2016

MODELLING THE DETERMINANTS OF COMMERCIAL VEHICLE OWNERSHIP: AN APPLICATION TO WINDSOR, ONTARIO, CANADA

Aya Hagag
University of Windsor

MODELLING THE DETERMINANTS OF COMMERCIAL VEHICLE OWNERSHIP: AN APPLICATION TO WINDSOR, ONTARIO, CANADA

By

Aya Hagag

A Thesis
Submitted to the Faculty of Graduate Studies
through the Department of Civil and Environmental Engineering
in Partial Fulfillment of the Requirements for
the Degree of Master of Applied Science
at the University of Windsor

Windsor, Ontario, Canada

2016

MODELLING THE DETERMINANTS OF COMMERCIAL VEHICLE OWNERSHIP: AN APPLICATION TO WINDSOR, ONTARIO, CANADA

by

Aya Hagag

APPROVED BY:

____

W. Anderson, Outside Program Reader
Department of Political Science

____

C. Lee, Department Reader
Department of Civil and Environmental Engineering

____

J. Tofflemire, Special Member
Department of Civil and Environmental Engineering

____

H. Maoh, Advisor
Department of Civil and Environmental Engineering

May 10, 2016

DECLARATION OF ORIGINALITY

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I declare that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

The majority of the existing transportation literature has been concerned with private travel activities. Fortunately, the importance of commercial vehicles and their movement has led to a surge of research activities to analyze and understand commercial vehicle movements. However, an absent theme in the emerging research activities is the process governing commercial vehicle ownership. This project attempts to fill the gap in the existing literature by developing a new model of commercial vehicle ownership. The focus is on studying the spatial prevalence of the various types of commercial vehicles, as derived from the Gross Vehicle Weight (GVW) classes, in a given traffic analysis zone (TAZ) within an urban area using various types of statistical methods. The results allow us to unravel the significant factors explaining the variability in the spatial distribution of commercial vehicles. The obtained statistical results form the basis for developing predictive urban commercial vehicle ownership location models.

ACKNOWLEDGEMENTS

First and foremost, I am grateful for the chance Dr. Hanna Maoh gave me to start my MASc under his supervision. I am thankful for Dr. Maoh's incredible amount of trust, encouragement and support throughout my graduate studies. His guidance both challenged and encouraged me at the same time which positively shaped and improved my graduate experience. Our enlightening conversations about transportation helped me think more conceptually, critically and quantitatively for which I am grateful. In addition, I am extremely thankful for the extra time he exerted to help me prepare and submit conference papers, presentations, and scholarships.

I would also like to extend a great thank you to my committee members Dr. Chris Lee, Dr. William Anderson and Mr. John Tofflemire for taking the time to review my work and provide insightful comments which ultimately strengthened this thesis. Also, thank you to Dr. Faouzi Ghrib for taking the time to chair my final defense meeting.

Additionally, I would like to acknowledge Georgina Madar, who developed the Business Establishments Commercial Travel Survey, without which this project would have been significantly more difficult to complete. I especially thank Mr. Shakil Khan for all his help and expertise in managing the TSI lab and my workstation.

To all my fellow lab mates, it was a pleasure working along such a great group of highly talented people. I am extremely grateful for Rahaf Husein and Terence Dimatulac for their continuous support and encouragement throughout my work.

I am thankful to the Natural Science and Engineering Research Council of Canada (NSERC) for supporting this research through a Discovery Research Grant. Also, this research was enabled through a Canada Foundation of Innovation (CFI) Infrastructure Research Grant.

Last, and certainly not least, I am especially thankful and am indebted to my family for their constant encouragement, for their love, support and for always being there for me - this thesis would not have been possible without you all. You have always believed that I can pursue and achieve my goals with great success.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 Overview

Freight movement is a major factor contributing to economic growth and development. In 2014, the transportation services sector represented 4.2 percent of Canada's GDP (Transport Canada, 2012). Freight movement originates from the distribution of raw materials to final products and services from producers, wholesalers and distribution centers to consumers at multiple spatial scales. This includes movements at the international, national, regional and urban level. In 2014, Canada was the 13th largest exporter and 12th largest importer in the world, where the United States ranks as its largest and most important trading partner (Central Intelligence Agency, 2013). Canada's freight is transported through multi-modal networks involving waterways, railways, highways, air-ways, and intermodal facilities. However, road transportation is the most dominant mode for freight movement in terms of value transported (Transport Canada, 2012). This dependency on road-based freight transport raises the need to understand commercial vehicle transportation activity to help planners and policy makers accommodate the travel demand needs for commercial vehicles along with private vehicle on the same road network.

In general, freight transportation is a very complex system unlike passenger transportation. Indeed, freight movement is a dynamic process that results from diverse interactions among many stakeholders that belong to heterogonous industries (Tavasszy et al., 1998). Stakeholders involved are classified either as public or private decision makers who generally share the same objective which is to transport and deliver goods and/or services. Also, commercial vehicle movements tend to vary on a daily basis and from season to season as a result of the rapid changes in the supply chain structures, logistics and technological advancements. For example, Just-in-time (JIT) delivery makes studying freight transport a complex process. On the contrary, private passenger vehicles tend to follow the same pattern for their main activities like driving to work, school or limited recreational locations.

Furthermore, particularly in urban areas, freight transportation exhibits some unique features. Among them, commercial vehicle fleets are composed of a wide range of vehicle sizes spanning from small sized vehicles to large multi-unit trucks (Hunt & Stefan, 2007). Unlike passenger vehicles, commercial vehicles have a higher impact on the performance of the transportation system. More specifically, commercial vehicles on average have significantly larger gross vehicle weight (GVW), and as such contribute to the deterioration of pavement and air quality. Another unique feature is tour chaining where commercial vehicle movements are not as simple as a two legs trip but rather a tour with multiple trips. For example, the vehicle starts at the establishment and in many cases makes several stops for different purposes such as delivering goods or services before returning to the establishment.

Initially, the conventional four-step approach, primarily designed for passenger transport modeling, was the most commonly used method by many practitioners and decision makers to model commercial vehicle movement. The main reasons for such practice are the simplicity of applying the model and the relatively inexpensive and low effort in data collection. However, the complexities of commercial vehicle movements, as mentioned earlier, have led to a steady methodological evolution that sought to overcome the known deficiencies inherited in the four-step approach. There has been a shift from the conventional approach to disaggregate models, for their ability to capture the fundamental characteristics of freight movement. For example, the supply-chain approach, explicitly accounts for the different component of the supply chain and their behavior. Also, such technique could be calibrated to capture new shipping behaviours such as the adoption of outsourcing, e-commerce, and JIT delivery systems. Moreover, the tour-based approach is able to represent commercial vehicle activities and activity chains using data collected in surveys or based on actual observed GPS data. It is worth noting that obtaining a representative sample to develop such models is hard and time demanding. As a result, in many cases, the available data eventually determines the appropriate technique to be used to study a specific aspect in commercial vehicle movement.

1.2 Statement of the Problem

In spite of the progress that commercial vehicle movement research has made in recent years, the state-of-the-art in freight modeling is far behind when compared to the work done on passenger vehicles (Samimi et al., 2012). For instance, the process governing commercial vehicle ownership is among the areas that are lacking within the freight modeling research. Therefore, this project intends to fill this gap by studying the determinants of commercial vehicle ownership with an application to Windsor Census Metropolitan Area (CMA). Our efforts will focus on studying the spatial distribution of commercial vehicles. This particularly important since the spatial distribution of where these vehicles are housed by their respective establishments (similar to private vehicles) is a key determinant of the level of traffic volume observed on an urban road network. In fact, the work presented by Madar (2014) confirms that the number of owned commercial vehicles is a significant factor that explains the number of generated commercial trips. Therefore, there is a need to identify the factors that explain the prevalence of specific types of commercial vehicles at a certain location in a city to help devise more effective travel demand models.

One of the main reasons for the underdevelopment of commercial vehicle ownership models in the literature is due to the lack of detailed commercial vehicle travel data. According to the literature, most of the existing efforts to collect detailed commercial travel data resulted in a low response rate (Samimi et al., 2012). Private establishments usually vacillate to share information related to their business freight and/or transportation activities. Therefore, this research also addresses the problem of data scarcity by employing synthetic population techniques to microsimulate the number of commercial vehicles owned by all individual business establishments that engage in delivering goods or services in the Windsor CMA. The case presented here uses the combinatorial optimization technique (CO) to synthesize the number of commercial vehicles owned by business establishments that engage in commercial travel activities (i.e. delivering goods and/or services).

The offered procedure to microsimulate the spatial distribution of commercial vehicles within an urban area is novel and has not been attempted in the past. Such analysis provides a basis for evaluating commercial data sources such as the R. L. Polk and Co. vehicle registry data and its potential in analyzing commercial vehicles in other areas where micro-datasets are not available. Accordingly, the acquired Polk records are used to identify the locational determinants which explain the spatial prevalence of specific types of commercial vehicles in a given census tract (zone). The choice decision is handled using the logit model modeling framework (i.e. Multinomial Logit and Mixed Logit) and the count models (i.e. Poisson and Negative Binomial Regression).

1.3 Objectives

The primary objectives of this project are:

1) Advance the current state of knowledge on urban commercial vehicle movement transportation research by focusing on commercial vehicle ownership and its spatial distribution
2) Determine the factors that lead to the prevalence of commercial vehicles by their GVW at the traffic analysis zoning (TAZ) level
3) Provide the basis for commercial vehicle ownership forecasting through the use of different statistical methods and techniques
4) Provide the foundation for performing micro-level future travel demand forecasting with the help of data synthesis techniques

1.4 Thesis Outline

The remainder of this thesis is organized as follows. The next chapter provides an overview of previous studies regarding commercial vehicle movement models, vehicle ownership models and population synthesis techniques. Chapter 3 provides a clear description of the study area, the different data sources and data treatment required to generate the synthetic population and the statistical models. Then chapter 4 highlights method of analysis including a description of the general means through which the goals of this research is achieved. The results from the synthetic population and the performed statistical analyses are presented in Chapter 5. This will be followed by a final chapter that provides a conclusion of the results achieved, contributions and policy implications, the limitation of the work conducted, and some considerations for future research. A list of references follows Chapter 6, along with appendices containing additional information.

CHAPTER 2: LITERATURE REVIEW

2.1 Commercial Vehicle Movement

Despite the importance of commercial vehicle movement and their impact on the transportation network, the topic has not received the same attention as passenger vehicle movement (Zhou & Dai, 2012). Fortunately, the ability to collect detailed information in recent years has led to a surge of studies on commercial vehicle movement. Yet, there is a need for more work on the topic. The development of effective models requires a thorough understanding of commercial vehicle movement. To gain such understanding researchers have applied a wide variety of methodological techniques over the years.

Reviews found in freight literature employ different frameworks to classify commercial vehicle movement models. For example, Chow et al. (2010) provide a review of different freight forecasting models with respect to their data requirement, model development and the objectives that could be achieved using each model. Similarly, Zhou and Dai (2012) review existing freight models with respect to major data requirements, procedures, techniques and real-world application cases. Alternatively, Anand et al. (2012), follow a framework that classifies the available urban goods movement models based on the stakeholders' involvement, the objective of modeling, the descriptors for modeling purpose and the view point for achieving the objective. Other reviews only focus on models developed at the urban and metropolitan level (Anand et al., 2012; Regan & Garrido, 2001; Zhou & Dai, 2012) or only at the national and international levels (Yang et al., 2010; De Jong et al., 2004)

The intent of this section is to provide a review of some modeling efforts in freight demand modeling by underlining the methods used, the nature of the data required and the limitations of each approach in terms of its strengths and weaknesses. The following is organized on the basis of three basic approaches: 1) Conventional Approaches, 2) Supply-Chain Approaches and 3) Simulation (Tour-Based) Approaches.

2.1.1 Conventional Approaches

Conventional approaches usually starts with a specified formulation or a defined model which is applied and adapted to accessible data. This category includes: 1) Factoring methods, 2) Truck or (Trip-based or Vehicle) method, 3) Commodity-based method and 4) Input-Output method. Most of the aforementioned techniques generally apply one or more element of the urban four-step travel demand forecast model (NCHRP 606, 2008). Applying the urban four-step modeling process to freight modeling was an ordinary transition since this method is well established in modeling private vehicles movement. This model is a sequence of four sub-models: trip generation, trip distribution, mode split and network assignment.

Assuming that the study area is divided into different geographic zones, the first sub-model trip generation, estimates the total number of production and attraction freight movement trips starting or ending in each zone most often as a function of employment or establishment size (NCHRP 606, 2008). The second sub-model, trip distribution, estimates the freight flows between all the geographic zones. For this step the most commonly used methods is the gravity model as a function of generation and attraction factors of the origin and destination zones weighted by an impedance term that represents transportation costs between zone pairs (De Jong et al., 2004). The third sub-model, mode split, designates the estimated freight flows for each distinct mode. Usually discrete choice models are employed either at the aggregate or the disaggregate level to estimate modal split usually as a function of cost and commodity classifications (De Jong et al., 2004). Finally, the fourth sub-modal, traffic assignment, assigns estimated flows by mode to individual links of the network. This process can be rule-based assignment, freight truck only network assignment or multiclass network assignment (NCHRP 606, 2008).

