The Effects of Regional and Neighborhood Conditions on Location Choice of New Business Establishments

DISTRIBUTION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By
Jae Teuk Chin
Graduate Program in City and Regional Planning

The Ohio State University
2013

Dissertation Committee:
Jennifer Evans-Cowley, Advisor
Rachel Garshick Kleit
Edward Malecki
Copyright by

Jae Teuk Chin

2013
Abstract

City governments and local municipalities have created and implemented a variety of strategies and incentives to stimulate start-up activity within their own jurisdictions. A policy of enhancing business climate is productive in some regions, while the same policy does not work for others. To understand these variations in outcomes, this research focuses on examining the relationship between the uniqueness of certain regions, spatially bounded characteristics, and how both affect where new establishments locate. Considering both regional and neighborhood determinants of new establishments using a single framework is necessary to contrast macro-scale with micro-scale effects. Moreover, it is useful to assess the effectiveness of business environment polices on start-ups because most strategies concentrate on one spatial level while ignoring the other level.

The combination of a quantitative approach of multilevel modeling and a qualitative approach of in-depth interviewing examines how the conditions of regions and neighborhoods affect where new establishments locate. Additionally, using mixed methods allows for an examination of both place-based and people-based factors. A two-level model employs the census tract as a spatial unit of analysis and analyzes new establishments within 27 medium-sized MSAs in the U.S. That quantitative model allows
this study to find key regional and neighborhood factors driving location decisions of new establishments. In addition, interviews with selected founders help this research explore the meaning of the outcomes of the quantitative model, understand the individual stories behind start-up businesses, and discover any placed-based variables that are not included in the model.

Existing employment and population density are two primary neighborhood level determinants influencing new establishments. Unexpectedly, neighborhood-level demographic factors influence the locations of new establishments. Higher proportions of non-minority populations and a higher median age both are also associated with more new establishments being in a neighborhood. Conversely, higher rates of vacant housing units are associated with fewer new establishments in a neighborhood, on average. The major regional determinant is the average January temperature, which is consistent with other research of regional growth, while the impacts of agglomeration economies are less dominant than previous empirical findings. This indicates that the extent of agglomeration economies, generally increase with scale, is likely affected by medium-sized MSA criteria. The results of in-depth interviews confirm the importance of economic, demographic, and geographic conditions in the model, but reveal the existence of unmeasured factors that influence location choice behaviors including the personal preference of founders regarding the quality of neighborhood.
Dedication

To my family
Acknowledgments

I have had the good fortune and privilege of working with wonderful researchers and mentors at The Ohio State University. I am particularly grateful to my advisor Jennifer Evans-Cowley for her advice, support, and guidance. I appreciate Rachel Garshick Kleit who has helped me learn how to tell a clear story with findings. The seminar with Ed Malecki at Derby Hall paved the way for developing ideas for this dissertation. I also thank Burkhard von Rabenau who taught me how to think logically and develop insights.

I would like to thank PhD colleagues of the planning program, especially my friend Chunhui with his encouragement. I am also thankful to my co-workers of Delaware County and Bennett and Williams. Finally, these acknowledgements would not be complete without an expression of deep gratitude to my parents, my parents in law, my brother Jae Bon, and my loving wife Sung Eun. Thanks for your endless patience, support, and love.
Vita

1999........................................B.S. Urban Engineering, Seoul National University
2001........................................M.S. Urban Design, Seoul National University
2002 to 2003 .........................Ensign, Construction Engineer Corps, ROK Navy
2003 to 2005 .........................Lieutenant Junior Grade, Facilities Engineering Bureau, ROK Navy Headquarters
2006 to 2007 .........................Intern, Delaware County Regional Planning Commission
2008 to 2009 .........................Part-time planner, Bennett &Williams Environmental Consultants, Inc.
2009........................................MCRP, The Ohio State University
2010 to present .......................Graduate Instructor, City and Regional Planning, The Ohio State University

Fields of Study

Major Field:  City and Regional Planning
# TABLE OF CONTENTS

Table of Contents ........................................................................................................... vii

List of Tables .................................................................................................................. xii

List of Figures .................................................................................................................. xiv

Chapter 1 Introduction .................................................................................................. 1

1.1 Aims of the study .................................................................................................... 2

1.2 Organization ............................................................................................................ 4

Chapter 2 Literature Review ......................................................................................... 7

2.1 Overview ................................................................................................................ 7

2.2 Urban spatial conditions, new establishments, and economic performance ....... 9

2.2.1 Spatial nature of new establishments and urban conditions ......................... 9

2.2.2 New establishments and entrepreneurship ....................................................... 11

2.2.3 New establishments and economic performance .............................................. 14

2.3 A historical review of early studies ...................................................................... 19

2.4 New establishments by industry sector ................................................................. 22

2.5 Determinants of new establishment location choice ............................................ 25
2.5.1 Supply side of new establishments ................................................................. 25
2.5.2 Demand side of new establishments ................................................................. 30
2.6 Geographic units of analysis ................................................................................. 31
2.7 Summary and Discussion ....................................................................................... 33
Chapter 3 Model and Methodology ............................................................................. 38
3.1 Background ............................................................................................................ 38
3.2 Typology of new establishment location choice behaviors .................................... 42
  3.2.1 Local businesses ............................................................................................... 44
  3.2.2 Home-based businesses .................................................................................. 45
  3.2.3 Spin-offs ......................................................................................................... 45
  3.2.4 National firms .................................................................................................. 46
3.3 A Multi-level approach to model formulation ......................................................... 48
Chapter 4 Variables, study areas, and data sources ..................................................... 53
4.1 Variables ................................................................................................................ 53
  4.1.1 Measuring new establishments ....................................................................... 53
  4.1.2 MSA level variables ......................................................................................... 56
  4.1.3 Census tract level variables ............................................................................ 62
4.2 Selection of study areas: 27 Medium-sized MSAs ............................................... 70
4.3 Building the Database ........................................................................................... 75
Appendix B: The two-level model.................................................................................. 169
Appendix C: Interview Questions.................................................................................. 172
Appendix D: List of open businesses.............................................................................. 177
Appendix E: Invitation Script.......................................................................................... 180
### LIST OF TABLES

Table 2.1 Summary of variables ........................................................................................................34

Table 3.1 Typology of new establishments .........................................................................................44

Table 4.1 Descriptive statistics ...........................................................................................................68

Table 4.2 U.S. MSAs by population groups .........................................................................................72

Table 4.3 27 MSA Summary ..............................................................................................................73

Table 5.1 New establishments by industry .........................................................................................82

Table 5.2 New establishments by organizational status .....................................................................87

Table 6.1 The results of the establishment model: The two-level Poisson estimation ..........95

Table 6.2 Establishment and Employment Model: The two-level linear estimation ......107

Table 6.3 Comparison of results across models ................................................................................111

Table 6.4 Standardized Coefficients: Census tract level results ...............................................112

Table 6.5 Standardized Coefficients: MSA level results ..............................................................112

Table 7.1 The top census tracts for each category ...........................................................................124

Table 7.2 The economic census tract variables ..............................................................................128

Table 7.3 The demographic census tract variables .......................................................................133
Table 7.4 Original establishment tract and relocated tract ........................................137

Table 7.5 The geographic census tract variables .........................................................138
LIST OF FIGURES

Figure 3.1 Four aspects of new establishments ................................................................. 40

Figure 3.2 Location decision-making process ................................................................. 47

Figure 3.3 Varying relationships between existing employment density and the number of
new establishments ........................................................................................................ 51

Figure 4.1 Map of 27 Medium-sized MSAs ....................................................................... 74

Figure 5.1 New establishments by industry ..................................................................... 83

Figure 5.2 New establishments by size ........................................................................... 85

Figure 5.3 Census tract size distributions ....................................................................... 89

Figure 7.1 Population and number of new establishments in 27 MSAs ......................... 122

Figure 7.2 Population and created employment in 27 MSAs ........................................ 122

Figure 7.3 The selected census tract for the interview of economic category .............. 127

Figure 7.4 The selected census tract for the interview of demographic category ......... 132

Figure 7.5 The selected census tract for the interview of geographic category .......... 137
CHAPTER 1 INTRODUCTION

Nearly every city and town seeks to attract new businesses to support job creation and have an impact on the health of the community. New business establishments do that, creating jobs immediately, disseminating ideas, and potentially drawing more businesses. Start-ups create far more jobs than incumbent firms. Between 1977 and 2005 in the U.S., start-ups created an average of three million new jobs annually, while existing firms eliminated, on average, one million jobs per year (Kane, 2010). New businesses are regarded as the economic engine of regional growth. However, start-up activities differ across regions, and across neighborhoods within a region. The primary purpose of this study is to examine what determines spatial variation in new firm formations.

I first became interested in the spatial distribution of new establishments, because it impacts land development patterns and shapes urban configuration. Little research has been conducted on the spatial nature of new establishments, even though there has been considerable research interest on the determinants of new establishment activities. Research has explored the determinants of new establishments around entrepreneurship theory in the field of business management. In this research, the primary focus is business organization and individual entrepreneurs so spatial factors have been noticed, yet little
studied. Another research strand focuses on the determinants of new establishments through economic growth theory in the field of economics. The empirical findings of the economics are too broad or not applicable since they do not fully consider characteristics of regions, but instead rely on economic conditions at the macro geographical level. Hence, spatial issues are noted, but still vastly underexplored.

Therefore, this study sheds light on the spatial positioning of new establishments, which depends on the uniqueness of certain areas. City governments and local municipalities have created and implemented a variety of strategies and incentives to stimulate more start-up activity within their own jurisdictions. Nevertheless, the results of the efforts to attract new businesses fall short of policy makers’ expectations. A policy of enhancing business climate is productive in some areas, while the same policy does not work for other areas. Central to this study is an effort to understand the relationship between the uniqueness of certain regions, spatially bounded characteristics, and how they both affect where new establishments locate, i.e. the effects of regional and neighborhood conditions on location choice of new business establishments.

1.1 Aims of the study

This dissertation aims to identify the spatial determinants of new establishment location decisions, and ultimately provide assistance to policy makers and planners. This study will clarify four key dimensions. First, I provide a hierarchical framework for modeling new establishment location choice behavior by applying a multilevel approach. This approach enables the empirical model to consider geographically macro-scale and
micro-scale effects simultaneously, whereas previous studies analyze only one spatial level, mostly the macro-scale effects. As spatial units of analysis in the U.S., the Metropolitan Statistical Area (MSA) for the macro regional scale and the census tract for the micro neighborhood scale are employed.

Second, the neighborhood level effects on new establishments are identified using variables at the census tract level. The advantage of the neighborhood level approach is empirical and practical. Empirically, I could quantify the effects of neighborhood characteristics on new establishments within a small geographic scope. This method helps to understand the spatial variation in new establishments between neighborhoods within a region, whereas most existing papers do not explore the neighborhood effects. In addition, one novelty of this study is including demographic characteristics that relate to the local demand for new establishments, which have largely been ignored in the literature. Practically speaking, geographically local factors have more direct effects on individual new establishments and can be improved easily compared to regional factors. Recent empirical studies have found that, for example, the average January temperature and the education level are significant regional factors that affect where new establishments locate. These findings are difficult to apply to local policies. Regional characteristics, such as the average January temperature and the education level, are not easily changed in the short term, since they have been formed over the long term. Furthermore, natural amenities such as a warm winter and proximity to the beach are unchangeable.

Third, I focus on the medium-sized metro areas in the U.S. by excluding large
metro regions and very small rural regions. By focusing on medium sized metro areas, my findings can more easily be applied to regions of similar size. Previous research has focused on larger areas and the findings were not necessarily applicable to mid-sized regions. Although the results of empirical models are significant, they do not guarantee to explain a phenomenon of a certain region. Most empirical studies using the U.S. data sets include all regions from the largest with a population in the tens of millions to a small city of less than ten thousand. To improve applicability, the sample needs to be more homogenous at least for the size of population.

Finally, this study uses mixed methods that integrate quantitative and qualitative approaches. On the one hand, a quantitative approach enables this study to measure and compare the magnitude of various factors at both regional and neighborhood levels. On the other hand, a qualitative approach of in-person interviews with start-up founders highlights the importance of understanding individual conditions when the start-up location choice is analyzed.

1.2 Organization

Chapter 2 reviews the literature on the factors that determine where new establishments locate. In particular, the spatial nature of new establishments is discussed in relationship with economic performance and urban characteristics. After reviewing early studies on new establishments chronologically in the U.S., I review recent empirical studies by demand and supply sides of new establishments, by industry, and by the geographical unit of analysis.
Chapter 3 conceptualizes the model, justifies the multilevel approach, and introduces the methodology. I introduce a typology of new establishments to model their location choice behaviors in response to regional and neighborhood characteristics. The two-level analysis framework is justified to simultaneously model new establishment location choice at the regional level and at the neighborhood level.

Chapter 4 introduces study areas, describes data sources, and discusses variables of the model. I explain the purpose of the selection criteria for 27 medium-sized Metropolitan Statistical Areas in the U.S. The Dun and Bradstreet database is introduced as the primary firm level data source. I also discuss how the selected variables impact where new establishments locate.

Chapter 5 analyzes the spatial distribution of new establishment activities by industry, size, and internal organization. This approach is a firm level descriptive analysis to convey an overall picture of spatial distribution patterns. In addition, the descriptive results of the census tract level examination are discussed, since the census tract is the primary spatial unit of analysis.

Chapter 6 discusses the results of the two-level model to specify and understand the effects of each variable on new establishment location choice. I pay attention to the census tract level results by taking a closer look at economic, demographic, and geographic environments. The difference between the factors influencing establishment and the factors influencing employment is also explored.

Chapter 7 examines new establishment location choice by using a qualitative approach to test how the model predicts actual cases properly and to comprehend who the
founders of start-ups are. In-person interviews with selected founders in the Columbus MSA clarify the context of the decision-making process at the individual level. In addition, I can verify how the factors considered in the model play a role in influencing the actual cases.

Chapter 8 summarizes research findings, discusses policy implications, and presents future research directions.
CHAPTER 2 LITERATURE REVIEW

The purpose of this chapter is to survey the literature on new establishments. To determine the scope of the literature review, the meaning of new establishments will be discussed first, in particular, as an association with economic performance and urban conditions. Two streams of literature were reviewed; the first examines determinants of new establishments and the second focuses on consequences of new establishments. The third section reviews early studies on establishments chronologically to examine traditions of start-up research based on industrial location theory in the U.S. context. The fourth section examines empirical studies across industries. The fifth section compares primary determinants studied in the literature. The sixth section discusses the unit of geographical analysis. The final section summarizes the limitations of the existing literature.

2.1 Overview

Understanding the relationship between new establishments and economic performance is key for planners and policy makers to implement strategies that stimulate economic growth. Although the relationship between new establishments and economic performance is noted, the two topics have not been much studied as an associated issue
especially in the planning field. A primary reason for this problem is that the two topics have been established in different academic foundations. While in the business management field start-ups are examined as realized outcomes of entrepreneurship, in the economics field economic performance is a primary theme of economic growth.

This review bridges the gap between the two topics by providing urban conditions as meeting grounds, since new firm formation and economic performance are ultimately spatial phenomena and urban settings are an essential space for businesses. For instance, every new establishment requires a physical location to launch its businesses, even in the cases of internet-based establishments or shell corporations. Economic performance is generally measured by the unit of certain geographic or administrative boundaries such as a country or a city. Hence, new establishments and economic performance are tied to spatial issues.

There are two main streams of research in the literature on new establishments. The first examines determinants of new establishments, and the second focuses on the consequences of new establishments. I will review mostly studies in the first stream, since the focus of this study is on regional and neighborhood conditions as primary determinants of new establishments’ location choice. First, I examine the historic origin of the new business research field in the US context. Second, more recent papers are discussed by the industry sector because most studies concentrate on a specific sector of industry and determinants of new establishments depend on characteristics of the industrial sector. Third, determinants of new establishments are discussed in more detail, by dividing them into supply and demand sides.
The second stream, on consequences of new establishments, includes multiple research fields\(^1\), and I do not review all issues that go beyond the interests of the current study. Instead, I focus on economic influences of new establishments, such as job creation and economic growth, for two reasons. First, enhanced economic performance as a positive outcome of new establishments is a central theme of economic development policy. Second, the economic outcomes of new establishments can be determinants of new establishments as well, because of a hypothetical endogenous relationship between them. For instance, newly created jobs from new businesses are in turn able to stimulate additional new businesses. Hence, economic performance as an outcome of new establishments needs to be reviewed as a potential determinant. After a brief discussion of the relationship between regional and neighborhood conditions, new establishments, and economic performance, the following sections review empirical studies on the determinants of new establishments.

2.2 Urban spatial conditions, new establishments, and economic performance

2.2.1 Spatial nature of new establishments and urban conditions

Spatial positioning is a significant concern for establishments to be activated (Andersson, 2005). Hence surrounding geographical environments now play a key role in

\(^{1}\) For example, the literature on innovation examines spinoff new establishments as key drivers of the dissemination of new knowledge (e.g. Klepper and Sleeper, 2005). See Davidsson (2008) for the overview of consequences of new establishments.
where new establishments locate (Malecki, 2009). In particular, clustered economic activities (Delgado et al. 2010) and proximity to knowledge sources (Baptista and Mendonca, 2010) are highlighted as necessary conditions in recent research. Urban settings provide an ideal environment to satisfy such conditions (Renski, 2008; Frenkel, 2001).

The strand of research in urban studies that has primarily explored the spatial conditions for start-ups is the agglomeration literature (Glaeser et al., 2010).

Opportunities perceived at cluster locations are pursued there because barriers to *entry* are lower than elsewhere. Needed assets, skills, inputs, and staff, often readily available at the cluster location, can be assembled more easily for *a new enterprise*. (emphasis added, Porter, 2000, p. 263)

Consider Porter’s argument about the benefit of clustering. It turns out that his argument deals with key issues of new establishments such as “*entry*” and “*a new enterprise*.” Although his focus is the cluster theory, the mechanisms associated with the entry of new businesses in agglomerated areas (Delgado et al., 2010) are similarly applicable to the spatial determinants of new establishments when it comes to location decision-making.

If that is the case, what factors regarding agglomeration relate to where start-ups locate? Since the seminal work of clustering theory by Marshall (1925), there are three main reasons of localization economies and co-location of firms (Krugman, 1991; Gordon and McCann, 2005). The three classes of mechanisms are more efficient sharing of local inputs, better matching between business partners, and learning from idea
spillovers (Duranton and Puga, 2004; Puga, 2010). All these topics directly relate to start-ups.

For example, this three mechanism framework can explain why high-tech start-ups are so abundant in Silicon Valley. New firms in the information technology industry prefer to locate there to pursue proximity to well equipped local suppliers and to tap into the labor pool that specializes in the IT industry. In addition, novel start-ups often arise from established incumbent firms through spin-offs in which apparent idea spillovers are observed.

2.2.2 New establishments and entrepreneurship

This section reviews a model of new establishments based on the entrepreneurial decision-making behavior. While the model delineates a choice of new establishments, the spatial nature of the start-up is missing even though that is a crucial stage for a new establishment to be active. Consequently, the importance of spatial positioning of new establishments is emphasized as a conclusion.

When can a new business be created? As a start-up is an outcome of entrepreneurial activities\(^2\), the domain of entrepreneurship research is briefly reviewed. There are diverse definitions of entrepreneurship such as “whether entrepreneurship should be defined in terms of dispositions, behavior, or outcomes; whether it belongs in

---

\(^2\) To be entrepreneurial is not a necessary condition for new establishments. Although many new establishments merely launch their businesses, this review focuses on entrepreneurship as a theoretical foundation of the start-up decision-making process.
the economic-commercial domain or can be exercised also in not-for-profit contexts; whether it belongs in only small and/or owner-managed firms or in any organizational context” (Davidsson, 2008, p. 13). Nevertheless, a consensus among different perspectives is that entrepreneurship is the phenomenon that relates to the “emergence of new economic activity” (emphasis added, Wiklund et al., 2011).

The next question is: what are the conditions for the “emergence?” To simplify economic activities of the emergence, an economic theory models the entrepreneurial activity as the market-entry decision for an individual’s occupational choice between entrepreneur and employee (Gries and Naude, 2009). The main idea of this approach is that a person can become an entrepreneur when the profits of opening a firm exceed the wages and other benefits of being an employee. Murphy et al. (1991) suggest the production function for an occupational choice which incorporates entrepreneurial ability $A$ that measures how much he can improve operating technology.

When $s$ is the common technology, $F$ is the relation between labor and output, $H$ is the human capital of workers or size of the new firm, and $w$ is the wage rate, the profits of being an entrepreneur are:

$$ y = s \cdot A \cdot F(H) - w \cdot H. $$  

(2.1)

When this same person would be hired as an employee, his income is $w \cdot A$. Accordingly, a person will make a decision to start a new firm when$^3$

---

$^3$ The size of firm is $H(A)$ according to the first-order condition with respect to $H$;

$$ s \cdot A \cdot F'(H) = w. $$
\[ s \cdot A \cdot F(H(A)) - w \cdot H(A) + \eta > w \cdot A + C. \] (2.2)

Gries and Naude (2009) expand this model by adding the non-pecuniary benefits of entrepreneurship \( \eta \) and new establishment opportunity costs \( C \) including additional benefits of being an employee.

The equation (2.2) abstracts key features of the start-up phenomenon. The entrepreneurial ability \( A \) shows the importance of the talent that belongs to entrepreneurs (Helfat and Lieberman, 2002). More abundant knowledge and skills could improve the marginal production from inputs and reduce the possibilities of failure (Lloyd-Ellis and Bernhardt, 2000). The wage rate \( w \) is another defining factor behind the entrepreneurial decision process. Higher expected income will generate fewer chances of start-ups. Finally, new business investment as a sunk cost explains risk taking and institutional issues such as legitimization.

The weakness of the model is in formulating entrepreneurial activities as a non-spatial process, although all factors in the model are essentially spatial. Wages and entry costs, for instance, vary across regions and the differences are not ignorable. It seems evident that the start-up process might be best viewed as a result of a combination of a non-spatial production equation and spatial attributes.
2.2.3 New establishments and economic performance

This section reviews two issues that link new establishments and economic performance. The first issue is to integrate the factors related to start-up activities into their overall economic performance by using the concept of entrepreneurship capital. The second issue is to examine the causal relationship between new establishments and economic performance.

2.2.3.1 Entrepreneurship capital

Based on the framework of new firm formation in the previous section, this section examines the role of founders that affect economic performance. Three roles of entrepreneurs are highlighted to examine the importance of decision-makers in the process of new establishments (Carree and Thurik, 2010). The first role is an innovator first conceptualized by Schumpeter. The concept of “creative destruction” is introduced, in which a turnover of economic units accelerates the process of structural change (Fritsch, 2008) by replacing incumbent firms with new establishments. The second role is a perceiver introduced by Kirzner (1997). The founder’s role is discovering profit opportunities that others do not recognize. The third role is a risk-taker of the Knightian entrepreneurship (Shane, 2000). Uncertainty is a key nature of any start-up businesses, and therefore the founder’s role as a sole decision-maker is critical.

When the roles of founders are emphasized, the concept of the start-up is not limited to the cases of new firms. Essentially, entrepreneurship is a behavioral characteristic of individuals. For instance, enterprising individuals in large corporations,
the so-called “intrapreneurs” and “corporate entrepreneurs,” behave similarly as actual individual entrepreneurs do (Carree and Thurik, 2010). This is an effort to mimic smallness or newness by using business units or subsidiaries.

Audretsch and Keilbach (2004) introduce the idea of *entrepreneurship capital* to conceptualize entrepreneurs’ role in the framework of microeconomics. This is “the capacity for geographically relevant spatial units of observation to generate the startup of new enterprise.” In their empirical model, the capital is measured by the number of start-ups relative to population in each region. They argue that Solow’s (1970) growth model misses a significant factor of economic growth, namely innovation generated by entrepreneurial activities. Such entrepreneurial activities could be considered to reflect the underlying stock of entrepreneurship capital.

