DEVELOPING AN EMPLOYMENT LOCATION MODEL FOR THE HAMILTON CENSUS METROPOLITAN AREA (CMA)

DEVELOPING AN EMPLOYMENT LOCATION MODEL FOR THE HAMILTON CENSUS METROPOLITAN AREA (CMA)

By

HANNA F. MAOH, B.Sc.

A thesis
Submitted to the School of Graduate Studies
in Partial Fulfillment of the Requirements for the degree

Master of Science

McMASTER UNIVERSITY

MASTER OF SCIENCE (1999) | McMASTER UNIVERSITY
(Geography) | Hamilton, Ontario

TITLE: Developing an Employment Location model for the Hamilton CMA

AUTHOR: Hanna F. Maoh, B.Sc. (Bethlehem University)

SUPERVISOR: Professor Pavlos Kanaroglou

NUMBER OF PAGES: xv, 177

ABSTRACT

In this Thesis, the spatial distribution of employment is modeled for the Hamilton CMA. A behavioral employment location model is constructed, estimated, implemented and tested. Employment mobility is associated with the redistribution of firms in the region, thus, the spatial distribution of firms is modeled and employment location is inferred from that. Manufacturing, construction, wholesale trade, retail trade and services sectors are modeled. Sectors like the communication, transportation, government services and educational services are given exogenously to the model.

Three models form the elements of the employment model. The first is an Input-Output Model, which captures the linkages between the different sectors of the economy and predicts the regional employment in each industry. The second is a destination choice model having in its elements a Multinomial Logit Model that predicts the choice probability of new and relocating firms. The third model is a regression model having in its elements both a spatial regression (SAR model) and non-spatial (classical) regression model. The model predicts the number of lost firms. The last two models predict firms at the census tract level.

Estimation results indicate that firms in the different industries show a systematic behavior in choosing a site to locate at. Factors such as the CBD proximity, highway proximity, mall proximity, population size, household density, and agglomeration economies affect the locational decision of the different types of firms. Moreover, the analysis shows that loss of firms is linearly related to the total number of firms at the census tract level.

The employment model is implemented using the GAUSS programming language. The model shows a significant goodness-of-fit when comparing the predicted values of employment with the observed with an r-square value of 0.91. Scenario simulation is also achieved using the implemented model. The model shows its capability of simulating certain types of scenarios.

ACKNOWLEDGMENTS

I extend my sincere appreciation and thanks to my supervisor, Dr. Pavlos Kanaroglou who has put a lot of effort in making me a mature researcher. Thank you Pavlos for sharing your experience and knowledge, it's been a pleasure to work with you for the past two years. I would not be able to achieve this fine work without your help and guidance. Thank you to Professors William Anderson and K. Lee Liaw for the information and knowledge gained from their courses. Thank you to Dr. Atif Kubursi for his assistance in tutoring some of the economic concepts that were a mystery to me.

I am thankful to the Planning and Development Department of the Regional Municipality of Hamilton-Wentworth for providing me with the employment databases used in this thesis. Specially, I would like to mention Angela Monaco and William Pearce, both of whom I was in contact with at various times.

There are many others who were not involved in the work of this thesis that deserve an acknowledgment for contribution of encouragement and friendship. Thank you to Darren Scott for the good friendship, useful discussions and guidance. Thank you to Angela Cuthbert for helping me in editing some parts of this thesis. Thank you to the Applied Research Institute - Jerusalem (ARIJ) for the encouragement in my academic career. Specially, I would like to mention Dr. Jad Isaac, who believed in me as a real researcher. To many other friends in Canada and Bethlehem, thank you all for your encouragement and the good times we have.

Finally, I would like to express my deepest and warmest thanks to my be loving family, for limitless patience and understanding, for supporting me in my life trip, for encouraging me all the way up till today, for the continuos prayers to succeed. Thanks mother, thanks to my sisters Antoinette and Mary, I really appreciate all what you did for me. Thank you for making me proud of myself.

Thank you God for every thing…

Hanna F. Maoh, April, 1999

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

Page

CHAPTER THREE cont'd

TABLE OF CONTENTS (continued)

Page

CHAPTER FIVE cont'd

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

LIST OF TABLES

LIST OF TABLES (Continued)

CHAPTER ONE

INTRODUCTION

1.1 Introduction

There are a large number of potential interactions between urban form, energy consumption and the environment. One of the most direct and profound interactions is the relationship of urban form with energy consumption and pollution generation in the transportation sector (Anderson et al., 1994). During the past four decades many theories and methodologies have been employed to help urban planners assess the impacts of transportation plans and polices that will support the evolution of more energy-efficient and less polluted cities, and to aid in the design of specific travel-reduction strategies (Southworth, 1995).

Integrated urban models *IUMs* can be used as decision support tools in designing policies that will lead the urban system in the desired direction. According to Anderson et. al. (1994), *IUMs* are a class of empirical computer simulation models that may be used to project changes in transportation flow and land-use patterns in the metropolitan area under various scenarios. However, predicting the changes in land-use involves modeling these changes. The main factors affecting land-use change are population and employment, therefore, modeling population and employment becomes a necessity and part of the whole modeling process.

One of the challenges facing these types of models is the ability to incorporate new theoretical developments and methodologies that reflects the present status of an urban area. Fortunately, rapid advances in information and computing technologies made these challenges easier from the technical side (Wegener, 1995).

In September 1992 an **I**ntegrated **M**odel for **U**rban **LA**nd-Use **T**ransportation and **E**nvironmental Analysis ***IMULATE*** was developed for the Hamilton Census Metropolitan Area (CMA). *IMULATE* is designed to simulate the spatial assignment of vehicle trips for the morning rush hour period in the region. The aim of each simulation is to assess the level of congestion on the transportation network and from that to infer the emissions and energy consumption from passenger vehicles.

Like any other integrated urban model, *IMULATE* includes a land use model within it that simulates the land use changes that occurs over time. The major component of the land use model is a submodel called *POPMOB*. *POPMOB* is used to simulate the process of residential mobility as well as the performance of housing market. The resultant data from this model is a set of matrices that includes a matrix of probabilities for place of residence-place of work. On the other hand, *POPMOB* estimates the household distribution in the region for the next simulation period. However, places of work remains unchanged in all simulation periods since employment mobility is not modeled in the land use model.

1.2 Problem Status and Research Objective

In its present status, *IMULATE* is governed by a number of limitations that include:

1. The assumption that the households are mobile, but that the spatial distribution of employers, commercial locations, recreational facilities etc. are exogenously determined.
2. The applications of the model only to the morning peak travel demand period on weekdays.
3. The use of somewhat crude methods for estimating energy use and emissions.

The previous limitations indicate some sort of weakness in the model that prevents it from being an effective tool for projection and forecasting. It is worth noting that the first limitation is the most urgent. *IMULATE* is designed with the assumption that jobs location are static-that is, they do not change with time. This assumption will create an artificially large number of work trips from growing peripheral areas to traditional employment areas (Anderson et al., 1994). As a result, the model will fail in projecting the accurate number of trips and this will affect the overall performance of the model.

In this context, the purpose of this research is to develop an employment location model that can be implemented within *IMULATE*. The model will predict the change in employment mobility. This will be linked with the *POPMOB* model to form a comprehensive land use model for the region. The model will be a behavioural one whereby employers' location choices are effected by changes in the spatial patterns of accessibility and cost.

1.3 Outline of Thesis

The reminder of this thesis includes five chapters. Chapter Two will review the literature on some employment location models used in integrated urban models. On the other hand, it will identify factors affecting both the gain and loss of firms in urban areas by reviewing most recent work in this field. Chapter Three will present the data and methodology of the thesis, which is the hardcore of this study. The databases on which the analysis is based will be discussed in this chapter. Moreover, we shall introduce a formal framework of the firms' location model to be developed in this research. Chapter Four will present the empirical work to be conducted in this study. Three models will be developed. The first is an input-output model, which will predict the firms at the regional level. The second is a destination choice model that will predict the location of new and relocating firms at the census tract level. The third is a lost firms model, which will predict the number of lost firms in each census tract. The three models will be linked together to provide the employment model. Chapter Five will present a computerised version of the model. The goodness-of-fit and the potential of simulating scenarios will be tested in the chapter. Finally, Chapter Six will give a summary and conclusion for the findings of the thesis with a direction for future research in the field of modeling employment in urban areas.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

The purpose of this chapter is to synthesize the theory underlying modeling employment in metropolitan areas. Concepts and methodologies used to develop employment location models are highlighted in this chapter. As discussed in the first chapter, the objective of this research is to develop a model that simulates the location of employment in an urban area. This model is to be used in an existing integrated urban model-IMULATE (Integrated Model of Urban LAnd use, Transportation, and Environment analysis). Thus, we review the literature on integrated urban models, emphasizing the different theories used to construct employment models. This is intended to inform the conceptualization of a modeling framework as described later in this thesis

The review will also identify the factors that influence the locational decision of firms in the city. Studies have shown that firms do not locate randomly in cities, but show a systematic behavior in choosing a site to locate. In this context, a number of push and pull factors that trigger this behavioral decision is highlighted and discussed.

2.2 Theories and Methodologies Used in Modeling Intraurban Employment

Integrated Urban Models (*IUMs*) comprises a class of empirical computer simulation models, which are used to project changes in transportation flows and land use patterns under various scenarios (Anderson *et al.*, 1994). They evolved from a more primitive modeling system called the *Urban Transportation Modeling System* (*UTMS*). Land-use activities represented by households (people) and jobs (employment) are not modeled and assumed fixed in this system. However, in the long run, both residential and commercial activities change spatial forms in the urban area. Transportation models that fail to take account of changes in land use structure will make erroneous transportation flow projections. Integrated urban models present a solution to the problem by modeling the changes in the urban activities in the long run through a land use modeling system of people and jobs and use the results generated as an input into the *UTMS* in an iterative fashion. Given that the objective of this study is to model the employment activities, the emphasis in this review will be on this part of the land use model of the *IUMs*.

In the field of integrated urban modeling, intraurban employment modeling is used to forecast the changes in the spatial distribution of employment over space. Employment distribution in the region is affected by the economic structure of the urban region where employment is divided into a number of different economic sectors. The fact that the economy comprises of different types of sectors, implies that interaction exists between these sectors or some pairs of them. Interaction between the different sectors of the economy affects the level and spatial distribution of employment in the region. Although spatial interaction models, such as gravity models, are capable of predicting the distribution of employment over a region divided into a finite number of zones, the fact that inter-linkages between the different economic sectors exist might lead to bias in predicting the long run future. An efficient system requires both capturing the aspect of inter-linkages between the different activities of the economy and then distributing these actives over space. Input-Output modeling is a powerful tool used to capture the inter-sectoral relation existing between the different sectors of the economy where the output of each sector is a function of the final demand of its own sector and other sectors of the economy. Thus, a good approach is to merge an input-output model with a spatial interaction model to account for the distribution of the diverse employment types in the region.

Putman (1990) developed the first operational integrated urban model *ITLUP*. The employment location model *EMPAL* of this system is a spatial interaction model as derived from the theory of entropy maximization. Employment at zone *i* is calculated as a function of the population, type of land use in zone *i*, zonal total employment in the base year, and travel cost in terms of distance between each two zones. The structure of this employment model indicates that it is a population driven model since population stands for the major factor that effects the distribution of employment in the geographical region. Also land use by type affects the redistribution of employment.

Wegener (1986) developed the *DORTMUND* model for the city of Dortmund in Germany. Modeling employment activities in the *DORTMUND* model is based on random utility theory where a nested logit model is employed to predict the spatial distribution of firms in Dortmund. The spatial distribution of urban activities is allowed to change within the modeling process in two ways. The first occurs through aging of firms, which in the model depends on time. The mechanism of the aging process involves the probabilistic Markov process, which is applied once in each model iteration. The other way is the recognition that the opening or closing of large industrial plants may not be predictable by any modeling system, thus, they are determined exogenously in the model. All other changes depend on accessibility based spatial choices generated explicitly within the nested logit model.

De la Barra (1984,1990) developed the TRANUS model. The model makes use of random utility theory to derive a multi-regional input-output model that incorporates the same results one can achieve using a traditional multi-regional input-output model. The model is driven by an exogenous final demand so that the total output of sector n (representing employment in sector n) in region i (representing zone i in the metropolitan area) is a function of the final demand and the technical coefficients estimated for zone i. However, the model makes use of discrete choice theory and links a multinomial logit model to the input-output model to predict the gross output for each activity. The elements forming the systematic utility of the logit model in TRANUS are the composite cost of transport for activity n from zone i to zone j and the value of land in zone j.

The modeling approach in TRANUS might not be practical for two main reasons. First, the multi-regional input-output model depends on the trade coefficients, which must be estimated for the whole set of zones (regions). Technically speaking, this will force the model to deal with a huge coefficient matrix that is not easy to handle, especially, when the urban area is divided into a fairly large number of zones and the economy into a large number of economic sector. Second, even if the approach is adequate, usually the limitation of data needed to estimate this type of model makes it an unpractical modeling approach.

Hunt and Simmonds (1993) discuss the structure of an integrated urban model *MEPLAN* of land use and transportation where they developed a model that predicts land use changes. The land use part of the model relies on three widely used but usually independent types of economic theory: input-output modeling, demand functions and random utility choice modeling in a spatial context. They noted that the mathematical structure of the input-output model is still used but the sector definition is altered so that the model predicts industries' demands for labour and residents' demands for services. Demand functions are used to make the coefficients describing the interactions between sectors elastic with respect to prices and income. This is important if the spatial model is to reproduce variations in the density of development and activities. The random utility discrete choice modeling approach is used to handle the spatial choice elements. In choice modeling it is assumed that economic actors select the alternatives (zones) they consider being the best, subject to specific constraints.

Apparently, all the employment models discussed above make use of random utility theory, spatial gravity and input-output analysis to predict employment at the zonal level. In this context, it becomes suitable to deploy our conceptual framework within the framework discussed above. Nevertheless, our model will not replicate any of the above models. Also the conceptualization will be affected by the available data for the region.

2.3 Factors Affecting the Choice Location of Firms in Urban Areas

Predicting employment at zonal level scale might not only be needed to serve transportation purposes, but also to monitor urban changes that will help to inform policies and set up plans for the development of cities and urban centers. In this context, a fair amount of research was found on this subject, which predicts employment location at different scales of zonal levels for different geographical regions. These are summarized in the following table:

Table 2.1 Summary of the Employment Location models

Author	*Study Area*	*Modeling approach and objective*
1. Vahaly (1976)	Census tracts of Nashville Davidson County, USA	Linear discriminant model; location of service and office activities
2. Clapp (1980)	Los Angeles, USA.	Linear regression model; intra-metropolitan location of offices.
3. Erickson and Wasylenko (1981)	66 municipalities of Milwaukee region, USA	Binary regression model; location of relocating firms in the major industries.
4. Lee (1981)	city of Bogota, Colombia.	MNL model; locational choice of different types of firms
5. Hansen (1986)	State of Sao Paulo, Brazil	Nested logit model; locational decision of manufacturing firms
6. Ihlanfeldt and Raper (1990)	299 census tracts of Metropolitan Atlanta, USA.	Tobit model; project new office sites.
7. Shukla and Waddell (1991)	141 zip code zone for the Dallas-Fort Worth area, USA.	Multinomial logit model; location of new firms in six different industries
8. Gottlieb (1995)	Northern New Jersey, USA	Regression model; intra-metropolitan firm location
9. Coffey *et al.* (1996)	Montreal (CMA), Canada	Binomial logit model; locational pattern of high order services firms
10. Yarish (1998)	Toronto (CMA), Canada	MNL model; office firms location

The literature presented in Table 2.1 will be synthesized according to the factors that influence the locational decision of firms in cities. Table 2.2 summarizes these factors with the purpose of using them in the different employment models.

2.3.1 The Effect of the CBD

In modeling the spatial distribution of firms in a metropolitan area, it is important to understand the structure of that area to allow for a better understanding of the process, thus, enabling the researcher to model it adequately. Urban formation is different from one city to the other and so it is impossible to devise the same type of model for two cities unless they are identical in structure and form. Despite the fact that urban areas maintain a unique structure, there are some common factors that exist between the different cities and are believed to have an influence on the decision shown by a firm when locating at a specific site. The first of these factors is the relative location of Central Business District (CBD) with respect to other patterns of the city. The CBD is considered as the heart of commercial activities in any city of the world. Despite development that takes place with time, the CBD retains its role as a major landmark in the urban area. Other patterns can always be measure relative to the core of the city either in terms of centralization or decentralization. The importance of the CBD as a factor influencing the location of firms in metropolitan areas is well documented.

Table 2.2 Intra-urban factors affecting the location of firms

Intra-Urban Factor	*Purpose of using the variable*	*Studies by authors name*
1. Central Business District (CBD) proximity	a. Monocentricity in the urban area b. Decentralization and land prices	• Vahaly (1976) • Clapp(1980) • Erickson and Wasylenko (1980) • Lee (1983) • Ihlanfeldt and Raper(1990) • Shukla and Waddell (1991) • Coffey et al. (1996) • Yarish (1998)
2. Highway and Expressway proximity	Accessibility and location decision	• Erickson and Wasylenko (1980) • Ihlanfeldt and Raper(1990) • Shukla and Waddell (1991) • Yarish (1998)
3. Mall and airport proximity	Multi-nucleated and location in urban center	• Ihlanfeldt and Raper(1990) • Shukla and Waddell (1991) • Yarish (1998)
4. Land Prices	Land price on location decision	• Clapp (1980) • Hansen (1987) • Yarish (1998)
5. Population Size and Income	Population size and income on profit and location	• Ihlanfeldt and Raper(1990) • Shukla and Waddell (1991) • Yarish (1998)
6. Residential Amenities	Relation between residential and commercial amenities on the location of firms	• Gottlieb (1995) • Yarish (1998)
7. Agglomeration Economies	Localization and urbanization economies	• Erickson and Wasylenko (1980) • Ihlanfeldt and Raper(1990) • Shukla and Waddell (1991) • Yarish (1998)

Vahaly (1976) found that access to the CBD is an important factor that affects the location of accounting firms, health and government services in the model he developed for Nashville-Davidson County. Erickson and Wasylenko (1981) tested the importance of the distance from the CBD in their model as a variable affecting the decision of location for firms in the City of Milwaukee. They used the reciprocal of the distance from the CBD and found that construction, manufacturing, transportation, wholesale, retail trade and services tend to locate on land away from the center of the city. Finance Insurance and Real Estate (FIRE) firms show good affiliation with the core of that city. Similar results with respect to the CBD effect on the locational decisions of firms is shown in a study completed by Shukla and Waddell (1991) where they found that firms in the different sectors, except the FIRE firms, escape the CBD in favor of the suburbs. Coffey *et al.* (1996) test the effect of the CBD on attracting high order firms. They found that these firms cluster in the CBD and in close proximity to it. Ihlanfeldt and Raper (1990) test the relative effect on the location decision of new office firms with respect to the CBD. The results show a decentralization trend with respect to the core.

In two studies with different employment sectors for Bogota and Cali, Lee (1981,1983) found that the net outflow of jobs from the CBD was relatively high, showing a decentralization trend from the core of the city. Yarish (1998) included the distance to the CBD in the model he developed for the intrametropolitan location of new office firms in Metropolitan Toronto. He found that distance to the CBD is a strong factor influencing the decision of location. Clapp (1980) found that office activities need access to the CBD in order to conduct business, which assumes that the CBD is an important factor influencing the decision of location for office firms.

The affect of the CBD on the locational decision of firms is evident as noted above. Different types of firms are sensitive to the CBD proximity either in locating at the core or far away from it. Thus employing such a factor in an employment location model is important and should be considered when developing the model.

2.3.2 The Highway and Expressway accessibility

Hanson (1986) states that the construction of highways and expressways after the World War II have worked on reshaping every corner of urban America as the new suburbs they engendered represented nothing less than the turning inside out of the historic metropolitan area. Anderson *et al.* (1996) argue that urban areas began to change shape at the onset of public transportation, such as cable cars, allowed for greater separation between land uses. As a result, cities became more dispersed and urban land uses began concentrating in specific areas, which acquires specific characters, such as commercial, industrial and/or residential.

Simmons *et al.* (1998) point that decentralization of the commercial market is accompanied by decentralization of commercial activities along major arterials such as major roads, highways and expressways. Erickson and Wasylenko (1981) tested for the importance of locating near highways. Their findings suggest that firms of different types tend to locate near highways. Shukla and Waddell (1991), Ihlanfeldt and Raper (1990) and Yarish (1998) achieved the same results when testing the influence of highways and expressways on the location of firms in the metropolitan area.

Coffey *et al.* (1996) argue that accessibility by automobile to firms in inner and outer parts of Montreal suburbs stands for as an important pull locational factor that influence firms in high services. This suggests that areas reachable by highways and expressways are more desirable places to locate at. Vahaley (1976) found that firms in FIRE, professional services, warehousing, manufacturing, health and government offices required proximity to the local interstate network because of easy access to population and workers residence.

2.3.3 Urban Nucleations as a Pull Factor

Decentralization is a common documented phenomenon observed in North American and industrialized cities (Bourne 1991; Yeates 1990; Hanson 1986; Wegener 1986; Lee 1985; Simmons *et al.* 1998 and Anderson *et al.* 1996). With development taking place in the urban areas decentralization from the core of the city is observed and a net loss of population and jobs at the core is evident over time with an attempt to occupy and settle in the suburbs. In the long run, this leads to a polycentric urban structure with the formation of clusters of mixed residential and commercial activities, called nucleations.

In a recent study by Jones *et al.* (1998) on the commercial structure in the Dallas Forth Worth CMSA, some interesting results were obtained in which they found that commercial firms tend to cluster in box stores of 1000 - 5000 sq. m around regional malls. Similar observations were attained by Simmons *et al.* (1998) when studying the commercial structure and change in metro Toronto. Yarish (1998) found that mall proximity is an important factor that attracts some types of office firms in metro Toronto. Clapp (1980), Danials (1982) and Ihlanfeldt and Raper (1990) also suggested this type of variable to have an influence on the decision of location for some types of firms. They proposed that proximity to malls will lower labor costs where employee amenities, such as restaurants, shopping, and entertainment exist.

In identifying the patterns of the city that drives a firm to locate at a site, Yarish (1998) and Shukla and Waddell (1991) use the proximity to the airport as an additional factor to proximity to the CBD or proximity to malls. Shukla and Waddell found that firms in wholesale, retail and services value accessibility to the airport. Yarish (1998) found that new office firms in financial services, travel services, and industrial services show high affinity in locating to the airport in Metro Toronto.

2.3.4 The Impact of Land Prices

From the hypothetical point of view, land prices tend to have a major influence on the decision of location in metropolitan areas. Firms in the different industries, other things being equal, seek to locate on land that is as inexpensive as possible. Nevertheless, this is not the case in some studies where factors other than the land price have more significant roles on location. Hansen (1987) found that land prices do not add any explanatory power in his model of industrialized location choice. Other studies show that land prices are important and influence the decision of location at some parts of the city. Coffey *et al.* (1996) report that land prices rank as the first factor affecting the location of firms in the CBD of Montreal, and rank as the third factor at the CMA level. Though land price at the CBD is the most important factor, it does not appear to be the most critical among the other factors. Other factors like accessibility for clients, and visible and prestigious location tend to have more influence on location choice in Montreal.

Expensive land at the core is always associated with a decentralization and preference of locating away from the CBD. In this context, CBD proximity is used as a surrogate for land prices. Erickson and Wasylenko (1980) used this type of measurement to capture the effect of land prices on the locational decision of firms in the municipalities that form the Milwaukee region. This was the case since they believed that land price values decline as distance increases from the CBD. Yarish (1998) employs the gross average rental rate in his location model to test for its location sensitivity on new firm offices. His findings suggest that except for manufacturing offices and law firms, all other office firms are sensitive to rental rates and tend to locate on less expensive land. Clapp (1980) used rental rate as a variable in the model he developed to test its effect on the intrametropolitan office location for Los Angeles. His findings overlap with what is suggested by Erickson and Wasylenko (1980) regarding the relation between land prices and the CBD. He found that buildings farther away from the CBD commanded less rent and so reflects inexpensive land.

2.3.5 The Forces of Agglomeration Economies

As mentioned earlier, modeling employment in an urban areas entails modeling the economic process that takes place over time in that particular area. In this context, agglomeration economies are defined and discussed. Selting *et al.* (1995) define four important components of agglomeration benefits: transfer economies, internal economies of scale, localization economies and urbanization economies. Transfer economies are defined as the transport savings a firm detains by locating close to its market or near other firms that supply it with inputs. Internal economies of scale are the reductions in the long-run average cost curve of a firm that occur when output expands (Selting *et al.*, 1995). Yarish (1998) defines localization economies as the benefits attained by co-location of competitors. These are advantages gained by locating in specialized clusters providing specialized labor and technological needs. Also, these economies reduce production costs by allowing firms to identify rapidly changing needs and allowing them to experiment with new products and production processes that require frequent face-to-face contact. Moreover, urbanization economies are the benefits a firm achieves by locating in metropolitan areas.

Selting *et al.* (1995) highlights two major issues that involve agglomeration economies. The first is that agglomeration economies are instrumental in explaining firm location and the importance of examining net social costs and benefits of firm location remain an important research task. The second issue is the need to isolate the importance of localization versus urbanization economies. Selting *et al.* (1995) argue that urbanization and localization economic benefits are strikingly dissimilar and their magnitude in firm location dictate different policy directions.

The effect of agglomeration economies is tested by many researchers to see their impact on the locational decision of firms in urban areas. Clapp (1980) found that agglomeration economies is effective in satisfying the needs of a group of office activities when clustering together on less expensive land. Erickson and Wasylenko (1980) test the effect of localization economies on the location decision of firms in the municipalities of Milwaukee. The agglomeration variable in their model was defined as the ratio of employees in a particular economic sector in each municipality to all non-central city employees in that specific sector. The findings suggest that agglomeration economies exist among manufacturing, construction, wholesale, retail and FIRE service industries.

Shukla and Waddell (1993) define three agglomeration variables in the employment model they developed for the Dallas-Worth Forth region. The variables are derived according to a spatial interaction formula. In building the agglomeration variables, they used accessibility to employment. The variables are defined so that they can capture urbanization and localization economies by accounting for intra-industry and inter-industry spatial linkage opportunities. Their findings suggest that agglomeration economies exist among different types of sectors of the economy and affect the decision of location. Yarish (1998) used accessibility to employment in different types of offices to test the effect of urbanization and localization economies on the location decision of offices. His findings suggest that offices of advertising agents, media and communication, financial services, other business services, artists, travel agents, personal services, associations and government services display an affinity for central locations mainly because these industries prefer the benefits of urbanization and localization economies.

2.3.6 Population size and Income Effects

Studies have shown that population size affect the location decision of firms in metropolitan areas. Some types of firms are population-oriented and attempt to locate in populated areas to enjoy the benefit of forming a market where clients (population) are available. Shukla and Waddell (1991) included the population size as a variable in their model. They found that all types of firms show a tendency towards locating in populated areas.

Population income was used in employment location models to test its affect on the location decision of firms in metropolitan areas. This is evident for population-oriented firms such as retail trade firms and services. Jones *et al.* (1998) note that household income affect firm location in Dallas-Fort Worth in an indirect way. They argue that all commercial activities are automobile oriented, since middle class households spend as much as 20% of their income on transportation. Ihlanfeldt and Raper (1990) used the number of households at the census tract level, which are below the poverty level to see the effect of income on the location decision of firms. The variable appeared to have a strong influence on location of branch offices. Hansen (1987) used the wage rate of semi-skilled workers by sector in the nested logit model he developed to predict the probability that a firm locates in a non-metro municipality. The variable turns out to be insignificant.

Yarish (1998) employs two variables to test the effect of income on the location decision of office firms in metro Toronto. The first variable is the number of households in zone below $10,000 income level; the other variable is the number of households in zone above $70,000 income level. His findings suggest that office firms in advertising agencies, computer services, law firms, media and communication, financial services, other business services, health services and personal services prefer to locate at zones with high income. Trade services show an affinity to locate in zones with low-income households.

Shukla and Waddell (1991) used the median household income in their model. Their results have shown that all firms except manufacturing and retail firms hold a positive relation with the income. However, they justify the negative relation of the income variable for manufacturing firms by noting that the income might have a negative sign if low rents compensate for proximity to incompatible or fiscally onerous non-residential uses.

2.3.7 Other Factors Affecting the Location of Firms

In modeling the changes in employment for an urban area, it is important to test the factors that drive a firm to escape a site and locate at different locations. Firms search the metropolitan area to locate at a site that will maximize their profit (Storey, 1990). Profit maximization is possible by selecting a site that can guarantee the highest amount of profit. At the zonal level, if a firm relocates to another site, a loss will take place. Loss in firms at the zonal level might also occur if a firm goes bankrupt. Bankruptcy occurs when the firm can not maintain the normal profit it is supposed to have. That is, a firm is said to go into bankruptcy if its profit becomes less than the normal profit. Miron (1982) defines the normal profit of a firm to be the flow of profit to a business enterprise that must be maintained in the long run for that firm to find it worthwhile to remain in business. In relocation and bankruptcy, loss of firms at a specific zone occur subject to certain push factors. These push factors work on declining the profit of the firm. When decline starts, some firms will take precaution procedures by relocating elsewhere whereas others will have a sharp decline in profit-they go into bankruptcy.

Some studies consider the factors causing firms profit to decline as push factors affecting the locational decision of firms. Tervo and Niittykangas (1994) point that firm closure may be following from changes in the structure and size of demand. They argue that firms that are small in size are more likely to close in suburban municipalities, noting that success is more likely to occur at the center of the urban area where customers are available. Coffey *et al.* (1996) show that insufficient space to expand rank as the first push factor influencing all high order service firms to leave an intrametropolitan site and relocate elsewhere. They also show that high land prices stand as an influential factor that effect the decision of some firms in Montreal to leave a site and relocate at a different location. None of the high-order services firms show sensitivity to high local taxes as a push factor effecting the decision to leave the previous site.

Competition is believed to have an influence on location decision in metropolitan areas. Jones *et al.* (1998) indicate that competition among retailers is intense in Dallas-Forth Worth, which affects shopping, and thus, affects sales that will increase the profit of one retailer and decrease others. Simmons *et al.* (1998) point that competition in Toronto GTA has created a hierarchy of commercial nucleations, where firms escape a site and establish a new market in disperse nucleations all over the GTA due to competition. Lea (1989) advises methods and theories that affect the location decision of retail firms. He lists a number of class variables that can be included in a retail location model where competition stands for one of the important factors affecting the location decision.

Gottleib (1995) conducted a regression analysis at the intrametropolitan level to test the effect of residential amenities on the location of firm in the professional services. His regression model included information on the levels of traffic congestion, crime, pollution, recreation, public education and public services. His findings suggest that professional service firms tend to locate in municipalities with high density of amusement employment. Municipalities with high level of crime and pollution are not a desired location.

2.4 Conclusion

In this chapter, the theories and methodologies used to develop employment models for both transportation modeling and urban modeling were discussed. Several methods were employed to develop models for the purpose of predicting the commercial land use changes in *IUMs*. Employment models in the *IUMs* depend on input-output modeling, discrete choice modeling and spatial interaction modeling.

Moreover, studies in urban modeling have shown that firms locate at a site that will maximize their profit. Location decision is influenced by the urban structure of the city and appears to affect the profit of the firm. In this context, Table 2.3 summarizes the empirical findings of this review in terms of the intra-urban factors affecting the locational decision of firms.

Table 2.3 Intraurban Location Factors - Empirical Findings

Factor	***Empirical Findings***
CBD Proximity	+ / -
Highway Proximity	++
Mall Proximity	+
Airport Proximity	+
Land Price Impact	+/ -
Population Size & Income	+
Residential Amenities	+
Agglomeration Economies	++

Key:
+ = positive effect on location, ++ = strong positive effect
- = negative effect on location, -- = strong negative effect
+/- = positive effect on some industries, negative effect on others

Table 2.3 indicates that the some factors stand for as strong variables affecting the location of firms in metropolitan area. Variables such as the highway proximity and agglomeration economies appear to have a very positive effect on the locational decision shown by firms. Mall and airport proximity appears to have a positive effect but not as much strong as that shown by the highway and agglomeration variables. Residential amenities along with population income exhibit a positive effect. CBD proximity and land prices show to have a controversial effect on location. Some studies shown that the CBD exerts a positive influence on the location of some types of firms, other studies shown that the CBD is not a desirable place to locate at. Moreover, land prices appeared in some studies to influence the locational decision of firms. Some other studies shown that the land price does not add any power to the modeling process.

The factors presented in this review will be used in our model according to their importance and availability. However, we will include any other variable we think it might have an influence on the location decision of firms in the region. As mentioned earlier, each urban center has a unique structure, therefore, variables in our model will be included after we analyze the spatial structure of the CMA to see what variables will be most suitable to include in the model as we will see in the coming chapters.

CHAPTER THREE

DATA AND METHODOLOGY

3.1 Introduction

The chapter starts by introducing the spatial and geographical characteristics of the study area chosen for this research. An overview of the urban structure is discussed with some emphasis on the major land uses in the area. Following this, the data sets used in the thesis are described, along with the methods used to extract and arrange the data. Kernel estimation as an exploratory tool is discussed and kernel maps are produced to show the clustering of firms in the region. General figures are extracted and discussed to give an idea on the employment trend in Hamilton-Wentworth. In the last part of this chapter, the conceptual framework of the employment model is developed. This is a formal structure in which the location and redistribution of firms infer the spatial distribution of employment. The core of the structure consists of three linked models: the input-output model, the destination choice model and the lost firm model. Each one of these models is developed to fulfill a specific purpose. The result of the linked models is the employment at the census tract level.

3.2 The Study Area

3.2.1 Spatial Units

The region chosen for this study is the Hamilton Census Metropolitan Area (CMA). The CMA is made up from the following municipalities: Ancaster, Burlington, Dundas, Flamborough, Glanbrook, Grimsby, Hamilton and Stoney Creek. The total amount of land in the CMA is approximately 1377 squared kilometers. In terms of distance, the east to west extent is approximately 61 kilometers and the north to south extent is approximately 48 kilometers. A map for the region is shown in Figure 3.1. The region under investigation was chosen for two main reasons. First, the model developed will supply an important subsystem to IMULATE, an integrated urban model developed for the Hamilton CMA. Second, the availability of data for this region makes it an appropriate environment for this type of research especially that no previous research involving the development of an employment location model has been conducted for this area.

The employment in the region is predicted at the census tract level because IMULATE is based and calibrated using the 1986 census data. Thus, the census tracts, as defined by statistics Canada, for the year 1986 census, are used (Figure 3.2). The CMA consists of 151 census tracts.

