A Firmographic Microsimulation Model of

Small and Medium-Sized Business Establishments:

Application to the City of Hamilton, Canada

A Firmographic Microsimulation Model of Small and Medium-Sized Business Establishments: Application to the City of Hamilton, Canada

By

Jia Yang, M.Sc.E.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Arts

McMaster University

MASTER OF ARTS (2011)	McMaster University
(School of Geography and Earth Sciences)	Hamilton, Ontario

TITLE: A Firmographic Microsimulation Model of Small and Medium-Sized Business Establishments: Application to the City of Hamilton, Canada

AUTHOR: Jia Yang, M.Sc.E. (China University of Geosciences (Wuhan))

B.E (China University of Geosciences (Wuhan))

SUPERVISOR: Dr. Pavlos S. Kanaroglou

NUMBER OF PAGES: x, 93

ABSTRACT

Previous research on firmography microsimulation model for the City of Hamilton, Ontario, has been conducted by estimating several crucial statistical models, which would serve as the basis for an operational computer simulation model.

Based on the previous research, this thesis illustrates the implementation of the firmography microsimulation model for the City of Hamilton. This implementation includes the development of separate computer modules for the survival submodel, the mobility submodel, the location choice submodel and the firm formation submodel, as well as the integration of all these submodels. Meanwhile, the data storage mechanism, the simulation results visualization and analysis functions have been implemented by the support of GIS technology.

The microsimulation model starts with the 1990 firm micro data for the City of Hamilton as the base year and proceeds year by year with the simulation. The simulation results of firm distribution are validated by 1997 firm micro data for the City of Hamilton. The validation has proved that the developed firmography model is able to capture the overall trend of urban development processes in terms of firms at the micro level.

The limitations of the current model, especially those caused by the requirement for detailed data, are discussed, and some directions for the future research are indicated.

ACKNOWLEDGEMENTS

I would like to sincerely appreciate my supervisor, Dr. Pavlos Kanaroglou for his limitless support and patience in mentoring me during my two years of graduate studies. Thank you for your knowledge to help me to have a deep understanding of statistical models, modelling skills and the way of being a good researcher. All of those will benefit me throughout my entire life.

I would like to express my appreciation to my committee members, Dr. Antonio Páez and Dr. Darren M. Scott, for giving me professional courses and helpful answers when I encountered confusion in the research.

I would like to thank Dr. Hanna Maoh for both your guidance of the establishment of the microsimulation model and evaluation works for the final simulation results. I also appreciate your previous research results, the basis to this thesis.

I would like to thank all my colleagues in the Centre of Spatial Analysis for your great support to my study and research. Special thanks to Justin Ryan for the pleasant days we had and to Tom, who gave me great suggestions during various periods of confusion.

I would like to thank my girlfriend, Pei Yu. Your great encouragement and support helped me overcome all the difficulties.

Thanks to all my friends in McMaster University; I won't have such vivid memories without your friendship. I would like to especially thank Yifei Wang, who gave me great support when I first arrived here.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Research Context

Over the last couple of decades, there has been increasing interest in developing Integrated Urban Models (IUMs) to assist the city planning process. As a crucial component of IUM, the land use model, which is especially crucial for firms, has emerged from various research fields. Recently, the new micro-analytical method has emerged, supplying better understanding of the urban development process (Miller, Hunt, Abraham, & Salvini, 2004).

Since the evolution process of land use of firms is complex, submodels using the concept of firmography are established for analyzing those different phases—including the birth, growth, migration and death processes. Firmography adopts a micro approach to study changes of business establishments over space and time. The firmographic method takes every individual business establishment into account and analyzes its behaviour to get a better understanding of establishments' land use propensity.

So far, there have been several publications focusing on firmographic microsimulation models. Van Wissen (2000) has made a positive start with SIMFIRMS. Other models include UrbanSim (Waddell et al., 2003), ILUMASS (Moeckel, Schürmann, & Wegener, 2002; Moeckel, 2005), SFM (De Bok & Bliemer, 2005; De Bok, 2006; De Bok, 2007; De Bok, 2009), ILUTE (Elgar, Miller, & Farooq, 2008), Hunt's model (Hunt & Simmons, 1993; Khan, Abraham, & Hunt, 2002; Hunt, Khan, & Abraham, 2003), and Kumar's model (Kumar & Kockelman, 2008). Maoh (2005) has also proposed a firmographic microsimulation model framework at the urban scale and has established several key statistical submodels, such as the survival, mobility and location choice submodels.

However, the work for establishing an operational firmographic microsimulation model has not been done yet. Thus, this research fills this gap by establishing an operational firmographic microsimulation model for the City of Hamilton, Ontario, Canada.

1.2 Research Objective

The aim of this research is to implement and test an operational grid-cell level firmographic microsimulation model that can simulate the firmography evolution processes, and store and visualize the simulation results. Based on this operational firmography microsimulation model, different policy scenarios could be projected and discussed to explore the relationship between firmography and urban form.

1.3 Thesis Structure

This thesis consists of five chapters.

Chapter one describes the research context and its objectives.

In chapter two, the overall theoretical background is provided. It begins with the introduction of integrated urban models, followed by an explanation of the firmography concept, which is the core idea of our microsimulation model. Then, a comprehensive understanding of the influential factors in a firmographic model is provided. Finally, we review the literature related to the conventional employment simulation efforts as well as the recent firmographic microsimulation efforts.

Chapter three focuses on the study area and model methodologies. First, we describe the study area, followed by a discussion of the firmographic, spatial and socioeconomic data. The methodology section provides discussions for model structure and data storage issues, followed by a description of the model development circumstance and the user interface. We then discuss the firm age synthesis algorithm, which is used to supplement the incomplete firm micro data. Then, submodel implementations are described for the survival, mobility, location choice and formation submodels.

Chapter four discusses the simulation results. Statistical validation by a simple linear regression analysis was applied to evaluate the simulation accuracy of firm distribution at the census tract level for the year 1997. Then, possible reasons for overestimation and underestimation are discussed for each economic sector.

Chapter five begins with the conclusion of this research, followed by the contributions to the literature. Then, the research limitations are discussed as well as the possible directions for the future research.

Chapter 2 Theoretical Background

2.1 Introduction

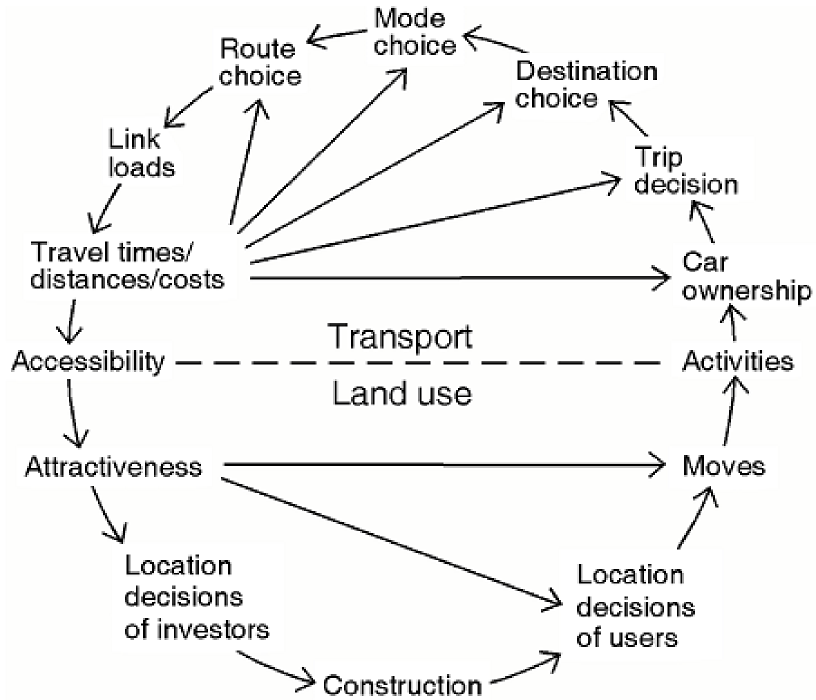
The model developed in this research is one component of the potential micro IUM for the City of Hamilton, Ontario, Canada, thus this chapter first introduces the concept of IUM and microsimulation in section 2.2. Then, the concept of firmography is provided in section 2.3. Section 2.4 introduces each individual firmographic phase along with its influential factors within the simulation framework. Sections 2.5 and 2.6 contain the literature reviews for conventional employment simulation and recent firmographic microsimulation efforts, respectively.

2.2 The Concept of IUMs and Microsimulation

Polycentric urban form, caused by high rates of urbanization as well as rapid suburbanization, has become the dominant urban morphology (Kanaroglou & Scott, 2002). Within this context, commercial and non-commercial behaviours in everyday life are much more complex than ever before.

A considerable body of literature has proposed that changes in land-use patterns influence several aspects of transportation. The variety of land-use patterns in the city affects spatial activities, generating the travel demand. On the contrary, transportation network which carries these spatial activities influences the accessibility of travel destinations. Moore and Thorsnes (1994) indicated that land use and transportation are mutually interdependent. A well-known circular interrelationship between transportation and land use, shown in Figure 2-1, was proposed by Weneger (1995) and has been considered a good source for understanding of the behaviour of actors in the urban system (Kanaroglou & Scott, 2002).

Figure 2-1: The land-use/transportation feedback cycle (Wegener, 1995)



Within these complex circumstances, urban development planning policy can have a wide range of unforeseen or even undesirable indirect consequences (Kanaroglou & Scott, 2002). IUMs are proposed to project and study the consequences of alternative policy decisions before those decisions become reality. Usually, IUMs consist of the several interdependent submodels, in which the land use and transportation models are two essential ones.

Previous efforts for IUMs development were based on the zone level. A number of operational IUMs for the cities around the world are in existence (Southworth, 1995; Miller et al., 2004; Kanaroglou & Scott, 2002; Wegener, 1995; Hunt et al., 2003). However, according to

Miller et al. (2004), those existing models have certain drawbacks: (1) they are aggregate models; (2) they model the entire system in a static manner and lack mechanisms for modelling agents' behaviour; (3) they are weak in importing policy input and responding to a long-term planning process.

Due to the drawbacks to the zone-level models, microsimulation models were proposed. Microsimulation is a modelling technique that operates at the level of agents such as a person, household, vehicle or firm. Within the model, each agent is represented by a record containing a unique identifier and a set of associated attributes. A set of rules (transition probabilities) are then applied to these units, leading to simulated changes in state and behaviour. These rules could be deterministic, in which the probability would be one or zero for a certain event, or stochastic, in which the probability would be between 0 and 1 (exclusive) for an uncertain event, such as the chance of dying or moving within the study area. Through microsimulation, dynamic behaviour for each agent in a certain event could be simulated explicitly over time and space to generate aggregate system behaviour (Miller et al. 2004). In our research, the firmography events are simulated within the microsimulation framework.

2.3 The Concept of Firmography

The interest in firmography has increased recently. According to Van Wissen (2000), firmography encompasses the following processes: firm formation, firm closure, firm growth and decline and firm relocation. He considered firmography to be similar to demography because firms experience processes similar to those faced by human beings: birth and death, as well as aging. Meanwhile, a firm's size and growth, as well as economic activity are major dimensions of the firm population structure, thus any research on firmographic analysis needs to take those factors into account. Firmography considers every firm individually and thereby could analyze those firms' behaviour to gain a better understanding of their land-use propensity. So far, there have been considerable researches making use of the firmographic microsimulation model.

2.4 Influential Factors in Firmographic Evolution Phases

Firmography consists of several subphases, so our research focused on different phases of the entire process.

Regarding the birth of firms, Parker (2009) stated that small firms produced more entrepreneurs, because these firms are more likely to transfer management skills to employees, who are in turn more likely to leave their current job to pursue better-paying employment or become entrepreneurs. As well, Van Wissen (2000) considered that those who were not employed yet, such as new graduates or laid-off workers, will be a source for new firm development. Van Wissen (2000) also considered that, based on the existing firms, new firms would be born according to a certain birth rate, representing events such as one firm splitting off into another, or starting a new branch. This point is quite similar with Parker's (2009) if we consider the situation that those small firms are generating entrepreneurs, especially those who quit their current job and set up new business establishments.

There are three explanations for the phenomenon of small firms producing more entrepreneurs. The first explanation is based on the idea that it is easier for small firms than large firms to transmit pro-entrepreneurship capabilities or attitudes to their employees. Parker

(2009) called this the 'transmission' theory. The second theory applies the concept of self-selection, whereby individuals with particular characteristics are more likely to choose to work for small firms and to engage in entrepreneurship. The third theory takes a 'dual labour market' perspective and argues that peripheral workers who become frustrated by blocked mobility in small firms and are excluded from large firms in the primary labour market are more likely to try self-employment as a last resort. Meanwhile, those relatively bigger firms, which generate new units to extend their profit, will be also a source of new firms. Thus, size is an important factor to be considered. Van Wissen (2004) thought firmography analysis should take carrying capacity or macroeconomic factors into account when modelling the number of new business establishment. Also, Kirchoff and Phillips (1988) suggested that the new firms were the main reason why jobs increased.

From the literature, the birth rate varies a lot among different economic sectors. For example, the retail sector may have relatively higher birth rate while manufacturing has a lower birth rate. Even in the same industry, the birth rate could vary significantly because of the size factor.

On the aspect of firm survival issue, according to Maoh's work (2005), there are plenty of attributes related to closure of firms. Failure of a business can be an involuntary liquidation or a voluntary exit. Those related factors used to explain firms' survival could be summarized into the following generic types: (1) firm-specific, (2) industry specific, (3) macro-economic and (4) spatial or geographic. Firm-specific factors usually contain establishment age, employment size, establishment growth, organizational structure and adaptability to the new technologies.

In particular, young firms tend to fail with a higher probability than large firms (Baldwin, Bian, Dupuy, & Gellatly, 2000). The industry-specific factors imply that different economic sectors present different survival rates. Baldwin et al. (2000) show that high industry growth can lead to more firms in existence. Thus, the industry types could be used as determinant variables. Regarding macro-economic conditions, past studies have used a number of indicators to account for the change in the business cycle over time. The most common indicators are unemployment rate, interest rate, exchange rate and growth rate in GDP. In particular, GDP could be presented by an input-output model, which could be used to compute the carrying capacity mentioned by Van Wissen (2000). Spatial or geological factors are variables relating the individual characteristics of an establishment or its industry to spatial settings. Examples include the distance to the CBD, the establishment's proximity to main transportation routes, or the agglomeration and competition effects in the surrounding urban area (Maoh, 2005).

When it comes to firms' growth and decline, Cassia, Colombelli, and Paleari (2009) state that small, young and independent business grow at the fastest rate. They also demonstrate that the knowledge-sharing relationship between firms and universities can lead to greater business innovation. Furthermore, Cassia et al. (2009) considered that a firm's size can be measured through economic factors other than employment size. Ghosh (2009) used regression to model the firms' growth, taking productivity into account, while also stating that using sales or assets to measure growth could be biased due to inflation. Ghosh (2009) and Van Wissen (2000) considered that previous size and age are important factors relating to the growth or decline of a firm. Nelson and Winter (1982) state that growth depends on investments in research and development, which is more concentrated in larger firms. In other words, Nelson and Winter (1982) argue larger firms tend to grow proportionally more. Another hypothesis (Van Wissen, 2000) considers firms will grow until they mature and reach a saturation level. Furthermore, firms' growth is not only related to their productivity, but also to the greater macroeconomic circumstances.

There are also several factors related to a firm's mobility. According to Maoh's (2005) work on the City of Hamilton, an increase in the size of an establishment would decrease its propensity to move. Mobility also decreases as an establishment ages. Van Wissen (2000) has reported similar results in his model SIMFIRMS: a positive growth rate increases the probability of relocation for certain economic sectors, however, businesses in some sectors (like retail trade) are likely to stay in the current location when they grow. Industry variation and location difference also affect firms' relocation decisions (Maoh, 2005). Furthermore, agglomeration of establishments will increase firm inertia.

When establishments decide to relocate, in most cases they are likely to move into a location with infrastructure and characteristics that help maximize their profit. The importance of the Central Business District (CBD) is well-reported (Shukla, 1991; Waddell & Shukla, 1993; Wu, 1999). The effect of the CBD has been used to test the centralization or decentralization of firms in the urban context. Some kinds of establishments such as retail units prefer to stay close to the CBD while manufacturers, for example, tend to locate on land away from the CBD. Another important factor affecting establishments' location decisions is the nearby transportation infrastructure, such as expressways, the effect of which was tested Shukla and

Waddell (1991). According to Maoh's (2005) work, the proximity to the regional malls in the urban area also has an effect on location choice. Researchers have tested the effects of economies of agglomeration (Shukla & Waddell, 1991; Waddell & Shukla, 1993; Wu, 1999), the concept of which is essential for many researchers' modelling work. Other factors influencing firm location are land use zone (Wu, 1999), as well as the industry and other characteristics such as firm age, size, and organizational structure (Maoh, 2005).

2.5 The Employment Simulation Efforts in Conventional IUMs

According to Timmermans (2003), there are three IUM development waves. The first wave includes the models based on aggregate data and principles of gravitation and entropy-maximization. The second wave contains models based on the principle of utility-maximization. The third wave consists of models using the micro data and micro simulation methods. The first two waves could be considered as conventional IUMs for the reason that they lack the ability to capture firms' behaviour. The third-wave models are still on the way and involve lots of research efforts.

We will first discuss the first wave: aggregate spatial interaction-based models. In 1964, Ira Lowry (1964) introduced the first urban land use model, which has evolved over the past few decades (Horowitz, 2004). The general idea of a Lowry-type model is to distribute a fixed amount of basic sector employment and its service employment to zones in the study area at the very beginning. This allocation changes by a recursive process that finds the minimum population difference between two successive allocation processes. Putman (1983) has tried to extend the original Lowry model to gravity-based models since the early 1970s. Nevertheless, this kind of model does not consider the relationship between the different industries, nor do they simulate the land market clearing processes. The theoretical foundation of such models is based on the assumption that the change in the employment of a zone is based on an attractive measure in the zone, such as the travel time between the residence zone and work zone (Maoh, 2005).

Regarding the second wave, utility-maximizing multinomial logit-based models, Macgill (1977) first tried to formulate the Lowry model as an input-output model that derived its final demand according to entropy maximization. De la Barra, Pérez, and Vera (1984) developed the TRANUS land use model within the framework of a random utility and input-output model. A relatively similar modelling approach was adopted to develop the MEPLAN model (Hunt & Simmonds, 1993).

The two waves mentioned above are considered to lack the ability to model firms' behaviour due to the aggregate nature of the employed models. Thus, researchers have been putting more efforts on microsimulating the firmography

2.6 Firmography Microsimulation Efforts

Despite the drawbacks discussed previously with the first two waves of IUM development, there are very few firmographic microsimulation studies found in the literature (Van Wissen, 2000; Khan et al., 2002; Waddell et al., 2003; Kumar & Kockelman, 2008).

Van Wissen (2004) established the SIMFIRMS model, a regional firmographic model including birth, death, growth, and migration components, to microsimulate firmography evolution process for firms in Netherlands. In particular, SIMFIRMS involves the concept of carrying capacity, which has been used before to model the individual behaviour of the firm, according to Hannan and Freeman's work (1989). The carrying capacity here is utilized as a macro economic indicator. If the carrying capacity is larger than the current number of firms, then more firms will be born and existing firms will grow.

De Bok and Bliemer (2005) used a similar method to Van Wissen (2000) to model the dynamics in the firm population in integrated land use and transport models. Their firmographic approach models the transitions in the state of individual firms by simulating events such as firm migration, firm growth or firm formation and dissolution. For model estimation, de Bok and Bliemer (2005) used LISA data, which provide a longitudinal micro level view of the firm population in the province of South Holland for the period from 1988 to 1997.