Eventually, they suggest a modified Cobb-Douglas production function that includes entrepreneurship capital as an additional variable to the traditional equation of physical capital and labor.

\[ Y_i = K_i^\alpha L_i^\beta E_i^\gamma \]  

(2.3)

where \( Y_i \) is economic performance of region \( i \), measured as GDP, \( K_i \) is region’s \( i \) endowment of capital, \( L_i \) is labor, and \( E_i \) represents the region’s endowment entrepreneurship capital. Hence, this specifies formally that entrepreneurship capital contributes to the economic output of regions. This approach is an extension to the one chosen by Mankiw, Romer and Weil (1992) who emphasize the impact of a region’s human capital while Audretsch and Keilbach (2004) focus on entrepreneurship capital.
This concept needs to be improved more theoretically as its approach of extending Romer’s growth model makes many assumptions (Audretsch and Keilbach, 2005). Nevertheless, this approach is useful to evaluate new establishments effects on economic growth empirically, since the effects can be easily added to the Cobb-Douglas functional form.

2.2.3.2 The causal relationship between new establishments and economic performance

The effects of new establishments on economic development have been discussed by many recent studies theoretically (Fritsch, 2008) and empirically (See van Praag and Versloot, 2008 for a review of recent research). There is an interesting question about the direction of causality between new establishments and economic development. Do start-ups influence economic development or is it the other way around?

First, the effects of new establishments on economic development are clear and immediate\(^4\). For instance, start-ups always generate new jobs, by definition, and this contribution to job creation is significant especially compared to incumbent firms (Kane, 2010, Haltiwanger et al. 2012). In addition, newly added establishments are conducive to advancement in creativity, variety, and efficiency in the market because of using new

\(^{4}\) Carree and Thurik (2008) argue that this is not immediate, but there is a time lag structure when it comes to the impact of changes in the number of business owners on three measures of economic performance; employment growth, GDP growth and labor productivity growth.
ideas, providing diversity, and increasing competition, respectively. Second, the effects of economic development on new establishments are also active even though the effects are not instant. One common index of economic growth is increased GDP, and this will raise overall demand for new services and products. Hence, there will be more chances of success for new entries due to results of an economic upturn.

Hartog et al. (2010) explain this two-way relationship by using a conceptual model as follows. When a region’s GDP is used as a measurement of economic performance, new establishments of time \( t \) will increase the GDP at the same time \( t \) by definition of the GDP. Since new establishments provide newly added products and services, these new additions increase total value of the GDP. However, when the GDP affects new establishments, it will take time to be effective such as one year \( t+1 \) or two years \( t+2 \) when the effect is quantified annually\(^5\). As a result, very few studies discuss this mutual relationship simultaneously because of the time-lag structure (van Stel and Storey, 2004).

Nevertheless, Audretsch and Keilbach (2005) is, to my knowledge, the only study that models the relationship empirically by using a simultaneous equation framework. They measure the entrepreneurship capital by the number of new businesses in the county.

\(^5\) If certain new establishments predict the GDP growth beforehand according to a business cycle forecast and behave proactive, the time lag could be one year \( (t-1) \) or two years \( (t-2) \) earlier. This current study sets the time lag as one year \( t+1 \), and details are discussed in the Chapter Four.
(Kreise) relative to its population in Germany during 1998 to 2000\(^6\). To include new businesses endogenously, two equations are suggested in the framework of the simultaneous equation model.

The first equation is a Cobb-Douglas form of an aggregate production function

\[
Y_i = K_i^\alpha L_i^\beta R_i^\gamma E_i^\delta 
\]

(2.4)

where \(Y_i\) is economic performance of county \(i\), measured as GDP, \(K_i\) is county’s capital, \(L_i\) is labor, \(R_i\) is R&D intensity and \(E_i\) represents the county’s entrepreneurship capital. The second equation is

\[
E_i = f(y_i, x_i) 
\]

(2.5)

where \(y_i\) is a vector of measures of a county’s GDP and \(x_i\) is a vector of other variables influencing new establishments. The 3SLS (three-stage least squares) approach is used to estimate these two equations.

The model results show that the new establishment rate increases the GDP of counties, whereas the effect of the GDP on new establishments is not statistically significant. However, the GDP’s growth, measuring a county’s average growth rate between years 1992 and 2000, has a positive effect on new establishments. An increase of

\(^6\) Audretsch and Keilbach (2005) measure the variables used in the model in the year of 2000 except new businesses.
the GDP's growth by one percentage point increases the new establishment rate by about 30%.

Hence, the model does not validate the two-way simultaneous relationship when economic performance is measured as a stock of the GDP. However, one-way direction effects of start-ups on the GDP and the GDP's growth on start-ups are estimated. Although this empirical model does not prove its original hypothesis of the relationship between new establishments and regional economic performance, its findings are valuable when I develop an empirical model. For instance, a comparison between stock of the GDP and the GDP's growth is necessary for the measurement of regional economic performance.

2.3 A historical review of early studies

The seminal empirical research on new establishment location decision-making in the U.S. is “Why industry moves South: a study of factors influencing the recent location of manufacturing plants in the South” by McLaughlin and Robock (1949). By using surveys, they examine the location choice behavior of manufacturing industries that locate in the southern states. They conclude that the primary factors driving business location decisions are proximity to increasing demand in the southern states and better labor market conditions such as less unionization.

In line with this research stream on the movement to the South, Jusenius and Ledebur (1977) examine the trends of the migration behavior of firms across industrial sectors by using Dun and Bradstreet data between 1969 and 1974. They find that the
employment problems of New England are not the consequence of outmigration of existing firms to southern states, which was a common explanation for employment losses in New England. In fact, New England states were a net beneficiary in terms of the migration of firms. The birth and death of firms shows that the number of start-up firms was greater than the number of closures in the Fire, Insurance, and Real Estate (FIRE) industry. While about 8,000 firms disappeared in the manufacturing industry, only 4,389 new firms emerged. In addition, the closed firms had, on average, a larger number of employees per firm than the start-ups. The primary cause of employment problems in New England was a relatively smaller start-up birth rate compared to the death rate of companies, particularly in the manufacturing sector.

In the 1980s, statistical approaches were used to model location choices of new establishments, whereas previously descriptive analysis (e.g. Mueller and Morgan, 1962) was widely used. There are two reasons for the popularity of statistical approaches. The first is the increase in data availability on new firm formation and the second is advancement in statistical methodology, in particular the discrete choice model (McFadden, 1974).

Carlton (1983) developed an econometric model that examines the location decisions of new manufacturing plants in the United States by using Dun and Bradstreet data between 1967 and 1971. The model estimates the determinants of location choices across SMSAs (Standard Metropolitan Statistical Areas) for three manufacturing industries: fabricated plastic products, communication transmitting equipment, and electronic components. Since the unit of analysis is a plant establishment, a conditional
logit model is used to examine the factors influencing new establishments in each SMSA. The primary findings are that high energy costs, including the cost of electricity and natural gas, can strongly deter new manufacturing plants from locating in a certain area. However, property and income taxes, and state incentives are not significant factors, although theory expects a negative tax effect and a positive incentives effect (Schmenner, 1982). Carlton measures state incentives by counting the number of incentives provided to new firms, for example, state loans for construction, sales tax exemption on new equipment, and revenue bond financing. Although it deals with only three manufacturing sectors, this research is important as a pioneer in using a statistical approach in measuring factors influencing start-up location decision.

Bartik (1985) builds on Carlton (1983) research by suggesting that the probability of location choice would increase proportionally to the land area of a geographic unit. Land area is a proxy for the number of potential sites, and this is the so-called “dartboard theory” in the location choice literature. “If two states are identical in observed characteristics, and the unobserved characteristics of sites are independent drawings from a distribution, the state with twice as many sites should have twice as great a probability

---

7 In conditional logit model, the expected utilities are modeled in terms of the characteristics of the location choice alternatives. Conversely, multinomial logit model models the expected utilities in terms of the characteristics of the firms.

8 See Ellison and Glaeser (1997) and Guimaraes et al. (2009) for recent development in “dartboard theory.”
of being chosen” (Bartik, 1985, p. 16). Using 1,607 new manufacturing plants of Fortune 500 companies between 1972 and 1978 in the U.S, a conditional logit model is used to examine the effects of states’ attributes on the location decision. It concludes that unionization, which is measured by percentage unionized in a state, is the most significant factor of a manufacturing plant’s choice of location across the U.S.

2.4 New establishments by industry sector

Empirical research has focused on the manufacturing sector (e.g. Bartik, 1988; Becker and Henderson, 2000) compared to other industrial sectors. In particular, U.S. empirical papers mostly focus on specific cases of manufacturing such as the auto industry (Smith and Florida, 1994; McConnell and Schwab, 1990) and FDI (Foreign Direct Investment) in manufacturing sector (Coughlin et al., 1991; Head et al., 1999; Luker, 1998).

Guimaraes et al. (2004) is first introduced as these authors develop the most general model that includes all manufacturing sectors and covers the entire 48 contiguous states between 1989 and 1997. The model quantifies new establishments as the number of new plants per county. By using the Standard Statistical Establishment List that is a special tabulation from the U.S. Census Bureau, a Poisson model estimates the effects of several explanatory variables on the number of plants in 20 two-digit SIC manufacturing sectors. They incorporate all manufacturing sectors in an aggregate model to concentrate on overall effects. However, this is a limitation of this study since the effects by
individual sector at two-digit SIC level in addition to the aggregated model could be examined due to their detailed dataset.

The explanatory variables include wage as a proxy for labor cost, population density as a proxy for land cost, total personal income as a proxy for market size, the density of existing firms in the same two-digit SIC as a proxy for localization economies, the density of existing service and manufacturing firms as a proxy for urbanization economies, and per capita property tax as a proxy for the tax business climate. The model also includes states’ dummy variables to control state-level characteristics; however the dummy variables are not significant. The results show that cost factors such as labor cost, land cost, and taxes have negative impacts on start-ups, but market size, localization and urbanization economies have positive impacts on establishments at the county level.

While Guimaraes et al. (2004) concentrate on the overall manufacturing industry, Papke (1991) examines individual five three-digit SIC sectors including Women’s and Misses’ Outerwear (SIC 233), Household Furniture (SIC 251), Book Publishing and Printing (SIC 273), Communication Equipment, including Radio and TV (SIC 366), and Electronic Components and Accessories (SIC 367). Using the Dun and Bradstreet database, new establishments are measured by the number of new plants at the state level between 1975 and 1982. A pooled model and five industry models are estimated by using a Poisson method. The principal finding is that land price has a

\[\text{land price} \]

\[\text{Papke (1991) uses land price value with the price in hundreds of dollars per acre of farm land that was constructed from the U.S. Department of Agriculture’s ‘Farm Real Estate Market Development.’}\]
negative impact on new establishments, whereas state population shows a positive impact.

There are several papers examining the aggregate manufacturing sector in Europe. Arauzo and Manjon (2004) use a Poisson model to analyze the location choice of manufacturing start-ups in Catalonia, Spain between 1987 and 1996. The unit of analysis is the municipality, and start-ups are measured by the number of new establishments in a municipality. The study found that employment density has a positive effect, while population density has a negative effect on the number of new establishments. They conclude that “the preferred environment of households and industrial establishments is not necessarily the same” (p.309).

While Arauzo and Manjon (2004) concentrate on aggregate manufacturing industry, Arauzo (2005) studies five sub-groups of manufacturing according to the OECD classification. The five classifications are natural resources, labor intensive, economies of scale, differentiated products, and R&D intensive sectors. For instance, natural resources intensive sectors include food, tobacco, and paper products. By using a Poisson model at the municipality level in the ten years between 1987 and 1996, this research finds that urbanization economies measured by the density of jobs (jobs/km$^2$), the ratio of employment in the industrial sector, and industrial diversity have positive effects, while commuting intensity measured by kilometers traveled daily per employee has a negative effect on the number of new establishments. In a breakdown across sectors, high levels

---

$^{10}$ In 1986, the average population of these municipalities was 6,367 (Arauzo 2005, p.110).
of commuting decrease\textsuperscript{11} the expected number of new establishments, especially in labor-intensive sectors, but firms in the research and development (R&D) sectors are less affected by commuting intensity.

### 2.5 Determinants of new establishment location choice

#### 2.5.1 Supply side of new establishments

The supply side of new establishments relates to start-up cost. Financially low barriers to entry make new establishments easier. Start-ups, for example, need accessible infrastructure, land or office space, license, as well as tax incentives.

*Sharing infrastructures*

Coughlin and Segev (2000) analyze location determinants of new foreign-owned manufacturing plants at the county level in the United States and find that higher levels of transportation infrastructure are associated with larger numbers of new plants\textsuperscript{12}. They

\textsuperscript{11} Coefficients for commuting intensity are -0.145 (natural resources intensive), -0.246 (labor intensive), -0.140 (scale economies intensive), -0.145 (differentiated products), and -0.106 (R&D intensive).

\textsuperscript{12} Theoretically, this association has been studied by Chandra and Thompson (2000), Duranton and Turner (2012), although their focus is the relationship between transportation and aggregate economic activities.
model transportation infrastructure via a dummy variable according to whether or not an interstate highway is located in the county. Although they find a positive relationship between existence of highways and the location choice of new plants, they did not consider quantifying accessibility of infrastructure such as the distance from the highway to the new plants.

A paper analyzing the effect of transportation infrastructure as a dominant factor on new establishments is Holl (2004a, b). Holl's research improves the measurement of accessibility to transportation compared to Coughlin and Segev (2000), as Holl's research includes tiers of zones to classify the distance from the highway by using multiple dummy variables. The years between 1986 and 1997 are selected, since there was a significant improvement to the highway network in Portugal. Looking at the manufacturing and service sectors at the municipality level, Holl shows that the number of start-ups drops beyond ten kilometers from the new highways. This means that the construction of new highways has a negative effect on new establishments beyond ten kilometers that is the geographical scope of any new highway effect, because new firms are attracted and concentrated near the new highways.

Specialization

An industrial structure based on specialization is conducive to sharing intermediate input suppliers. The concept of specialization has changed from traditional sector specialization to functional specialization (Duranton and Puga, 2005). The first focuses on sharing similar suppliers, while the second relates to interaction with other
firms that are functionally alike. Another important aspect of specialization is the knowledge spillover, but this topic will be dealt with in detail later when I discuss a geographic scope of spillover effects.

Garofoli (1994) finds in the Italian case that provinces more specialized in a specific manufacturing sector attract more new businesses overall, and the specialization effect on manufacturing start-ups is about 1.4 times stronger than the effect on start-ups of all industries. However, this result is quite predictable because this study measures specialization by including only the manufacturing industry without considering other industries, since the specialization effect would be stronger on the same industry than the effect on the overall industries.

Qian and Li (2008) develop an empirical sector specialization model of entrepreneurial activities measured by the number of new firms per person in MSAs. The model is an extension of Garofoli’s approach (1994) because it includes multiple industry sectors. Location quotients of eight two-digit NAICS industries, used as proxies for sector specialization, are primary explanatory variables in the model. The result shows that the location quotients for construction, information, and the real estate industry positively affect new firm formation. However, the location quotient for retail trade shows a significant negative effect on new establishments. Other location quotients are insignificant.

A recent article by Audretsch et al. (2011) classifies 71 West German planning regions into three categories according to the functional specialization model (Duranton and Puga, 2005). The three classes are headquarters’ cities dominated by service firms,
industrial cities with large stand-alone plants in one sector, and cities with integrated smaller firms. They conclude that the third type of city is friendlier to new firms, whereas the first and second type gives small chances of entries.

*A pool of workers*

For a start-up firm, a pool of workers is an essential factor. As uncertainty is a fundamental characteristic of start-ups, new businesses reduce uncertainty by hiring experienced and skilled personnel (Angel, 1991; Takatsuka, 2011). Audretsch et al. (2011) find an evidence to support this relationship. Increasing the share of workers with engineering or natural science degrees by one percentage point would raise the number of start-ups by 8.02% in the manufacturing sector. Conversely, a one percentage point increase in the share of humanities degrees would decrease manufacturing start-ups by 4.3%.

Arauzo and Viladecans (2009) develop an empirical model that highlights the relationship between the stock of human capital and new establishments. Two proxies are selected to measure human capital: the percentage of the population with a bachelor’s degree and the percent who completed secondary school and intermediate level education. Interestingly, the university-level variable and intermediate-level variable show contrary effects on start-ups. Intermediate-level education increases start-ups, whereas more university-level education has a negative effect on start-ups. They argue that new firms would prefer to avoid higher wages, but this model is not enough to confirm this hypothesis. There are many other possible effects of the education level
differences other than the wage difference. For example, university degree job seekers prefer established incumbent firms to newly started firms.

Another factor that relates to the labor pool is the unemployment rate. Davidsson et al. (1994) find in the Swedish case at the regional level that a higher unemployment rate is likely to result in less new firm formation. Woodward (1992) discovers a similar result in the case of Japanese manufacturing establishments in the United States. Woodward’s empirical model at the county level confirms the negative impact of unemployment on start-ups in the manufacturing sector. Moreover, the impact is about 1.5 times stronger in case of the “auto alley” states.\(^\text{13}\)

*Local Policy*

Public-policy determinants are rarely studied due to the difficulty in collecting policy relevant data and in quantifying the effect on new establishments. The most typical variable considered is the tax rate. Most studies confirm that lower corporate tax rates would attract more new establishments at the state level (Papke, 1991) and at the county level in California (Coughlin and Segev, 2000; List, 2001), and in Maine (Gabe, 2003). One unique study on policy intervention is List and McHone (2000) that considers environmental regulation in the state of New York. By using a dummy variable at the county level that indicates the existence of pollution regulations for ozone, they find a

\(^{13}\) The study includes regional dummy variables to divide the U.S into eight regions. The author calls East North Central and East South Central regions as “auto alley,” where many Japanese suppliers opened greenfield plants during the 1980s. (p.701)
negative effect of environmental regulation on the number of new establishments in seven pollution-intensive two-digit SIC industries.

Chrisman et al. (2002) examine the effectiveness of the Small Business Development Center (SBDC) program at the state level in the United States by using the result of a survey that was sent to all firms who received five or more hours of counseling assistance from SBDCs in 1994. The main finding of this research is that rural states, where more than 40% of population live in non-metropolitan counties, have more new businesses than non-rural states.

2.5.2 Demand side of new establishments

The demand side of new establishments relates to the expected benefits of selling their products or services. Catchment area, competition or cooperation with existing businesses, and demographic distribution are examples of demand-related factors.

Population

Many papers include population as a control variable or proxy for market size, and it always shows a positive sign coefficient (e.g. Luker, 1998). However, population density shows varying signs. Woodward (1992) finds that population density is a positive factor for new establishments of Japanese plants at the state level in the United States, but Arauzo and Manjon (2004) show a negative effect of density on new manufacturing establishments at the province level in Spain.
GDP and wages

GDP or wages is a measure of consumers’ purchasing power. Disdier and Mayer (2004) find that for French FDI (Foreign Direct Investment) firms’ GDP has a positive effect on the number of new firms. However, the paper shows that wages have a negative effect. Couglin et al. (1991) and Friedman et al (1992) find a positive effect, but Head et al. (1999) show a negative effect on start-ups. This is because Head et al. (1999) model new establishments for Japanese FDI plants, and for foreign companies the wage factor could work differently compared to domestic companies. This needs more information to confirm which direction is correct for the wage effect on new establishments.

2.6 Geographic units of analysis

The state is the sub-national administrative unit in the United States and most data are collected and available at this geographic unit. Bartik (1985) and Levinson (1996) model the location choice of manufacturing plants across states by using a logit model. Lugar and Shetty (1985) and Couglin et al. (1991) look more specifically at foreign firms investing in manufacturing plants at the states level with a logit model as well. Their common finding is that more existing manufacturing activity attracts more establishments.

The number of MSA level studies is small compared to the state level. Carlton (1983) and Glaeser and Kerr (2009) focus only on manufacturing sectors, but Strauss-Kahn and Vives (2009) examine relocation of headquarters for nine industries. Carlton (1983) highlights the size of the market as measured by population, Glaeser and Kerr
(2009) emphasize abundant workers in relevant occupations, and Strauss-Kahn and Vives (2009) find that airport facilities are a positive, and corporate taxes are a negative, factor in attracting headquarters.


In European studies, the municipality level is similar to the county level of the United States. For example, Guimarães et al. (2000) and Figueiredo et al. (2002) are papers for Portugal, and Arauzo (2005), Manjon and Arauzo (2006) are for Spain. All these studies examine the manufacturing sector.

Finally, smaller than municipality-level studies are very rare because of data availability. Rosenthal and Strange (2003) is the only study that models new firm formation at the zip code level. With the number of new establishments as a dependent
variable, six industries\textsuperscript{14} are modeled by using the Tobit count model. Since they use a finer unit of analysis, the focus is to measure the geographic scope of spillover externalities that decrease as distance increases, according to the New Economic Geography theory (Fujita, Krugman, and Venables, 1999). Their primary result is that the urbanization effect works within one mile, then vanishes, but the localization effect\textsuperscript{15} is significant up to five miles.

To sum up: in general, analysis at smaller spatial scales, such as the county level, utilizes count data models (CDM). Analysis at a larger spatial scale, such as the state level, uses discrete choice models (DCM) including the logit. This is because DCM has a difficulty in calculating the likelihood function when there are too many choices of location. The reason for focusing on the manufacturing industry study across states is due to the size of the market: manufacturing firms cover a national market, while service or retail firms tend to provide their products in, at most, regional markets.

\textbf{2.7 Summary and Discussion}

Table 2.1 on the following page is a summary of variables found to be significant in previous studies organized by the spatial level.

\textsuperscript{14}Six industries are software, food products, apparel, printing & publishing, fabricated metal, and machinery.

\textsuperscript{15}Urbanization effects are measured by the effects of other industries except for the own industry, while location effects are measured by the effect of other firms in the own industry.
<table>
<thead>
<tr>
<th>Variables</th>
<th>State</th>
<th>County/Metro</th>
<th>Local</th>
<th>Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Population</td>
<td>+</td>
<td>+</td>
<td></td>
<td>Coughlin and Segev (2000)</td>
</tr>
<tr>
<td>Population density</td>
<td></td>
<td>+</td>
<td></td>
<td>List (2001)</td>
</tr>
<tr>
<td>Existing employment</td>
<td></td>
<td></td>
<td>+</td>
<td>Rosenthal and Strange (2003)</td>
</tr>
<tr>
<td>Unemployment</td>
<td></td>
<td></td>
<td>–</td>
<td>Coughlin et al. (1991)</td>
</tr>
<tr>
<td>Skill/Education</td>
<td>+,–</td>
<td>+</td>
<td></td>
<td>Smith and Florida (1994)</td>
</tr>
<tr>
<td>Land cost</td>
<td>–</td>
<td>–</td>
<td></td>
<td>Guimarães et al. (2004)</td>
</tr>
<tr>
<td>Energy cost</td>
<td>–</td>
<td></td>
<td></td>
<td>Carlton (1983)</td>
</tr>
<tr>
<td>Environmental regulation</td>
<td></td>
<td></td>
<td>–</td>
<td>List and McHon (2000)</td>
</tr>
<tr>
<td>Business climate, public subsidies</td>
<td>+</td>
<td>+</td>
<td></td>
<td>Luker (1998)</td>
</tr>
<tr>
<td>Unionization</td>
<td>–</td>
<td></td>
<td></td>
<td>Head et al. (1999)</td>
</tr>
<tr>
<td>Distance to highway/airport</td>
<td></td>
<td></td>
<td>–</td>
<td>Gabe and Bell (2004)</td>
</tr>
<tr>
<td>Transportation infrastructures</td>
<td>+</td>
<td></td>
<td></td>
<td>Smith and Florida (1994)</td>
</tr>
</tbody>
</table>

Table 2.1 Summary of variables
There have been very few studies at local level. Therefore, the effects of most variables used at state or county level are not empirically tested at the local level. Comparing the effect at the state and county levels, I note there is a different set of variables that affect location choice according to geographical unit. For example, distance to highway and environmental regulations are meaningful variables only at county level. Conversely, unionization and energy cost are active at state level. The only controversial variables are related to the quality of employees. Since a firm that uses higher techniques prefer skilled workers, that means higher wages. However, labor intensive firms that do not need skilled workers hire low-skilled and therefore low income labors. Therefore the signs of skill and education variables could go in either direction.