Factoring method includes different ways to estimate truck flows or commodity flows. This includes (i) direct facility flow factor method and (ii) Origin – Destination (O-D) factoring method. The direct facility flow factor method forecasts future flows based on existing base year data. This approach could be used either to estimate future flow on a facility by applying growth factors to the flow on that facility or by applying factors that account for the diversion of flow from that facility to other routes or modes. This approach relies on time series analysis and economic analysis to forecast flows based on historical data and change in the level of economic activities (Yang et al., 2010; NCHRP 606, 2008).

Alternatively, the O-D factoring method uses existing O-D freight flow data to forecast future estimates of flow either using the Fratar expansion technique or the entropy maximization mathematical programming process (NCHRP 606, 2008; Stefan & Hunt, 2004). The former is an iterative proportional fitting technique applied to freight tables to predict freight flows based on predicted production and attraction growth rates. Afterwards, the factored O-D matrix is used for mode split and traffic assignment model components.

Truck-based model in essence is a three-step model; trip generation, trip distribution, and traffic assignment, rather than a four-step model since it only addresses trucks. Consequently, the model estimates aggregate truck flows and assign them to links of the road network. Alternatively, other models incorporate different truck classes such as light, medium and heavy trucks based on gross vehicle weight within the trip generation component (NCHRP 606, 2008). Such models use the trip generation and distribution component to produce flows by vehicle class and then assign flows to the road network.

In the literature, Commodity-based models are also named the four-step commodity model since they utilize the same model structure as the passenger forecasting models. In the first stage freight flow production and attraction are estimated in ton or dollars at a predefined geographic level. In the second stage, the flows between distinct zones within the study area are estimated. Although, trucks are considered the dominant mode for commodity transportation, the third stage which is modal split is applied to assign commodity flows to different available modes. Finally, the fourth stage handles the flows to be assigned to the roadway network by estimating traffic volumes on routes between different zones for specific time frames.

Input-output economic model generates commodity flows and truck flows O-D directly from land use activities and zonal data. This model focuses on the layer of trade by incorporating input-output (I/O) and land use–transport interaction models to explain the interaction between trade, transport and the economy (Anon, 2013). Accordingly, changes in economic activities and land-use patterns influence the extent and distribution of freight flows on the transportation network and corresponding transportation performance. Cases studies for each of the model classes defined under the conventional approaches can be found in the statewide freight forecasting toolkit (NCHRP 606, 2008).

Conventional approaches are comparatively straightforward and thus are easy to perform and can be relatively inexpensive. The sources of information needed to formulate such models are usually surveys or roadside counts. The former are surveys conducted to gather general information about shippers and carriers within the study area such as their current address and the total number of employees. Also information collected include the number of trips, the time of each trip, the origin and destination of each trip, the type of commodity or service shipped, the cost of the shipment/service, the weight of the shipment, the distance travelled, the mode choice, the route choice and other general information (NCHRP 606, 2008; Chow et al., 2010; Regan & Garrido, 2001; Gonzalez-Feliu & Routhier, 2012). On the other hand, roadside counts only provide information such as vehicle count, type, speed and weight but not trip purpose, routing, duration etc. This process is challenging since it is hard to determine which vehicles are truly commercial vehicles (Stefan & Hunt, 2004).

These data sources thus provide the data needed to establish the trip generation, trip distribution, mode split, and traffic assignment. However, often times, the data collected to study commercial vehicle movement are limited due to the low response of establishments, where they are hesitant to share information regarding their firms and/or commercial vehicle activities. Other data sources that are used are the Census Data and Network Data. The former is a data source where information about the population, employment and other socioeconomic factors are available. The latter include network physical information such as the segment capacity, volume, free flow speed, and travel time for the all the existing networks within a study area.

Although these models are accepted as useful toolkits and have been widely used, they are still limited. The factoring methods, for instance, can only be applied to forecast freight flow for the near future for a small study area. Also, such technique lacks the response to policy changes. Moreover, the structures of the four-step freight travel demand models significantly limit their ability to address some of the challenges to urban freight research (Anon, 2013). For example, vehicle-based models are usually formulated as a function of zonal employment or firm's size and therefore insensitive to the true economic behavior of commodity movement. Similarly, commodity-based models have limited ability to address the actual trip chain of commodity flows, services trips, local pickup and delivery trips, truck trips with less-than-a-truck loads, and empty truck trips (Zhou & Dai, 2012). Therefore, such techniques are widely used to simulate zone-to-zone commodity flow data rather than analyzing public policies.

2.1.2 Supply-Chain (Logistics) Approach

Freight transportation is, in nature, a very complex system unlike passenger transportation. Indeed, freight movement is a dynamic process that results from diverse and complicated interactions among many stakeholders. This dynamic process includes the economic activities underlying the complete supply chain starting from the movement of raw goods to the production, sales and sourcing, inventory and distribution of finished products (Tavasszy et al., 1998). In general, the logistical organization of a firm can be divided into two types: product logistics and transport logistics (Tavasszy et al., 1998).

Product logistics involves connecting demand and supply of goods, resulting in trade relations between origin and destinations of goods based on goods prices, availability and accessibility. Transport logistics involves the optimization of goods movement between locations utilizing the available transport modes and services by considering costs and quality elements such as reliability and speed (Boerkamps et al., 2000).

In this process, stakeholders involved are classified as public or private decision makers. Although, in general they have the same objective which is to transport goods and/or services, each has their own interests. Public stakeholders are the administrators who apply different policy measures to optimize the movement of commodities to improve the environmental condition, alleviate the traffic congestion, and any other enhancement to the society. While private stakeholders are shippers, carriers or receivers who take decisions that aim to maximize their profit by minimizing the cost of pickup or delivery the products. Receivers have the strongest influence on the demand of goods. This demand involves the types, volumes and delivery frequency depending on the characteristics of the goods. On the other hand, shippers are often responsible for transportation and therefore have to decide on mode choice, vehicle type, and vehicle size. They also decide on grouping of goods type with different logistical characteristics (Boerkamps et al., 2000).

The model developed by Tavasszy et al. (1998), Strategic model for integrated logistic evaluations (SMILE), consists of three stages; (i) production, (ii) inventory and (iii) transportation. In the first stage, make/use tables are used to form a production function that generates the volume of commodities produced and consumed for each sector. Then based on production chains, sales, sourcing processes at each location, the spatial distribution of the flows between these locations is determined. Afterward, at the second stage inventory chain are obtained using two steps. The first step involves finding optimal distribution location given three alternative channel types; direct, one distribution center or two distribution center. Then the second step involves using a choice model to assign flows to these alternative channel types based the total logistic costs (i.e. handling inventory and transport costs). Lastly, at the third stage, six modes of transportation are considered in a mode choice model, using the shortest route per mode for the choice disutility.

The model developed by Boerkamps et al. (2000), GoodTrip, is a modification of the four stage model to incorporate the supply chain. The supply chains are constructed by linking different activity types by distribution channels. Activity types can be consumers, supermarkets, stores, offices, distribution centers or factories. Distribution channel can be direct with one vehicle type or complex with multiple distribution centres, transport modes and transport companies. Based on consumer demand the model estimates the volume of goods produced in each zone per goods types. Goods flows are estimated by linking demands of activities to their supplying activities based on receiver choice and the spatial distribution of activities. The estimated flows are then combined using a groupage probability. Afterwards, each combined goods flow is assigned to vehicle tours by mode. Depending on origin's activity type, the transport mode, vehicle capacity, maximum load factor, and maximum number of stops per tour are determined. Finally, based on the destination activity type the minimal activity delivery frequency are determined. This model was calibrated and used to compare the logistical performance of three types of urban distribution systems in the city of Groningen in Netherland.

The GoodTrip model framework was the basis for developing a micro-simulation urban truck logistics model which was tested for Tokyo, Japan by Wisetjindawat and Sano (2003). Later, Wisetjindawat et al. (2006) extended the model to incorporate the fractional split distribution method. The model structure consists of two stages; commodity production and commodity distribution. At the first stage, the model utilizes regression techniques to estimate the production and consumption amounts based on firm characteristics such as number of employees and floor area. At the second stage, a spatial mixed logit model is utilized to incorporate the complex interactions among freight agents in a supply chain and the spatial interactions affecting each agent behavior. Hence, the fraction of a commodity *k* assigned to a customer *j* from shipper *i*, is the product of 1) distribution channel probability, 2) zone choice probability, and 3) shipper choice probability. Then the multiplication of the calculated fraction and the total amount of commodity consumption of customer *j* yields the commodity flow from shipper *i* to customer *j*. Summing all commodity flows among firms of each zone will produce the commodity OD matrix.

In view of that, De Jong and Ben-Akiva (2007) also developed a structure and identified the data sources needed for a new logistics module for Norway and Sweden to be included in the existing freight demand model. However, the model has not been estimated on disaggregate data. A recent study by Samimi et al. (2012) also proposed a framework to provide a behavioral picture of the current and future modal split in the U.S. freight transport market. Other logistics models presented an agent-based microsimulation that accounts for logistics reaction patterns (Liedtke & Schepperle, 2004; Roorda et al., 2010).

Data quality and availability is a vital key to develop and validate the supply-chain models and all other approaches in general (Samimi et al., 2012). This approach in particular, requires detailed information about the actual location of producers, distributors and consumer within a study area. Also, detailed information on the volume and frequency, the mode choice and route of each commodity flow for each origin-destination pairs are required (Zhou & Dai, 2012). As mentioned earlier, obtaining such information via surveys is mainly dependent on the cooperation of private stakeholders, who most of the time vacillate to share information related to their business freight and/or transportation activities such as logistics cost and the origins of raw material. As a results most of the existing efforts to collect detailed commercial travel data resulted in a low response rate.

Supply-chain model, attempts to explicitly account for the different components of the supply chain to understand, describe and predict their behavior under different scenarios. This dynamic approach makes it possible to analyse the interaction between economics and freight transport. Such models are able to provide valuable information about system mechanics and its activities to explain how changes in external factors such as socio-economic trends can affect the performance of logistics and transport system. Also, they are developed to answer questions related to measures to improve the system's performance and the impacts of these measures (Tavasszy et al., 1998). This framework makes it possible to analyze and evaluate logistical choices of different stakeholder before implementation. Last but not least, these models can be designed and calibrated to capture new shipping behaviours such as the adoption of outsourcing, e-commerce, (JIT) delivery systems (Samimi et al., 2012). Indeed, developing such framework will aid obtaining insights for future policy making.

Then again, the issue remain on how to overcome the limitation of the available survey data. Hence, these models are often suitable to model the supply chain of limited number of industries where reliable data exists. Therefore, most of the existing models are used in conjunction with freight model such as I/O models to understand the complete picture of freight flows within a study area (Zhou & Dai, 2012) Additionally, supply chain models generally focus on commodities flows were service delivery tours are not included.

2.1.3 Tour-Based Microsimulation Approach

Beside the fundamental characteristics mentioned before, freight transportation, particularly in urban area, exhibits some unique features. Among them, tour chaining where commercial vehicle movement are not as simple as two legs trip but rather manifest themselves in the form of tours. For example, when a commercial vehicle movement takes place to give rise to commercial trips, the vehicle starts at the establishment and in many cases make several stops before it returns to the establishment. Also, commercial vehicle movement tend to vary from day to day and season to season, unlike private vehicle passengers were they tend to follow the same pattern for their main activities like driving to work, school or limited recreational locations. In recognition of these characteristics, there has been a shift toward adopting the micro-approach instead of the conventional approaches to model commercial vehicle movement (Hunt et al., 2004).

This approach construct tours to simulate the salient chaining behaviour of commercial vehicle movement trips.

Hunt and Stefan (2007) where one of the first pioneers who developed a tour-based microsimulation for the city of Calgary, Alberta. First, using an aggregate trip generation model the number of tours generated by each category of establishment for each time period per zone is calculated. Then, using Monte Carlo techniques and single level logit model with utility functions that include zonal-level land use, establishment location and accessibility, each tour on the list for each zone is assigned a primary tour purpose and a vehicle type. Then using a Monte Carlo process each tour in the list for each time period is assigned a precise start time. Finally, again using a Monte Carlo technique and single level logit model, the next stop purpose, location and duration are assigned iteratively until the tour ends. This novel technique provides a very useful tool, to understand and simulate urban commercial movements well beyond the ‘freight only’, ‘large-truck-only’ and ‘regional-level’ approaches used previously. Later Ferguson et al. (2012), extended this study and examined the transferability of this framework to estimate the movements of commercial vehicles in other urban areas. The model was calibrated, implemented and validated for the Greater Toronto and Hamilton Area (GTHA) indicating that modeling framework from Calgary can be transferred and used to depict the travel patterns of commercial vehicle movements in other urban areas.