The work of Khan et al. (2002) focused on the firmographic behaviour of business establishments in urban areas. In this model, a city is divided into a number of discrete zones which constitute markets where transactions take place. A firm's decision to choose a certain market to make a transaction is decided by a two-level nested logit model. Thus, the probability of choosing a particular market for one specific transaction depends on choosing a destination zone. For migration, the business establishment first decides whether to stay at the existing location, to move to a new location or to move out of the city. If it decides to move, its destination choice will be based on the maximum utility of all possible locations. Khan et al.'s model (2002) also creates new business establishments every year by a fixed growth rate; other components, like the growth and death modules, are still under development.

Kumar and Kockelman (2008) proposed a basic framework for modelling firm demographics by using a microsimulation approach. They started with employment data for the Austin, Texas region, and then seven years of aggregate data from the Statistics of U.S. Businesses were used to simulate firm entry, exit, and evolution over time and space. The authors used a Markov model to anticipate firm growth and contraction, along with logit and Poisson models for firm location choice.

Another effort is ILUTE, by Elgar et al. (2008), who modelled mobility and location choice behaviour of small office firms in the Greater Toronto Area (GTA) in Ontario, Canada, utilizing hazard model for mobility behaviour and the multinominal logit (MNL) model for location choice model. They especially tried to use the anchor concept to reduce the choice set while selecting location alternatives for the MNL estimation process, as this is closer to a firm's relocation behaviour in reality.

Moeckel (2005) simulated the firmography in the urban region of Dormund, Germany, and his research was integrated in the research project ILUMASS. The firms' location choices in his model are represented by logit models. Other firmographic components such as birth, growth, decline and closure of business are simulated by Markov models, since he considered those firmographic decisions were made by entrepreneurs rather than depending on the local development.

Waddel et al. (2003) developed the UrbanSim land use model to microsimulate the change in the location choices of households and firms as well as the change in the real estate market. UrbanSim microsimulates the job population rather than firm population. Although

UrbanSim has never considered taking the firm as the basic unit for the firmographic simulation process, Waddel et al. (2003) considered the job-based method was more suitable for modelling location of jobs of large business establishments. In UrbanSim, the number of new firms and dead firm depends on the economic transition model, which serves to add or remove jobs from the UrbanSim database to achieve consistency with the external assumptions about economic growth or decline in each economic sector over the period of one year. The employment relocation model also predicts the probability that a job in a given sector and location will be moved from that location during any given year, so the employment location choice model predicts a job's probability of choosing a new location from the location choice set. Because UrbanSim takes jobs as the basic analytical unit, it is not necessary to estimate a growth model for firms, and such a model does not exist in the UrbanSim framework. However, Maoh (2005) believed that this is an unrealistic assumption since the firms, and not the jobs, are what drive the location choice behaviour.

Maoh (2005) used principles from firm demography to develop a framework for an agent-based microsimulation model for the City of Hamilton, Ontario, Canada. This framework is used to simulate the evolution of small- and medium-sized business establishments as an outcome of processes related to the birth, failure, growth, mobility and location choice of establishments in the city. The statistical models underlying these processes are discussed and estimated using data extracted from the Statistics Canada Business Register (BR) for the period 1990-2002. The parameters of these models attempt to identify and quantify the causes associated with firmographic processes in Hamilton and will form the basis towards implementing an agent-based microsimulation model.

Chapter 3 Study Area and Methodologies

3.1 Study Area and Data Description

3.1.1 Study Area

This study focuses on business establishments located in the City of Hamilton, which is a port city at the west end of Lake Ontario in an area known as the Golden Horseshoe. The City of Hamilton, as shown in Figure 3-1, contains the “old” city of Hamilton as well as the former municipalities of Ancaster, Dundas, Flamborough, Glanbrook, and Stoney Creek, which are considered the suburban areas in our research. The former city of Hamilton can be considered to be divided geographically by the Niagara Escarpment into two parts. The northern part below the escarpment is adjacent to the west of Lake Ontario, and is also considered to be the old city core both in residential and commercial aspects. The southern part has thrived in the past decades due to the growing population and economic development, and connects to additional highways leading into the United States. The most important traditional economic sector in the city is manufacturing, the employment size of which has experienced gradual decline over time. On the other hand, the service sector has become more significant.

Figure 3-1: The City of Hamilton

3.1.2 Data Description

This study is based on the individual business establishment data from the City of Hamilton in 1990 and 1997. The list of the 1990 business establishments was used as the base year data, while the 1997 firm data were used for model validation purposes. Both datasets have similar data fields including:

- A unique record ID for each business establishment
- Geographical information such as coordinates and municipal name
- Standard Industrial Classification (SIC) for Establishments in terms of SIC name, two-digit code and three-digit code

The only difference between the two datasets is that the 1997 set does not record the employment size information. Unfortunately, business establishment age, an important variable, is not recorded in either one of the datasets.

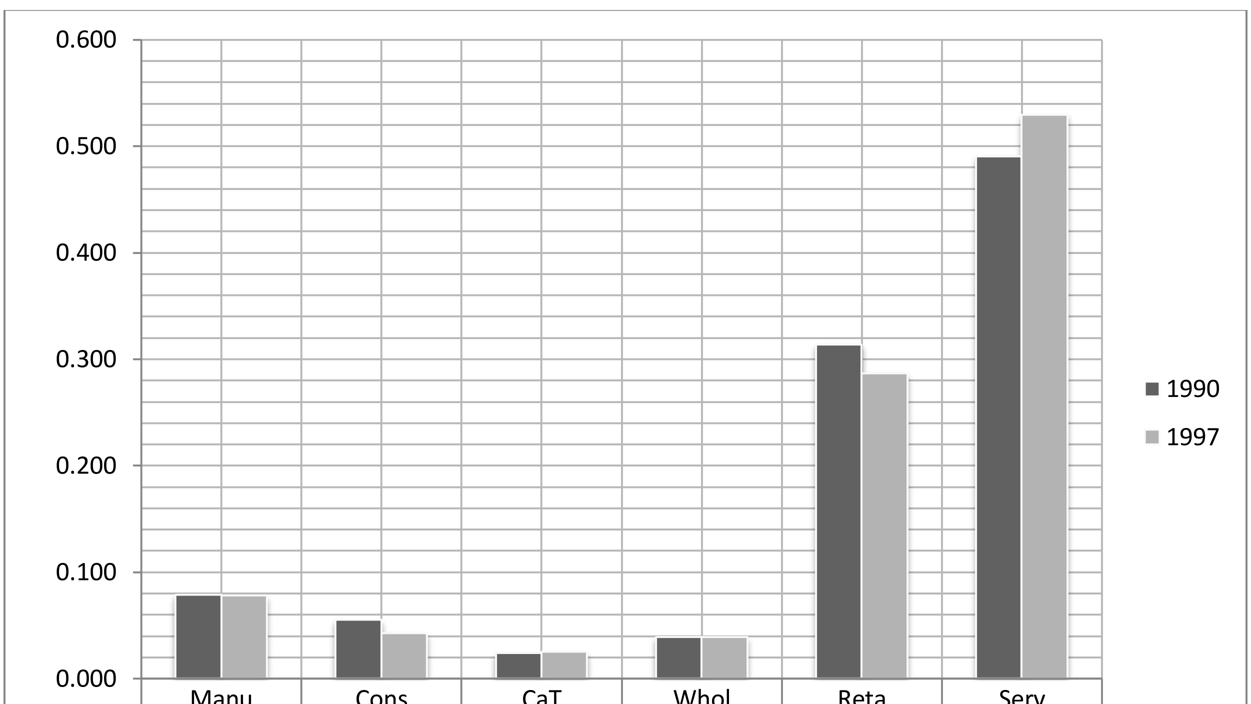
Table 3-1 shows the distribution of the business establishments by sector for 1990 and 1997.

Table 3-1: Distribution of Business Establishments by Sector

	1990	Percentage		1997	Percentage
Manufacturing	905	0.079		928	0.078
Construction	633	0.055		509	0.043
Communication and Transportation	275	0.024		297	0.025
Wholesale	445	0.039		462	0.039
Retail	3604	0.313		3421	0.287
Service	5637	0.490		6319	0.529
Total	11499	1		11936	1

The total numbers of business establishments for 1990 and 1997 are nearly the same; there are 437 more establishments in 1997, an increase of just 3%. However, the distribution of establishments by industry varies for these two years. Figure 3-2 shows that the manufacturing, wholesale and communication and transportation sectors keep almost the same proportion for both 1990 and 1997, whereas construction and retail sectors decreased from 5.5% to 4.3% and 31.3% to 28.7% respectively in 1997. The positive change happened in the service sector, the proportion of which increased from 49% to 52.9%, showing that the service sector has occupied a more important position in economic development.

Figure 3-2: Business Industrial Proportion for 1990 and 1997



	Manu	Cons	CaT	Whol	Reta	Serv
1990	0.079	0.055	0.024	0.039	0.313	0.490
1997	0.078	0.043	0.025	0.039	0.287	0.529

Several spatial data layers, in the form of ESRI Shapefiles, are involved in the firmography microsimulation model:

- The mall layer includes the name and geographic location of the main malls and Business Improvement Areas (BIAs) in the City of Hamilton.
- The major roads and highway layer includes a road net, which exists as a number of line features, extracted from DMTI data (2001), describing the transportation atmosphere in the study area.
- The CBD centroid layer includes a single point feature, centred on the city centre of Hamilton.
- The census tract boundary layer was generated from the 1996 census tracts of Statistics Canada. The layer records the census tract shape as a polygon and also maintains the attribute fields such as census tract ID, municipal name and geographical information. In total, there are 127 records in this file corresponding to 127 census tracts from 1996 in the study area. The geographic information specifies the area of the city: for example, Hamilton – Lower and Hamilton – Upper, which are distinguished by the Niagara Escarpment, and the suburban communities (Ancaster, Dundas, Flamborough, Glanbrook, and Stoney Creek).
- The land use layer was clipped from the DMTI data (2001), which presents seven major land use types in the City of Hamilton: Commercial, Government and Institutional, Open Area, Parks and Recreational, Residential, Resources and Industrial, and Waterbody.

Besides the firmographic data and spatial data, the microsimulation model also includes socioeconomic data such as annual unemployment rate and annual average total household income from Statistics Canada's CANSIM database. These data are available at the census metropolitan area (CMA) level; the Hamilton CMA includes the entire City of Hamilton as shown in Figure 3-1 as well as two nearby municipalities, Burlington and Grimsby.

The annual unemployment rates for the adult population (15 years and older) for the years 1990 to 1996 were extracted from CANSIM Table 282-0053, and are shown in Table 3-2.

Table 3-2: Annual Unemployment Rates in the Hamilton CMA

Year	1990	1991	1992	1993	1994	1995	1996
Unemployment rate (%)	6.2	9.9	10.6	11.8	8.2	6.6	7.4

Table 3-3 shows the arithmetical average total income for all family units in the Hamilton CMA from 1990 to 1996, as extracted from CANSIM Table 202-0401.

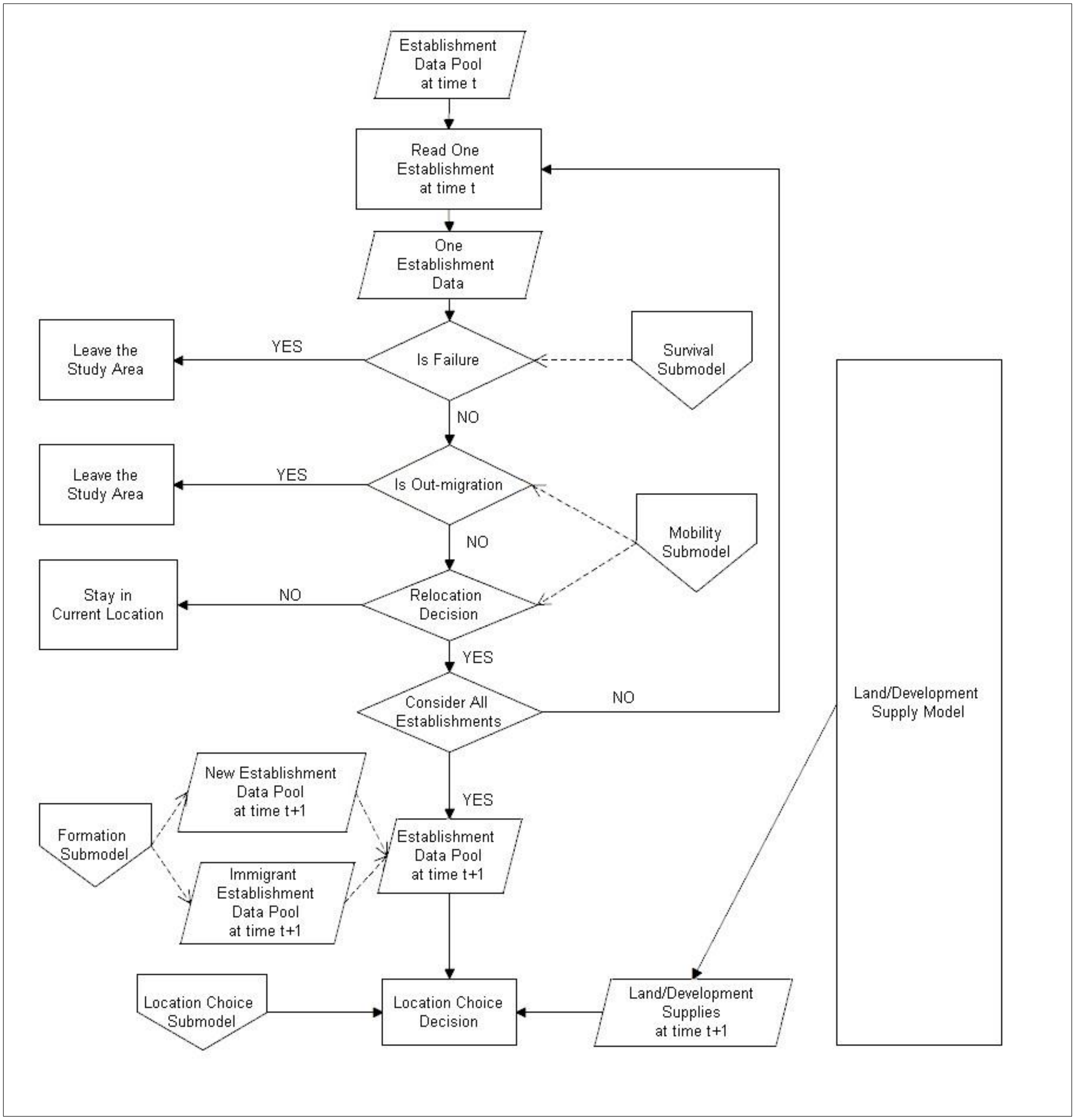
Table 3-3: Annual Average Family Income in the Hamilton CMA

Year	1990	1991	1992	1993	1994	1995	1996
Total income ($)	72,700	64,000	62,400	68,900	65,700	66,500	70,600

3.2 Methodologies

3.2.1 Model Structure and Data Storage

Figure 3-3: Firmographic Microsimulation Model Structure Design



Every year, each existing establishment in the dataset ages, and a new cohort of establishments are established and added to the population. The relatively old firms die out due to their self-capability problems or external conditions, while some establishments move out of the study area and are thus subtracted from our population pool. The rest of the non-stationary establishments in the study area, which include both surviving establishments that have decided to relocate as well as establishments that are just moving into the study area, would choose new locations to maximize their profits. To microsimulate the firmographic event, a set of submodels (survival, mobility, location choice and formation) are established for the entire microsimulation model.

According to Figure 3-3, the microsimulation model is executed through the following steps:

1. An establishment is read from the firmographic dataset at time *t*.
2. The survival submodel decides if this establishment will survive; if not, this establishment will be tagged with a death year at time *t+1*.
3. The mobility submodel deals with potential location decisions of surviving establishments, by deciding an establishment will stay in its current location, relocate within the study area, or leave the study area altogether. Relocating establishments will be tagged with a relocating year at time *t+1*.
4. If all of the establishments in the dataset have been considered, the formation submodel is called to generate new establishments, which are either newborn or migrating into the study area. Otherwise, the model returns to step 1.

5. Once the data pool for the relocating establishments has been formed, the location choice submodel is called to generate a location for each establishment at time *t+1*.

For step 5, we assume that a land/development supply model, which would provide business location units, exists and operates in parallel with our firmographic model. The business location units for each grid cell are estimated by the interaction of the number of business establishments in each grid cell by a supply constraint *n*. The supply constraint *n* here presents a strict or loose land/development circumstance when we set relatively smaller or larger values in the model. The simulation results from using two different scenarios (supply constraints 1.1 and 1.3) are discussed in Chapter 4.

Due to the data availability, we make a strong assumption that the employment size for each establishment is constant during the simulation period. This assumption is made for the reason that our framework so far does not have a growth submodel, which, if it existed here, would estimate the employment size for each establishment for every year in the simulation.

The number of new establishments is generated on the basis of aggregate establishment counts at time *t* and tine *t+1* rather than in a microsimulation way; the formation submodel performs as a macro model for this purpose. The detailed implementation for this submodel is addressed in Section 3.3.7.

With regard to data storage, the firmographic data are stored in the ESRI shapefiles. In each simulated year, a series of new firmographic shapefiles will be created for to record the existing firms. Then, firmographic events as they happen are recorded in those files. In our research, seven new shapefiles are created, one per year from 1991 to 1997. Of particular note is that the dead establishment records would not be removed from the firmographic dataset but labelled with the death year in the field DelFlg. This field did not exist originally in the firmographic dataset and was created to capture more dynamic firmographic information. Another field called Relocation was also added to record the relocating event that happened during the simulation. An age field was also created to keep the synthetic age information from the base year 1990.

Spatial data, such as mall distribution layer, CBD centroid layer, census tract boundaries layer and land use layer, were also kept in a series of shapefiles.

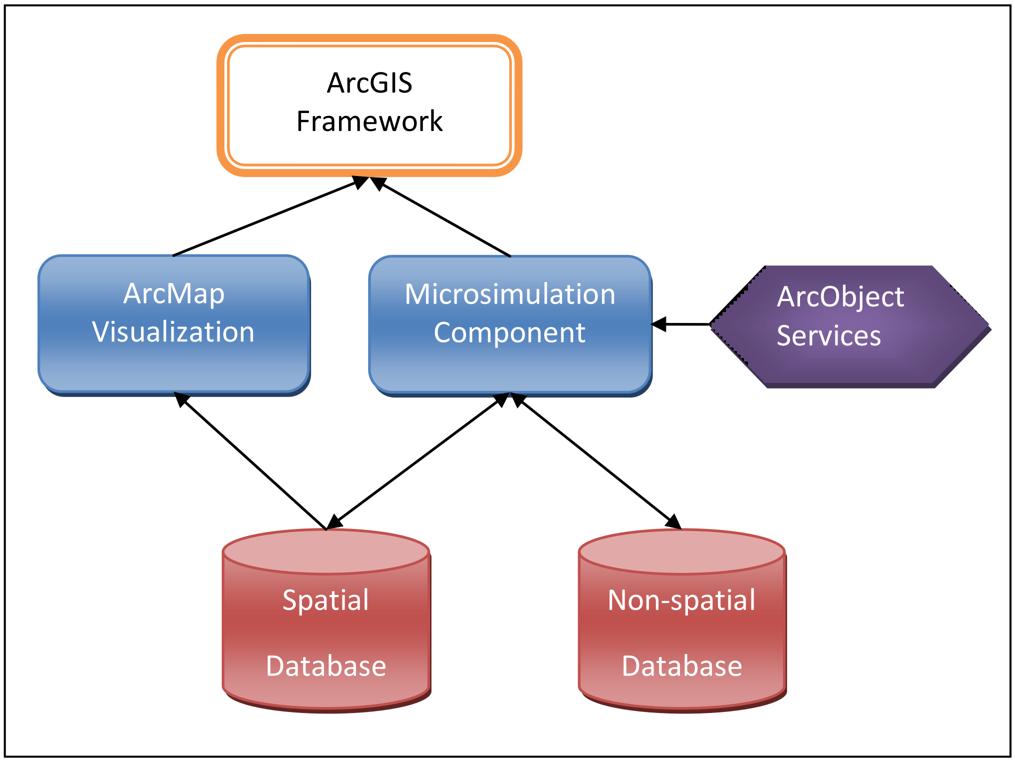
Social economic data, such as annual unemployment rate data and annual average total income, were stored in a Microsoft Access database. Moreover, this Access database also records the total number of firms by economic sector, data used by the formation submodel.