Recent papers examine new establishments at a smaller unit of analysis such as the local level, whereas earlier studies used national or sub-national level as a spatial unit of analysis. However, there are few studies that consider more than one level simultaneously. Although the unit of analysis is determined by the interests of researchers, the geographic unit is typically selected by available data for empirical research (Arauzo and Manjon, 2011). However, a change in the level of unit presents different results for determinants of start-ups. At a macro level like states in the U.S., all states could have similar attributes because any variations inside the states are averaged

\(^{16}\) Arauzo-Carod (2008) models manufacturing startups at three different levels that are 941 Municipalities, 41 Comarques (county), 74 TTWAs (travel-to-work area) in Catalonia, but the three levels are actually three ways of aggregate that varies in its unit size by including the identical independent variables in the model.
out at the state level. When I consider the relationship between the unit of analysis and choice of methodology, in general, the smaller unit of analysis such as the county utilizes count data models (CDM), while a bigger unit of the state uses discrete choice models (DCM) such as the logit. This is because DCM has a difficulty in calculating the likelihood function when there are too many choices of location like the county level. The reason that only the manufacturing industry has been studied across states is due to the size of the market. In general, manufacturing firms cover national market, while service or retail firms provide their products a local or regional market at most.

The literature review shows that most papers focus on one or two specific determinants of new establishments. Additional determinants are discussed jointly but those are included as control variables. I suggest determinants can be thought of as falling into two categories; the supply side and the demand side. On the one hand, start-up cost, available land, and licensing are examples of supply factors. On the other hand, catchment area, consumers’ income levels, and population growth are examples of demand factors. However, the result of my survey shows that most papers deal with mostly supply factors but few papers include demand factors in their empirical models. This concentration on the supply side research can be explained as follows. Most researchers are interested in firms’ utility that determines whether or not, an entrepreneur can launch a new firm, but underestimate the potential consumers of the new products or services that will be provided by start-ups.

The manufacturing sector is the most common type of industry in the literature. I note that there are few papers that deal with manufacturing and other sectors jointly in a
model (e.g. Strauss-Kahn and Vives, 2009). The popularity of manufacturing research in the literature is mainly due to easier data availability, because manufacturing is traditionally the most important sector as the economic base of a nation or a region, and so most countries collect and provide manufacturing establishments information. Additionally, location choices of manufacturing establishments are independent, and so their behaviors are to be analyzed by means of empirical models, whereas retail establishments are, in general, spatially distributed simply along with the population distribution. As a result, manufacturing establishments consider their location choice in a national model when its size is large enough to have a national or international market. But, smaller plants choose its facility location within sub-national or regional range.
CHAPTER 3 MODEL AND METHODOLOGY

The purpose of this chapter is to present the model and to discuss methodology. The first section discusses the background of the location choice behavior of new establishments by examining a multi-stage decision-making process. The second section introduces typology of new establishment location choices to develop a conceptual model. The third section introduces and justifies the multilevel framework of this study to model location choice behaviors of new establishments as a function of regional and neighborhood conditions.

3.1 Background

Where a company decides to locate is not a simple choice. The behavior is a result of a multi-stage decision-making process that involves a set of decisions that make it iteratively, but at each stage the prior decision can be reexamined. Where to locate depends on the following considerations. A fundamental question is who is starting the company, since the founder, as the primary decision maker, determines whether or not a company will be established. In addition, the timing of when it will be started and the products of what it will sell should be considered. In other words, a start-up happens "when an alert individual notices and comes to believe that the output or value of a given
resource could be higher if sold at another time, in another form, or in another location” (emphasis added, Plummer and Pe’er, 2010).

Who, where and when a business locates and what the company produces, under certain conditions, cannot be distinguished but should be dealt with concurrently. For example, when I do an empirical analysis of start-ups, available observations are established start-ups. The decision-making step of whether or not is already determined and I could detect new establishments by using the when question. Statistically, when is same as whether, plus the time dimension. Hence, if I conduct a cross-section analysis of a certain time, when replaces whether. Figure 3.1 shows conceptual relationships between four aspects of new establishments: location, founder, timing, and products. This study focuses on where to locate by examining regional and local conditions, while recognizing the other three aspects are not independent from the location decision. Following discusses how the four aspects are correlated around the location choice.

A. Where and Who

In investigating where founders locate their establishments, a region could have more start-ups than other regions due to the higher education level of founders, since highly educated people are more likely to be entrepreneurial (Audretsch et al., 2005; O’Shea et al., 2005). Accordingly, a variable measuring the education level of founders is to be included in the model because the variable factors into where founders locate. However, founders' propensity for risk-taking is not included in the model, although it is a noteworthy characteristic of founders that relates to new establishments. Propensity is not
measurable, and therefore I do not know if there is a disparity in risk taking across locations.

Figure 3.1 Four aspects of new establishments. The shaded area represents research focus
B. Where and When

The relationship between where and when a company locates depends on macroeconomic conditions involving the business cycle. The business cycle varies across regions according to the trend of growth or decline of the regions. Consider a declining metropolitan area such as Detroit and Cleveland. The number of new establishments in these declining regions will also tend to drop. Conversely, there will be increasing numbers of new establishments in growing regions such as Atlanta and Phoenix. Furthermore, where and when a company starts up is related to the characteristics of founders as well. As discussed in Chapter Two, empirical observations show that there is always a time-lag between the business cycle and start-up activities. The effect of economic growth (or decline) on additional (or reduced) establishments (van Stel and Storey, 2004) is not immediate. Start-ups are followers of the business cycle. On the other hand, proactive entrepreneurs forecast the business cycle and establish new businesses earlier than the peak of the cycle to respond to anticipated market expansion. In this way, new establishments are leaders of the business cycle.

C. Where and What (Products)

The products and services created by new establishments factor into where a company locates. Specialized industrial clusters are examples of start-ups’ dependency on where to locate. High-technology at Silicon Valley (Feldman et al, 2005; Angel, 1991) and biotechnology in the Greater Boston area (Reynolds, 2012) show how the dependency could drive a virtuous cycle in which agglomeration of start-ups, new
technology adoption, and regional growth are mutually enhancing. To examine how the virtuous cycle can occur, the process of spinning off is receiving attention because spin-off establishments tend to be located close to their parent companies. How spin-offs decide where to locate will be discussed more in the following section.

3.2 Typology of new establishment location choice behaviors

This section discusses a typology of new establishments that behave differently in terms of where they locate. Although different patterns of the decision-making process are observed, all types of new establishments can be modeled in a framework of a two-stage model. The two-stage location choice model consists of regional level decision and local level decision. In addition, the step of whether decision is also included along with the two-stage model, since the step shows contrast in location choice behaviors across different types.

The issue of where an establishment locates has different dimensions of choice sets, depending on who is making the decision. Unlike the assumption of homogenous firms in most models of microeconomics, companies are heterogeneous in size, industry, and organization. When they examine start-up activities, many papers, especially in the business literature, emphasize the role of firm-level attributes. The founder of the company is a significant factor because the founder or the organization is a decision-making entity for every aspect of start-up activities.

For instance, where a start-up locates hinges in part on the size of the new establishment. The size of businesses is a common indicator of internal scale economies
and access to financial capital. On the one hand for big international firms, new establishments are the result of expansions, such as additional regional branches or manufacturing plants. The primary decision step is a macro level location choice that is to find the best country or region for new establishments. On the other hand, small local businesses of one or two employees generally take the macro level location choice as given. They do not consider other regions for new establishments other than the region where the founders live.

Since new establishment is a comprehensive and inclusive definition of newly started business activities, more criteria need to be discussed to conceptualize a typology of new establishments. Luger and Koo (2005) argue that start-ups are *new*, *active*, and *independent* establishments\(^{17}\). For example, a national firm opens a branch office in a certain city is not a start-up because this establishment is not independent, but a case of new establishments creating new jobs and influencing economic outcomes of the city.

Following sections conceptualize the decision-making process of location choice behavior by utilizing the two-stage model in company with the *whether* step question. To provide a typology of new establishments, four representative cases are introduced: (I) local businesses, (II) home-based businesses, (III) spin-offs, and (IV) national firms. Figure 3.2 categorizes them based on the definition of start-ups by Luger and Koo (2005).

\(^{17}\) “which did not exist before during a given time period (new), which starts hiring at least one paid employee during the given time period (active), and which is neither a subsidiary nor a branch of an existing firm (independent)” (Luger and Koo, 2005, p.19).
New establishments

<table>
<thead>
<tr>
<th>Start-ups</th>
<th>Non start-ups</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I) Local businesses</td>
<td>(IV) Branches/Plants/Facilities of existing national firms</td>
</tr>
<tr>
<td>(II) Home-based businesses</td>
<td></td>
</tr>
<tr>
<td>(III-I) Independent Spin-off businesses</td>
<td>(III-D) Dependent Spin-off businesses</td>
</tr>
</tbody>
</table>

Table 3.1 Typology of new establishments

### 3.2.1 Local businesses

Local small establishments typically do not look to other regions, but locate close to the entrepreneurs’ existing location. For these local businesses, the region is given. Therefore, their location decisions mostly depend on local conditions, such as available office space and infrastructure, catchment area, existing peer-companies nearby, accessibility, like distance to highway, and physical conditions of surrounding neighborhoods. Nevertheless, regional conditions still impact the likelihood of new establishments. If a region has a high ethnic population and no existing ethnic grocery stores, a founder may make a decision to open a store. If the founder lives in a region with a saturated market or not enough demand, a new business might not be considered. The business climate, such as access to financing and low corporate taxes for small businesses, is another example of regional conditions for local businesses. In short, regional attributes affect new establishment decisions.
3.2.2 Home-based businesses

Increasingly, home-based businesses are being started up, in particular by self-employed founders. The choice of owning a house as an office space depends on the following considerations. The most common reason is to save on office rent, which is usually a large portion of the operating costs. Moreover, finding a preferred location is another effort in which founders have to make multiple decisions as I discussed so far. The increase in e-commerce and internet marketing also make it easier to start home businesses. If face-to-face interactions with customers or other businesses are not necessary, e-commerce can replace traditional transactions. Likewise, exposure to passing pedestrian and car traffic can be replaced by internet marketing. In this case, there is no new location choice for business start-ups. Instead, the housing choice plays a role in the location decision model. Of course, the two-stage model is still active since consideration of both regional and local conditions is a common approach in housing studies.

3.2.3 Spin-offs

"Spin-offs" are businesses created by an employee or employees of a firm who decide to branch off and start their own companies. In general, spin-offs locate near their parent companies to retain externalities and networks acquired when the founders were employees (Jofre-Monseny et al., 2011; Kerr and Kominers, 2012). A spin-off could be a start-up or a non start-up depending on the relationship with the parent firms according to the independency criterion of start-ups. An example of a non start-up spin-off would be a
franchise. In a franchise the parent firm, for example McDonald’s, plays a key role in siting decisions, competition in the region, marketing, and supplies.

To form a start-up spin-off business, first, potential founders of the spin-off need to be employees of parent firms. Spin-offs are prevalent where innovative industries are clustered, such as the IT industry in Silicon Valley (Klepper, 2010). Consider a graduate student of computer science, who is a potential founder of a spin-off in the future, looking for her first job. She will face a choice between existing IT firms, for example, in Silicon Valley, California and in Research Triangle, North Carolina. This step of a regional search for the first job is equivalent to a regional location decision of the two-stage model. Then, the decision whether to stay or spin off is the step of deciding whether or not to start-up a business. Later she may establish a spin-off, but typically within the region she chose for her first job. This step is a local location decision that is very similar to general start-ups. In short, this regional job search and local spin-off process is statistically the same as the location choice behavior in that it involves considering both regional and local conditions.

3.2.4 National firms

Consider this scenario: a national or international firm expands its business by opening new branches or by constructing new factories in a region or a county other than the location of the headquarters. In this case, the whether step is not really to start a new business, hence a case of non start-ups. Rather, the question is if additional branches or
factories are necessary in other regions. Then, it follows the common two-stage process in the literature: first choose a region, then locate at a site within the selected region (ex. Lambert et al., 2006). For a national firm seeking to locate a new manufacturing factory, deciding on a region from many candidate regions is a more important step than finding a site locally within the selected region. As a result, the location decisions of national firms mostly depend on regional conditions as a framework, such as business climate, market size, regulatory conditions, amenities, and demographics.

![Decision-making process diagram](image)

**Figure 3.2 Location decision-making process**

18 Although the location choice of international firms likely follows a three-stage process including country, regional, and local level decision, the country level stage is not considered to simplify the model.
3.3 A Multi-level approach to model formulation

The previous section justifies the two-stage model for all types of new establishments. Simply put, new establishments can be modeled as a function of regional and local conditions as follows. The most common approach in the literature is McFadden’s Maximum Utility approach that is developed by Carlton (1983) for location choice behavior of new establishments. As I discussed in the previous chapter, the expected profit function of a start-up firm can be estimated equivalently by using a count data model (Guimaraes et al. 2003).

\[ \text{Number (Start-ups)} = f(\text{regional attributes, local attributes}) \]  \hspace{1cm} (3.1)

When I estimate regional and local coefficients of the model by using a standard OLS, one problem of this method is that regional and local attributes are not independent. Rather, local attributes are nested across regions and they share the same regional attributes if they come from the same region. For example, if a new establishment locates at a northern neighborhood and another establishment locates at a southern neighborhood both should be under same conditions in terms of the regional context such as weather conditions.

As an alternative, multilevel modeling allows the regional level and local level to be modeled simultaneously. When I collect establishment data from \( J \) MSA groups at the regional level, each MSA includes a different number of census tracts \( n_j \). The number of new establishments at a census tract \( i \) in MSA \( j \) is the dependent variable \( Y_{ij} \). For a simple
case, the census tract level includes an independent variable \(X_{ij\,(CT)}\) such as population density, and there is one MSA level independent variable, \(Z_j\,(MSA)\).

**[Census Tract Level]**

\[Y_{ij} = \beta_{0j\,(CT)} + \beta_{1j\,(CT)} X_{ij\,(CT)} + e_{ij}.\] (3.2)

The intercept coefficients \(\beta_{0j\,(CT)}\) at the census tract level becomes an outcome variable in the MSA level regression model (3.3). Likewise, the slop coefficient \(\beta_{1j\,(CT)}\) becomes an outcome variable in the MSA level regression model (3.4).

**[MSA Level]**

**Intercept:** \(\beta_{0j\,(CT)} = \gamma_{00} + \gamma_{01} Z_j\,(MSA) + u_{0j},\) (3.3)

**Slope:** \(\beta_{1j\,(CT)} = \gamma_{10} + \gamma_{11} Z_j\,(MSA) + u_{1j}.\) (3.4)

The census tract-level residuals \(e_{ij}\) follows a normal distribution \(N(0, \sigma^2_e)\). The MSA-level residuals are distributed as multivariate normal, with each element of \(u_j\) having a mean of zero. The main difference with a usual regression model is that each MSA has a different set of intercept and slope coefficients. They are random coefficients interacting with the census tract level regression (3.2). See Appendix B for details of model specification.
Figure 3.3 illustrates cases of varying relationship between existing employment density and the number of new establishments\textsuperscript{19}. Figure 3.3 (a) shows a linear relationship with a positive slope. The number of new establishments is positively associated with existing employment. The positive relationship is consistent with the general idea of agglomeration economies. In this case, MSA level factors are not considered. Figure 3.3 (b) shows a “random intercepts” model and each of different MSAs has their own relationship between the existing employment density and new establishments at the census tract level. The parallel lines indicate that certain MSAs have uniformly higher establishment rates than other MSAs. Nevertheless, the relationship between the number of new establishments and existing employment density is the same across MSAs, as the slopes of lines are same. Figure 3.3 (c) illustrates that low employment density census tracts have a similar effect on new establishments across MSAs, while high density census tracts show very different performance in attracting new establishments across MSAs. Conversely, Figure 3.3 (d) shows that low density census tracts have a different effect on new establishment across MSAs, while high density census tracts influencing new establishments are alike.

\textsuperscript{19} The explanation of varying relationships is originally suggested by Duncan et al. (1998).
These four types of relationships can be applied to my selection of nested model approaches as follows. The relationship (a) justifies a pooled model by no consideration of differences across MSAs. In this case, the main hypothesis is that new establishments are not affected by regional conditions, and the framework effect at the MSA level is not significant. Consequently, I do not need to use a nested modeling. The relationship (b) presents the case of the relationship between the number of new establishments and local-

\[ \text{Number of new establishments} \]

\[ \text{Existing employment density} \]

Figure 3.3\textsuperscript{20} Varying relationships between existing employment density and the number of new establishments

\[ \text{Number of new establishments} \]

\[ \text{Existing employment density} \]

\[ \text{Figure 3.3} \]

\[ \text{A modification of Figure 1 from Duncan et al. (1998, p.99).} \]

\[ \text{51} \]
level conditions are the same, but the intercepts vary across MSAs. This case justifies a nesting model that explains the reasons for why the intercepts vary and the meaning of the differences across MSAs. The relationship (c) shows the case in which the intercept is the same but the slopes vary across MSAs. The explanation of nesting for this case is to look for reasons of variation in slopes that may depend on different conditions across census tracts. The relationship (d) combines (b) and (c) and it represents two types of variation across MSAs as well as across census tracts. Thus, this case is the most generalized model and therefore a nested model considers two variations simultaneously in the slope as well as in the intercept. Because there is at least one variation across MSAs or census tracts, I will model (b), (c), (d) cases by using multilevel nested structures.
CHAPTER 4 VARIABLES, STUDY AREAS, AND DATA SOURCES

This chapter discusses the choice of variables, study areas and data sources. First, the variables measuring new establishments and variable selection at the MSA level and at the census tract level are introduced. Second, I attempt to justify my focus on the medium-sized metropolitan areas as the study areas throughout this research compared to previous studies in the literature. Third, primary data sources and how they are processed will be discussed.

4.1 Variables

4.1.1 Measuring new establishments

This section defines the measurement of new establishments, the dependent variable of the model. Besides a direct count, there are several indirect approaches to quantify new establishments, including the self-employment rate (e.g., Gentry and Hubbard 2004), the firm birth rate (e.g., van Stel and Suddle, 2008), and the number of business owners (e.g., Carree et al., 2007). By reviewing previous studies, it is necessary to determine two aspects of the measurement. The first is the subject of observation such as self-employment, start-ups, and business owners. The second is scale of quantification
that can be the absolute number, such as counts or relative terms such as counts per area or per population.

Earlier empirical studies were interested in the people who founded and operated independent businesses. As a result, self-employment rates have been widely suggested as a measurement of new establishments (e.g., Gentry and Hubbard 2004; Lazear 2004). Even though self-employment rates are popular in empirical papers partly because of easier availability, some researchers criticize it as a measurement due to a potential bias towards small-size establishments (Glaeser and Kerr 2009).

Alternatively, the number of new establishments and the amount of new employment within those new firms has been suggested in recent papers (e.g., Kolko and Neumark 2010. These proxies have the advantage of capturing behavioral aspects of start-up activities, such as the unit of decision-making and the contrast between new entries and created jobs from them. On the one hand, a count of new establishments is realistic in terms of the behavioral unit of decision-making as well as start-up activities. It is obvious that a firm is the unit that decides start-up activities and the location choice and not individual employees in the firm. On the other hand, the second metric of employment in the new firms could capture the size of the enterprise and show the number of jobs created. In addition, for a regional economic policy implication, the amount of job creation is one of the key indicators and it is related to other important indexes such as the unemployment rate.
In this study, I consider both metrics, and I normalize them by the size of census tract that is the spatial unit of observation. The following paragraphs discuss the rationale for this normalization.

The first reason is the wide disparity in the sizes of census tracts: they vary from .02 square miles (.05 sq. km or 51,800 sq. m) to 8,689 square miles (22,504 sq. km) with a mean of 14.5 square miles (37.6 sq. km). If I use a count metric for measuring new establishments, larger census tracts will have a higher probability of the counts in proportion to the area of census tract according to the so-called “dartboard theory” (Bartik, 1985). Hence, there is a need to adjust the area effect.

In addition, per-area normalization will give us less biased information on new establishment location choices compared to other common normalization approaches, such as population or housing units. If the unit of geographic analysis is a region, normalization by population can make sense, since a start-up's location choice is based on population within the same region in terms of demand and supply sides. As an example of the supply side of the labor force, most employees of new establishments live in the same region. However, the area of census tract is quite small, thus an employee of new establishments in a tract mostly does not live in the same tract. Hence, if I use a population normalization at the census tract level, there is a mismatch problem between the locations of employees and new establishments. Moreover, population normalization could generate biased information on how land use affects new establishments. For example, there might be fewer chances of start-ups in areas with a high concentration of since less land is available for businesses. Hypothetically, populous area may have small
numbers of new establishments because of a crowding out effect from dominant residential land uses. I will test the existence of this effect later in the model.

4.1.2 MSA level variables

MSA level variables are collected to represent a set of characteristics related to new establishments. Two recent studies (Glaeser, 2007; Glaeser and Kerr, 2009) describe several theories that explain the variation in new establishments across regions. They discuss agglomeration economies, demographics, and natural advantages as key drivers of start-ups at the metro level. This study refers to their theoretical discussion to drive my choice of MSA level variables. In addition, I include some business climate variables to inspect any effects on new establishments.

4.1.2.1 Agglomeration economies variables

Agglomeration economies reflect that “incumbent industrial structures of each city shape the availability and flow of goods, people, and ideas to new ventures” (Glaeser and Kerr, 2009). I create four variables to characterize agglomeration economies at the MSA level: total employment density (MSA_EMPDEN), GDP per capita (GDP), and average commuting minutes (COMM). Initially, additional variables that are commonly used in empirical studies of agglomeration economies (e.g., Combes et al., 2009; Puga,
2010; Behrens et al., 2012) were tested, for instance, population density. But these three variables were selected to avoid multicollinearity\textsuperscript{21}.

*MSA employment density:* MSA\_EMPDEN measures total employment density at the MSA level. The employment numbers are from CTPP 2000, and the scale of density is per square mile. The variable is intended to capture an overall regional clustering effect. Dense areas tend to be more productive in terms of economic performance, and many empirical studies support a positive relationship between density and economic performance, as measured here by new establishments. As reviewed in Chapter Two, previous studies have shown a positive effect on new establishments at the city and regional levels.

*GDP per capita\textsuperscript{22}:* GDP per capita by MSA is collected from the BEA (Bureau of Economic Analysis) regional economic accounts. The BEA estimates real GDP an inflation-adjusted value measured in chained year 2005 U.S. dollars. I use year 2001 data because year 2000 data is not available, but BEA’s estimation uses midyear numbers, so this is not a significant issue. Given a positive correlation between the GDP level and

\textsuperscript{21} For example, population density and employment density are highly correlated.

\textsuperscript{22} Real GDP by metropolitan area is an inflation-adjusted measure of each metropolitan area's gross product that is based on national prices for the goods and services produced within the area. Real GDP by metropolitan area is measured in chained (2005) dollars. Per capita, real GDP by metropolitan area is calculated by dividing the real GDP for a metropolitan area by the resident population of the area. In its calculation, BEA uses the Census Bureau's annual midyear population estimate. Per capita, real GDP indicates the trend in output as it relates to population.
purchasing power, per capita GDP proxies the demand side for new establishments and indicates latent post-startup performance of new establishments. Hence, a positive effect on new establishment is expected.

*Average commuting minutes*: COMM shows the average commuting minutes within an MSA as a proxy for a region’s accessibility. This variable has been included to indicate the speed of moving people in a region that is conducive to business performance. The cost of moving people is still a significant component of the production cost, whereas the cost of moving goods has become relatively cheap. New establishments prefer less commuting time for their employees. In addition, the faster employees are moved, the easier they firm will be able to interact with existing firms and attract potential customers. COMM is expected to have a negative impact on start-ups.

### 4.1.2.2 Demographics variables

Demographics explain both the demand and supply sides of start-ups. On the demand side, an existing population will be potential local customers of the newly created goods and services. This aspect is already considered by using the GDP per capita variable. On the supply side, demographics matter in two ways: First, a straightforward explanation of new establishments is a supply of prospective entrepreneurs. More entrepreneurial people in a region will increase the chances for start-ups. Second, demographics affect the labor pool, which is important because new establishments tend to be labor intensive. Hence, when certain cities have more entrepreneurial people and suitable workers, there would be more active start-ups.
College degree 25 plus: COL25PLUS is the percent with a bachelor’s degree or higher in the population aged 25 years or older. This variable measures the amount of educated human capital, according to the assumption of a positive relationship between workers’ education level and start-up activities. A positive effect on new establishments is expected.