Joubert and Axhausen (2009) represented another novel approach in extracting commercial vehicle tours from raw global positioning system (GPS) data in Gauteng, South Africa. Building on that foundation, and using the same dataset, Joubert et al. (2010) developed an agent-based approach to reconstruct commercial vehicle chains and simulate them along with private vehicles. They adopted the same framework proposed by (Tavasszy, 2007) that extends the conventional four-step models to account for the decisions and issues relevant to freight modeling. This framework consists of five sub-modals: Production and Consumption, Trade (sale and sourcing), Logistics, Transportation and Network sourcing. Starting with synthesized population of agents, each selects a single plan from its set of plans. A plan consists of a set of sequential activities, each with a location that is associated with a given network. These activities are connected with a route through the network and the mode of transport. These plans are developed for each commercial vehicle based on conditional probabilities calculated for four chain parameters: the start location, start time, chain duration and number of activities per chain. When the model is executed, information is recorded in the form of discrete events such as the start and end of an activity, entering and leaving a link in the network, or waiting to access a link. Based on the selected plan simulation events are then interpreted to derive a score. Depending on the rewarded score these plan can be modified by changing the start time/location of the activities, or the mode/routing of the travel legs during the re-planning step. After a specified number of iterations, the achieved results are then compared to observed counts data. The achieved results were validated to be geographically and temporally accurate representation of observed vehicle chains.

These models usually require an extensive amount of data that are usually gathered from surveys. For the model developed by Hunt and Stefan (2007) a massive survey was conducted for Calgary and Edmonton regions, in the province of Alberta, Canada, to collect information on tours made on a typical weekday in the year 2001

(Hunt et al., 2006). The surveyed establishments provided information on the transportation activity of their entire fleet over a 24-h period, including origin, destination, purpose, fleet, commodity information and descriptions of the associated person and vehicle movements arising with this activity. Total survey costs including survey planning and design, execution, data coding and verification, and expansion were about $800,000. Other studies employed GPS data to understand and simulate commercial vehicle tours. This detailed and disaggregate data was used to represent the activity and chain durations, number of activities per chain and the spatial extent of the activity chain (Joubert & Axhausen, 2009).

Tour based microsimulation approach provide new insight in modeling urban commercial vehicle movement. It highlights the significance of light commercial vehicle and the prevalence of service trips opposed to good delivery within an urban area. This approach makes it possible to provide detailed representation of commercial vehicle tours without having to deal with shipment and the related complexities regarding conversion of commodity flows to shipment, the allocation of shipments to vehicle or routing (Hunt & Stefan, 2007). The model permits “feedback effects” to effectively assess response to changes in policies and hence leading to better planning of urban freight activities. For instance, it can be used to examine modal changes and rail planning (Chow et al., 2010). Moreover, it can be used to assess the impact of commercial vehicle movement on the environment and to provide measures to elevate system’s performance. It is worth mentioning again that the major constraint in developing such model is data availability.

2.2 Vehicle Ownership Modeling

Vehicle ownership models are typically used as an external subsequent model before modeling travel demand (Miller et al., 1998). Vehicle ownership has been considered as one of the important explanatory socio-economic factors that affect trip distribution and mode choice models in passenger travel demand models. Moreover, automobile ownership is a required input to most land use models. Therefore, estimating such variable is critical and can affect the ability to perform accurate future travel demand forecasting. More specifically, vehicle ownership levels can affect the prediction of vehicle miles of travel, traffic congestion, and air quality emission (Chu, 2002).

Our review to the literature indicates that the existing body of literature on vehicle ownership modeling has been focused on private vehicles. In comparison, no efforts have been conducted in the past to model commercial vehicle ownership. Given the scarcity of studies on commercial vehicle ownership, the remainder of this section highlights the work that has been done for the case of private vehicle ownership. While the ownership process for commercial vehicles differs from private vehicles, some of the general principles governing the ownership process hold in the case of the former. Also, the methods used to model private vehicle ownership could be used in the case of commercial vehicles.

The studies found in literature employ different frameworks to classify car ownership models. Jong et al. (2004) provide an audit of a wide range of models developed for public sector planning since 1995. They classified car ownership models into nine types and compared them based on sixteen different criteria. Potoglou and

Kanaroglou (2008a) complement the aforementioned study and focus on the specific aspects of disaggregate automobile demand. They provide a review of different studies with respect to data requirement, modelling approach, and the relevant explanatory variables. These studies include car ownership level, vehicle-type choice, vehicle holdings, and vehicle transactions models. They also discuss models that assess the potential demand of alternative fueled vehicles. A more recent study by Anowar et al. (2014) has reviewed the most noticeable disaggregate models that were developed within the past two decades using a four-way classification of the modeling framework. These four models categories are exogenous static, endogenous static, exogenous dynamic and endogenous dynamic. They also provide a decision matrix to aid researchers and practitioners to determine the appropriate model frameworks for conducting vehicle ownership analysis.

A variety of methodological techniques have been applied to analyze automobile ownership data. According to the available data, two broad modeling approaches can be identified: aggregate analysis at the regional level, and disaggregate analysis at the household level (Li et al., 2010; Chu, 2002). The statistical methods employed by researchers have primarily relied on the nature of the modeled dependent variable. The dependent variable of existing vehicle ownership models are typically either a binary response outcome (e.g., owning a car and not owning a car) (Karlaftis & Golias, 2002; Li et al., 2010), or a multiple response outcome (e.g. zero automobiles, one automobile, two automobiles, and three or more automobiles) (Potoglou & Kanaroglou, 2008a). Dependent variables with multiple response outcomes has been treated as ordinal (accounting for the ordinal nature of ownership data), nominal (no physical relation between the various ownership level) or count (representing an actual count of the number of vehicles owned)

According to Li et al. (2010), earlier approaches examined vehicle ownership as the accumulation of household decisions at a more aggregate geographic level that pertained to a region or a country. Here, different aggregate-based methods (e.g. aggregate OLS, aggregate time series models, aggregate cohort models, pseudo-panel data analysis, and longitudinal data analysis) were reported in the literature (Gómez-Gélvez & Obando, 2013; Thakuriah et al., 2010). However, the general consensus is that aggregate-based models tend to lack a behavioral basis to capture the variation among households. Also, aggregate models have been criticized to lack the proper policy sensitivity needed for practical urban planning applications (Anowar et al., 2014).

Recognizing the need to address the different behavioral aspects associated with the decision of a typical household to own a certain type of vehicle, the use of disaggregate modeling techniques became widespread. Examples are the work done by (Schimek, 1996; Ryan & Han, 1999; Wu et al., 1999; Hanly & Dargay, 2000; Baldwin Hess & Ong, 2002; Chu, 2002; Prillwitz et al., 2006; Sillaparcharn, 2007; Potoglou & Kanaroglou, 2008b; Li et al., 2010; Ma & Srinivasan, 2010; Thakuriah et al., 2010; ter Schure et al., 2012; Bhat et al., 2013; Dash et al., 2013; Gómez-Gélvez & Obando, 2013; Klincevicius et al., 2014). Across these studies, the discrete choice approach is widely used to model car ownership at the household level where two general decision mechanisms have been used; the unordered and the ordered-response responses.

The Multinomial Logit (MNL) model is the most well-known un-ordered response approach, with the assumption that an individual (i.e. decision maker) is rational and his/her vehicle type ownership decision is the one that will maximize his/her choice utility. On the other hand, the ordered logit (ORL) and ordered probit (ORP) models capture the ordered nature of vehicle ownership data. Unlike the MNL model, these models assume that a single continuous variable represents the propensity of a household to own a certain number of vehicles. Thresholds are then estimated to distinguish the different ownership levels (Gómez-Gélvez & Obando, 2013). Potoglou and Susilo (2008) have compared different car ownership models including MNL, ORL and ORP, and found that the MNL should be selected for modeling household car ownership. Despite the popularity of using the ordered and unordered discrete choice modeling to examine disaggregate vehicle ownership data, some studies acknowledged that ownership data represent typical count data. As such, several authors have applied count data regression models to model passenger car ownership levels (Karlaftis & Golias, 2002; Shay & Khattak, 2005). However, the application of such models remains uncommon in the literature (Anowar et al., 2014). Furthermore, others have applied spatial models to account for spatial autocorrelation across observational units. For example the Poisson-lognormal conditional autoregressive (CAR) model implemented by Chen et al. (2015) to examine Prius hybrid electric vehicles, other electric vehicles, and conventional vehicle ownership patterns at a neighborhood level. The results of this study emphasize on the importance of considering spatial auto-correlation patterns to avoid biased parameters.

Car ownership models have employed various explanatory variables to replicate the mechanism that a household experience when choosing a type of vehicle to own or when determining the number of vehicles to own. In general, these explanatory variables can be classified into the following categories: socio-economic and demographic characteristics, vehicle-specific characteristics, housing attributes and urban form characteristics. Based on the reviewed studies, household income has been reported as one of the most important socio-economic explanatory variables for car ownership. Another important determinant that has a strong influence on the number of vehicles maintained by the household is the number of household members and their designation (e.g. working adults, children, etc.) (Gómez-Gélvez & Obando, 2013). Other socioeconomic and demographic variables that were reported include driving license, ethnicity, gender, and occupation type (Chu, 2002; Karlaftis & Golias, 2002).

Automobile characteristics include variables that represent quality of vehicles, vehicle fuel efficiency and ownership costs. These variables are expected to increase vehicle ownership, as a result of the decreased per-kilometer cost of driving and increased quality of vehicles such as comfort and roominess (Schimek, 1996). Ownership costs including the expenditures of purchasing or leasing the vehicle, maintaining and operating the vehicle are expected to decrease vehicle ownership level as they increase (Ryan & Han, 1999). In term of housing attributes, variables representing the type of housing (e.g. single-family house detached or multi-family) have been used (Chu, 2002). Four types of urban form measurements have been considered in disaggregate car ownership models: land use measurement (e.g. population and residential density), urban design (e.g. land use mix, street width, and pedestrian connectivity), transit accessibility and location effect (Ryan & Han, 1999; Baldwin Hess & Ong, 2002; Li et al., 2010).

2.3 Population Synthesis

Population synthesis has been widely used in transportation research to develop activity-based microsimulation models and disaggregate land use models to address several policy relevant issues. It has been also incorporated as part of comprehensive socioeconomic and demographic model system and econometric micro-simulator for urban systems (Eluru et al., 2008; Pendyala et al., 2012). In particular, population synthesis is used as a preliminary step to construct the micro dataset that represent the characteristics of the agents used in a microsimulation. For these models the decision agents to be micro-simulated may include individual, households, dwelling or establishment populations (Ryan et al., 2009). In general, to develop such models, substantial amount of disaggregate data is required. However, almost all the available data are anonymized, geographically diluted, or generalized to specific spatial areas to protect the privacy of individuals (Voas & Williamson, 2000). In fact, when detailed information about individual demographics exists, their spatial location is diluted to maintain confidentiality (Frick & Axhausen, 2004). Hence, to overcome the tedious effort and long-time associated with data collection and the availability of data, synthetic populations for transportation research has received more attention over the past two decades (Arentze et al., 2007).

The data used to create synthetic populations are usually; aggregate zonal population data and disaggregate sample data. The former are available in terms of summary cross-tabulations of demographics represented as one-way, two-way, or multiway cross-tabulations that describe the joint aggregate distribution of relevant demographic and socioeconomic variables at the zonal level (Arentze et al., 2007; Ryan et al., 2009). For example, the Summary Files (SFs) used in the United States and the Small Area Statistics (SAS) file used in the United Kingdom (Beckman et al., 1996; Voas & Williamson, 2000). On the other hand, the disaggregate data represents a sample of individuals with information about the characteristics of each individual in it, excluding addresses and unique identifier. Examples are the Public-Use Microdata Samples (PUMS) in U.S. and the Sample of Anonymized Records (SAR) in UK (Beckman et al., 1996; Voas & Williamson, 2000).

A wide variety of techniques exist in the literature to estimate detailed microdata such as stratified sampling, geodemographic profiling, data fusion, data merging, reweighting, iterative proportional fitting synthetic reconstruction (IPFSR) and combinatorial optimization (CO) (Huang & Williamson, 2001). However, both IPFSR and CO have been identified as the most dominant techniques in the recreation of synthetic population microdata, although the IPFSR have been used more widely (Ryan et al., 2009). The synthetic reconstruction techniques presented by Wilson and Pownall (1976) current state-of-the-art, make use of the iterative proportional fitting (IPF) technique to create multiway tables of proportions that are consistent with the aggregate data totals. Then synthetic population of households is drawn from the microdata to match the proportions in the estimated multiway table (Beckman et al., 1996). In fact, with the variations in the types of input and how certain synthesis routines are carried out, a wide variety of the current population synthesizers involve the use of the IPF technique. Examples are the work done by (Arentze et al., 2007; Auld et al., 2009; Guo & Bhat, 2007; Martin Frick et al., 2004; Simpson & Tranmer, 2005).