As well, there are a set of text files which store the variable names and parameter values for the statistical models. The age distributions by economic sector and size, which are used to synthesize firms' age, are also recorded in text files.

Finally, all those files or databases are read or updated by the microsimulation add-on for ArcMap.

3.2.2 Development Circumstance and User Interface

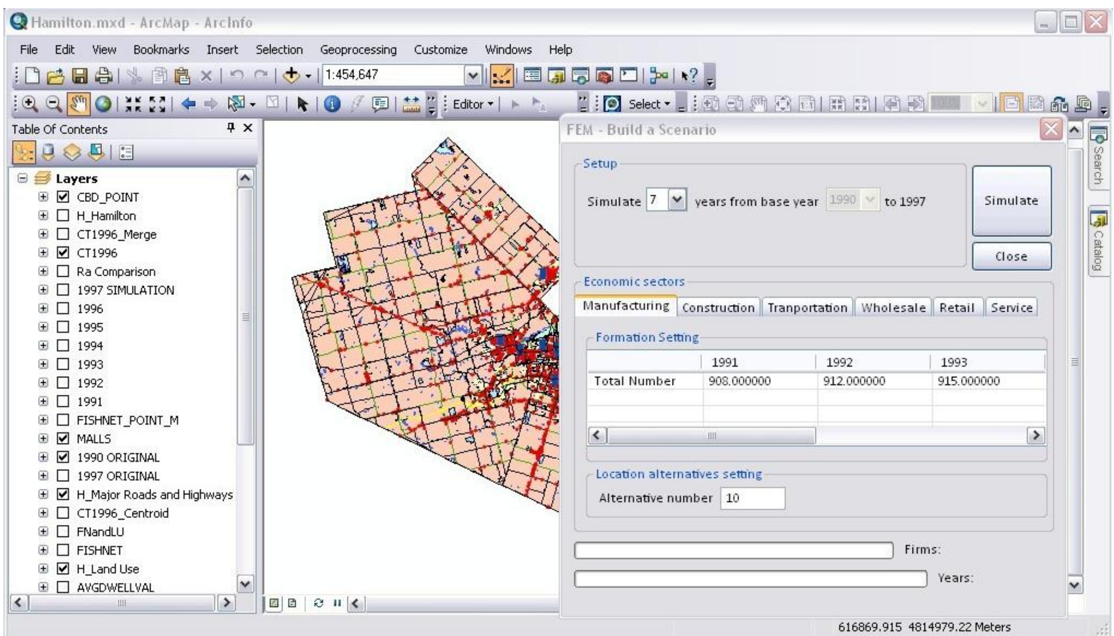
Figure 3-4: Software Structure Design



The microsimulation development is based on Windows and ArcGIS platform support. Figure 3-4 shows the software structure design. The firmographic microsimulation model component implements major firmographic events such as survival, mobility decision, location choice and formation procedure, while utilizing the GIS services from ArcObjects. Spatial and non-spatial data are read and updated by the firmographic microsimulation model component during the simulation period. Then, the model component is seamlessly embedded into ArcGIS framework. Finally, the ArcMap component provides visualization functionality based on updated microsimulation results in the spatial database.

The reason for choosing ArcGIS as the GIS development platform is twofold. First, the original spatial data in this research is stored in Shapefile format. Second, the ArcGIS support add-on development mode is widely used for plenty of GIS development where those applications are to include functionalities of ArcGIS framework. Because our research frequently uses several ArcGIS geoprocessing functionalities, the microsimulation model in the research is in the form of an add-on component to ArcGIS rather than stand-alone software (see Figure 3-5).

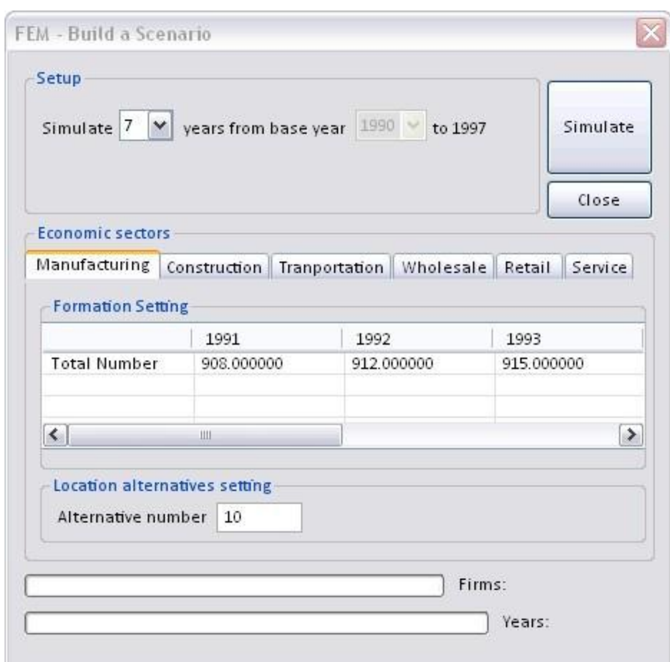
Figure 3-5: Add-on Component for ArcGIS



To develop this add-on for the microsimulation model, the C++ language was chosen, for the reason that it is a compiled language rather than an interpretative language such as Python. This means that the application as developed in C++ will have better execution efficiency than other high level programming languages. As an object-oriented programming (OOP) language, C++ also has a strong ability to describe the various agents in our microsimulation model. Furthermore, C++ has an effective memory management strategy, which could be significantly useful since microsimulation models often consume extreme amounts of computer resources.

The user interface (see Figure 3-6) is contained in a regular pop-up window. Modellers of the firmographic microsimulation model can adjust a set of parameters here before starting the simulation process. The list box in the Setup section allows modellers to set how many years to be simulated. Economic sectors section provides a list control for modellers to view the total number of establishments of each economic sector for each simulation year. The Location alternatives setting box allows the modeller to set the number of alternate locations in the choice set; the default setting is 10.

Figure 3-6: User Interface of Firmographic Microsimulation Model



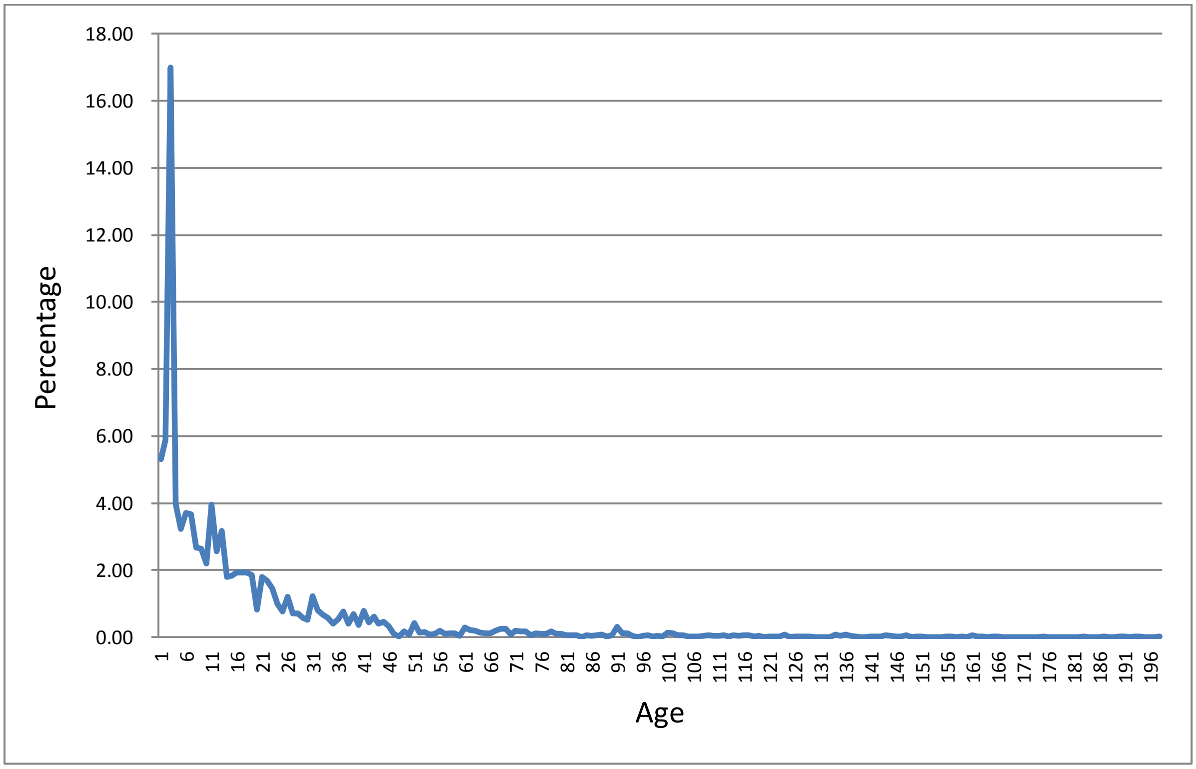
3.2.3 Firm Age synthesis

Our microsimulation model takes 1990 data as the base year; however, our dataset does not include the age information for each firm record. Thus, before microsimulating the firmography, we need to synthesize the age information for every establishment in the dataset.

Two steps are needed to synthesize the firm age information. First, the age distribution by industry and size must be established, because age distribution varies considerably among the different industries as well as among different-sized groups in the same industry. The second step is to utilize the Monte Carlo simulation by drawing from the distribution and comparing against the industry and employment size for each firm. If the random draw value is less than or equal to the threshold probability, the firm will be assigned the corresponding age.

To synthesize the age distribution, we used the 2003 Hamilton Business Directory, which maintains detailed business establishments' information, as of 2003. This dataset contains the establishment year, employment size, industry type, and other related variables. Although it is not a perfect dataset since not all business establishments were registered in that directory, it still gives us a relatively good overview of the age distribution of Hamilton businesses. Firm records without an establishment year were removed from the dataset. As we can see from the figue 3-7, the majority of the firms in the City of Hamilton are less than 50 years old.

Figure 3-7: Age Distribution of Firms, Hamilton, 1990 (Hamilton Business Directory)



According to the 1980 Standard Industrial Classification - Establishments (SIC-E) from Statistics Canada, there are 18 industry types in total. Four of them – Agricultural and Related Service Industries, Fishing and Trapping Industries, Logging and Forestry Industries and Mining (including Milling), Quarrying and Oil Well Industries – were removed from our research since our data from the base year of 1990 did not include any establishments of those types in Hamilton.

On the other hand, those firm records which doesn't maintain the establishment year are also removed from the dataset, because establishment year information is used to compute the age for those establishments.

With regard to size group, the 2003 Hamilton Business Directory includes several size group categories. To be consistent with the previous research, we reorganized those original categories into three new categories: small (1-4 employees), medium (5-49 employees) and large (50-200 employees).

Based on the processes outlined above, we have 14 industry categories, each of which has three size groups. The age synthesis algorithm can then run for each business establishment to generate synthetic age information for the 1990 firm dataset.

3.2.4 Implementation of Survival Submodel

The survival submodel is utilized to determine whether a firm will survive or die in the simulating year. The core of the survival submodel is a statistical model which produces the probability of death for a certain firm. Monte Carlo simulation then generates a random draw; if the random value is smaller or equal to the death probability, this firm is considered "dead." A firm that does not survive would not participate in the simulation in the next year. However, instead of removing those dead firms from the firm dataset, a new field called DelFlg is added to the firm shapefiles. The DelFlg field will be equal to the year the firm died if it did not survive; otherwise, DelFlg is zero. Initially, the DelFlg is set to be zero for every firm in the base year of 1990.

According to Maoh (2005), the statistical model to be used is a binomial logit model:

$$P_{it} = \frac{1}{1 + \exp[-(\alpha + \beta x_{it})]}$$

Here P_{it} is the death probability as a dependent variable for a certain firm *i* at time t, α is the constant, β is a set of coefficients, and x_{it} is a set of independent variables.

The independent variables (see Table 3-4) can be classified into four types of generic factors (Maoh, 2005): (1) firm-specific, (2) macro-economic, (3) industry-specific and (4) spatial or geographic. However, due to the insignificance of the previous research into geography-specific factors (Maoh, 2005), they are not included in the simulation model.

Table 3-4: Coefficients for the Binomial Logit Model for All Industries (Maoh, 2005)

Covariate name	Covariate definition	Value
Constant		6.44990
Firm Specific factors		
Age1	1 if the business establishment is 1 year old	0.99940
Age2	1 if the business establishment is 2 years old	0.78510
Age3	1 if the business establishment is 3 years old	0.73550
Age4	1 if the business establishment is 4 years old	0.55640
Age5	1 if the business establishment is 5 years old	0.31730
Age6	1 if the business establishment is 6 years old	0.26180
$\mathrm{Size(t)_i}$	Employment size of business establishment i in the year t	-0.07150
$\mathrm{Size(t)_i^2}$	Employment size of business establishment i in the year t, squared	0.00176
$\mathrm{Growth(t)_i}$	$\frac{\mathrm{Size(t)_i - Size(t-1)_i}}{\mathrm{Size(t-1)_i}} * 100$	-0.00181
$\mathrm{Relocation(t)_i}$	1 if business establishment i relocated within the City of Hamilton between 1996-2002, 0 otherwise	-0.69850

Macro-economic factors		
UEMR (t)	Annual unemployment rate (%) in the Hamilton region	0.1201
ATINC (t)	Annual average total income (CAD $) in the Hamilton region	-0.00025
Industry specific factors		
$\mathrm{RAZE}\ (t)_i^k$	$\frac{\text{Average employment size of k}}{\text{the number of employees of i (i belongs to k)}}$	0.00573
F (construction)	1 if the business establishment I belongs to construction sector, 0 otherwise	0.1203
H (communication)	1 if the business establishment I belongs to communication sector, 0 otherwise	0.2421
J (retail trade)	1 if the business establishment I belongs to retail trade sector, 0 otherwise	-0.2389
K (finance insurance)	1 if the business establishment I belongs to finance insurance sector, 0 otherwise	0.571
P (health and social services)	1 if the business establishment I belongs to health and social service sector, 0 otherwise	-0.3853
Q (accommodation food and beverage services)	1 if the business establishment I belongs to accommodation food and beverage services sector, 0 otherwise	0.3737

Firm-specific factors include firm age, size, growth, and relocation variables. Firm age is represented as one of six dummy variables (Age1 to Age6), corresponding to the age of the establishment (one year to six-plus years). The coefficient for age variables are all positive, and the coefficients' values are from bigger to smaller, which can be interpreted to mean that younger firms have higher probabilities of death. The firm size factor is described by Size and Size^2; the Size variable has a negative coefficient since smaller establishments fail at higher rates. Size^2 is introduced for a non-linear effect that the size variable may exhibit. The growth variable represents the establishments' performance; however, the input value for this variable is consistently zero due to the lack of a growth submodel. The relocation variable is time covariate to exhibit the effect of death on a change in business location. The coefficient of this variable is negative, which means establishments with a relocating history have a higher probability of surviving.

Due to the insignificance of the previous research (Maoh 2005), the geographic specific factors are not included in the simulation model.

The annual unemployment rate and the annual average family total income are included as macro-economic variables in the model. The Statistics Canada data from 1990 to 1997 shown in Table 3-2 and Table 3-3, are stored in the Access database. During the simulation, the add-on reads those data and applies them to the statistical model.

The industry-specific factors here contain a set of dummy variables that classify the establishments according to the main industry to which it belongs. We also include a variable called RAZE to reflect the effect of industry size. According to the significance of the previous estimation results (Maoh, 2005), only six dummy industry variables are introduced, covering the construction, communication, retail trade, finance insurance, health and social services and accommodation food and beverage services. The value of those dummy variables would be 1 if the establishment SIC belongs to any of those industry groups and 0 otherwise. Another industry-specific RAZE, which is the ratio of the average employment size of the two-digit SIC containing establishment *i* to the employment size of establishment *i*, is introduced in the model. To implement RAZE, we calculate the average employment at the beginning of the every simulation year for each industry. During the simulation, this ratio will be derived dynamically. The sign of the RAZE coefficient is positive here, because a higher RAZE value means a large gap between the average industry employment size and a certain establishment's employment size, which indicates there is room in this industry for that establishment to grow.

3.3.5 Implementation of Mobility Submodel

Some of the survival establishments prefer to relocate their business to maximize their profits. The mobility submodel here is utilized to decide if a firm will relocate in the study area. The core of the submodel is a multinomial logit model which generates the probability for certain establishment i for three mobility choices: stay, relocate and leave. If a firm is chosen to stay in the current location, it will not be taken into the location choice model after the mobility model. If a firm is chosen to relocate, it will be marked as a mover in the study area and a new business location will be designated to this firm. Otherwise, if a firm is chosen to leave, it will be removed from the study area and will not participate in the next year's simulation.

$$P_i(m) = \frac{\exp(V_{im})}{\exp(V_{iS}) + \exp(V_{iR}) + \exp(V_{iL})}$$

Here, $P_i(m)$ is the probability of business establishment i choosing mobility choice m, where m represents one of the mobility choices in the set {Stay, Relocate, Leave}. V_{im} is the utility function of business establishment *i* choosing mobility choice m, whereas V_{iS}, V_{iR} and V_{iL} are respectively the utility functions of establishment *i* choosing to stay, relocate, and leave.

A business establishment's final mobility decision is determined by the Monte Carlo simulation. Based on the probability of each choice generated from the multinomial logit model, a random draw will be applied and business establishment *i* will be assigned a mobility decision (stay, relocate, leave) based on the thresholds for each selection probability.

Mobility characteristics vary among different industries, so there are six multinomial logit models, one each for the manufacturing, construction, communication and transportation, wholesale, retail and services industries. For each industry, the three mobility alternatives are included in the specific multinomial logit model.

For the utility of the "Stay" alternative, we have several independent variables including industry dummy variables, a geographic dummy variable, and an agglomeration variable. Industry dummies indicate the economic sector to which a business establishment belongs, and are equal to 1 whenever an establishment belongs to the industry represented by the variable. Geographic dummies include lower Hamilton and upper Hamilton; if a firm locates in either of them, the value will be 1. An agglomeration variable is used to capture the establishments' preference for clustering with other establishments in the same industry. We implement this variable, with the support of ArcGIS, by creating a circle buffer with a radius of 1500 meters centred on establishment *i*. Then, to get the value for the agglomeration variable, we count the number of establishments from the same industry within the 1500-metre radius of establishment *i*.

For the utility of the "Relocate" alternative, the independent variables of the utility function are size, size squared, the logarithm of age, industry dummies, growth and D_{od}/LC. Those firm specific variable values could be directly read from the firm dataset or by some computing process. The values of the industry dummies depend on which industry is home to a certain business establishment. The covariate D_{od}/LC is a ratio, dividing the average distance from the current location of firm i to the potential moving destinations by a measure of the local competition. Local competition is measured as the ratio of the total employment of establishment *i* rivals to the employment of establishment *I*, where rivals are those establishments with the same two-digit industry as establishment *i*, located within 1500 metres of establishment *i*. The positive sign of this variable for each industry here indicates that if establishment *i* relocates, it will choose a relatively far place to settle down, to gain an easier (i.e., less competitive) business environment.

For the utility of the "Leave" alternative, the independent variables of the utility function are size, size squared and the logarithm of age. Those firm specific variable values could be directly read from the firm dataset or by some simple computing process.

Table 3-5 to Table 3-10 state the detailed covariate values in each industry-specific multinomial logit model.