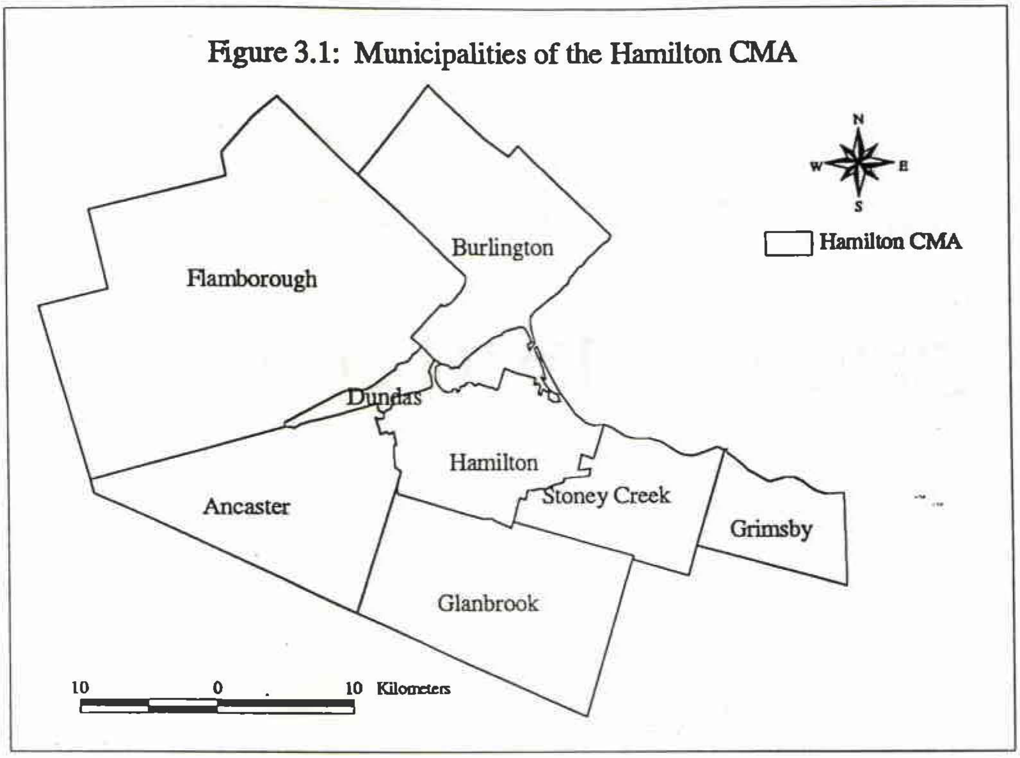


Figure 3.1: Municipalities of the Hamilton CMA
N
W
E
S
Hamilton CMA
Flamborough
Burlington
Dundas
Hamilton
Stoney Creek
Grimsby
Ancaster
Glanbrook
10
0
10 Kilometers

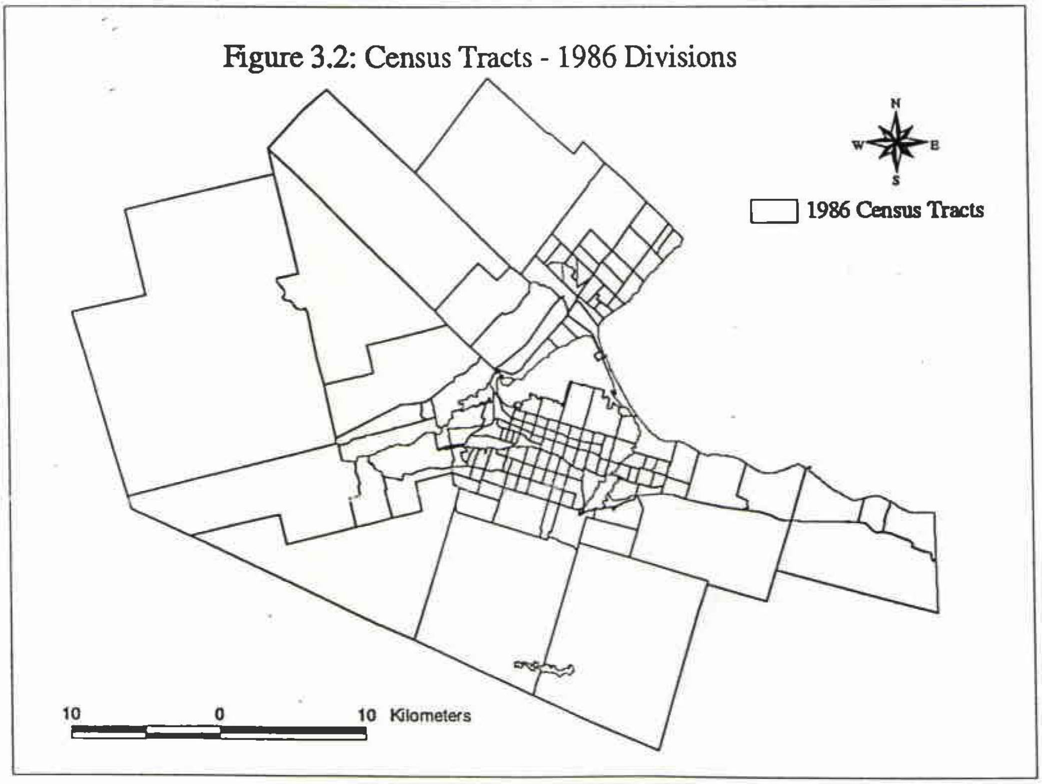


Figure 3.2: Census Tracts - 1986 Divisions
N
W
E
S
1986 Census Tracts
10
0
10 Kilometers

3.2.2 Urban and Economic Structure

As shown in Figure 3.2, the city of Hamilton has a large number of census tracts compared to other municipalities of the CMA, there are 92 census tracts, or 61 % of the total number of tracts in the CMA. On the other hand, the census tracts for the city of Hamilton are smaller in size than those for other municipalities. This is because the population at these tracts maintains high density. Therefore, most of the CMA population is concentrated in the city of Hamilton, especially, in the lower part of the city. However, with development taking place, an increase in population density in the upper part of the city, as well as the other municipalities is expected over time. Figure 3.3 shows the household density over the various census tracts in terms of low, medium and high density. Also Figure 3.4 shows the population density in the CMA. Like many North American cities, Hamilton enjoys having a developed network of highways and expressways that links the different parts of the CMA with each other via circumferential and radial highways and expressways. This can be seen in Figure 3.5.

Hamilton is considered as one of the heavy industrial areas in the Province of Ontario. The steel industry plays a major role in the economy of the city with two major industrial establishments - Stelco and Dofasco - occupying the census tract land around Hamilton Harbor. Figure 3.6 shows the land occupied by manufacturing industries in the region. Furthermore, despite the fact that Hamilton is a heavy-industry area, retailing and services play a major role in the economy. With five major shopping malls in the region, Hamilton is considered as a multinucleated city rather than a monocentric one. This gives a special urban form where businesses in different industries cluster around these nodes

Figure 3.3: 1996 Household density in Hamilton-Wentworth

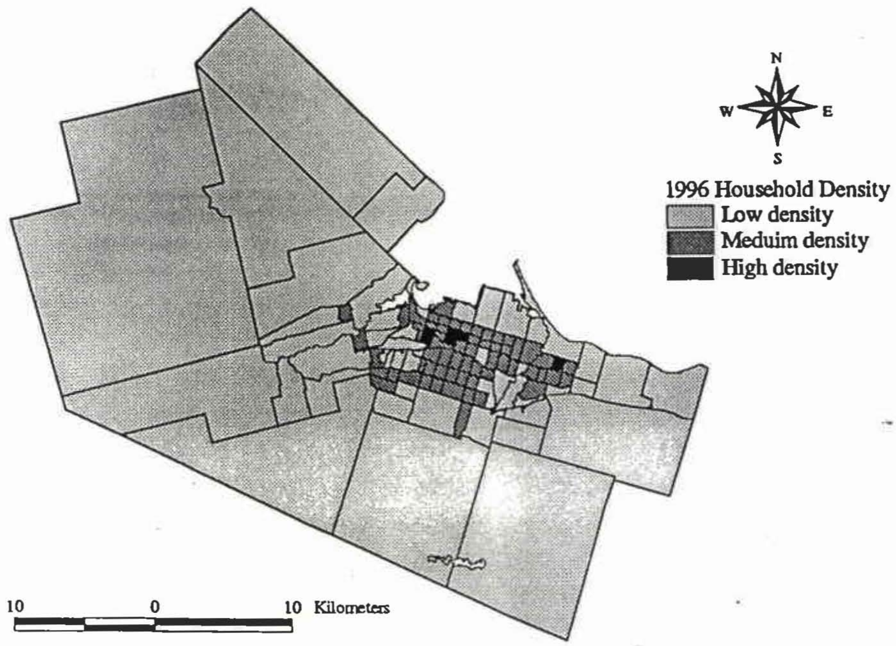


Figure 3.4: Spatial Distribution of 1996 Population

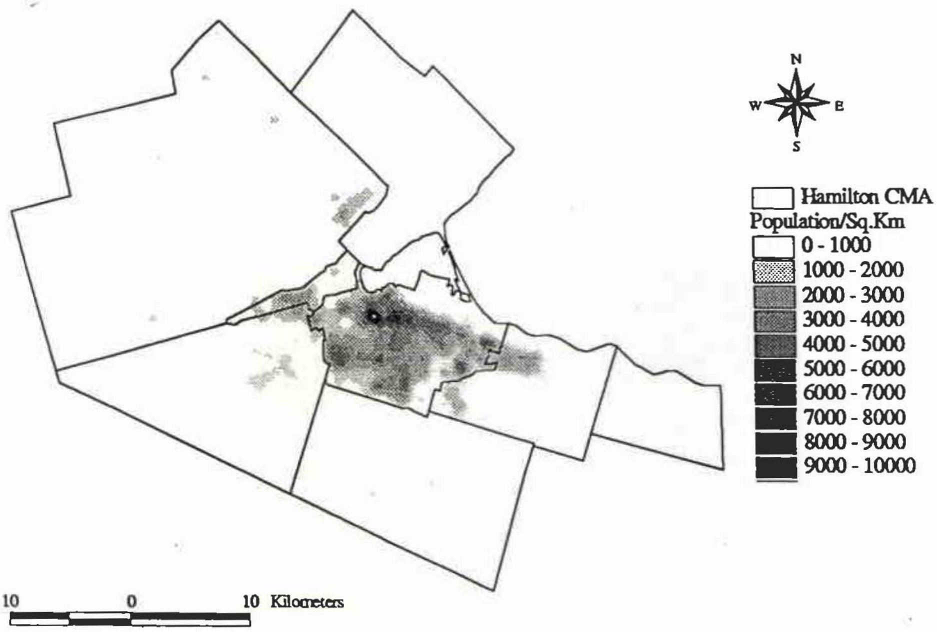


Figure 3.5: Highways and Expressways in Hamilton CMA

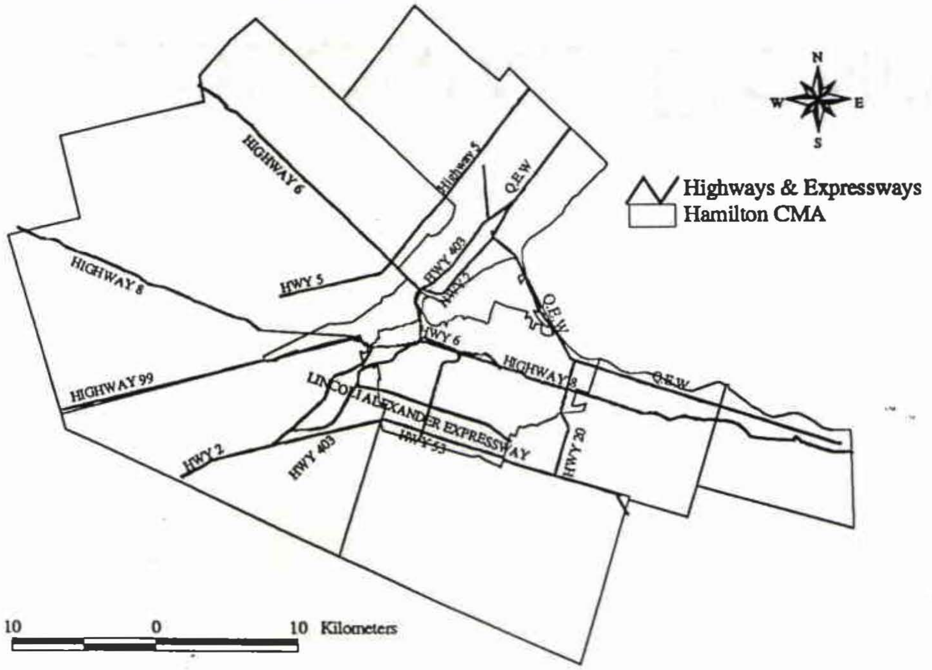
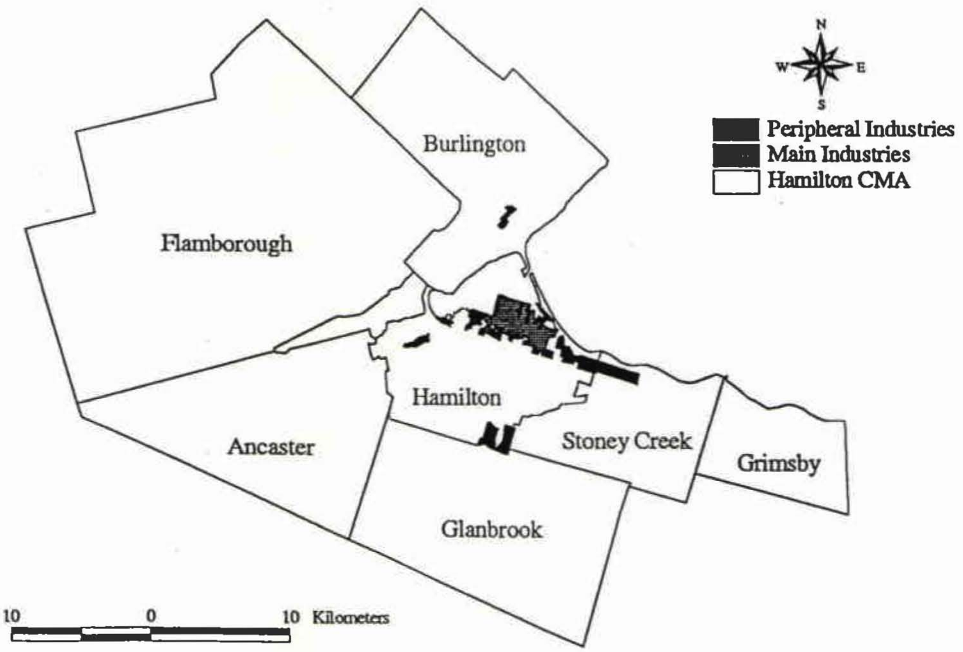


Figure 3.6: Manufacturing Landuse in the Hamilton CMA



rather than locating and clustering only at the CBD. Figure 3.7 shows the CBD and the major shopping malls in the region.

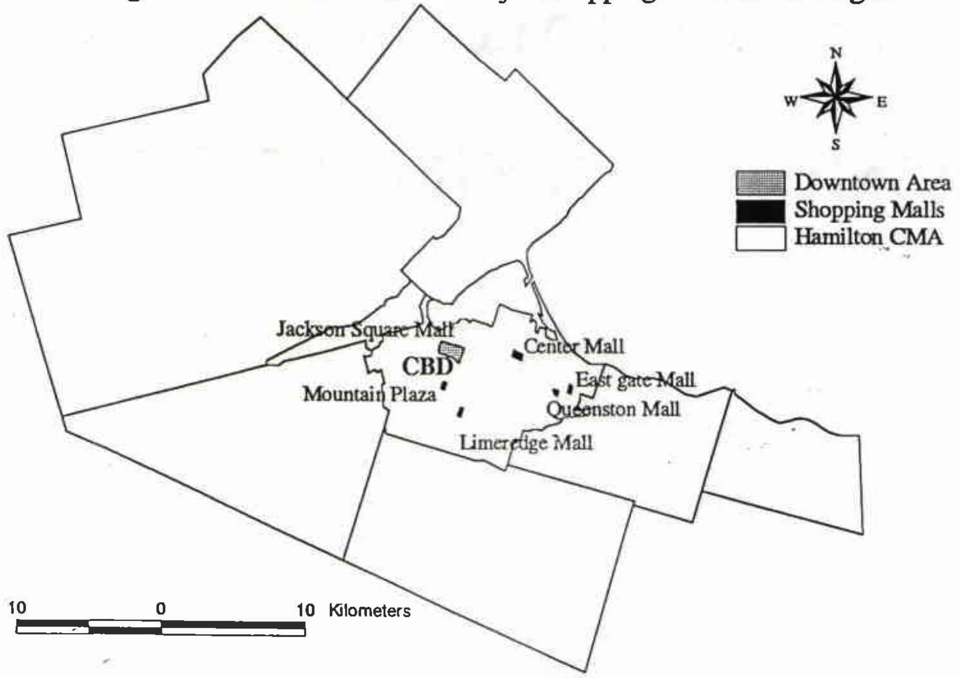


Figure 3.7: The CBD & The Major Shopping Malls in the Region

3.3 The Data Set

Three sets of data are used in this thesis. They are employment data, socioeconomic data and economic data. The employment data are obtained from the Regional Municipality of Hamilton-Wentworth for the years 1990 and 1997. The Hamilton-Wentworth region encompasses all the municipalities of the CMA with the exception of Burlington and Grimsby. Unfortunately, employment data for these two municipalities are not attained. Therefore, the study area will be restricted to the Hamilton-Wentworth region. The available data are based on two employment surveys conducted by the Regional Municipality of Hamilton-Wentworth. The 1990 employment survey was conducted over a seven month period between May and December 1990. It aimed at identifying the type of business activities located in Hamilton-Wentworth and the number of people employed in these activities. The survey covered the entire population of firms in Hamilton-Wentworth (Regional Municipality of Hamilton-Wentworth, 1992) with the exception of farming activities. Although it is difficult to estimate total agriculture employment, approximately 4000 people were listed in the 1986 census as living on a farm in Hamilton-Wentworth. Some businesses were probably missed, but this is a small percentage of total employment and the affect on the aggregate result is small. On the other hand, because the survey was conducted over a 7 month time period there was the potential for missing businesses that relocated. It is also possible that some firms were surveyed twice. Therefore, the 1990 employment survey probably undercounted the total number of people working in the region by an estimated 5000 people. The total number of firms surveyed was 11570 and only 45 declined to complete the survey. The response rate for the survey was extremely high with almost 99% of employers contacted, agreeing to complete the survey. Furthermore, each business in the survey was categorized using both the 1980 version of the Standard Industrial Classification (SIC) and the Regional Information System's Committee (RISC) codes. The database included the total number of employees in each firm and the geographical coordinates represented in the UTM coordinate system. The Regional Municipality provided an updated database for the businesses in the year 1997. The 1997 database did not include the number of employees in each firm, otherwise, it has the same characteristics as the 1990 database.

Socioeconomic data represented by the household numbers, population size and employment by place of residence for Ontario and the Hamilton CMA for the year 1991 and 1996 were obtained from Statistics Canada. The spatial distribution of the household and population among the various census tracts is given. The economic data used in this thesis is the transaction tables for firms in different industries in the year 1990. This was also obtained from Statistics Canada. These transaction tables are for the province of Ontario. However, for the sake of the model, these are scaled down to the level of Hamilton-Wentworth. Methods for estimating the regional transaction tables from the provincial accounts are discussed in the next chapter.

3.3.1 Extracting the employment data

Next the methods used to extract the required data to develop the model are discussed. The focus of this section is a discussion of the method used to determine the stayer firms-those that remained in the same location under the same SIC number between 1990 and 1997. As mentioned earlier, the employment databases for both years 1990 and 1997 are characterized by the geographic location (x and y coordinates) and by the SIC code. We made the assumption that firms, which maintained the same geographical location with the same SIC number are the firms that remained in the same place between 1990 and 1997. We will call these stayer firms. A script in the Visual Basic 4.0 programming language was written by the author in order to compare the two databases and extract this type of information. The program was written in a way that reads a record from the 1990 database then reads a record in the 1997 database. If those two records match by the coordinates and the SIC number, the program saves the record in an output file and removes it out from the original databases to count for duplicates. If they do not match, the 1990 record is removed and a new record is compared with the 1997 record. This was done until the last record in both databases is checked. The output of this process is a new database file with the number of remained firms between time period 1990 and 1997. The script also enabled us to extract the gain firms at time period 1997 by intersecting the resultant database of the remaining firms from the 1997 database.

ArcView 3.0a GIS was used in this section for two reasons. On the one hand, it was use to break down the employment database by the type of industry. This was done using the query builder in the database package of ArcView. The result is several databases that correspond to the various types of industries of the economy. Table 3.1 lists the various types of industrial sectors with the SIC ids for each industry. On the other hand, ArcView was used to aggregate the individual firms for each census tract. The database that contains the x and y coordinates was superimposed on the census tract GIS coverage and a link between the two coverages was established. Next a tool in ArcView was used to aggregate the points that lie on each census tract. The result is a new database with the census tract id and the total number of firms associated with that tract. This is repeated for the various industrial sectors.

Table 3.1 Industrial Sectors and the 2 digit Standard Industrial Codes

Industry	*SIC ID's*
Other Industries	**01 - 09**
Manufacturing Industries	**10 - 39**
Construction Industries	**40 - 44**
Transportation and Storage Industries	**45 - 47**
Communication and other Utility Industries	**48 - 49**
Wholesale Trade Industries	**50 - 59**
Retail Trade Industries	**60 - 69**
FIRE Industries	**70 - 76**
Business Service Industries	**77**
Government Service Industries	**81 - 84**
Educational Industries	**85**
Health and Social Service Industries	**86**
Accommodation, food and beverage service Industries	**91 - 92**
Other service industries	**96 - 99**

3.3.2 Kernel Estimation

Visualizing trends in the data under investigation is an important part of a modeling exercise. In order for us to develop an effective location model, it is important to look at the trend in our data and try to observe any kind of clustering in the region. In geography, there are many ways to visualize the data by producing different types of maps. One simple way is to plot the points representing each firm on the map and look for an evident trend. This is called a dot map where we look directly at the spatial distribution of events (firms) in space to observe any kind of clustering. However, there are more effective statistical methods to accomplish this. Kernel estimation, a well-known method, enables the researcher to observe the clustering of events in a geographical area. According to Bailey and Gatrell (1995) estimating the intensity of a spatial point pattern is like estimating a bivariate probability density. More formally, if **s** represents a general location in $\Re$ and $s_1, s_2, \ldots, s_n$ are the locations of the n observed events then the intensity $\lambda(s)$ at s is estimated by

$$\lambda_\tau(\mathbf{s}) = \frac{1}{\delta_\tau(\mathbf{s})}\sum_{i=1}^{n}\frac{1}{\tau^2}k\left(\frac{(\mathbf{s}-\mathbf{s}_i)}{\tau}\right)$$

Where:

- $k()$ - a suitably chosen bivariate probability density function known as the *kernel*, which is symmetric about the origin.
- τ - a positive parameter, called the bandwidth. It determines the amount of smoothing. It is the radius of a disc centered on s within which points s_i will contribute significantly to $\lambda_\tau(s)$.
- $\delta_\tau(s)$ - an edge correction function. It is the volume under the scaled kernel centered on s that lies inside $\Re$.

The edge correction function is given as follows

$$\delta_\tau(\mathbf{s}) = \int_{\Re}\frac{1}{\tau^2}k\left(\frac{(\mathbf{s}-\mathbf{u})}{\tau}\right)\, d\mathbf{u}$$

Figure 3.8 illustrates how the kernel estimation is obtained for a point pattern in region $\Re$.

Figure 3.8: (a) Slice through a quartic kernel, (b) Kernel estimation of a point pattern.

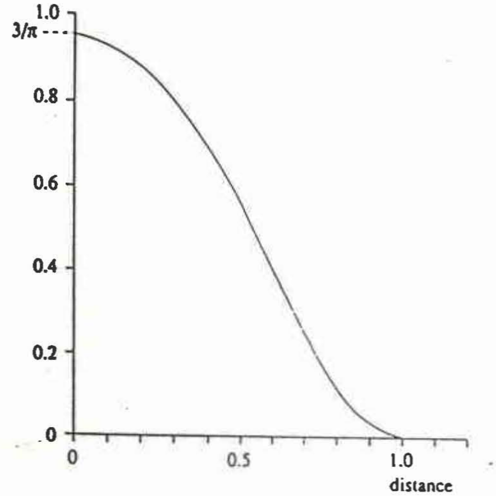


Figure 3.8: (a)

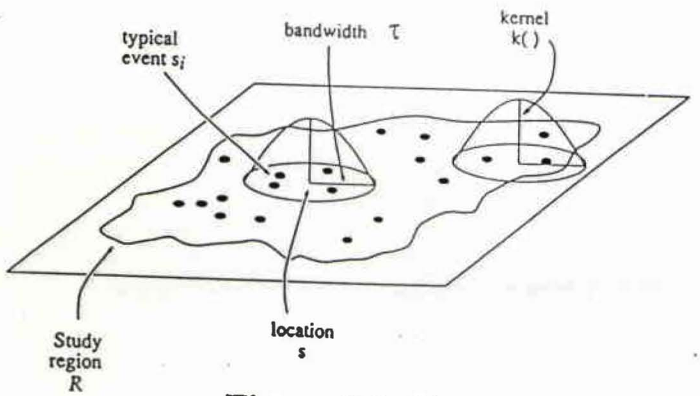


Figure 3.8: (b)

It is worth noting that the bandwidth τ effects the intensity and thus the clustering pattern. Therefore, it is the researcher who can determine what bandwidths will mostly suite the data to get a better visualization of the trend. For the purpose of this research, kernel estimates were achieved using the point pattern analysis "Splancs" library and the "Spatial" module of the Splus 4.5 statistical package. Results of the estimation are shown in Appendix A (Figures A.1 to A.25). Following are the major clustering trends shown by the different firms in the different industries. More details are provided into Appendix A.

- Manufacturing firms show high clustering in the harbor of Hamilton city and northern part of Stoney Creek. Also clustering of these firms is evident in the Southeast part of Hamilton.
- Construction firms show the same clustering trend shown by manufacturing firms.
- Transportation firms cluster in the northern part of the city of Hamilton especially in the northeast part close to the Q.E.W highway.
- Communication firms cluster in the northwest part of Hamilton. Clustering is more intensified in the core of the city.
- Wholesale Trade firms show the same clustering trend shown by manufacturing firms
- Retail Trade firms show high affiliation with the core of the city. However, retail firms exhibits a more dispersed clustering pattern in areas occupied by population.
- FIRE firms are clustered in the CBD area and around it.
- Business firms follow the trend shown by FIRE firms. Clustering is highly at the core of the city.

- Government firms cluster at the core of the city.
- Education firms show high clustering at the core. Also, education firms tend to locate in populated areas.
- Health and Social insurance firms cluster at the core of the city and in close proximity to it.
- Accommodation, food and beverage firms cluster at the core of the city and in close proximity to it.
- Other services follow the pattern of clustering shown by most service firms by locating at the core and in close proximity to it.

3.4 Employment location Model

3.4.1 Conceptual Framework

In this section, we introduce the conceptual framework and modeling approach we will use to develop the employment location model. We will start by looking at the general structure of the model and then explain the different components of it. The model is developed as a mathematical program that predicts the employment at the census tract level. Since employment is associated with firms, the objective is to predict the location of firms and from that to infer the employment.

In this context, consider a zone i of the metropolitan area. Assume that change in employment will take place between time periods t and $t + 1$. At time period t zone i holds F firms. As we move in time from t to $t + 1$ change in employment will take place. The change process is best described according to the following diagram:

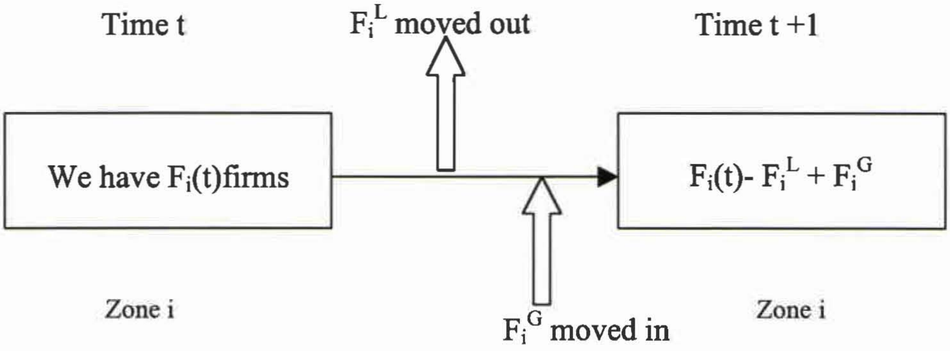


Figure 3.9 Change process in firms over time interval [t, t+1]

As can be seen from the chart, during the time period [*t*, *t+1*], F_i^L firms moved out from zone *i*. During the same period F_i^G firms moved in. The net change in the number of firms at zone *i* at time *t* + *1* can be thought of as a loss and gain process. The move out from the zone is considered as a loss to that zone and the move into the zone is considered as a gain to the zone. It becomes obvious that in order to develop a model that accounts for the net change in the number of firms at time *t* + *1* it is important to model the two processes. In order to model these processes, it is important to understand the reasons that drive a firm to move out from a zone or move into another. Firms move out from a zone for three reasons: the first is the move out to relocate elsewhere in the region, the second is the move out because the firm is bankrupted, and the third is the move out to relocate in a different region. In the first, the firm looks for a new location that will maximize its profit. In the second and third, the firm vanishes from the zone and the entire region. The model will not differentiate between the three processes and treat them as one-the loss of firms from the zone. Firms move into a zone for one explicit reason, profit maximization. Therefore, firms search the metropolitan area for a location that will insure the maximum amount of profit.

The modeling process involves modeling employment both at the aggregate and the disagregate levels[1]. Three different models for the different industrial sectors will be developed. An Input-Output model will be used to predict firms between time t and $t + 1$ at the aggregate level. Those have to be broken to the disagregate level. For this we make use of the random utility theory and employ a discrete choice model like a logit model, where we predict the probability that a firm in an industrial sector n will choose zone i as its destination. This will account for the gain process that takes place between time t and $t + 1$. The loss process will also be modeled by employing certain types of regression models. These models then will be linked together to get totals of firms at the census tract level. Finally, employment will be determined from those predicted firms as we will see in the following section.

3.4.2 Modeling Approach - Mathematical Structure

Consider a metropolitan area divided into M zones. In this study the metropolitan area is the Hamilton CMA and the zones are the census tract (CTs). Assume that the economy in the region is divided into N sectors. Let $F_i^n(t)$ represents the total number of firms belonging to sector n of the economy in zone i at time t. One has to keep in mind that the total number of firms at time t is the basis of the model and is used as an input for the base year. Also define:

[1] We mean by the aggregate level the Hamilton CMA region, the disagregate level is the census tarct.

- $F_i^{nL}(t, t+1)$ – number of lost firms in sector n of the economy that was located at zone i between time period t and (t + 1).
- $F_i^{nR}(t, t+1)$ – number of remaining firms in sector n of the economy that are located at zone i between the time period t and (t + 1)
- $F_i^{nG}(t+1)$ – number of firms in sector n of the economy that move into zone i at time period (t + 1), those are firms gained at zone i.

Then the total number of firms in sector n of the economy located in zone i at time period t is:

$$F_i^n(t) = F_i^{nR}(t, t+1) + F_i^{nL}(t, t+1) \qquad ...(3.1)$$

Also, the total number of firms in sector n of the economy located in zone i at time period (t + 1) is:

$$F_i^n(t+1) = F_i^{nR}(t, t+1) + F_i^{nG}(t, t+1) \qquad ...(3.2)$$

All terms in equations 3.1 and 3.2 are previously defined. One can make all terms in equation (3.2) represented in terms of the lost and gained by subtracting equation 3.2 from equation 3.1. The result is:

$$F_i^n(t+1) = F_i^n(t) - F_i^{nL}(t, t+1) + F_i^{nG}(t, t+1) \qquad ...(3.3)$$

Equation (3.3) represents firms at time $t + 1$ in terms of firms at time t, lost firms and gained firms. Given that firms at time t are known, the objective becomes to predict the lost firms as well as the gained ones. Once the lost and gained firms are determined, we can use the average number of employment in each industrial sector $AVE^n(t+1)$ to get the total number of employment in each zone. Finally, we can aggregate over all the modeled sectors to get the total employment at zone i:

$$ME_i = \sum_n AVE^n F_i^n(t) - \sum_n AVE^n F_i^{nL}(t,t+1) + \sum_n AVE^n F_i^{nG}(t,t+1) \quad \ldots(3.4)$$

In equation (3.4) we will use regression analysis to predict the total number of lost firms in each zone. This will be done for all the modeled sectors. As for the gained firms term, a multinomial logit model will be used to estimate the probability that a firm in sector n of the economy will choose zone i to locate at, that is $P^n(i)$. If we know the total number of gained firms at time t + 1 $F^{nG}(t+1)$, then gained firms at zone i in time period t +1 can be represented as a function of the predicted probability

$$F_i^{nG}(t, t+1) = F^{nG}(t+1)\, P^n(i) \quad \ldots(3.5)$$

The Input-Output model will be used to determine the total number of firms $F^n(t + 1)$ at time t + 1 for each industrial sector n at the regional level. Given that we are modeling firms in the different sectors of the economy, an input-output model is an efficient tool that capture the linkages that exist between the different sectors of the economy. Firms $F^n(t)$ at time t in the region are basically the aggregates of all firms in all zones, that is:

$$F^n(t) = \sum_{i=1}^{M} F^n_i(t)$$

Substituting for the value of $F^n_i(t)$ in equation (3.1) yields:

$$F^n(t) = \sum_{i=1}^{M} F^{nR}_i(t,t+1) + \sum_{i=1}^{M} F^{nL}_i(t,t+1) \quad \ldots(3.6)$$

In the same manner, total firms $F^n(t+1)$ at time t+1 are the aggregates of all firms in all zones:

$$F^n(t+1) = \sum_{i=1}^{M} F^n_i(t+1)$$

Substituting for the value of $F^n_i(t+1)$ in equation (3.2) yields:

$$F^n(t+1) = \sum_{i=1}^{M} F^{nR}_i(t,t+1) + \sum_{i=1}^{M} F^{nG}_i(t,t+1) \qquad \text{...(3.7)}$$

Subtracting equation (3.7) from equation (3.6) yields:

$$F^n(t+1) - F^n(t) = \sum_{i=1}^{M} F^{nG}_i(t,t+1) - \sum_{i=1}^{M} F^{nL}_i(t,t+1)$$

Rearranging we get:

$$\sum_{i=1}^{M} F^{nG}_i(t,t+1) = F^n(t+1) - F^n(t) + \sum_{i=1}^{M} F^{nL}_i(t,t+1)$$

Or:

$$F^{nG}(t+1) = F^n(t+1) - F^n(t) + \sum_{i=1}^{M} F^{nL}_i(t,t+1) \qquad \text{...(3.8)}$$

Using equations (3.8), (3.5) and (3.4) the total number of employment can be expressed in terms of the modeled elements $F^n(t+1)$, $P^n(i)$, $F^{nL}_i(t,t+1)$. Thus, The general form of the employment location model will be given according to the following equation:

$$ME_i(t+1) = \sum_n AVE^n F^n_i(t) - \sum_n AVE^n F^{nL}_i(t,t+1) + \sum_n AVE^n P^n(i)\left(F^n(t+1) - F^n(t) + \sum_{i=1}^{M} F^{nL}_i(t,t+1)\right)$$

...(3.9)

3.4.3 Exogenous Sectors of the Model

We have seen from the previous section that employment at time t + 1 is a function of the lost and gained firms (equation 3.3). That is, the lost and gained firms represent the net change in employment at zone i between time t and t + 1. In this context, it is crucial to look at the net change in the number of firms to determine which sectors should be modeled and which should not. In other words, determine what the exogenous sectors, if any, will be? The data extracted in section 3.3.1 can be used for this purpose. Table 3.2 summarizes the total number of firms in both time periods. It also provides the employment for the year 1990.