Some of the more recent studies listed above attempted to address some of the shortcomings in the method presented in Beckman et al. (1996). For instance, the method in Beckman et al. (1996) does not address the zero-cell-value problem and the inability to control for statistical distributions of both household- and individual-level attributes. Guo & Bhat (2007) introduces a new population synthesis procedure that allows the user to adjust the choice of control variables and the class definition of these variables at run time to avoid initial incorrect value of zero in the contingency tables. That is when a specific demographic group in the population is represented in the aggregate data but not represented in the sample of the disaggregate data. Auld et al. (2009) also addressed this problem and developed a routine that allows for the aggregation of control variable categories during execution at the sub-regional level based on a user-controlled aggregation threshold parameter. The inability for controlling attributes on multiple analysis levels in a population synthesis program was also tackled by Auld et al. (2010). Their methodology is implemented within a population synthesizer to allow multiple-level synthetic populations such as household- and person-level, establishment and employee or household and vehicle estimation.

The combinatorial optimization (CO) technique has been used as an alternative to the IPFSR method. This iterative technique is more computationally efficient as it synthesizes the population on a zone-by-zone basis. That is, synthesis for multiple zones could be executed at the same time using multiple processes or workstations. For a given zone, the CO method starts by choosing a random set of households from a sample microdata with replacement from the micro list, then assessing the effects of replacing one of the selected households. Accordingly, the replacement will be made only if the swap improves the fit. Williamson et al. (1998) presented different techniques to solve the combinatorial optimization problem pertaining to synthesizing small-area micro population estimates. These included: 1) hill climbing approach, 2) simulated annealing (SA) approach, and 3) genetic algorithm approach. It was found that the SA is the best performing approach. Later, Voas & Williamson (2000) assessed the implementation of the CO technique and proposed a sequential fitting procedure to further improve the quality of the synthetic microdata. The simulated annealing approach is robust for its ability to find an optimal solution much faster than the classical hill-climbing method.

SA has been proposed to solve optimization problems that may have several local minima. According to Yang (2010), SA can be used to solve global optimization problems by mimicking the annealing process of solid material. Annealing in the physical world entails liquefying a solid material (e.g. metal) then solidifying or crystalizing it at a low temperature. Crystallization at a low temperature tends to reduce the defects in the metallic structure of the material. In such case, global minimum energy is used to produce the solid state of the atoms forming the material. As noted in Yang (2010): "*the annealing process involves the careful control of temperature and cooling rate, often called annealing schedule*". Unlike the hill-climbing method, SA avoids being trapped in local minima, thus speeding up the process of finding an optimal solution.

The change in state (i.e. from liquid to crystal) is driven by a probability function that is analogous to the Boltzmann probability from the physical world:

$$P = \exp\left(-\frac{\Delta E}{kT}\right) \tag{2.1}$$

where ΔE is the change in energy level, T is the temperature for controlling the annealing process and k is Boltzmann constant which is assumed to equal 1. In the physical word, if the liquid is cooled at a low temperature, then the energy spent to change from the liquid state to the solid state will be optimal (i.e. global minimum energy). In such case, the formed solid will have the least amount of structural defect. The transition in state occurs over time and is subjected to a particular cooling schedule that could be either linear or geometric. In the linear case, $T = T_0 - \alpha t$, where T_0 is the initial cooling temperature, α is the cooling rate and t is a time instant. On the other hand, a geometric cooling temperature function takes the form $T = T_0 \alpha^t$. An advantage of the geometric form is that T will approach 0 when time reaches infinity. The cooling factor α is usually set to a value in the range of 0.7 – 0.99 to reflect a slow cooling process that will enable the system to stabilize. Usually, changes in energy level at time t (i.e. for $\Delta E = E_{t+1} - E_t$) will be accepted if the Boltzman probability P is greater than some random number r.

In CO problems, the objective function Δf is directly related to ΔE. That is, $\Delta E = \Delta f$ and this give rise to the transition probability

$$P = \exp\left(-\frac{\Delta f}{T}\right) \qquad (2.2)$$

When running the CO, the choice of the initial temperature T_0 and cooling rate α are critical in achieving the equivalent of an optimal minimum energy (i.e. global minimum solution). Very high temperature represents a system at a very high energy state that makes it hard to achieve an optimal solution. On the other hand, very low temperature represents a situation where the system energy is not enough to jump out of a local minimum (i.e. system is trapped in a local minima state).

The CO technique is used to synthesize a disaggregate list of individuals with attributes that when aggregated will conform to a predefined zonal totals. For the case presented here, the simulated annealing (SA) algorithm starts by drawing N records from the available micro-sample at random and with replacement. The drawn N records are used to form aggregate tables that are then compared to the predefined zonal totals. The comparison, which could be the absolute difference between the generated aggregates from the random draw and the predefined aggregates, give rise to some cost function f_t. Next, the SA algorithm will generate another random draw from the micro-sample by arbitrarily swapping one of the records drawn from the previous iteration with a new record from the micro-sample. The new list is then used to calculate a new cost function f_{t+1}. Next, the algorithm will evaluate the fitness of f_{t+1} and f_t relative to the predefined zonal totals. If f_{t+1} provides a better solution f_t then the swaps is accepted, otherwise, the algorithm calculates the transition probability P given the Δf ($\Delta f = f_{t+1} - f_t$), α and the initial temperature T_0. If the calculated probability P is greater than some random number r then the list used to calculate f_{t+1} is accepted even though it did not provide a better fit compared to f_t. This enables the algorithm to move more freely within the solution space to arrive at the global minima in much faster computer run-time. Note that the hill-climbing method will only consider the swap at instant (iteration) t+1 if and only if f_{t+1}provides a superior fit over f_t. The SA will keep iterating until the synthesized list is able to mimic the predefined zonal aggregates or when a maximum number of iterations is reached.

Given their popularity, the CO and IPFSR techniques have been compared to find which would achieve better results. According to Ryan et al. (2009), both techniques are capable of producing synthetic microdata that fit constraining tables extremely well. However, the CO technique is deemed superior for its ability to produce more accurate results with the variation in tabulation details and input sample size. Also, the utilization of the SA algorithm provides a further advantage given its computational superiority when compared to the IPFSR technique.

CHAPTER 3: STUDY AREA AND DATA DESCRIPTION

3.1 Study Area

The analysis in this research is focused on the Windsor CMA located on the south shore of the Detroit River and Lake St. Clair. Windsor occupies approximately 1022.31 square kilometers of Canada's land (Statistics Canada, 2011). The strategic location of Windsor across from Detroit Michigan, U.S.A. makes it an international gateway for people and commerce. According to the 2011 Canadian Census, the CMA housed 323,342 people, 126,845 households and 123,305 jobs. The manufacturing sector accounted for 21 percent of all jobs in 2011, followed by retail trade and health care with each accounting for a respective 12 percent of the total jobs. Educational services ranked 3rd with just below 8.9 percent. Windsor's job distribution by industry is fairly similar to the total national distribution except for the manufacturing sector. Excluding Windsor, the national share of jobs pertaining to the manufacturing industry was just below 9.5 percent. Windsor's population and employment size and the strong presence of the manufacturing sector guarantee the presence of commercial vehicles (CVs) and the generation of CV trip activities.

3.2 Data Description

The data used for this research was acquired from different sources: R. L. Polk and Co., InfoCanada, Business Establishment Commercial Travel Survey (BECTS), and other sources. As an input to population synthesis process 1) the InfoCanada dataset is used to create representative aggregate cross-tabulations for the total firm population that engage in commercial activities, 2) the BECTS results is used to extract a representative microsample and 3) the Polk dataset is used to validate the results achieved as will be discussed in the next section. On the other hand, for the modeling framework, the Polk dataset was also used as the dependent variable within the models and the other sources such as the 2011 Canadian Census data was used to provide independent parameters as input variables within the models.

3.2.1 InfoCanada Dataset

The first dataset was acquired from InfoCanada and consists of all 10,771 establishments registered in the Windsor CMA in 2013. The attributes provided for each establishment includes; the InfoUSA (IUSA) code designation for each establishment, contact information (i.e. telephone number and full address), employment size and industry classification according to the North American Industry Classification System (NAICS) six-digit code and the Standard Industry Classification (SIC) code.

3.2.1.1 Geocoding

To get the zonal location of the business establishments within the Windsor CMA, this dataset was uploaded into ArcGIS and geocoded using the street addresses and postal codes. Initially, all business establishments with no addresses and postal codes (i.e. 606) were dropped from the analysis, accounting for approximately 5.6% of all business establishments. Then, an address locater was created using the Desktop Mapping Technology Inc (DMTI) road network. Out of the 10,165 business establishments considered, 8,971 (i.e. 88%) were successfully geocoded. To further increase the percent of records that are spatially located, a shapefile with all the local delivery units (LDU) in the Windsor CMA from CHASS was used. This file had three records: Postal Code, Longitude and the Latitude for each delivery unit in the CMA. Accordingly, an additional 968 establishments were geocoded by joining the postal codes address of the business establishments with the postal codes in the LDU shapefile.

As a result of the geocoding process, 9,939 business establishments were geocoded, representing (97%) of all the addresses considered. Figure 3-1 illustrates the location of the geocoded establishments within the Windsor CMA, where establishments mainly operate from 65 zones rather than all the 73 zones comprising the study area.

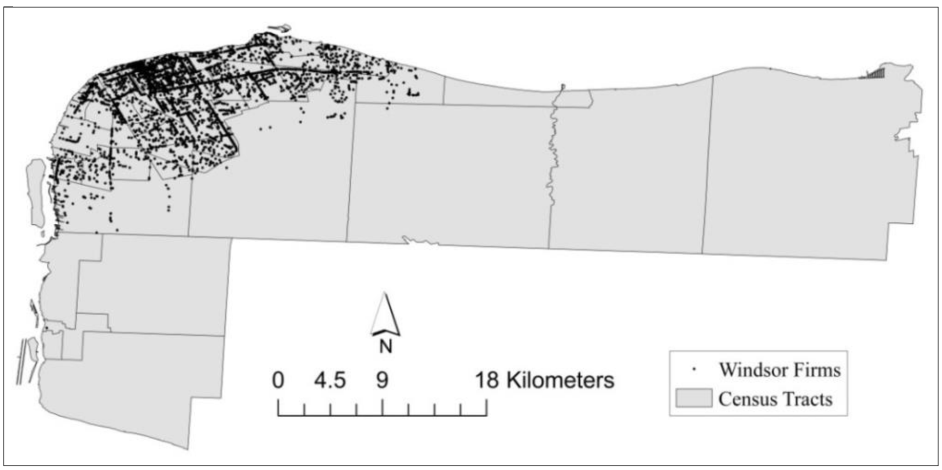


Figure 3 - 1: Establishment Spatial Distribution in the Windsor Census Metropolitan Area, 2013

3.2.2 Business Establishment Commercial Travel Survey

The second data source is the Business Establishment Commercial Travel Survey that was conducted by the University of Windsor in 2013 (Madar, 2014). This survey was undertaken in two stages; a telephone based survey followed by a web based survey. During stage one, businesses were contacted through a short telephone survey to know whether they engage in shipping and/or receiving goods and services and if yes, whether the establishment would partake in the online survey. A total of 6,740 establishments where contacted in the Windsor CMA, where 1,461 reported that they engage in shipping or receiving goods and/or services. Furthermore, from the 1,461 establishments, 681 agreed to participate in the online survey.

During stage two, a link to the web survey was provided for those establishments that agreed to participate in the online survey. Then reminder emails were sent to establishments' representatives after 2, 7 and 14 days. Recruited establishments were given 40 questions about the general establishment characteristics and inbound and outbound commercial activities on the day the survey. The data collection process is described in details in (Madar, 2014). Out the establishments who provided contact information to receive the survey, 171 completed the survey. However, only 162 establishments completed the questions that pertain to the attributes required to create the microsample that will be used for the synthesizing process as will be discussed in the following chapter of this thesis. These attributes provide information that characterizes the surveyed establishments (e.g. location, industry type, employment size and number of owned vehicles).

The industry type and the employment size reported were then checked, verified and corrected according to the information provided in the InfoCanada dataset. Since the values provided in the InfoCanada dataset are accurate and verified, whereas the survey responses were dependent on the respondents (i.e. establishments' representatives) judgment and knowledge. Figure 3-2 represents the number of employees reported by each firms, it can be seen that the majority of the establishments (i.e. 66%) have a total of ten employees or less.