Table 3-5: Covariates for the Multinomial Logit Model for Manufacturing (Maoh, 2005)

Stay			Relocate			Leave	
Covariate name	Value		Covariate name	Value		Covariate name	Value
Industry1	2.174		Constant	1.378			
Lower Hamilton	2.147		$Size_i$	-0.081		$Size_i$	-0.081
Upper Hamilton	1.821		$Size_i^2$	0.000343		$Size_i^2$	0.000343
Agglom	0.025		Log (Age)	-1.488		Log (Age)	-1.488
			$Growth$	0.778			
			Suburbs	-1.767			
			D_{od}/LC	0.000875			

Table 3-6: Covariates for the Multinomial Logit Model for Construction (Maoh, 2005)

Stay			Relocate			Leave	
Covariate name	Value		Covariate name	Value		Covariate name	Value
Industry 2	1.278		Constant	2.092			
Lower Hamilton	1.155		$Size_i$	-0.572		$Size_i$	-0.572
Upper Hamilton	1.810		$Size_i^2$	0.00725		$Size_i^2$	0.00725
Agglom	0.035		Log (Age)	-1.086		Log (Age)	-1.086
			$Growth$	-0.111		$Growth$	-0.111
			Industry3	0.781			
			Suburbs	-3.515			
			D_{od}/LC	0.018385			

Table 3-7: Covariates for the Multinomial Logit Model for Communication & Transportation (Maoh, 2005)

Stay			Relocate			Leave	
Covariate name	Value		Covariate name	Value		Covariate name	Value
Industry4	1.5464		Constant	1.909			
Industry5	1.6774		$Size_i$	-0.331		$Size_i$	-0.331
Lower Hamilton	1.084		$Size_i^2$	0.001342		$Size_i^2$	0.001342
Upper Hamilton	4.107		Log (Age)	-0.943		Log (Age)	-0.943
Agglom	0.066		$Growth$	-2.146		$Growth$	-2.146
			Suburbs	-3.019			
			D_{od}/LC	0.020842			

Table 3-8: Covariates for the Multinomial Logit Model for Wholesale (Maoh, 2005)

Stay			Relocate			Leave	
Covariate name	Value		Covariate name	Value		Covariate name	Value
Industry6	2.028		Constant	1.642			
Lower Hamilton	1.429		$Size_i$	-0.332		$Size_i$	-0.332
Upper Hamilton	3.508		$Size_i^2$	0.002874		$Size_i^2$	0.002874
Agglom	0.071		Log (Age)	-0.818		Log (Age)	-0.818
			$Growth$	-0.829		$Growth$	-0.829
			Suburbs	-2.053			
			D_{od}/LC	0.003308			

Table 3-9: Covariates for the Multinomial Logit Model for Retail (Maoh, 2005)

Stay			Relocate			Leave	
Covariate name	Value		Covariate name	Value		Covariate name	Value
Lower Hamilton	2.459		Constant	1.481			
Upper Hamilton	2.794		$Size_i$	-0.216		$Size_i$	-0.216
Agglom	0.037		$Size_i^2$	0.002207		$Size_i^2$	0.002207
			Log (Age)	-1.188		Log (Age)	-1.188
			$Growth$	-0.431		$Growth$	-0.431
			Industry7	1.143			
			Suburbs	-4.462			
			D_{od}/LC	0.044522			

Table 3-10: Covariates for the Multinomial Logit Model for Services (Maoh, 2005)

Stay			Relocate			Leave	
Covariate name	Value		Covariate name	Value		Covariate name	Value
Industry8	1.936		Constant	2.083			
Industry9	1.851		$Size_i$	-0.041		$Size_i$	-0.041
Industry10	1.168		$Size_i^2$	0.000162		$Size_i^2$	0.000162
Industry11	2.704		Log (Age)	-0.887		Log (Age)	-0.887
Industry12	1.070		$Growth$	-0.080		$Growth$	-0.080
Industry13	1.911		Suburbs	-2.398			
Lower Hamilton	1.846		D_{od}/LC	0.001739			
Upper Hamilton	2.047						
Agglom	0.008						

3.3.6 Implementation of Location Choice Submodel

Once a business establishment is selected for relocation, the location choice submodel decides where the establishment will locate, considering business owners' desire to maximize their profits. The relocating establishments in any given year include both the ones selected by the mobility submodel, and those new or in-migrating establishments from the formation submodel.

The location choice submodel contains two steps for making a relocating decision (Maoh, 2005). First, a sample of ten locations is randomly selected from the full set of alternative locations. The location choice probability is then computed for the ten locations using the multinomial logit model that corresponds to the industry to which the establishment belongs:

$$P_i(n) = \frac{\exp(V_{in})}{\sum_{m=1}^{m_i} \exp(V_{im})}$$

Here $V_{in}(V_{im})$ is a linear-in-parameter function that depends on covariates characterizing establishment *i* and the attributes of site *n (m)*. m_i is the number of alternative locations for the business establishment *i*, which is set to 10 in our model. After calculating the location probabilities, Monte Carlo simulation is used to select the specific alternative location to be assigned to establishment *i*.

For each of the six industries introduced in Section 3.3.5, there is a multinomial logit model for computing alternative locations' probabilities (see Table 3-11 to Table 3-16).

.

Table 3-11: Covariates for the Multinomial Logit Model for Manufacturing (Maoh, 2005)

Covariate name	Value
CBDPRO	0.000110
HWYMRPRO	0.663567
AGGLOM1	0.047804
POPDENS	-0.000258
LANDUSE2	0.750878
NEWBORN*CBDPRO	-0.000058
INDUSTRY1*CBDPRO	-0.000110
INDUSTRY2* POPDENS	0.000339

Table 3-12: Covariates for the Multinomial Logit Model for Construction (Maoh, 2005)

Covariate name	Value
CBDPRO	0.0000970
HWYMRPRO	-0.362410
MALLPRO	-0.646879
AGGLOM2	0.080863
MOUNTAIN	0.783139
NEWDEVLOP	0.004611
OLDDEVLOP	-0.001936
AVGDWELLVAL	-0.000002
HHLDDENS	0.000801
LANDUSE1	1.559011
INDUSTRY3*HWYMRPRO	0.806584
INDUSTRY3*MOUNTAIN	-0.891513
INDUSTRY4*LANDUSE1	-0.918098
INDUSTRY5*LANDUSE1	1.753567

Table 3-13: Covariates for the Multinomial Logit Model for Communication & Transportation (Maoh, 2005)

Covariate name	Value
CBDPRO	0.000092
HWYMRPRO	0.580206
MALLPRO	-0.959108
AGGLOM3	0.234618
LANDUSE1	0.984063
LANDUSE2	2.136029
INDUSTRY6*LANDUSE2	-2.205140
INDUSTRY7*CBDPRO	-0.000136

Table 3-14: Covariates for the Multinomial Logit Model for Wholesale (Maoh, 2005)

Covariate name	Value
CBDPRO	0.000061
HWYMRPRO	0.498985
AGGLOM4	0.094018
SUBURBS	0.618772
NEWDEVLOP	0.002791
INDUSTRY8*CBDPRO	-0.000159
INDUSTRY9*HWYMRPRO	-0.763197

Table 3-15: Covariates for the Multinomial Logit Model for Retail (Maoh, 2005)

Covariate name	Value
HWYMRPRO	0.376421
AGGLOM5	0.022177
SUBURBS	0.426167
NEWDEVLOP	0.003345
OLDDEVLOP	-0.000629
HHLDINCDENS	0.000002
LANDUSE4	-0.559876
LANDUSE5	1.152924
INDUSTRY10*MALLPRO	0.650485
INDUSTRY11*HHLDINCDENS	-0.000002
INDUSTRY12* LANDUSE4	0.643616
INDUSTRY13* SUBURBS	-0.425008

Table 3-16: Covariates for the Multinomial Logit Model for Services (Maoh, 2005)

Covariate name	Value
CBDPRO	0.000055
HWYMRPRO	0.332224
AGGLOM6	0.005285
HHLDINCDENS	0.000001
LANDUSE1	0.661705
LANDUSE3	-0.771938
LANDUSE5	0.628379
LANDUSE6	0.398759
INDUSTRY14*CBDPRO	-0.000028
INDUSTRY15*MOUNTAIN	0.286058
INDUSTRY16*NEWDEVLOP	0.001207
INDUSTRY17*NEWBORN*HWYMRPRO	0.882275
INDUSTRY18*NEWBORN*HWYMRPRO	0.546541

The independent variables in the logit model contain two categories: location-specific factors and firm-specific factors.

There are a number of location-specific factors in the logit model. One of them is the distance to the CBD which is measured by the Euclidean distance between the CBD core and the centroid of the grid cell containing establishment *i*. Another factor is proximity to highway;

the value of this dummy covariate is 1 if the grid cell intersects with highways or main roads in the city, and is otherwise 0. The mall proximity dummy plays a significant role in the model. First we generate a circle buffer around the centroid of the grid cell within 1500 meters, then if there are one or more malls within the buffer, the value of the dummy would be 1 (0 otherwise).

There are six agglomeration variables, one for each industry models. Agglomeration variables are used to capture establishments' preference for clustering with other establishments in the same industry. Such agglomeration usually plays a positive role in a firm's survival. To implement this variable, we count the number of other establishments from the same industry as establishment *i* that are located within 1500 metres of establishment i.

The MOUNTAIN and SUBURBS dummy variables are introduced in the model to account for the fact that establishments from different industries have different geographic preferences. The SUBURBS variable is 1 for any location within the suburban communities, and the MOUNTAIN variable similarly corresponds to the city above the escarpment, as shown in Figure 3-1. These two covariates are implemented by a pre-set GIS layer which contains the geographical information for each grid cell. The geographical information for this grid cell is retrieved from both the original 1990 firm dataset and the 1996 census tract layers.

The values of another six factors, which are NEWDEVLOP, OLDDEVLOP, POPDENS, HHLDINCDENS, HHLDDENS and AVGDWELLVAL, are retrieved from 1996 census tract data. The first step is to geocode the census tract value to the centroid of each census tract, stored as a layer in ArcGIS. The second step is to make use of the kernel estimation functionality of ArcGIS:

a continuous surface of population density is generated for each covariate, and this surface is then superimposed on the grid cell layer. The grid cell layer will then contain six fields storing the values of each of these covariates.

A set of land-use factors exists in the submodel. The land use categories are derived from DMTI data which include six land types as follows: open space, resource and industrial, park and recreational, residential, commercial, and governmental land use. The value of the covariates is represented by the percentage of the grid cell area assigned to each land use type. To obtain the land use percentage for each grid cell, we over-clipped DMTI land use data against grid cells to generate a number of new small land use tiles. According to the original grid cell ID stored in these tiles, we are able to find the base grid cell for each tile and get the percentages, which are then stored in the grid cell fields. Thus, during simulation, the value of those covariates is read directly from the dataset, to save time compared to computing dynamically.

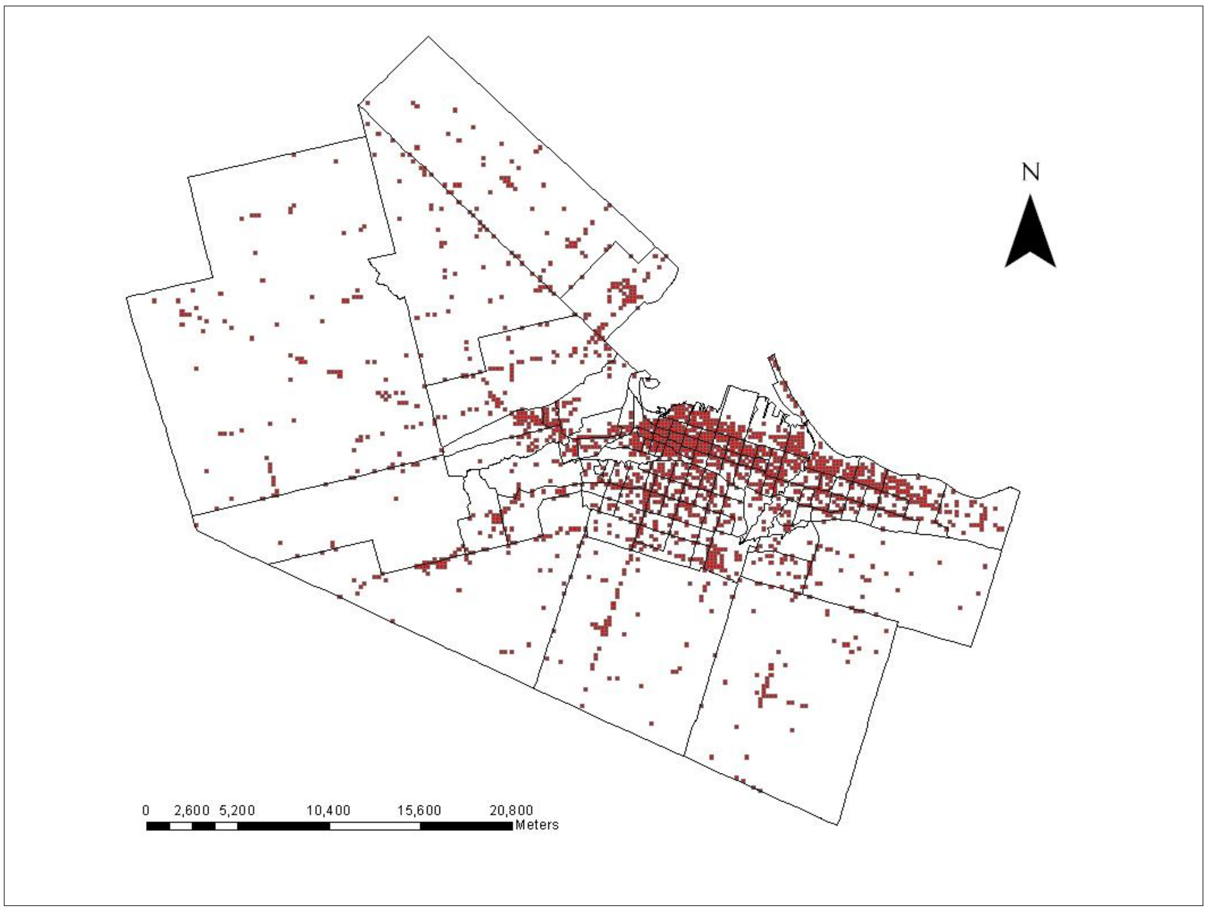
For the firm-specific factors, a dummy variable (NEWBORN) is introduced and set to be 1 if an establishment is newly formed (0 otherwise), and a number of industry dummy variables (see Table 3-17) are involved in the submodel to represent the different location preferences for establishments from different industries. The value of those variables is specified by SIC code from each establishment in dataset. Furthermore, the interaction terms in the model are derived by multiplying the corresponding variables together.

Table 3-17: Industrial Covariate Definitions (Maoh, 2005)

NAME	SIC CODE	Industry
Industry 1	SIC 28	Manufacturing
Industry 2	SIC 28, 39	Manufacturing
Industry 3	SIC 426	Construction
Industry 4	SIC 401, 421, 422 or 427	Construction
Industry 5	SIC 424	Construction
Industry 6	SIC 45 - 47	Communication & Transportation
Industry 7	SIC 48 - 49	Communication & Transportation
Industry 8	SIC 52	Wholesale
Industry 9	SIC 59	Wholesale
Industry 10	SIC 60, 61 or 65	Retail
Industry 11	SIC 61, 62 or 63	Retail
Industry 12	SIC 64 or 65	Retail
Industry 13	SIC 601, 633, 635 or 641	Retail
Industry 14	SIC 70 – 74, 77, 91 – 92, 96 – 99	Services
Industry 15	SIC 96 - 99	Services
Industry 16	SIC 75 – 76, 77, 91 – 92, 96 – 99	Services
Industry 17	SIC 86	Services
Industry 18	SIC 91 - 92	Services

To generate location alternatives, we first divide the whole study area into grid cells of 200 by 200 meters (Maoh, 2005) with the support of ArcGIS. And then we intersect those grid cells with the spatial coordinates occupied by the existing establishments in 1997. In doing so, we gain 2066 grid cells (see Figure 3-8) as the location alternatives.

Figure 3-8: Location Alternative Distribution



3.3.7 Implementation of Formation Submodel

The formation submodel in the framework is used for simulating the number and distribution of new business establishments in the study area.

Unlike human fertility, firm formation is not a straightforward concept. The process of the birth of a new firm is complex. According to the literature (Van Wissen, 2000), there are mainly two reasons a new business firm is established. The first theory relates the number of births to the existing population of firms, which leads to the concept of firm formation rate. This approach assumes that it is the firm that is at risk of giving birth to a new firm, for instance by splitting off, starting a new branch or establishment, and so on. The second theory contends that new business establishments are created by individuals, either after they graduate from school and enter the workforce, or while unemployed. Important attributes of future entrepreneurs include age, sex, and education. The state of the economy also has an influence on firm formation.

However, with our data restrictions, the formation submodel in current framework could not be implemented as a microsimulation model. Thus, we established a macro model for both newly-formed and in-migrating establishments.

The business establishment data we have for 1990 and 1997 were obtained from the City of Hamilton. These two datasets contain basic attributes for establishments, such as the geographic information, employment size and industry information. However, for the microsimulation framework, the time step is one year, thus we have to make an assumption that the number of total establishments between 1990 and 1997 increases or decreases uniformly for each industry. The number of establishments in each industry in each year from 1991 to 1996 is then defined as follows:

$$n_{i,y} = n_{i,y-1} + \frac{n_{i,1997} - n_{i,1990}}{7}$$

where *i* is one of the six industries listed in Table 3-1 and *y* is the year.

Using this method, we can estimate the total number of total establishments for each industry for each year from 1991 to 1996, inclusive (see Table 3-18).

Table 3-18: Number of establishments by industry and year

	1990	1991	1992	1993	1994	1995	1996	1997
Manu	905	908	912	915	918	921	925	928
Cons	633	615	598	580	562	544	527	509
CaT	275	278	281	284	288	291	294	297
Whol	445	447	450	452	455	457	460	462
Reta	3604	3578	3552	3526	3499	3473	3447	3421
Serv	5637	5734	5832	5929	6027	6124	6222	6319

Based on the total number of establishments in each industry by year we are able to derive the number of new establishments by the following equation:

$$New_{i,y+1} = N_{i,y+1} - N_{i,y} + Dead_{i,y+1}$$

where *i* is the industry index and *y* is the year index. $New_{i,y+1}$ is the number of new establishments in industry *i* in the time period from *y* to *y+1*. $N_{i,y+1}$ and $N_{i,y}$ are the numbers of establishments in industry *i*, year *y+1* and *y* respectively. $Dead_{i,y+1}$ is the number of dead establishments in industry *i* in in the time period from *y* to *y+1*.

The basic attributes of newly-formed establishments in year *y+1*, such as employment size and industry information, will be randomly drawn from the establishment dataset set in year *y*, which includes both surviving and dead establishments. The age of each newly-formed establishment will be set to be 1.

The new business establishments in the study area will also need to be assigned locations. Thus, after we synthesize their attributes as outlined above, the formation submodel will call the location choice submodel to simulate their location decisions. Once an establishment has chosen a location, its geographic information will be stored in the establishment dataset.

Chapter 4 Results and Discussions

4.1 Simulation Results Validation by Simple Linear Regression Analysis

The City of Hamilton business establishment micro data for the year 1997 is used to validate the microsimulation results at the census tract level. 1996 census tract data, which includes 127 census tracts in the City of Hamilton, was involved for the purpose of defining the geographic boundaries.

The number of business establishments was first aggregated to the 1996 census tract level and 1997 economic sector for both observation and simulation results. Then, a simple linear regression model, shown in the following equation, was used to evaluate how well the simulated results fit the observed data.

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Here *y* is the dependent variable, representing the simulated number of business establishments in each census tract, and *x* is the observed number of business establishments in that census tract. If the simulation results are perfect, all the *x*-*y* pairs should lie on the line $y = x$. In other words, the accuracy of the simulation could be tested by how close the slope parameter β_1 is to 1 and the constant parameter β_0 is to 0.