4.1.2.3 Environmental amenities variables

Environmental amenities become increasingly important factors in where new establishments decide to locate. A start-up's success relies partly on the value of human capital (Rotemberg and Saloner, 2000) and increasingly on the quality of life, particularly for young and creative employees. Therefore, amenities have a role to play in attracting a desirable labor force (Cherry and Rickman, 2009)\textsuperscript{23}.

January temperature: Generally, environmental amenities are hard to compare across regions because amenities are qualitative and locally unique. To include a measurable indicator of amenities that can be calculated everywhere, I use the mean January temperature (JANTEMP) between 1971 and 2000. The mean January temperature is commonly known as a strong predictor of area growth as well. Previous studies estimate that the higher the mean January temperature, the more regional growth (e.g. Glaeser, 2011). I expect a high JANTEMP to show a positive effect on new establishments.

\textsuperscript{23} Environmental Amenities and Regional Economic Development (Cherry and Rickman, 2009)
4.1.2.4 Business climate variables

The effects of business climate on new establishments are tested by using the BHI Metro Area Competitiveness index 2002 (Beacon Hill Institute, 2002). I choose the BHI index for the following three reasons. First, its unit of analysis, the MSA level, is exactly same as my model. I can use their variables directly without any adjustment of geographical unit, though the MSA, as compared to city or state, is not the best spatial unit of analysis for business climate as discussed in Chapter Three. Second, their data set of the largest fifty metro areas in the U.S. of year 2000 includes all 27 MSAs of this study. Finally, they use data collected in year 2000 or 2001 to create indices of the report. This also fits into my model since I use data from 2000 for the independent variables and from 2001 for the dependent variables.

BHI created nine normalized indicators with means of five, standard deviations of one, and a range from zero to ten. Each indicator is calculated as the average of component sub-indicators. I use four indicators among nine indicators, because five indicators are similar to some of my existing variables. For example, Human resources in the BHI index is similar to College degree 25 plus (COL25PLUS) of the demographics variables. The following now describes the details of the four indicators that are included in the model to estimate business climate effects.

Technology: This is the average of six sub-indicators; academic R&D funding relative to employment, NIH support to institutions per capita, new patents issued per 100,000 inhabitants, science and engineering graduate students per capita, scientists and engineers as a percent of the labor force, and high-tech payroll as a percent of the total
payroll. Technology should be a critical ingredient of a start-up, if the start-up is innovative like spin-off firms in new industries. However, many new establishments are not innovative, and are simply additions similar to the existing businesses. This suggests that the overall positive effect is expected, but the magnitude of the effect may be small because many new establishments do not require new technologies.

**Finance:** This is the average of three sub-indicators; bank deposits per capita, venture capital as percent of gross metropolitan product, and cost of living. One essential concern for every founder of a new business is how to raise start-up capital. Easy access to business loans and investors are necessary, particularly for small enterprises. This variable is to assess the effect of financial services on start-up activities.

**Competition:** This is the average of three sub-indicators; net firm creation per 100,000 inhabitants, The Cognetics entrepreneurial hot spots index, and the number of new publicly traded companies. The effect of competition on new establishments is twofold: First, as a positive impact, competition fosters entrepreneurial culture, and it increases the likelihood of start-ups. Successful industrial clusters, such as Silicon Valley, are good examples of this. Second, as a negative impact, excessive competition is a barrier for new entries as a market gets saturated. Hence, both a positive and a negative effect are possible.

**Environmental policy:** This is the average of four sub-indicators; electricity prices, toxic release pounds per capita, pollution standards index, serious pollution days per year. In general, environment-related policy and regulation do not attract new businesses. Particularly, pollution intensive manufacturing firms are sensitive to
environmental regulations. However, this environmental policy could have no impact on new establishments, since most new establishments are not related to pollution or environmental issues. When environmental policy reflects the degree of general regulation of public sectors, overall a negative effect is expected.

4.1.3 Census tract level variables

4.1.3.1 Economic environment variables

*Employment density:* CT_EMPDEN measures existing employment density at the census tract level, and is calculated by dividing the employee count of a census tract by the area of the census tract, in acres. This is a representative variable to characterize the existing economic environment that new establishments would have taken as a given when deciding where to open a new establishment within a MSA. In particular, employment density is a common proxy for measuring the degree of agglomeration economies. Hence, CT_EMPDEN will be used to test any effects of agglomeration economies and co-location attractiveness on start-up activities.

* Nearby employment within three miles:* The second variable in the economic environment category is EMP3MI, which measures total employment within three miles of a census tract. This metric has been selected to examine the geographic scope of spillover effects of nearby employment on start-ups’ location choice. Two recent papers argue that spillover effects are no longer effective beyond five miles (Rosenthal and
Strange, 2008) or beyond a half-mile (Arzaghi and Henderson, 2008). However, Rosenthal and Strange include rural areas as well as urban areas. Their focus is not consistent with my concentration on agglomeration economies within urban areas. Arzaghi and Henderson’s conclusion is based on a case involving the advertising industry in south Manhattan in New York City.

This study follows the methodology suggested by Rosenthal and Strange to measure the scope of spillover effects. I first construct a three-mile, about the average of five miles (Rosenthal and Strange, 2008) and half-mile (Arzaghi and Henderson, 2008), ring buffer from each census tract. Assuming employment to be uniformly distributed throughout each census tract, EMP3MI is calculated by constructing a proportional summation of employment for those portions of the census tracts intersected by the ring. For example, if a three-mile radius ring includes all of census tract A and 20% of the area of census tract B, then RING3M equals the employment in tract A plus 20% of the employment in tract B.

*Industrial dominance:* DOMIN measures dominance of industry, calculated by the sum of squared shares of each industry. At the city or regional level, debates about the relationship between industry dominance, specialization (or diversification) strategy and economic performance have been a primary topic in the literature (e.g., Glaeser et al. 1992; Duranton and Puga 2000; Donegan et al. 2008; Drucker 2010; Franz and Hornych 2010). However, few studies have dealt with the issue of industry composition at the micro (ZIP code) level industrial diversity on entrepreneurial activities in an empirical model.

---

24 Rosenthal and Strange (2003) is, to our knowledge, the only paper which measures the effect of
micro or neighborhood level, even though the effect on new establishments is geographically immediate compared to the city level effect. At the macro city level, a specialization or diversification index represents the overall industry mix in a city. But at the census tract level, the mechanism of industry composition works differently. The area of a census tract is quite small, particularly near downtown areas, so when the diversity of industry composition is very small, I tend to observe that the tract is dominated by massive employers such as hospitals, universities, and plants. If that is the case, dominant employers can crowd out new establishments around there. This is why I use the term *dominance* rather than the phrase less *diversity* which is a more common index in city level studies since first suggested by Jacobs (1970).

A simple way of measuring dominance of a census tract is to pay attention to the share of the dominant industry. A census tract could have a single dominant industry, or it could also have two or more dominant industries. First of all, I need to decide how many industries will be included in calculating the dominance variable. Different numbers of top industries included in the calculation of dominance will represent different aspects of the dominance. A more general method of calculating the dominance is to include all fourteen industrial sectors from the data source. One advantage of this method is to be able to conduct a comparative analysis across census tracts.

Another issue is calculating the variable when I include all industries. Since the simple summation of the shares is always 1.0, I suggest using the sum of squared shares.  

---  

25 Many downtown census tracts show approximately 100 acres size which equals about a 2,000ft by 2,000ft or a 600m by 600m square.
Squaring the shares will weigh the more dominant industry. When one or two industries are dominant, it will generate a number that is larger than it would be if all industries were equivalent\textsuperscript{26}. The possible maximum number of $DOMIN$ is 1 when only one industry consists of all employment in the census tract. The minimum case is when all 14 industries have the same share of the industrial composition. In this case, the share of each industry is 7.14\%, and the $DOMIN$ will be $0.071(=14*(0.0714)^2)$. So the range of $DOMIN$ of 14 industries is from 0.071 to 1.0.

*Population/Employment:* POPEMP is a population/employment ratio that quantifies the amount of land that's residential in an area. In a census tract, the higher the ratio, the more land use is residential, rather than commercial/industrial. As this variable is created to measure the degree of residential, commercial and industrial land use, it enables the model to interpret the effect of current land uses on new establishment location choices.

### 4.1.3.2 Demographic environment variables

Where a new establishment locates depends on a demographic environment, which receives less attention in the literature compared to the economic environment. While the economic environment can explain the supply side of new establishments,

\textsuperscript{26} The idea of squared summation is similar to Hirschman-Herfindahl index (HHI), which measures market concentration between competing firms. Refer to Henderson et al. (1995, p. 1075).
demographic characteristics in the vicinity of a census tract can affect the demand side of
new establishments. This is a plausible hypothesis when new establishments target local
customers. However, national or international firms do not consider local demand. Rather,
they may prefer to co-locate with existing firms at regional employment centers. This
suggests that their decisions are either not related or negatively correlated with the
demographic environment at the census tract level. Hence, it is possible for the
demographic environment to have both positive and negative effects on new
establishments. To analyze a mixture of various aspects of population demographics, I
include density, the percentage of non-minority populations, median age, and the
percentage of vacant housing units.

*Population density: CT_POPDEN* is population density calculated by dividing the
population of a census tract by the area of the census tract, in acres. This is an indicator of
the existing demographic environment. For new establishments that respond to local
demand, proximity to potential clients is a common strategy; hence densely populated
areas will be attractive. An example of this case is that a similar distributional pattern
exists between retails and residential developments.

*The percentage of whites: WHITE* is the white population proportion, calculated
by dividing the size of the white population by the total population in the census tract.
This index is generally correlated with income and purchasing power.

*Median age: MEDAGE* is the median age of the population and has been included
to show a picture of what the age distribution looks like. However, in the literature, the
theoretical debate on this variable is not enough to associate it with start-up activities.
The percentage of vacant housing units: VACANT is the percentage of vacant housing units that is calculated by dividing the vacant housing units by the total number of housing units in a census tract. This variable could show the overall status of neighborhood conditions including physical conditions, since higher vacancy rates likely stem from problems in neighborhood management. Furthermore, it also shows whether or not the neighborhood is declining. Hence, for my interest, this variable is another indicator along with WHITE, but high vacancy rates harm a start-up's ability to capture potential demand for new products and services.

4.1.3.3 Geographic environment variables

Distance to nearest highway: NEARHI measures the distance between a census tract and the nearest highway. Accessibility to a highway is important for any new establishment because a company's clients and its employees need to get to and from the business. For the manufacturing sector, traditionally proximity to a highway has been an important factor in the choice of location, which is supported by previous empirical studies.
<table>
<thead>
<tr>
<th>Levels</th>
<th>Categories</th>
<th>Variables</th>
<th>Expected effect</th>
<th>Unit</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Census Tract Level</td>
<td>Economic Environment</td>
<td>CT employment density</td>
<td>+</td>
<td>people/acre</td>
<td>2.814</td>
<td>8.10</td>
<td>0</td>
<td>209.53</td>
</tr>
<tr>
<td>N=8514</td>
<td>Employment within 3miles</td>
<td>+ people ('000)</td>
<td>34.2</td>
<td>38.53</td>
<td>0</td>
<td>235.935</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=8514</td>
<td>Industrial dominance</td>
<td>− index</td>
<td>0.2189</td>
<td>0.17</td>
<td>0.0898</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=8514</td>
<td>Population/Employment</td>
<td>− ratio</td>
<td>6.00</td>
<td>8.79</td>
<td>0</td>
<td>330.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=8514</td>
<td>Demographic Environment</td>
<td>CT population density</td>
<td>+</td>
<td>people/acre</td>
<td>5.475</td>
<td>5.50</td>
<td>0</td>
<td>66.146</td>
</tr>
<tr>
<td>N=8514</td>
<td>Percent of white</td>
<td>+ ratio</td>
<td>0.73</td>
<td>0.27</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=8514</td>
<td>Median age</td>
<td>+/− years old</td>
<td>35.3</td>
<td>0.01</td>
<td>15</td>
<td>87.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=8514</td>
<td>Percent of vacant housing units</td>
<td>− ratio</td>
<td>0.075</td>
<td>0.07</td>
<td>0</td>
<td>0.939</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=8514</td>
<td>Geographic Environment</td>
<td>Distance to nearest highway</td>
<td>−</td>
<td>miles</td>
<td>2.64</td>
<td>4.63</td>
<td>0</td>
<td>154.66</td>
</tr>
<tr>
<td>MSA Level</td>
<td>Agglomeration Economies</td>
<td>MSA employment density</td>
<td>+</td>
<td>people/sq mi</td>
<td>221.67</td>
<td>95.24</td>
<td>17.44</td>
<td>456.71</td>
</tr>
<tr>
<td>N=27</td>
<td>Average commuting minutes</td>
<td>− minutes</td>
<td>24.7</td>
<td>2.12</td>
<td>20.9</td>
<td>28.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=27</td>
<td>GDP per capita</td>
<td>+ dollars</td>
<td>43,000</td>
<td>6,676.61</td>
<td>31,160</td>
<td>61,165</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=27</td>
<td>Business Climate</td>
<td>Technology</td>
<td>+</td>
<td>0-10index</td>
<td>4.76</td>
<td>0.96</td>
<td>3.21</td>
<td>7.46</td>
</tr>
<tr>
<td>N=27</td>
<td>Finance</td>
<td>+ 0-10index</td>
<td>4.87</td>
<td>0.88</td>
<td>3.11</td>
<td>6.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=27</td>
<td>Competition</td>
<td>+/− 0-10index</td>
<td>4.84</td>
<td>1.01</td>
<td>2.89</td>
<td>7.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=27</td>
<td>Environmental policy</td>
<td>− 0-10index</td>
<td>5.03</td>
<td>0.88</td>
<td>2.92</td>
<td>6.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=27</td>
<td>Amenities</td>
<td>January temperature</td>
<td>+</td>
<td>° F</td>
<td>37.73</td>
<td>12.28</td>
<td>20.7</td>
<td>66.2</td>
</tr>
<tr>
<td>N=27</td>
<td>Demographics</td>
<td>Percent of college degree 25plus</td>
<td>+</td>
<td>percent</td>
<td>25.77</td>
<td>4.43</td>
<td>16.4</td>
<td>38.9</td>
</tr>
</tbody>
</table>

Table 4.1 Descriptive statistics

Continued
<table>
<thead>
<tr>
<th>Levels</th>
<th>Categories</th>
<th>Variables</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Census Tract Level</strong></td>
<td>Economic Environment</td>
<td>CT employment density</td>
<td>Census Transportation Planning Package (CTPP) 2000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Employment within 3miles</td>
<td>Census Transportation Planning Package (CTPP) 2000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Industrial dominance</td>
<td>Census Transportation Planning Package (CTPP) 2000</td>
</tr>
<tr>
<td>N=8514</td>
<td>Demographic Environment</td>
<td>CT population density</td>
<td>2000 Census of Population and Housing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percent of white</td>
<td>2000 Census of Population and Housing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Median age</td>
<td>2000 Census of Population and Housing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percent of vacant housing units</td>
<td>2000 Census of Population and Housing</td>
</tr>
<tr>
<td></td>
<td>Geographic Environment</td>
<td>Distance to nearest highway</td>
<td>Created by the author using Esri StreetMap Premium 2010</td>
</tr>
<tr>
<td><strong>MSA Level</strong></td>
<td>Agglomeration Economies</td>
<td>MSA employment density</td>
<td>Census Transportation Planning Package 2000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average commuting minutes</td>
<td>Census Transportation Planning Package 2000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GDP per capita</td>
<td>Regional Economic Information, Bureau of Economic Analysis</td>
</tr>
<tr>
<td>N=27</td>
<td>Business Climate</td>
<td>Technology</td>
<td>BHI Metro Area and State Competitiveness Report 2002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Finance</td>
<td>BHI Metro Area and State Competitiveness Report 2002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Competition</td>
<td>BHI Metro Area and State Competitiveness Report 2002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Environmental policy</td>
<td>BHI Metro Area and State Competitiveness Report 2002</td>
</tr>
<tr>
<td></td>
<td>Amenities</td>
<td>January temperature</td>
<td>Normal Daily Mean Temperature, National Climatic Data Center</td>
</tr>
<tr>
<td></td>
<td>Demographics</td>
<td>Percent of college degree 25plus</td>
<td>2000 Census of Population and Housing</td>
</tr>
</tbody>
</table>
4.2 Selection of study areas: 27 Medium-sized MSAs

This research explores the effects of regional and neighborhood conditions on where new establishments locate in medium-sized U.S. metropolitan statistical areas (MSAs). I define medium-sized metropolitan areas as those with populations of one to two million in the Census 2000. As shown in Table 4.1, 27 MSAs fit this category. They include the largest Cincinnati MSA with a population just below two million (ranked the 23rd largest MSA in the U.S.) down to the smallest, Louisville MSA, with a population just above one million (ranked the 49th largest MSA in the U.S.).

These MSAs behave differently compared to large-sized or small-sized ones when it comes to new establishment location choice. First, a medium-sized MSA represents one set of regional attributes as assumed in the model, while a larger MSA represents multiple sets. Large-sized MSAs are commonly described as having polycentric configurations (Fujita and Ogawa, 1982; Henderson and Becker, 2000; McMillen and Smith, 2003) which stand for several business climates. This polycentric structure means that more than one sub-region exists in the MSA, and it violates my assumption of one regional attribute per a MSA. McMillen and Smith (2003) and Lee (2006) suggest that polycentric structures can be generated when a MSA has a population over two million.

---

27 Henderson (1997) defines and investigates medium-sized cities (population of 50,000 to 0.5 million) first in the literature. Florida (2002) defines medium-sized metro areas with populations of half to one million.
Second, a medium-sized MSA is often considered for new establishments or new branches of national firms as discussed in the previous chapter, while small-size MSAs are often neglected. In general, small-sized regions attract only local businesses. I use a population of one million as the criterion for small-sized MSAs, although there is no objective standard to define small-sized MSAs.

Finally, researchers have paid attention to specific mega metropolitan areas like New York City (Rosenthal and Strange, 2005) and Los Angeles (Hackler, 2003), but medium-sized regions are rarely studied for new establishments. This is unfortunate when I consider the roles of medium-sized metropolitan areas in many states in the U.S. In the World City Network theory, Derudder et al. (2003) says that medium-sized regions have more direct effects on the domestic economy at the state-level than larger metropolitan regions, since large mega-regions are more specialized and closely connected to the international economy, competing with other mega regions outside the U.S. Geographically, large-regions are distributed mostly along the coastal areas, but medium-regions are more uniformly distributed nationally. Accordingly, the 27 MSAs considered here cover all regions of the U.S., and are distributed over 19 states. It is important to cover the national market to estimate variations across regional divisions, for example between New England and Pacific regional effects. By my definition, the top 25 large MSAs are located in 12 states, but the top 26-50 MSAs are distributed over 19 states. Hence, medium-sized MSAs are closely related to the domestic economy at the state level.
According to the U.S. Office of Management and Budget (OMB), there were 258 MSAs, and 18 Consolidated MSA(CMSA)s comprising 73 Primary MSA(PMSA)s. As our model deals with CMSAs as same as MSAs, total number of MSAs are 276. We do not include four MSAs in Puerto Rico.
<table>
<thead>
<tr>
<th>Code</th>
<th>MSA</th>
<th>Population 2000</th>
<th>Area square miles</th>
<th>Counties</th>
<th>Census Tracts</th>
<th>Establishments 2001</th>
</tr>
</thead>
<tbody>
<tr>
<td>1642</td>
<td>Cincinnati, OH</td>
<td>1,979,202</td>
<td>3,866</td>
<td>13</td>
<td>478</td>
<td>3,058</td>
</tr>
<tr>
<td>6922</td>
<td>Sacramento, CA</td>
<td>1,796,857</td>
<td>5,309</td>
<td>4</td>
<td>403</td>
<td>3,566</td>
</tr>
<tr>
<td>3760</td>
<td>Kansas City, KS</td>
<td>1,776,062</td>
<td>5,478</td>
<td>11</td>
<td>500</td>
<td>3,172</td>
</tr>
<tr>
<td>5082</td>
<td>Milwaukee, WI</td>
<td>1,689,572</td>
<td>1,835</td>
<td>5</td>
<td>454</td>
<td>2,002</td>
</tr>
<tr>
<td>5960</td>
<td>Orlando, FL</td>
<td>1,644,561</td>
<td>4,010</td>
<td>4</td>
<td>328</td>
<td>6,348</td>
</tr>
<tr>
<td>3480</td>
<td>Indianapolis, IN</td>
<td>1,607,486</td>
<td>3,541</td>
<td>9</td>
<td>340</td>
<td>2,841</td>
</tr>
<tr>
<td>7240</td>
<td>San Antonio, TX</td>
<td>1,592,383</td>
<td>3,354</td>
<td>4</td>
<td>317</td>
<td>3,078</td>
</tr>
<tr>
<td>5720</td>
<td>Norfolk, VA</td>
<td>1,569,541</td>
<td>2,634</td>
<td>15</td>
<td>364</td>
<td>2,560</td>
</tr>
<tr>
<td>4120</td>
<td>Las Vegas, NV</td>
<td>1,563,282</td>
<td>39,719</td>
<td>3</td>
<td>383</td>
<td>3,525</td>
</tr>
<tr>
<td>1840</td>
<td>Columbus, OH</td>
<td>1,540,157</td>
<td>3,170</td>
<td>6</td>
<td>372</td>
<td>2,850</td>
</tr>
<tr>
<td>1520</td>
<td>Charlotte, NC</td>
<td>1,499,293</td>
<td>3,441</td>
<td>7</td>
<td>301</td>
<td>3,274</td>
</tr>
<tr>
<td>5560</td>
<td>New Orleans, LA</td>
<td>1,337,726</td>
<td>4,492</td>
<td>8</td>
<td>395</td>
<td>2,340</td>
</tr>
<tr>
<td>7160</td>
<td>Salt Lake City, UT</td>
<td>1,333,914</td>
<td>2,101</td>
<td>3</td>
<td>284</td>
<td>6,035</td>
</tr>
<tr>
<td>3120</td>
<td>Greensboro, NC</td>
<td>1,251,509</td>
<td>3,922</td>
<td>8</td>
<td>263</td>
<td>2,454</td>
</tr>
<tr>
<td>640</td>
<td>Austin, TX</td>
<td>1,249,763</td>
<td>4,280</td>
<td>5</td>
<td>256</td>
<td>3,954</td>
</tr>
<tr>
<td>5360</td>
<td>Nashville, TN</td>
<td>1,231,311</td>
<td>4,135</td>
<td>8</td>
<td>248</td>
<td>2,677</td>
</tr>
<tr>
<td>6480</td>
<td>Providence, RI</td>
<td>1,188,613</td>
<td>1,088</td>
<td>5</td>
<td>232</td>
<td>1,757</td>
</tr>
<tr>
<td>6640</td>
<td>Raleigh, NC</td>
<td>1,187,941</td>
<td>3,555</td>
<td>6</td>
<td>211</td>
<td>2,978</td>
</tr>
<tr>
<td>3280</td>
<td>Hartford, CT</td>
<td>1,183,110</td>
<td>1,617</td>
<td>5</td>
<td>282</td>
<td>3,355</td>
</tr>
<tr>
<td>1280</td>
<td>Buffalo, NY</td>
<td>1,170,111</td>
<td>1,573</td>
<td>2</td>
<td>302</td>
<td>1,784</td>
</tr>
<tr>
<td>4920</td>
<td>Memphis, TN</td>
<td>1,135,614</td>
<td>3,098</td>
<td>5</td>
<td>274</td>
<td>1,835</td>
</tr>
<tr>
<td>8960</td>
<td>West Palm Beach, FL</td>
<td>1,131,184</td>
<td>2,219</td>
<td>1</td>
<td>265</td>
<td>6,257</td>
</tr>
<tr>
<td>3600</td>
<td>Jacksonville, FL</td>
<td>1,100,491</td>
<td>2,819</td>
<td>4</td>
<td>197</td>
<td>3,286</td>
</tr>
<tr>
<td>6840</td>
<td>Rochester, NY</td>
<td>1,098,201</td>
<td>3,464</td>
<td>6</td>
<td>268</td>
<td>1,515</td>
</tr>
<tr>
<td>3000</td>
<td>Grand Rapids, MI</td>
<td>1,088,514</td>
<td>2,819</td>
<td>4</td>
<td>225</td>
<td>1,898</td>
</tr>
<tr>
<td>5880</td>
<td>Oklahoma City, OK</td>
<td>1,083,346</td>
<td>4,303</td>
<td>6</td>
<td>331</td>
<td>1,917</td>
</tr>
<tr>
<td>4520</td>
<td>Louisville, KY</td>
<td>1,025,598</td>
<td>2,099</td>
<td>7</td>
<td>241</td>
<td>2,742</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>37,055,342</td>
<td>123,941</td>
<td>164</td>
<td>8,514</td>
<td>83,058</td>
</tr>
</tbody>
</table>

Table 4.3 27 MSA Summary
Figure 4.1 Map of 27 Medium-sized MSAs (Source: ESRI, 2010)
4.3 Building the Database

This section introduces the primary data source of the study and explains how this data has been treated to conduct further analysis. In particular, the reliability of a database and the justification of concentration on certain data will be discussed. Additionally, I will delineate the data processing procedure in order to clarify how the data set has been created for further descriptive and analytical research.