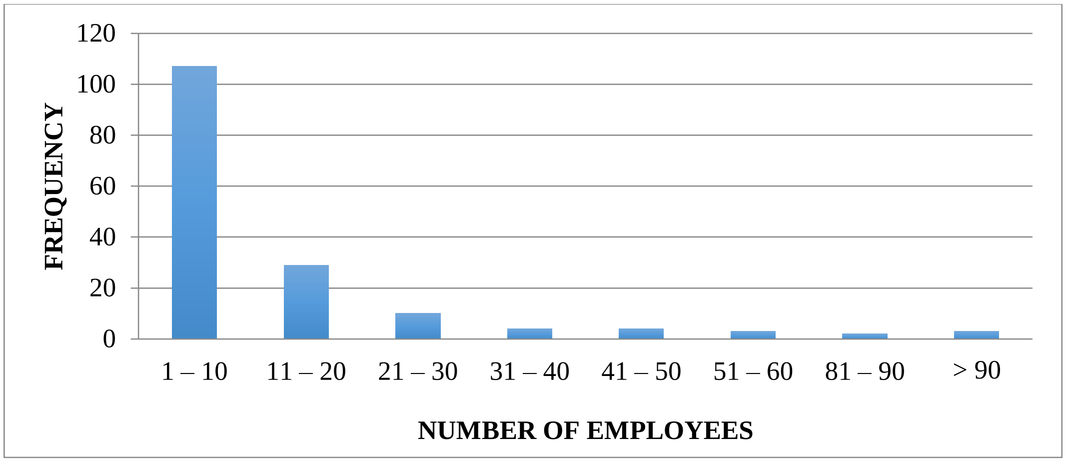


Figure 3 - 2: Frequency of Reported Total Number of Employees for Survey Respondents

Figure 3-3 represent the frequency distribution of the establishments by their two-digit SIC industry category and Table 3-1 provides a list all of the SIC categories to which survey respondent firms belonged with a general description of each category. A detailed description of each industry class is shown in Appendix A. Most of the respondent establishments belonged to “Retail Trade”, “Manufacturing” and “Wholesale Trade” accounting for 36%, 33% and 15% of all respondents, respectively.

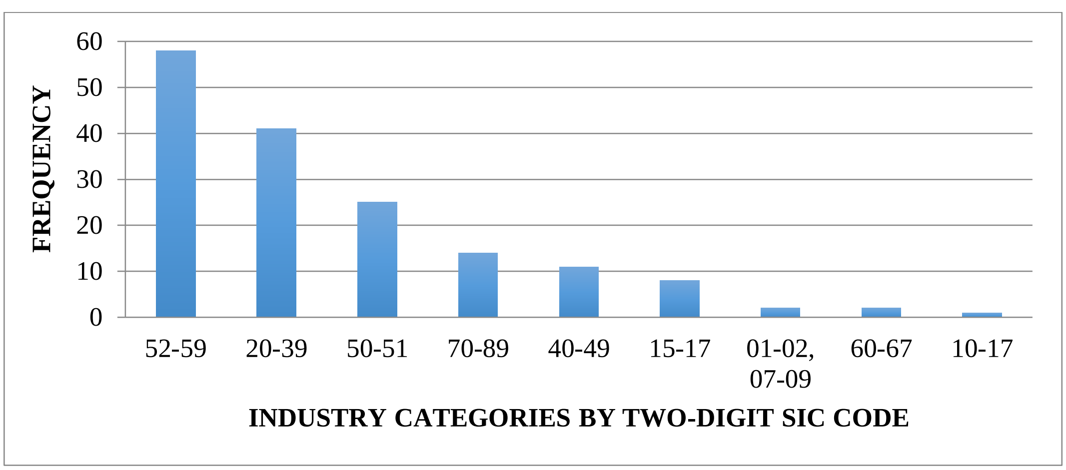


Figure 3 - 3: Frequency of Industry Classifications of Survey Respondents

Table 3 - 1: Two - digits SIC codes and description of Survey Respondents

Industry Code	Description
01 – 02, 07 – 09	Agriculture, Forestry, Fishing
10 – 14	Mining
15 – 17	Construction
20 – 39	Manufacturing
40 – 49	Transportation & Public Utilities
50 – 51	Wholesale Trade
52 – 59	Retail Trade
60 – 67	Finance, Insurance, Real Estate
70 – 89	Services
91 – 99	Public Administration

The diversity in the distribution of firms per industry classes and firm sizes lead to a varied CVs ownership across the firms. Figure 3-4 shows the distribution of the reported total number of vehicles owned by each firm. It can be observed that most establishment own four vehicles or less. A possible explanation for such trend is the large portion of establishments with low employment.

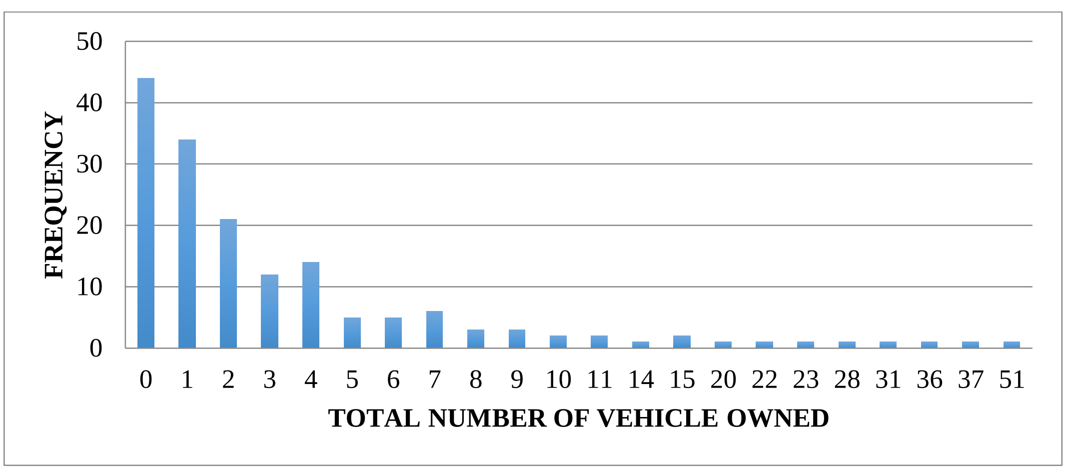


Figure 3 - 4: Frequency of Total Number of Vehicles Owned by Survey Respondents

3.2.3 Polk Dataset

The third dataset is the Polk data for the year 2013 acquired from R. L. Polk and Co. This dataset consists of all registered CVs at the census tract level in the CMA, where commercial vehicles are classified into eight different classes according to their Gross Vehicle Weight (GVW) and geo-referenced to the census tract level. Table 3-2 provides the weight ranges in pounds for each GVW class.

Table 3 - 2: Gross Vehicle Weight Classes and Corresponding Weight Ranges in Pounds

GVW	Wight Ranges
Class 1	6,000 lbs. or less
Class 2	6,001 – 10,000 lbs.
Class 3	10,001– 14,000 lbs.
Class 4	14,001– 16,000 lbs.
Class 5	16,001– 19,500 lbs.
Class 6	19,501– 26,000 lbs.
Class 7	26,001– 33,000 lbs.
Class 8	33,001 lbs. and over
Class 9	Cars

Based on the acquired data, the total number of commercial vehicles in 2013 in the 73 zones comprising the CMA was found to be 13,983 vehicles. However, vehicles registered outside the 65 zones (that were identified through the geo-coding process) were excluded from the analysis. Also, vehicles for 2 of the 65 zones were excluded since these had no establishments engaging in shipping goods and/or services. As a result, the total number of registered vehicles was reduced from 13,983 to 12,240 vehicles and those were attributed to 63 zones.

The Polk dataset provided information with regards to the model, make and year of each registered vehicle. Given this information, the counts of registered vehicles in each zone by their year, model and make were considered. The exploration indicated the existence of large counts of the same vehicle spanning from 23 to 220 vehicles per unique year, make and model in zone 5590033. For instance, this is the only zone that has a count of 220 vehicles of “2013 Chrysler 200”. Consequently, this zone was also dropped from the analysis since it represents a clear outlier compared to all other zones. This is particularly the case because the count of vehicles pertaining to a unique year, make and model class in any of the other zones was fairly low with an average of 5 vehicles. In fact, the average among 59 zones was 3 vehicles and only 5 zones had counts of 30, 24, 22, 18 and 13 vehicles per unique year, make and model class, respectively as illustrated in Figure 3-5. After dropping zone 5590033, the total number of registered vehicles was further reduced from 12,240 to 8,869 vehicles. The latter are housed in 62 zones.

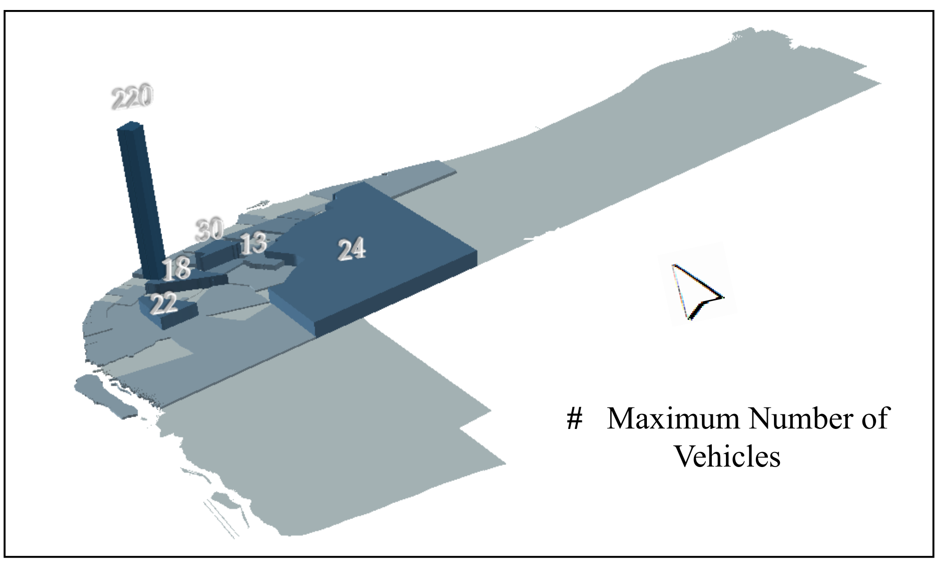


Figure 3 - 5: Max Number of Vehicle per Make, Model and Year per Zone

Figure 3-6 presents the breakdown of the 8,869 CVs registered in the 62 zones considered for the analysis. As can be seen, the majority (78%) of all the considered registered CVs are either cars or light duty trucks (i.e. GVW 1 and GVW 2).

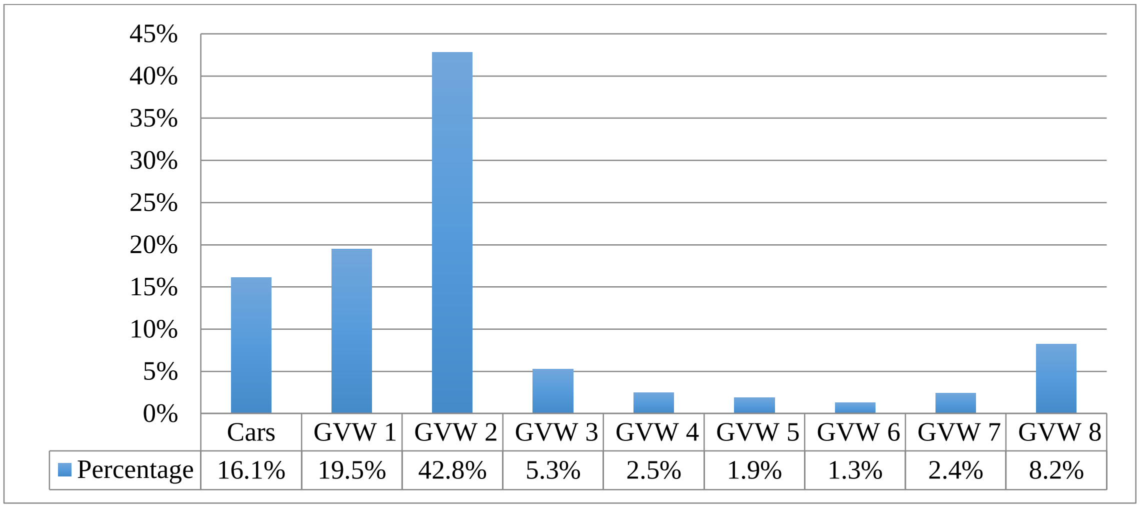


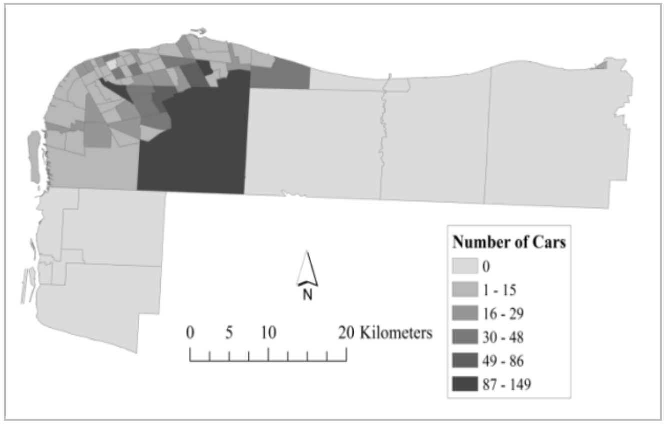
Figure 3 - 6: Distribution of Commercial Vehicle Classes

3.2.3.1 Spatial Autocorrelation (Moran's I Statistics)

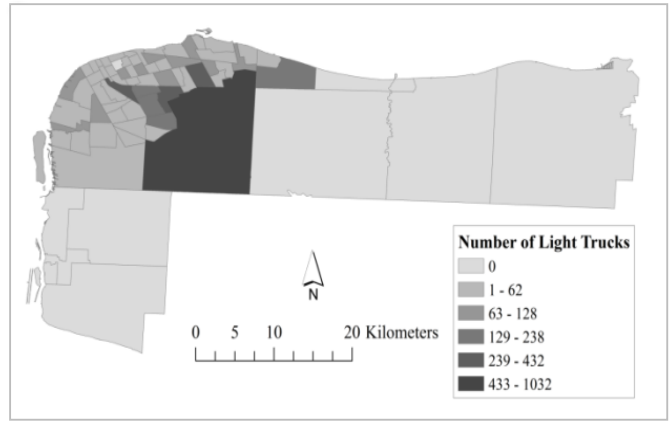
Moran's I Statistic was calculated in ArcGIS to test if the spatial distribution of commercial vehicles by class exhibits geographic clustering. The eight GVW classes were grouped based on the Federal Highway Administration (FHWA) classifications to: (i) Cars; (ii) Light Duty Trucks (GVW 1-2); (iii) Medium Duty Trucks (GVW 3-6); and (iv) Heavy Duty (GVW 7-8). Figure 3-7 highlights the spatial distribution of these four classes. The Global *Moran's I* statistic tested the null hypothesis that the attribute being analyzed is randomly distributed in the study area (i.e. no clustering).