This linear regression analysis method is applied to all the industries together as well as separately to each of the six major economic sectors: manufacturing, construction, communication and transportation, wholesale, retail and service industries.

The overall simulation results (see Table 4-1), obtained by applying supply constraint 1.1 to the firmographic model, are encouraging.

Table 4-1: Regression Results between Simulation Results and Observation Data by Supply Constraint 1.1, Hamilton, 1997

Covariate	All Industries	Manufacturing	Construction	Communication & Transportation	Wholesale Trade	Retail Trade	Services
β_0(Intercept)[a, c]	**- 1.6661 (-0.48)**	**0.0916 (0.20)**	**- 0.3616 (-0.67)**	**1.7093 (5.35)**	**1.1967 (2.57)**	**- 2.8744 (-1.20)**	**2.223 (0.78)**
β_1[b, c]	**1.0223 (0.83)**	**0.9875 (-0.49)**	**1.2081 (3.19)**	**0.2691 (-13.023)**	**0.6689 (-6.87)**	**1.1055 (1.78)**	**0.9552 (-1.12)**
No. of Observation	**127**	**127**	**127**	**127**	**127**	**127**	**127**
R^2	**0.9217**	**0.923**	**0.7336**	**0.1553**	**0.6065**	**0.7360**	**0.8195**

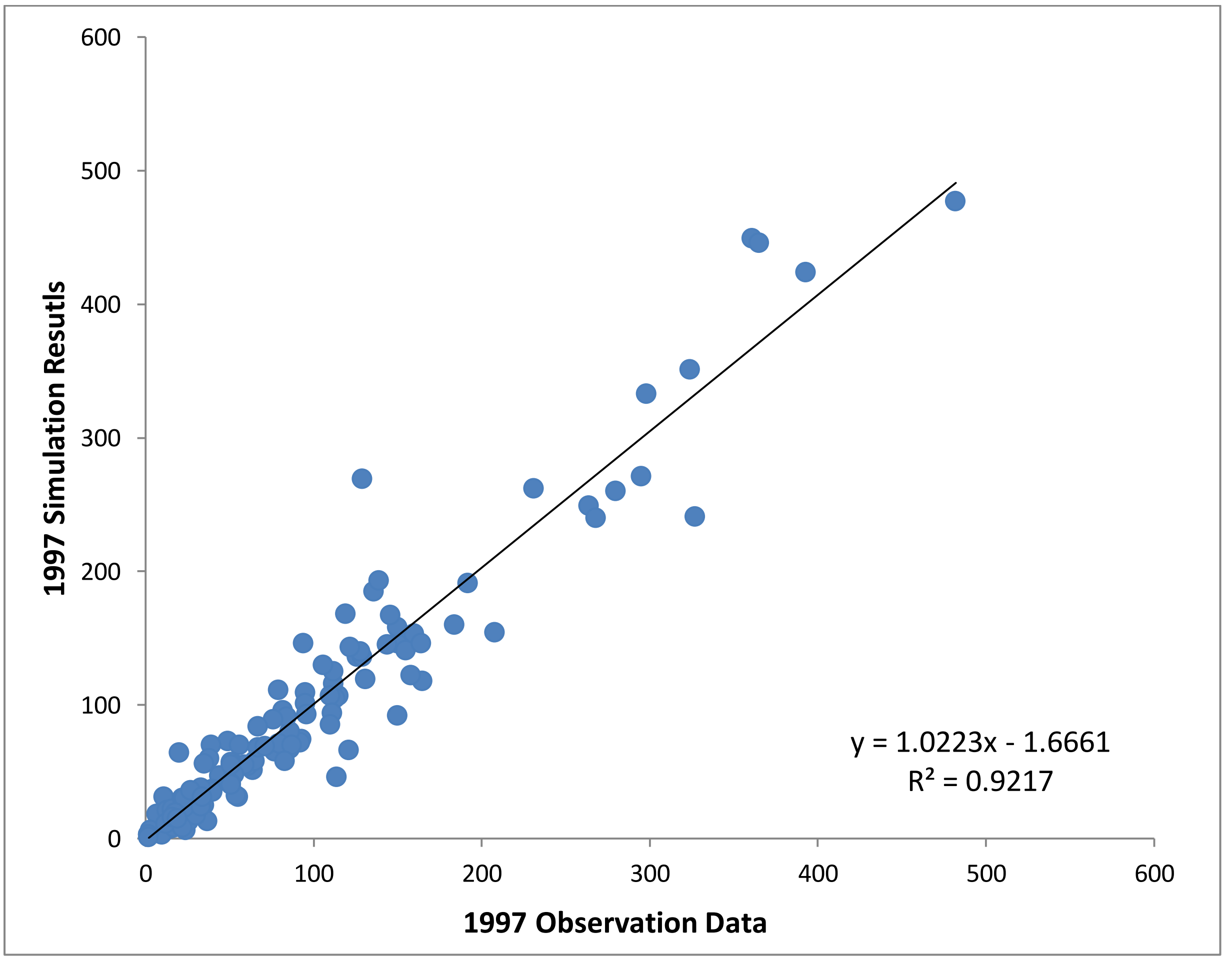
a. In parentheses, the table presents the estimated coefficients with their t-stats related to the H_0 hypothesis that the Intercept (β_0) is 0

b. In parentheses, the table presents the estimated coefficients with their t-stats related to the H_0 hypothesis that the slope (β_1) is 1

c. The critical value is 1.97 for the t distribution with 126 degrees of freedom at 0.05 level of significance

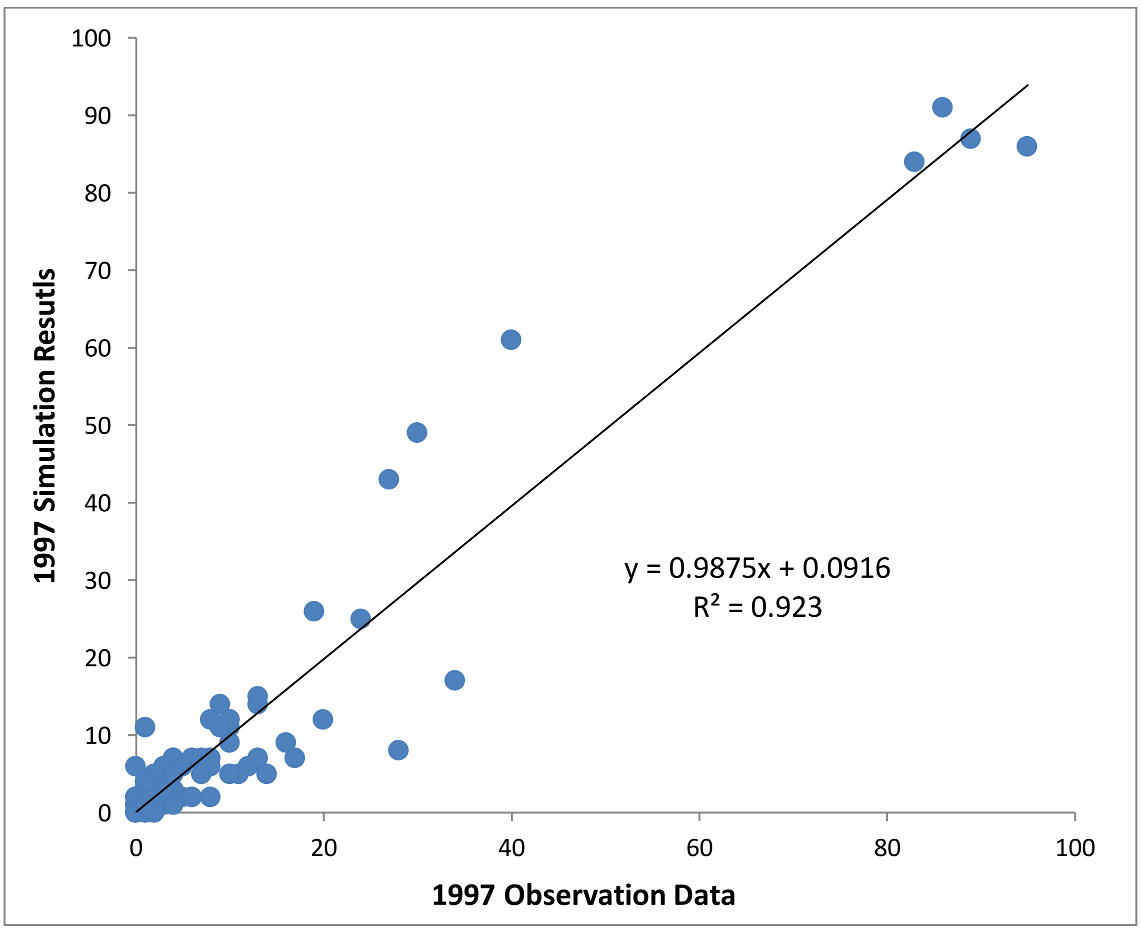
For all industries, the scatter plot in Figure 4-1 shows that the majority of points are distributed along the regression line. From Table 4-1, the slope β_1 of this line is 1.0223 and the intercept (the constant β_0) is - 1.6661. The t-stats indicate that β_1 is insignificantly different than 1 at the 0.05 significance level while β_0 is insignificantly different than 0. Meanwhile, the R^2 is 0.9217. All of those statistical values show that the results from our microsimulation model can reasonably explain the relocation of business establishments at the census tract level.

Figure 4-1: The Business Establishment Distribution Correlation between Simulation Results and Observation Data for All Industries, Hamilton 1997



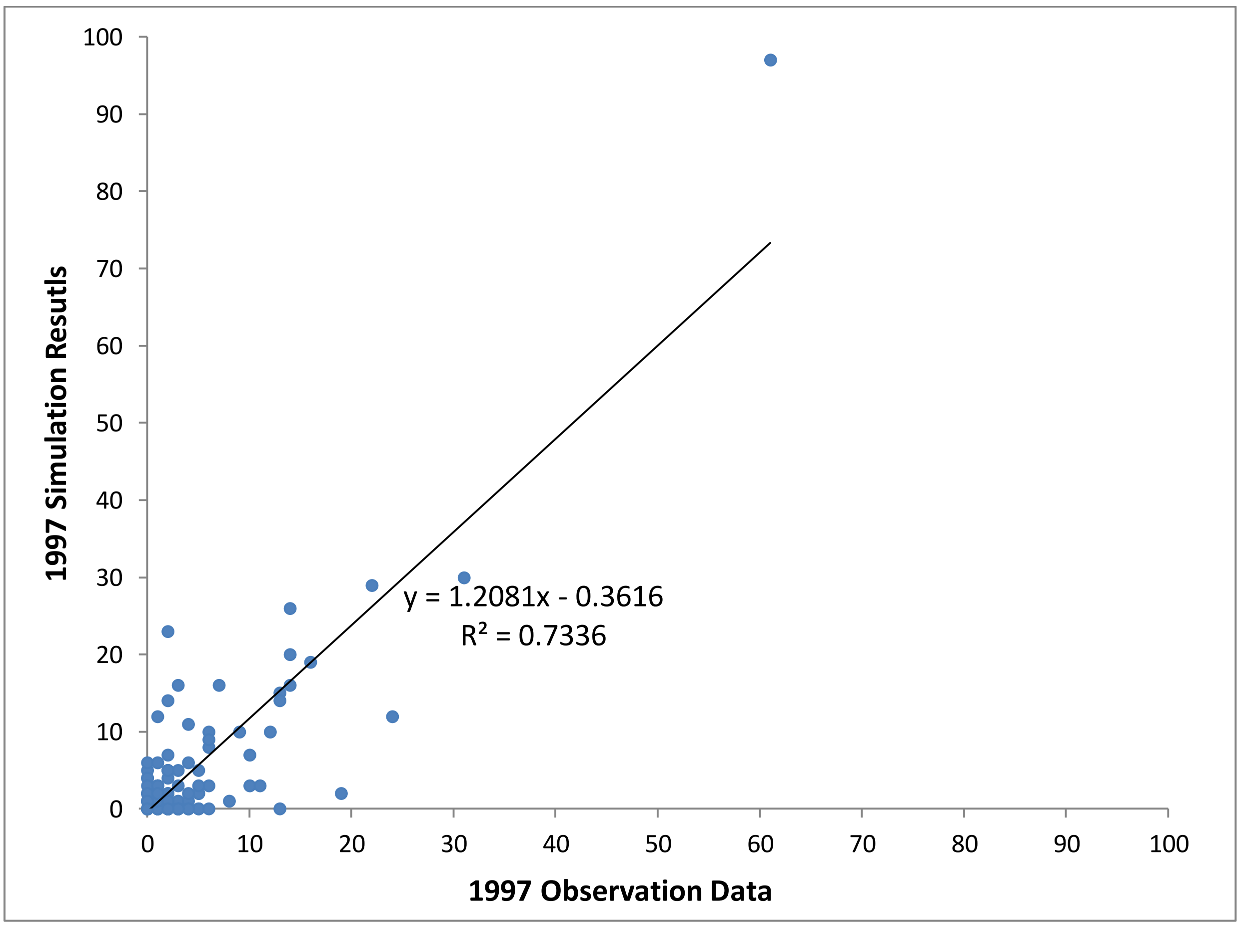
For manufacturing, the scatterplot in Figure 4-2 shows that most points are distributed along the regression line. From Table 4-1, the slope β_1 is 0.9875 and the constant β_0 is 0.0916. The t-stats indicate that β_1 is insignificantly different than 1 and β_0 is insignificantly different than 0, suggesting that the simulation results have correlated well with observed data along the line $y = x$. Meanwhile, the R^2 value is 0.923 indicating that our microsimulation model has explained well the relocation of manufacturing business establishments at the census tract level.

Figure 4-2: The Business Establishment Distribution Correlation between Simulation Results and Observation Data for Manufacturing Industry, Hamilton 1997



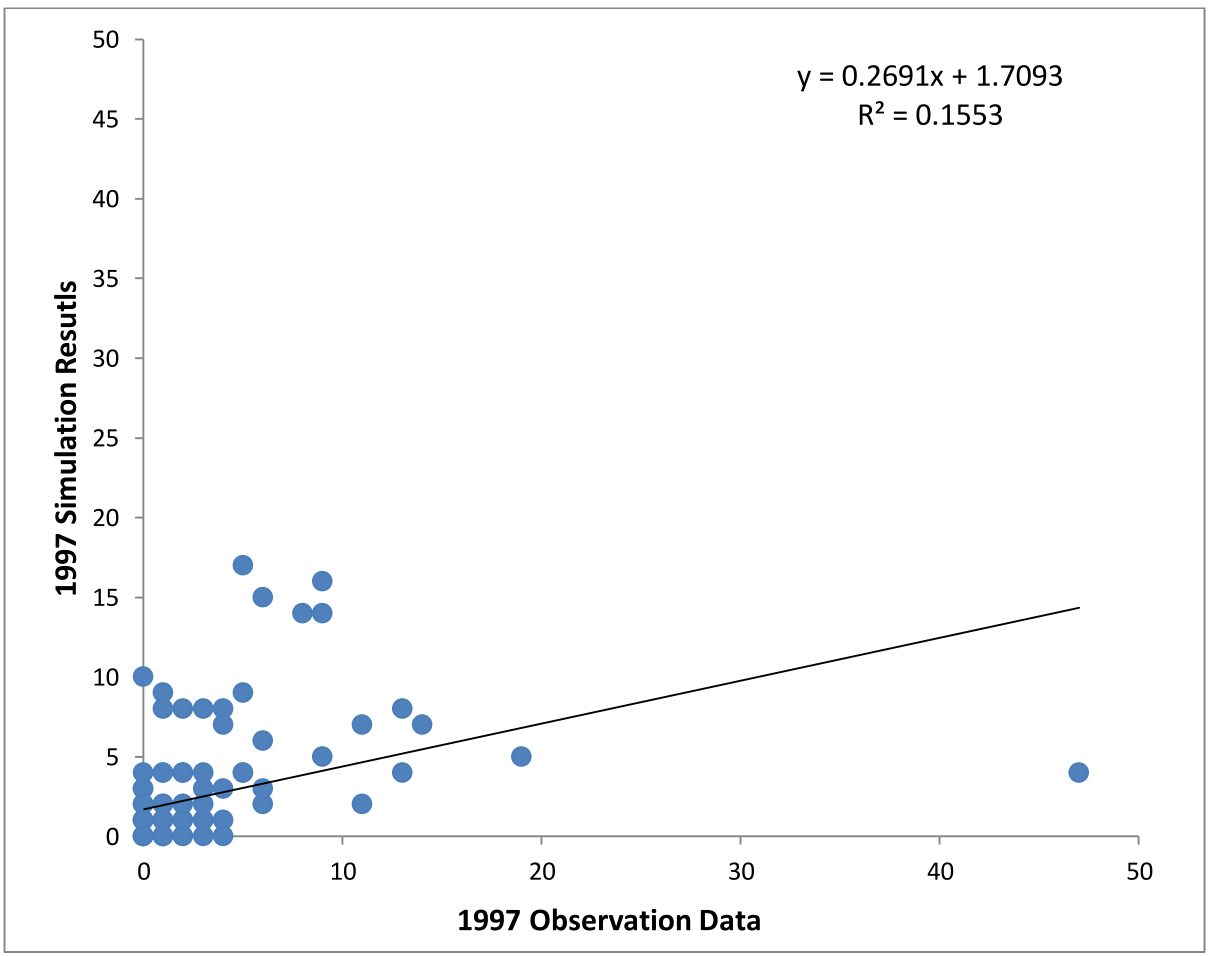
For the construction industry, the scatterplot of which is in Figure 4-3, the regression coefficient β_1 is 1.2081 and the constant β_0 is 0.3616 from Table 4-1. β_1 is very close to 1 while β_0 is close to 0, but the t-stats indicate that only β_0 is insignificantly different than 0. This suggests that the slope β_1 does not distribute along $y = x$ very well. The R^2 of 0.7336 here is also relatively low. These statistical values demonstrate our model has relatively low simulation capability for establishment relocation in the construction industry.