Table 3.2 Total Employment and Firms in the Hamilton-Wentworth Region

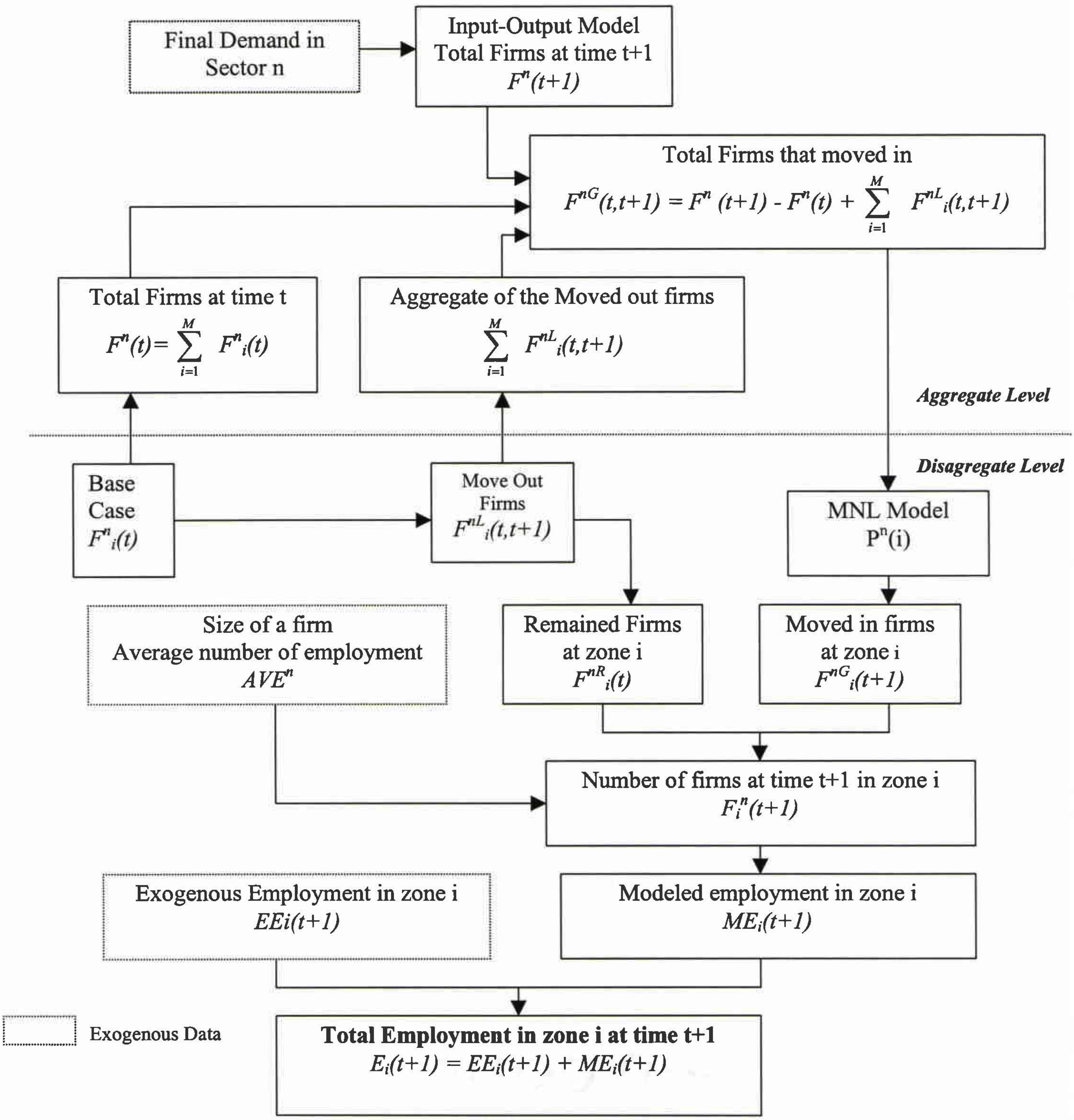
Industry	*Employment 1990*	*% Employment 1990*	*Firms 1990*	*Firms 1997*	*% Firms 1990*	*%Firms 1997*
Other Industries	980	0.50	82	122	0.71	1.01
Manufacturing	51255	26.40	905	928	7.81	7.69
Construction	6781	3.49	633	509	5.46	4.22
Transportation	4054	2.09	157	152	1.35	1.26
Communication	4886	2.52	119	145	1.03	1.20
Wholesale Trade	6830	3.52	445	462	3.84	3.83
Retail Trade	29826	15.36	3607	3426	31.13	28.40
All Services*	89377	46.04	*5639*	*6321*	*48.67*	*52.39*

***** All services sector includes: FIRE, Businesses, Government, Education, Health and Social, Accommodation, Food & Beverages and other services.***

Table 3.2 shows that the services of all types and retail trade sector exhibit the highest proportion of establishments in both years. These account for 31.13% and 48.67% for both sectors in the year 1990. For the year 1997, the total number of establishments in retailing tends to decline by 2.73% where as services grows by 3.72%. Manufacturing ranks third in terms of number of establishments. It can be seen that services, retail trade and manufacturing plays the major role in forming an economic market and job opportunity in the region. From the 1990 database percentages of total employment in these three sectors are 46.03%, 15.36% and 26.40%, respectively. Thus, the three sectors account for 87.79% of the total employment in the year 1990 in the region. With these figures it becomes mandatory to model these sectors since they count for most of the employment and establishments. Construction and transportation represent less than 5% of firms and employment in the region, thus we will assume that these two sectors are given exogenous to the model. In the services sector, the firms in the government services and education services will also be given exogenously to the model. These firms are not profit maximizing firms and thus do not follow the trend shown by other types of firms in which profit plays the major role in the location decision. The percentages of firms in government and education services are 1.95% and 3.6%, respectively.

The general structure of the employment model is given in Figure 3.10. As can be seen, the modeling process involves predicting firms at both the aggregate level and dissaggregate level. At the aggregate level, the input-output model predicts the total number of firms in the region for each sector of the economy. Exogenous final demand is used as input for the input-output model to predict firms at the regional level. At the dissaggregate level, the move out firms model predicts the lost firms at the zonal level. These are aggregated for each sector of the economy and combined with the results predicted by the input-output model. The result of this combination is the totals of new and relocated firms in the region. Totals of new and relocated firms are used as input for the destination choice model. The destination choice model is a model that predicts the choice probability at the dissaggregate level. Once the choice probability is calculated at the dissaggregate level, it is translated into the actual number of new and relocated firms by multiplying the choice probability by the totals of new and relocated firms given at the aggregate level. Results from the lost firm model are combined with the total number of firms at the base year to yield the remained firms at the dissaggregate level. These are combined with the new and relocated firms predicted by the destination choice model to yield the total number of firms at the census tract level. Finally, firms are translated into employment using the exogenous size of firms in each industry. Total employment at census tract *i* is the sum of the modeled and exogenous employment.

Figure 3.10 General Structure of the Employment Location Model



3.5 Conclusion

The formal structure of an employment location model is introduced in this chapter. Three models are proposed and linked together to form the components of the model. The three models are the input-output model, the destination choice model and the lost firms model. The spatial distribution of employment at the census tract level is based on the spatial distribution of firms at the same geographical scale. Thus, the model predicts firms and infers employment based on the redistribution of firms.

An empirical analysis was done to identify the sectors to be modeled. Results have shown that manufacturing, retail trade and services industries forms more than 87% of the total employment and firms in the region. Other sectors like transportation and communication have less presence in the region. The percentage of firms in these two sectors is 1.35% and 1.03%, respectively. This suggests that the redistribution of firms in transportation and communication will not have a significant influence on the spatial distribution of employment in the region and so will be given exogenously to the model. Moreover, firms in the government services and educational services will be taken exogenously in the model. Firms in these two types of services are not profit seeking firms and thus do not behave in the same way other services firms do.

CHAPTER FOUR

ESTIMATION OF THE EMPLOYMENT MODEL

4.1 Introduction

In this chapter of the thesis we explore the methods used to estimate the parameters of the different models discussed in the previous chapter. The input-output model, the destination choice model and the lost firm model form the elements of the employment location model. The input-output model predicts the total number of firms in each industry at time $t + 1$. The destination choice model distributes the new and relocating firms to the different census tracts of the CMA. The lost firms model predicts those firms that are lost in each of the census tracts between time period t and $t + 1$. The models are then linked together according to equation (3.9) and the total employment of each census tract is predicted.

The remainder of this chapter is organized as follows. First, we evaluate the input-output model for the Hamilton-Wentworth region using the provincial input-output transaction tables. Next, the elements of the destination choice model are introduced. The model is based on the assumption that new and relocating firms search the metropolitan area for a location that maximizes their profit. The method used here is similar in structure to the one proposed by Erickson and Wasylenko (1980), Shukla and Waddell (1991), and Yarish (1998). Models are developed for firms in the different economic sectors. Characteristics of the census tracts and firms are used as the explanatory variables in the logit formula. Estimated coefficients associated with each of the explanatory variables are shown. Finally, the lost firms model is introduced and estimated. Different types of regression models such as simple regression and Simultaneous Autoregressive models are employed. As in the destination choice model, separate models for the different modeled sectors are developed and estimated. Results are provided and discussed.

4.2 The Input - Output Model

4.2.1 The Basic Model - Firms at the Regional Level

The input-output model is used at the highest level of hierarchy in the employment model. It attempts to predict the total number of firms at the level of Hamilton-Wentworth. According to Miller and Blair (1985), such a model captures the linkages between the economic sectors. Thus, input-output modeling is a proper tool for achieving our goal.

The starting point of the model is a set of input-output accounts, which summarizes all the flows of goods and services that occur within an economy. Each firm in the economy is assigned to one of a mutually exclusive and collectively exhaustive set of *n* production sectors (Anderson et. al., 1994). The basic logic of these accounts is that the output of each sector is made up of two aggregate components: intermediate demand, which is the sale of goods and services from one production sector to another, and the final demand, which is the sales of goods and services for a final purpose such as consumption, investment, export, or sales to the public sector. Correspondingly, the expenditure of each sector can be divided into two components: intermediate expenditures, which is the purchase of goods and services from other sectors, and the value added, which includes payment to labor, capital(debt services and profits), and the public sector (taxes). Table 4.1 demonstrates the general form of a typical input-output transactions table.

Table 4.1 An input-output account table

		Purchasing Sector					***Final Demand(Y)***				***Total Output***
		1	***2***	***...***	***i***	***n***	***C***	***I***	***G***	***E***	***(X)***
	1	z_{11}	z_{12}		z_{1i} ...	z_{1n}	C_1	I_1	G_1	E_1	X_1
	2	z_{21}	z_{22}		z_{2i} ...	z_{2n}	C_2	I_2	G_2	E_2	X_2
Selling	.										
Sector	***i***	z_{i1}	z_{i2}		z_{ii} ...	z_{in}	C_i	I_i	G_i	E_i	X_i
	.										
	n	z_{n1}	z_{n1}		z_{ni} ...	z_{nn}	C_n	I_n	G_n	E_n	X_n
Value	***l***	w_1	w_2		w_i ...	w_n	w_C	w_i	w_G	w_E	W
Added	***k***	k_1	k_2		k_i ...	k_n	k_C	k_i	k_G	k_E	K
	t	t_1	t_2		t_i ...	t_n	t_C	t_i	t_G	t_E	T
		X_1	X_2		X_i	X_n	C	I	G	M	X

The table shows that for all sectors, the value of output and expenditures balance so that the sum of all elements in a row (output) is equivalent to the sum of all elements in the corresponding column (expenditure). In the input-output analysis the sum of rows is important and is mathematically summarized as follows:

$$x_i = \sum_{j=1}^{n} z_{ij} + y_i, \qquad ,i = 1,..., n \qquad ...(4.1)$$

$$y_i = c_i + i_i + g_i + e_i, \qquad ...(4.2)$$

Where:

- x_i - the total output of industry *i*
- z_{ij} - the amount of output of industry *i* sold to industry *j* (intermediate demand)
- y_i - the final demand for the output of industry *i*
- c_i - the consumer purchases
- i_i - the investment
- g_i - the government purchases
- e_i - the net export, (the difference between the export of the outputs of sector *i* and the imports of like goods).

The input-output model is known to be a demand driven model; that is, the total output of sector *i* will change if the final demand changes. Hence, the model assumes the final demand values to be given exogenously to it. One further assumption is built into this model; technology remains constant. This can lead us to the definition of what is so called the *"technical coefficient"*, which is defined as

$$a_{ij} = \frac{z_{ij}}{x_j} \qquad ...(4.3)$$

The technical coefficient represents the value in dollars of sector *i* output used as input in producing one dollar's worth of sector *j* output. Based on the assumptions of exogenous final demand and fixed sectoral technologies, a model can be specified to project the total output of each sector as a function of its own final demand and the final demand of all other sectors. Combining equations (4.3) and (4.1) yields:

$$x_i = \sum_{j=1}^{n} a_{ij} x_j + y_j \qquad i = 1,2,..., n$$

This equation can be written in matrix form as:

$$\boldsymbol{X = A\,X + Y}$$

Or

$$\boldsymbol{X = (I - A)^{-1}\,Y} \qquad \ldots(4.4)$$

Where:

- $\boldsymbol{X}$ - an $(n \times 1)$ column vector representing the total output in each sector of the economy.
- $\boldsymbol{A}$ - an $(n \times n)$ matrix having in its elements the technical coefficients of sectors.
- $\boldsymbol{(I - A)^{-1}}$ - the Leontief Inverse. (*I* is called the identity matrix)
- **Y** - an $(n \times 1)$ column vector having in it the final demand in each sector of the economy.

In many cases, the dollar value of sectoral total output may not ultimately be the most important measure of the economic impact following a change in exogenous demand. Total output requirements could be translated into coefficients of firm per dollar of sectoral output. Assuming that these coefficients are denoted by (f_1, f_2, ..., f_n) then a diagonal matrix of those coefficients can be constructed to have these coefficients on the diagonal of the matrix as follows:

$$\hat{\mathbf{F}} = \begin{bmatrix} f_1 & 0 & \dots & 0 \\ 0 & f_2 & \dots & 0 \\ . & . & \dots & . \\ 0 & 0 & \dots & f_n \end{bmatrix}$$

The total number of firms in each sector can be obtained by multiplying matrix ***F*** by the total output we have in equation (4.4) as follows:

$$f = F\,X$$

or

$$f = F[(I - A)^{-1}\ Y] \qquad \dots(4.5)$$

The result is an (*n x 1*) matrix with the total number of firms in each sector of the economy. In the employment model, these sectors are the sectors that form the economy of the Hamilton-Wentworth region. In conclusion, projections of the total number of firms in each sector at the regional level can be achieved if we use the technical coefficients, the firm coefficients, and the final demand at the regional level.

The input-output model proposed above suffer some shortcomings that prevent it from being an effective tool in long-run projections. The assumption of fixed technology ignores economies of scales in production and thus does not meet the scope of long-term projections since technology changes over time. However, Miller and Blair (1985) suggest techniques to over come the technology problem persisting in the input-output model. Among these is the "*Marginal Input Coefficient Technique*", Miller and Blair (1985, page 274) and the "*RAS Technique*", Miller and Blair (1985, page 276). Toyomane (1989) used discrete choice theory to model the change in the technical coefficients that enables the input-output model to be used in long-run simulations.

The problem of the technical coefficient and change in technology in our model will be handled exogenously through scenarios where the coefficient values should be altered when required.

4.2.2 Evaluating the Input-Output Model

One of the obstacles confronting the design of an input-output model at the regional level is the limitation of data. The model not only requires measures of output and firms in individual sector but also measures of the degree of interaction between n^2 sectoral pairs. Such data can only be obtained via costly surveys of individual businesses. Fortunately, science tackles these obstacle by developing methods which enables the researcher to estimate the local or regional data from National data. Statistics Canada provides input-output transaction tables at both the national level and the provincial level. For the purpose of our model, the transaction table is needed at the regional level. Thus, we employ the data provided for Ontario (provincial data) to derive a transaction table for Hamilton-Wentworth.

4.2.2.1 The Commodity-By-Industry Input-Output Accounts

Input-output transactions for the year 1990 for the province of Ontario were obtained from *Statistics Canada*. However, the form of data came in a shape that is not the same as the one we introduced in the previous section, which represents the transaction table in terms of a square matrix, known as the square system because the flow of money is given on the basis of flows between industries. Therefore, the system is a square matrix of industries by industries in dimension. This type of system, represented by a square matrix, does not account for secondary production (*commodities*) which is important and affects the economic system (Miller and Blair, 1985). Most modern surveys overcome this problem by constructing what is called the rectangular system. In this system rather than having one square matrix, we have two matrices: the ***make*** matrix and the ***use*** matrix. The rows of the ***make*** matrix (also called the production output matrix) are the commodities produced by industries in the economy, whereas the columns describe the industry sources of commodity production. The diagonal elements of the ***make*** matrix are the primary products (which define the industries in the first place) while the off-diagonal elements are the secondary products. However, the matrix does not provide a complete picture of the inter-industry activity in an economy since inputs to an industry production process includes not only commodities but also value-added as described in section 4.2.1. In order to complete this picture of sources and disposition of industry output, another matrix is constructed, this is known as the ***use*** matrix. It records the commodity inputs to an industrial process to form a complete transaction system with the ***make*** matrix. The new system of transaction accounts will take the following form:

Table 4.2 Summary of commodity and industry accounts

	Commodities	***Industries***	***Final Demand***	***Total Output***
Commodities		*Use Matrix* ***U*** ***(m x n)***	***E*** ***(m x 1)***	***Q*** ***(m x 1)***
Industries	*Make Matrix* ***V*** ***(n x m)***			***X*** ***(n x 1)***
Value Added		***W*** ***(1 x n)***		
Total Inputs	***Q'***	***X'***		

*Note: The term "**m**" stands for the number of commodities in an economy, "**n**" stands for the number of industries*

In the table, V is the ***make*** matrix, an element v_{ij} represents the amount of commodity j produced by industry i, U is the ***use*** matrix, an element u_{ij} represents the amount of commodity i bought by industry j, E is the vector of commodities deliveries to the final demand, W is a vector of industry value-added inputs and X is a vector of industry outputs as defined in equation (4.4). This new system of industries and commodities can be manipulated through linear algebra to yield the same system presented in equation (4.4).

In table 4.2, we can normalize the ***make*** matrix V by its column sum resulting the fraction of total production of commodity j in the economy produced by industry i:

$$d_{ij} = \frac{v_{ij}}{Q_j}$$

where d_{ij} is referred to as the *commodity output proportion*. In matrix notation the equation is as follows:

$$\boldsymbol{D} = \boldsymbol{V}\,\boldsymbol{Q}^{-1}$$

Or

$$\boldsymbol{V} = \boldsymbol{DQ} \qquad \ldots(4.6)$$

D is of dimension (*n x m*) corresponding to *n* industries by *m* commodities. In equation (4.6) we assume that industries are fixed, so we are assuming that the total output of a commodity is provided by industries in fixed proportions; this is often referred to as an industry-based-technology assumption. Recalling the definition of the technical coefficient in the base model (equation (4.3)) in terms of industry-by-industry, we can define a similar coefficient in terms of commodity-by-industry as follows:

$$b_{ij} = \frac{u_{ij}}{X_j}$$

Here, b_{ij} is the dollar's worth of commodity *i* required to produce one dollar's worth of industry *j*'s output. In matrix terms the above equation is represented as:

$$\boldsymbol{B} = \boldsymbol{U}\,\boldsymbol{X}^{-1} \qquad \ldots(4.7)$$

In the first row of Table 4.2 we can see that the total output of a commodity *i* is the sum of that commodity consumed by industries in the economy plus any sales of that commodity to final customers, that is:

$$Q_i = u_{i1} + u_{i2} + \ldots + u_{in} + E_i$$

In matrix notation, this is represented by:

$$Q = U + E$$

Using the value of **U** from equation (4.7) we get:

$$Q = BX + E \qquad \ldots(4.8)$$

In the table the total output X associated with each industry can be expressed in terms of the ***make*** matrix since $X = V_i$. Substituting for the value of V in equation (4.6) by X yields:

$$X = DQ \qquad \ldots(4.9)$$

Substituting for the value of Q from equation (4.8) into equation (4.9) yields:

$$X = D[BX + E] \qquad \ldots(4.10)$$

Manipulating equation (4.10) yields:

$$X - DBX = DE$$

or

$$(I - DB)X = DE$$

or

$$X = (I - DB)^{-1} (DE) \qquad \ldots(4.11)$$

The final demand of commodities, E, can be redefined in terms of industry output as opposed to commodity output. An element of the matrix D that was defined earlier gives the proportion of commodity j produced by industry i so that

$$Y_i = d_{ij} E_j$$

And in matrix terms

$$Y = DE \qquad \ldots(4.12)$$

The term ***DB*** is a matrix of (***n x n***) in dimension, which represents the technical coefficient matrix of an n^2 interaction industrial pairs. If we let the term ***DB*** = ***A*** and substitute for the value of ***Y*** from equation (4.12) into equation (4.11) the result is $X = (I - A)^{-1} Y$ which is the square model we introduced in equation (4.4).

Given both the ***use*** matrix and the ***make*** matrix for the Province of Ontario for the year 1990, we can apply the technique described above to form the square system from the rectangular system. The matrices ***D, B and E*** are prepared and matrix algebra is applied to get the values of ***A = DB*** and ***Y = DE***. The results are shown in tables 4.3a and 4.3b.

Table 4.3a 1990 technical coefficients (***A = DB***) for Ontario, Canada

	Sector1	*Sector2*	*Sector3*	*Sector4*	*Sector5*	*Sector6*	*Sector7*	*Sector8*
Sector1	0.40476	0.29736	0.12147	0.03433	0.04925	0.03265	0.03139	0.28195
Sector2	0.00395	0.00100	0.02586	0.01930	0.00221	0.00403	0.01862	0.00505
Sector3	0.00615	0.00314	0.12080	0.00990	0.01467	0.00794	0.00234	0.20828
Sector4	0.01965	0.00337	0.02673	0.05008	0.03170	0.04425	0.03016	0.03676
Sector5	0.03125	0.04646	0.02287	0.00608	0.01220	0.00482	0.00639	0.04377
Sector6	0.00578	0.01685	0.00837	0.00534	0.00289	0.00454	0.00525	0.02680
Sector7	0.04521	0.09087	0.10348	0.07574	0.10966	0.14045	0.14781	0.11992
Sector8	0.13909	0.03484	0.04501	0.07393	0.07321	0.06896	0.04822	0.07083

Table Key:
Sector 1 = Manufacturing
Sector 2 = Construction
Sector 3 = Transportation
Sector 4 = Communication
Sector 5 = Wholesale Trade
Sector 6 = Retail Trade
Sector 7 = All Services
Sector 8 = Other sectors

Table 4.3b 1990 final demand (*Y= DE*) for Ontario, Canada

Industrial Sector	1990 Final Demand
Manufacturing	58165038.03
Construction	33105460.15
Transportation	3863158.28
Communication	7292943.40
Wholesale Trade	14485872.57
Retail Trade	21131396.68
All Services	74854921.15
Other sectors	3606635.35

4.2.2.2 Location Quotients and Regional Estimates

Tables 4.3a and 4.3b gives the technical coefficient matrix and the final demand corresponding to each sector of the economy at the provincial level. However, the same matrices are required at the regional level in order to apply the system introduced in equation (4.4) for the study area. A method called the "*location quotients technique*" is used to get figures at the regional level. The values are based on local sectoral employment data. These will be used to calculate location quotients for sector *i* in region *r* as:

$$LQ_i^r = \frac{E_i^r / E^r}{E_i^P / E^P} \qquad \text{...(4.13)}$$

Where E_i^r and E_i^P are sector *i*'s employment at the local and provincial levels respectively and E^r and E^P are total employment at the regional and provincial levels respectively. In other words, the location quotient is the ratio of sectors *i*'s employment share at the local level to its employment share at the provincial level (Kanaroglou et. al., 1998). If the location quotient is greater than 1, sector *i* has a strong presence in region *r*, and as a result, it should be able to meet all local intermediate demand. The basic rule adopted is to set local technical coefficients to the provincial value in those cases where the location quotient is greater than 1, and to scale it down proportionately in those cases where the location quotient is less than 1. For this, the adjusted regional technical coefficients in terms of the adjusted location quotients become:

$$a_{ij}^{r} = \begin{cases} a_{ij}^{P}(LQ_i^r), & if \quad LQ_i^r < 1 \\ a_{ij}^{P}, & if \quad LQ_i^r \geq 1 \end{cases} \qquad \ldots(4.14)$$

Location quotients for all sectors at the regional level were calculated using the sectoral employment for both Hamilton-Wentworth and Ontario. Employment data for Hamilton were obtained from the 1990 employment survey. The data for the province of Ontario were obtained from the 1991 census data. The results in Table 4.4 show that manufacturing and retail trade are more localized at Hamilton-Wentworth with their employment location quotients being greater than 1.

Table 4.4 Employment location quotients for Hamilton-Wentworth

Sectors	E_i^P	E_i^r	LQ_i^r	a_{ij}^r	Concentration in the region
Manufacturing	942995	51255	1.54	a_{ij}^P	*More Localized in the region*
Construction	358890	6781	0.54	$LQ_i^r \times a_{ij}^P$	*Less Localized in the region*
Transportation	187830	4054	0.61	$LQ_i^r \times a_{ij}^P$	*Less Localized in the region*
Communication	188630	4886	0.74	$LQ_i^r \times a_{ij}^P$	*Less Localized in the region*
Wholesale trade	233910	6830	0.83	$LQ_i^r \times a_{ij}^P$	*Less Localized in the region*
Retail trade	700925	29826	1.21	a_{ij}^P	*More Localized in the region*
Services	2632485	89377	0.96	$LQ_i^r \times a_{ij}^P$	*Less Localized in the region*
Other Sectors	265570	980	0.10	$LQ_i^r \times a_{ij}^P$	*Less Localized in the region*

Other sectors show less localization at the region and so they can not meet all the local intermediate demand. In the later sectors, the provincial technical coefficients are scaled down to meet the regional requirements. Using the values of the regional location quotients and the technical coefficients from Table 4.3a we can calculate the technical coefficient matrix for Hamilton-Wentworth region. The results are provided in Table 4.5. Although the final demand in the input-output model should be given exogenously to the model, we took the task of computing the final demand corresponding to each sector for Hamilton-Wentworth region for the year 1990. This is done to form a compatible system that will match the figures at provincial level.

Table 4.5 Evaluated technical coefficient matrix for Hamilton-Wentworth

	Sector1	***Sector2***	***Sector3***	***Sector4***	***Sector5***	***Sector6***	***Sector7***	***Sector8***
Sector1	0.40476	0.29736	0.12147	0.03433	0.04925	0.03265	0.03139	0.28195
Sector2	0.00212	0.00053	0.01387	0.01035	0.00118	0.00216	0.00998	0.00271
Sector3	0.00376	0.00192	0.07401	0.00607	0.00899	0.00487	0.00143	0.12761
Sector4	0.01445	0.00248	0.01965	0.03682	0.02331	0.03253	0.02218	0.02703
Sector5	0.02590	0.03851	0.01896	0.00504	0.01011	0.00399	0.00529	0.03628
Sector6	0.00578	0.01685	0.00837	0.00534	0.00289	0.00454	0.00525	0.02680
Sector7	0.00435	0.00875	0.00997	0.00730	0.01057	0.01353	0.01424	0.01155
Sector8	0.01457	0.00365	0.00471	0.00774	0.00767	0.00722	0.00505	0.00742

Note: The sectors listed in this table are the same as those in table 4.3a

In order to calculate the regional final demand we first calculate the total output at the provincial level using equation 4.4. The total output for sector i at the regional level X^r_i is evaluated by assuming that the ratio of regional employment to provincial employment in sector i equals the ratio of the regional gross output to the provincial gross output for that specific sector as follows:

$$\frac{E_i^r}{E_i^P} = \frac{X_i^r}{X_i^P}$$

And so:

$$X_i^r = \frac{E_i^r X_i^P}{E_i^P}$$

Where all terms are previously defined. The above relation is based on two important assumptions. The first suggests that the total output of a certain sector i depends on the labor supply given in that sector. This relation is true at both the regional level as well as the provincial. The second assumption is that the supply of labor for production at the regional level is of the same magnitude as the supply at the provincial level. The above assumptions also suggest the same level of productivity in the different regions of the Province of Ontario. Computed values of both the total output and final demand for Hamilton-Wentworth are shown in Table 4.6.

Table 4.6 Total sectoral output and final demand for Hamilton-Wentworth

Sector	Gross Output	Final Demand
Manufacturing	8086712.32	4312870.39
Construction	699442.64	629268.61
Transportation	347176.76	250966.68
Communication	479580.45	194477.62
Wholesale trade	707336.48	424379.00
Retail trade	1051108.10	955653.25
Services	3939548.41	3811703.83
Other Sectors	145982.99	0.00

Finally, firm coefficients are calculated based on the total output given in Table 4.6 and the number of establishments we extracted from the 1990 regional employment database. The results are shown in Table 4.7.

Table 4.7 Firm coefficients for the different sectors in the year 1990

Sector1	***Sector2***	***Sector3***	***Sector4***	***Sector5***	***Sector6***	***Sector7***	***Sector8***
0.00011	0.00072	0.00043	0.00030	0.00065	0.00325	0.00156	0.00083

It is worth noting that ***sector7*** in all the tables above represents all service sectors. This is the case since the tables provided by Statistics Canada are at the L-Level of aggregation (The Input-Output Structure of the Canadian Economy, 1989) where all types of services are grouped as one sector. However, not all of these sectors are modeled as indicated earlier. Therefore, once the total firms in each sector is predicted according to equation (4.5), number of firms in ***sector7*** are adjusted to fit the modeled sectors. This is done by subtracting the number of firms in both the "*government service*" sector and "*educational*" sector at time $t + 1$ from the figure we obtain from equation (4.5) and that corresponds to ***sector7***. This yields the total number of firms for the modeled services sectors at time $t + 1$.

4.3 The Destination Choice Model

In this section we discuss the method used to construct the destination choice model. This part of the employment model is used to project the gained firms at zone i at time $t + 1$. Gained firms include new firms established between t and $t + 1$ and chose to locate at census tract i, as well as those that moved into census tract i from elsewhere in the region. Moreover, the choice of location is considered to be behavioral, where firms tend to locate at the site that will maximize their profit. In this context, discrete choice theory can be employed to develop the model and carry out the task of predicting the choice. Next, we introduce the concept of the discrete choice theory that gives the multinomial logit model its general form. The discussion then focuses on destination choice models. The model we aim to develop is similar in structure to the theory proposed by Yarish (1998), Shukla and Waddell (1991) and Erickson and Wasylenko (1980). The definition and hypothesis of the explanatory variables in the multinomial logit model are presented. Finally, the parameters associated with the variables are estimated and results are presented and discussed.

4.3.1 Discrete Choice Theory

Ben-Akiva and Lerman (1985) identify the basic problem confronted by discrete choice analysis as the modeling of choice from a set of mutually exclusive and collectively exhaustive alternatives. McFadden (1975) uses the principle of utility maximization to provide a complete model of individual choice. He defines the utility of individual n facing choice i as a function of two types of variables: attributes of the choice alternative and characteristics of choice makers.

$$U_{ni} = U(x_{ni}, s_n)$$

Where x_{ni} are the attributes and s_{ni} are the characteristics. In the principle of utility maximization the decision-maker[1] is modeled as selecting the alternative with the highest utility among those available at the time the choice is made. An operational model consists of parameterized utility functions in terms of observed independent variables and unknown parameters. The values of these parameters are estimated from a sample of observed choices made by decision-makers when confronted with the choice situation. Since it is impossible to specify and estimate a discrete choice model that will always succeed in predicting the chosen alternatives by all individuals the concept of random utility is used.

The true utility of the alternatives are considered random variables, so the probability that an alternative is chosen is defined as the probability that it has the greatest utility among the available alternatives. The choice probability is derived by assuming a joint probability distribution for the set of random utilities. Ben-Akiva and Lerman (1985) identifies four different sources of randomness that effect utility function. The first is the presence of unobserved attributes. Here, the analyst does not have complete information about the attributes affecting the decision. The second source of randomness is the presence of unobserved taste variation among decision-makers. The third source of randomness is named by measurement errors of attributes and imperfect information. The fourth source of randomness is the instrumental errors. In this case the analyst does not have a complete information about the variables related to the actual attributes.

---

[1] The decision-maker might be an individual person, a household or a firm.

Based on the previous four sources of randomness, the random utility of an alternative can be represented as the sum of two components:

1. A systematic (or deterministic) component $V(x_{ni}, s_{ni})$
2. A random component (or disturbance) that captures additive effects of unobserved characteristics of individuals and alternatives $\varepsilon(x_{ni}, s_{ni})$ (Daganzo, 1979)

Mathematically, the utility can be written as follows:

$$U_{ni} = V(x_{ni}, s_{ni}) + \varepsilon(x_{ni}, s_{ni}) \qquad \ldots(4.15)$$

If we assume that individual n is going to choose alternative *i* from a set of discrete alternatives J_n, then the probability that alternative *i* is chosen is the probability that the utility of alternative *i* exceeds the utility of all other variable alternatives. Thus,

$$P_{ni} = Pr(U_{ni} > U_{nj} \text{ for all } i \neq j)$$

The multinomial logit model is derived under certain assumptions about the joint distributions and the disturbance. If we assume that the disturbances ε are independently and identically distributed and follow a Gumbel distribution, the probability can be worked out to get the following formula:

$$P_n(i) = \frac{e^{Vni}}{\sum_{j=1}^{J_n} e^{Vjn}} \qquad \ldots(4.16)$$

This equation is the general formula of the multinomial logit model. More complicated forms of the probability can be obtained if other assumptions on the joint distributions and the disturbances are assumed. One well known assumption is that the disturbances ε

are normally distributed and dependent resulting in correlations across the errors (Yarish, 1998). This is known as the probit model. Unfortunately, the equation of the normal distribution is complex and makes estimation difficult. Equation (4.16) is easy to handle and estimation of the parameters that form the utility function can be easily achieved via maximum likelihood estimation.

The basic assumption of the multinomial logit model is the "*Independence from Irrelevant Alternatives*" IIA property. The principle states that the ratio of probabilities of choosing one alternative over another is unaffected by the presence or absence of any additional alternative in the choice set. The IIA property assumes that the random elements in the utility function are independent across alternatives (Ben-Akiva and Lerman, 1985). This independence from irrelevant alternatives provides a computationally easy model that allows the elimination or addition of alternatives in the choice set without re-estimation (Yarish, 1998). However, the logit model will overpredict the choice probability if two alternatives are identical or very similar in terms of their characteristics. In such a case the disturbance of both alternatives will tend to be correlated with each other which might lead to biased estimates of the modeled parameters.

4.3.2 Intrametropolitan Location of New and Relocating Firms

In this section we present the theory proposed by Yarish (1998), Shukla and Waddell (1991) and Erickson and Wasylenko (1980) to model the intrametropolitan location of firms. New and relocating firms are those firms a census tract gains between time t and t + 1. Relocating firms can be thought of as totally new firms since both new and relocating firms tend to locate at the site that will maximize their profit. Thus, firms are modeled based on the assumption that they search the metropolitan area to locate at the site that will ensure the maximum profit. Shukla and Waddell (1991) believe that a firm location decision is motivated as an outcome of the following maximization problem over possible locations *i*

Maximize $$[P\,F(X_i, L_i, \boldsymbol{Z}_i) - P_x X_i - R_i L_i]$$

Where:

- X_i - the purchased non-land
- L_i - the land inputs
- $\boldsymbol{Z}_i$ -a vector of non-purchased inputs having to do primarily with various type of access opportunity as spatial agglomeration attributes
- P - the free on board (f.o.b) output price
- P_x -the price of variable inputs such as labour and capital price
- R_i -the land price

The prices P and P_x are assumed invariant with location within the metropolitan area, while the land price R_i varies spatially. The firm chooses the optimal quantities of X_i, L_i, and Z_i so that an indirect profit function in factor prices and non-price attributes of location *i* can be specified (Shukla and Waddell, 1991):

$$U_i = U_i(P, P_x, R_i, Z_i)$$

If firms of any type face the same non-land prices within the urban area and that the input employment is governed by technology considerations, then the prices P and P_x can be considered as constants and the above function can be summarized as the industry specific formulation of the maximum profit attainable by a firm in industry type n at location i

$$U_{in} = V_{in}(R_i, Z_i) + \varepsilon_{in} \qquad \ldots(4.17)$$

A firm in sector n of the industry locates in j provided that profits are highest there, that is $U_j = \max U_i$ for all i. This profit maximizing selection can be cast as a random utility process subject to stochastic error. Assuming that the error term ε_{in} follows a Gumbel distribution then equation (4.17) becomes identical to equation (4.15). This gives the formulation of the logit model-equation (4.16). In this case n represents any firm in industry n and i represents a census tract to be chosen.