Figure 3-8 provides the estimation results for the commercial vehicles at the census tract level using a first order rook contiguity spatial weight matrix. The estimated p-value is statistically significant for all vehicle classes, suggesting that the null hypothesis of no spatial autocorrelation can be rejected. The p-value of 0.041 estimated for cars indicates that there is less than 5% likelihood that this clustered pattern could be the result of random chance. While, the p-value of 0.002, 0.004 and 0.001 for light, medium and heavy duty trucks, respectively, indicate that there is less than 1% likelihood that this clustered pattern could be the result of random chance.

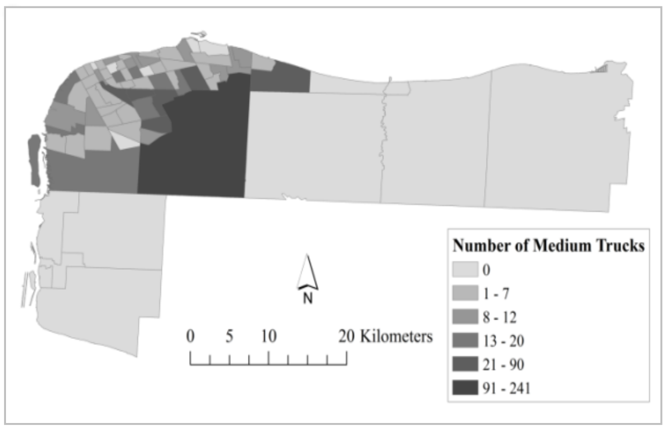
Therefore, the spatial distributions of all vehicle classes in the study area exhibit the presence of clustering over space. Note that individual GVW classes were also tested for clustering and the results were equivalent to their aggregates. Results of these individual tests are shown in Appendix B. Accordingly, these clustering patterns suggest an underlying process that gives rise to the prevalence of commercial vehicle per class in particular census tracts.



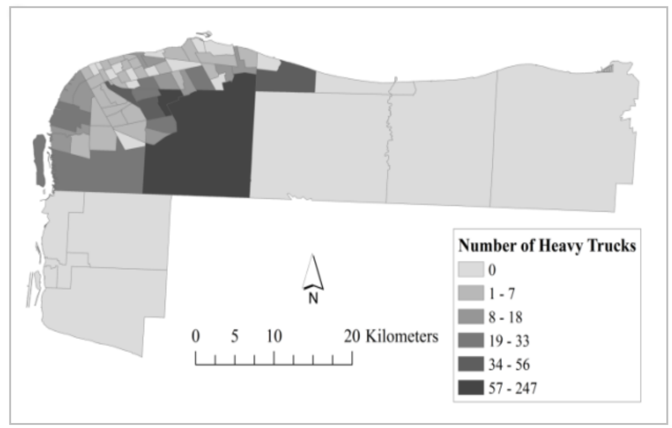
(a)



(b)



(c)



(d)

Figure 3 - 7: Spatial Distribution of (a) Cars, (b) Light Duty Trucks, (c) Medium Duty Trucks and (d) Heavy Duty Trucks by Place of Registration in Windsor in 2013

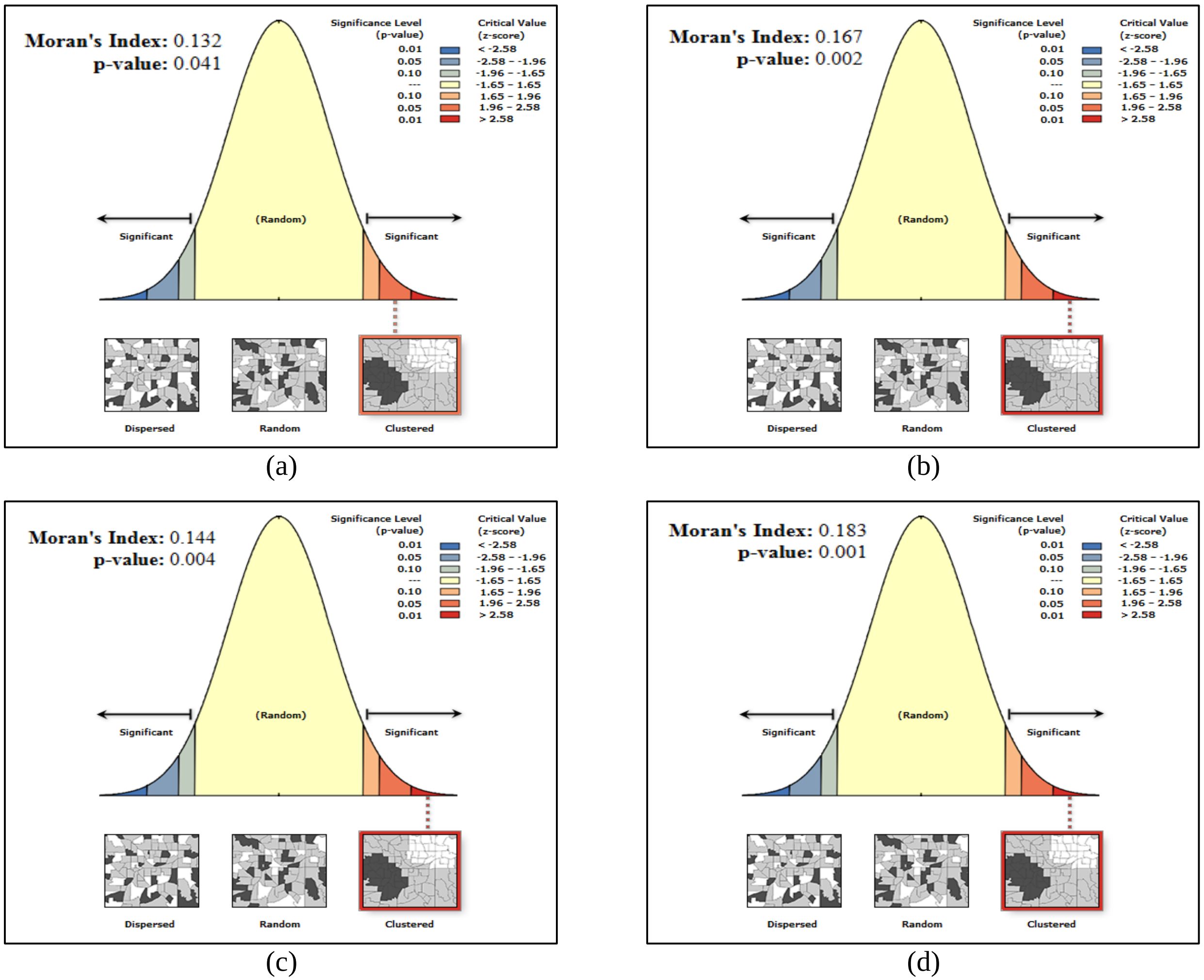
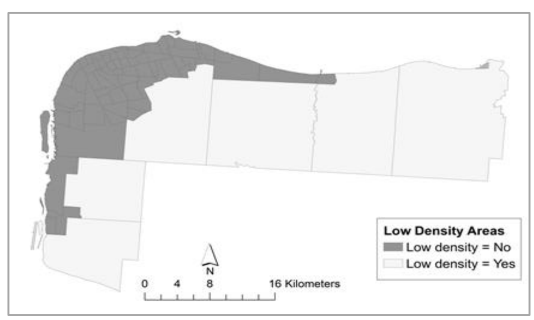


Figure 3-8: Moran's I result for (a) Cars, (b) Light Duty Trucks, (c) Medium Duty Trucks and (d) Heavy Duty Trucks

3.2.4 Other Data Sources

Other data sources were used to provide input variables in the models. This includes the National Household Survey (NHS) of the 2011 Canadian Census and different ESRI shapefile layers created by DMTI. The NHS dataset was acquired from the Computing in the Humanities and Social Sciences (CHASS) census analyzer from the University of Toronto, it provides demographic information such as the population and the employment numbers by industry type (NAICS) for each census tract. On the other hand, the shapefile layers obtained from DMTI were used in ArcGIS to identify and create various zonal variables needed to estimate the discrete choice models and the count models. For instance, ArcGIS was used to generate variables that 1) identify low density zones (i.e. less than hundred jobs per km^2 and less than hundred residents per km^2), 2) locate zones in proximity with highways (i.e. highway infrastructures passes through the zone), and 3) calculate the equidistance in kilometres between the centroid of different zones and the Central Business District (CBD) of the study area. Figure 3-9 shows zones in close proximity with highways and low density suburban zones, respectively.

(a)



(b)

Figure 3 - 9: (a) Zones in Proximity to Highways and (b) Low Density Suburban Zones

CHAPTER 4: METHOD OF ANALYSIS

The framework followed for this research is illustrated in Figure 4-1. Synthetic population techniques are first employed to micro-simulate the number of commercial vehicles owned by all individual business establishments that engage in delivering goods or services in the Windsor CMA. This process consists of three consecutive steps: 1) identifying all the establishments that engage in shipping goods and/or services in the CMA, 2) applying the Simulated Annealing Procedure to solve the Combinatorial Optimization (CO) problem of assigning commercial vehicles to business establishments, and 3) validating the synthesized records using an external dataset (Polk data) that was not initially used in the population synthesis procedure.

This process provides the basis to justify the use of zonal data from Polk to analyze the spatial prevalence (or assignment) of commercial vehicles (CVs) in a given census tract. The main concern is that some CVs that are registered to the establishment in the census tract as given in the Polk data might not be physically located or operating from that location. The hypothesis is that synthesized aggregates are a true representation of the vehicles registered and operated from the same zone. Hence achieving comparable results would provide a basis to use the of Polk data to examine the spatial distribution of commercial vehicles. Accordingly, for the zones with Polk totals less than the synthesized totals, the Polk totals are used. However, if the Polk zonal totals were greater than the synthesized totals, the synthesized totals are used. In view of that, discrete choice models and count models are used to study the spatial distribution of commercial vehicles by their GVW classes.