Figure 4-3: The Business Establishment Distribution Correlation between Simulation Results and Observation Data for Construction Industry, Hamilton, 1997



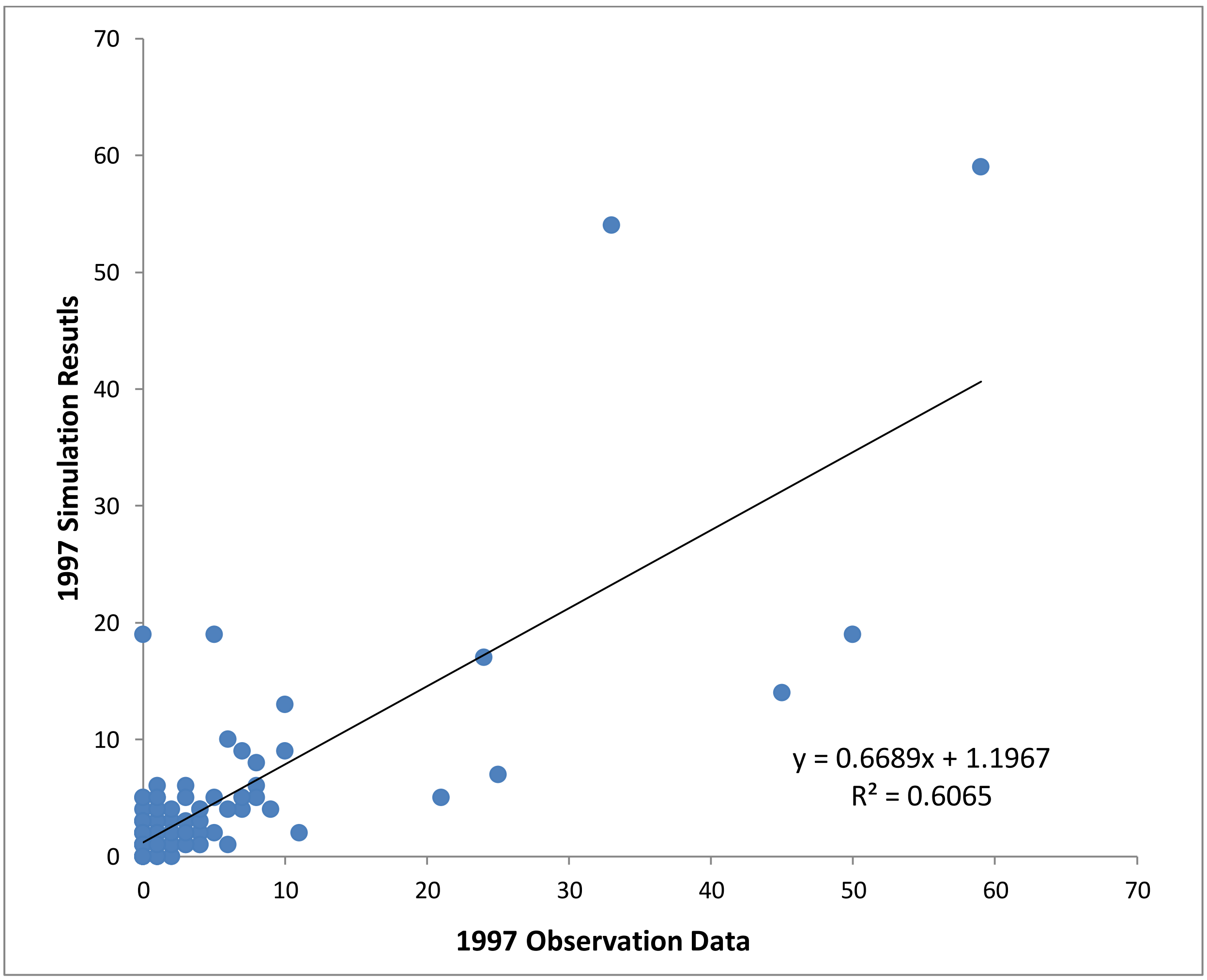
For the communication and transportation industries, Figure 4-4 shows that most points do not distribute along the $y = x$ line. From Table 4-1, the slope β_1 is significantly different than 1 and constant β_0 is significantly different than 0, and the R^2 of 0.1553 is very low. All of those statistical values indicate that our model failed to simulate the relocation activities for this industry sector.

Figure 4-4: The Business Establishment Distribution Correlation between Simulation Results and Observation Data for Communication & Transportation Industries, Hamilton, 1997



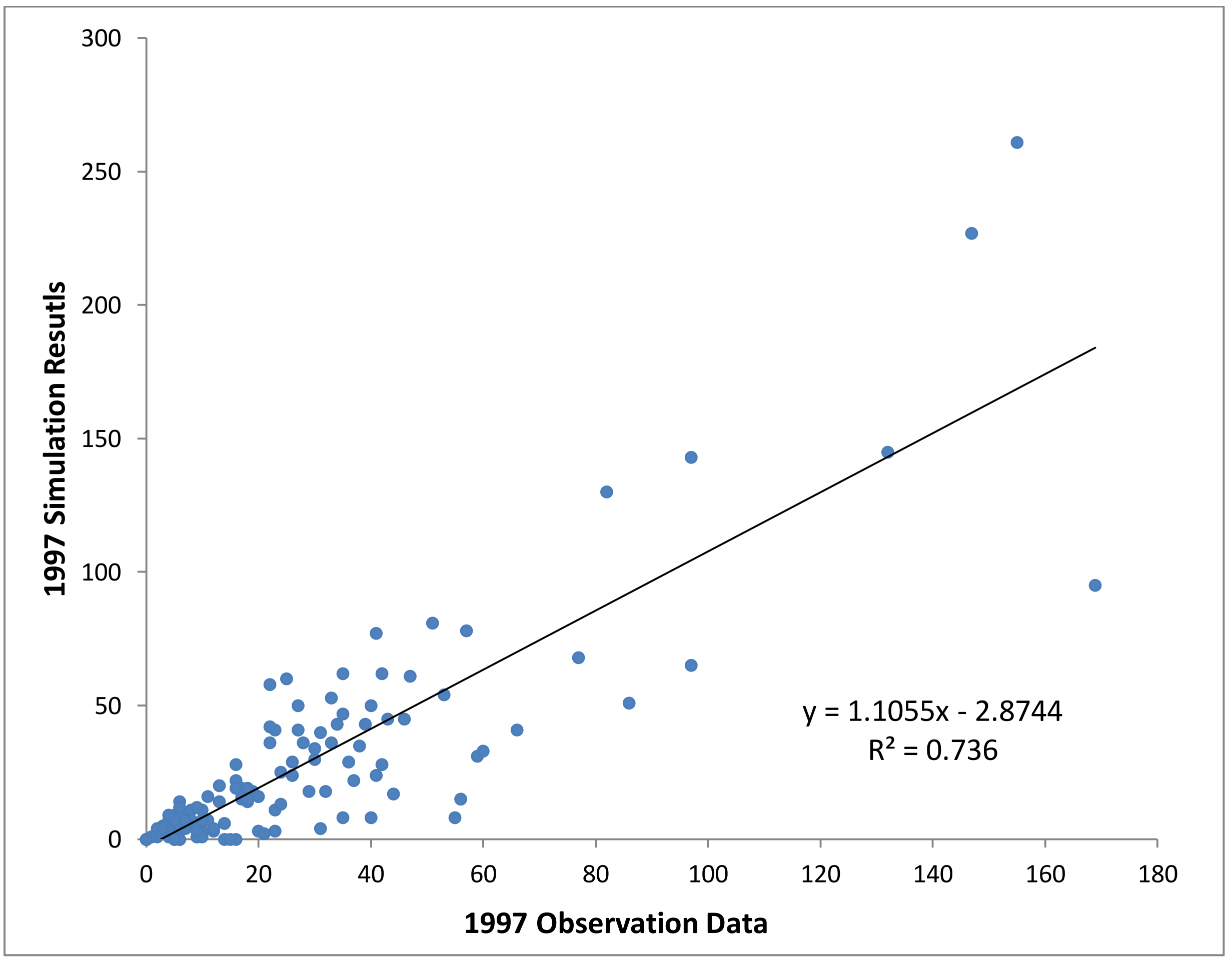
For the wholesale trade industry (see Figure 4-5), only a low percentage of points are distributed along the $y = x$ line. From Table 4-1, the slope β_1 is 0.6689, which is significantly different than 1, and the constant β_0 is 1.1967, which is also significantly different than its desired value of 0. This suggests our model produced simulation results which have low correlation with the observed data. Meanwhile, the R^2 of 0.6065 also indicates inaccurate simulation results were generated from our wholesale industry model.

Figure 4-5: The Business Establishment Distribution Correlation between Simulation Results and Observation Data for Wholesale Industry, Hamilton, 1997



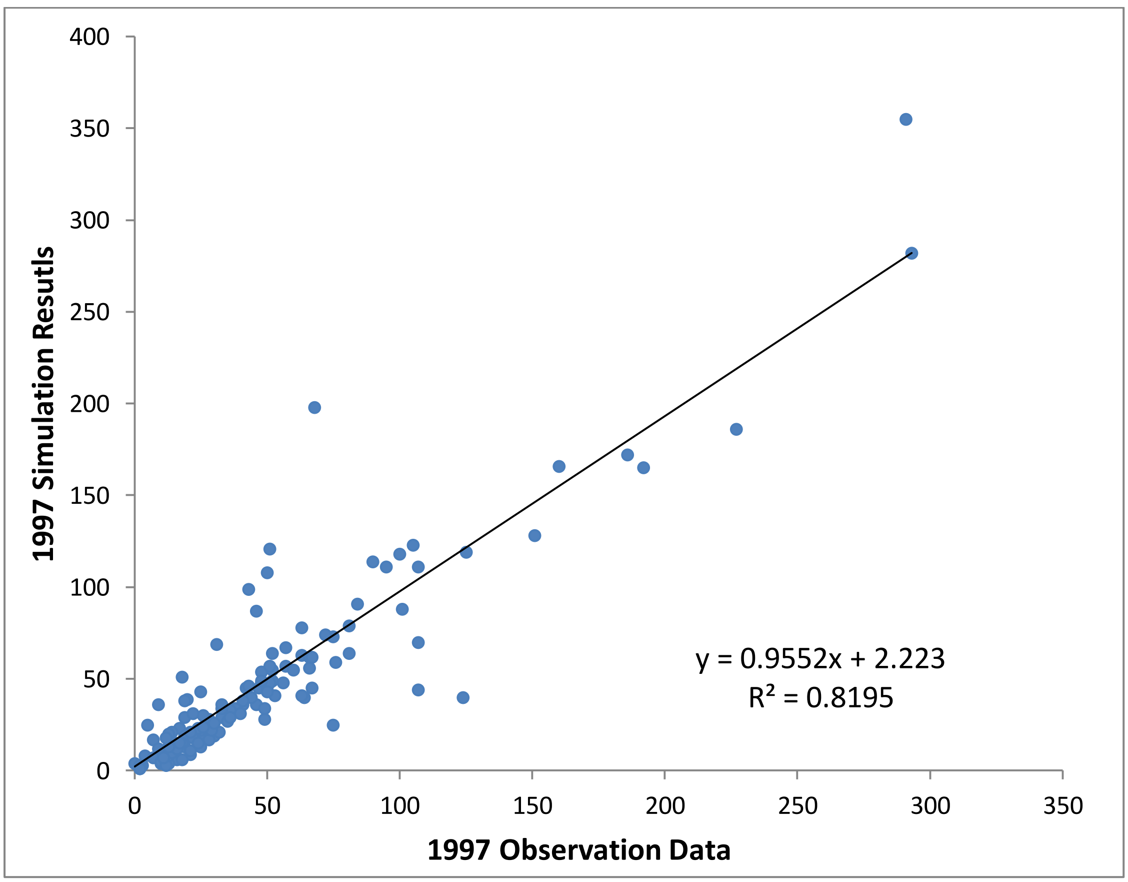
For the retail industry, shown in Figure 4-6, the majority of points distribute along the line $y = x$. From Table 4-1, the slope β_1 is 1.1055 and the constant β_0 is -2.8744. β_1 is insignificantly different than 1 and β_0 is insignificantly different than 0. Meanwhile, the R^2 value of 0.736 is relatively high indicating that our simulation model could simulate the major trends of the relocation behaviour of retail establishments.

Figure 4-6: The Business Establishment Distribution Correlation between Simulation Results and Observation Data for Retail Industry, Hamilton, 1997



For the service industry (see Figure 4-7), the slope β_1 is 0.9552 and the constant β_0 is 2.223. From Table 4-1, we learn that β_1 is insignificantly different than 1 and β_0 is insignificantly different than 0. The scatterplot in Figure 4-7 indicates that the majority of the points lie along the line $y = x$, except for a few outliers located quite far away. Also, the R^2 of 0.8195 demonstrates that our model has relatively high capability to simulate the relocation behaviour in the services industry, which contains more than half of all business establishments.

Figure 4-7: The Business Establishment Distribution Correlation between Simulation Results and Observation Data for Service Industry, Hamilton, 1997



For comparison purposes, the firmographic model was applied to another scenario, which used supply constraint 1.3, assumed to be a less accurate location supply circumstance than supply constraint 1.1. Table 4-2 represents the overall simulation results for supply constraint 1.3.

The results for supply constraint 1.3 are generally worse than the previous results from supply constraint 1.1. For all industries, the slope β_1 increases to 1.0802 and the constant β_0 decreases to -7.2636. The slope β_1 is significantly different than 1, while the constant β_0 is almost significantly different than 0 at the 0.05 significance level. We can compare those two coefficients to their values under supply constraint 1.1, which were 1.0223 and -1.6661, respectively. Furthermore, the R^2decreases to 0.8871, which compares to 0.9217 under the previous scenario.

Three industries, manufacturing, retail trade and services, have better simulation results compared against other industries in this scenario. The slopes and constant parameters from these three industries are insignificantly different than 1 and 0 respectively. However, the two predominant industries, retail trade and services, received worse simulation results compared to the previous supply constraint scenario, with respective decreases in R^2s to 0.6103 and 0.7723. This is the major reason for the decline of R^2 for all industries.

The slope β_1 and constant β_0 for the remaining three industries, construction, communication and transportation and wholesale trade, are still significantly different than 1 and 0 respectively, even though their R^2 values have increased slightly compared to the previous scenario: 0.7670 (construction), 0.2743 (communication & transportation), 0.7949 (wholesale).

Table 4-2: Regression Results between Simulation Results and Observation Data by Supply Constraint 1.3, Hamilton, 1997

Covariate	All Industries	Manufacturing	Construction	Communication & Transportation	Wholesale Trade	Retail Trade	Services
β_0(Intercept)	**- 7.2636 (-1.61)**	**- 0.1356 (-0.31)**	**- 0.8751 (-1.65)**	**1.5317 (4.04)**	**0.8596 (2.56)**	**- 3.0968 (-0.96)**	**3.6607 (0.90)**
β_1	**1.0802 (2.33)**	**1.0186 (0.77)**	**1.295 (4.62)**	**0.4562 (-9.82)**	**0.7637 (-6.81)**	**1.1144 (1.44)**	**1.0475 (-1.29)**
No. of Observation	**127**	**127**	**127**	**127**	**127**	**127**	**127**
R^2	**0.8871**	**0.9346**	**0.7670**	**0.2743**	**0.7949**	**0.6103**	**0.7723**

a. In parentheses, the table presents the estimated coefficients with their t-stats of the H_0 hypothesis that the Intercept (β_0) is 0

b. In parentheses, the table presents the estimated coefficients with their t-stats of the H_0 hypothesis that the slope (β_1) is 1

c. The critical value is 1.97 for the t distribution with 126 degrees of freedom at 0.05 level of significance

4.2 Overestimation and Underestimation Analysis

To capture and understand outliers from the regression analysis, the precision for each census tract is measured by the following equation:

$$The\ Measurement\ of\ Precision_{ij} = \frac{Simulated\ Firm\ Population_{ij} - Observed\ Firm\ Population_{ij}}{Observed\ Firm\ Population_{ij} + 1}$$

Here *i* is the index for the census tract and *j* is the index for the industry. (In the event that a census tract *i* does not have any business establishment from industry *j*, we add 1 to the denominator to avoid division by zero.) A positive value of measurement means an overestimated firm population for census tract i while a negative value of measurement means an underestimated firm population.

Based on this method, we generated a distribution map for measurement of the over- and underestimation for each major economic sector, using the results under supply constraint 1.1. Figure 4-8 through Figure 4-13 present these distribution maps for the manufacturing, construction, communication and transportation, wholesale, retail and services economic sectors respectively. According to the comparison with the land use distribution within the City of Hamilton, we have found strong spatial patterns for overestimation and underestimation of each economic sector.

Figure 4-8: The Measurement of Precision Distribution from the Comparison between Simulation Results and Observation Data for the Manufacturing Industry, Hamilton, 1997

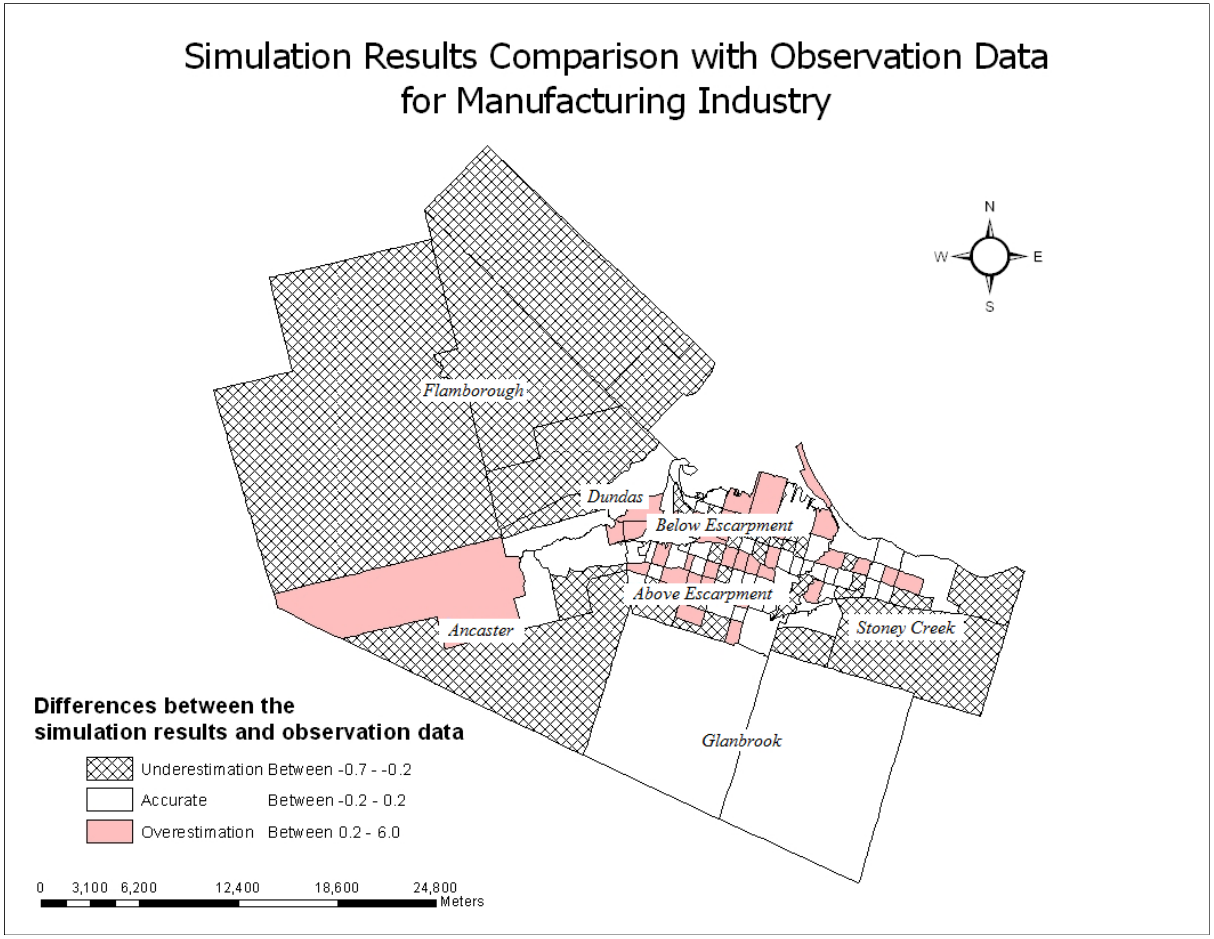


Figure 4-9: The Measurement of Precision Distribution from the Comparison between Simulation Results and Observation Data for the Construction Industry, Hamilton, 1997