4.3.1 Existing economic conditions: Census Transportation Planning Package

The data Census Transportation Planning Package 2000 (CTPP 2000) is tabulated from answers to the Census 2000 long form questionnaire mailed to one in six U.S. households. Because of the large sample size, the data are reliable and accurate, according to the U.S. Department of Transportation. Originally, the data was designed for transportation planners to understand Journey-to-Work behavior. Therefore, CTPP 2000 is divided into three parts: Place of Residence (origin), Place of Work (destination), and Journey-to-Work flow (origin-destination pair).

This study relies on Part 2 data to figure out the employment distribution at the census tract level, even though the data are also available at the Transportation Analysis Zones (TAZs) level, which is typically smaller than census tracts. However, my analysis requires demographic attributes from the 2000 Census, and these attributes are not available at the TAZ level. More importantly, TAZs are not appropriate for a comparative study across regions because TAZs are defined by Metropolitan Planning Organizations
based on their local transportation conditions. Conversely, the boundaries of census tracts are consistently decided across the nation by the Census Bureau, and therefore are expected to be homogeneous. In addition to this, economic attribute of employees who work there, demographic variables will be included as well at the same census tract level to consider factors related to characteristics of residents who live there.

4.3.2 New business establishments: Dun and Bradstreet database

The second set of data, the Dun and Bradstreet database, contains attributes of individual firms which were established in the year 2001 (See Appendix A for the details). The D&B database is one of few firm-level data sets which cover the entire United States. While there is a firm-level establishment database\textsuperscript{29} from the U.S. Census Bureau, the database is not open to the public because of nondisclosure requirements. Accordingly, except for papers affiliated with the U.S. Census Bureau's Center for Economic Studies (CES), most of the literature conducts firm-level research by using the D&B.

Since the database has been created by a private company, I need to check its reliability. D&B has collected and tracked establishments’ information since 1933\textsuperscript{30}. According to the D&B, they have the quality assurance process of collecting data and

\textsuperscript{29} The Longitudinal Business Database from the Center for Economic Studies (CES), the U.S. Census Bureau. See Jarmin and Miranda (2002) for details.

verifying data from thousands of sources to enhance the accuracy and completeness of the database. Of course, the D&B database is not flawless, and there is missing data which is common for massive databases. Nevertheless, Rosenthal and Strange (2003), after interviewing analysts at the D&B and conducting their research by using the D&B database, support the availability of the database for empirical studies when it comes to the spatial distribution of establishments. They say “the omissions from the data set are sufficiently random that the D&B database is representative of the spatial distribution of establishments in the United States.”

However, there are several limitations of the D&B database according to previous empirical studies. Luger and Koo (2005) compare Dun and Bradstreet’s DMI file and the Employment Services’s unemployment insurance file (ES 202) to examine pros and cons of establishment databases. When a researcher needs to analyze establishments based on detailed firm attributes, using the DMI is necessary, since the ES202 provides only limited firm level data and available only at the aggregated level in many states. However, sometimes the DMI does not reflect an ownership change properly, and thus new establishment information can be biased. Early papers in the 1980’s criticize incomplete establishment information of the DMI file. Verway (1980) criticizes that several new firms were not listed in the DMI file, but the Yellow Page includes more establishment cases in the Harrison village, Michigan in 1975. Aldrich et al. (1989)

31 These sources are supplemented with the information D&B obtains annually through millions of on-site and telephone contacts with business owners and managers.

argues that the DMI file is better than the ES202 file since it provides extensive firm attributes information such as number of employees. However, the DMI file missed several new businesses found in the ES202 file.

As far as the data attributes are concerned, D&B provides the following information about company profiles: establishment year; industry classification (NAICS Codes); the number of employees; annual sales; whether it is a headquarters or branch; and X-Y Coordinates. There are more available attributes in the raw data, but I have only listed the attributes related to my research focus.

4.3.3 Data selection

Since the D&B database covers millions of businesses in the United States, I should identify the research focus so as to conduct an empirical analysis by selecting a data set. This study pays attention to new establishments during the year 2001. The rationale for selecting this year is simple. Eventually, I want to build a model that will be able to clarify the causality between economic environment and start-up activities. To avoid an endogeneity problem, the causal direction should be one way, and the existing economic conditions should be exogenous. As I will use the data sets of Census 2000 as the economic environment, new establishments in the year 2001 can be treated as consequences of conditions in the year 2000. This one year time lag enables us to justify one way causal direction, and there is no reverse causation, in which new establishments would cause the economic environment.
4.3.4 Data processing procedure: the geocoding process

Before turning to further descriptive and analytical research, it is helpful to first understand the data processing procedure. In the framework of the 27 MSA locations in the year 2001, there are more than eighty-thousand new establishments in the D&B database. As my focus is the location choice, I need to first associate the D&B database with spatial references, the so-called geocoding process. As X-Y coordinates are required for the process, I exclude the firms’ data that is missing X-Y coordinates\(^{32}\).

The spatial distribution of new establishments is a type of point pattern. Overall, the distribution does not appear random\(^{33}\). Instead, I observe a clear tendency that shows the number of events is inversely proportional to the distance from the city center; more events happen near the city center. Of course, this tendency is not surprising because almost every event occurring in medium-sized urban areas has a tendency to be monocentric. Another remarkable thing is that there are numerous new establishments in residential areas. This phenomenon indicates that many new establishments begin by using their own house for an office. As this tendency and phenomenon are clearly not accidental, I will discuss what the drivers of location choice behaviors are in next chapters.

\(^{32}\) The percentage of firms that include X-Y coordinates is 96.4% (=80,105/83,058).

\(^{33}\) To do a statistical analysis of the randomness of the spatial points, the hypothesis of Complete Spatial Randomness (CSR) needs to be tested.
CHAPTER 5 SPATIAL DISTRIBUTION OF NEW ESTABLISHMENTS

This chapter investigates spatial distribution of new business establishments. A hypothesis is that the distribution is not random but rather that there are principles of start-up activities for the location choice behavior\(^{34}\). Following sections discuss the results of descriptive analysis at the firm and census tract levels. Understanding the descriptive statistics of individual firms, census tracts, and the relationship between the two are crucial, since the model is built on the interaction between individual firms and census tracts. I estimate location choices of new establishments by utilizing spatial data aggregated in given spatial zones as units of analysis. However, disaggregated firm-level data is also necessary to conduct a refined investigation at the local level. Because a firm is the unit that decides location choice, the firm-level approach is useful to understand the decision-making process of individual establishments. In addition, the census tract level analysis examines how observed patterns of start-ups are associated with the unit of spatial analysis.

\(^{34}\) \(H_0\): new establishment events exhibit complete spatial randomness (CSR), \(H_a\): new establishment events are spatially clustered or dispersed, for example.
5.1 Firm level descriptive analysis

5.1.1 New establishments by industry

This section reports the results of the distribution of new establishments by industry at NAICS code two-digit level (see Table 5.1). New establishments by industry range from the lowest 0.1% of Utilities (NAICS 22) to the highest 22% of Administrative and Support industry (NAICS 56).

Since most new establishments locate in urban areas and my focus is metropolitan areas, there are very small portions of new establishments from Agriculture (NAICS 11) and Mining (NAICS 21) sectors. Conversely, Administrative and Support (NAICS 56) and Professional, Scientific, and Technical Services (NAICS 54) industries report noticeably high start-ups, in both number and as a share of all new establishments.
<table>
<thead>
<tr>
<th>NAICS Code</th>
<th>Industry</th>
<th>New establishments</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>56</td>
<td>Administrative and support and waste management and remediation services</td>
<td>17,793</td>
<td>22.20%</td>
</tr>
<tr>
<td>54</td>
<td>Professional, Scientific, and Technical services</td>
<td>10,590</td>
<td>13.20%</td>
</tr>
<tr>
<td>23</td>
<td>Construction</td>
<td>8,411</td>
<td>10.50%</td>
</tr>
<tr>
<td>81</td>
<td>Other services (except Public Administration)</td>
<td>7,139</td>
<td>8.90%</td>
</tr>
<tr>
<td>62</td>
<td>Health care and Social assistance</td>
<td>7,093</td>
<td>8.80%</td>
</tr>
<tr>
<td></td>
<td><strong>Sum</strong></td>
<td><strong>51,026</strong></td>
<td><strong>63.60%</strong></td>
</tr>
</tbody>
</table>

Table 5.1 New establishments by industry (Top 5)
5.1.2 New establishments by size

The firm size, measured by the number of employees, is the most important attribute of individual firms. First, it represents the internal structure or organization of
start-up firms, so its effect on location choice behavior is significant. For example, when the firm size is very small, like one or two, the decision-making process should be simpler than for larger firms. Second, the smaller the size of the firm, the smaller the legal burden. Size also affects the amount of available tax subsidies from the government, financing opportunities from banks, and investment costs. Recently, the number of self-employment and SOHO (small office/home office) increases, and increasingly there are many types of financial supports from the public sector, especially federal taxation. The 2010 Tax Relief Act\textsuperscript{35} for self-employment is one example of this policy.

The firm size distribution (FSD) of the sample shows a typical FSD case. The dominance of small start-up firms is a well known fact in empirical studies of new firm formation. In general, “the size distribution for very young firms is highly concentrated in small values and is far more skewed than the overall firm size distribution” (Cabral and Mata, 2003, p. 1077). Since the new establishments of this study are one year old of very young firms, the graph of Figure 5.2 shows exactly the expected distributional pattern.

One interesting result of the distribution is that a firm size of two employees has the largest share, 41.3%. This tendency is equivalent across all 27 MSA case study areas. Although this is an interesting observation, unfortunately I do not have any theory to explain this phenomenon. When I cluster sizes by small (1-3), medium (4-10), and large size groups (more than 10), roughly the small-sized group shares 75%, the medium-sized group shares 15%, and the large-sized group shares 10%.

\textsuperscript{35} The 2010 Tax Relief Act reduced the self-employment tax by 2% for self-employment income earned in calendar year 2011.
The firm size distribution needs to be further discussed by referring to a relationship with industry classification. Whereas Professional and Technical service (NAICS 54) and Administrative and Support (NAICS 56) are likely in the small-sized group, Manufacturing (NAICS 31-34) and Transportation and Warehousing (NAICS 48-49) are expected to be in the large-sized group.

![New establishments by size](image)

**Figure 5.2 New establishments by size**

### 5.1.3 New establishments by organizational status

The Dun and Bradstreet database, the data source for firm level information, contains one essential attribute that is organization status. The attribute is expected to affect where new establishments locate, as I discussed in the section 3.2. Large national
firms (Headquarters in Table 5.2) behave differently than small local (Single Location in Table 5.2) when it comes to where they locate. Whereas big firms or national firms depend on national or international markets rather than the local market, small local firms entirely rely on local consumers. As a result, the geographical distribution of small local firms likely follows the population distribution of a region because of their need for proximity to local demand. For large firms, consideration of easier accessibility to other regions results in proximity to highways or the airport. When face-to-face interaction with peer existing companies or related government authorities is a significant matter for their business operation, they prefer to locate at the central location of a region like downtowns or suburban employment centers.

Table 5.2 reports a similar pattern across 27 MSA that shows about 1% of Headquarters compared to about 99% of Single Location establishments. The number of Headquarters ranges from the smallest, 7 new establishments of Rochester MSA, to the largest, 42 new establishments of Orlando MSA. When I look at the frequency in a relative term of the percentage, Louisville MSA and Hartford MSA report the lowest 0.3% and Oklahoma City MSA reports the highest 1.4%.
| MSA       | Total | Headquarters | Single Location | | % | % |
|-----------|-------|--------------|-----------------|---|---|
| Austin    | 3,857 | 26           | 3,831           | 99.3% | |
| Buffalo   | 1,747 | 12           | 1,735           | 99.3% | |
| Charlotte | 3,176 | 28           | 3,148           | 99.1% | |
| Cincinnati| 3,003 | 23           | 2,980           | 99.2% | |
| Columbus  | 2,761 | 19           | 2,742           | 99.3% | |
| Grand Rapids | 1,868 | 12           | 1,856           | 99.4% | |
| Greensboro| 2,382 | 24           | 2,358           | 99.0% | |
| Hartford  | 2,477 | 8            | 2,469           | 99.7% | |
| Indianapolis | 2,768 | 34           | 2,734           | 98.8% | |
| Jacksonvill | 3,217 | 17           | 3,200           | 99.5% | |
| Kansas City | 3,100 | 34           | 3,066           | 98.9% | |
| Las Vegas | 3,353 | 29           | 3,324           | 99.1% | |
| Louisville| 2,690 | 9            | 2,681           | 99.7% | |
| Memphis   | 1,792 | 11           | 1,781           | 99.4% | |
| Milwaukee | 1,939 | 26           | 1,913           | 98.7% | |
| Nashville | 2,592 | 25           | 2,567           | 99.0% | |
| New Orleans | 2,285 | 11           | 2,274           | 99.5% | |
| Norfolk   | 2,492 | 20           | 2,472           | 99.2% | |
| Oklahoma City | 1,960 | 27           | 1,933           | 98.6% | |
| Orlando   | 6,143 | 42           | 6,101           | 99.3% | |
| Providence| 1,722 | 10           | 1,712           | 99.4% | |
| Raleigh   | 2,893 | 22           | 2,871           | 99.2% | |
| Rochester | 1,489 | 7            | 1,482           | 99.5% | |
| Sacramento| 3,473 | 19           | 3,454           | 99.5% | |
| Salt Lake | 5,917 | 26           | 5,891           | 99.6% | |
| San Antonio | 3,009 | 25           | 2,984           | 99.2% | |
| West Palm | 6,141 | 36           | 6,105           | 99.4% | |
| Total     | 80,246| 582          | 79,664          | 99.3% | |

Table 5.2 New establishments by organizational status
5.2 Census tract level descriptive analysis

A census tract is the spatial unit of the current study, and hence understanding overall spatial patterns at the census tract level is crucial to model the location choice behavior of new establishments. Using the census tract for a neighborhood scale analysis is unavoidable because consistent data, such as Census results, is available through census tracts in the US. The concept of neighborhood is originally based on demographics instead of local economic characteristics that closely associate with business establishments.

Figure 5.3 (a) shows an overall relationship between population 2000 and census tract size. Even after excluding outlier tracts (See the 99.5% N=8471 sample case), census tract sizes show a wide variation. This wide disparity in size requires normalization per square mile to count the number of new establishments when I model the location choice behavior. When I compare distributions of (a) population and (b) employment, I could note that population distribution presents more scattered patterns. This phenomenon can be explained by our common observations around any urban settings. Employment distributions are geographically clustered and the intense case is in a downtown or central business district. On the other hand, residential areas are everywhere even though the density could be dissimilar. Figure 5.3 (a) represents this tendency clearly.

Figure 5.3 (c) describes the relationship between the census tract size and number of new establishments. In addition, Figure 5.3 (d) shows a comparable relationship
(N=8514. Min=0.02, Max=8689.1, Mean=14.5 SqMi)  
**a) Population 2000 and Tract Area**

(99.5% N=8471. Min=0.02, Max=207.5, Mean=9.1 SqMi)  
**b) Employment 2000 and Tract Area**

Figure 5.3 Census tract size distributions  
Continued
Figure 5.3 Continued
(N=8514. Min=0.02, Max=8689.1, Mean=14.5 SqMi)
(99.5% N=8471. Min=0.02, Max=207.5, Mean=9.1 SqMi)

(c) New establishments 2001 and Tract Area

(d) Establishment Employment 2001 and Tract Area
between the census tract size and number of employees that come from new establishments. While the number of new establishments ranges from zero to about one hundred, the number of employees ranges from zero to about three thousand. Both cases show a similar negative relationship between the size and the numbers. Overall, smaller census tracts show a large number of start-ups and more employees. In general, the area of census tracts is proportionally correlated with the distance to the center of a city. While downtown census tracts are quite small with less than five downtown blocks, suburban tracts are vast because of low density development patterns. Hence, the results of Figure 5.3 (c) and (d) is consistent with a theory in the literature that denser areas are more attractive to new establishments, which is a principal idea of agglomeration economies.
CHAPTER 6 THE TWO-LEVEL MODEL AND RESULTS

This chapter presents and discusses the model results to clarify the key factors driving where new establishments locate at the MSA level and at the census tract level. The first section introduces the establishment model and the employment model to validate the purpose of using the two models in the framework of the two-level approach. The second section discusses the findings of the establishment model, and the third section contrasts the differences in the key determinants of establishment and employment location. The fourth section synthesizes the results across the models as well as across the geographic levels, seeking to discover the implications of the model for planning practices. The fifth section summarizes and discusses findings of the model and, suggests how to improve the model for future research.

6.1 Overview

I developed two regression models, the establishment model to estimate the number of new establishments, and the employment model to estimate the number of created jobs from the new establishments. A comparison between the two enables us to understand the effect of a firm's size on its location choice, since there is considerable
variation in the size of new establishments. The establishment model uses both Poisson estimation and two-level linear modeling estimation; the employment model employs only two-level linear modeling. Since the purpose of the establishment model is to examine the regional and neighborhood conditions affecting the number of new establishments, a count data model (CDM) is employed in the framework of the two-level model. As reviewed in Chapter Three, a Poisson regression approach is common and appropriate for the establishment model because it quantifies the probability of a start-up’s events on each census tract, which is a spatial unit of analysis. For the employment model, the focus is on the quantity of new jobs that are created from the new establishments in the establishment model. There is a disparity between the amount of jobs small and large firms create. Consider that there are two firms; a small self-employed firm of one employee, and a large firm of one hundred employees. In the case of the establishment model, they both are just one establishment and they are the same in terms of quantity. In the case of the employment model, the large firm is one hundred times bigger than the self-employed firm. Hence, the distribution of created jobs on each census tract in the employment model is more divergent than the distribution of the number of new establishments in the establishment model. As a result, for the employment model, two-level linear modeling is used to estimate the conditions affecting the amount of created jobs from new establishments. In addition, for the establishment model, both two-level linear modeling and Poisson estimations are used. By using two-level linear modeling estimations for both the establishment and the employment models, I compare the results of the two models. The model results are directly comparable when
the two models employ the same method. See Appendix B for details of model specification.

6.2 The establishment model

The results of the two-level Poisson estimations for the establishment model are presented in Table 6.1. Census tract level results are first introduced, and then MSA level results are discussed. Census tract level findings are a more significant contribution of to the current literature, since there have been a very small number of empirical studies accomplished at the neighborhood level compared to at the city or regional levels.

6.2.1 Census tract level results

Overall, census tract level variables are significant at the 1% levels. The primary finding is that the effects of the demographic environment as well as the economic environment on new establishments are significant. In the literature on entrepreneurship and start-up location choice, the importance of economic factors such as employment density is repeatedly emphasized, according to the agglomeration economies. However, demographic factors are rarely studied at the census tract or neighborhood level. Therefore, the causal relationship between demographic conditions and new establishments is examined in more detail, after reviewing economic conditions first. Finally, the geographic environment variable, the distance to nearest highway, is discussed.
| Level | Category                | Variable                        | Coefficient | T value | Pr>|t| |
|-------|-------------------------|---------------------------------|-------------|---------|-----|
|       | **Intercept**           | 0.687***                        | 4.681       | <0.001  |
|       | Economic Environment    | CT employment density           | 0.015***    | 12.351  | <0.001|
|       |                         | Employment within 3miles        | 0.006***    | 8.924   | <0.001|
| Census|                         | Industrial dominance            | -0.554***   | -3.279  | 0.001 |
| Tract |                         | Population/Employment           | -0.029***   | -6.584  | <0.001|
| Level | Demographic Environment | CT population density           | 0.053***    | 10.061  | <0.001|
|       |                         | Percent of white                | 0.495***    | 3.233   | 0.002 |
|       |                         | Median age                      | 0.015***    | 3.536   | 0.001 |
|       |                         | Percent of vacant housing units  | -1.015***   | -3.750  | <0.001|
|       | Geographic Environment  | Distance to nearest highway     | -0.068 **   | -7.887  | <0.001|
|       | Agglomeration Economies | MSA employment density          | 0.002***    | 2.960   | 0.009 |
| MSA   |                         | Average commuting minutes       | -0.109***   | -3.202  | 0.006 |
| Level | Business Climate        | Finance                         | -0.039      | -0.681  | 0.505 |
|       |                         | Competition                     | 0.117**     | 2.616   | 0.018 |
|       |                         | Environmental policy            | -0.205**    | -2.551  | 0.021 |
|       | Amenities               | January temperature             | 0.030***    | 4.431   | <0.001|
|       | Demographics            | Percent of college degree 25 plus| 0.037*      | 1.848   | 0.082 |

The symbols ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 6.1 The results of the establishment model: The two-level Poisson estimation
6.2.1.1 Economic environment

Factors explaining the economic environment include employment density, employment within three miles, industrial dominance, and population/employment ratio. Two positive effects and two negative effects are estimated on new establishments. The positive directions of the coefficients include employment density and population within three miles, confirming the existence of the cluster effect within each census tract and within three miles. These results of the positive cluster effects are consistent with existing empirical findings. The negative directions of the coefficients include the dominance index and the population/employment ratio, explaining the effect of the industry mix and land use. As these two variables are newly added by this study, I pay more attention to interpreting the meanings of these two variables.

Employment density and employment within three miles: As expected, higher existing employment density (EMP_DENS) and higher employment within three miles (RING3M) attract more arrivals of new firms. The smaller coefficient (0.006) of RING3M than employment density (0.015) confirms the existence of the spillover effect of employment density as well as the attenuation of the spillover effect. It means that new establishments are affected by the locations of incumbent firms within the census tract as well as the locations of existing firms nearby up to three miles from the census tract. Because the cluster effect comes from geographical proximity, the magnitude of the

---

36 Rosenthal and Strange (2003) considers the industrial composition at the ZIP code level, but the focus is different than ours.
cluster effect attenuates when the distance between existing firms and newly added start-ups increases. However, this model cannot estimate the scope of the spillover effect, since I only employ a three-mile ring variable. When I include additional distance ring variables, the scope of this spillover effect can be measured more precisely.

A one unit increase of employment density (the number of employment per acre) increases the log number of new establishments per square miles by .015, which means 1.5 percent increase. The RING3M coefficient .006 indicates that a thousand more people employed within three miles in the year 2000 creates .6 percent more establishments in the year 2001. One example of this is the spin-off establishments that prefer the proximity to the parent's firm, in which a bigger influential company contributes to create more numbers of spin-off businesses like Silicon Valley (Kerr and Kominers, 2012).