4.3.3 Explanatory Variables and Hypothesis

For this study the choice set for the logit model is the set of 116 census tracts that form the Hamilton-Wentworth region. n represents the following modeled sectors: manufacturing, construction, wholesale trade, retail trade, and certain services of the economy. Given that we have five sectors and 116 census tracts to be modeled, our task is to specify and estimate five MNL models, one for each of the modeled sectors where the number of choices facing the individual (firm) is 116. Next we define the explanatory variables of the logit model. These are the major elements in the logit formula, which give it the power to predict. It is crucial to specify a good selection of variables that will enable the model to do a good job in forecasting the choice decision. Recalling equation (4.17) one can see that the logit formulation depends on land inputs. In our case the land inputs is reflected by the characteristics of each census tract. Moreover, characteristics of firms are used in some of the modeled sectors, as we will see next.

4.3.3.1 Dependent variable specification

In order to model the sectors, we define five multinomial logit models, one for each industry. The dependent variable represents the proportion of gained firms in each census tract between time t and t + 1. The number of gained firms in each tract from the total number of new and relocating firms in the region is used to represent the choice. This approach is used to model the manufacturing, construction and wholesale sectors. In this case we are using group data to represent the dependent variable. However, this is not the case for retail trade and the services sectors. In the later, individual data representing the gained firms at each census tract is used to designate the dependent variable. The locational pattern of firms is used to identify the types of firms and thus enables us to use individual data. Kernel maps for firms in the subsectors of each of the modeled sectors are produced. The maps are presented in Appendix B with an analysis of the locational pattern. The results show that different types of firms within manufacturing, construction and wholesale trade sectors show the same general locational trend, thus, no distinction between the different types of firms within the one sector can be made based on the locational pattern. Firms within retail trade and service sectors show a diverse locational pattern, thus, categorical variables are assigned to the different types of firms to identify the firms' type. The identification of the types of firms allows for using individual data to represent the dependent variable in the logit model.

4.3.3.2 Independent Variable Specification

As mentioned earlier, we are modeling the choice of firms at the census tract level. The linear-in-parameter form is specified for all the systematic utilities. The independent variables used are shown in Table 4.8. Characteristics of the census tracts and firms are used to represent the utility function. In the model we account for several factors that we think would effect the location choice of the firms. These are structural variables, control variables, agglomeration variables and categorical variables.

4.3.3.2.1 Structural Variables

Structural variables are used to reflect the structure of the urban area in terms of the general form. Distance to the CBD is considered to be an important factor that affects the location of firms. The CBD or the core is traditionally known as the commercial heart of the urban area. If firms tend to locate in close proximity to the CBD, we then observe a centralization trend. In contrast, a decentralization trend takes place where firms try to escape the core of the city and locate in the fringes. In this context, it becomes important to test for this type of variable to see if firms of certain industries try to escape the core and enjoy the amenities of locating in the suburban areas.

Table 4.8 Explanatory variables used in the logit models

I - Structural Variables	
CBDPRO	Linear distance from the census tract centroid to the center of the zone containing the Central Business District (CBD) Area
LNCBDPRO	Natural logarithm (log) of *CBDPRO*
HWYPRO	Dummy variable = 1 if a highway or an express way pass by or intersect the zone, 0 otherwise
MALLPRO	Dummy variable = 1 if a mall drops in a census tract or lie in proximity to a census tract, 0 otherwise
II- Control Variables	
HHLDDENS	Household density in a census tract (HHLD/Sq. Km)
POP96	1996 population size at the census tract
MLAND	Dummy variable = 1 if a census tract land use is manufacturing, 0 otherwise.
III- Agglomeration Variables	
MCTW_EMPA	Accessibility to employment in manufacturing, construction, transportation and wholesale trade industries
RS_EMPA	Accessibility to employment in retail trade and services
S_EMPA	Accessibility to employment in services
IV- Categorical Variables	
RCAT1	Categorical variable = 1 if a firm is in SIC60, SIC64 or SIC65, 0 otherwise
RCAT2	Categorical variable = 1 if a firm is in SIC62 or SIC63, 0 otherwise
RCAT3	Categorical variable = 1 if a firm is in SIC61 or SIC62, 0 otherwise
SCAT1	Categorical variable = 1 if a firm belongs to FIRE or Business, 0 otherwise
SCAT2	Categorical variable = 1 if a firm belongs to Health & Social, Accommodation, food & beverages and Other services, 0 otherwise

The linear distance from the centroid of each census tract to the centroid of the CBD zone is calculated using TransCAD 3.1 GIS system. The census tract coverage is modified to include the CBD zone to it using ArcView GIS. The modified coverage is then imported into the TransCAD GIS to calculate the aerial distance in Kilometers from the CBD. Distance from the CBD is included in all of the five models. For some sectors-Retail trade and Services-the natural logarithm of the distance is used rather than the distance itself, as it displays better statistical properties in terms of significance. The hypothesized sign of the parameter associated with this variable is negative indicating high affinity to locate at the core.

Distance from the CBD is not the only structural factor that drives the decision of location in the urban area. Firms might find it more suitable to locate in close proximity to shopping malls if this guarantees a higher profit. This type of clustering around malls is an observable trend in the Hamilton-Wentworth context. Figure B.22 and B.23 show that certain retail firms in ***SIC61*** and ***SIC62*** show the tendency to cluster in and around malls. Therefore proximity to malls in the region for these types of firms does influence the locational decision. For these types of firms proximity to malls ***MALLPRO*** is included as an independent variable. This is represented as a dummy variable that holds the value 1 if a tract contains a mall in it or neighbors a tract that contains a mall, otherwise it is 0. The hypothesized sign of the ***MALLPRO*** is positive indicating the preference of locating in proximity to malls.

Highway and Expressway accessibility plays a major role in the location decision since it affects decentralization. Simmons et al. (1998) argues that the decentralization of the commercial market is accompanied by decentralization of commercial activities along major arterials. This suggests that proximity to highways and expressways is a factor that might attract firms to locate on land accessible by car. Highway and Expressway proximity is also portrayed as a dummy variable that holds the value 1 if a highway or an expressway passes through or intersects a census tract, and 0 otherwise. This was prepared through the Spatial Analyst of the ArcView GIS system. For our purpose, highway proximity is used in all the logit formulas. The hypothesized sign for this variable is positive suggesting high likelihood of locating near highways and expressways.

4.3.3.2.2 Control Variables

Control variables are variable that regulates the spatial distribution of firms in the region. These include population size and household density by census tract. The first is used to test the effect of population size on the location decision since certain types of firms are expected to follow population and others tend to avoid expensive inhabited land. Retail trade and service sectors are expected to follow population since they are population-oriented. Kernel maps for certain retailing and service firms (see Appendix B) show this type of trend when compared with the population map in Figure 3.4. Population size is also used to capture the aspect of urbanization in the region. The hypothesized sign of this variable is positive showing the tendency to follow population.

Household density is used as a surrogate to capture the effect of newly developed land and its influence on location decision. It is also used to capture the effect of decentralization in the region. Figure 3.3 shows the household density in terms of low, medium and high density. The core of the city exhibits the highest household density. Other parts of the city of Hamilton such as the lower part as well as the northern part of the mountain escarpment have the medium density. The outer strips of the CMA exhibits the lowest household density. It becomes clear that household density is a good proxy for newly developed land when looking at the spatial trend it shows in the region. Other things being equal in the logit model, the amount of newly developed land is used to test the locational decision choice in the peripheries versus locating at the core. The core of the city exhibits the highest amount of population reflected by the high household density in this part of the CMA. With development taking place, population is expected to occupy land in the periphery of the CMA.

However, this would have lower household density in the short run than that of the core given that the land area of the census tracts in the periphery are larger as shown in Figure 3.2. Thus for those new and relocating firms that seek to follow development and exit the congested core, it is expected that locating at the periphery would be a favorable choice. This choice is portrayed in two ways. First, by locating at periphery the firm will form a new accessible market for the clients at these areas. In particular, this is expected for firms in retailing and services. Secondly, the firm will benefit from paying less rent for land in these areas. This is expected for all types of firms. In both cases, the firms are able to earn a higher profit. Thus, the likelihood of locating at these areas is expected. The hypothesized sign for the ***HHLDDENS*** variable is negative indicating the attractiveness to land with lower household density, which will reflect the decentralization of employment in the region.

Land use by type is another factor that might influence the location decision of firms. It intends to test for the effect of specialization at certain locations of the region.

However, land use data were not available for commercial establishments. The only type of land use data available is the land uses in manufacturing industries given in Figure 3.6. Manufacturing land use is identified since this type of land use tends to specialize at certain locations of the CMA. Thus, for the manufacturing MNL model land use is taken into consideration to see what is the effect of land use specialization on the decision of location. The hypothesized signs for this variable is positive indicating the favor of locating on specialized land.

4.3.3.2.3 Agglomeration Variables

Agglomeration variables are used to capture the effect of agglomeration economies on the location decision. Selting et al. (1995) argue that agglomeration economies are instrumental in explaining firm location. According to Yarish (1998), agglomeration economies are locational advantages gained by firms when choosing to locate and cluster in large metropolitan cities over smaller towns. There are several ways to represent the effect of agglomeration economies in a model. One simple way to create an agglomeration index is to sum the proportion of firms in different industries with respect to the total number of industries. This was proposed by Wheaton and Shishido (1981) where they used this type of agglomeration measure to test the assumption that heavy concentration of industries is equivalent to industries being tightly interwoven; that is, they exhibit strong forward and backward linkages. So locating together helps capture the external benefits of agglomeration.

Selting et al. (1995) suggest that clusters of firms can be obtained by using correlation coefficients or factor analysis. However, they argue that correlation expresses geographical tendency rather than functional association since consistent spatial associations do not imply that there are economic linkages between sectors. To overcome this type of deficiency input-output analysis can be used. The flow of money from one sector to the other in an input-output table can give a more precise idea on the type of linkages that persist among the different sectors. Those sectors with higher flows tend to interact more than those with low flows. On this basis, firms in sector *i* that interact mostly with firms in sector *j* might find it useful to locate in proximity to those firms to minimize transportation cost.

Gravity models as a spatial interaction measurements was suggested by both Selting et al. (1995) and Shukla and Waddell (1991) to measure agglomeration of economies the zonal level. This measurement can be represented as an accessibility measure to certain zone to capture agglomeration forces influenced by other firms in the surrounding neighborhood. We took the task of measuring agglomeration economies in terms of what has been presented above to define suitable agglomeration variables to use in the different logit models. Correlation coefficients of the number of firms locating at the census tracts for each pair of the sectors in the economy are calculated. However due to the fact that some tracts are small in size the weighted average number of firms in a tract was used instead of the actual number of firms in that tract. This average is calculated by counting the number of firms in a tract and the surrounding tracts and then dividing by the total number of counted tracts. The reason why this kind of measurement is attained is because we want to capture the effect of the nearest neighbor that may persist for small census tracts. The results of this calculation indicate that there are agglomeration economies in the region. The figures in Table 4.9 show that manufacturing industries interact highly with construction, transportation, communication and wholesale trade. However, wholesale trade firms interact the most with manufacturing having a correlation value of 0.94. Also firms in construction correlate with firms in transportation and wholesale trade. All of the previous firms show a weak affiliation with retail trade and services firms. Retail trade firms show high affiliation with service firms and vise-versa.

Table 4.9 Correlation coefficient of the "*weighted average number*" of firms in census tracts

	Sector1	***Sector2***	***Sector3***	***Sector4***	***Sector5***	***Sector6***	***Sector7***
Sector1	*1.00*	**0.86**	**0.71**	**0.63**	**0.94**	0.51	0.39
Sector2		*1.00*	**0.68**	0.46	**0.89**	0.58	0.42
Sector3			*1.00*	0.64	**0.81**	0.27	0.31
Sector4				*1.00*	**0.73**	0.40	0.18
Sector5					*1.00*	0.47	0.33
Sector6						*1.00*	**0.74**
Sector7							*1.00*

Sector1= Manufacturing, Sector2 = Construction, Sector3 = Transportation, Sector4 = Communication, Sector5= Wholesale Trade, Sector6 = Retail Trade, Sector7 = Services

The result of the correlation coefficient can also be achieved by looking and comparing the kernel maps we produce for the different sectors in Chapter 3. The results of the input-output table indicate some of the correlations between sectors but with some discrepancies. The actual figures in the input-output table are calculated from the technical coefficient matrix and total output we evaluated in section 4.2. The results are summarized in table 4.10. In the input-output framework, firms in sector *i* will tend to locate in proximity to firms in sector *j* if this helps minimize the cost of transportation. This kind of assumption suggests that the purchasing sector will tend to locate in proximity to the selling sector to enjoy the benefit of agglomeration and minimize the cost of transportation.

Table 4.10: 1990 transaction table for the Hamilton-Wentworth region

	Sector1	***Sector2***	***Sector3***	***Sector4***	***Sector5***	***Sector6***	***Sector7***
Sector1	3273226	207992	42173	16468	34836	34320	123666
Sector2	17159	376	4817	4965	840	2275	39347
Sector3	30481	1347	25696	2912	6359	5119	5665
Sector4	116861	1737	6824	17662	16488	34200	87384
Sector5	209490	26936	6584	2417	7157	4204	20873
Sector6	46757	11788	2907	2565	2047	4775	20702
Sector7	35238	6126	3463	3501	7476	14229	56124

Sector1= Manufacturing, Sector2 = Construction, Sector3 = Transportation, Sector4 = Communication, Sector5= Wholesale Trade, Sector6 = Retail Trade, Sector7 = Services

Table 4.10 indicates that the flow of money from the manufacturing sector is directed towards the construction sector and the wholesale trade sector. Also manufacturing tends to interlink with its own sector, this result is expected since sub-sectors within the one sector might agglomerate within the same area for the same reasons discussed above. Thus, firms in manufacturing will tend to agglomerate in close proximity to other manufacturing firms, and firms in construction and wholesale will tend to cluster around manufacturing firms. As for the construction sector, it pours money into the manufacturing sector, transportation and communication sectors. This suggests that firms of these types will tend to cluster in close proximity to construction firms. Wholesale trade interlinks with manufacturing. The flow of money from this sector into the manufacturing sector is the highest. Additionally, wholesale money flow into the construction direction. Again, this implies that firms in manufacturing and wholesale sectors locate in close proximity to construction firms. Retail trade firms show the tendency to direct money into the services sector, the manufacturing and construction sectors. However, this is not compatible with the results from Table 4.9 or the kernel maps (Figures B.17 - B.27 in Appendix B) for these sectors. Table 4.9 and the kernel maps in Figures B.17 to B.27 show a strong relation between retail trade and services firms. Nevertheless, the relation with respect to manufacturing and construction is not as strong as suggested in the input-output table. Services sector flow of money is directed towards the services sector, retailing and manufacturing industries. As in the case of retailing, this shows to have a poor relation with manufacturing when looking at the results in Table 4.9 or the kernel maps. These results suggest that the services sector is highly interacted with both retail trade and itself.

The previous discussion indicates that agglomeration economies do exist for Hamilton-Wentworth. In this context, it is important to include agglomeration economies measurement in the logit formula to account for the powers of agglomeration that influence the choice decision. Agglomeration in the model is represented as a spatial interaction measure index. Accessibility is used to represent agglomeration. The important question that arises here is: "What kind of measurement is efficient in representing the effect of agglomeration induced by certain types of firms at the census tract level?" A direct and straightforward answer would be accessibility to firms. In reality, this is true. Imagine a census tract in a geographical area and assume that firms are attracted to that specific tract due to the effect of agglomeration. In this case, it would be important to take into consideration the effect of locating in close proximity to the tract. Firms might not locate exactly in that tract to enjoy the benefit of agglomeration since census boundaries are nothing but artificial walls imposed by the available data and they do not account for a firm looking for a site that would maximize its profit. A firm may find it useful to locate in close proximity to a firm at a distance of 2 kilometers or so; that is, in a neighboring tract. However, to have a meaningful measurement we will use accessibility to employment rather than accessibility to firms. This kind of measurement is proposed by Ihlanfeldt and Raper (1990), Shukla and Waddle (1991) and Yarish (1998) to capture the effect of agglomeration economies.

Accessibility accounts for neighborhood effect since it is a distance decay function. The general form of accessibility is as follows:

$$A_i = \sum_{j=1}^{N} \frac{E_j}{D_{ij}^{\beta}} \qquad \ldots(4.18)$$

Where:

- A_i - accessibility to census tract i
- E_j - Employment in the census tract j.

- D_{ij} - is the linear distance between the centroid of census tract *i* to centroid of census tract *j*
- β - is a parameter associated with the speed of decay of the agglomeration force, usually this equals to 2.

Three different types of independent variables are constructed to capture the effect of agglomeration:

1. ***MCTW_EMPA***: Accessibility to employment in manufacturing, construction, transportation and wholesale trade industries
2. ***RS_EMPA***: Accessibility to employment in retail trade and services
3. ***S_EMPA***: Accessibility to employment in services

The hypothesized sign of the parameters associated with the previous three variables is positive indicating the interlinkages and agglomeration powers that might exist between the different sectors of the economy.

4.3.3.2.4 Categorical Variables

As pointed in Section 4.3.3.1 individual data is used to represent the dependent variable in the logit model for both retail trade and services. Since the different types of firms in the one sector show a deviant trend of clustering, new variables are constructed to categorize firms within the one sector. This type of variables is called ***categorical variables***. These variables are based on the trend shown in the kernel maps of the subsectors within retailing and services. For instance, Figures (B.17, B.21 and B.22) show dispersed locational trend in the region that matches the population distribution from Figure (3.4). Therefore, a categorical variable ***RCAT1*** that holds the value 1 is assigned if the firm is in any of these three sectors, and 0, otherwise. Figures (B.19 and B.20) show low intensity in the downtown core area whereas other types of firms in retailing have high intensity at the downtown area. As a result the variable ***RCAT2*** is created. This variable equals 1 if the firm is in SIC62 or SIC63 and is 0, otherwise. Moreover, Figures (B.18 and B.19) show the tendency to locate in and around malls. This locational tendency is evident in the high clustering around the malls as can be seen in Figures (B.18 and B.19). For this the variable ***RCAT3*** is defined. ***RCAT3*** hold the value 1 if the firm belongs to SIC61 or SIC62 and is 0, otherwise. The first variable ***RCAT1*** is used to interact with the population size variable; ***RCAT2*** is used to interact with the log of the distance from the CBD and ***RCAT3*** is used to interact with the proximity to the mall variable. A final word about categorical variables is that unlike the last three groups of variables, which are ***generic*** and available to all choices, the categorical variables are not. They are specified for certain tracts (choices) which make them ***alternative specific variables***.

Categorical variables for the subsectors comprising the services sector are defined based on the clustering trend observed in the kernel maps. In this context, two variables are constructed. The first is ***SCAT1*** that holds the value 1, if the firm is in SIC86, SIC9192 or SIC9699 and a 0 if not. This variable is used to interact with the population size variable. Figures (B.25, B.26 and B.27) indicate that firms of those types tend to be dispersed in the region and follow a pattern that is similar to the spatial distribution of population in the region. The second variable is ***SCAT2*** that takes the value 1 if the firm is in SIC7076 or SIC77 and is 0, otherwise. This variable is used to interact with the log of the distance to the CBD. Although all types of services tend to have high clustering pattern in the core of the city, FIRE and businesses have the highest presence at the core. The hypothesized sign of the categorical variables will depend on the type of variables they will interact with.

4.3.3.3 Results

The estimates of the MNL models are presented in Tables 4.11 for the five-modeled sectors. Parameters associated with the explanatory variables that form the utility function of the logit model are estimated using a program developed by Ferguson (1995). Unlike other models available in the market, the model is a powerful tool that enables the user to estimate a multinomial logit model with a large number of choices. In our case, the numbers of choices available to the decision-makers (firms) is 116. Numerous runs of the model were performed to get the best combination of explanatory variables and to obtain the best fit for the models. In the estimation process, variables other than those defined in Table 4.8 that were noted in the literature review were included to test for their effect on the model and its fit. These variables include: labor force, represented by employment by place of residence at the census tract level; average income of household in each census tract; and average monthly dwelling rent as a surrogate for land prices. Nevertheless, these additional variables showed to be insignificant in all runs and do not add any improvement to the statistical result of the models, thus, they were omitted from the model specification. Given the fact the multinomial logit model comprises 116 choices, each model can have up to 115 constants. Again, not all of these constants came out to be significant and as a result they were omitted. The estimated parameters for the five models with the t-statistics values are reported below.

In the manufacturing logit model, distance to the CBD exhibits a significant and positive sign implying a decentralized locational trend from the core. The highway and expressway proximity is also significant with a positive sign indicating the preference of locating in close proximity to highways and expressways. This high affiliation with the highway proximity is evident with the t-statistic being the most significant of all the variables. This indicates that accessibility is important for manufacturing firms. The household density retains a significant and negative sign indicating the tendency of locating on land with low household density. This can be interpreted as the preference of locating on large vacant land farther from the CBD. This locational behavior for manufacturing firms is expected since these firms seek to locate on large and vacant land if they are to expand in the future. Manufacturing firms do not seem to follow population, this is evident with the negative sign of *POP96* variable. This suggests the trend of specializing in areas with less population size.

Table 4.11 Logit Estimated Parameters for the Different Industrial Sectors

Variable	*Manufacturing Industries*		*Construction Industries*		*Wholesale Trade Industries*		*Retail Trade Industries*		*Services Industries*	
	Beta	***T-stat***	***Beta***	***T-stat***	***Beta***	***T-stat***	***Beta***	***T-Stat***	***Beta***	***T-stat***
CBDPRO	0.061	4.870	-0.057	-3.326	0.039	2.606				
LN(CBDPRO)							-0.417	-4.128	-0.008	-1.768
HWYPRO	1.284	9.225	1.194	7.157	1.022	6.038	0.981	9.602	1.033	15.316
HHLDDENS	(-0.796)	-7.957	(-1.595)	-13.145	(-2.228)	-15.435	(-1.741)	-16.680	(-0.078)	-3.004
POP96	(-0.093)	-5.265			(0.090)	5.143	(0.099)	11.584	(0.136)	10.017
MLAND	1.126	9.993								
MCTW_EMPA	(0.138)	4.608	(0.237)	7.515	(0.408)	12.127				
RS_EMPA							(0.269)	8.168		
S_EMPA									(0.057)	15.646
RCAT1*POP96							(0.094)	4.588		
RCAT2*LN(CBDPRO)							0.055	5.455		
RCAT3*MALLPRO							0.902	6.502		
SCAT1*LN(CBDPRO)									-0.910	-3.638
SCAT2*POP96									(0.054)	4.046
L(0)		-2856.908		-1578.192		-1616.221		-7939.496		-12958.287
L(β)		-2489.871		-1422.361		-1371.929		-7278.273		-11655.769
ρ^2		0.128		0.099		0.151		0.083		0.101
$\bar{\rho}^2$		0.123		0.092		0.144		0.076		0.097

Note: Values in ( ) are in thousands

Specialization at certain locations is evident for manufacturing industries. The high significance and positive sign of *MLAND* variable suggests this. Specialization is also deduced from the few numbers of significant constants in the logit model. Out the 115 constants only 9 came out to be significant reflecting the market share of new and relocating firms at the certain tracts of the CMA. Localization economies are evident for manufacturing firms. The sign of *MCTW_EMPA* is positive and significant, this suggests that manufacturing firms appreciate the forces of agglomeration economies.

Construction firms show some affinity to locate in close proximity to the core of the city. Similar to the manufacturing firms, construction firms show a tendency to locate near highways and expressways. Locating in the fringes is evident with the negative sign of the variable *HHLDDENS.* Nevertheless, construction firms seem to benefit from the effect of agglomeration. This is detected in the positive sign of the *MCTW_EMPA* variable. Population size appears to be insignificant in the final run, so it was omitted from the model specification. As in the case of manufacturing firms, only seven of the 115 constants appear to be significant. This is due to the fact that firms in this sector try to cluster and specialized in a limited number of census tracts.

Wholesale trade firms results reveal the fact that the core is not a desirable place to locate at. The MNL model shows the best fit of data among all the modeled sectors with rho-square value of 0.151. Wholesale trade firms escape the congested core and tend to locate in the suburbs to enjoy the available amount of vacant, inexpensive land. Moreover, firms in wholesale trade sector benefit the effect of localization imposed by the agglomeration of economies. The positive sign of *MCTW_EMPA* variable illustrates localization effect. Moreover, the positive sign of the *POP96* variable reflects urbanization, as firms in this industry tend to locate in populated areas. However, the population variable is marginally significant when comparing its significance to the retail and services logit models. Again, firms in this sector of the economy show the tendency to cluster in specialized areas in the region. This is observed in the few numbers of significant constants (seven) in the last run.

Retail trade firms display a natural propensity to locate in close proximity to the core of the CMA. This is evident in the significant and negative sign of the *LN(CBDPRO)* variable. Accessibility is important to retail trade firms. The *HWYPRO* variable is very significant and holds the expected positive sign. The *POP96* variable is very significant indicating a strong relation between population size and location decision for firms in retailing. This implies that retail trade firms are population oriented in Hamilton-Wentworth. The resultant parameter associated with *(RCAT1*POP96)* appears to be significant with the expected sign. This indicates that firms in (1) food, beverage and drug industries (SIC60); (2) shoe, apparel, fabric and yarn industries (SIC61); (3) general retail merchandising industries; and (4) other retail store industries are more population oriented in the region.

Interaction term *(RCAT2*LN(CBDPRO))* indicates that not all firms in retailing have the same locational behavior with respect to locating at the core. This is evident since the associated parameter with the previous interaction variable is positive, indicating that other things being equal in the model, firms in (1) household furniture, appliances and furnishing industries (SIC62) and (2) automotive vehicles, parts and accessories industries (SIC63) are attracted to the core while the rest of retail firms escape the core and locate elsewhere in the region. Moreover, the variable *(RCAT3*MALLPRO)* is significant and retains the required positive sign. This significance suggests that firms in (SIC61) and (SIC63) cluster in malls and in close proximity to them.

The *HHLDDENS* variable holds a negative sign indicating the preference to locate on land with low household density. This is expected since locating at the periphery will tend to form a new market for clients (population) located there. Also, by choosing the suburban land, the firm will enjoy the benefit of locating on low price land, which allow for increasing the profit. Retail firms enjoy the benefit of localization and agglomeration of economies. They show high tendency of clustering in areas, which are occupied by both retail, and services firms. This is evident in the significant value of *RS_EMPA* variable used to capture the effect of agglomeration. Due to the fact that retail firms are more dispersed in the region, the market share of firms in the CMA represented by the constants associated with the choices are apparent in the model. Out of the 115 indulged constants in the model, 48 turned out to be significant.

The negative sign of the *CBDPRO* in the services sector model postulate the affinity of services firms to locate at the core of the CMA. However, firms in FIRE and businesses services industries are more centralized compared to other types of services firms. This is evident in the negative sign associated with the *(SCAT1*LN(CBDPRO))* variable. Accessibility stands as an important factor influencing location decision of firms in services industries. This is seen in the high significance of the *HWYPRO* variable. Population size *POP96* turned out to be very significant. This indicates as in the case of retail trade firms that firms in services industries have a high affiliation towards populated areas. The interaction term *(SCAT2*POP96)* indicates that firms in (1) heath and insurance service industries, (2) accommodation, food and beverages service industries (SIC91-92) and (3) other industries (SIC96-99) are more population oriented than FIRE and business firms. The positive sign of the *POP96* variable emphasize the forces of urbanization economies in the region.

The negative sign of the *HHLDDENS* variable indicates that firms in the service sector tend to locate on less expensive land away from the CBD. However, the low value of the *HHLDDENS* variable when compared with other sectors indicate that the trend of locating on less expensive land away from the core is minimum for firms in the service sector. This suggests that service sector firms are the most centralized types of firms in the region. Agglomeration economies powers are important and influence the decision of location for service firms as suggested by the positive sign associated with the significant *S_EMPA* variable.

4.4 Lost firms model

Lost are considered the firms that left the census tract either to relocate elsewhere (either in or outside the CMA) or because they go bankrupt. As in the case of the destination choice model, lost firms are modeled at the sectoral level. Thus five different models are constructed and estimated for the five sectors. It is worthwhile noting that our first intention was to apply a binomial logit or probit model where firms are faced with two choices: either remain or depart the census tract. The dependent variable in the model would be the proportion of lost firms in each census tract. Unfortunately, such a model did not produce any significant associations between the dependent and the set of independent variables. Further analysis revealed that the proportions exhibit low variation over space. We found that 31% of the manufacturing proportions hold the value 0, 22% of the proportions hold the value 1. In the construction sector, 24% of the proportions hold the value 0 and 32% hold the value 1. Wholesale trade figures show that 43% of the observations hold the value 0 and 25% hold the value 1. Retail trade figures show that 54% of the proportion are between 0.45 and 0.65. This is also noticed in the services sector where 62% of the observation have a proportion value between 0.3 and 0.5.

The variables used in the binary logit specification are CBD proximity, highway proximity, mall proximity, monthly average household rents as a surrogate for land price, household income, population and household densities. The CBD proximity is used to capture the effect of decentralization, which we expect to be an evident trend in the region. Highway proximity tests for accessibility since firms of all types appreciate better accessibility. Locations with low accessibility will drive the firm to relocate to a different location. Land prices and property taxes for tracts are important factors that influence the decision of remaining or leaving a census tract. Because no data were attainable on land prices, the monthly average household rent as a surrogate for land price and local taxes is employed.

Household average income is used as one of the explanatory variables since income is assumed to have an influence on sales, specially for the population oriented industries such as retail trade industries and services industries. Thus tracts with higher income are less likely to have firms moving out. Household density is used to capture the effect of urbanization in the region where firms of certain types tend to cluster and locate in close proximity to populated areas. Thus, locations with high population size and household density are less likely to lose firms.

The low variability in the proportion of lost firms over space led us to model the number of lost firms as a function of the total number of firms in a census tract within a linear relationship. We thus fitted regression models, taking into account whenever necessary spatial autocorrelation.

4.4.1 Simultaneous Autoregressive models

Non-spatial linear regression models are powerful tools that enable the researcher to predict the future of the geographical phenomena from an observed trend that took place in the past. This is the case when the relation between the resultant phenomena and the factors driving that phenomenon to occur is linear. Green (1993) represents this relation mathematically as follows:

$$Y = X\beta + \varepsilon \qquad \ldots(4.19)$$

Where

- Y - a column vector of dimension $(n \times 1)$ representing the dependent variable
- X - a matrix of dimension $(n \times p)$ representing the independent variables
- β - a column vector of dimension $(p \times 1)$
- ε - a column vector of dimension $(n \times 1)$ representing the disturbances or error terms.

The basic set of assumptions that comprise the classical (non-spatial) linear regression model are as follows (Green, 1993):

1. Functional form as shown in equation (4.19)
2. Zero mean of the disturbance: $E[\varepsilon] = 0$
3. Homoscedasticity: $Var[\varepsilon] = \sigma^2 \boldsymbol{I}$,
4. Non-autocorrelation: $Cov[\varepsilon\, \varepsilon^T] = 0$
5. Uncorrelatedness of regressor and disturbance: $Cov[X,\varepsilon] = 0$
6. Normality: the elements in vector ε follow the normal distribution, i.e. $\varepsilon_i \sim N[0, \sigma^2]$ for all i

If the disturbances are normally distributed, then assumption 4 implies that they are independent as well.

Despite the fact that the model in equation (4.19) might be capable of fitting the data under investigation, the results might be misleading since we are trying to model spatial zones and data. To account for this type of spatial nature, one has to test for the spatial autocorrelation that involves the correlation between the values of the same variable at different spatial locations where assumption 4 does not hold anymore. One of the most widely used measures that test for the spatial autocorrelation is called the

Moran's I test statistic. Spatial correlation in attribute values y_i for a spatial proximity matrix $\boldsymbol{W}$ is estimated as

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right)\left(\sum_{i \neq j}\sum w_{ij}\right)}$$

Where w_{ij} is an element in an adjusted first order neighbor weighted matrix $\boldsymbol{W}$. In our analysis, the weighted matrix is created using both ArcView 3.1 and Splus 4.5 GIS systems. The calculation of this matrix is done according to the following steps (Mathsoft, 1996):

1. A temporary matrix is created so that an element x_{ij} equals 1 if census tract j is adjacent to census tract i, and zero otherwise. The diagonal of the matrix is zero.
2. Then, an average centroid-to-centroid distance between the neighboring census tracts and census tract i is calculated. A census tract not among those defined as neighbors already will be admitted as a neighboring census tract if its spatial distance (centroid-to-centroid) from the census tract i is less than this average distance.

An element x_{ij} of the resultant matrix is 1 if the census tract j is a neighbor to census tract i under the above criteria, and is 0 otherwise. If the *Moran I* test statistic is significant then an autocorrelation problem exists among the data set and a different model other than that in equation (4.19) should be used to model the process. This leads us to the definition of the spatial regression model.

The spatial regression model is an extension from the non-spatial regression model presented above. Mathematically, the model is written as follows:

$$Y = X\beta + U \qquad \ldots(4.20)$$

Where:

- U - is a column vector of dimension $(n \times 1)$ representing the autocorrelated disturbances

Unlike the classical regression model which uses the least square method to estimate its parameters, the spatial regression model uses the method of generalized least squares which allows the model to relax the assumption of spatial interdependence inherited in the simple regression model. Using this modification the disturbance U becomes a zero-mean vector of errors with variance-covariance matrix C, so that $E[U]=0$ and $E[UU^T]=C$. The estimated parameters represented by the matrix β depend on the variance-covariance matrix C since matrix β becomes a function of the matrix C (Bailey and Gatrell, 1995). Formally, the estimator β is given according to the following equation:

$$\beta = (X^T C^{-1} X)^{-1} X^T C^{-1} y$$

However, this appears to cause a problem in estimating β's since C is unknown and not easy to model. Bailey and Gatrell (1995) overcome this problem by trying to specify C indirectly by an interaction scheme. This is done by including in the model, relationships between variables and their neighboring values, which indirectly specify particular form of C. They propose a variate interaction model of the form:

$$Y = X\beta + U$$

$$U = \rho WU + \varepsilon$$

Where:

- $X\beta$ - is the set of explanatory variables and associated parameters
- U - is the matrix of correlated disturbances
- ρ - is a parameter to be estimated
- ε - is a vector of independent random errors with constant variance σ^2.
- W - is the weighted matrix defined earlier.

The previous model can be written as

$$Y = X\beta + \rho WY - \rho WX\beta + \varepsilon$$

Or

$$Y = (I - \rho W)^{-1}[X\beta - \rho WX\beta + \varepsilon]$$

Using the previous specification, the variance-covariance matrix can be obtained in terms of σ^2, ρ and W. This will take the form $C = \sigma^2[(I - \rho W)^T (I - \rho W)]^{-1}$ and thus the parameters in matrix β can be estimated.

4.4.2 Estimation of the Parameters

As mentioned earlier, due to the fact that we are dealing with spatial data, we must test for spatial autocorrelation. Therefore, the *Moran I* test is held for the five sectors using the statistical package Splus. The results in Table 4.12 indicates that spatial autocorrelation exists among firms in the retail trade and services sectors. This implies that a Simultaneous Autoregressive (SAR) model should be applied to estimate these sectors. As for the other three sectors, non-spatial regression model is sufficient. The estimation of the (SAR) models for both the retail and services sectors were achieved using the Spatial module of the Splus statistical package. The estimation of the classical regression model is also achieved using the Splus statistical package.