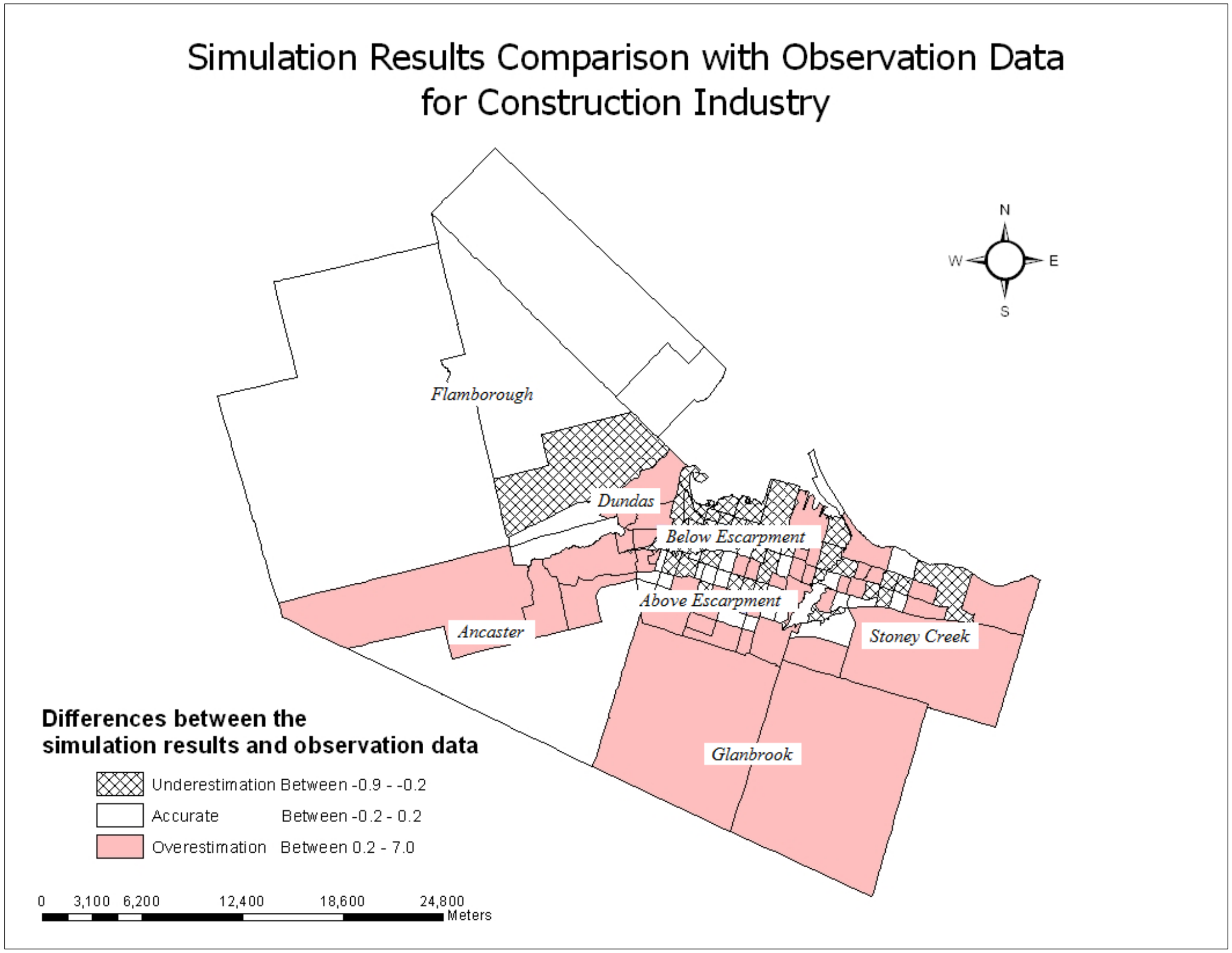


Figure 4-10: The Measurement of Precision Distribution from the Comparison between Simulation Results and Observation Data for the Communication and Transportation Industries, Hamilton, 1997

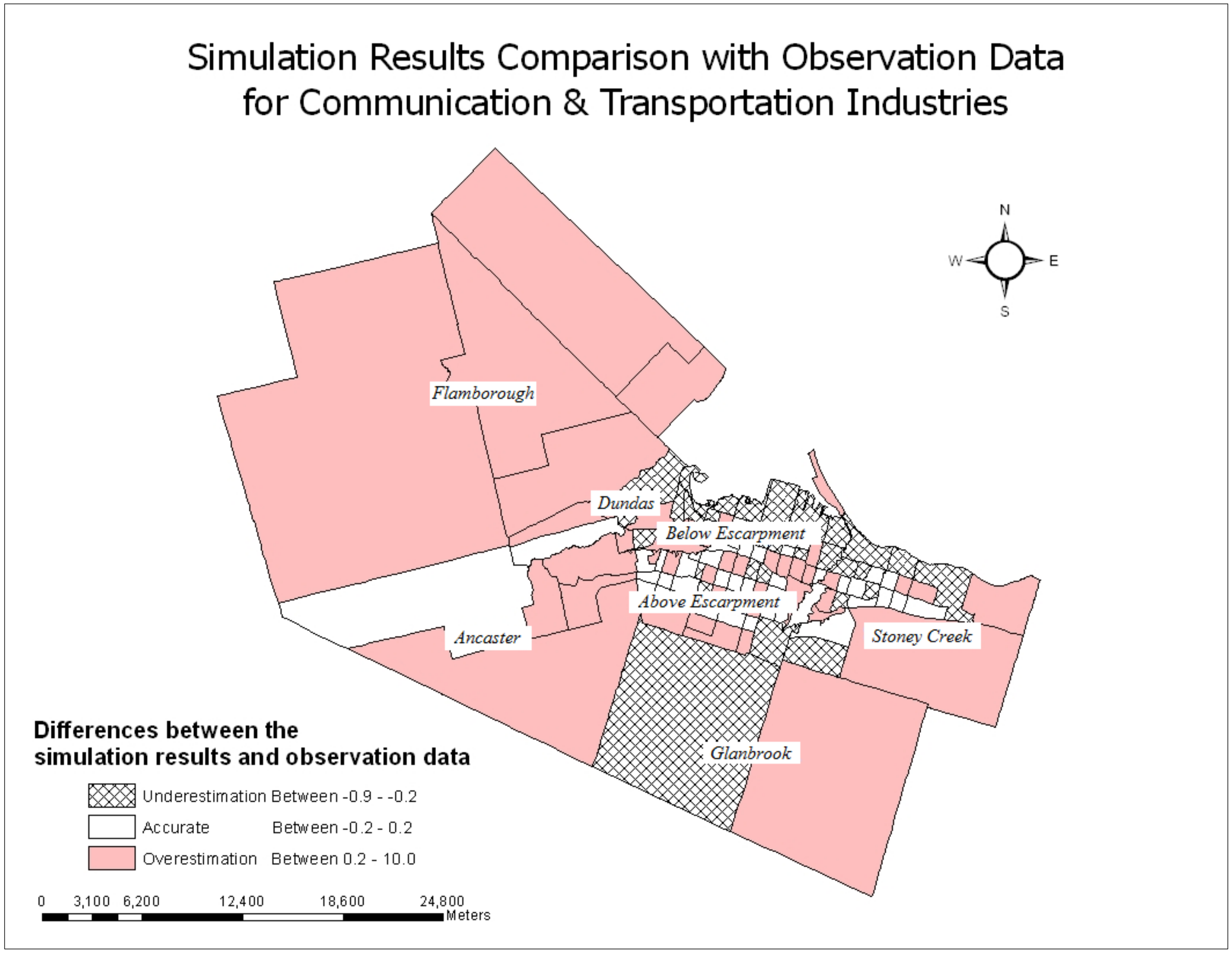


Figure 4-11: The Measurement of Precision Distribution from the Comparison between Simulation Results and Observation Data for the Wholesale Trade Industry, Hamilton, 1997

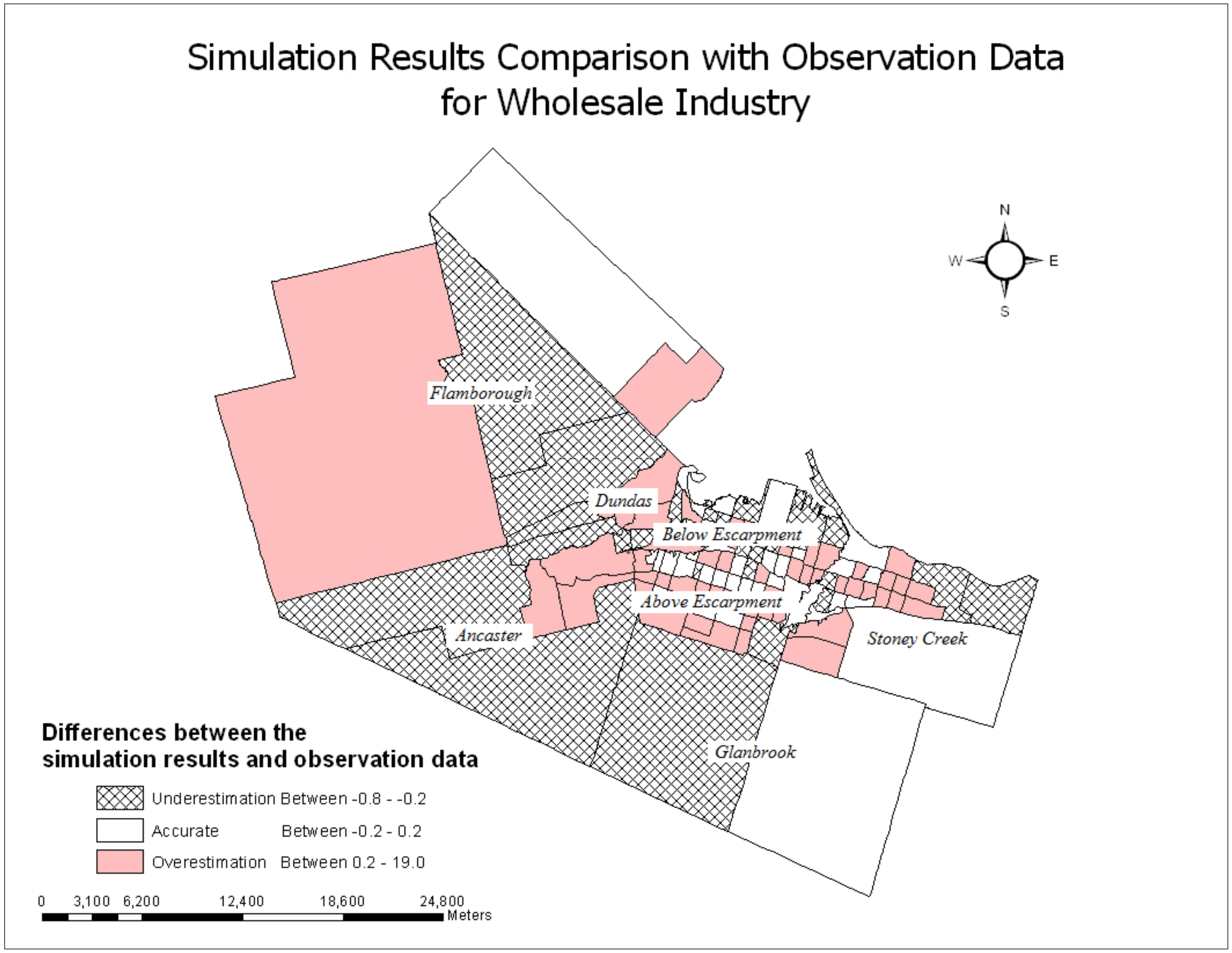


Figure 4-12: The Measurement of Precision Distribution from the Comparison between Simulation Results and Observe Data for the Retail Trade Industry, Hamilton, 1997

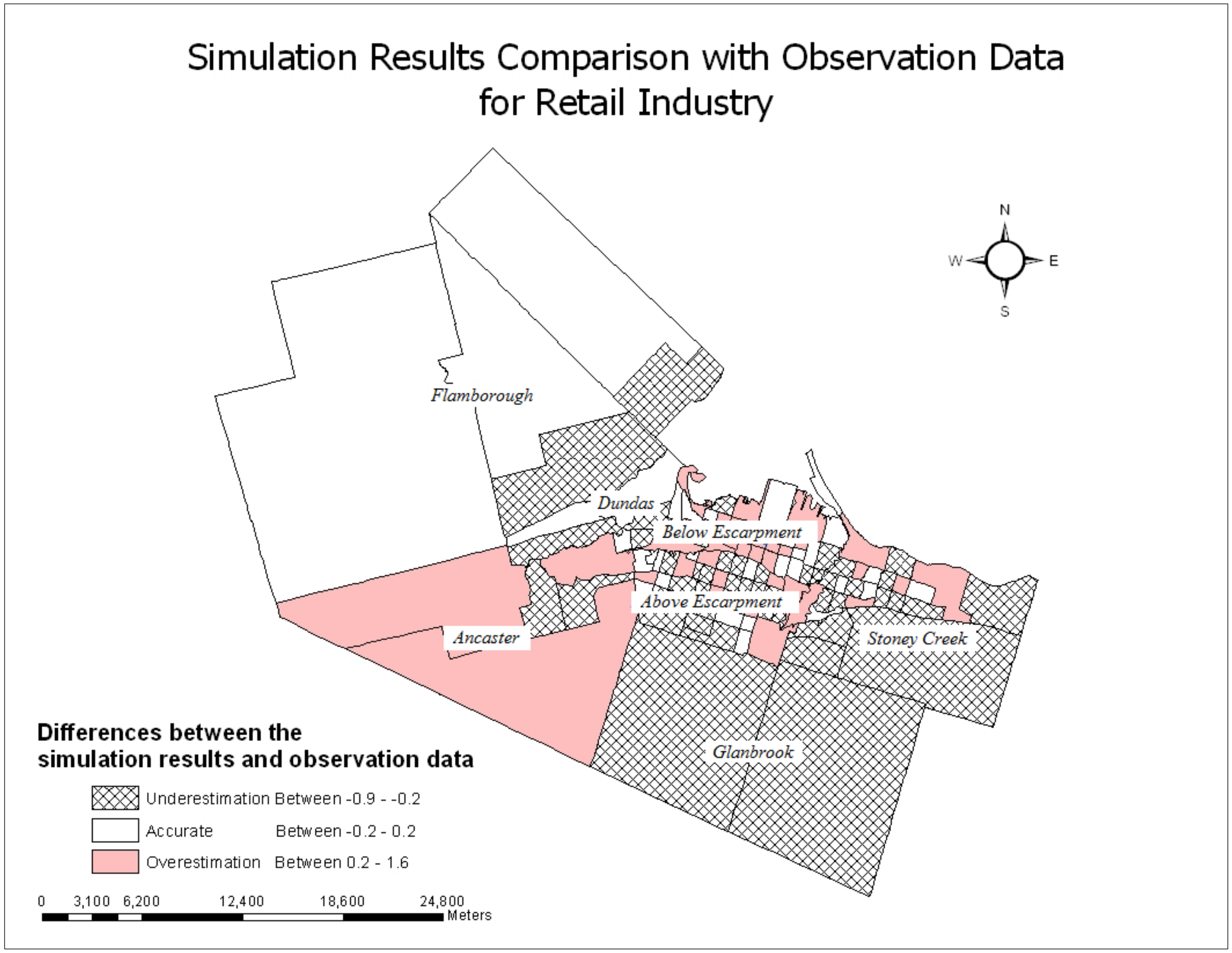
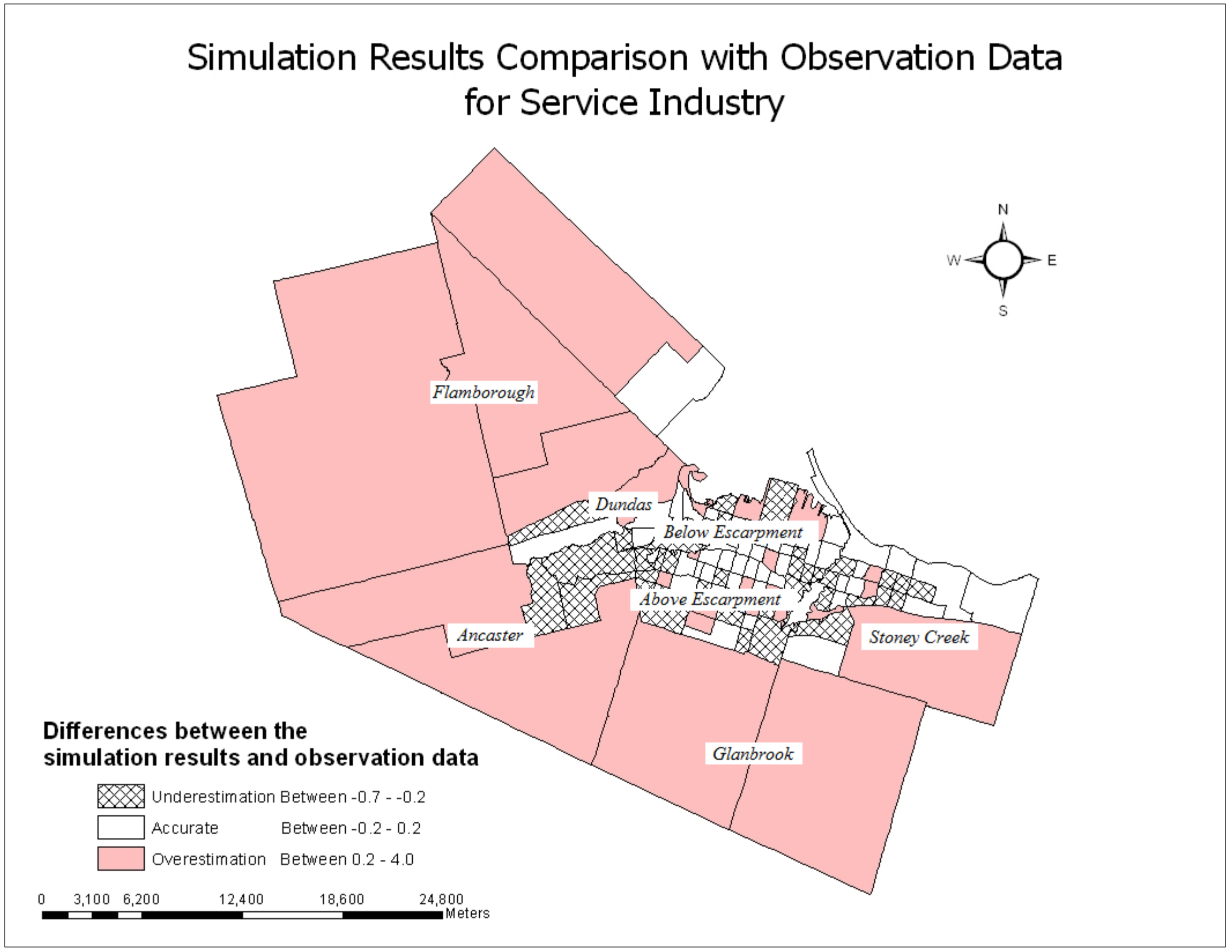


Figure 4-13: The Measurement of Precision Distribution from the Comparison between Simulation Results and Observation Data for the Services Industry, Hamilton, 1997



For the manufacturing industry, from Table 4-3, we find the LANDUSE2 (resource and industrial land use) covariate plays a dominant role in the location choice model. As we see in Figure 4-8, the most overestimated census tracts are located in Hamilton Harbour and parts of lower, upper and eastern Hamilton areas. These areas contain a set of grid cells which have a high proportion of resource and industrial land use type. As well, some census tracts, such as suburban areas in Flamborough, Ancaster and Stoney Creek, were underestimated because of the lower proportion of resource and industrial land use and the lower utility values of the grid cells belonging to these census tracts.

For the construction industry (see Table 4-4), we find the LANDUSE1 (open space land use) covariate plays a dominant role in the location choice model. The most overestimated census tracts, such as those shown in Figure 4-9 in Ancaster, Dundas, Glanbrook and Stoney Creek, contain a set of grid cells with a high proportion of open space land use. Similarly, some census tracts, such as those in the lower city and downtown areas, were underestimated because of the lower proportion of open space land use, and thus the lower utility values of the grid cells belonging to these census tracts.

For the communication and transportation Industries, as shown in Table 4-5, we find that both LANDUSE1 (open space land use) and LANDUSE2 (resource and industrial land use) play dominant roles in the location choice model. The most overestimated census tracts, such as most of those in the suburban communities (see Figure 4-10), contain a set of grid cells with a high proportion of open space, resource or industrial land use. On the contrary, some census tracts, such as some Hamilton downtown areas, have a lower proportion of land allocated to these two land uses or even no such kinds of land uses were underestimated because of the lower utility values of the grid cells belonging to these census tracts.