*Industrial dominance:* The DOMIN variable that quantifies industry mix, which is higher when one or two industries are dominant at the census tract level, has a negative effect on the number of new establishments. It can be interpreted as start-up businesses likely locate away from areas occupied by dominant industries. When dominant firms and vertically related firms are co-located, new establishments that are irrelevant to the dominant firms have been crowded out and excluded from the geographic integration. This result relates to supply-chain externalities (Renski, 2011) by locating near interconnected industries to reduce the cost of obtaining resources or shipping goods. This finding reveals a different aspect of the industry mix effect in case at the micro scale

---

*Benchmark Input-Output Data* by the Bureau of Economic Analysis (BEA) is an example of interconnected industries.
geographical unit such as at the neighborhood level. For example, if there is a massive manufacturing company in a certain neighborhood, less relevant new establishments are not likely to locate nearby. That is because firms that are closely related, in terms of their similarity in industry or their supply-chain relationship with the existing firm, are more likely to locate close to the existing dominant firm. If I have additional variables, such as very detailed industry classifications of existing and new establishments, I can examine how a link between the dominant industry and supply-chain externalities affects a start-up's location choice.\textsuperscript{38} It will verify the mechanism of the relationship between the dominant industry and the start-up location choice. In addition, there is another possible reason for the dominant industry's effect on new establishments. Because the size of the census tract is small, there are not enough available sites for new establishments amid large dominant businesses. A good example of this is the university campus census tract that is already crowded with university buildings and little space is available for new establishments.

*Population/employment ratio:* The last variable in the economic environment category is the population/employment ratio that quantifies the level of land uses for residential areas. A higher number indicates that the census tract has more residential than commercial or industrial development. I create this variable to estimate the land use effect on new establishments, since the actual land use or zoning variables are not easily collectable at the micro scale level. It provides the information whether new

\textsuperscript{38}An example study is “The mechanisms of agglomeration: Evidence from the effect of inter-industry relations on the location of new firms” (Jofre-Monseny et al., 2011).
establishments prefer residential areas for proximity to customers, or commercial/industrial areas for a closer relationship with existing businesses. This is a way to compare the demand side and supply side of new establishments. Therefore, the negative relationship between the population/employment ratio and the location choice reveals that new establishments prefer existing commercial/industrial areas to residential areas. However, this result does not mean that new establishments ignore residential areas nearby population density has a positive effect that will be examined in the next section in the results of the demographic environment. I use three variables to measure employment and population distributions at the census tract level; employment density, population density, and population/employment ratio. I am able to include all three variables because statistically they are not correlated. The inter-relationship of three variables and a comparison between the coefficients of three variables are further discussed when I employ the standardized coefficients in the section 6.4.

### 6.2.1.2 Demographic environment

With regard to the demographic environment, I employ four variables including population density, percent of whites, median age, and percent of vacant housing units. The only variable that has a negative effect is the percent of vacant housing units. Previous studies have explored the importance of the demographic environment as a demand factor of new establishments, such as new services and products. Although the literature supports the significance of the demand factor to attract start-ups in theory, empirical studies testing the existence of the demographic effect at the neighborhood
level are very limited\textsuperscript{39}. Furthermore, this study first introduces the percent of whites, median age, and the percent of vacant housing units, while previous works have considered only the population density among demographic factors.

*Population density:* As expected, population density has a positive effect on a new establishment's location choice. A one unit increase in population density (people per acre) increases the log number of new establishments per square miles by .052, which means a 5.2 percent increase. This result is consistent with both theoretical and empirical findings in the literature. Most new establishments are small businesses, and if they are community-based operations such as mom-and-pop retail establishments, their target consumers are likely population nearby.\textsuperscript{40} Therefore the population density is a consequential predictor of a start-up's location choice. However, there are many new establishments that do not depend on nearby clienteles.

*Percent of whites:* This variable is included to estimate the effect of the racial mix at the neighborhood level has on a new establishment's location choice. In addition to the original measurement of the percent of whites, this variable can be a proxy for other correlated factors such as the average income in the census tract. I originally included income level in the model but later excluded it because of its high correlation with the percent of whites. The home ownership is also a highly correlated variable with the income and the percent of whites. I selected Percent of white instead of income level, since Percent of white improves the two-level model. Using Percent of white increases

\textsuperscript{39} One example is “Are poor neighborhoods “retail deserts”? (Schuetz et al., 2012)”

\textsuperscript{40} The relationship between firm size and catchment area is discussed in the Chapter Five.
the number of significant variables at the MSA level compared to including other correlated variables at the census tract level. This interdependence between the MSA level and the census tract level confirms the importance of a multilevel perspective to analyze location choice of new establishments.

The positive effect of the variable shows new establishments in favor of the neighborhood having more whites, higher income, and more home ownership. The average income in the census tract is likely the most significant factor in attracting new establishments, since typically income is a proxy for purchasing power. Schuetz et al. (2012) focuses on the relationship between neighborhood income and retail density at the zip code level, and finds that high poverty neighborhoods have lower employment density for retail overall. Although Schuetz et al. investigate the neighborhood income effect on the existing establishments instead of my focus new establishments, they demonstrate how demand for retail goods increases with the average income of the neighborhood.

Median age: Although the median age shows a positive effect on new establishments, it is not easy to see why age would have a positive effect. Originally, I included age to test the effect of the age distribution within the metro area on new establishments. The median age can affect new establishments in two ways; as a supply factor and as a demand factor. As a supply factor approach, one possible explanation for the positive effect is that their house is used for office space and older people are more likely to start home-based businesses. Many small-sized establishments that have less than five employees locate in single–family residential areas, and small-sized
establishments make up a large portion of the entire dataset. This is an interesting finding, highlighting a new aspect of a start-up's location choice\textsuperscript{41}, since residential areas are studied as mostly for the residence rather than the work place. As a result, business location theory does not look at residential areas as a possible site for opening up businesses because of the traditional perspective of “Home/Work Separation” dichotomy (Gurstein, 1996). In addition, new businesses are likely initiated by experienced people who have savings and knowledge rather than younger less experienced people. As a demand factor approach, older people are likely to be better clients for new establishments, since they are more financially established. Their expenses for basic needs, such as a home mortgage, are already paid off, and the older age group is more available for new services offered by new establishments.

\textit{Percent of vacant housing units:} The last variable in the demographic environment category is the percent of vacant housing units. The significant negative coefficient of the variable is consistent with an interpretation that new establishments tend to avoid declining neighborhoods for their business locations. A high proportion of vacant housing units is a key indicator of neighborhood decline. Since new businesses are established to satisfy increased demand or expected demand in the near future, the declining neighborhoods are not attractive locations to start a new business. Furthermore,}

\textsuperscript{41} This research is, to our knowledge, the first to pay attention to a new establishment's location choice from the perspective of home-based businesses. There are several papers about home workers in the literature on housing choice. For instance, Moos and Skaburskis (2008) find that home workers occupy larger single-family detached houses for home businesses in Canada.
the variable represents the physical conditions of a neighborhood that generally affects the image of neighborhood, for example. The image of the neighborhood is not quantifiable, but typically a significant factor that affects new firm formations. In general, vacancy rates are negatively correlated with the image of neighborhood and hence can be a proxy for the image.

6.2.1.3 Geographic environment

The geographic environment is generally the most important and noteworthy factor for any location choices and new establishments are no exception. The main focus of the geographic environment is accessibility to new establishments, therefore I employ the distance to the nearest highway to quantify overall accessibility. Initially, another variable of accessibility, the distance to the downtown of the central city, is considered as well, but it turns out to be highly correlated with the distance to the nearest highway. For medium-sized metropolitan areas in the U.S., the density of highways near downtown areas is high. Hence the two variables are highly correlated.

As expected, the coefficient on distance to the nearest highway is negative and significant at the 1% level, which means highway accessibility is an important factor for location choice of new establishment. This phenomenon can be interpreted in two ways. First, the importance of physical interaction with a clientele still exists although internet-based transactions are widely available. Second, face-to-face contacts with co-workers and other related businesses are necessary for start-up businesses (Berg et al., 2010). That is especially true in the beginning stage of new establishments, such as in their first year.
when interactions with other related businesses are frequent and essential for the founders to learn and apply what other successful businesses do.

6.2.2 MSA level results

By including the MSA level variables, the two-level model, in the framework of the multilevel regression, estimates weighted random coefficients of regional attributes to control regional variances. This is one advantage of the multilevel approach compared to general regressions that use regional dummies to control regional variance. As a result, the MSA level factors are not just control variables in the model. Rather, estimated coefficients reveal the regional business environment for start-up activities. However, 27 regional observations are not large enough to estimate regional attributes statistically significant. Because the regional observations are limited to 27 cases, the overall level of significance at the MSA level is lower than the level at the census tract level.

At the MSA level, four categories of variables are included; agglomeration economies, business climate, amenities, and regional demographics. Among nine explanatory variables, MSA employment density, Average commuting minutes, and January temperature are significant at the 1% levels, three business climate variables are significant at the 5% levels, and percent of college degrees 25 plus is significant at the 10% level.

Agglomeration economies: Agglomeration economies variables are used to estimate the regional clustering effect; these are commonly used variables across regions. MSA employment density has a positive coefficient as expected. Since employment
density also has a positive effect at the census tract level, positive coefficients of both levels confirm that a more compact development pattern, in terms of employment distribution, is conducive to attracting new establishments locally as well as regionally (Lee and Gordon, 2011).

The second statistically significant variable in the agglomeration economies is average commuting minutes. The longer commuting times, the less likely a new establishment is to locate in the MSA. This variable can be considered as an indicator of the overall traffic conditions of each MSA. Better transportation conditions can encourage moving people and more face-to-face meetings within a MSA, for example, between the down and suburban employment centers. The results suggest that new establishments tend to avoid the environments with longer commuting times and congested traffic conditions.

*Business climate:* Many start-up research papers and policy makers emphasize the importance business climate to be a better place for new businesses. However, in most cases it is hard to quantify business climate, so index indicators or ranking systems are widely used to compare business environments across regions. I also use the BHI business climate index to include measurable variables at the MSA level.

Among four business climate variables, Technology, Competition, and Environmental policy, all show significant coefficients at the 5% level. Finance is not statistically significant, although financing initial investment is one barrier for every start-up to overcome. The technology variable has a negative effect that is the opposite of what's expected. In general, a higher level of technology within a region is regarded as a
better condition to stimulate entrepreneurial activities (Bade and Nerlinger, 2000). However, not every establishment is entrepreneurial or innovative. Among the data sets, the most frequent industrial sector is *Administrative and support and waste management and remediation services* (NAICS 56) that account for 22.2% of the entire new establishments, in which technology is not really necessary to start a new business. Conversely, the technology-based industry of *Professional, Scientific, and Technical Services* (NAICS 54) shows only 13.2%. Thus, the level of technology is not a necessary condition for most new establishments because they are not based on advanced new technologies. One explanation for the expected positive sign is that most start-up studies are focused on new industries such as computer and IT technologies. This focused industry analysis overestimates the importance of technology to start a new business.

*Competition and Environmental Policy:* New establishments increase in a competitive business climate, which is consistent with the findings of previous empirical studies that argue competition between firms is necessary to nurture start-up activities. It also happens the other way around, in which new establishments increase competition and therefore overall enhanced competitiveness is anticipated.

As expected, environmental policies have a negative influence on new establishments. However, most new establishments are not in pollution intensive industries. Only 3.3 percent of the start-ups in 2001 were from the manufacturing sector (NAICS 31-34). If the environmental policy indicates the degree of the overall public sector regulations, this negative effect is likely relevant for other industries. In general, tighter regulations will reduce new establishments.
New establishments increase with the average January temperature that measures regional natural amenities as expected. This finding exemplifies the significance of the warm winter effect on new establishments as well, since the January temperature is one of the strongest predictors of regional growth in empirical studies in the U.S. As a result, I expect a positive relationship between new establishments and regional growth within 27 MSAs, although the relationship between two is not the main focus of this study.

In the final category at the MSA level of Demographics, new establishments are positively associated with the MSA education level. It may tell us that highly educated people tend to be more entrepreneurial than the less educated. In addition to the entrepreneurs’ perspective, the education level could characterize an aspect of labor pooling. When the education level indicates the level of skills in line with research by Florida et al. (2011) and Glaeser and Resseger (2010), start-up businesses prefer well trained workers to novices. New establishments try to reduce any risks of entrepreneurial activities by having skilled workers. Since the education level is also associated with the level of income, higher incomes will increase the number of new establishments because of the increase in purchasing power for novel goods from new establishments. However, the income level on the demand side is already controlled by GDP per capita variable, although the GDP per capita is not statistically significant in the model.
<table>
<thead>
<tr>
<th>Level</th>
<th>Category</th>
<th>Variable</th>
<th>Establishment Model</th>
<th>Employment Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Coefficient</td>
<td>Pr&gt;</td>
<td>t</td>
</tr>
<tr>
<td>Census</td>
<td>Economic</td>
<td>Intercept</td>
<td>-5.008***</td>
<td>0.001</td>
</tr>
<tr>
<td>Tract</td>
<td>Environment</td>
<td>CT employment density</td>
<td>0.622***</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Level</td>
<td>Economic</td>
<td>Employment within 3miles</td>
<td>0.040***</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Demographic</td>
<td>Environment</td>
<td>Industrial dominance</td>
<td>-5.662***</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Level</td>
<td>Economic</td>
<td>Population/Employment</td>
<td>-0.027**</td>
<td>0.013</td>
</tr>
<tr>
<td>Geographic</td>
<td></td>
<td>CT population density</td>
<td>0.405***</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>MSA</td>
<td>Agglomeration</td>
<td>Percent of white</td>
<td>3.745***</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Level</td>
<td>Economies</td>
<td>Median age</td>
<td>0.153***</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percent of vacant housing units</td>
<td>-6.500***</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Geographic</td>
<td>Distance to nearest highway</td>
<td>-0.043**</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSA employment density</td>
<td>0.010*</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average commuting minutes</td>
<td>-0.885**</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GDP per capita</td>
<td>0.000</td>
<td>0.126</td>
</tr>
<tr>
<td></td>
<td>Technology</td>
<td>-1.734**</td>
<td>0.037</td>
<td>-5.661***</td>
</tr>
<tr>
<td></td>
<td>Finance</td>
<td>-0.327</td>
<td>0.483</td>
<td>-1.638</td>
</tr>
<tr>
<td></td>
<td>Competition</td>
<td>0.973**</td>
<td>0.046</td>
<td>4.175***</td>
</tr>
<tr>
<td></td>
<td>Environmental policy</td>
<td>-1.511*</td>
<td>0.078</td>
<td>-4.343*</td>
</tr>
<tr>
<td></td>
<td>Amenities</td>
<td>January temperature</td>
<td>0.217***</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>Demographics</td>
<td>Percent of college degree 25</td>
<td>0.273</td>
<td>0.156</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2*LogLikelihood</td>
<td>55836.076</td>
<td>99017.595</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Census Tract Levelvariance (σ rij)</td>
<td>40.59 (90.0%)</td>
<td>6584.70 (99.5%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSA Levelvariance (u 0)</td>
<td>4.51 (10.0%)</td>
<td>31.52 (0.5%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ICC</td>
<td>10.0%</td>
<td>0.7%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AIC</td>
<td>55840.1</td>
<td>99075.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BIC</td>
<td>55854.2</td>
<td>99089.2</td>
</tr>
</tbody>
</table>

Table 6.2 Establishment and Employment Model: The two-level linear modeling estimation (The symbols ***,**, and * denote significance at the 1%, 5%, and 10% levels, respectively. ICC : Intraclass Coefficient, AIC: Akaike’s Information Criterion, BIC: Schwarz’s Bayesian Criterion)
6.3 A comparison between the establishment model and the employment model

This section contrasts the meanings and the effects of determinants between the employment model and the establishment model. As there is considerable variation in the size of new establishments, the employment model is necessary to capture the firm size effect on new establishments. Furthermore, the amount of created jobs from new establishments is the direct and immediate consequence of new establishments, and one important indicator of regional economic growth. Thus, another purpose of the employment model is to focus on the relationship between the quantity of the created jobs and a start-up's location choice.

Table 6.2 contrasts the estimated coefficients of determinants between the establishment model and the employment model. Since I find similar results across two models and the establishment model results are already discussed, I look at the variables presenting different results using the employment model.

First at the census tract level, the Population/employment ratio and Population density are not predictors of neighborhood employment, although they do predict the number of new establishments. Second at the MSA level, Total employment density is not a predictor of created jobs from new establishments, however it is associated with an increased number of start-ups.

The fundamental reason for these different effects between two modes lies in the firm's size which affects the number of employees as discussed in the section 6.1. The larger firms with many employees tend to behave differently in location choice decisions...
compared to small firms with one and two employees. The employment considers the firm size effect, while the establishment does not.

Population density and the population/employment ratio do not influence the neighborhood employment because the large firms, which have a stronger impact than small firms in the employment model, are not affected by neighborhood population. In general, smaller size establishments are likely more dominant in populated residential areas, while larger size start-ups tend to locate in less residential areas, and instead preferring commercial or industrial areas. Small firms are more community-demand based businesses, so proximity to clients is important. Conversely, large firms are more regional-demand based or inter-business operations within supply chain relationships between firms. Thus, they prefer to locate at clustered employment areas such as downtown or suburban employment centers, and these areas are not populated areas.

If small firms are more community-demand based businesses and large firms are more regional-demand based, this finding is consistent with my hypothesis that more small size establishments locate in residential areas. To clarify these variations in location choices depending on the firm size, future research can use the employment data on a firm to examine the impact of a firm's size on its location choice.
<table>
<thead>
<tr>
<th>Level</th>
<th>Category</th>
<th>Variable</th>
<th>Establishment</th>
<th>Establishment</th>
<th>Employment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Poisson</td>
<td>Two-level linear</td>
<td>Two-level linear</td>
</tr>
<tr>
<td>Census Tract</td>
<td>Economic Environment</td>
<td>CT employment density</td>
<td>+ (*** )</td>
<td>+ (*** )</td>
<td>+ (*** )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Employment within 3 miles</td>
<td>+ (*** )</td>
<td>+ (*** )</td>
<td>+ (*** )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Industrial dominance</td>
<td>− (*** )</td>
<td>− (*** )</td>
<td>− (*** )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Population/Employment</td>
<td>− (*** )</td>
<td>− (** )</td>
<td>− (** )</td>
</tr>
<tr>
<td>Level</td>
<td>Demographic Environment</td>
<td>CT population density</td>
<td>+ (*** )</td>
<td>+ (*** )</td>
<td>+ (*** )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percent of white</td>
<td>+ (*** )</td>
<td>+ (*** )</td>
<td>+ (*** )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Median age</td>
<td>+ (*** )</td>
<td>+ (*** )</td>
<td>+ (*** )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percent of vacant housing units</td>
<td>− (*** )</td>
<td>− (*** )</td>
<td>− (*** )</td>
</tr>
<tr>
<td>Geographic Env.</td>
<td>Distance to nearest highway</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSA Level</td>
<td>Agglomeration Economies</td>
<td>MSA employment density</td>
<td>+ (*** )</td>
<td>+ (*)</td>
<td>+ (*** )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average commuting minutes</td>
<td>− (*** )</td>
<td>− (** )</td>
<td>− (** )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GDP per capita</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business Climate</td>
<td>Technology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Finance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Competition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Environmental policy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amenities</td>
<td>January temperature</td>
<td>+ (*** )</td>
<td>+ (*** )</td>
<td>+ (*** )</td>
<td></td>
</tr>
<tr>
<td>Demographics</td>
<td>Percent of college degree 25 plus</td>
<td>+ (*)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The symbols ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 6.3 Comparison of results across models
<table>
<thead>
<tr>
<th>Category</th>
<th>Variable</th>
<th>Establishment Model Two-level linear</th>
<th>Employment Model Two-level linear</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standardized Coefficient (1)</td>
<td>Coefficient (2)</td>
<td>Standardized Coefficient (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4)</td>
<td>(3)</td>
</tr>
<tr>
<td>Economic</td>
<td>CT employment density</td>
<td>0.525***</td>
<td>0.498***</td>
</tr>
<tr>
<td></td>
<td>Employment within 3miles</td>
<td>0.160***</td>
<td>0.097***</td>
</tr>
<tr>
<td></td>
<td>Industrial dominance</td>
<td>-0.069***</td>
<td>-0.037***</td>
</tr>
<tr>
<td></td>
<td>Population/Employment</td>
<td>-0.025**</td>
<td>-0.016</td>
</tr>
<tr>
<td>Demographic</td>
<td>CT population density</td>
<td>0.232***</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>Percent of white</td>
<td>0.105***</td>
<td>0.042***</td>
</tr>
<tr>
<td></td>
<td>Median age</td>
<td>0.106***</td>
<td>0.044***</td>
</tr>
<tr>
<td></td>
<td>Percent of vacant housing</td>
<td>-0.046***</td>
<td>-0.015**</td>
</tr>
<tr>
<td>Geographic</td>
<td>Distance to nearest highway</td>
<td>-0.020**</td>
<td>-0.334**</td>
</tr>
</tbody>
</table>

Table 6.4 Standardized Coefficients: Census tract level results

<table>
<thead>
<tr>
<th>Category</th>
<th>Variable</th>
<th>Establishment Model Two-level linear</th>
<th>Employment Model Two-level linear</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standardized Coefficient (1)</td>
<td>Coefficient (2)</td>
<td>Standardized Coefficient (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4)</td>
<td>(3)</td>
</tr>
<tr>
<td>Agglomeration</td>
<td>MSA employment density</td>
<td>0.099*</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>Average commuting minutes</td>
<td>-0.196**</td>
<td>-0.059**</td>
</tr>
<tr>
<td></td>
<td>GDP per capita</td>
<td>0.058</td>
<td>0.009</td>
</tr>
<tr>
<td>Business</td>
<td>Technology</td>
<td>-0.174**</td>
<td>-0.056**</td>
</tr>
<tr>
<td></td>
<td>Finance</td>
<td>-0.030</td>
<td>-0.015</td>
</tr>
<tr>
<td></td>
<td>Competition</td>
<td>0.102**</td>
<td>0.043**</td>
</tr>
<tr>
<td></td>
<td>Environmental policy</td>
<td>-0.139*</td>
<td>-0.039*</td>
</tr>
<tr>
<td>Amenities</td>
<td>January temperature</td>
<td>0.277***</td>
<td>0.083***</td>
</tr>
<tr>
<td>Demographics</td>
<td>Percent of college degree 25</td>
<td>0.126</td>
<td>0.042</td>
</tr>
</tbody>
</table>

Table 6.5 Standardized Coefficients: MSA level results
6.4 A comparison of results across models

One advantage of using the multilevel model is the capability of comparing variance across levels, and this comparison leads to identifying whether the effects at the MSA level or at the census level have more of an influence on where new establishments locate. The intra-class correlation coefficient (ICC) shown in Table 6.2 provides a comparison between the MSA variance and the census tract variance. The ICC ($\rho$) measures the proportion of variance at the MSA level.

The ICC of 10% in the establishment model indicates that 10% of variance comes from MSA level effects, whereas the census tract variance is 90%. The ICC in the employment model is much smaller 0.7%, because between MSA variance is 0.5% and between census tract variance is 99.5%. In other words, new establishment location choice is mainly affected by the census tract level neighborhood effects (90% and 99.5%) rather than the MSA level regional effects (10% and 0.5%).

In addition, I find that the MSA level accounts for less of the variance in the employment model (0.7%) than the establishment model (10%). Hence, the MSA level factors do matter, but have limited effects on a start-up's location choice when they are estimated by the employment model. Since the employment model considers the firm size effect and the quantity of created employment, this result demonstrates that the spatial distribution of jobs is more affected by the census tract level variance than by the spatial distribution of establishments. It is plausible that large establishments that created more employment are more clustered at employment centers, and hence the between census tract variance of employment is larger than the variance of establishment.
The results of standardized coefficients reported in Table 6.4 and Table 6.5 compare the magnitude of the determinants across the levels and the models. Standardized coefficients are calculated to control the differences in measurement in the unit of observation for each variable. Because the unit of observation determines the magnitude of coefficients in the model, direct interpretation of the number of coefficients could be confusing if the size of effects is compared. Since standardized coefficients consider the unit effects of each variable in terms of standard deviation, a direct comparison of coefficients is possible.