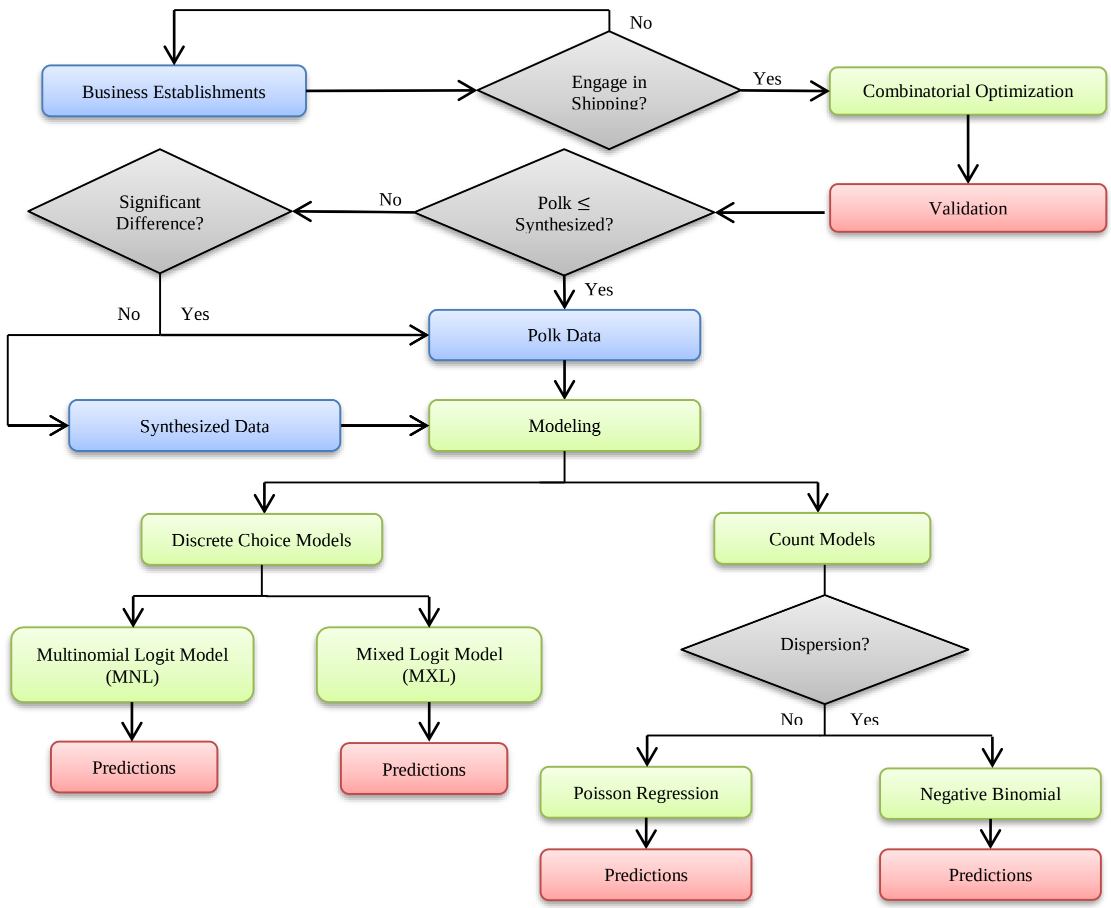


Figure 4 - 1: Methodology Flow Chart

4.1 Population Synthesis

The CO technique is used to synthesize a disaggregate list of establishments with attributes, when aggregated conform to predefined zonal totals. For the case presented here, the simulated annealing approach was used to reach to the optimal solution. The annealing parameters used to synthesize the population are as follows: initial temperature (T_0)= 400, cooling parameter (α) = 0.999, final temperature for convergence *(T)* = 0.001 and number of swaps = 1. Then the goodness of fit was assessed using the Relative Sum of Squared Z-Scores (RSSZ) documented in Ryan et al. (2009). For more information on the simulated annealing approach in the context of CO method see (Williamson et al., 1998). This process was repeated ten times to confirm the consistency of these populations. The program to execute this process was written in C# and is called the Combinatorial Optimizer program.

To generate the aggregate cross-tabulations, as an input to synthesizing process, first the list of individual business establishments that engaged in shipping and receiving had to be determined. This is performed using a participation quotient (PQ) approach. The latter is based on the number of establishments that reported to engage in shipping and/or receiving goods in the BECTS. The PQ index is calculated as the ratio of (F_n^s) the number of businesses that reported to engage in shipping or receiving within a specific industry category n, to ($\sum_n F_n^s$) the total number of establishments who reported engaging in shipping or receiving, divided by the ratio of (F_n) the number of businesses in the respective industry to ($\sum_n F_n$) the number of establishments in the entire population of establishments as illustrated in Equation 4.1. Note that the firms were categorized to the most detailed industry classification (i.e. six-digit SIC industry classification)

$$PQ_S = \frac{F_n^S / \sum_n F_n^S}{F_n / \sum_n F_n} \tag{4.1}$$

This method is inspired by the Location Quotient technique, a well-established method that has been used in economics and economic geography (Miller and Blair, 2009). The rational is, a PQ value greater than or equal to 1 indicates that the industry category being investigated is more likely to engage in shipping and/or receiving as the industry will have higher concentration of its firms in the sample relative to the entire firm population. However in the case presented here, a threshold value of 0.7 was chosen after considering different values. The 0.7 cut-off value achieved the most reliable outcomes as will be discussed in the results section.

Accordingly, all establishments that belong to industries with PQ greater than or equal to 0.7 were kept for the analysis, while establishments that belong to industries with PQ less than 0.7 were dropped except for those establishments who were originally surveyed. Out of 9,939 establishments in the study area, a total of 3,478 establishments were found to engage in shipping and/or receiving goods and/or services (35%).

To create representative aggregate cross-tabulations for the 3,478 establishments, the attributes considered are: census tract ID (CTID), number of employees and a two-digit SIC industry classification. The values from the number of employees attribute were reclassified into ten categories representing discrete employment ranges as illustrated in Table 4-1. The two-digit SIC industry categories to which establishment belonged are 62 mutually exclusive categories. Besides, the CTID attribute with 62 categories.

Table 4 - 1: Categories for Number of Employees

Category	Range (Number of Employees)
1	1 – 10
2	11 – 20
3	21 – 30
4	31 – 40
5	41 – 50
6	51 – 60
7	61 – 70
8	71 – 80
9	81 – 90
10	> 90

In what follows, two cross-tabulations for the CO were derived based on the two-digit SIC industry classes (SIC-2D) and employment size categories (Emp-Cat). The first tabulation represents the breakdown of zonal firms by Industry class and the second by employment size categories as shown in Table 4-2 and Table 4-3, respectively.

Table 4 - 2: SIC-2D x CTID Tabulations Derived from the Windsor Firm Population

Industry	Zone 1	Zone 2	Zone 3	...	Total
SIC – 57	2	13	1	...	193
SIC – 58	6	17	1	...	595
SIC – 59	8	14	0	...	442
⋮	⋮	⋮	⋮	...	⋮
Total	40	121	11	...	3478

Table 4 - 3: Emp-Cat x CTID Tabulations Derived from the Windsor Firm Population

Employment	Zone 1	Zone 2	Zone 3	...	Total
Cat – 1	30	90	9	...	2615
Cat – 2	6	9	2	...	436
Cat – 3	0	5	0	...	137
⋮	⋮	⋮	⋮	...	⋮
Total	40	121	11	...	3478

Next, a micro-sample of 162 establishments were extracted from the survey responses were information about the two-digit SIC industry classification and employment category and the number of commercial vehicles owned are stated, as shown in Table 4-4.

Table 4 - 4: Microsample Derived from the BECTS

ID	SIC-2D	Emp-Cat	Vehicle Owned
Firm 1	42	2	22
Firm 2	57	3	5
Firm 3	50	6	36
⋮	⋮	⋮	⋮
Firm 162	35	4	0

The CO method was then used to create a list of 3,478 establishments, where each establishment in the list has attributes that represent the employment size class and corresponding two-digit SIC category. Each synthesized establishment was linked directly to an establishment from the micro-sample, and as such the values for the number of vehicles owned were assigned.

4.2 Modeling the Spatial Distribution of Commercial Vehicles

This research will make use of two major methodological approaches to study the spatial distribution of commercial vehicles: 1) discrete choice modeling, and 2) count models.

4.2.1 Discrete Choice Modeling Approach

The discrete choice modeling techniques namely, the MNL and the MXL models are used to capture the effect of different zonal and vehicle characteristics on the spatial prevalence of commercial vehicles in zones comprising the study area. The econometric analysis for the discrete choice models is performed in the NLOGIT 5 software.

Given the population of CVs, the MNL or the MXL model predicts the probability of finding a given commercial vehicle *c* in one of the 62 zones encompassing firms that engage in shipping and receiving goods and/or services. Although the choice probability could be modeled across all 62 zones, we opted for generating a smaller choice set of ten alternative zones. The later was formed by considering the actual zone *i* where vehicle type *c* is observed and adding to it 9 randomly selected zones from the remaining 61 zones as shown in Figure 4-2. Such practice has been used before to overcome the problem of dealing with large choice sets and it is proven to produce reliable coefficient estimates for the logit model (McFadden, 1978; Maoh & Kanaroglou, 2009). As shown by McFadden (1978), the random sampling of alternatives will still produce efficient and unbiased parameters with the smaller choice set.

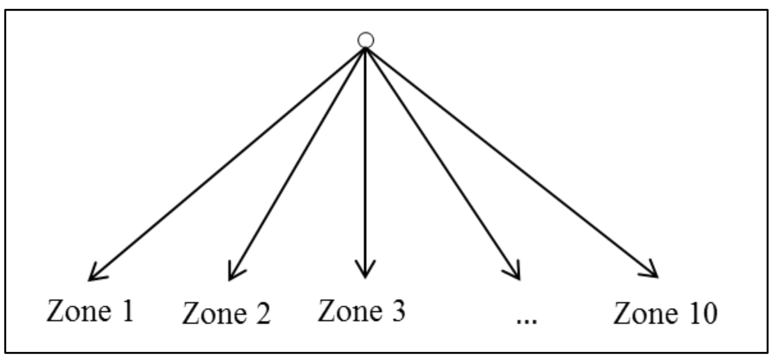


Figure 4 - 2: Discrete Choice Model Structure

Each zone *i* is associated with a utility function, U_{ic} that can be expressed as follows:

$$U_{ic} = V_{ic} + \varepsilon_{ic} \tag{4.2}$$

$$V_{ic} = \sum_k \beta_k \cdot Xk_{ic} \tag{4.3}$$

where V_{ic} is a linear-in-parameter deterministic function characterizing the nature of alternative zone *i* and the attributes of commercial vehicle *c*. For the MNL, characteristics Xk_{ic} of alternative *i* and vehicle *c* will be associated with a single point coefficient β_k that represents the influence of Xk_{ic} on the probability of finding vehicle *c* in zone *i*. The random error terms ε_{ic} (for all *i* and *c*) are assumed to be independently and identically distributed (iid) across alternatives and observations according to a Gumbel probability density function. Accordingly, the location of a given commercial vehicle *c* is modelled by calculating the probability that vehicle *c* will be located in specific zone *i* such that:

$$P_{ic} = P\,r\left(\beta_k \cdot Xk_{ic} + \varepsilon_{ic} \geq \beta_k \cdot Xk_{jc} + \varepsilon_{jc}\right) \text{ for all } i \neq j \text{ and } i;\ j \in C \qquad (4.4)$$

Where the MNL choice probability can be represented as follows:

$$P_{ic} = \frac{\exp(V_{ic})}{\sum_{j=1}^{10} \exp(V_{jc})} \qquad (4.5)$$

Despite its widespread use in travel demand modeling, the MNL model has been replaced by the MXL model in recent years. Unlike the MNL model, the MXL model provides more flexibility when modeling the choice data since it assumes that parameter β_k associated with covariate Xk_{ic} is random and varies across the modeled commercial vehicles (Train, 2009). Therefore, the MXL model accounts for unobserved heterogeneity in the modeled sample. Following this assumption P_{ic} can be formulated as the weighted probability across all possible β's that are drawn from a known probability distribution (Φ), that is:

$$P_{ic} = \int_{\beta_c} \frac{\exp\left(V_{ic/\beta_c}\right)}{\sum_j \exp\left(V_{jc/\beta_c}\right)}\ P(\beta|\Phi)\ d\beta \qquad (4.6)$$

4.2.1.1 Model Specification

The starting point for specifying the deterministic utility function V_{ic} is to consider jobs in the various zones. These are the locations where CVs are housed by their respective firms. The rationale followed is based on the premises that locations where jobs are found give rise to the existence of CVs since firms housing these jobs will need to own vehicles for their business transportation activities (i.e. delivering goods and/or providing services). Given the diversity in the clustering pattern of industries over space (Maoh & Kanaroglou, 2006), variability in the spatial distribution of CVs in the various zones across the city of these vehicles are expected.

Intuitively, not all industries will be dependent on the same types of vehicles for their business transportation activities. For instance, the basic industry is more likely to own heavy duty trucks while firms from the services sector are more prone to own and use small cars and light commercial trucks. The inclusion of a list of variables which represent the number of jobs (by zone) from a particular industry would capture these differences especially in the presence of interaction terms characterizing the class of commercial vehicles. For the latter, the GVW classification is utilized. More specifically, as mentioned earlier, the nine GVW classes are grouped to: (i) Cars; (ii) Light Duty Trucks; (iii) Medium Duty Trucks; and (iv) Heavy Duty.

Zonal jobs are classified into 5 major industrial groups following the categorization used in Hunt and Stefan (2007). That is, basic industry, wholesale, retail, transportation and services. Figure 4-3 illustrates the spatial distribution of the zonal jobs per class. Table 4-5 lists the variables used in the specification of the utility equation with their description.