Table 4.12 *Moran I* Test Statistic for Spatial Autocorrelation, (*y = lost firms)*

Sector	Normal Statistic	Normal p-value	Null Hypothesis*
Manufacturing	-0.605	0.5455	*Accepted*
Construction	0.536	0.5917	*Accepted*
Wholesale Trade	-0.593	0.5534	*Accepted*
Retail Trade	1.904	0.0569	*Rejected*
Services	8.662	0.0000	*Rejected*

Null Hypothesis: No Spatial Autocorrelation

In examining the performance of the SAR models we noticed that the estimated model for the retail trade sector performs well except for some outliers. Further analysis was conducted to check for these outliers. We found that the model underpredicts or overpredicts the number of losses corresponding to tracts with shopping malls within them. The model underpredicts the losses that correspond to the CBD zone namely census tracts 1511 and 1512 containing the Jackson Square shopping mall. Also it underpredicts the number of losses for census tract number 3542 that lies in the northeast side of the city of Hamilton. Moreover, we found that the model overpredicts the losses for census tract 3544 that contains the East Gate shopping mall; census tract 3534 that contains the Limeridge shopping mall; and census tract 1521, which contains the Centre Mall. Apparently, the model underpredicts the lost firms for the zones having the Jackson Square shopping mall because the number of closures in that mall is much higher than the general trend of closures in the region. Furthermore, the model overpredicts the number of closures for the census tracts containing the Limeridge mall, East Gate mall and the Centre mall due to the fact that the census tracts with malls have larger number of firms within them.

Loss of firms in a census tract is proportionally related to the total number of firms in that tract, and since these are relatively larger than the total number of firms in other census tracts, the model will overestimate them. To overcome this problem and develop a more precise model, we defined a number of dummy variables that can reduce the effect of the overprediction or underprediction of the model.

The service sector model seemed to underpredict some of the values for certain tracts and overpredict some others. The model underpredict the values for the following tracts: 1499, 1500, 1503, 1509, 1511, 1512, 1513, 1514, and 1527, all of which are located either in the downtown area or in close proximity to it. Also the model underpredicts the lost firms for the census tracts located in the northwest part of the city of Hamilton (census tracts number 3542 and 3545). It is believed that the model underpredicts the values for these tracts since they have more presence at the core but apparently show higher rate of losses when compared with other tracts. Again and as in the case of the retail trade sector, the model overpredicts the number of losses for the census tracts containing Limeridge mall, Centre mall, East Gate mall, Queens malls, and the University Plaza mall located in the town of Dundas. As in the retail sector, dummy variables were introduced in the model to account for these deficiencies. In this context, the explanatory variables used to form the different five models are given in Table 4.13.

Table 4.13 Explanatory Variables for the Lost Firm Model

1) Manufacturing Industries	
M_FIRMS90	Manufacturing firms in the year 1990
2) Construction Industries	
C_FIRMS90	Construction firms in the year 1990
3) Wholesale Trade Industries	
W_FIRMS90	Wholesale Trade firms in the year 1990
4) Retail Trade Industries	
R_FIRMS90	Retail Trade firms in the year 1990
R_DTN	Dummy variable =1 if the census tract is located in downtown, 0 otherwise
R_MALLS	Dummy variable = 1 if the census tract contains Limredge mall, East Gate mall or Center mall
R_NW	Dummy variable = 1 if the census tract is in the north west side of the city of Hamilton, 0 otherwise
5) Services Sectors	
S_FIRMS90	Services firms in the year 1990
S_DTN	Dummy variable = 1 if the census tract is located in the downtown area or in close proximity to it, 0 otherwise
S_MALLS	Dummy variable = 1 if the census tract contains Limredge mall, East Gate mall, Queens mall, Center mall and the mall in the town of Dundas, 0 otherwise
S_NW	Dummy Variable = 1 if the census tract is located in the northwest side of the city of Hamilton, 0 otherwise

The first three sectors are calibrated using the regression model built in the Splus statistical package. On the other hand, the (*SAR*) models for the retail and services sector where estimated using the SLM function of the spatial module of Splus. Results of the estimations are reported in Tables 4.14 and 4.15 for the simple regression model and the spatial models (*SAR*), respectively. Results for the lost firms model show a significant fit in the five models estimated with R-square value being greater than 0.95 for all of them. However, the constant parameter appears to be insignificant in all of the models. This insignificance indicates that when a tract contains 0 firms then the number of lost firms will be zero. The variable representing the number of the firms in the different industries is significant and holds a positive sign for the different models.

Table 4.14 Lost Firm Model - Simple Regression Results

Sector	*Variables*	*Estimated values*	*Standard Error*	*T-Ratio*
Manufacturing	***Constant***	-0.0196	0.1494	-0.1310
	M_FIRMS90	0.644	0.008395	76.674
	R^2	0.981		
Construction	***Constant***	-0.1223	0.1337	-0.915
	C_FIRMS90	0.7424	0.0127	58.574
	R^2	0.967		
Wholesale Trade	***Constant***	0.0096	0.0966	0.0990
	W_FIRMS90	0.7233	0.0101	71.098
	R^2	0.978		

(*R_DTN* R_FIRMS90*) variable is significant with a positive sign indicating that the number of loses associated with tracts in the core is higher than that in other places of the CMA. The variable (*R_NW * R_FIRMS90)* came out significant with a positive sign indicating that tracts located at the northwest of the city of Hamilton loses more firms than other census tracts. The negative sign of (*R_MALL* R_FIRMS90)* suggests that census tracts having regional malls in them does not loose as much as other census tracts in the region. Service firms show the same trend shown by retail firms where we see that services firms in the core of city are lost at a higher rate than that shown at other tracts, this is indicated by the positive sign given by *(S_DTN* S_FIRMS90)* variable. Also the northwest part of the city of Hamilton is loosing firms at a higher rate than other areas, this is proposed by the positive sign of *(S_NW* S_FIRMS90)*. Census tracts with malls in them does not seem to loose as much as other tracts do, this is reflected by the negative sign of *(S_MALL* S_FIRMS90)*.

Table 4.15 Lost Firm Model - Simultaneous Autoregressive (SAR) Model Results

Sector	*Variables*	*Estimated values*	*Standard Error*	*T-Ratio*	*Estimated Rho parameter*
Retail Trade	*Constant*	-0.2545	0.4482	-0.5679	
	R_FIRMS90	0.5284	0.0136	38.7548	
	*R_DTN*R_FIRMS90*	0.0687	0.0148	4.6534	
	*R_MALL*R_FIRMS90*	-0.1600	0.0155	-10.3189	
	*R_NW*R_FIRMS90*	0.1513	0.0386	3.9240	
					0.02133
	R^2	0.9854			
Services	*Constant*	0.4384	0.4037	1.0861	
	S_FIRMS90	0.3910	0.0079	49.2289	
	*S_DTN*S_FIRMS90*	0.1322	0.0094	14.0234	
	*S_MALL*S_FIRMS90*	-0.0396	0.0174	-2.2669	
	S_NW S_FIRMS90*	0.1862	0.0776	2.3988	
					-0.00884
	R^2	0.9843			

4.5 Conclusion

In this Chapter, the empirical work of this study is completed. The three models that form the elements of the employment model is evaluated and calibrated. The first model is the input-output model. Technical coefficients for this model are evaluated from those given for the province of Ontario. The location quotient technique is used to evaluate the regional technical coefficients. The location quotients are calculated using the employment data at both the regional and provincial level for the year 1990. The location quotient values indicate that manufacturing and retail trade industries are more localized in the region and so provincial figures are used at the regional level. Construction, wholesale trade and services industries have less presence and so the location quotient values were used to scale down the provincial figures to the regional figures.

The second model is a multinomial logit model. This model is used to distribute the new and relocating firms at the different census tracts that forms the region, thus, is considered as a destination choice model. Five different multinomial logit models are calibrated for the five modeled economic sectors. Characteristics of the census tracts are used to form the linear in parameter utility function. Variables such as distance from the CBD, highway proximity, mall proximity, population size, household density and agglomeration economies shown to have an influence on the locational decision of the different firms. However, due to the fact that firms in the one industry have a different locational pattern, we used categorical variables to reflect the different types of firms in the models. This is used for retail and services industries. Results shown that firms within the one industry have a diverse locational pattern in the region.

The third model used is a regression model, which predicts the number of losses in firms at the census tract level. As in the case of the destination choice model, five different regression models are calibrated for the five modeled sectors. Non-spatial (Simple) regression models are used for the manufacturing, construction and wholesale trade sectors. Spatial regression (SAR) models are used for the retail trade and services sector. The Spatial regression models are used for retail trade and services because the analysis shows that number of losses within retail trade firms and within services firms are correlated, thus, using a simple regression model will lead to bias in the estimation. The analysis shows that tracts in the downtown area and northwest part of Hamilton City loose more that other tract in the region. Moreover, the analysis indicates that census tracts containing regional malls in them do not loose as much as other tracts do.

CHAPTER FIVE

IMPLEMENTATION, VALIDATION AND SIMULATIONS

5.1 Introduction

This chapter discusses the implementation of the framework introduced in the previous two chapters into a computer simulation model. For validation purposes, the predicted employment distribution from the simulation model is contrasted against the 1997 observed employment distribution. Furthermore, to examine the sensitivity of the model to economic conditions and changes in the transportation infrastructure, several scenarios are simulated and the associated results are presented.

5.2 Model Implementation

To test the significance of the employment model developed in chapter three and estimated in chapter four, a computerized version of the model is developed through the design of programming scripts. The programming scripts are developed using the Gauss programming language. The reason behind using gauss is in two folds. On the first hand, Gauss is a very powerful programming tool used to develop programs that involves matrix algebra. In Gauss, the programmer is able of reading in large dimensional matrices, apply algebraic operations and retrieve output results easily. On the other hand, Gauss is chosen other than any programming language because the original model *IMULATE* was implemented using the Gauss programming language.

Scripts for each of the submodels forming the employment model are developed and linked together. The result is an automated model linking the input-output model, the destination choice mode and the lost firm model into one system. The model is written in a way that enables the user to design and simulate scenarios to project the status of employment in the Hamilton-Wentworth region at the census tract level. The model is calibrated for the years 1990 and 1997 allowing the user to project the employment future in the region for every seven years. The base case is to project the employment for the year 1997. The scripts for the input-output model, the destination choice model and the lost firm model are presented in Part 3 of Appendix C of this thesis. A user manual and a technical manual are also developed to help the user running the model. These are also presented in Appendix C.

5.3 Model Validation (Goodness-of-fit)

In this section, we test the performance of employment model by running the model for the base case. The result is the spatial distribution of employment at the census tract level for the year 1997. This is compared with the observed data for the same year. However, it is also important to test the validity of the results generated by each of the models designed to predict the redistribution of firms in a certain industry.

The employment model is run for the base case and the results are obtained. The correlation coefficient values between the predicted and observed number of firms in manufacturing industries, construction industries, wholesale trade industries, retail trade industries and services industries are calculated. Results are shown in Table 5.1.

Table 5.1 Correlation coefficient values between predicted and observed number of firms for the different modeled sectors (1997)

Manufacturing	Construction	Wholesale trade	Retail Trade	Services
0.82	0.83	0.81	0.91	0.99

Results in Table 5.1 indicate that the different models perform well in predicting the total firms at the census tract level for the year 1997. The models for retail trade and services industries predict better than the models for manufacturing, construction and wholesale trade industries. This is seen in the high values of the correlation coefficients for retail and services industries, namely 0.91 and 0.99 for retail and services, respectively.

The modeled employment at the census tract level is calculated by transforming firms in each industry into employment using the exogenous firm size of each industry and aggregating over all the modeled sectors. This is achieved for both the observed and predicted values. The correlation coefficient between the observed and predicted modeled employment is 0.91. The observed and predicted employment density at the census tract level is shown in Figures 5.1 and 5.2.

The Figures indicate that the model does a significant job in predicting the distribution of employment in the region. The performance of the model leads us to deduce that it is a good tool that can be used for forecasting the future of employment in the region.

Figure 5.1: Observed Employment Density - 1997

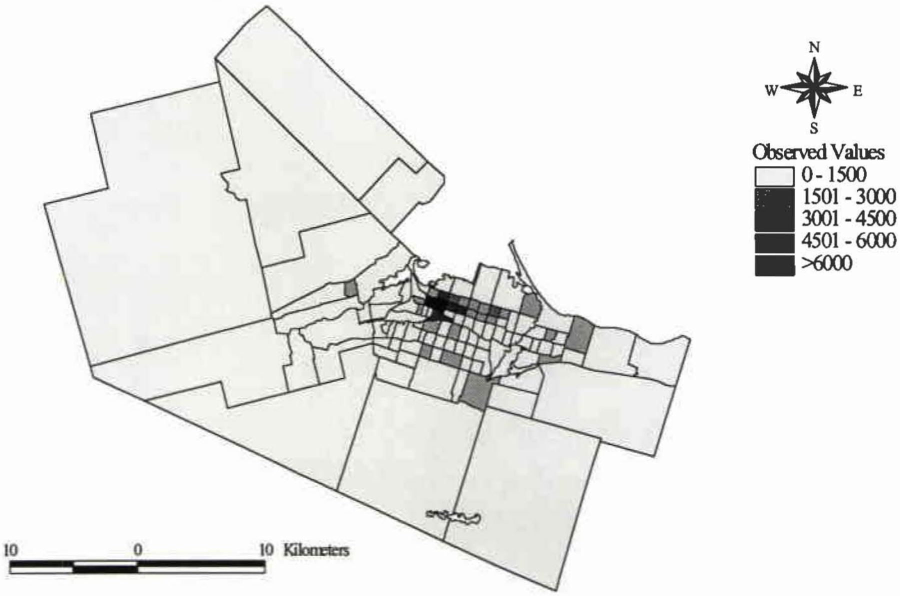


Figure 5.2: Predicted Employment Density - 1997

5.4 Scenario Simulation

In this section we are going to test the employment model and its potential of simulating scenarios. The purpose will be to show that the employment model is capable of simulating the employment mobility and generating plausible results under certain circumstances. The reader should keep in mind that changes in land use requires feed back loops between residential and commercial changes since residential land use change affect commercial land use and vise-versa is also true. Thus, the scenarios in this chapter will test how sensitive the employment model is if certain changes in the urban structure and form occur.

In scenario simulation we are testing the base scenario against an alternative one. The base scenario represents the status quo when projected in future. The alternative scenario reflects the situation that can take place if inputs to the system based on a certain criteria is to occur. For the purpose of illustrating the potential of the model we define a number of alternative scenarios and compare them to the base scenario we already run in the previous section.

The employment model we developed is considered a demand driven model; that is, a change in the final demand in the input-output model will affect the distribution of employment in the region. Moreover, the model is affected by the change in population, household density, highway and expressway proximity, mall proximity, and the total number of firms in the base year for each simulation. Therefore, the alternative scenarios can be defined so that all, some or one of the factors mentioned earlier is changed. The model is designed in a way that allows the user to simulate two types of scenario. The first type of scenarios is time independent, where we assume that changes occur in the urban structure with no time framework involved in that. The other type of scenario, time is involved and the model project for the consecutive seven years in future. Here, the first type of scenarios is constructed and simulated to test the impact of the urban and economic structure on the spatial distribution of firms in the region. Three groups of scenarios are defined and simulated. The first tests the effect induced on the spatial distribution of employment in the region by a change in the final demand for some sectors of the economy. The second tests the effect of constructing a new expressway on the spatial distribution of employment. The third tests the effect of sprawl on the redistribution of firms in the region.

5.4.1 Final Demand Scenario Simulation

The scenario aims on identifying the effect of demand change on the spatial distribution of jobs in the region. The input-output model predicts the total output of all firms at the regional level using the exogenous final demand given for each industrial sector. Thus, the number of firms in each industrial sector is a function of its own demand and the demand of other sectors. In other words, a change in the demand for one or more sectors will result in the change of the number of firms in all the industrial sectors. To attach the same meaning to this scenario, we assume that the effect of globalization will be a shift in manufacturing to newly industrialized countries. The implication for Hamilton-Wentworth will be a shift in demand from manufacturing to the services sector. In this context, we assume that the final demand for the manufacturing sector is decreased by 50% and that of services sector increases by the same percentages. This kind of change is expected to influence not only the distribution of firms in the region, but also the total employment. The results show that employment in the region will increase by 6.13%.

Figure 5.3 Change in Employment Level due to Change in Demand

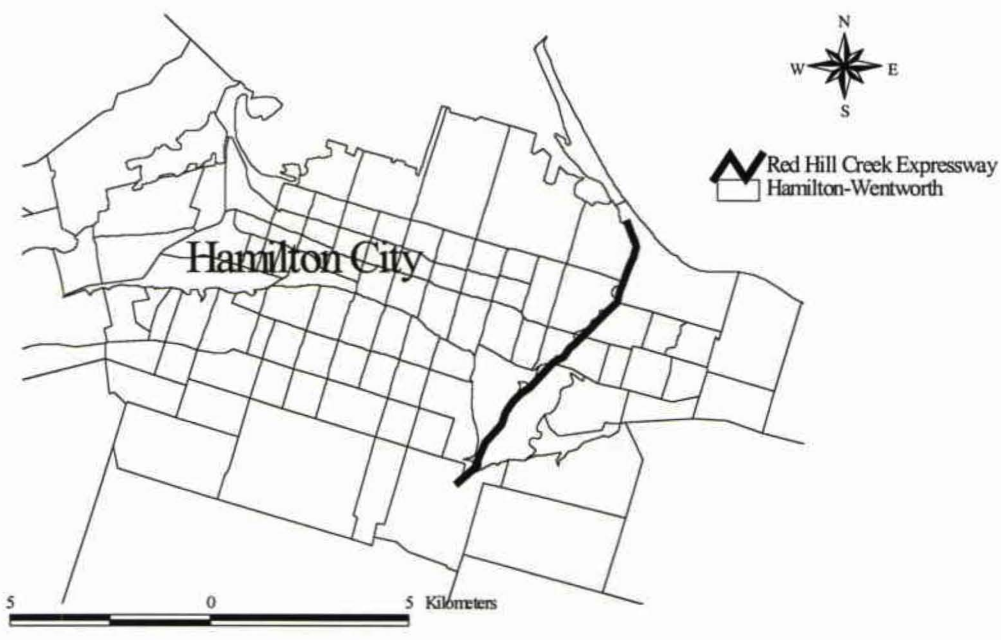
Apparently, the spatial distribution of employment changes as the final demand change. This is evident in Figures 5.3. Employment level appears to increase in central areas, tracts in proximity to malls and highways. This is expected since services are predominately centralized in the region and show high affinity towards locating near highways. Also, the fact that the linkages between retail trade firms and services firms are strong cause an increase in the employment level at areas having regional malls.

Moreover, Figure 5.3 indicates that a loss in jobs occur in areas where most manufacturing firms are concentrated.

5.4.2 Highway and Expressway Scenario Simulation

As shown in the previous chapter, employment location is sensitive to highway and expressway proximity. Thus, we developed this scenario to examine the impact of constructing a new expressway on the redistribution of employment in the region. The widely discussed Red Hill Creek (RHC) expressway is added to the transportation network of the model for this purpose. This is represented in the model as a categorical variable that holds the value 1 if an expressway passes through or intersects a census tract and 0, otherwise. The location of the proposed expressway is shown in Figure 5.4.

Figure 5.4 Red Hill Creek Expressway



Results from the scenario indicate that other things being equal in the model, construction of the expressway will attract firms to the tracts adjacent to it. Given that total employment remains the same as in the base scenario, construction of the expressway will result in a shift of employment toward census tracts closer to it. Figure 5.5 shows the increase in number of jobs in the due to the construction of the expressway.

Figure 5.5 Increase in Jobs - Base Scenario versus Alternative Scenario

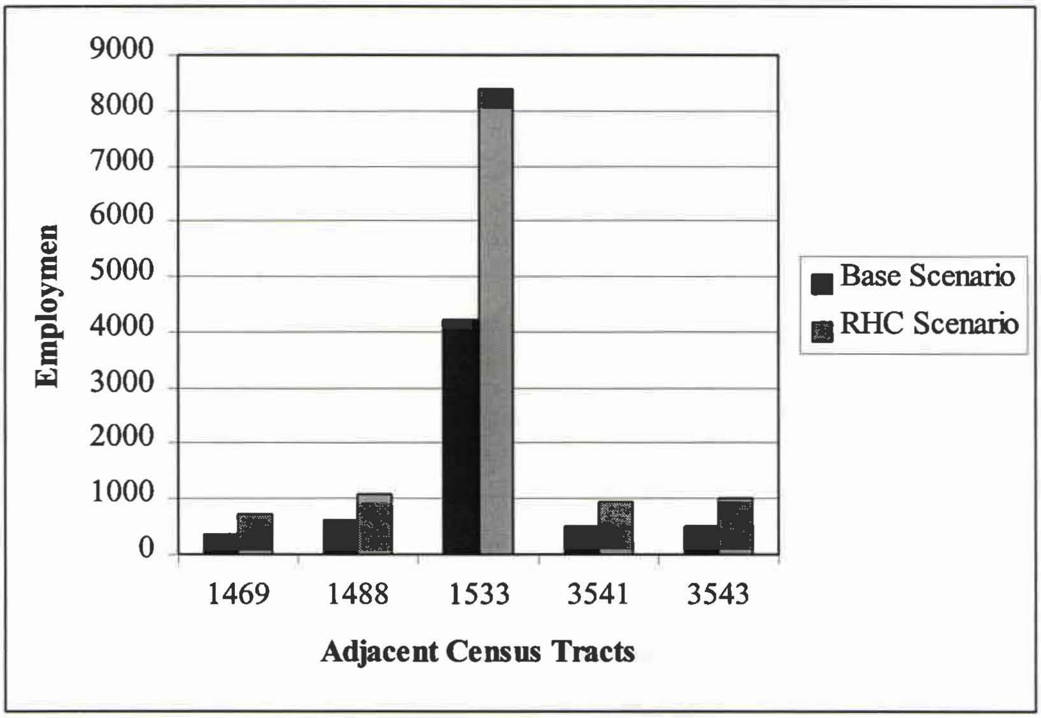
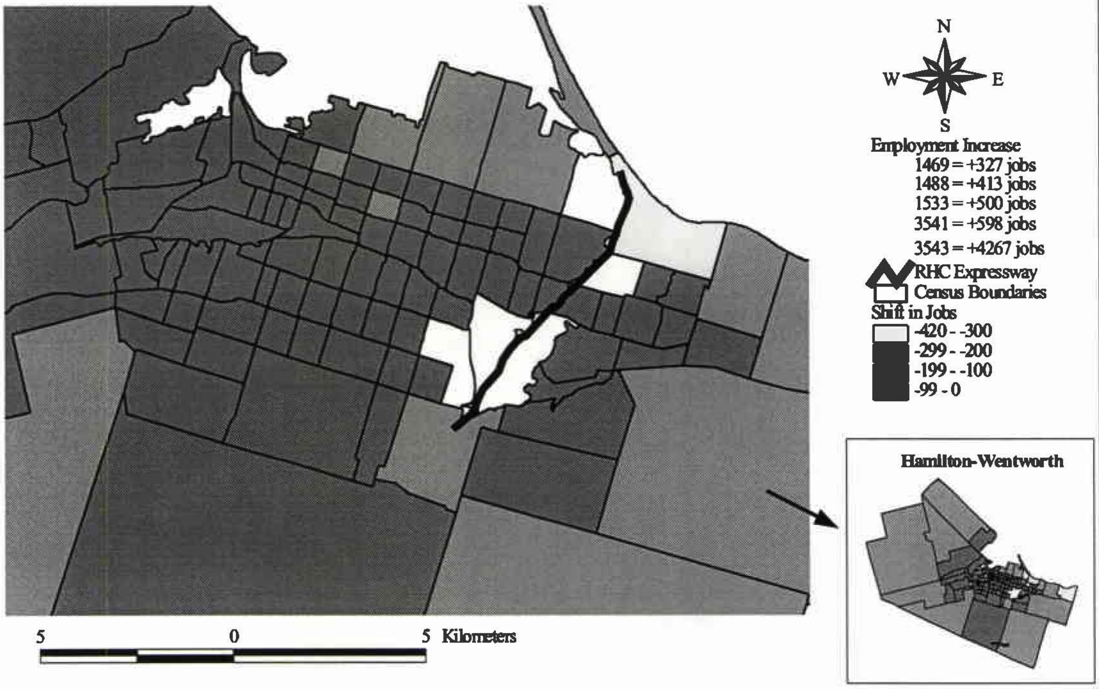


Figure 5.5 indicate that the numbers of jobs are more likely to increase in the census tracts lying in close proximity to the expressway. In fact the construction of the expressway result is a substantial change in the number of jobs that ranges between 80.51% to 95.97% for those zones. These are attracted from other census tracts. Figure 5.6 shows the shift in employment that accompanies the construction of the expressway.

The figure shows the tracts that gain jobs and are adjacent to the expressway. It also shows that the core and the suburbs experience a loss in jobs due to the construction of the expressway. Results in chapter four indicate that retail and services firms are more sensitive to highway proximity than other types of firms, thus, it is expected that the highest level of employment shift in the region will be in retail and services firms.

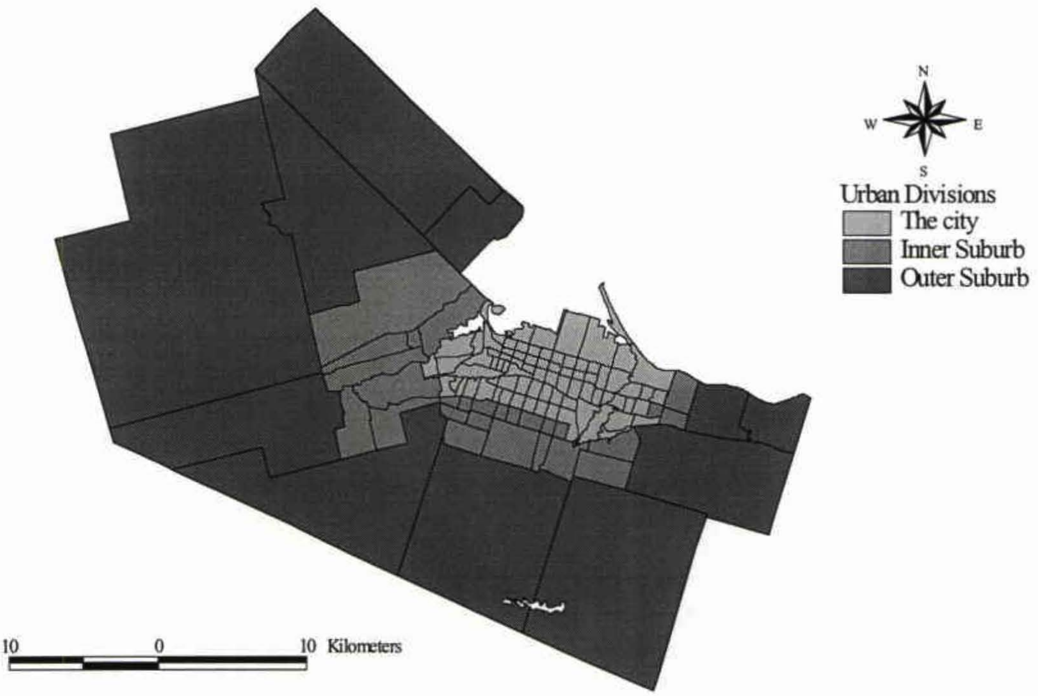
Figure 5.6 Shift in Jobs due to the construction of the Red Hill Creek Expressway



5.4.3 Sprawl Scenario Simulation

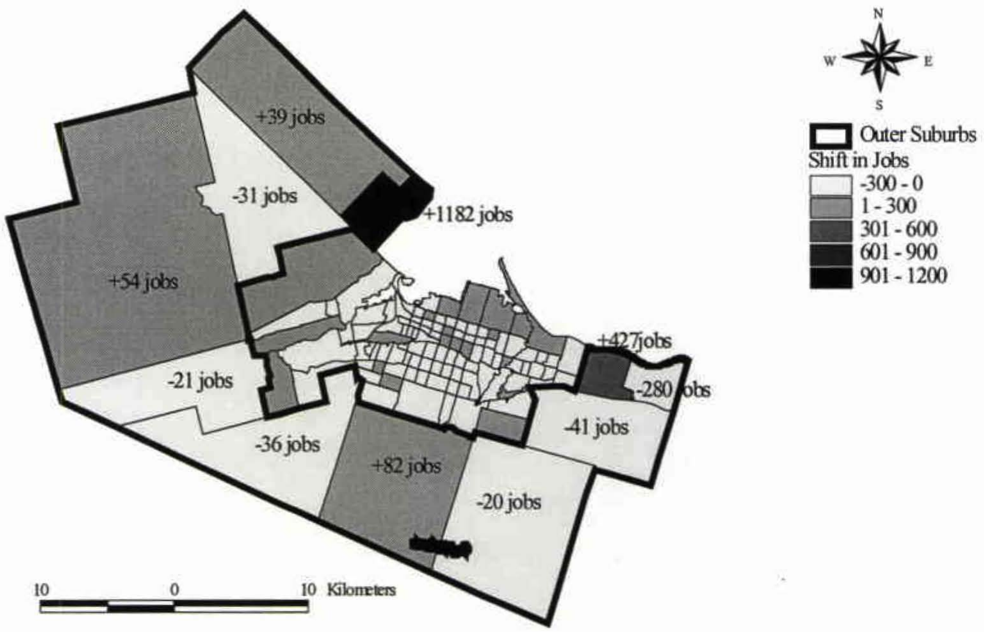
Urban sprawl is a common phenomenon, taking place in many North American cities. According to Darovny (1998) urban sprawl is characterized by large tracts of low density housing in the fringe of the city. It results from the influence of the transportation system where the dependence on automobile has promoted development at greater distance from city centers. To represent sprawl in the model and test its impact on the redistribution of firms in the region, we will assume that each census tract in the outer suburbs of the region will experience an increase in population by 50%. Given that the area of tracts in the fringe is large, the household density will remain low. Census tracts experience the change in population are the tracts that lie in the outer suburb section of the region. Divisions of the region in terms of the city, the inner suburb and the outer suburb are presented in Figure 5.7.

Figure 5.7 The Divisions of the Urban City



The results indicate how sprawl affects the redistribution of firms in the region. This is evident in Figure 5.8 where a shift in employment from the core of the city to the suburb takes place. The results from the scenario show that the employment in the outer suburb will increase by 4.16% for some tracts. Given that the total employment remains the same from the base scenario, sprawl caused a shift in employment from the different parts of the inner suburb and the city towards the outer-suburb. The 4.16% increase in jobs corresponds to 1355 new jobs created in the outer suburb due to the shift in employment from the city and the inner suburb. The city experiences 895 jobs shifting out from it towards the outer suburb. Moreover, the inner suburb experiences 460 jobs shifting out towards the outer suburb. These figures indicate that net outflow of jobs is higher with respect to the city. Thus, sprawl will affect the redistribution of firms and will cause decentralization. Figure 5.8 presents the shift in employment that accompanies sprawl.

Figure 5.8 Shift in Employment Due to Sprawl



Nevertheless, some other tracts in the outer-suburb experience loss in jobs. This loss is not large and is believed to occur because the model is sensitive to high household density. In chapter four, it is shown that the parameter associated with the household density is larger in magnitude than that associated with the population density. In this scenario, the relation between the population variable and household variable is evident in the loosing tracts in the outer suburb. In the scenario, an increase in population is associated with an increase in the household density, thus, tracts that exhibits large household density will tend to loose jobs rather than gaining. Moreover, some industries are sensitive to population. For example, manufacturing firms in the model tend to avoid populated areas, thus assuming that an increase in population in some tracts of the outer suburb will result in loss of jobs if these already contain manufacturing firms.

5.5 Conclusion

In this chapter of the thesis the employment model is implemented and tested. The model was implemented using the GAUSS programming language. The implemented version of the model allowed for testing the goodness-of-fit of the model. The model appears to have a significant goodness-of-fit with an r-square value of 0.91 between the predicted and observed employment.

Scenario simulation was held using the implemented version of the model. Three groups of scenarios were constructed and tested. The first tested the impact of the change in the final demand of certain sectors of the economy on the redistribution of firms and employment in the region. The results came out as expected. A decrease in the demand for the manufacturing sector yields a decline in employment at the areas that is predominantly in manufacturing industries. An associated increase in the demand for services, increases the employment level at certain areas in close proximity to malls and highways.

The second scenario tested the effect of building a new expressway on the redistribution of employment in Hamilton-Wentworth. The proposed Red Hill Creek expressway is included in the model. Again the results came out as expected where a shift in employment towards the tracts that lie in close proximity to the expressway is observed. This matches the findings of chapter four that firms in the different industries are sensitive to highway and expressway proximity.

The third scenario tested the effect of sprawl on the redistribution of firms in the region. Results have shown that firms are sensitive to development, thus as sprawl occur, more firms are likely to move to the outer suburbs. This is evident in the results drawn from the third scenario where we noticed that a shift in employment from the core and the inner suburb to the outer suburb is evident due to the increase of population in those areas.

CHAPTER SIX

THESIS CONCLUSION

6.1 Findings

In this thesis, an employment location model is developed and implemented for the Hamilton CMA. The redistribution of employment is associated with the location of firms, thus, firm mobility is modeled and the employment location is inferred from that. The model is calibrated for the time period 1990 to 1997. The formal structure of the model combines three inter-linked models that operate together to project employment at the census tract level. For a given census tract level, the number of firms at time t + 1 is a function of those firms at time t, firms that are lost between time t and t + 1, and firms that are gained in the same time period. Thus, the model deals with the employment change on the basis of the net change in employment between time t and t + 1. The net change is the difference between the gained and lost firms at each census tract.

The first model is an input-output model evaluated for the Hamilton-Wentworth region. Provincial input-output transaction accounts for Ontario are scaled down to local accounts using the location quotient technique. The model is used to project the total employment in each industrial sector of the economy using an exogenous final demand.

The second model is a destination choice model based on discrete choice theory. A Multinomial Logit (MNL) model is used to distribute new and relocating firms to the different census tracts. Profit maximization is assumed to influence the locational decision of different types of firms. A different MNL model is estimated for each of the five sectors modeled. These are the manufacturing, construction, wholesale, retail and services. The findings of the logit models are summarized in the following table:

Table 6.1 Location Factors for New and Relocating Firms in Hamilton-Wentworth by economic sector

Intra-urban Factor	*Manu*[1]	*Const*[2]	*Wholesale Trade*	*Retail Trade*	*Services*
CBD: Close proximity	--	--	-	++	++
Highway: Close proximity	++	+	+	++	+++
Malls: Close proximity	-	-	-	+	-
Population Size (large)	-	--	+	++	++
Household Density(high)	-	--	--	--	-
Agglomeration Economies	+	++	++	++	++

Key:
+ = positive influence; ++ = strong positive influence; +++ = very strong positive influence
- = negative influence; -- = strong negative influence; --- = very strong negative influence
1 Manufacturing industries
2 Construction industries

The third model is a combination of spatial and non-spatial regression models. It is used to predict the number of lost firms at the census tract level. Number of lost firms and total firms appear to be associated through a linear relationship and therefore, a regression model is sufficient to model the phenomenon. However, due to the spatial autocorrelation that expresses the lost number of firms in the retail and services sectors, a spatial-regression model (SAR model) is used to model the phenomenon.

Finally, the three models are linked together to form the employment model. The model is implemented using the GAUSS programming language. The goodness-of-fit of the model is tested. The employment location model developed in this thesis shows a significant goodness-of-fit between the predicted and observed values with an r-square value of 0.91. Scenario simulation was conducted using the implemented version of the model. The model showed that it is capable of simulating certain types of scenarios that will result in a change in the spatial distribution of firms.

