supply demand balance indicators with floating catchment area

alter table bos_tr00_54ca add
(
g2mile number(38,8),
den_g2mile number(38,8),
g4mile number(38,8),
den_g4mile number(38,8),
g6mile number(38,8),
den_g6mile number(38,8)
);


update bos_tr00_54ca set g2mile=d2mile_ca-s2mile_ca;
update bos_tr00_54ca set den_g2mile=den_d2mile_ca-den_s2mile_ca;
update bos_tr00_54ca set g4mile=d4mile_ca-s4mile_ca;
update bos_tr00_54ca set den_g4mile=den_d4mile_ca-den_s4mile_ca;
update bos_tr00_54ca set g6mile=d6mile_ca-s6mile_ca;
update bos_tr00_54ca set den_g6mile=den_d6mile_ca-den_s6mile_ca;

commit;

--------------------------------------------------------------------------------

--Calculate areas of circles with various radii

alter table bos_tr00_54ca add
(
area_2mile number(38,8),
area_4mile number(38,8),
area_6mile number(38,8)
);


declare
  cursor geo_cursor is
   select * from ctpp.boston_centroid_00;
  geo_val geo_cursor%ROWTYPE;

  tempval2 number;
  tempval4 number;
  tempval6 number;
```

```
Begin
 open geo_cursor;
 loop
  fetch geo_cursor into geo_val;
  exit when geo_cursor%NOTFOUND;

    select sum(b.dry_sqmi) into tempval2
    from ctpp.boston_centroid_00 t, bos_tr00_54ca b
    where t.tract=b.ct_id and
    sdo_within_distance (t.shape, geo_val.shape, 'distance=3219')='TRUE';

    select sum(b.dry_sqmi) into tempval4
    from ctpp.boston_centroid_00 t, bos_tr00_54ca b
    where t.tract=b.ct_id and
    sdo_within_distance (t.shape, geo_val.shape, 'distance=6437')='TRUE';

    select sum(b.dry_sqmi) into tempval6
    from ctpp.boston_centroid_00 t, bos_tr00_54ca b
    where t.tract=b.ct_id and
    sdo_within_distance (t.shape, geo_val.shape, 'distance=9656')='TRUE';

    update bos_tr00_54ca set area_2mile=tempval2 where ct_id=geo_val.tract;
    update bos_tr00_54ca set area_4mile=tempval4 where ct_id=geo_val.tract;
    update bos_tr00_54ca set area_6mile=tempval6 where ct_id=geo_val.tract;

 end loop;
 close geo_cursor;
end;
.
run;
commit;

--------------------------------------------------------------------------------

--Calculate the retail gaps (adjusted demand) within buffer rings

alter table bos_tr00_54ca add
(
d_ex_g2mi number(38,8),
d_g2_4mi number(38,8),
d_g4_6mi number(38,8)
);


update bos_tr00_54ca
set d_ex_g2mi=((d2mile_ca-demand_CA)-(s2mile_ca-supply_ca))/(area_2mile-dry_sqmi)
where area_2mile>dry_sqmi;

update bos_tr00_54ca
```

```
set d_g2_4mi=((d4mile_ca-d2mile_ca)-(s4mile_ca-s2mile_ca))/(area_4mile-area_2mile)
where area_4mile>area_2mile;

update bos_tr00_54ca
set d_g4_6mi=((d6mile_ca-d4mile_ca)-(s6mile_ca-s4mile_ca))/(area_6mile-area_4mile)
where area_6mile>area_4mile;

commit;

------------------------------------------------------------------------------------

--Set the null entry in the ring density gap fields to be 0

update bos_tr00_54ca
set d_ex_g2mi=0
where d_ex_g2mi is null;

update bos_tr00_54ca
set d_g2_4mi=0
where d_g2_4mi is null;

update bos_tr00_54ca
set d_g4_6mi=0
where d_g4_6mi is null;
```