First of all, at the census tract level of the establishment model shown in Column 1 of Table 6.4, employment density and population density are two dominant factors influencing which neighborhood a new establishment chooses to move. Employment density (0.525) has approximately twice the effect of the population density (0.232). However, using the employment model shown in Column 3 of Table 6.4, the primary impact of employment density (0.498) still exists, but the population density is not a significant factor. Instead, using the employment model, existing employment within three miles (0.097) becomes the secondary factor influencing a start-up's location choice. The distribution of created employment is affected by existing employment of nearby neighborhoods up to three miles away, whereas population density does not affect employment in the same neighborhood. In addition, the standardized coefficients of demographic variables, including the percent of whites, median age, and vacant housing units, show fewer impacts in the employment model and about half of the results in the
establishment model. Thus, demographic characteristics have only a small effect on employment stemming from new establishments.

Secondly at the MSA level shown in table 6.5, the largest effect comes from the natural amenity of a warm winter. These results are consistent across the establishment model and the employment model. Since the January temperature has been a strong predictor of regional growth in the existing empirical studies, regional start-ups' rate can be an indicator of regional growth. Inclusion of business climate variables at the regional level is one unique contribution of this study to empirical studies on new establishments. Among business climate variables, technology shows the largest effect across two models. In general, a higher level of technology is associated with a more entrepreneurial atmosphere in regions, and as a result I expect a positive effect of technology on creating new establishments. However, the results of negative signs of both two models suggest that the level of technology is negatively correlated with attracting new establishments regionally. As discussed in the section 6.2, technology-based establishments are a small portion of all new establishments. The technology coefficient is negative, indicating that a common regional growth approach focusing on new technology-based development policy needs to be reconsidered to stimulate start-up activities. Finally, I need to pay attention to the standardized coefficients of agglomeration economies, having relatively small effects compared to other variables, although agglomeration economies are regarded as the key factor of regional growth. It is possible that the population size criteria of the medium-sized MSA could affect this smaller effect of agglomeration economies, because the extent of the urbanization of economies is mostly determined by
the metro size. Thus, the effect of agglomeration economies on new establishments in the medium-sized regions can be smaller than larger sized regions.

6.5 Summary and Discussion

This chapter uses the two-level model to analyze three major issues including neighborhood and regional level determinants as well as the difference of location choice between new establishments and created jobs. First, I find that existing employment and population density are two primary neighborhood level factors driving new establishments. In addition, this study sheds light on unexplored demographic factors, finding that new establishments increase with higher rates of whites and a higher median age, but decrease with higher rates of vacant housing units. Second, I find that the major regional determinant is the average January temperature, which is consistent with other work regarding regional growth. The impacts of agglomeration economies are less dominant than previous empirical findings, indicating that the extent of agglomeration economies, generally increase with scale, is likely affected by medium-sized MSA criteria. Conversely, existing empirical studies include larger regions and hence agglomeration economies are a dominant factor. The business climate with more competition and less environmental regulation attracts more establishments, but technology is associated with fewer establishments. Third, the location of created employment is not affected by the existing population distribution within an MSA, but the population nearby is a dominant factor in a start-up’s location choice. Employment’s less dependency on population distribution reveals that new establishments follow the
existing population, but the new jobs are likely clustered at non-residential areas such as employment centers.

Several issues are still unclear even though the model predicts and explains factors that affect a start-up's location choice. The fundamental reason for this obscurity comes from the assumption that new establishments are homogeneous though they are individually unique. In Chapter Three, I introduce four types of new establishments that include national firms, local businesses, spin-offs, and home-based businesses. However, there are more diverse types of new establishments that cannot be modeled in the four types of establishments. In addition, there are many unobservable and immeasurable factors that cannot be included, although these are included as the error terms in the model. Unobservable variables, such as the personal preference of start-up founders, are sometimes a dominant factor in the actual decision on where to locate. The next chapter will complement the two-level model by using a qualitative approach. In-person interviews with founders of new establishments will provide more contextual information in addition to the observable factors analyzed by the two-level model. To keep the research framework consistent, the same categories of explanatory variables of the model and the same data set will be used in Chapter Seven.
CHAPTER 7 INTERVIEWS WITH FOUNDERS OF NEW ESTABLISHMENTS

This chapter examines location choices of new establishments using a qualitative approach\textsuperscript{42} to test whether the quantitative model predicts actual cases accurately and to understand the individual stories behind start-up businesses within the data set. In-person interviews with selected start-ups in the Columbus MSA to clarify how the three categories of the census tract level variables; economic, demographic, and geographic factors, influence a new establishment's location decision. In addition, it is useful to understand any place-based variables critical to the decision to site a new establishment not included in the model. The first section introduces how the interviewees were selected. The following sections discuss the results of interviews with individual founders of new establishments that represent cases within the three categories of variables. The final section synthesizes the results of the quantitative approach and the qualitative approach emphasizing the complementary relationship between the two.

\textsuperscript{42} As this qualitative approach is a behavioral research involving human subjects, contents and methods are screened and approved by the IRB protocol review (IRB Protocol # 2012E0582).
7.1 Interviewee selection criteria and in-person site interview

Three criteria were used to select the interviewees for the case study. The first is a regional criterion, the second is a survival criterion, and the third is a model prediction criterion. For the first regional criterion, Columbus MSA, Ohio has been selected among 27 medium-sized metropolitan areas. The Columbus area is becoming one of the top destinations for start-ups, especially for knowledge-based and emerging industry businesses, as described in the following section. In addition, this location offered the best opportunity for in-person interviews, since I am a student at The Ohio State University.

For the survival criterion, the focus was to find businesses that have been successful and are still in business today. A successful business is defined as one that has been in operation more than ten years (2001-2012). Furthermore, twelve years of being in business since 2001 proves that they are survivors and their establishment location choice may have influenced their success. The founders of these successful start-ups are expected to tell the story about how the location decision for their business has actually affected their business operations compared to what they anticipated when they started the business.

The third model prediction criterion is employed to verify the estimation results of the two-level model of location choice. The census tracts in the Columbus MSA having the best economic, demographic, and geographic conditions were identified respectively using the model. New establishments still in business within the top census tracts were
contacted for an interview. The details of the selection procedure, which uses the third criterion, are introduced in the section 7.1.2.

Interviews with the founders of start-ups are necessary to understand firm level behaviors, since the decision maker is able to provide information about the motivation behind the location decision, as well as the context for this decision. All interviews were conducted at the business location in an in-person meeting with the founder of the start-up. Visiting business spaces allowed me to have a chance to observe the surrounding environments such as neighborhood physical settings as well as look around their working conditions, including building amenities. A copy of the prepared interview questions (See the Appendix B for the details) was provided at the interview, but interviewees were encouraged to speak freely about their businesses even if the issues went beyond the location decision.

7.1.1 The Columbus metropolitan area and new establishments

This section introduces the regional characteristics of the Columbus metropolitan area as a selected region for the interviews with founders of new establishments. Various recent data and indices have ranked the Columbus metropolitan area as a top market for new businesses in new industry\textsuperscript{43}. The national news media has focused attention on the success story of the Columbus region while most Midwestern regions are declining

\textsuperscript{43} For example, Forbes ranked #1 Up-and-Coming Tech City (2008).
There is additional evidence directly related to the MSA level variables of the model. Business Week ranked Columbus the 9th Best City for New College Grads (2011) that relates to the demographics category variable of the \textit{percent of people 25 or older with a college degree}. Kiplinger's ranked Columbus as the 2nd Best City for Commuters (2011), relating to the \textit{average commuting minutes} variable in the agglomeration economies category. In addition, the city has a stable housing market and large government and public-education sector, high-quality health care services,\textsuperscript{45} and it is home to the headquarters of fifteen Fortune 1000 companies. All of these features create a competitive business climate for cultivating start-ups in spite of the recession.

The Columbus MSA shows average performance regarding the volume of new establishments, and the size of the market among 27 medium-sized MSAs according to the primary data source of this study, the Dun and Bradstreet database. Figure 7.1 and Figure 7.2 illustrate that the number of new establishments and total created jobs are about average among the 27 MSAs. The 1.5 million (1,540,157) population for the Columbus MSA is in the middle of the sample. Hence, the Columbus MSA can be a representative region for the case study, when the quantity of new establishments and population size are considered.

\textsuperscript{44} “How Columbus, Ohio Bounced Back from the Recession,” TIME (Sept. 27, 2012).

\textsuperscript{45} Greater Columbus Metropolitan Area, 2012 Economic Overview, CBRE
Figure 7.1 Population and number of new establishments in 27 MSAs: The larger black point indicates the Columbus MSA.

Figure 7.2 Population and created employment in 27 MSAs: The larger black point indicates the Columbus MSA.
7.1.2 Interviewee selection procedure using the likelihood of new establishments

This section describes how the interviewees were selected by using the two-level model to consider the three categories of census tract level variables including economic, demographic, and geographic environments. The database for this study consists of the MSA level and the census tract level data sets. Since the Columbus MSA has been already selected at the MSA level, only the census tract level factors are used to select the interviewees for the case study. Using the results of the establishment model in Chapter Six, I could identify the census tracts that are expected to show the highest likelihood for attracting new establishments for the three categories of census tract level variables. The purpose of this selection process is to look for the three neighborhood census tracts that represent the best economic, demographic, and geographic conditions. The census tract with the best economic conditions is determined by the estimated score according to the formula (7.1).

\[
\text{The estimated score} = \sum \left( \text{the value of each variable} \times \text{the coefficient of the establishment model} \right) \tag{7.1}
\]

Other variables in the demographic and geographic categories are not considered for the best economic conditions. Hence, the four variables of the economic category are included: employment density, employment within three miles, industrial dominance, and the population/employment. When the estimated score of a census tract is a high number, it indicates the census tract provides superior economic conditions for start-ups. By using
the same estimation methodology, the census tracts with the best demographic conditions and the best geographic conditions are identified. Table 7.1 summarizes the three census tracts identified, the number of start-ups launched, and the number of jobs created in 2001.

To search the interviewees for the case study, all 372 census tracts in the Columbus MSA are ranked based on the estimated likelihood for attracting start-ups using the establishment model with the formula (7.1). Then I took the top census tracts for each category and looked through to find which new businesses established in 2001 are still open in 2012. Ten open businesses in the top census tracts were identified as candidates for interviews in each category. I called or emailed businesses one by one in each category until a business agreed to an interview. I contacted 17 businesses. Six of the 17 open businesses for the economic category, seven for the demographic category, and four for the geographic category.

<table>
<thead>
<tr>
<th>Category</th>
<th>Tract FIPS code</th>
<th>Location</th>
<th>New establishments</th>
<th>Employment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic Conditions</td>
<td>39049004000</td>
<td>Downtown Columbus</td>
<td>38</td>
<td>171</td>
</tr>
<tr>
<td>Demographic Conditions</td>
<td>39049001300</td>
<td>University District</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Geographic Conditions</td>
<td>39045031500</td>
<td>Lancaster</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 7.1 The top census tract for each category
For the economic category, there were 38 new establishments within the top first census tract in 2001 and we may expect there are more than ten open businesses in 2012. However, in the geographic category, there was only one start-up in the top first census tract. To identify ten open businesses in the geographic category, the top three census tracts are searched to have enough candidates for the interview. Likewise, in the demographic category, the top eleven census tracts are reviewed to discover ten open businesses.

The number of open businesses shows how many firms are survived since their openings 2001. New business survival is another interesting topic of start-up research that is associated with firm growth and sustainable job creation (Renski, 2008). Considering about 20 percent of new firms die within first year (Christie and Sjoquist, 2012), survival rate is likely a more important indicator of economic growth than the number of opening of new businesses. Appendix C summarizes the results of survival ratio for each category. The survival rate for the economic category within the top census tract is 26.3%, the demographic category within the top 11 census tracts is 33.3%, and the geographic category within the top 3 census tracts is 28.6%.

7.2 The case of economic conditions

For the case of economic conditions, a law firm within a census tract in the downtown Columbus area was interviewed. First, the characteristics of the census tract where the firm is located are introduced; the results of the interview, including the company profile, are discussed.
7.2.1 A neighborhood in the downtown area

The economic case study of a new establishment is located in downtown Columbus, which has the state office buildings and the highest employment density area in the Columbus metro area. The highest number of existing establishments and the highest employment proves that this neighborhood is part of the central business district; the downtown contains the most vibrant economic activities in the Columbus area. Therefore, residential development is limited, and the small Population/Employment ratio, as shown in Table 7.2, indicates that this area is a dominant place for work. The western portion near the Scioto River is full of high-rise office buildings, while the eastern portion includes relatively low density development, including the Discovery District.
Figure 7.3 The selected census tract for the interview of economic category (Background image source: Google Maps, 2012)
<table>
<thead>
<tr>
<th>Year 2000</th>
<th>Population: 2,195</th>
<th>Employment: 38,055</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 2001</td>
<td>New establishments: 38</td>
<td>Created employment: 171</td>
</tr>
<tr>
<td><strong>Category</strong></td>
<td><strong>Variable</strong></td>
<td><strong>Value</strong></td>
</tr>
<tr>
<td>Economic</td>
<td>Employment density</td>
<td>41,818.68 employees/sq mi</td>
</tr>
<tr>
<td></td>
<td>Employment within three miles</td>
<td>151,692 employees</td>
</tr>
<tr>
<td></td>
<td>Industrial dominance</td>
<td>0.143 (0-1 index)</td>
</tr>
<tr>
<td></td>
<td>Population/Employment ratio</td>
<td>0.058</td>
</tr>
<tr>
<td>Demographic</td>
<td>Population density</td>
<td>2412.1 people/sq mi</td>
</tr>
<tr>
<td></td>
<td>Percent of white</td>
<td>69.7 %</td>
</tr>
<tr>
<td></td>
<td>Median age</td>
<td>42 years old</td>
</tr>
<tr>
<td></td>
<td>Percent of vacant housing units</td>
<td>8.65 %</td>
</tr>
<tr>
<td>Geographic</td>
<td>Distance to nearest highway</td>
<td>0.3 miles</td>
</tr>
</tbody>
</table>

Table 7.2 The economic census tract variables (FIPS code: 39049004000)

### 7.2.2 Interview results of the economic category

This section discusses the interview with the founder of a downtown law firm. I discuss the motivation of the founder to start their business and their decision on where to locate the new establishment. Since this firm has been selected according to the economic condition criteria, economic-related topics are analyzed first, and then demographic and geographic factors are examined.

The founder worked for a national law firm’s branch in downtown Columbus right after graduating from law school. After moving to a second job for another national
law firm also in the downtown Columbus, she became a partner and specialized in bankruptcy cases. At that point, she thought that she could become independent because of having enough of her own clients. Consequently, she founded a bankruptcy law firm that “combines the breadth of a large firm experience with the personalization of a small firm practice.” Thus, this is a case of the Spin-offs (Type III) among the four types of the new establishments in Chapter Three. Her explanation of the spin-off process and the location decision nearby the parent firm confirms that the Type III models the location choice behaviors of spin-off cases adequately. She co-founded the firm with another attorney in 2001. Since the co-founder left years later, she has worked with an associate attorney and a legal assistant. Hence, three people work for the firm.

She mentions three primary reasons for her location choice. The first is its proximity to the Bankruptcy Court that is located in downtown Columbus. The Court is less than a mile and she needs to visit the Court two to three times a week. The second reason for her choosing that location is its proximity to the law firms of her colleague attorneys. She reported that most legal businesses strongly prefer a downtown location, whereas numerous firms tend to choose the suburbs instead. Furthermore, legal-related events and meetings are frequently held in the downtown. The third reason she chose her location is that she prefers the neighborhood atmosphere including iconic buildings such as the Columbus Metropolitan Library and the open green space, such as the Topiary Park. Because the office building is located at the border between downtown high-rise buildings and the low-density development in the Discovery District, the office site is at an ideal location to enjoy both urban amenities and openness. As a result, she has actively
participated in the neighborhood meetings of the Discovery District and enjoyed her community involvement.

The economic factors of the model matter for her decision making as she highlighted the importance of proximity to the Court and other legal firms. The literature attributes this preference for downtowns locations to the benefits of face-to-face interaction. She confirmed that face-to-face meetings are still essential in the legal industry, since people feel secure when they talk about private problems in person. The high-density downtown development pattern is an outcome of businesses opting for proximity between economic activities. The employment density variable in the model predicts this tendency.

The demographic factors affect her decision making as well although that is not expected according to the model. Her preference to surrounding neighborhoods is a demographic factor to be considered, but the model does not include those factors such as quality of neighborhoods and historic characteristics. In general, demographic characteristics are analyzed in a demand factor, since the residents are likely customers of start-ups. However, this interview reveals that the amenity of a surrounding neighborhood is also important as working conditions affected by surrounding neighborhood characteristics and environments.

Finally, the geographic factors impact this location choice because they provide convenient accessibility for clients. Her clients mostly come from all around the Columbus metro areas such as Dublin and Bexley, or small cities in central Ohio. The
central downtown location in the Columbus area as well as in the state of Ohio is a plus in attracting clients.

7.3 The case of demographic conditions

For the case of demographic conditions, a home-based design firm within a census tract in the neighborhood of Victorian Village was interviewed. First, the characteristics of the census tract where the firm is located are introduced, and the results of the interview, including the company profile, are discussed.

7.3.1 A neighborhood in Victorian Village

The demographic case of a new establishment is within Victorian Village, which is one of the historic districts in Columbus. The high population/employment ratio variable, as shown in Table 7.3, confirms that this is a largely residential neighborhood, whereas the Employment within three miles reflects the proximity to downtown Columbus. Additionally, the distance to the nearest highway (0.38 miles) is very close, and this is another advantage of this neighborhood in terms of accessibility from other areas. The low vacancy rate indicates that this neighborhood is a healthy neighborhood and has good housing conditions.
Figure 7.4 The selected census tract for the interview of the demographic category

(Background image source: Google Maps, 2012)
<table>
<thead>
<tr>
<th>Category</th>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic</td>
<td>Employment density</td>
<td>1,518.52 employees/sq mi</td>
</tr>
<tr>
<td></td>
<td>Employment within three miles</td>
<td>180,763 employees</td>
</tr>
<tr>
<td></td>
<td>. Industrial dominance</td>
<td>0.107 (0-1 index)</td>
</tr>
<tr>
<td></td>
<td>. Population/Employment ratio</td>
<td>7.651</td>
</tr>
<tr>
<td>Demographic</td>
<td>Population density</td>
<td>11,618.5 people/sq mi</td>
</tr>
<tr>
<td></td>
<td>Percent of white</td>
<td>84.0 %</td>
</tr>
<tr>
<td></td>
<td>Median age</td>
<td>30.5 years old</td>
</tr>
<tr>
<td></td>
<td>Percent of vacant housing units</td>
<td>6.0 %</td>
</tr>
<tr>
<td>Geographic</td>
<td>Distance to nearest highway</td>
<td>0.38 miles</td>
</tr>
</tbody>
</table>

Table 7.3 The demographic census tract variables (FIPS code: 39049002000)

7.3.2 Interview results of the demographic category

This section discusses the interview results with a design firm in Victorian Village. I discuss the motivation of the founder to start their business and their decision on where to locate the new establishment. Since this firm has been selected according to the demographic condition criteria, demographic-related topics are analyzed first, and then economic and geographic factors are examined.

The founder's first job was with a county government, but he wanted to find a way to use his education in fine arts. After completing a degree in a technical college to
advance his design skills, one of his colleagues in the program referred a client to him for design work involving a kids' magazine. As he got a reputation in the magazine industry, he realized that he could have his own clients, then started his own business. However, he has never rented office space. Rather, he uses a floor of his house as a design studio. Thus, his design company is a home-based start-up that is a case of the Home businesses (Type II) among the four types of the start-ups in Chapter Three. His explanation of the search process for the house confirms that the Type II models the location choice behaviors of home-based start-ups properly. One difference between the Type II and his decision is that he considered his start-up and the work space within a house as an essential factor when he purchased the house. However, the Type II explains the housing location choice without considering start-ups, and does not consider the work space for start-ups. According to the Type II, the founder purchases a house only for a residence, but later decides to start a business. He chose this current site for living and for working. He liked the Victorian Village neighborhood, but, at first, he could not afford to buy the house. So initially he rented the first floor only, while the owner still lived on the second floor. At that time, he used a portion of the first floor as his working studio. As his business grew and he needed more space for it, he purchased the entire house, and the second floor became his office. As an office location choice, the Victorian Village provides a proper
environment for creative workers. Although he does not have a close relationship with other design businesses in the community, the office location within the neighborhood enhances his corporate image because of the artistic environment of surroundings such as the Short North Arts District. The historic district of the neighborhood has a positive impact on advertising his businesses.

The interview confirms why the demographic conditions are significant factors with home-based start-ups, because the office location decision is also a housing location decision. The consideration of work space is influential as well, as he uses the second floor for a design studio. He looked for a spacious house with big rooms that is commonly available in the Victorian Village, where the average size of a detached house is larger than in other nearby neighborhoods. This preference for a bigger house for home-based start-ups is consistent with the findings of the literature (Moos and Skaburskis, 2008).

The economic factors of the neighborhood are not considerable issues for this company, since the neighborhood is mostly a residential area, and he rarely interacts with other businesses nearby. However, the geographic location relatively close to the city center does matter, as most of his clients come from downtown area businesses. He is satisfied with the location within an inner city neighborhood and said “eighty percent of

46 The research on the neighborhood built environment and home-based business is limited. One example is “The impact of the residential built environment on work at home adoption frequency: An example from Northern California” (Tang et al., 2011).
people I worked with locate within a mile away from the downtown, travel and tourism, consultancy, and PR firms.”

### 7.4 The case of geographic conditions

For the case of geographic conditions, a franchise firm specializing in foot care services within a census tract in the Northland area was interviewed. Since the firm relocated recently in 2012, this case analyzes the motivation of relocation as well. First, the characteristics of the census tract where the firm is located are introduced, then the results of the interview, including the company profile, are discussed.

### 7.4.1 A neighborhood in the Northland area

The original location of the geographic case of a new establishment in 2001 was on Cleveland Avenue in the Northland area. The current location in 2012 is in the Polaris area, one of the fastest-growing suburban employment centers in the Columbus metro area. In addition to their proximity to the highway, both locations have many similar attributes as Table 7.4 shows. All demographic characteristics are alike except the owner-occupancy. The higher owner-occupancy of the Polaris area (63%) reflects that the average income is likely higher than the Northland area (43.9%).
Figure 7.5 The selected census tract for the interview of geographic conditions case (2001: the initial census tract, 2012: the tract after relocation) (Background image source: Google Maps, 2012)

<table>
<thead>
<tr>
<th>Tract</th>
<th>Original Tract (#71.14)</th>
<th>Current after relocation (#124)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>4,823</td>
<td>4,815</td>
</tr>
<tr>
<td>Median age</td>
<td>35.2 years old</td>
<td>33 years old</td>
</tr>
<tr>
<td>Housing units</td>
<td>2,204</td>
<td>2,403</td>
</tr>
<tr>
<td>Vacant percent</td>
<td>8.5%</td>
<td>9.2%</td>
</tr>
<tr>
<td>Owner-occupied</td>
<td>43.9%</td>
<td>63%</td>
</tr>
<tr>
<td>White percent</td>
<td>64.9%</td>
<td>78.5%</td>
</tr>
</tbody>
</table>

Table 7.4 Original establishment tract and relocated tract (Source: Census 2010)
<table>
<thead>
<tr>
<th>Year 2000</th>
<th>Population: 2,787</th>
<th>Employment: 7,860</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 2001</td>
<td>New establishments: 20</td>
<td>Created employment: 116</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category</th>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic</td>
<td>Employment density</td>
<td>5,420.69 employees/sq mi</td>
</tr>
<tr>
<td></td>
<td>Employment within three miles</td>
<td>34,283 employees</td>
</tr>
<tr>
<td></td>
<td>. Industrial dominance</td>
<td>0.225 (0-1 index)</td>
</tr>
<tr>
<td></td>
<td>. Population/Employment ratio</td>
<td>0.355</td>
</tr>
<tr>
<td>Demographic</td>
<td>Population density</td>
<td>1,922.1 people/sq mi</td>
</tr>
<tr>
<td></td>
<td>Percent of white</td>
<td>76.5 %</td>
</tr>
<tr>
<td></td>
<td>Median age</td>
<td>30.7 years old</td>
</tr>
<tr>
<td></td>
<td>Percent of vacant housing units</td>
<td>6.6 %</td>
</tr>
<tr>
<td>Geographic</td>
<td>Distance to nearest highway</td>
<td>0.02 miles</td>
</tr>
</tbody>
</table>

Table 7.5 The geographic census tract attributes (FIPS code: 39049007114)

7.4.2 Interview results of the geographic category

This section discusses the interview results with a foot care service firm originally in the Northland area and currently in the Polaris area. I discuss the motivation of the founder to start their business and their decision on where to locate the new establishment. Since this firm has been selected according to the geographic condition criteria,
geographic-related topics are analyzed first, and then economic and demographic factors are examined.