For the wholesale trade industry (Table 4-6), CBDPRO and AGGLOM4 have a relatively heavier weight to attract business establishments to locate in census tracts which are far away from the CBD (explaining some overestimation in areas such as Flamborough) or have high agglomeration effects (explaining the overestimation in central Hamilton). (See Figure 4-11.) On the other hand, some census tracts located in suburban areas, such as parts of Ancaster and Glanbrook, have neither low agglomeration effects nor long distance from CBD and so would attract fewer wholesale business establishments.

For the retail trade Industry, shown in Table 4-7, we find that the LANDUSE4 (residential land use) and LANDUSE5 (commercial land use) covariates play dominant roles in the location choice model. The most overestimated census tracts, such as those in Ancaster, Dundas, and central parts of Hamilton (see Figure 4-12), contain a set of grid cells which have high proportions of residential and commercial land uses. On the contrary, some census tracts, such as those in Glanbrook and Stoney Creek, have lower proportion of residential and commercial land uses, and their retail establishments were thus underestimated because of the lower utility values of the grid cells belonging to these census tracts.

For the services industry, from Table 4-8, we find LANDUSE1 (open space land use), LANDUSE3 (park and recreational land use), LANDUSE5 (commercial land use) and LANDSUE6 (government and institutional land use) all play dominant roles in the location choice model. The most overestimated census tracts, as shown in Figure 4-13, are mostly in Flamborough, Dundas, Ancaster, Glanbrook, Stoney Creek and part of central Hamilton. Those census tracts contain a set of grid cells which have high proportions of open space, commercial, government or institutional land uses, which serve to attract business establishments in this industry. On the contrary, for most census tracts in the central Hamilton areas, which mostly contain residential land use, the services industry was consequently underestimated due to the lower utility values of the gridcells belonging to these census tracts.

Table 4-3: Interaction Values of Coefficient by Covariate in Location Choice Model for Manufacturing Industry, Hamilton

Covariate	Minimum	Maximum
CBDPRO	0	0.000110*30417.38 = 3.35
HWYMRPRO	0	0.663567*1 = 0.66
AGGLOM1[a]	0	0.047804*55 = 2.63
POPDENS	-0.000258*10083.89 = -2.60	0
LANDUSE2	0	0.750878*100 = 75.09
NEWBORN*CBDPRO	-0.000058*1*30417.38 = -1.76	0
INDUSTRY1*CBDPRO	-0.000110*1*30417.38 = -3.35	0
INDUSTRY2*POPDENS	0	0.000339*1*10083.89 = 3.42

a. The value of AGGLOM1 is computed using 1990 firmographic data from the City of Hamilton

Table 4-4: Interaction Values of Coefficient by Covariate in Location Choice Model for Construction Industry, Hamilton

Covariate	Minimum	Maximum
CBDPRO	0	0.0000970*30417.38 = 2.95
HWYMRPRO	-0.362410*1 = -0.36	0
MALLPRO	-0.646879*1 = -0.65	0
AGGLOM2[a]	0	0.080863*38 = 3.07
MOUNTAIN	0	0.783139*1 = 0.78
NEWDEVLOP	0	0.004611*1126.98 = 5.20
OLDDEVLOP	-0.001936*1003.38 = -1.94	0
AVGDWELLVAL	-0.000002*451121.22 = -0.90	0
HHLDDENS	0	0.000801*5806.75 = 4.65
LANDUSE1	0	1.559011*100 = 155.90
INDUSTRY3*HWYMRPRO	0	0.806584*1*1 = 0.81
INDUSTRY3*MOUNTAIN	-0.891513*1*1 = -0.89	0
INDUSTRY4*LANDUSE1	-0.918098*100 = -91.81	0
INDUSTRY5*LANDUSE1	0	1.753567*100 = 175.36

a. The value of AGGLOM2 is computed using 1990 firmographic data from the City of Hamilton

Table 4-5: Interaction Values of Coefficient by Covariate in Location Choice Model for Communication and Transportation Industries, Hamilton

Covariate	Minimum	Maximum
CBDPRO	0	0.000092*30417.38 = 2.80
HWYMRPRO	0	0.580206*1 = 0.58
MALLPRO	-0.959108*1 = -0.96	0
AGGLOM3[a]	0	0.234618*15 = 3.52
LANDUSE1	0	0.984063*100 = 98.41
LANDUSE2	0	2.136029*100 = 213.60
INDUSTRY6*LANDUSE2	-2.205140*100 = -220.51	0
INDUSTRY7*CBDPRO	-0.000136*30417.38= -4.14	0

a. The value of AGGLOM3 is computed using 1990 firmographic data from the City of Hamilton

Table 4-6: Interaction Values of Coefficient by Covariate in Location Choice Model for Wholesale Trade Industry, Hamilton

Covariate	Minimum	Maximum
CBDPRO	0	0.000061*30417.38 = 1.86
HWYMRPRO	0	0.498985*1 = 0.50
AGGLOM4[a]	0	0.094018*31 = 2.91
SUBURBS	0	0.618772*1 = 0.62
NEWDEVLOP	0	0.002791*1126.98 = 3.15
INDUSTRY8*CBDPRO	-0.000159*1*30417.38 = -4.84	0
INDUSTRY9*HWYMRPRO	-0.763197*1*1 = -0.76	0

a. The value of AGGLOM4 is computed using 1990 firmographic data from the City of Hamilton

Table 4-7: Interaction Values of Coefficient by Covariate in Location Choice Model for Retail Trade Industry, Hamilton

Covariate	*Minimum*	Maximum
HWYMRPRO	*0*	0.376421*1 = 0.38
AGGLOM5[a]	*0*	0.022177*400 = 8.87
SUBURBS	*0*	0.426167*1 = 0.43
NEWDEVLOP	*0*	0.003345*1126.98 = 3.77
OLDDEVLOP	*-0.000629*1003.38 = -0.63*	0
HHLDINCDENS	*0*	0.000002*91768.52 = 0.18
LANDUSE4	*-0.559876*100 = -55.99*	0
LANDUSE5		1.152924*100 = 115.29
INDUSTRY10*MALLPRO	*0*	0.650485*1 = 0.65
INDUSTRY11*HHLDINCDENS	*-0.000002*1**	0
INDUSTRY12*LANDUSE4	*0*	0.643616*1*100 = 64.36
INDUSTRY13*SUBURBS	*-0.425008*1*1 = -0.43*	0

a. The value of AGGLOM5 is computed using 1990 firmographic data from the City of Hamilton

Table 4-8: Interaction Values of Coefficient by Covariate in Location Choice Model for Service Industry, Hamilton

Covariate	Minimum	Maximum
CBDPRO	0	0.000055*30417.38 = 1.67
HWYMRPRO	0	0.332224*1 = 0.33
AGGLOM6[a]	0	0.005285*983 = 5.20
HHLDINCDENS	0	0.000001* 91768.52 = 0.09
LANDUSE1	0	0.661705*100 = 66.17
LANDUSE3	-0.771938*100 = -77.19	0
LANDUSE5	0	0.628379*100 = 62.84
LANDUSE6	0	0.398759*100 = 39.88
INDUSTRY14*CBDPRO	-0.000028*1*30417.38 = -0.85	0
INDUSTRY15*MOUNTAIN	0	0.286058*1*1 = 0.29
INDUSTRY16*NEWDEVLOP	0	0.001207*1126.98 = 1.36
INDUSTRY17*NEWBORN *HWYMRPRO	0	0.882275*1*1 = 0.88
INDUSTRY18*NEWBORN *HWYMRPRO	0	0.546541*1*1 = 0.55

a. The value of AGGLOM6 is computed using 1990 firmographic data from the City of Hamilton

Chapter 5 Conclusion

5.1 Thesis Conclusion

In the present day, modelers of IUMs have shown an increasing interest in the agent-based microsimulation approach. This methodology has advantages such as disaggregate nature, dynamic and behavioural realism and responding capability for long-term planning policy. (Miller et al., 1998).

This thesis describes a firmographic microsimulation model that simulates small- and medium-sized business establishments' survival, mobility, location choice and formation events for the City of Hamilton, Ontario, Canada. Our research also provides some operational software based on the GIS technologies, which offers a user-friendly interface, interactive tool and visualization services.

The results from a validation test show that the decentralization and urban sprawl phenomena, which are particularly true for wholesale and retail trade firms in Hamilton (Maoh, 2005), are captured by the model. Meanwhile, the regression analysis comparing the simulation results and the observation data for the year 1997 suggests that the model has a relatively strong ability for predicting the distribution of business establishments by industry sector at the census tract level.

5.2 Research Contribution

Existing efforts for modelling employment mostly use the aggregate models. Firmographic modelling methodology is still a relatively new field, although some disaggregate models, for example SIMFIRMS (Van Wissen, 2000), have incorporated the firmography concept into their research, and are used at the regional level. Other urban-level models, such as UrbanSim (Waddell, 2003), treat jobs rather than business establishments as the smallest unit of interest. However, our research applies the firmographic methodology at the urban scale and considers individual business establishments as the unit of interest, the latter of which we believe is more realistic for simulating real-world economic activities.

Moreover, this research fills an existing gap, as the City of Hamilton does not have a microsimulation model for the commercial industry. The final model based on our research could replace the existing aggregate employment submodel in IMULATE, providing urban planners with a more policy-sensitive decision support tool.

Furthermore, this research integrates economic models with GIS technologies, supplying solid geoprocessing and visualization functionalities, and it has produced a piece of operational software which provides a convenient way to run the model.

5.3 Thesis Limitation

Although the current firmographic model is capable to generate fairly good results, there are still several limitations.

The most significant limitation is the lack of data. Age information carries important influence and is part of both the survival and mobility submodels, but the firmographic data available for our base year of 1990 do not contain age information for each individual business establishment. However, age information, which carries important influence, appears in both survival submodel and mobility submodel. Although, this problem is solved, to some extent, by an age synthesis algorithm based on the 2003 Hamilton Business Directory, though the poor data quality of this dataset still restricts the algorithm's ability to generate better age information. Moreover, the relocation history is not included in the 1990 base year data, thus the business establishments' survival and failure events during our simulation, based on models which include a relocation parameter, will not take into account the pre-1990 relocation of any establishment.

Another data issue is related to land use. In the location choice model, land-use variables play a significant role in influencing an establishment location choice decision. However, the earliest land-use data in the DMTI dataset is from 2001, whereas our model simulates the years 1991 to 1997. Although we have tried to incorporate some other land-use data to establish the ideal dataset, we cannot upgrade effectively the land-use data at present.

Furthermore, we do not estimate a growth submodel, which is used for employment size prediction. Our lack of growth information would bias some of our simulation results. In the survival submodel, employment growth would increase the probability of a business establishment's survival in the next year; similarly, a decline in employment would make survival less likely. In the mobility submodel, positive growth would increase the tendency to stay in the current location for the construction, communication, transportation, wholesale, retail and services industries, while it would motivate manufacturing establishments to relocate. Thus, assuming no growth for each individual establishment will lead the simulation process to disregard the above trends.

Currently, the formation submodel is not a pure microsimulation model. The number of new business establishments each year is instead estimated from the seven-year difference between 1990 and 1997.

The present model framework considers that the execution sequence of business establishments' relocation decision is fixed in every simulation year. However, this is not realistic, as some small establishments will consider others' relocation decisions when they decide whether to relocate.

5.4 Directions for Future Research

Currently, better solutions for the formation and growth submodel are needed. The absence of the growth submodel has brought a certain degree of bias to the simulation while the formation submodel lacks the ability to predict the number of new business establishments in the study area each year.

A supply model for location alternatives also needs to be established. This model is responsible for simulating the number of available location units for a business establishment looking to relocate. In the current framework, the location alternatives are imported as an exogenous dataset. However, to project a scenario within the next few years, those datasets must be endogenous.

Residential models should be also incorporated into the current framework. The value of the variables used in the current location choice model, such as household density and household income density, need to be simulated by these residential models. The incorporation of residential models will also arouse a competition between residential and commercial land use development. Because land developers in most cases would like to maximize their profit, they will look to make decisions for the right development type.

If the data are available, some exiting firmographic models could be updated. For example, the location choice model could be updated to reflect the tendency of an owner of a small- or medium-sized business establishment to consider the distance between the new business location and his or her home when relocating the business (Elgar et al.,2008). Another example is that the same land use variables could have different weights depending on the urban area (for example, downtown or suburban). Proximity variables such as the distance between CBD and a location alternative could be measured by travel time rather than Euclidean distance. Highway proximity for a location alternative could be measured by its distance to the nearest highway interchange, rather than by a binary variable based on the existence of a section of highway in the grid cell. The mere presence of a highway in a grid cell does not imply that there is highway access, or that it is convenient.

Using dynamic displaying technology for firmographic agents in the model would provide better visual effects for understanding the entire firmographic evolution process. Currently, our model only shows the snapshots for final simulation results for the end of every simulation year.

After including all these improvements, the firmographic microsimulation model would gain strong projection capabilities for setting policy in the future. This research should also be able to serve as an effective and efficient decision support tool for urban planning.

REFERENCES

Baldwin, J. R., Bian, L., Dupuy, R., & Gellatly, G. (2000). "Failure rates for new Canadian firms: New perspectives on entry and exit." Catalogue no. 61-526-XIE. Ottawa: Statistics Canada.

Brouwer, A. E., Mariotti, I., & van Ommeren, J.N. (2004). "The firm relocation decision: an empirical investigation." *The Annals of Regional Science*, *38*, 335-347.

Cassia, L., Colombelli, A., & Paleari, S. (2009). "Firms' growth: Does the innovation system matter?" *Structural Change and Economic Dynamics, 20,* 211-220.

De Bok, M. (2006). "Infrastructure and firm dynamics: Calibration of microsimulation model for firms in the Netherlands." *Transportation Research Record: Journal of the Transportation Research Board*, *1977,* 132-144.

De Bok, M. (2007). Infrastructure and firm dynamics. (Ph.D. thesis). Delft University of Technology.

De Bok, M. (2009). "Estimation and validation of a microscopic model for spatial economic effects of transport infrastructure." *Transportation Research Part A: Policy and Practice, 43*, 44-59.

De Bok, M., & Bliemer, M. (2005). "Using firm demographic microsimulation to evaluate land use and transportation scenario evaluation – Model calibration." 45th ERSA CONGRESS.

De Bok, M., & Sanders, F. (2005). "Firm relocation and accessibility of locations: Empirical results from the Netherlands." *Transportation Research Record: Journal of the Transportation Research Board, 1902*, 35-43.

De la Barra, T., Pérez, B., & Vera, N. (1984). "TRANUS-J: Putting large models into small computers." *Environment and Planning B: Planning and Design, 11,* 87-101.

Elgar, I., Miller, E. J., & Farooq, B. (2008). "Development of an operational integrated urban model system, volume VI: Modelling firm mobility & location choice." ILUTE Project, Urban Transportation Research & Advancement Centre, University of Toronto.

Ghosh, S. (2009). "Do productivity and ownership really matter for growth? Firm-level evidence." *Economic Modeling, 26,* 1403-1413.

Hannan, M. T., & Freeman, J. (1989). "The population ecology of organizations." *American Journal of Sociology, 82,* 929-964.

Hayter, R. (1997). *The dynamics of industrial location: The factory, the firm and the production system.* Chichester, UK: Wiley.

Horowitz, A. J. (2004). Lowry-type land use models. In: D.A. Hensher, K.J. Button, K.E. Haynes, & P.R. Stopher (Eds.), *Handbook of transport geography and spatial systems* (pp. 167-183). Amsterdam: Emerald.

Hunt, J. D., Khan, J. A. S., Abraham, J. E. (2003). "Microsimulating firm spatial behaviour." Proceedings from *The 8th International Conference on Computers in Urban Planning and Urban Management*, Sendai, Japan.

Hunt, J. D., & Simmonds, D. C. (1993). "Theory and application of an integrated land-use and transport modelling framework." *Enviroment and Planning B: Planning and Design, 20,* 221-224.

Kanaroglou, P. and Scott, D. (2002). Integrated urban transportation and land-use models for policy analysis. In M.J. Dijst, W. Schenkel, & I. Thomas (eds.), *Governing Cities on the Move: Funcitonal and Management Perspectives on Transformations of European Urban Infrastructures* (pp. 43-73), Aldershot, UK: Ashgate.

Khan, A. S., Abraham, J. E., & Hunt, J. D. (2002). "Agent-based micro-simulation of business establishments." Proceedings from the *42nd Congress of the European Regional Science Association.* Dortmund, Germany.

Kirchoff, B. A., & Phillips, B. D. (1988). "The effect of firm formation and growth on job creation in the United States." *Journal of Business Venturing, 3*, 261-272.

Kumar, S., & Kockelman, K. M. (2008). "Tracking size, location, and interactions of businesses: Microsimulation of firm behavior in Austin, Texas." *Transportation Research Record: Journal of the Transportation Research Board, 2077*, 113-121.

Lowry, I. (1964). "A model of metropolis." Santa Monica, Calif, Rand Corp.

Macgill, S. M. (1977). "The Lowry model as an input-output model and its extension to incorporate full intersectoral relations." *Regional Studies, 11,* 337-354.

Maoh, H. (2005). Modeling firm demography in urban areas with an application to Hamilton, Ontario: Towards an agent-based microsimulation model." (Ph.D. thesis). Centre for Spatial Analysis, McMaster University.

Miller, E. J., Hunt, J. D., Abraham, J. E., & Salvini, P. A. (2004). "Microsimulating urban systems." *Computers, Environment and Urban Systems, 28*: 9-44.

Moeckel, R. (2005). "Microsimulation of Firm Location Decisions." *Proceedings from the 9th International Conference on Computers in Urban Planning and Urban Managaement*, London, England.

Moeckel, R., Schürmann, C., & Wegener, M. (2002). "Microsimulation of urban land use." Proceedings from the *42nd Congress of the European Regional Science Association.* Dortmund, Germany.

Moore, T., & Thorsnes, P. (1994). *The transportation/land use connection*. Chicago: American Planning Association, Planning Advisory Service.

Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change*. Cambridge: Harvard University Press.

Parker, S. C. (2009). "Why do small firms produce the entrepreneurs?" *Journal of Socio-Economics, 38*, 484-494.

Putman, S. (1983). *Integrated urban models: Analysis of transportation and land use*. London: Pion.

Shukla, V., & Waddell, P. (1991). "Firm location and land use in discrete urban space: A study of the spatial structure of Dallas-Fort Worth." *Regional Science and Urban Economics, 21*, 225-253.

Southworth, F. (1995). *A technical review of urban land-use transportation models as tools for evaluating vehicle travel reduction strategies*. Center for Transportation Analysis, Energy Division, Oak Ridge, Tennessee.

Timmermans, H. (2003). "The Saga of Integrated Land Use-Transport Modeling: How Many More Dreams Before We Wake Up?" Keynote paper from the *10th International Conference on Travel Behaviour Research*. Lucerne, Switzerland.

Van Wissen, L. (2000). "SIMFIRMS: Simulating the spatial demography of firms, with an application in the Netherlands." Proceedings from *Computing in Economics and Finance 2000*. Barcelona, Spain.

Van Wissen, L. J. G. (2004). "Demography of the firm: A useful metaphor?" *European Journal of Population, 18,* 263-279.

Waddell, P., Borning, A., Noth, M., Freier, N., Becke., M, & Ulfarsson, G. (2003). "Microsimulation of Urban Development and Location Choices: Design and Implementation of UrbanSim." *Networks and Spatial Economics, 3* 43-67.

Waddell, P., & Shukla, V. (1993). "Manufacturing location in a polycentric urban area: a study in the composition and attractivness of employment subcenters." *Urban Geography, 14*, 277-296.

Wegener, M. (1995). "Current and future land use models". Proceedings from the *Land Use Model Conference*, Dallas, Texas.

Wu, F. (1999). "Intrametropolitan FDI firm location in Guangzhou, China: A Poisson and negative binomial analysis." *The Annals of Regional Science, 33*, 535-555.