6.2 Direction for Future Research

The motivation for developing model systems is to analyze the effects of alternative polices, plans and scenarios. Often, the limitation of data sources forces the researcher to compensate by applying theoretical models to synthesize the missing data. Modeling a geographical phenomenon is an art that requires building a comprehensive model in which a series of linked submodels form the main model. Once an operational version is run, extensions and refinements can be applied to improve the model and use it for long-term projections and policy analysis.

In the context of our employment model, a couple of drawbacks are worth noted. Firstly, the input-output model used in the employment model assumes fixed technology. In reality, the technical coefficients in the input-output model do change with time (Miller and Blair, 1985; Toyomane, 1988). Thus, revising the coefficients is a major methodological task to be undertaken before any input-output model can qualify for policy analysis involving long-run simulations. The change in technology in our model is handled exogenously through the design of scenarios. Nevertheless, an ambitious goal is to deal with the change in technology endogenously.

Toyomane (1988) suggests methods that enable a multi-regional input-output model to be used in long-term projections. Perhaps, employing a multi-regional input-output model within our employment model can meet the methods suggested by Toyomane. Thus, dividing the CMA into regions (say the major municipalities) and develop a multi-regional input-output model within our model may overcome the problem confronting the input-output models as a weak tools used in long-term projections.

Secondly, the relation between residential and commercial land use in the model is treated in one direction. The model assumes that residential development occurs first and commercial development follows it. A more realistic picture of the process involves a feed back loop between these two types of developments. In other words, residential development may occur due to commercial development and vise versa is also true. Thus, in order to simulate the change in land use structure, the employment model should be linked to the POPMOB model of the IMULATE system to achieve the feed back loops and obtains a more realistic simulation land use model.

APPENDIX A

CLUSTERING OF FIRMS IN THE DIFFERENT INDUSTRIES IN HAMILTON-WENTWORTH (1990 - 1997)

The kernel maps are produced for all the industrial sectors and for the two time periods 1990 and 1997. Figures A.1 and A.2 show the cluster pattern of the manufacturing firms in the region. The figures reveal the fact that manufacturing cluster at specific sites. These types of firms are located at the harbor, in the northern part of the Hamilton and Stoney Creek. The Southeast of the city of Hamilton shows high intensity of firm cluster. The intensity is shown to increase in the year 1997. Construction firms are shown in figures A.3 and A.4. Clustering of firms in this type of industry appears to take place in the lower part of the city of Hamilton nearby the harbor strip. Also, the Southeast of the city exhibits high intensity. This even shows to become more intense in the year 1997. The towns of Flamborough and Ancaster seem to have some firms cluster in them, this appears to be in places close to major highways. The city of Stoney Creek shows an increase in the clustering intensity between the two years.

Transportation firms show high clustering in the harbor area, close to the Q.E.W highway that links between Hamilton and Burlington. Also, some clusters show to take place in the western part of the city of Hamilton. Other municipalities show a pattern of clustering in proximity to highways in the year 1997. Nevertheless, the clustering of transportation firms appears to decrease in the year 1997. Communication firms tend to agglomerate in the lower part of the city of Hamilton, also, some parts of mountain show clustering pattern. The city of Stoney Creek shows the same type of clustering. Wholesale trade seems to cluster and agglomerate in areas that have manufacturing firms in them, in fact, the clustering pattern of these firms is very similar to the pattern shown by the manufacturing firms in the region. In addition, the towns of Flamborough, Ancaster and Glanbrook show a pioneer of clustering in them in the year 1997. This tends to be in high proximity to highways.

Figures A.11 and A.12 show the clustering of retail trade firm in the CMA. A glance on the kernel maps for this industry indicates that the core of the city maintains the highest amount of intensity among other areas. Areas occupied by shopping malls tend to have cluster around them, this is evident in both the lower and the upper part of the city. The town of Dundas appears to have a clustering pattern as well. Moreover, the general trend of clustering seems to be dispersed in the city, this appears to have high affiliation with areas with high and medium household density when comparing figure A.12 with Figure 3.3.

Services firms of the different types show high affinity of clustering in the core. FIRE, business and government firms tend to cluster in the core and around it. Education firms are more dispersed in the region but still show high linkage with the core of the city. Heath and social insurance firms cluster in the lower part of the city of Hamilton, the inner part of the mountain and in the Town of Dundas. Accommodation, food and beverages services firms show the same trend shown by the health and social insurance firms, however, the latest seems to have more dispersed pattern in the region, especially on the mountain of the city of Hamilton. Other services firms in Figure A.25 and A.26 show a pattern of clustering similar to that shown by the retail trade firms. Cluster at the core of the city is predominantly, however, the clustering seems to be more dispersed and have high affiliation with the populated areas.

Figure A.1: Clustering pattern of firms in manufacturing industries (1990)

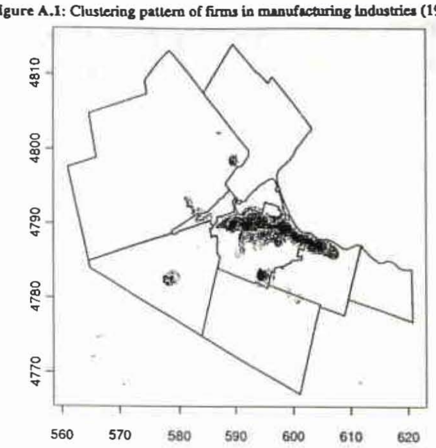


Figure A.2: Clustering pattern of firms in manufacturing industries (1997)

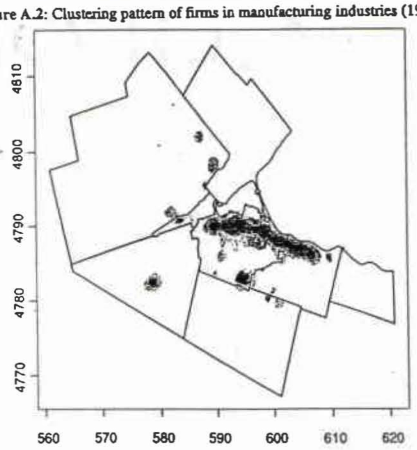


Figure A.3: Clustering pattern of firms in construction industries (1990)

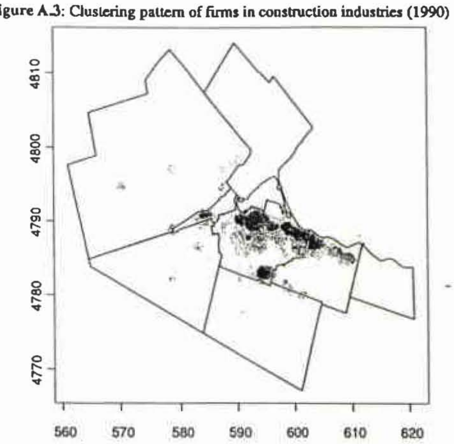


Figure A.4: Clustering pattern of firms in construction industries (1997)

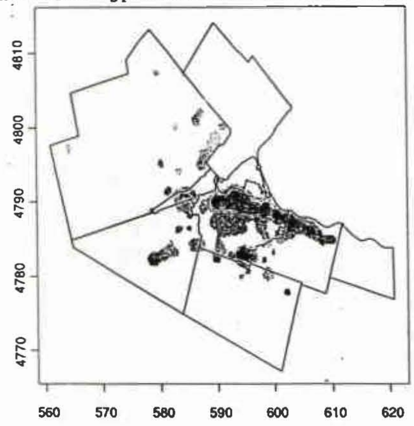


Figure A.5: Clustering pattern of firms in transportation industries (1990)

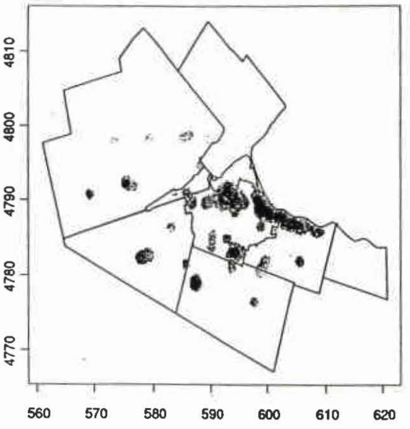


Figure A.6: Clustering pattern of firms in transportation industries (1997)

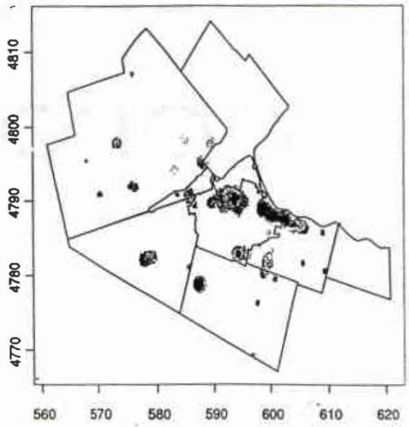


Figure A.7: Clustering pattern of firms in communication industries (1990)

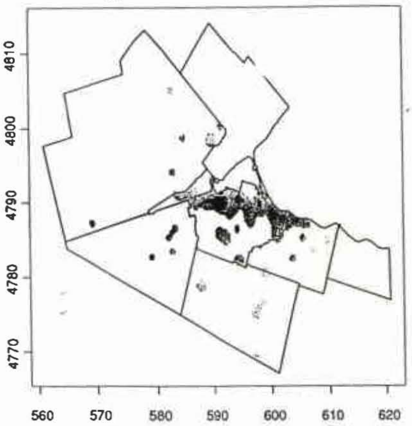


Figure A.8: Clustering pattern of firms in communication industries (1997)

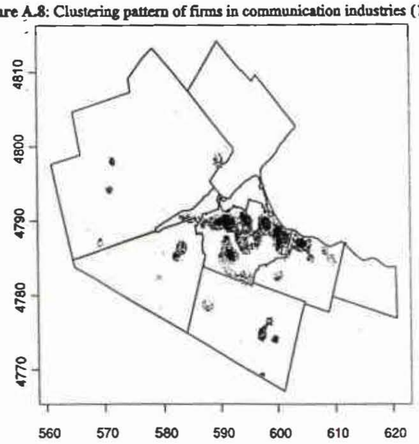


Figure A.9: Clustering pattern of firms in wholesale trade industries (1990)

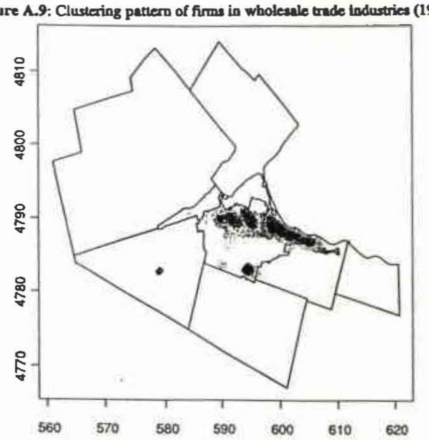


Figure A.10: Clustering pattern of firms in wholesale trade industries (1997)

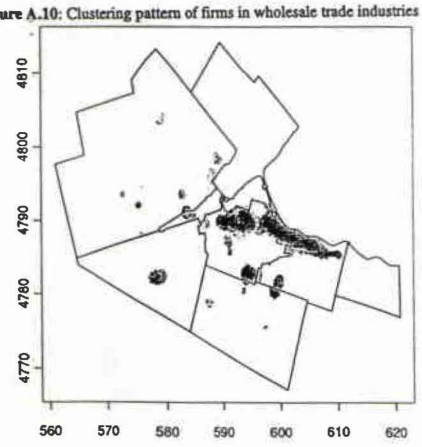


Figure A.11: Clustering pattern of firms in retail trade industries (1990)

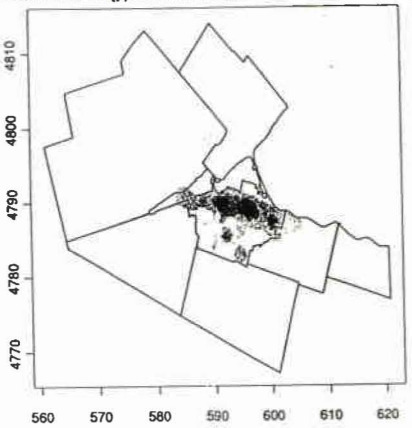


Figure A.12: Clustering pattern of firms in retail trade industries (1997)

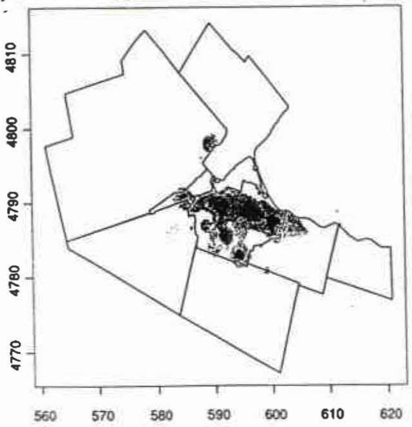


Figure A.13: Clustering pattern of firms in FIRE services industries (1990)

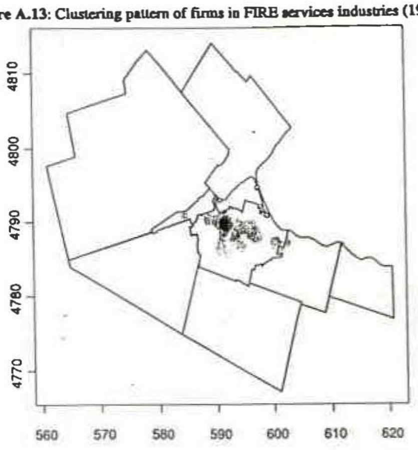


Figure A.14: Clustering pattern of firms in FIRE services industries (1997)

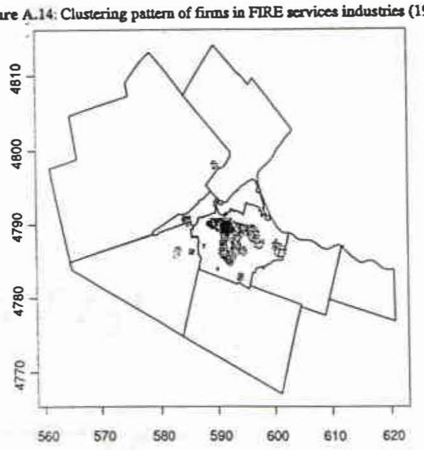


Figure A.15: Clustering pattern of firms in business services industries (1990)

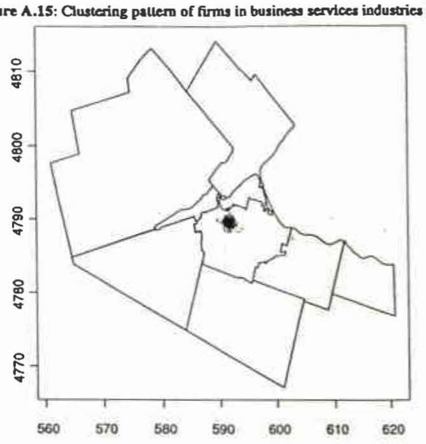
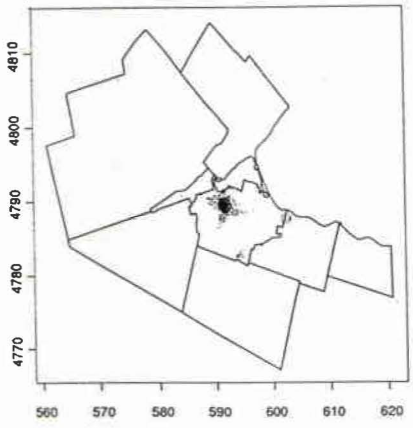


Figure A.16: Clustering pattern of firms in business services industries (1997)



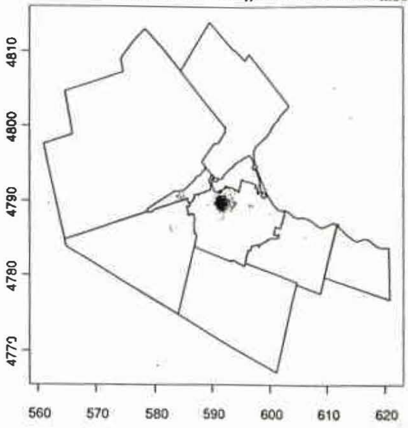


Figure A.17: Clustering pattern of firms in government services industries (1990)

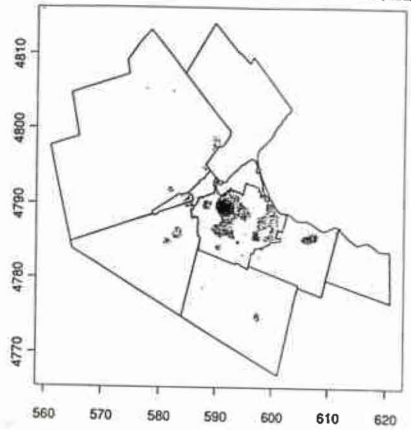


Figure A.18: Clustering pattern of firms in government services industries (1997)

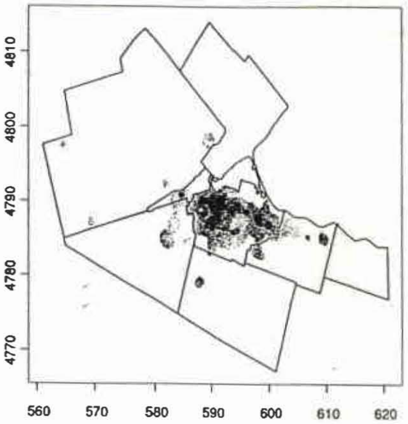


Figure A.19: Clustering pattern of firms in education services industries (1990)

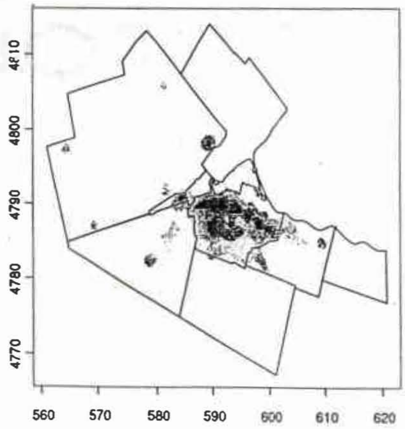


Figure A.20: Clustering pattern of firms in education services industries (1997)

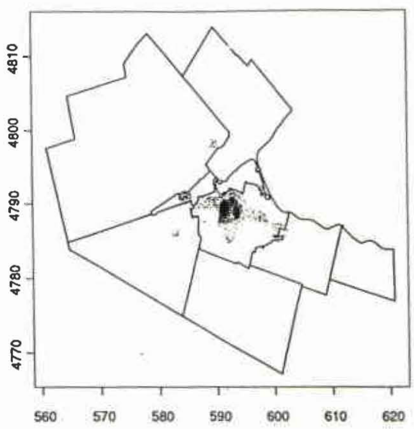


Figure A.21: Clustering pattern of firms in health and social insurance services industries (1990)

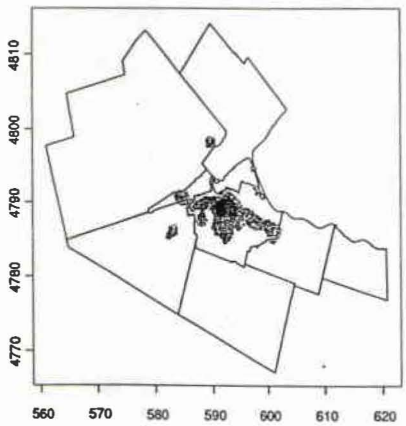


Figure A.22: Clustering pattern of firms in health and social insurance services industries (1997)

Figure A.23: Clustering pattern of firms in accommodation, food and beverages services industries (1990)

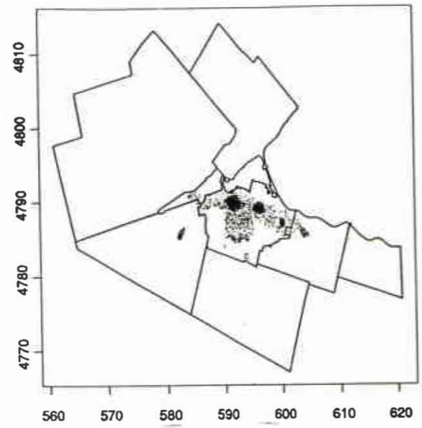


Figure A.24: Clustering pattern of firms in accommodation, food and beverages services industries (1997)

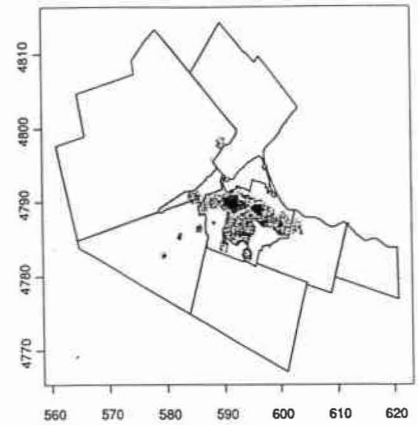


Figure A.25: Clustering pattern of firms in other services industries (1990)

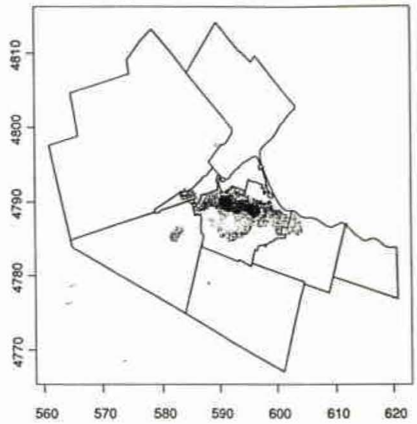
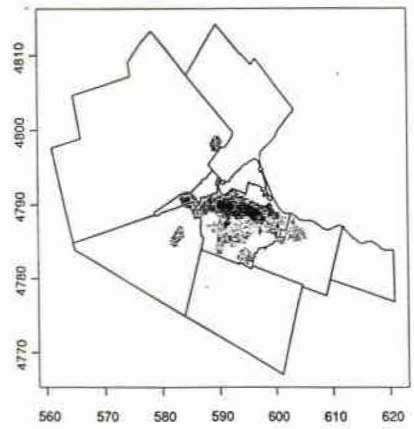


Figure A.26: Clustering pattern of firms in other services industries (1997)



APPENDIX B

CLUSTERING OF NEW AND RELOCATING FIRMS IN THE DIFFERENT INDUSTRIES BETWEEN 1990 AND 1997

B.1 Specification of the dependent Variable in the Logit models

The purpose of this section is to use the kernel estimation as an exploratory tool to build the dependent variable of the five-logit models to be estimated in Chapter 4. We will use the locational pattern of the new and relocating firms in the subsector of each industrial sector to build categorical variables that can be used to identify the firms as individuals rather than groups at the census tracts level. This can be done by assigning the value 1 if two or more subsectors of one sector show the same locational pattern, 0 otherwise. Thus, kernel maps are produced and analyzed, these are listed in Section B.2 of this appendix.

The new and relocating firms in manufacturing tend to locate and cluster in areas that are already occupied by manufacturing firms; this can be seen in Figures (B.1 - B.10) in Appendix B. Kernel maps were produced for those sectors that have more than 20 firms. The different types of manufacturing firms show the same locational pattern. Thus, grouped data represented by the proportion of new and relocating firms is used to represent the choice. This is also the case for construction and wholesale trade firms. The construction sector is divided into four subsectors: building, development and general contracting (BDG) industries; industrial and heavy construction (IHC) industries; trade contracting (TC) industries; and service industries incidental to construction (SIIC). Out of 374 new and relocating firms 24% belong to the (BDG) subsector and 62% belong to (TC) subsector. These two types of firms have similar locational pattern as shown in Figures (B.11, B.12). Again, this led us to use group data to present the choice. Wholesale trade firms comprises 340 gained firms in total, all of which tend to have the same general locational trend. The majority of the wholesale trade firms are in ***SIC56***, ***SIC57***, ***SIC58*** and ***SIC59***. The percentages corresponding to these types of firms are 15%, 30%, 26% and 17% respectively, all of which show the same trend of clustering in the region as can be seen in Figures B.13, B.14, B.15 and B.16.

Retail Trade and services firms put together make-up the vast majority of firms in the region. Therefore the spatial distribution of firms in these types of sectors are expected to have a more diverse trend within the sub-sectoral types of firms for both industries. On this basis, a distinction between the different types of firms in the one sector can be formed. This allows for a more flexible model and aims to capture the phenomena modeled in an efficient manner.

Retail trade sector emanates from seven subsectors at the two digits ***SIC*** code level. These are listed in Table B.1 with the total number of gained firms in each sector. The table reveals that retail trade sector is a large and diverse sector. A clearer picture can be drawn by looking at the kernel maps that correspond to the gained firms in the different sectors of retailing. Figures (B.17 - B.22) show the clustering pattern of the different types of firms within retailing in the entire region. The maps show a diverse pattern of clustering. For example, firms in ***SIC61*** and ***SIC62*** have the tendency to locate in close proximity to malls. Moreover, firms in ***SIC60***, ***SIC64*** and ***SIC65*** show the propensity of locating in areas with population. By comparing the intensity in the kernel maps for those types of firms with the population map Figure 3.4, these firms seem to follow people. The results suggest that firms in the different sectors can be categorized by the locational pattern they show. Information about the decision-maker in the logit model can be included to represent the individual (firm) characteristics.

Table B.1 Gained firms in the Retail sector

Sector	*Gained Firms*	*Percentages*
SIC60	163	9.74
SIC61	172	10.28
SIC62	176	10.52
SIC63	434	25.94
SIC64	125	7.47
SIC65	600	35.86
SIC69	3	0.18

Key:
SIC60 = Food, beverage and drug industries, SIC61 = Shoe, apparel, fabric and yarn industries, SIC62 = Household furniture, appliances and furnishing industries, SIC63 = Automotive vehicles, parts and accessories, sales and services, SIC64 = General retail merchandising, SIC65 = Other retail store industries, and SIC69 = Non-store retail industries.

A similar locational pattern exists between the service sector firms. The services sectors to be modeled are grouped in one sector. With the opportunity to model individual data it is more efficient to merge different types of sectors into one sector.

This will allow for sufficient large sample, thereby increasing the flexibility of the statistical power of the logit model. The different sectors that form the services sector to be modeled are shown in Table B.2 with the total number of gained firms. Kernel maps showing the clustering of firms in the different sectors are produced for the sake of capturing the general pattern of location in the region. Figures B.23 to B.27 shows the pattern under investigation. As in the case of retailing firms, different firms in the different sectors of services show a different clustering in the region. However, some sectors share some general trends. Figures B.23 and B.24 show that firms in FIRE and Businesses sector have high affiliation with the core. Furthermore, firms in the last three sectors listed in Table B.2 seem to be dispersed in the region and appear to follow population when compared with figure 3.4.

Table B.2 Gained firms in the services sector

Sector	*Gained Firms*	*Percentages*
FIRE	348	12.77
Businesses	540	19.81
Health & Social	498	18.27
Accommodation, food and beverages	393	14.42
Other services	947	34.74

B.2 Kernel Estimation of New and Relocating Firms

Here, the kernel estimation of the new and relocating firms between the period 1990 and 1997 is achieved using the Splus 4.5 software package. Results are introduced in Figures B.1 - B.27

Figure B.1: New and Relocating firms - Food Industries (SIC10)

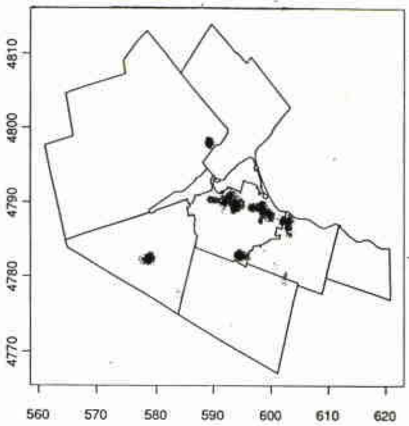


Figure B.2: New and Relocating firms - Beverage Industries (SIC11)

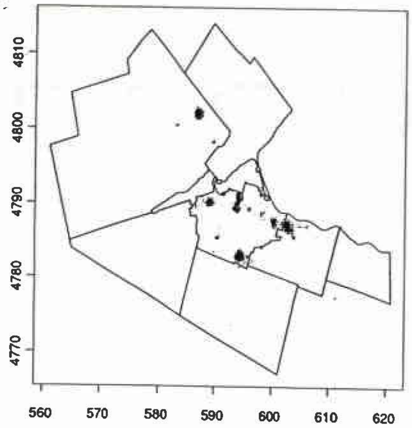


Figure B.3: New and Relocating firms - Plastic Products Industries (SIC16)

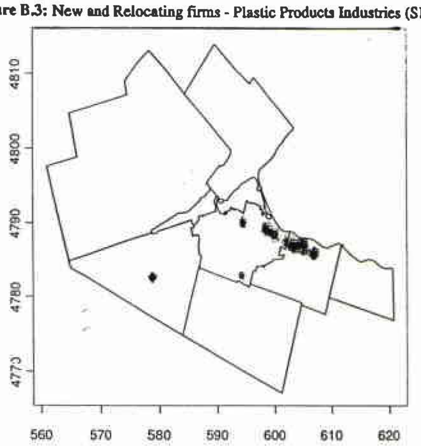


Figure B.4: New and Relocating firms - Textile Products Industries (SIC19)

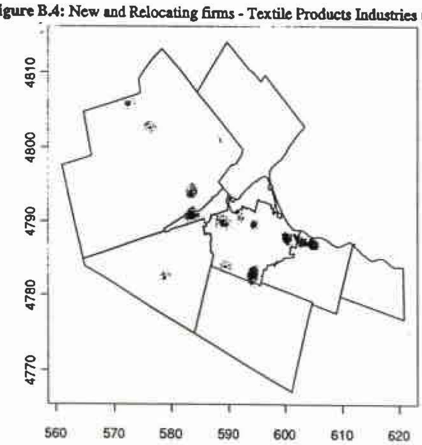


Figure B.5: New and Relocating firms - Wood Industries (SIC25)

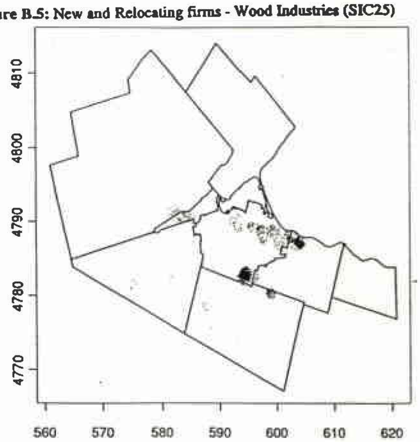


Figure B.6: New and Relocating firms - Printing Industries (SIC28)

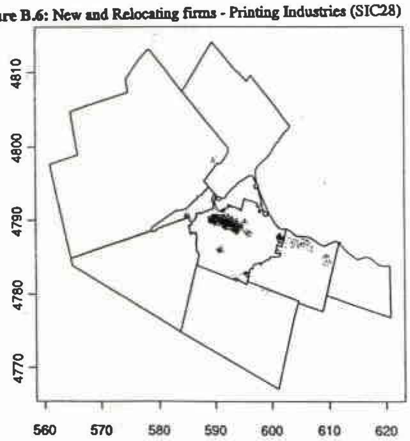


Figure B.7: New and Relocating firms - Primary Metal Industries (SIC29)

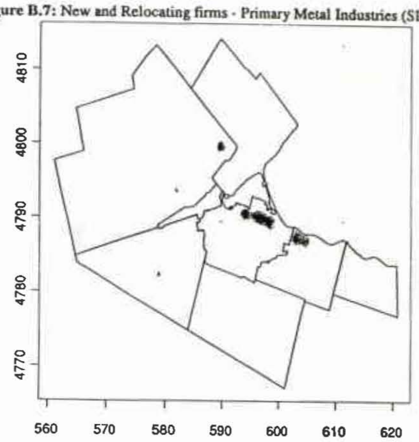


Figure B.8: New and Relocating firms - Fabricated Metal Products Industries (SIC30)

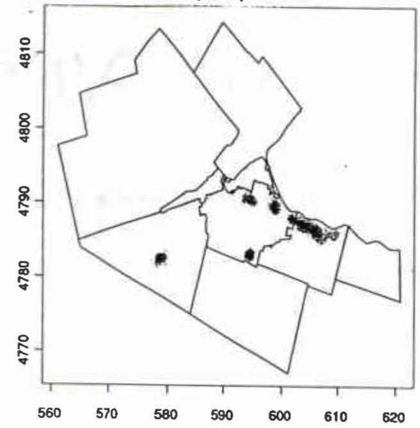


Figure B.9: New and Relocating firms - Machinery Industries (SIC31)

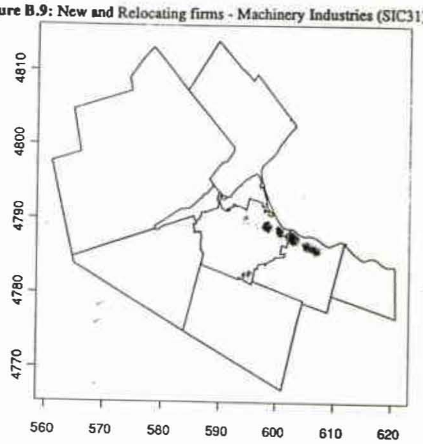


Figure B.10: New and Relocating firms - Other Manufacturing Industries (SIC39)

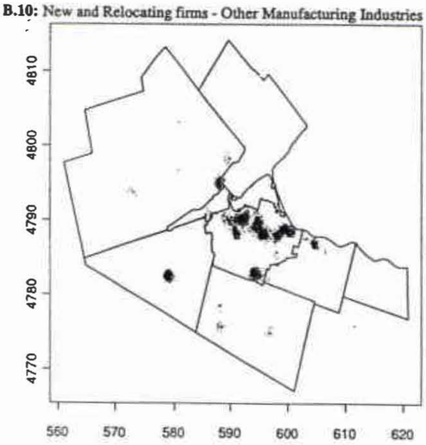


Figure B.11: New and Relocating firms - Building, Developing and General Contracting Industries (SIC40)

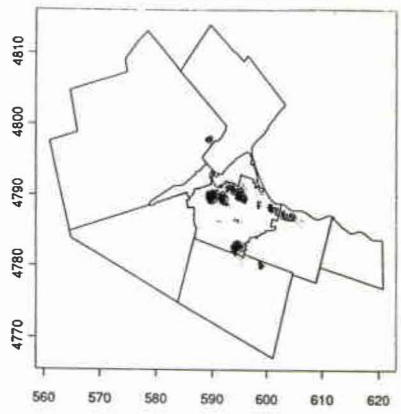


Figure B.12: New and Relocating firms - Trade Contracting Industries (SIC42)

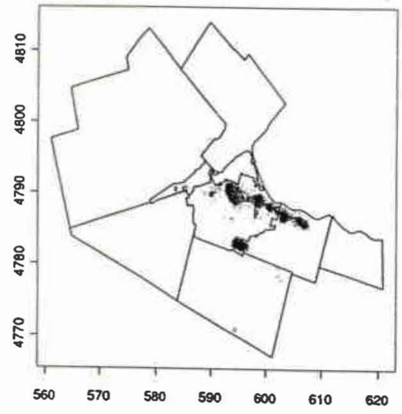


Figure B.13: New and Relocating firms - Motor Vehicles, Parts and Accessories Industries (SIC55)

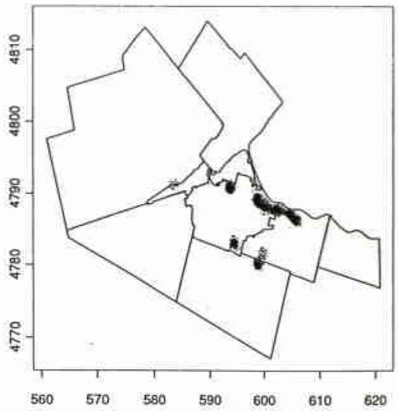


Figure B.14: New and Relocating firms - Metals, Hardware, Plumbing, Heating and Building Material Industries (SIC56)