The firm specializes in foot care services by creating customized foot arch supports. The headquarters is in California, and this office is a Columbus branch of a national franchise company. Thus, this is a branch case of the National firms (Type IV) among the four types of the start-ups in Chapter Three. The founder was initially a customer of the foot service when he worked for an electric company in California. Because he was satisfied with the services and the products, he decided to open a branch in Columbus since his family lives nearby. Hence, when he chose a region, his personal background was a dominant factor rather than other regional conditions.

The geographic factor was the most important factor for both the first location in the Northland area and the second location in the Polaris area. Both locations are very close to the highway I-270, the beltway of the Columbus metro area. “Since this is a very specialized business and we don’t have direct competition,” easy access to the shop through the beltway is significant in attracting customers regionally. Customers come from within fifty miles around Columbus including small rural cities outside of the Columbus metro.

The demographic factors seem to be irrelevant to the location decision, as most customers come from various areas in central Ohio. Nevertheless, the demographic factors influence the location choice in a different way. Although surrounding neighborhoods are not considered as a factor contributing to the demand, the physical amenity of the neighborhood is important. The reason for the first location is its
proximity to the highway, but also the rent was affordable since the neighborhood in the Northland area is of moderate quality area. The business has been successful, in spite of neighborhood decline over the last decade, and its owner is able to afford a new location in a high rent area in the Columbus metro area. He noted that “this location is much more expensive than the original location.”

The relocation behavior of this firm proves the importance of the economic factors. The variables of employment density and employment within three miles are well-selected factors to explain the relocation. The old location is a low-density employment area with mostly residential development within three miles. The new location is within one of the fastest growing suburban employment centers in the Columbus metropolitan area.

7.5 Discussion

In-person interviews with selected founders of start-ups spell out how the variables in the model influence where new establishments locate. In addition, I found other significant factors missed in the model, for example the personal preference of founders. The interviewees agreed with the overall effects of neighborhood factors that I employed in the model. For the economic condition case, the employment density and the employment within three miles variables are considered for the location choice of the law firm. For the demographic condition case, the population density and the percent of vacant housing units are considered. For the geographic condition case, the distance to the nearest highway is a dominant factor for the location choice.
Although they are not affected by regional factors according to the interview results, they reported it is important but they did not recognize them. In addition, they emphasized the importance of a quality neighborhood, including its natural amenities, historical atmosphere, and physical settings; all factored in to where they decided to locate. Some of these qualitative factors can be quantifiable and included in the model for future research. Many approaches are already developed to measure the value of natural environments or physical conditions. For example, the average age of housing units can be used to measure historic characteristic of neighborhoods. The distance to nearest open space and community parks can be a proxy for natural amenity. Nevertheless, there are still immeasurable factors that influence location choice behaviors, for instance the personal taste of decision-makers regarding the locations. The results of the interviews show that this qualitative approach plays a key role in complementing the quantitative model.
CHAPTER 8 CONCLUSION

8.1 Summary of findings

This dissertation aims to identify the spatial determinants of new establishment location decisions in the U.S. on two different geographic levels, examining the regional and neighborhood influences. To look at the effects of the two levels, the MSA level variables and the census tract level variables are derived taking into account theoretical debates and prior empirical findings. The combination of a quantitative approach of multilevel modeling and a qualitative approach of in-depth interviewing sheds light on how the conditions of regions and neighborhoods affect where new establishments locate.

An examination of the neighborhood level determinants is a primary contribution of this study, since most new establishments are micro-enterprises with a small number of employees targeting local markets. Since local markets are associated with the neighborhood conditions, neighborhood factors are more influential for local businesses than regional factors. Regional factors affect new establishments, albeit indirectly. Existing employment and population density are two major neighborhood level determinants positively influencing new establishments. The employment density effect has about twice as much of an impact as the population density effect in predicting the number of new establishments. This study empirically tests the existence of the
population density effect on new establishments at the neighborhood scale; the effect is only hypothesized in the literature. In addition, industrial dominance is negatively associated with new establishments that may be industries outside of those dominated in a geographic area. When dominant firms and vertically related firms are co-located, new establishments that are irrelevant to the dominant firms have been crowded out and excluded from the geographic integration. This result is inconsistent with the cluster strategy that encourages co-location and spatial proximity among related industries, although the geographic scale of the industry cluster is generally larger than a census tract. Therefore, this finding can initiate the debate on the optimal degree of industry mix at the neighborhood level between specialization and diversification strategies. Finally, higher rates of vacant housing units, an indicator of a declining neighborhood, are associated with fewer establishments in a neighborhood, although the effect of housing vacancy is the smallest within the demographic category.

The results of the MSA level analysis including the business climate and amenities are mostly consistent with the empirical findings of previous studies. Among the MSA level variables, the average January temperature has the most significant positive effect on new establishment location choice. This confirms the presence of the warm winter effect, warmer regions grow quickly. That is a well-known empirical finding of recent regional growth studies. One interesting result is that the impacts of agglomeration economies are less dominant, except the negative effect of commuting minutes, than previous empirical findings. For medium-sized metro areas, it may be that the effect of urbanization economies is smaller than in larger metro areas as the
population and market size of a certain region determines the extent of its urbanization economies.

In-person interviews with selected founders help this research explore the meaning of the outcomes of the quantitative model, understand the individual stories behind start-up businesses, and discover any place-based variables that are not in the model. In addition, the results of interviews clarify how the variables in the model influence new establishment location decisions, along with others that the founder believes are critical to the decision to site their business. The interviewed founders agreed with the overall effects of neighborhood factors including economic, demographic, and geographic conditions that I employed in the model. Although they are not directly affected by regional factors for their establishments, they think regional factors are important to enhance the business climate. In addition, they emphasized how the quality of neighborhood conditions including natural amenities, historical atmosphere, and physical settings factored into where they decided to locate, which are usually unmeasured in business location choice models.

8.2 Policy implications

First, the economic value of neighborhoods deserves more attention from local policy makers and planners. When I started this research, I hypothesized that neighborhood demographic factors play a key role in influencing the demand side of new businesses. My findings confirm the impacts of demographic conditions on new establishments. In addition, the number of home-based establishments is not small,
especially for the small size cases. Although home-based work has a positive effect on the local economy and vitality of communities, the zoning laws try to control them, rather than encourage them. For example, the City of Sacramento, which is one of 27 MSAs in the data, requires Home Occupation Permits to run a business in a residential area. The purpose of the code\textsuperscript{47} is, “to reduce the impact of the business on the surrounding neighborhood.” The traditional zoning perspective of “Home/Work Separation” (Gurstein, 1996; Garnett, 2001) needs to be improved to consider increasing numbers and portions of home-based employment including start-ups, teleworkers, and occasional homeworkers. In addition, the official data sources from the government that surveys the current status of home-based businesses including population, types, and industry sectors is very limited. Therefore this needs to be developed for further research.

Second, planners and policy makers need to distinguish the emphasis of start-ups' policy for local municipalities from those of regional authorities, such as economic development authorities. For instance, improving the business climate is a more effective policy at the regional level, but at the neighborhood level, providing a better quality of infrastructure and physical amenities is essential for attracting and enhancing entrepreneurial activities. Furthermore, city or state governments have created policy tools to attract new establishments, however neighborhood or community level approaches are underestimated and underdeveloped. One reason for this lack of policy at the neighborhood level is that the issues of small-scale establishments are managed principally by the economic development policies of municipal development departments.

\textsuperscript{47} Title 17.224 of the Sacramento City Code
They are more interested in an average level of start-ups within their administrative boundary, rather than paying attention to individual characteristics of communities. One remedy for solving this problem is re-organizing city government departments so that economic development is managed in closer connection with neighborhood, planning, and housing-related departments.

Third, medium-sized MSAs need to pursue a different path for encouraging new establishments instead of following the strategies developed for larger metro areas. Most regional economic development approaches are primarily developed and applied in practice by the largest metro areas in the U.S. When some of these approaches were successful, other cities adopted them. Simultaneously, small rural towns have also developed their own ways of attracting businesses, because the elimination of small businesses has been occurring in some communities. The results of MSA level analysis reveal that the impacts of agglomeration economies vary depending on the size of regions. It is necessary for medium-sized regions to create their own paths of economic development, through policies that support start-ups, learning from the policies of other regions similar in size.

8.3 Future research directions

While this dissertation only focuses on the determinants of new establishments, the consequences of new establishments are sometimes even more significant issues and therefore should be examined for their implications on future policies. A common direction is to examine the effects of new establishments on the economic performance of
certain regions, such as an increased income level and a decreased unemployment rate. Another direction is to analyze how the spatial distribution of new establishments influences land uses and development patterns. That direction will look at the spatial distribution of new establishments, i.e. the extent to which they choose to locate in central cities and suburban employment centers, or whether they disperse outside existing centers, perhaps contributing to the evolution of future new centers.

One limitation of the model in this research is that new establishments are regarded as a homogeneous business even though they are each different. A comparison of the factors influencing establishment and the factors influencing created employment partly captures the size effect from the number of employees in each establishment, but explanations of the results are mostly based on assumptions. To overcome this limitation of the aggregate model, development of separate models by industry, size, and establishment year, will provide more information that resolves questions raised through the current model. For instance, a separate model looks at the industry composition of new establishments, and how the industry composition varies according to its spatial distribution. New establishments’ location choices could duplicate the existing industry composition at their choice of location; that is they might co-locate with similar establishments already in place. They might locate elsewhere, thereby contributing to change in the industrial structure, and the development of new specializations or perhaps the diversification of the sectoral composition.

A firm-level approach is also a desirable direction for future research, though such an approach cannot be accomplished without a lot of data management. Since the
raw data of the Dun and Bradstreet database is firm-level data, technically it is possible to do a firm-level research based on this massive database. For example, a point pattern analysis contributes to our understanding of how existing firms and new additions of establishments are spatially interacted and influence each other. The firm level data also enables the development of a three-level model that includes the firm level, the neighborhood level, and the MSA level. It can do this by using the same two-level model developed for this study.

Finally, using administrative boundaries as spatial units of analysis rather than census tracts will evaluate the effects of current policies on new establishments. Even though this is a more plausible approach to testing the impacts of existing policies on efforts to attract new establishments such as tax subsidies and to suggest policy implications, the data for analysis is not available across regions or across cities; that data is necessary to develop such a model. The actual governmental boundaries are not MSAs or census tracts. Rather, a city or a municipality has its own independent authorities for dealing with issues associated with new establishments.
REFERENCES


ESRI, (2010). Esri StreetMap Premium [Data file].


Haltiwanger, J., Jarmin, R. S., & Miranda, J. (2012). *Business Dynamics Statistics Briefing: Where Have All the Young Firms Gone?* Ewing Marion Kauffman Foundation.


Kane, T. J. (2010). *The Importance of Startups in Job Creation and Job Destruction*. Ewing Marion Kauffman Foundation.


APPENDIX A: DUN AND BRADSTREET DATABASE ATTRIBUTES
### Dun and Bradstreet Database Attributes

<table>
<thead>
<tr>
<th>Type</th>
<th>Variable</th>
<th>Unit</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>XY-Coordinate</td>
<td>(X, Y)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Physical Address</td>
<td>Address</td>
<td></td>
</tr>
<tr>
<td>Numerical Variable</td>
<td>Employees</td>
<td>Number of Person</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Annual Sales</td>
<td>US Dollars</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Facility Size</td>
<td>Square Feet</td>
<td></td>
</tr>
<tr>
<td>Categorical Variable</td>
<td>Location Type</td>
<td>HQs / Branch / Single</td>
<td>A company’s Location Type will be one of the three.</td>
</tr>
<tr>
<td></td>
<td>NAICS Codes</td>
<td>Six digits</td>
<td>A company’s primary NAICS code</td>
</tr>
<tr>
<td></td>
<td>Year Established</td>
<td>Year</td>
<td>The year that the business was first established.</td>
</tr>
<tr>
<td></td>
<td>D&amp;B Prescreen Score</td>
<td>High / Medium / Low Risk</td>
<td>To assess risk by predicting the likelihood of a firm paying in a severely delinquent manner over the next 12 months.</td>
</tr>
</tbody>
</table>
I. The two-level linear model:

[Level 1: Census Tract $i$]

$$Y_{ij} = \beta_{0j} + \sum \beta_{qj}X_{qij} + e_{ij}$$

$\beta_{qj}$ (q=0,1,...,Q) are census tract level coefficients;

$X_{qij}$ is the census tract level predictor $q$ for census tract $i$ in MSA $j$;

$e_{ij} \sim N(0, \sigma^2)$ is the census tract level random effect.

[Level 2: MSA $j$]

$$\beta_{qj} = \gamma_{q0} + \sum \gamma_{qs}Z_{sj} + u_{qj}$$

$\gamma_{qs}$ (q=0,1,...,Q) are MSA level coefficients;

$Z_{sj}$ is a MSA level predictor;

$u_{qj}$ is a MSA level random effect;

For each unit $j$, the vector $(u_{0j}, u_{1j}, ..., u_{Qj})'$ is distributed as multivariate normal.

II. The two-level Poisson model:

The establishment model in Section 6.2 is based on the Poisson distribution for the census tract level estimation. The two-level Poisson model takes the following form:

$$E(Y_{ij} | \beta_j) = \text{Var}(Y_{ij} | \beta_j) = \lambda_{ij}$$

where the count $Y_{ij}$ is the number of new establishments in a census tract $i$ in MSA $j$. The link function for the Poisson distribution is the logarithm;

$$\log(\lambda_{ij}) = \eta_{ij}$$
[Level 1: Census Tract $i$]
$\eta_{ij} = \beta_{0i} + \beta_{1j} (CT \text{ employment density}) + \beta_{2j} (Employment \text{ within 3miles}) + \beta_{3j} (Industrial \text{ dominance}) + \beta_{4j} (Population/Employment) + \beta_{5j} (CT \text{ population density}) + \beta_{6j} (Percent \text{ of white}) + \beta_{7j} (Median \text{ age}) + \beta_{8j} (Percent \text{ of vacant housing units}) + \beta_{9j} (Distance \text{ to nearest highway}) + e_{ij}$

[Level 2: MSA $j$]
$\beta_{0j} = \gamma_{00} + \gamma_{01}(MSA \text{ employment density}) + \gamma_{02}(Average \text{ commuting minutes}) + \gamma_{03}(GDP \text{ per capita}) + \gamma_{04}(Technology) + \gamma_{05}(Finance) + \gamma_{06}(Competition) + \gamma_{07}(Environmental \text{ policy}) + \gamma_{08}(January \text{ temperature}) + \gamma_{09}(Percent \text{ of college degree 25 plus}) + u_{0j}$

$\beta_{1j} = \gamma_{10} + u_{1j}; \beta_{2j} = \gamma_{20} + u_{2j}; \beta_{3j} = \gamma_{30} + u_{3j}; \beta_{4j} = \gamma_{40} + u_{4j}$;

$\beta_{5j} = \gamma_{50} + u_{5j}; \beta_{6j} = \gamma_{60} + u_{6j}; \beta_{7j} = \gamma_{70} + u_{7j}; \beta_{8j} = \gamma_{80} + u_{8j}; \beta_{9j} = \gamma_{90} + u_{9j}$. 
APPENDIX C: INTERVIEW QUESTIONS
Interview Questions

Opening Script:

Hello, my name is Jae Teuk Chin and I am conducting research about how businesses decide where to locate their business for my doctoral dissertation in city and regional planning at The Ohio State University. Thank you for agreeing to participate in this interview. This interview will take approximately thirty minutes.

As a successful business owner, I’m interested in learning about the factors that lead to your decision to open and locate your business. Your participation in this research study will contribute to understanding business location decisions in the Columbus area. Your decision to participate is voluntary. You may choose not to answer any question during this interview or withdraw a response after the fact.

May I record this interview? The recording will be transcribed and will assist me in reviewing the results of this interview.

Each person will be handed a sheet with the following information:

Should you have any questions on this study or wish to withdraw any of your responses, you may contact my advisor Dr. Jennifer Cowley at 614-XXX-XXXX or myself directly.
For questions about your rights as a participant in this study or to discuss other study-related concerns or complaints with someone who is not part of the research team, you may contact Ms. Sandra Meadows in the Office of Responsible Research Practices at 1-800-678-6251.

1. Motivation:

What made you decide to start your own business?
What kind of background did you have in this industry before starting your business (if not offered in the initial question)?

2. Timing:

How long had you been thinking about starting your business prior to opening in 2001?
What made you decide that 2001 was the right time?

3. Location:

Thinking back to 2001, what factors did you consider when deciding where to locate your new business?
Thinking to the future, say ten years from now. Do you anticipate that your business will be in this location?
If you could choose to locate your business anywhere, where would it be?
3a. Variables

In making the decision to locate your business here:

How important was it that you have a location close to where other people work?

How important was it that the location be close to where people live?

How important was it that your business be located near a highway?

4. Operation and employees:

How has your company grown?

How many people are employed with your business now? Are these full or part time employees (if needed)?

5. Customers:

Can you tell me about the types of products or services that you provide to your customers?

Who are your customers? Are your customers mostly from this neighborhood, the Columbus region, or from farther away?

How do your customers find you?

How often do you meet face to face with your customers? (Especially for home-based businesses)

6. Independence:

Do you have any other locations of this business?
7. **Community involvement:**

When you think about this neighborhood, what is it that you like most about being located here?

Is there anything about the neighborhood you wish would change?

8. **Business climate and financing:**

Do you believe the business climate that we have here in Columbus area is supportive of your business?

Over the last ten years how easy or difficult has it been to obtain financing when you needed it?

9. **Concluding questions:**

Is there anything else I should know about your business?

Can I followup with you if I have any further questions later?
APPENDIX D: LIST OF OPEN BUSINESSES
### Economic category

<table>
<thead>
<tr>
<th>Census tract</th>
<th>Survival ratio</th>
<th>Employment</th>
<th>Line of business</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 1</td>
<td>10/38</td>
<td>2</td>
<td>Public Relations Services</td>
</tr>
<tr>
<td>Top 1</td>
<td>10/38</td>
<td>8</td>
<td>Medical Doctor's Office</td>
</tr>
<tr>
<td>Top 1</td>
<td>10/38</td>
<td>8</td>
<td>Architectural Services</td>
</tr>
<tr>
<td>Top 1</td>
<td>10/38</td>
<td>5</td>
<td>Staffing &amp; Job Board</td>
</tr>
<tr>
<td>Top 1</td>
<td>10/38</td>
<td>15</td>
<td>Secretarial/Court Reporting</td>
</tr>
<tr>
<td>Top 1</td>
<td>10/38</td>
<td>1</td>
<td>Legal Services Office</td>
</tr>
<tr>
<td>Top 1</td>
<td>10/38</td>
<td>3</td>
<td>Legal Services Office</td>
</tr>
<tr>
<td>Top 1</td>
<td>10/38</td>
<td>8</td>
<td>Business Support Services</td>
</tr>
<tr>
<td>Top 1</td>
<td>10/38</td>
<td>1</td>
<td>Legal Services Office</td>
</tr>
<tr>
<td>Top 1</td>
<td>10/38</td>
<td>21</td>
<td>Condominium Association</td>
</tr>
</tbody>
</table>

### Demographic category

<table>
<thead>
<tr>
<th>Census tract</th>
<th>Survival ratio</th>
<th>Employment</th>
<th>Line of business</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 2</td>
<td>2/4</td>
<td>3</td>
<td>Civic Association</td>
</tr>
<tr>
<td>Top 2</td>
<td>2/4</td>
<td>2</td>
<td>Retail Gifts/Novelties</td>
</tr>
<tr>
<td>Top 3</td>
<td>1/5</td>
<td>1</td>
<td>Business Consulting Services</td>
</tr>
<tr>
<td>Top 5</td>
<td>1/3</td>
<td>30</td>
<td>Eating Place</td>
</tr>
<tr>
<td>Top 6</td>
<td>1/5</td>
<td>1</td>
<td>Business Services Commercial Art/Graphic Design</td>
</tr>
<tr>
<td>Top 7</td>
<td>2/3</td>
<td>3</td>
<td>Retail Groceries</td>
</tr>
<tr>
<td>Top 7</td>
<td>2/3</td>
<td>6</td>
<td>Business Consulting Services</td>
</tr>
<tr>
<td>Top 8</td>
<td>1/5</td>
<td>5</td>
<td>Contractor - Heating &amp; Air Conditioning Systems</td>
</tr>
<tr>
<td>Top 10</td>
<td>1/1</td>
<td>2</td>
<td>Religious Organization</td>
</tr>
<tr>
<td>Top 11</td>
<td>1/2</td>
<td>5</td>
<td>Beauty Shop</td>
</tr>
</tbody>
</table>

Top 1-11 10/30
<table>
<thead>
<tr>
<th>Geographic category</th>
<th>Census tract</th>
<th>Survival ratio</th>
<th>Employment</th>
<th>Line of business</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 1</td>
<td>1/1</td>
<td>7</td>
<td>Retail Furniture</td>
<td></td>
</tr>
<tr>
<td>Top 2</td>
<td>5/14</td>
<td>2</td>
<td>Nonresidential Building Operator</td>
<td></td>
</tr>
<tr>
<td>Top 2</td>
<td>5/14</td>
<td>2</td>
<td>Chiropractor's Office</td>
<td></td>
</tr>
<tr>
<td>Top 2</td>
<td>5/14</td>
<td>65</td>
<td>Medical Doctor's Office</td>
<td></td>
</tr>
<tr>
<td>Top 2</td>
<td>5/14</td>
<td>10</td>
<td>Animal Services</td>
<td></td>
</tr>
<tr>
<td>Top 2</td>
<td>5/14</td>
<td>10</td>
<td>Accounting/Auditing/Bookkeeping</td>
<td></td>
</tr>
<tr>
<td>Top 3</td>
<td>4/20</td>
<td>2</td>
<td>Individual/Family Services</td>
<td></td>
</tr>
<tr>
<td>Top 3</td>
<td>4/20</td>
<td>10</td>
<td>Business Consulting Services</td>
<td></td>
</tr>
<tr>
<td>Top 3</td>
<td>4/20</td>
<td>15</td>
<td>Insurance Agencies</td>
<td></td>
</tr>
<tr>
<td>Top 3</td>
<td>4/20</td>
<td>6</td>
<td>Retail Misc Merchandise</td>
<td></td>
</tr>
<tr>
<td>Top 1-3</td>
<td>10/35</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX E: INVITATION SCRIPT
Dear INTERVIEWEE NAME HERE,

Hello, my name is Jae Teuk Chin and I am conducting research about how businesses decide where to locate their business for my doctoral dissertation in city and regional planning at The Ohio State University. I would like to invite you to participate in an interview that will take approximately thirty minutes. I found your business through the Dun and Bradstreet business database, which indicates that you started your business in 2001. As a successful business owner, I’m interested in learning about the factors that lead to your decision to open and locate your business. Your participation in this research study will contribute to understanding business location decisions in the Columbus area. Your decision to participate is voluntary. I would be happy to set up a time and location that is convenient for you. Please respond by email or call me at 614-XXX-XXXX and we can discuss setting up an appointment.

Should you have any questions on this study, you may contact my advisor Dr. Jennifer Cowley at 614-XXX-XXXX or myself directly at 614-XXX-XXXX. For questions about your rights as a participant in this study or to discuss other study-related concerns or
complaints with someone who is not part of the research team, you may contact Ms. Sandra Meadows in the Office of Responsible Research Practices at 1-800-678-6251.

Sincerely,

Jae Teuk Chin