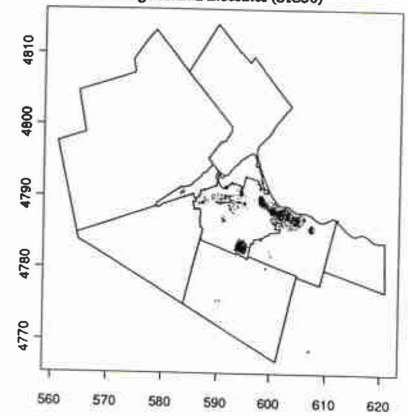


Figure B.15: New and Relocating firms - Machinery, Equipment and Supplies Industries (SIC57)

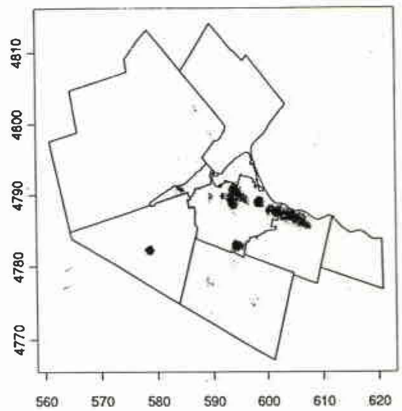


Figure B.16: New and Relocating firms - Other Products Industries (SIC59)

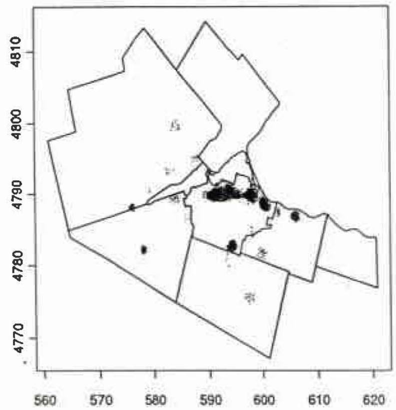


Figure B.17: New and Relocating firms - Food, Beverage and drug Industries (SIC60)

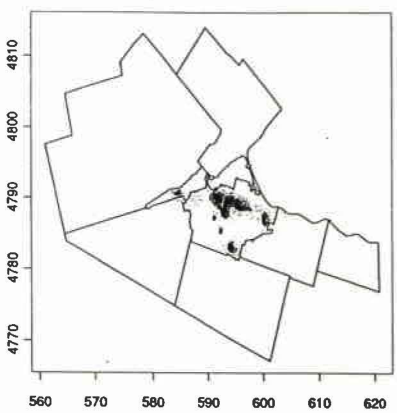


Figure B.18: New and Relocating firms - Shoe, Apparel, Fabric and Yarn Industries(SIC61)

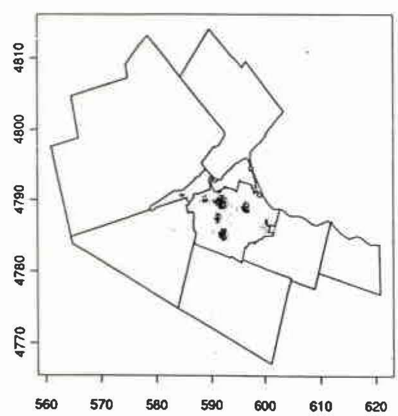


Figure B.19: New and Relocating firms - Household Furniture, Appliances and Furniture Industries (SIC62)

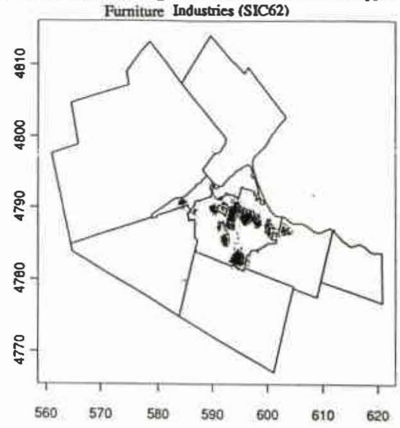


Figure B.20: New and Relocating firms - Automotive Vehicles, Parts and Accessories, Sales and Service Industries (SIC63)

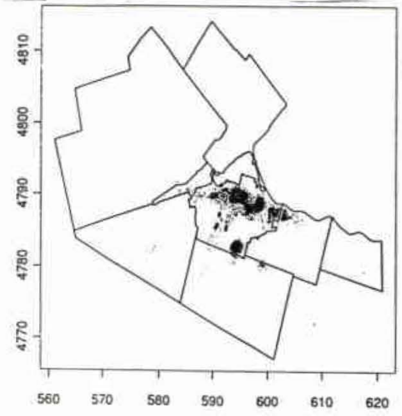


Figure B.21: New and Relocating firms - General Retail Merchandising Industries (SIC64)

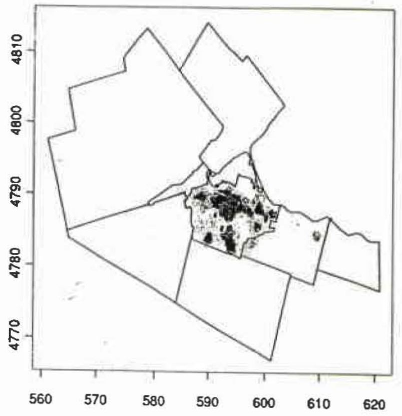


Figure B.22: New and Relocating firms - Other Retail Store Industries (SIC65)

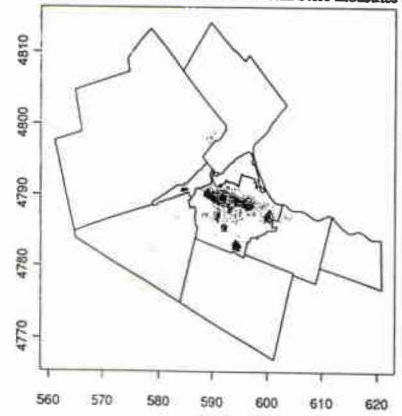


Figure B.23: New and Relocating firms - F.I.R.E Industries (SIC70 - SIC76)

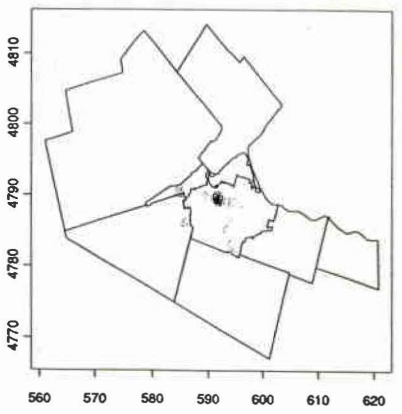


Figure B.24: New and Relocating firms - Businesses service Industries (SIC77)

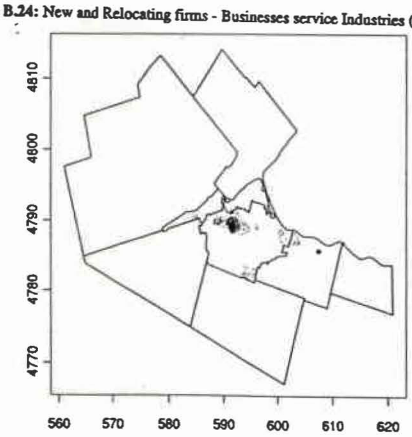


Figure B.25: New and Relocating firms - Health and Social Service Industries (SIC86)

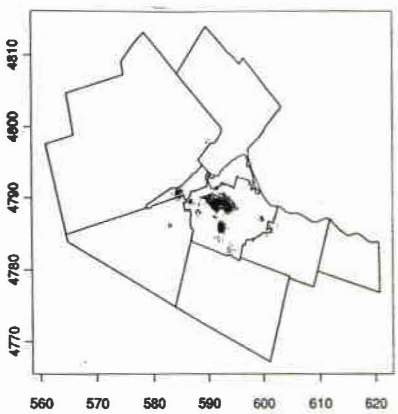


Figure B.26: New and Relocating firms - Accommodation, Food and Beverage Service Industries (SIC91 - SIC92)

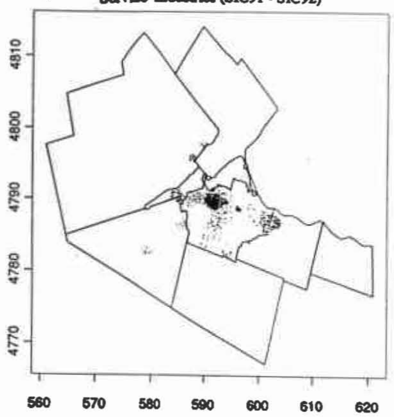
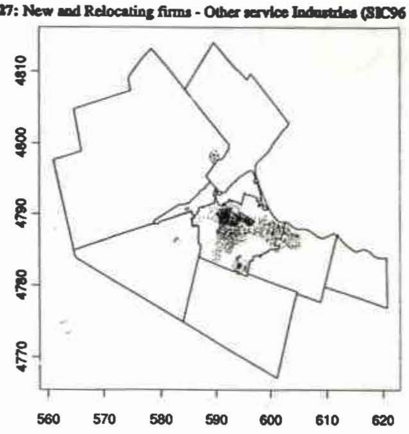


Figure B.27: New and Relocating firms - Other service Industries (SIC96 - SIC99)



Appendix C

Model Implementation - EMPLOC

C.1 User Manual

C.1.1 What does the model requires?

The model is designed in a way that allows the user to simulate various scenarios after running the base scenario which represent the status quo if every thing remain the same between time t and t+1. Running the base scenario is a requirement if the user wishes to run various scenarios for future projections. The model was calibrated for the year 1990 - 1997, thus, for the base case the model will use data from 1990 to project employment in the year 1997. The program will prompt the user if he wants to run the base scenario or not. If the user never run a base scenario he is obliged to do that to create the required data to simulate other scenarios.

C.1.2 What is exogenous in the model?

C.1.2.1 Particular sectors of the Economy

The model is designed to simulate employment at the census tract level for the Hamilton-Wentworth region. The sectors that form the economy in this region are the following:

Table C.1 Sectors that forms the economy

SECTOR	*Standard Industrial Code*
Manufacturing	**SIC1039**
Construction	**SIC4044**
Transportation	**SIC4547**
Communication	**SIC4849**
Wholesale Trade	**SIC5059**
Retail Trade	**SIC6069**
FIRE	**SIC7076**
Businesses	**SIC77**
Government	**SIC8184**
Education	**SIC85**
Health and Social Insurance	**SIC86**
Food and Beverages	**SIC9192**
Other Services	**SIC9699**

The EMPLOC model is designed to simulate the major sectors of the economy. All of the above sectors are modeled except the Transportation and Communication sectors. In the services sector all sectors are grouped into one sector and modeled except the following two: the Governmental and Educational Sectors. Those sectors that are not modeled will be taken exogenously. This leaves us with the following sectors to model:

Table C.2 Sectors that are modeled

SECTOR	Standard Industrial Code
Manufacturing	SIC1039
Construction	SIC4044
Wholesale Trade	SIC5059
Retail Trade	SIC6069
Particular Services	SIC7099

However, the total employment at time t+1 is a mixture of those sectors who will be modeled and those who will be taken exogenously. Therefore, employment and firms for those exogenous sectors should be supplied to the model in order to get the required result.

C.1.2.2 Final Demand

The final demand in the Input-Output Model should be given exogenously to the model. This should be supported for each run to insure that the EMPLOC model will run with no problems.

C.1.2.3 The exogenous Data

The following files should be prepared and saved in the "emploc/database/exog" directory:

Table C.3 Exogenous data in the model

File Name	Description
Avgemp.dat	(5 x 1) vector that contains the average number of employment in the sectors to be modeled
Sic8084.dat	(116 x 1) Number of firms in each census tract for the Government sector at time t (1990)
Sic85.dat	(116 x 1) Number of firms in each census tract for the Education sector at time t (1990)
Emp4547.dat	(116 x 1) Number of employment in each census tract for the Transportation sector at time t+1 (1997)
Emp4849.dat	(116 x 1) Number of employment in each census tract for the Communication sector at time t+1 (1997)
Emp8184.dat	(116 x 1) Number of employment in each census tract for the Government sector at time t+1 (1997)
Emp85.dat	(116 x 1) Number of employment in each census tract for the Education sector at time t+1 (1997)

As for the final demand, the user should provide a file which is (8 x 1) vector including the final demand for the following sectors: manufacturing, construction, transportation, communication, wholesale trade, retail trade, all services and other sectors respectively. This file should be saved under the path "emploc\database\exog ".

C.1.3 What types of Scenarios can be simulated?

The user can develop two different types of scenarios and simulate them with the EMPLOC model. In the first type the scenarios neglect the effect of time and run the base scenarios for all the different scenarios by altering the final demand values and maintain the exogenous files as they are. In this case the user should always choose the based scenario when the program prompt him (Answer=1) . The other type of scenarios is to take the effect of time into consideration. In this case and as mentioned earlier, it is compulsory to run the base scenario for the first time. After that the user can choose the Alternative scenario (Answer = 0) only if he do the following:

1. Save a file that contains the final demand "finald.dat" under the "emploc\change" directory
2. Given that the characteristics of zones might change with time the model assumes that the following information will change with time: highway proximity, mall proximity, population size, household density and the employment accessibility. For this the user should prepare the following files and save them under the "emploc\change" directory:
   a) hwypro.dat, mallpro.dat, popsize.dat, hhlddens.dat, mctwemp.dat, rsemp.dat, semp.dat. These files should be identical to the file in the “emploc\database\destch” in structure to guarantee no errors when running the model.
   b) If some of the previous information is changed and the other remained, the use still has to support these files under "emploc\change" path. If necessary copy those from the "emploc\database\destch" to the change directory and provide others that are would change.
3. The user should also support the exogenous data in both the base and alternative scenarios.

C.1.4 Output of the Model

The model calculates the total employment at the census tract level, also it calculates the number of establishments that is suppose to occupy each of the 116 zones. The output files are saved under the path "emploc\change" with the following names: tot1039.dat, tot4044.dat, tot5059.dat, tot6069.dat, tot7099.dat and TOTEMP.DAT. The first five files correspond to the total number of firms at time t+1. The last file is the total number of employment at the census tract level. Its worth mentioning that all the previous files except for the TOTEMP.DAT should be kept in the change directory with the same names if the user is to run alternative scenario. This is important since these files will be used as a feed in for time period t+1 to project for t+2.

Table C.4 provides a spatial index of the census tracts as given in the EMPLOC model. Associated location maps are also provided below.

Table C.4 Spatial Index with census tracts Ids as given in the EMPLOC model

Order	ID	Municipality	Order	ID	Municipality	Order	ID	Municipality
1	1469	*Hamilton*	40	1509	*Hamilton*	79	2623	*Dundas*
2	1470	*Hamilton*	41	1510	*Hamilton*	80	2624	*Dundas*
3	1471	*Hamilton*	42	1511	*Hamilton*	81	2625	*Dundas*
4	1472	*Hamilton*	43	1512	*Hamilton*	82	2626	*Flamborogh*
5	1473	*Hamilton*	44	1513	*Hamilton*	83	2627	*Flamborogh*
6	1474	*Hamilton*	45	1514	*Hamilton*	84	2628	*Flamborogh*
7	1475	*Hamilton*	46	1515	*Hamilton*	85	2629	*Flamborogh*
8	1476	*Hamilton*	47	1516	*Hamilton*	86	2630	*Flamborogh*
9	1477	*Hamilton*	48	1517	*Hamilton*	87	2788	*Glanbrook*
10	1478	*Hamilton*	49	1518	*Hamilton*	88	3522	*Hamilton*
11	1479	*Hamilton*	50	1519	*Hamilton*	89	3523	*Hamilton*
12	1480	*Hamilton*	51	1520	*Hamilton*	90	3524	*Hamilton*
13	1481	*Hamilton*	52	1521	*Hamilton*	91	3525	*Hamilton*
14	1482	*Hamilton*	53	1522	*Hamilton*	92	3526	*Hamilton*
15	1483	*Hamilton*	54	1523	*Hamilton*	93	3527	*Hamilton*
16	1484	*Hamilton*	55	1524	*Hamilton*	94	3528	*Hamilton*
17	1485	*Hamilton*	56	1525	*Hamilton*	95	3529	*Hamilton*
18	1486	*Hamilton*	57	1526	*Hamilton*	96	3530	*Hamilton*
19	1487	*Hamilton*	58	1527	*Hamilton*	97	3531	*Hamilton*
20	1488	*Hamilton*	59	1528	*Hamilton*	98	3532	*Hamilton*
21	1490	*Hamilton*	60	1529	*Hamilton*	99	3533	*Hamilton*
22	1491	*Hamilton*	61	1530	*Hamilton*	100	3534	*Hamilton*
23	1492	*Hamilton*	62	1531	*Hamilton*	101	3535	*Hamilton*
24	1493	*Hamilton*	63	1532	*Hamilton*	102	3536	*Hamilton*
25	1494	*Hamilton*	64	1533	*Hamilton*	103	3537	*Hamilton*
26	1495	*Hamilton*	65	1534	*Hamilton*	104	3538	*Hamilton*
27	1496	*Hamilton*	66	1536	*Hamilton*	105	3539	*Hamilton*
28	1497	*Hamilton*	67	1539	*Glanbrook*	106	3540	*Hamilton*
29	1498	*Hamilton*	68	2611	*Hamilton*	107	3541	*Hamilton*
30	1499	*Hamilton*	69	2612	*Hamilton*	108	3542	*Hamilton*
31	1500	*Hamilton*	70	2613	*Stoney Creek*	109	3543	*Hamilton*
32	1501	*Hamilton*	71	2615	*Stoney Creek*	110	3544	*Hamilton*
33	1502	*Hamilton*	72	2616	*Stoney Creek*	111	3545	*Hamilton*
34	1503	*Hamilton*	73	2617	*Ancaster*	112	3546	*Stoney Creek*
35	1504	*Hamilton*	74	2618	*Ancaster*	113	3547	*Stoney Creek*
36	1505	*Hamilton*	75	2619	*Ancaster*	114	3548	*Stoney Creek*
37	1506	*Hamilton*	76	2620	*Ancaster*	115	3549	*Stoney Creek*
38	1507	*Hamilton*	77	2621	*Ancaster*	116	4121	*Stoney Creek*
39	1508	*Hamilton*	78	2622	*Ancaster*			

Figure C.1 Ancaster - 1986 census tracts

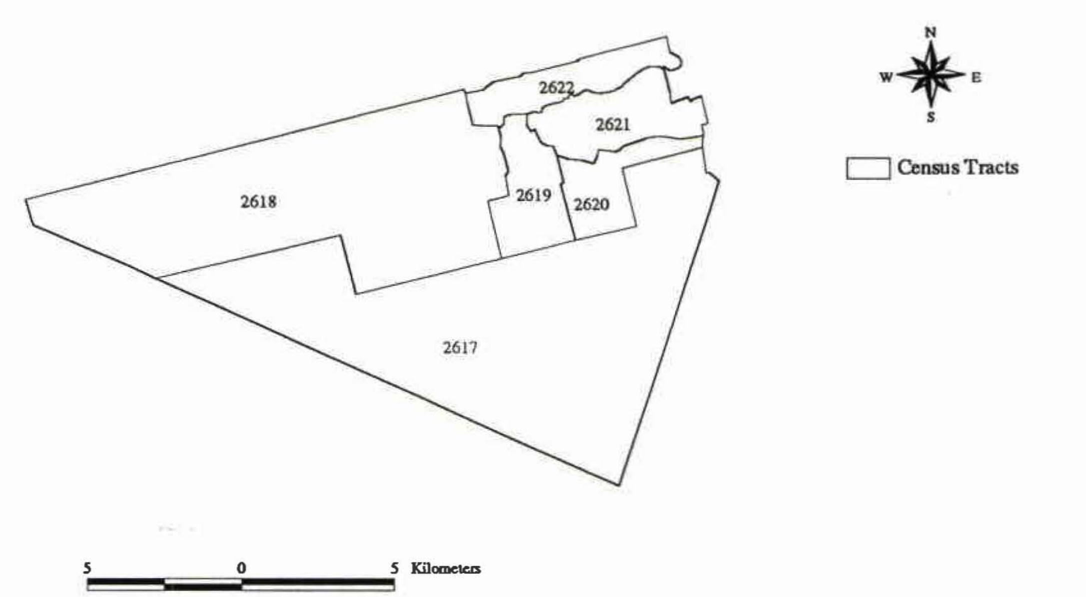


Figure C.2 Dundas - 1986 census tracts

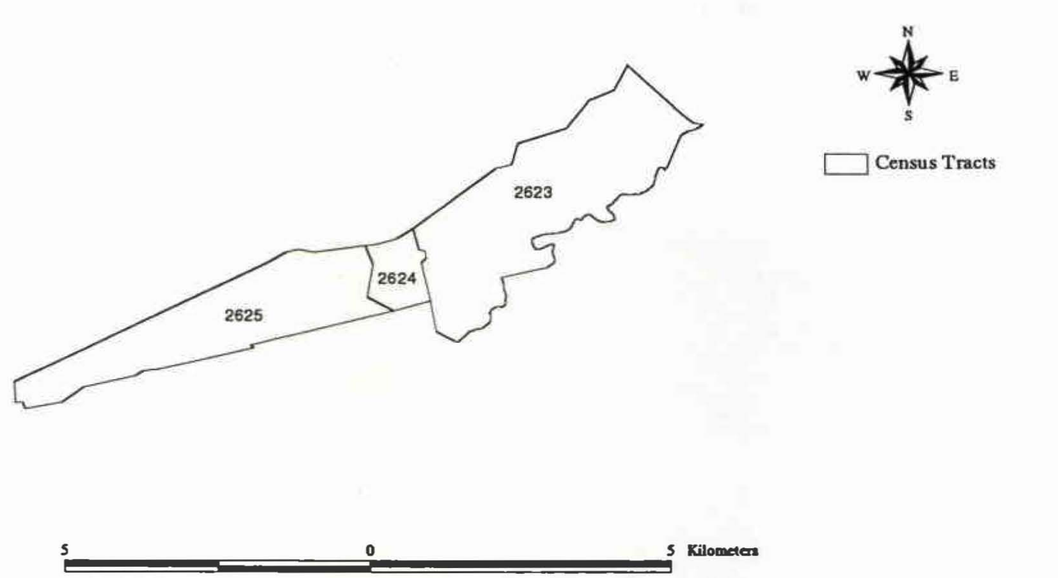


Figure C.3 Flamborogh - 1986 census tracts

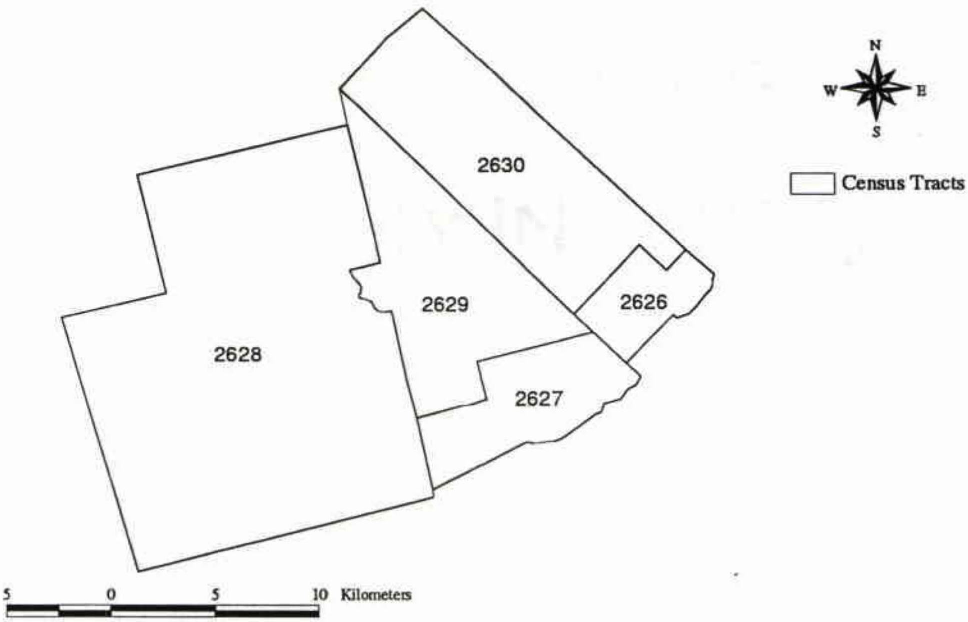


Figure C.4 Glanbrook - 1986 census tracts

Figure C.5 Hamilton - 1986 census tracts

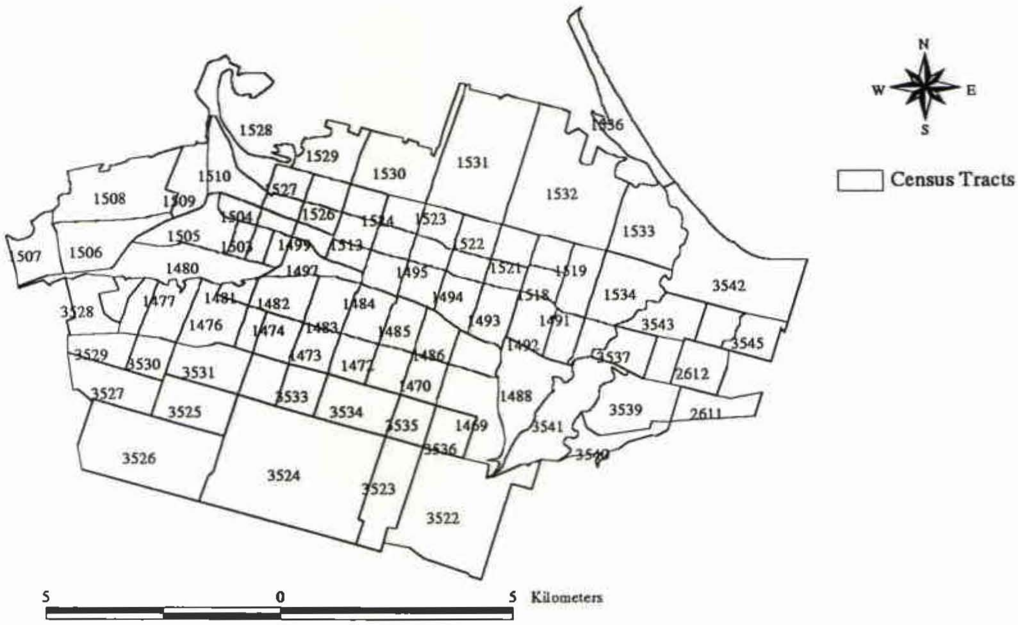
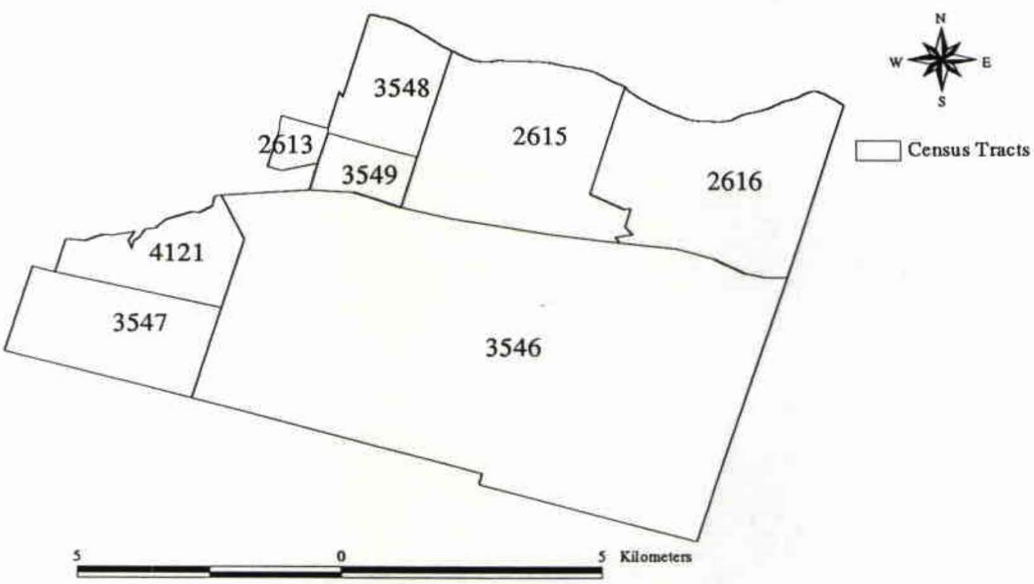


Figure C.6 Stoney Creek - 1986 census tracts



C.2 Technical Manual

C.2.1 EMPLOC Structure

The **EMPLOC** model is basically a number of programs written in gauss programming language that are linked together. Each program performs a specific task. The programs are listed according to its hierarchical structure as follows: ***EMPLOC.PRG***, ***INOUT.PRG***, ***MNL.PRG***, ***LOST.PRG*** and ***LINK.PRG***. Following is an overview of the functionality of each program.

- ***EMPLOC.PRG***: This program is used to check if the user wish to run a base scenario or an alternative scenario. Based on the answer, the program will run the ***INOUT.PRG*** program from the directory that corresponds to the users' choice. If the choice is the base scenario then the EMPLOC.PRG will run INOUT.PRG from "emploc\prog" path, otherwise it will run it from "emploc\prog\scenario". It is worth noting that the programs in both paths are identical in their structure. The only difference is the loading file statements since the files used for the base case are not the same as for the alternative scenarios.
- ***INOUT.PRG***: This program is used to read in an input-output technical coefficient matrix, final demand column vector and a firm coefficient matrix to calculate the total number of firms at the regional level for each of the modeled sectors. Once this is done the ***INOUT.PRG*** program will run the ***MNL.PRG***
- ***MNL.PRG***: This program is used to calculate the probability that a firm will choose a certain zone to locate at. This is done for all zones and all sectors. The output in this program is a set of column vectors with the probabilities for all zones. Each column

vector corresponds to one of the modeled sectors. Once the ***MNL.PRG*** generates these probabilities it runs the ***LOST.PRG*** program.

- ***LOST.PRG***: This program is used to calculate the number of lost firms from each zone between time t and t+1. The output is a set of column vectors with the number of lost firms in each zone. This is also done for all the modeled sectors. Once ***LOST.PRG*** generates these values it runs ***LINK.PRG***
- ***LINK.PRG***: This program is used to link the results generated from the previous three models. Therefore, ***LINK.PRG*** reads the total number of firms generated by ***INOUT.PRG***, the destination choice probabilities generated by ***MNL.PRG*** and the lost firms generated by ***LOST.PRG***. The model also loads some exogenous data needed for the calculation. It's the responsibility of the user to insure that the exogenous data is available within the database of the model. The output of this model is the total number of establishments in each of the modeled sectors. Also the model generate a column vector with the total number of employment at the census tract level for the time period t+1.

C.2.2 Database of EMPLOC

Next the set of databases (files) used in the different programs of the employment model is presented. A description of the input and output files with the location in which they should be saved is given.

C.2.2.1 The Input-Output Model (INOUT.PRG):

Table C.5 In-files for the Input-Output Model

Name	Location	Description
Finald.dat	Emploc\database\exog	A vector of 8 rows containing the final demand for Manufacturing, Construction, Transportation, Communication, Wholesale Trade, Retail Trade, All services and other services respectively.
Tech.dat	Emploc\database\inout	An (8 x 8) matrix containing the technical coefficients of the previous 8-sectors
Firmcoef.dat	Emploc\database\inout	Firm coefficients for the previous 8-sectors

Table C.6 The out-files of the Input-Output Model

Name	Location	Description
Firms.dat	Emploc\change	A vector of 8 rows containing the total number of firms for the 8-sectors at the regional level

C.2.2.2 The Destination Choice Model (MNL.PRG)

Table C.7 In-files for the Destination choice model

Name	Location	Description
Betaxxxx.dat[1]	Emploc\database\destch	(1 x 15) row vector contain the estimated parameters of the MNL model for various sectors.
Cnstxxxx.dat	Emploc\database\destch	(116 x 1) column vector contain the values of constants in the MNL models
Cbddist.dat	Emploc\database\destch	116 x 2 matrix. 1st column is distance from CBD, 2nd column is the log of distance
Hwypro.dat	Emploc\database\destch	116 x 1 vector. Highway proximity=1 if highway passes threw or by a zone, else 0

Table C.7 (Continue)

Name	Location	Description
Mallpro.dat	Emploc\database\destch	116 x 1 vector. Mall proximity=1 if the zone contains a mall or lie beside a zone that contains a mall, 0 otherwise
Popsize.dat	Emploc\database\destch	116 x 1 vector. Population size
Hhlddens.dat	Emploc\database\destch	116 x 1 vector. Household density (Household / Zone Area[sq -Km])
Mlanduse.dat	Emploc\database\destch	116 x 1 vector. Dummy variable = 1 if the majority of zones land is manufacturing, 0 otherwise
Mctwemp.dat	Emploc\database\destch	116 x 1 vector. Accessibility to employment in manufacturing, construction, transportation and wholesale
Rsemp.dat	Emploc\database\destch	116 x 1 vector. Accessibility to employment in retail and services
Servemp.dat	Emploc\database\destch	116 x 1 vector. Accessibility to employment in services
Rint1.dat	Emploc\database\destch	116 x 1 vector. Alternative specific variable representing specific sectors within retailing. This is used to form an interaction term with popsize variable
Rint2.dat	Emploc\database\destch	116 x 1 vector. Alternative specific variable representing specific sectors within retailing. This is used to form an interaction term with with log of the "distance to the CBD" variable
Rint3.dat	Emploc\database\destch	116 x 1 vector. Alternative specific variable representing specific sectors within retailing. This is used to form an interaction term with mallpro variable
Sint1.dat	Emploc\database\destch	116 x 1 vector. Alternative specific variable representing specific sectors within services. This is used to form an interaction term with log of the "distance to the CBD" variable
Sint2.dat	Emploc\database\destch	116 x 1 vector. Alternative specific variable representing specific sectors within services. This is used to form an interaction term with popsize variable

1) xxxx= 1039 is Manufacturing; 4044 is Construction;5059 is Wholesale Trade ;6069 is Retail Trade; 7099 is Services

Table C.8 Out-files for the Destination Choice Model

Name	Location	Description
Probxxxx.dat	Emploc\database\destch\prob	(116 x 1) vector. Destination Probabilities for the various xxxx sectors.

C.2.2.3 The Lost Firm Model (LOST.PRG)

Table C.9 In-files for the Lost firm Model

Name	Location	Description
Beta1039.dat	Emploc\database\lost	A (2 x1) column vector containing the estimated parameters for the regression model for the manufacturing industry
Beta4044.dat	Emploc\database\lost	A (2 x1) column vector containing the estimated parameters for the regression model for the construction industry
Beta5059.dat	Emploc\database\lost	A (2 x1) column vector containing the estimated parameters for the regression model for the manufacturing industry
Beta6069.dat	Emploc\database\lost	A (4 x1) column vector containing the estimated parameters for (SAR model) for the retail trade industry
Beta7099.dat	Emploc\database\lost	A (4 x1) column vector containing the estimated parameters for (SAR model) for the services industry
Sicxxxx.dat	Emploc\database\firms	(116 x 1) vector. The total number of firms at time t for each industry. xxxx refers to the different industries (see previous table).
Dumm6069.dat	Emploc\database\lost	(116 x 3) matrix. Each column represents a dummy variable that corresponds to downtown, malls, Northwest of Hamilton City respectively.
Dumm7099.dat	Emploc\database\lost	(116 x 3) matrix. Each column represents a dummy variable that corresponds to downtown, malls, Northwest of Hamilton City respectively.

Table C.9 (Continue)

Name	Location	Description
Weight1.dat	Emploc\database\lost	(116 x 116) matrix. The term rho x weighted matrix used in the SAR model. Rho here is the value estimated for the retail sector.
Weight2.dat	Emploc\database\lost	(116 x 116) matrix. The term rho x weighted matrix used in the SAR model. Rho here is the value estimated for the services sector.

Table C.10 Out-files for the Lost firm Model

Name	Location	Description
Lostxxxx.dat	Emploc\database\lost\result	(116x 1) vector. Number of lost firms in each zone for the various sectors.

C.2.2.4 The Linkage model (LINK.PRG)

Table C.11 In-files for the Linkage model

Name	Location	Description
Sicyyyy.dat	Emploc\database\firms	(116 x 1) vector. Number of firms at time t. These are given at time t and will be estimated at time t+1 for the modeled sectors.
Siczzzz.dat	Emploc\database\firms	(116 x 1) vector. Number of firms at time t. The files are given for the base case but should be supported for future projections
Avgemp.dat	Emploc\database\exog	(5 x 1) vector. Contains the average size of firms for the five modeled sectors
Empzzzz.dat	Emploc\database\exog	(116 x 1) vector. Contains the total employment in the zzzz sectors at time t+1. The user should support these exogenously.

yyyy= 1039 is Manufacturing, 4044 is Construction,5059 is Wholesale Trade,6069 is Retail Trade 7099 is Services
zzzz= 4547 is Transportation(exogenously determined), 4849 is communication (exogenously determined), 8184 is Government (exogenously determined) 85 is Education (exogenously determined)

Table C.12 Out-files for the Linkage Model

Name	Location	Description
EMPTOT.DAT	Emploc\change	(116 x 1) vector. Predicted total employment at the zonal level at time t+1.
Totxxxx.dat	Emploc\change	(116 x 1) vector. Predicted total firms at time t+1 for the modeled sectors

C.2.3 Model Directory Structure and Program Hierarchy

The model should be installed so that it has the directory structure shown in Figure C.7. Figure C.8 provides a flow chart that describes the hierarchy and function of each program in the model.

Figure C.7: Directory Structure of the model

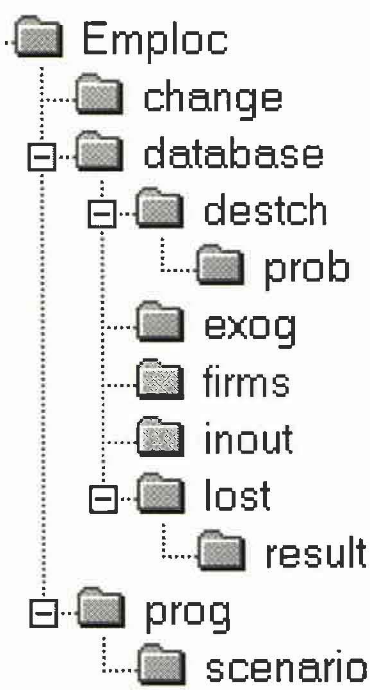
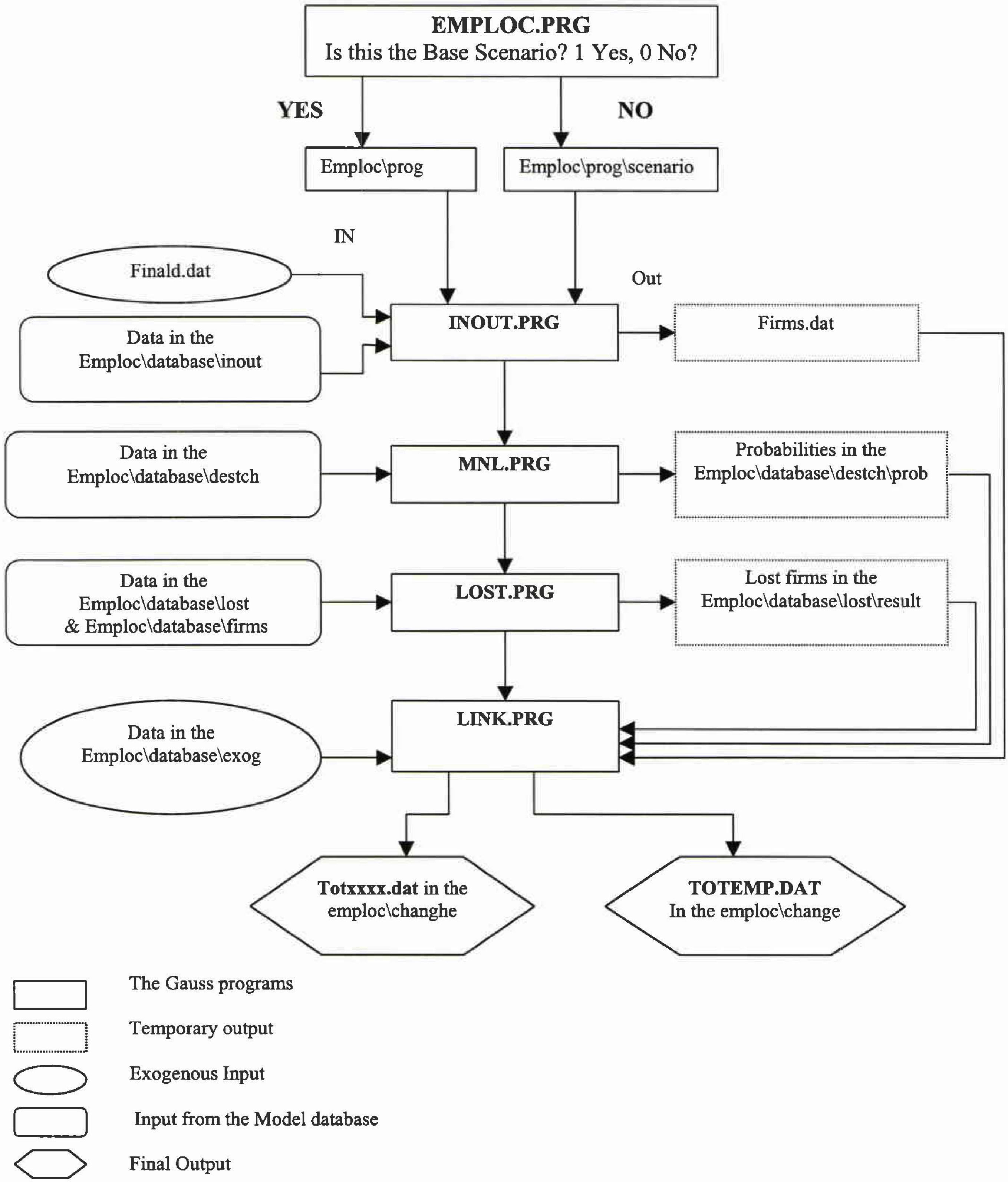


Figure C.8: Program Structure and Hierarchy



C.3 Model Scripts

C.3.1 EMPLOC.PRG

```
format /rd 8,0;
dos cls;
"*********************************************************************";
"  The Employment Location Model for the Hamilton-Wentworth Region    ";
"              EMPLOC PROGRAM (Version 1, February 1999)              ";
"                      Written by Hanna F. MAOH                       ";
"                   School of Geography and Geology                   ";
"               McMaster University, Hamilton, Ontario, Canada        ";
"*********************************************************************";

"*********************************************************************";
"If you are running scenarios, the final demand for the Input-Output  ";
" Model    ";
"should be saved under the path /EMPLOC/CHANGE/";

"";"";
"Now Model1 (the Input-Output Model) is running";

@*********************************************************************
Model 1: Input-Output Model.  This Model calculates the total number of
firms
at time t+1 at the regional Level.
*********************************************************************@

"Is this your Base Scenario, enter 1 if Yes, 0 if No";
ans=con(1,1);
h=ans;
if ans .ne 0;
"";"";
run d:\emploc\prog\inout.prg; @Call and run the Input-Output Model@
else;
run d:\emploc\prog\scenario\inout.prg; @Run the Input-Output Model for
Scenarios@
endif;
```

C.3.2 INOUT.PRG

```
format /rd 8,0;
dos cls;
"";"";
@*********************************************************************
The loading of data into the program: finald.dat, tech.dat, firmcoef.
For the first simulation period (1990-1997) the original files are read
from
Directory inout.  If the user wants to change the final demand he/she
must
save it in the scenario directory under the same name.  The files in the
inout
directory are:
1)finald.dat-The final demand for each sector of the economy
2)tech.dat-The technical coefficient matrix
3)firmcoef-Firm Coefficient Matrix
The output file is the total number of firms in each sector. This will
be saved
in the change directory under the name firms.dat
*********************************************************************@

load fnld[8,1]=d:\EMPLOC\database\exog\finald.dat;

load dat[8,8]=d:\EMPLOC\database\inout\tech.dat;
load fcoef[8,8]=d:\EMPLOC\database\inout\firmcoef.dat;
i=eye(8);
Leo=i-dat;
D=det(Leo);
If D .eq 0;
"You can not get an inverse of a singular matrix, please check your data
and";
"try again";
else;
LEOINV=inv(Leo);
gout= LEOINV*fnld;
firm=fcoef*gout;
screen off;
output file=D:\EMPLOC\change\firms.dat reset;
print firm;
output off;
endif;

"";"";
screen on;
"";"";
"*********************************************************************";
"Calculations of Modell is completed";
"*********************************************************************";
"";"";
"*********************************************************************";
```

```
"Now Running the Destination Choice Model";
"";"";
run d:\emploc\prog\MNL.prg; @Call and run the destination Choice model@
```

C3.3 MNL.PRG

```
format /rd 8,5;
dos cls;
"";"";
"**********************************************************************";
@EMPLOC PROGRAM (Version 1, February 1999), written by Hanna F. MAOH,
School of Geography and Geology, McMaster University, Hamilton, Ontario,
Canada @
"";"";
"The Model is now calculating the Choice probabilities              ";
screen off;
@*********************************************************************
Loading of data into
program:cbddist.dat,hhlddens.dat,hwypro.dat,mallpro.dat,
popsize.dat,mlanduse.dat, mctwemp.dat, rsemp.dat, servemp.dat,
rint1.dat, rint2.dat, rint3.dat, sint1.dat, sint2.dat.  These files
contain information needed to calculate the destination choice
probabilities.  All these files are save under the diretcory
(database/destch/).  The definition of each file is:

1)cbddist.dat-The linear distance between the centroid of each zone and
the CBD
2)hhlddens.dat-The household density in each zone
3)hwypro.dat-Highway proximity (1 if highway intersect or pass by a
zone, 0
otherwise
4)mallpro.dat-Malls proximity (1 if a mall is a zone or close to other
zones, 0 otherwise.
5)popsize.dat-Population size in each zone
6)mlanduse.dat-Dummy variable hold the value 1 if major proportion of
the zones land is manufacturing, 0 otherwise
7)mctwemp.dat-Accessibility to employment in manufacturing,
construction, transportation and Whole Sale trade.  To capture the
effect of agglomeration of economies
8)rsemp.dat-Accessibility to employment in Retail and Services
9)servemp.dat-Accessibility to employmenmt in Services
10)rint1.dat-Alternative specific variable reflecting particular types
of firms in retailing.  This is use to form interaction term with
popsize
11)rint2.dat-Alternative specific variable reflecting particular types
of firms in retailing.  This is use to form interaction term with
popsize
```

```
12)rint3.dat-Alternative specific variable reflecting particular types
of firms in retailing.  This is use to form interaction term with mall
proximity
13)sint1.dat-Alternative specific variable reflecting particular types
of firms in services.  This is use to form interaction term with
log(cbddist)
14)sint2.dat-Alternative specific variable reflecting particular types
of firms in services.  This is use to form interaction term with popsize
***********************************************************************
Also the estimated parameters and constants for each sector is saved
under:
beta1039.dat, cnst1039.dat, beta4044.dat, cnst4044.dat, beta5059.dat,
cnst5059.dat, beta6069.dat, cnst6069.dat, beta7099.dat, cnst7099.dat.
These are the betas and constants in the five sectors to be modeled

Note: The result of this program is saved under the path destch/prob
**********************************************************************@

load cbdpro[116,2]=d:\EMPLOC\database\destch\cbddist.dat;
load hhlddens[116,1]=d:\EMPLOC\database\destch\hhlddens.dat;
load hwypro[116,1]=d:\EMPLOC\database\destch\hwypro.dat;
load mallpro[116,1]=d:\EMPLOC\database\destch\mallpro.dat;
load popsize[116,1]=d:\EMPLOC\database\destch\popsize.dat;
load mlanduse[116,1]=d:\EMPLOC\database\destch\mlanduse.dat;
load mctwemp[116,1]=d:\EMPLOC\database\destch\mctwemp.dat;
load rsemp[116,1]=d:\EMPLOC\database\destch\rsemp.dat;
load servemp[116,1]=d:\EMPLOC\database\destch\servemp.dat;
load rint1[116,1]=d:\EMPLOC\database\destch\rint1.dat;
load rint2[116,1]=d:\EMPLOC\database\destch\rint2.dat;
load rint3[116,1]=d:\EMPLOC\database\destch\rint3.dat;
load sint1[116,1]=d:\EMPLOC\database\destch\sint1.dat;
load sint2[116,1]=d:\EMPLOC\database\destch\sint2.dat;


sec=1;
do while sec .le 5;


if sec .eq 1;
load beta[1,15]=d:\EMPLOC\database\destch\beta1039.dat;
load const[116,1]=d:\EMPLOC\database\destch\cnst1039.dat;

else;
  if sec .eq 2;
  load beta[1,15]=d:\EMPLOC\database\destch\beta4044.dat;
  load const[116,1]=d:\EMPLOC\database\destch\cnst4044.dat;

  else;
       if sec .eq 3;
       load beta[1,15]=d:\EMPLOC\database\destch\beta5059.dat;
       load const[116,1]=d:\EMPLOC\database\destch\cnst5059.dat;

       else;
            if sec .eq 4;
```

```
          load beta[1,15]=d:\EMPLOC\database\destch\beta6069.dat;
          load const[116,1]=d:\EMPLOC\database\destch\cnst6069.dat;

            else;
            if sec .eq 5;
            load beta[1,15]=d:\EMPLOC\database\destch\beta7099.dat;
            load const[116,1]=d:\EMPLOC\database\destch\cnst7099.dat;

              else;
            endif;

          endif;

      endif;

 endif;

endif;

@Calculating v-the systematic utility function for the 116 choices@

zrv=zeros(116,1);
h=1;
do while h .le 116;
t1=const[h,1]+beta[1,1]*cbdpro[h,1]+beta[1,2]*cbdpro[h,2]+beta[1,3]*hwyp
ro[h,1];
t2=beta[1,4]*mallpro[h,1]+beta[1,5]*popsize[h,1]+beta[1,6]*hhlddens[h,1]
;
t3=beta[1,7]*mctwemp[h,1]+beta[1,8]*rsemp[h,1]+beta[1,9]*servemp[h,1];
t4=beta[1,10]*mlanduse[h,1]+beta[1,11]*rint1[h,1]*popsize[h,1];
t5=beta[1,12]*rint2[h,1]*popsize[h,1]+beta[1,13]*rint3[h,1]*mallpro[h,1]
;
t6=beta[1,14]*sint1[h,1]*cbdpro[h,2]+beta[1,15]*rint2[h,1]*popsize[h,1];
v=t1+t2+t3+t4+t5+t6;
zrv[h,1]=zrv[h,1]+v;
h=h+1;
endo;


@Calculating the denominator of the logit formula@
sum=0;
n=1;
do while n .le 116;
sum=sum+exp(zrv[n,1]);
n=n+1;
endo;
@Calculating the logit formula P(i)@

zrp=zeros(116,1);
k=1;
do while k .le 116;
p=exp(zrv[k,1])/sum;
zrp[k,1]=zrp[k,1]+p;
k=k+1;
```

```
endo;

@SET Output files for the different industries@
screen off;
if sec .eq 1;
output file=d:\EMPLOC\database\destch\prob\prob1039.dat reset;
print zrp;
output off;
else;
     if sec .eq 2;
     output file=d:\EMPLOC\database\destch\prob\prob4044.dat reset;
     print zrp;
     output off;
     else;
          if sec .eq 3;
        output file=d:\EMPLOC\database\destch\prob\prob5059.dat reset;
        print zrp;
        output off;
        else;
             if sec .eq 4;
             output file=d:\EMPLOC\database\destch\prob\prob6069.dat
reset;
             print zrp;
             output off;
             else;
              if sec .eq 5;
              output file=d:\EMPLOC\database\destch\prob\prob7099.dat
reset;
              print zrp;
              output off;
              else;
              endif;
             endif;
       endif;
    endif;
endif;
sec = sec +1;
endo;
screen on;
"";"";
"*********************************************************************";
"Calculation of probabilities in Model 2 is Completed";
"*********************************************************************";
run d:\emploc\prog\lost.prg;@Now loading the lost.prg to calculate the
lost firms@
```

C3.4 LOST.PRG

```
format /rd 8,0;
dos cls;
"*********************************************************************";
@EMPLOC PROGRAM (Version 1, February 1999), written by Hanna F. MAOH,
School of Geography and Geology, McMaster University, Hamilton, Ontario,
Canada     @
"";"";
"The Model is now calculating the lost firms
";

@*********************************************************************
the loading of data into the program: sic1039.dat, sic4044.dat,
sic5059.dat,sic6069.dat,sic7099.dat.  These are the total firms at time
(t).  The regression model is based on the assumption that the lost
number of firms between time t and (t+1) is a function of the total
firms at time t, therefore, these are loaded from the directory
database\firms.

For Retail and Services an Simultaneous autoregressive model is used to
compute the number of lost firms, for Manufacturing, Construction and
Wholesale a simple regression model is used.

the files beta1039.dat,beta4044.dat,beta5059.dat,beta6069.dat
beta7099.dat are the estimated parameters for the different sectors.
dumm6069.dat and dumm7099.dat are dummy variables used in the retail and
services models.
weight1.dat and weigh2.dat are the rho * weight matrix.  This is given
for both retail and services respectively.

The result of this model is saved under lost\result with the name
lost1039.dat, lost4044.dat,lost5059.dat,lost6069.dat,lost7099.dat for
the five sectors, respectively
*********************************************************************@

Screen off;
l=1;

do while l .le 5; @Loading files for different sectors@


if l .eq 1;
load indp[116,1]=d:\emploc\database\firms\sic1039.dat;
load betas[2,1]=d:\emploc\database\lost\beta1039.dat;
else;
    if l.eq 2;
    load indp[116,1]=d:\emploc\database\firms\sic4044.dat;
    load betas[2,1]=d:\emploc\database\lost\beta4044.dat;
    else;
```

```
        if l.eq 3;
        load indp[116,1]=d:\emploc\database\firms\sic5059.dat;
        load betas[2,1]=d:\emploc\database\lost\beta5059.dat;
        else;
            if l .eq 4;
              load indp[116,1]=d:\emploc\database\firms\sic6069.dat;
            load dummy[116,3]=d:\emploc\database\lost\dumm6069.dat;
            load betas[5,1]=d:\emploc\database\lost\beta6069.dat;
              load pW[116,116]=d:\emploc\database\lost\weight1.dat;
              else;
              if l .eq 5;
              load indp[116,1]=d:\emploc\database\firms\sic7099.dat;
              load dummy[116,3]=d:\emploc\database\lost\dumm7099.dat;
              load betas[5,1]=d:\emploc\database\lost\beta7099.dat;
                  load pW[116,116]=d:\emploc\database\lost\weight2.dat;
              else;
              endif;
              endif;
        endif;
        endif;
endif;

if l .lt 4; @Simple Regression Model@

EX= zeros(116,2);
c=1;
do while c .le 116;
EX[c,1]=1;
EX[c,2]=indp[c,1];
c=c+1;
endo;
lost= EX * betas;
k=1;
do while k .le 116;
if lost[k,1] .lt 0;
lost[k,1]=0;
else;
endif;
k=k+1;
ENDO;

if l .eq 1;

output file=d:\emploc\database\lost\result\lost1039.dat reset;
print lost;
output off;
else;
    if l .eq 2;
    output file=d:\emploc\database\lost\result\lost4044.dat reset;
    print lost;
    output off;
    else;
      if l .eq 3;
      output file=d:\emploc\database\lost\result\lost5059.dat reset;
```

```
      print lost;
      output off;
      else;
      endif;
    endif;
endif;




else; @Calculating results for the SAR MODEL@

iden=eye(116);
i=iden - pW;
iv=inv(i);

EX= zeros(116,5);
c=1;
do while c .le 116;
EX[c,1]=1;
EX[c,2]=indp[c,1];
EX[c,3]=dummy[c,1]*indp[c,1];
EX[c,4]=dummy[c,2]*indp[c,1];
EX[c,5]=dummy[c,3]*indp[c,1];
c=c+1;
endo;

trend= EX * Betas;
crt=pW * trend;
rh=trend - crt;
lost = iv * rh;

k=1;
do while k .le 116;
if lost[k,1] .lt 0;
lost[k,1]=0;
else;
endif;
k=k+1;
ENDO;

if l .eq 4;
output file=d:\emploc\database\lost\result\lost6069.dat reset;
print lost;
output off;
else;
     if l .eq 5;
     output file=d:\emploc\database\lost\result\lost7099.dat reset;
     print lost;
     output off;
     else;
     endif;
endif;
```

```
endif;
l=l+1;
endo;
screen on;
"*********************************************************************";
"Calculation of Lost firms is completed";
"*********************************************************************";
"";"";
@Loading the link.prg to link the results of the previous three models@
run d:\emploc\prog\link.prg;
```

C3.5 LINK.PRG

```
format /rd 8,5;
dos cls;
"*********************************************************************";
@EMPLOC PROGRAM (Version 1, February 1999), written by Hanna F. MAOH,
School of Geography and Geology, McMaster University, Hamilton, Ontario,
Canada      @
"";"";
"The model is calculating the total firms in each sector at the zonal
level";
Screen off;
@*********************************************************************
the loading of data into the
program:lost1039.dat,lost4044.dat,lost5059.dat,
lost6069.dat,lost7099.dat.  Those are generated by the lost.prg program.

Since we have four sectors to be exogenous (Communication,
transportation, Education and Government, these should be supported for
the time period t+1 for the last two sectors to calculate the total of
firms in services (the modeled sectors).

The exogenous data is to be saved under the path (database\exog).

The destination probabilities generated from MNL.prg will also be loaded
to calculate the total number of firms that moved in a zone.  Those are
loaded from the path destch\prob.  Those are: prob1039.dat,prob4044.dat,
prob5059.dat, prob6069.dat and prob7099.dat

The average size of a firm is given exogenously to the model.  Files
contains this information should be saved under database\exog.  The
average size of employment for the modeled sectors should be given as a
column vector file of 5 x 1 in dimension.  Each row represent the
```

```
average number of employment in the sectors to be modeled These are
manufacturing, construction, wholesale, retail, services.
The file should be saved under database\exog and with the name
"avgemp.dat".

For the exogenous sectors (Transportation, Communication, Government,
and Education) the total employment at the zonal level at time t+1
should be supported to calculate total employment in a zone since it is
a mixture between those who are modeled and who will be determined
exogenously. The files should be saved under database\exog with the
names: emp4547.dat, emp4849.dat, emp8184.dat and emp85.dat

The total employment per zone will be saved under the EMPLOC\CHANGE
directory under the name TOTEMP.dat
*********************************************************************@

load firms[8,1]=d:\emploc\change\firms.dat;

@Calculate the remaining firms@
count=1;
do while count .le 5;
if count .eq 1;
load lost[116,1]=d:\emploc\database\lost\result\lost1039.dat;
load totfirm[116,1]=d:\emploc\database\firms\sic1039.dat;
else;
     if count .eq 2;
     load lost[116,1]=d:\emploc\database\lost\result\lost4044.dat;
     load totfirm[116,1]=d:\emploc\database\firms\sic4044.dat;
     else;
        if count .eq 3;
        load lost[116,1]=d:\emploc\database\lost\result\lost5059.dat;
        load totfirm[116,1]=d:\emploc\database\firms\sic5059.dat;
        else;
            if count .eq 4;
            load
lost[116,1]=d:\emploc\database\lost\result\lost6069.dat;
            load totfirm[116,1]=d:\emploc\database\firms\sic6069.dat;
            else;
             if count .eq 5;
             load
lost[116,1]=d:\emploc\database\lost\result\lost7099.dat;
             load totfirm[116,1]=d:\emploc\database\firms\sic7099.dat;
             load sic8084[117,1]=d:\emploc\database\exog\sic8084.dat;
             load sic85[117,1]=d:\emploc\database\exog\sic85.dat;
             else;
             endif;
            endif;
        endif;
     endif;
endif;
remain=totfirm - lost;

@Aggregating the total firms for the base year 1990@
sum1=0;
```

```
M=1;
do while M .le 116;
sum1 = sum1 + totfirm[M,1];
M=M+1;
endo;

@Aggregating the lost firms. This is done to link the model to the
Input-Output Model@

sum2=0;
M=1;
do while M .le 116;
sum2 = sum2 + lost[M,1];
M=M+1;
endo;

if count .eq 1;
nrfirm1=firms[1,1]-sum1+sum2;
remain1=remain;
else;
     if count .eq 2;
     nrfirm2=firms[2,1]-sum1+sum2;
     remain2=remain;
     else;
        if count .eq 3;
        nrfirm3=firms[5,1]-sum1+sum2;
        remain3=remain;
        else;
             if count .eq 4;
             nrfirm4=firms[6,1]-sum1+sum2;
             remain4=remain;
             else;
                if count .eq 5;
                nrfirm5=firms[7,1]-sic8084[117,1]-sic85[117,1]-sum1+sum2;
                remain5=remain;
                else;
                endif;
            endif;
       endif;
   endif;
endif;
count=count +1;
endo;
@Determing total new and relocating from nrtot matrix and the
probablities
in MNL model@
load prob1039[116,1]=d:\emploc\database\destch\prob\prob1039.dat;
load prob4044[116,1]=d:\emploc\database\destch\prob\prob4044.dat;
load prob5059[116,1]=d:\emploc\database\destch\prob\prob5059.dat;
load prob6069[116,1]=d:\emploc\database\destch\prob\prob6069.dat;
load prob7099[116,1]=d:\emploc\database\destch\prob\prob7099.dat;
load avgemp[5,1]=d:\emploc\database\exog\avgemp.dat;
load emp4547[116,1]=d:\emploc\database\exog\emp4547.dat;
load emp4849[116,1]=d:\emploc\database\exog\emp4849.dat;
```

```
load emp8184[116,1]=d:\emploc\database\exog\emp8184.dat;
load emp85[116,1]=d:\emploc\database\exog\emp85.dat;

nr1039=nrfirm1*prob1039;
nr4044=nrfirm2*prob4044;
nr5059=nrfirm3*prob5059;
nr6069=nrfirm4*prob6069;
nr7099=nrfirm5*prob7099;
tot1039=remain1+nr1039;
tot4044=remain2+nr4044;
tot5059=remain3+nr5059;
tot6069=remain4+nr6069;
tot7099=remain5+nr7099;
emp1039=tot1039*avgemp[1,1];
emp4044=tot4044*avgemp[2,1];
emp5059=tot5059*avgemp[3,1];
emp6069=tot6069*avgemp[4,1];
emp7099=tot7099*avgemp[5,1];
TOTEMP=emp1039+emp4044+emp5059+emp6069+emp7099+emp4547+emp4849+emp8184
+emp85;
format /rd 8,0;
output file=d:\emploc\change\tot1039.dat reset;
print tot1039;
output off;

output file=d:\emploc\change\tot4044.dat reset;
print tot4044;
output off;

output file=d:\emploc\change\tot5059.dat reset;
print tot5059;
output off;

output file=d:\emploc\change\tot6069.dat reset;
print tot6069;
output off;

output file=d:\emploc\change\tot7099.dat reset;
print tot7099;
output off;

output file=d:\emploc\change\totemp.dat reset;
print TOTEMP;
output off;

screen on;
"";"";
"**********************************************************************";
"Simulation is completed.  Please check the (change) directory ";
"for results ";
"The Total employment in each zone is saved under the name TOTEMP.DAT ";
"**********************************************************************";
```

REFERENCES

Anderson, W., Kanaroglou, P. and Miller, E. (1994). "Integrated Land Use and Transportation Models for Energy and Environmental Analysis: A report on Design and Implementation" Department of Geography, McMaster University, Ontario, Canada

Anderson, W., Kanaroglou, P. and Miller, E. (1996). "Urban Forms, Energy and The Environment: A Review of Issues, evidence and Policy", *Urban Studies*, 33, pp. 7 - 35

Bailey, T. and Gatrell, A. (1995). "*Interactive Spatial Data Analysis*", Longman Scientific and Technical, New York.

Ben-Akiva, M. and Lerman, S. (1985). "*Discrete choice analysis: Theory and Application to Travel Demand*". MIT, ed. M. Manheim, Cambridge.

Bourne, L.S. (1991). "Addressing the Canadian City: Contemporary Perspectives, Trends, and Issues", in *Canadian Cities in Transition 25 - 44 T.* Bunting and P. Filion, editors. Toronto: Oxford University Press

Clapp, J M. (1980). "The Intrametropolitan Location of Office Activities." *Journal of Regional Science*, 20, pp. 387 - 399

Coffey, W., Drolet, R., and Polese, M. (1996). "The Intrametropolitan Location Of High Order Services: Patterns, Factors and Mobility in Montreal", *Papers in Regional Science*, 75 (3): pp. 293 - 323

Daganzo, C. (1979). "*Multinomial Probit: The Theory and Its Application to Demand Forecasting.*", Academic Press, New York

Danials, P. W. (1982). "*Service Industries: Growth and Location*" Cambridge University Press

Darovny, M. (1998). "*Residential Development in Hamilton-Wentworth, 1980 to 1994*", M.A Dissertation Thesis, School of Geography and Geology, McMaster University, Ontario, Canada

De La Barra and Vera (1984). "TRANUS-J: Putting Large Models into Small Computers", *Environmental and Planning B: Planning and Design*, 11, pp. 87 - 101

De La Barra (1989). "*Integrated Land Use and Transport Modeling*", Cambridge University.

Erickson and Wasylenko (1980). "Firm Relocation and Site Selection in Suburban Municipalities", *Journal of Urban Economics*, 8, pp. 69 - 85

Ferguson M. (1995). "*The Aggregated Spatial Logit Model*", Phd Dissertation Thesis, School of Geography and Geology, McMaster University, Ontario, Canada

Gottlieb, D. (1995). "Residential Amenities, Firm Location and Economic Development", *Urban Studies*, 32, (9), pp. 1413 - 1436

Green W. (1993). "*Econometric Analysis*", 2nd Edition, Macmillan Publishing Company, New York

Hansen E. (1987). "Industrial Location Choice In SAO Paulo, Brazil, A Nested Logit Model", *Regional and Urban Economics*, 17, pp. 89 - 108

Hanson, S. (1986). "*The Geography of Urban Transportation*", The Guilford Press, New York, London

Hunt, J.D and Simmonds, D.C (1993). "Theory and application of an integrated land use and transportation modeling framework" *Environment and Planning*, 20B, pp. 221 - 244

Ihlanfeldt K.R and Raper M.D. (1990). "The Intrametropolitan Location of New offices Firms", *Land Economics*, 66 (2), pp. 182 - 198

Jones, K., Kamikihara, S., and Simmons, J. (1998). "Dallas-Fort Worth: Commercial Structure and Change", *Progress in Planning*, 50 (4), pp. 273 - 289

Kanaroglou, P., Anderson, W. and Kazakov, A. (1998). "Economic Impacts of Highway Infrastructure Improvements, Part 1. Conceptual Framework", *Journal of Transport Geography*, 6, (3), pp. 203 - 210

Lea A. (1989). "An Overview of Formal Methods for Retailing Site Evaluation and Sales Forecasting: Part 1.", *The Operational Geographer*, 7 (2), pp. 8 - 17

Lee K. (1982). "A Model of Intraurban Employment Location: An Application to Bogota, Colombia", *Journal of Urban Economics*, 12, pp. 263 - 279

Lee K. (1985). "Decentralization Trends of Employment Location and Spatial Policies in LDC Cities", *Urban Studies*, 22, pp. 151 - 162

Mathsoft. (1996) "*S+SPATIALSTATS Users Manual*", version 1.0, Seattle: MathSoft, Inc.

McFadden, D. (1975). "Aggregate Travel Demand Forecasting from Disaggregated Behavioral Models.", *Trans Research Records*, 534, pp. 24 - 37

Miller R. E. and Blair P. D. (1985). "*Input-Output Analysis - Foundations and Extensions*" Prentice-Hall, Inc., Englewood Cliffs, New Jersey

Miron, J. R (1982). "Economic Equilibrium in Urban Land Use" In *Internal Structure of the City - Readings on Urban forms, Growth and Policy*, 2nd Edition, Larry S. Bourne, Editor.

Pascal and McCall (1980). "Agglomeration Economies, Search Cost and Industrial Location", *Journal of Urban Economics*, 8, pp. 383 - 388

Peddle M. (1990) "Industrial Park Location: Do Firms Characteristics Matter?" *Regional Science Perspective*, 20 (2), pp. 26 - 35

Putman, S (1990). "*Integrated Urban Models 2*", Pion Limited, London

Hamilton-Wentworth Planning and Development Department (1992). "*Hamilton-Wentworth Employment Trends, 1982 to 1990*", Hamilton: Planning and Development department, Regional Municipality of Hamilton-Wentworth

Selting, A., Allanach, C., and Loveridge, S. (1995). "The Economics of Agglomeration and Firm Location", *The Great Lakes Geographer*, 2 (1), pp. 17-29

Shukla and Waddle (1991). "Firm Location and Land Use in Discrete Urban Space - A Study of the Spatial Structure of Dallas-Fort Worth", *Regional and Urban Economics*, 21, pp. 225 - 253

Simmons, J. and Yeates, M. (1998). "Toronto: Commercial Stricture and Change", *Progress in Planning*, 50 (4), pp. 253 - 272

Southworth, F. (1995). "A Technical Review of Urban Land Use- Transportation Models as Tools for Evaluating Vehicle Travel Reduction Strategies", The office of Environmental Analysis and Sustainable Development, U.S Department of Energy.

Statistics Canada (1991). Census Data, Ottawa

Statistics Canada (1996). Census Data, Ottawa

Statistics Canada (1990). Transaction Tables for Province of Ontario, Ottawa.

Statistics Canada (1989). "The Input-Output Structure of the Canadian Economy", Cat. No. 15 - 201

Storey, D. J (1990). "Firm Performance and Size", In *The Economics of small firms - A European Challenge*, Zoltan J. Acs and David B. Audretsch, Editors

Tervo, H. and Niittykangas, H. (1994). "Developments in The Growth, Location and Turbulence of Business Services in Finland", *Papers in Regional Science*, 73 (1), pp. 73 - 95

Toyomane, N. (1988). "*Multiregional Input-Output Models in Long-Run Simulation*", Kluwer Academic Publishers, Dordrecht, Boston, Lancaster.

Watts (1990). "Plant Closure in Urban Areas: Towards a Local Policy Response", *Urban Studies*, 28 (5), pp. 803 - 817

Wegener M. (1986). "Transportation Network Equilibrium and Regional Deconcentration", *Environment and Planning A*, 19, pp. 437 - 456

Yarish M. (1998). "*The Intrametropolitan Location of New Office Firms in Metropolitan Toronto*", M.A Dissertation Thesis, School of Geography and Geology, McMaster University, Ontario, Canada

Yeates, M. (1990). "*The North American City*", 4th Edition, New York: Harper and Row

Vahaly, J. (1976). "The location of Services and Office Activities in Nashville-Davidson County, 1970", *Land Economics*, 52 (4), pp. 479 - 492

Wheaton, C. and Shishido, H. (1981). "Urban Concentration, Agglomeration Economies, and the Level of Economic Development", Economic Development and Culture Change, 30(1), pp. 17 - 30

Xu, P., Tsanis, I., Anderson, W., and Kanaroglou, P. (1994). "An Economic Input-Output Analysis for Urban Storm water Quality Planning", *Water Resource Management*, 8, pp. 155 - 170