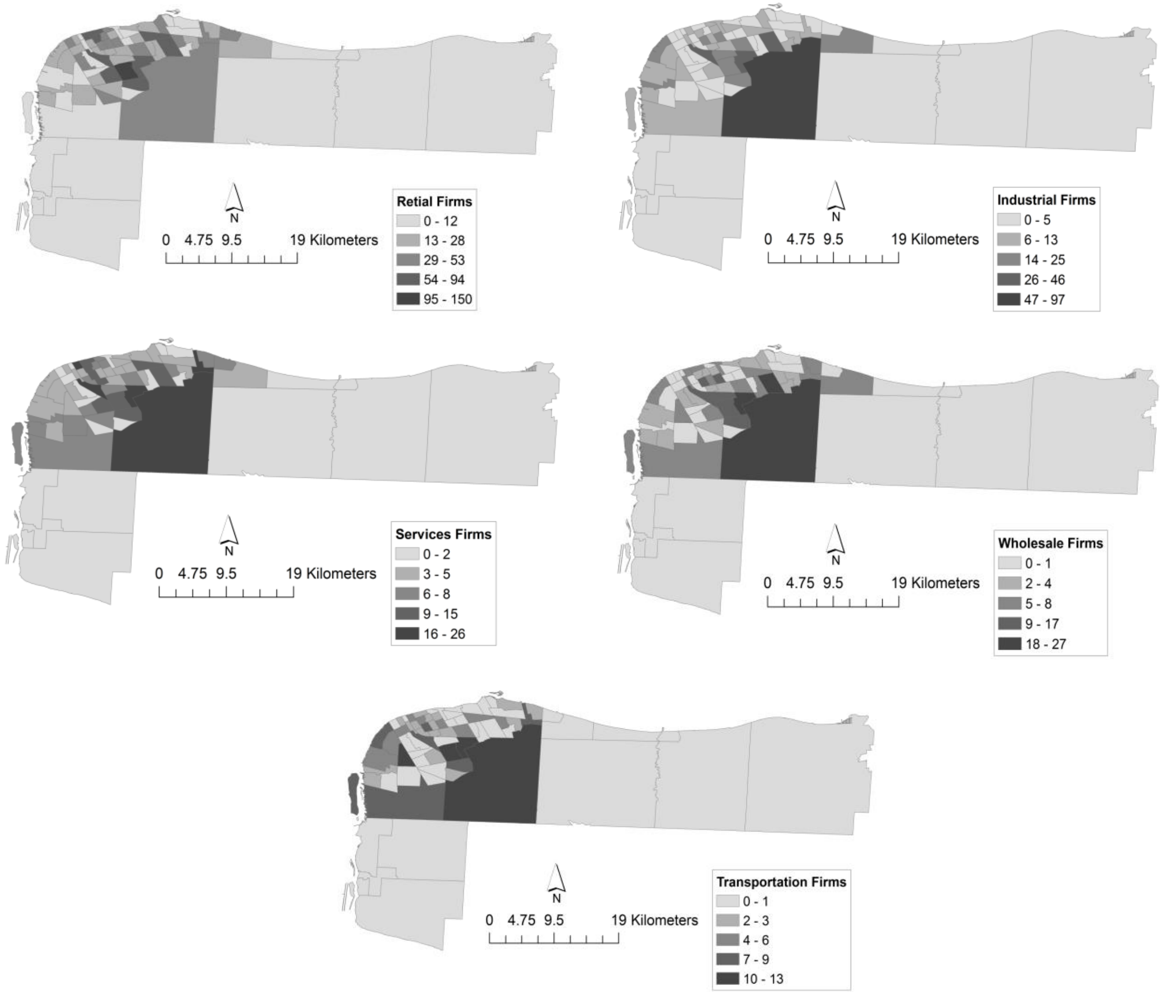


Figure 4 - 3: Spatial distribution of jobs by economic sector in Windsor in 2011

Table 4 - 5: Description of Explanatory Variables

Variable	Description
$\ln(IND_i)$	The natural log of the total number of Basic Industrial jobs in each zone *i*
$\ln(WHL_i)$	The natural log of the total number of Wholesale jobs in each zone *i*
$\ln(TRA_i)$	The natural log of the total number of Transportation jobs in each zone *i*
$\ln(RET_i)$	The natural log of the total number of Retail jobs for each zone *i*
$\ln(SER_i)$	The natural log of the total number of Service jobs in each zone *i*
$\ln(AREA_i)$	The natural log of the area in kilometers squared for each zone *i*
$HWYPRO_i$	1 if the zone *i* is in close proximity with highways, 0 otherwise
$\ln(IND_i) \times C$	Interaction term between the natural log of basic industrial jobs in each zone *i* and Cars
$\ln(IND_i) \times L$	Interaction term between the natural log of basic industrial jobs in each zone *i* and Light Duty Trucks
$\ln(IND_i) \times M$	Interaction term between the natural log of industrial jobs in each zone *i* and Medium Duty Trucks
$\ln(WHL_i) \times M$	Interaction term between the natural log of Wholesale jobs in each zone *i* and Medium Duty Trucks
$\ln(TRA_i) \times L$	Interaction term between the natural log of Transportation jobs in each zone *i* and Light Trucks
$\ln(TRA_i) \times M$	Interaction term between the natural log of Transportation jobs in each zone *i* and Medium Duty Trucks
$\ln(RET_i) \times H$	Interaction term between the natural log of Retail jobs in each zone *i* and Heavy Duty Trucks
$\ln(AREA_i) \times C$	Interaction term between the natural log of Area of each zone *i* and Cars
$\ln(SER_i) \times L$	Interaction term between the natural log of Service jobs in each zone *i* and Light Trucks
$HWYPRO_i \times C$	Interaction terms between Light Trucks and zones in close proximity with highways
$CBD_i \times C$	Interaction terms between Cars and the distance to Central Business District for each zone in kilometres
$ZONE_i$	Location dummy variables: 1 if Commercial Vehicle is registered in census tract *i*, 0 otherwise (see Figure 4-4 for the list of tracts *i*)

The natural logarithm is applied to the calculated zonal jobs to capture the non-linear linear effect between the size of jobs in a given zone and the probability of finding a particular type of CV in that zone. This transformation produced more stable results in the models.

The first a priori expectation is that zones with basic industry firms will be more prone to housing medium and heavy vehicles, other things being equal. On the other hand, these zones would be less likely to house cars and light trucks since their use to deliver goods is not common. When it comes to the wholesale industry, zones with firms pertaining to this industry are more likely to house various types of commercial vehicles. However, it is expected that firms from this industry will be more affiliated with medium and heavy trucks. Likewise, we expect to see a strong affiliation between zones housing the transportation industry and the prevalence of heavy trucks. At the same time, this industry is expected to show preference to other types of commercial vehicles albeit we do not expect the same influence as in the case of heavy trucks. Zones housing the retail industry are more likely to be affiliated with the prevalence of cars and light trucks. In contrast, we do not expect retail firms to own medium and heavy trucks in general. The same could be said about firms from the services sector.

Next, we hypothesize that zones with larger land areas will house more commercial vehicles, other things being equal. The rationale here is that smaller zones are typically associated with high population densities and not as many business establishments especially in a sprawled city like Windsor. Similarly, zone closer to the CBD are expected to be more specialized in terms of their firm population (e.g. zones housing services jobs) and as such are more likely to house cars and light commercial vehicles, other things being equal. Furthermore, zones in close proximity to highways and interchanges are also expected to attract more commercial vehicles since locations in proximity to transportation infrastructure are normally considered prime sites for business establishments to locate. Finally, location specific variables were introduced to two separate models (i.e. model 2 and model 4) to capture the added utility or disutility associated with the existence of commercial vehicles in particular zones. Figure 4-4 highlights the spatial variables used in the specification of the choice utilities.

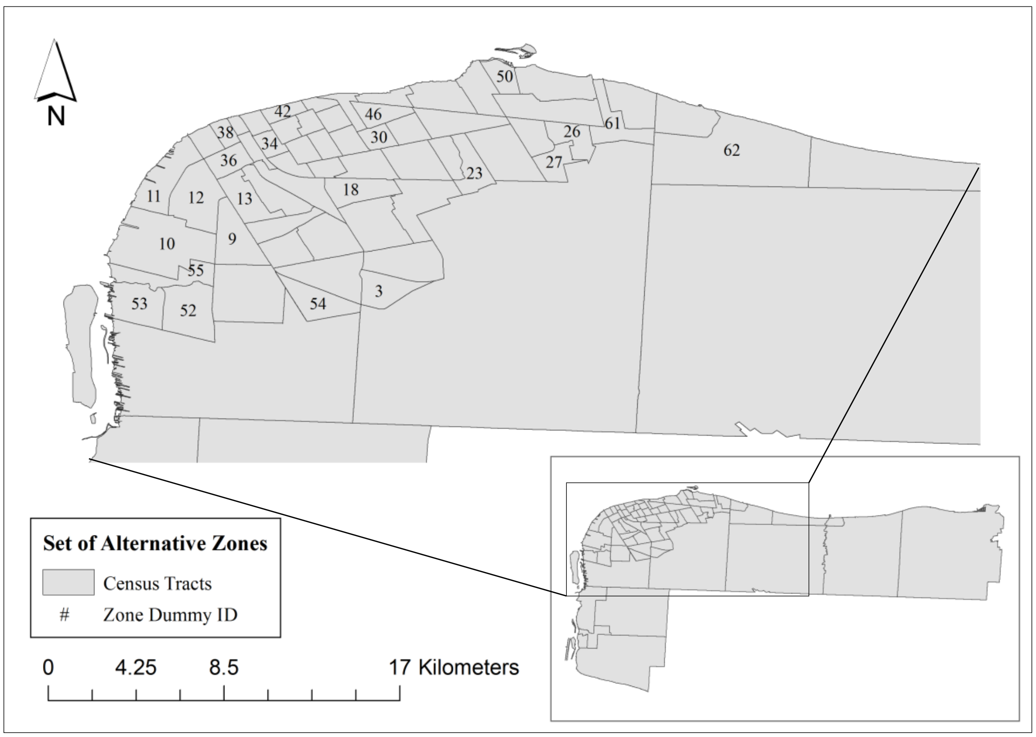


Figure 4 - 4: Spatial Variables Used in Model Specification

4.2.1.2 Spatial Effects

Two separate models are developed (i.e. MNL 3 and MXL 3) to account for the spatial autocorrelation in the modeled data. The latter effect was confirmed by the *Moran's I* results shown in section 3.2.3.1. A spatial variable S_i for zone i is introduced to the logit model to control for the impact of having vehicles of the same class type locating in neighbouring zones *j* (i.e. sharing the same edge with zone *i*). The assumption is that spatial autocorrelation will increase the probability of finding the same type of vehicle class in zone *i*, other things being equal. Quantitatively, S_i can be expressed as follows:

$$S_i = \sum_{j=1}^{n} \delta_{ij}\, C_{j/q} \tag{4.7}$$

where, $\delta_{ij} = 1$ when zone *i* and its neighboring zones *j* are in contiguity (i.e. share an edge), 0 otherwise. $C_{j/q}$ is the count of vehicle of type q in neighboring zones j. For example, if the modeled commercial vehicle type q is a light duty truck, then the count of light duty trucks in neighbouring zones *j* are used to represent $C_{j/q}$ in equation 4.7.

4.2.2 Count Modeling Approach

By considering the count nature of the available data, the Poisson and Negative Binomial (NB) regression models can be used to predict the number of commercial vehicles per census tract. Given that the commercial vehicle ownership data is non-negative, the Poisson regression model can be used as a starting point to predict the number of commercial vehicles per census tract. The parameter estimates for the count models is performed in the SAS 9.2 software.

In this model, the probability of specific zone *i* having count (y_i) commercial vehicles (where y_i is non-negative integer) is given by:

$$P(y_i|X_i) = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!} \tag{4.8}$$

where λ_i is the expected number of vehicles per zones, $E[y_i]$. The Poisson regression model is estimated by specifying the Poisson parameter λ_i as a function of explanatory variables as illustrated in equation 4.9.

$$\lambda_i = E[y_i] = \exp(\beta_1 X1_i + \beta_2 X2_i + \cdots + \beta_k Xk_i) \tag{4.9}$$

X_i represent a given explanatory variables that is associated with zone i and $\beta's$ are estimable parameters. Based on equation 4.9, λ_i is the mean of the Poisson distribution conditional on explanatory variable X_i. One of the main properties of the Poisson distribution is the equidispersion property that is the equality of the mean and the variance. Therefore, if the variance is over or under-dispersed relative to the mean, the estimated parameters β from the Poisson regression will be biased.

To remedy the equidispersion constraint problem the negative binomial (Poisson-gamma or NB) model are used. The NB model assumes that the Poisson parameter follows a gamma probability distribution. This model is derived by rewriting the Poisson parameter for each zone *i* as follows:

$$\lambda_i = E[y_i] = \exp(\beta_1 X1_i + \beta_2 X2_i + \cdots + \beta_k Xk_i + \varepsilon_i)$$

$$\lambda_i = \exp(\beta_1 X1_i + \beta_2 X2_i + \cdots + \beta_k Xk_i) \exp(\varepsilon_i) \tag{4.10}$$

where $\exp(\varepsilon_i)$ is a gamma-distributed error term with mean 1 and variance α. The addition of this term allows the variance to differ from the mean such that:

$$VAR[y_i] = E[y_i]\big[1 + aE[y_i]\big] \tag{4.11}$$

This yields the resulting conditional probability:

$$P(y_i|\varepsilon) = \frac{\exp[-\lambda_i \exp(\varepsilon)][\lambda_i \exp(\varepsilon)]^{y_i}}{y_i!} \qquad (4.12)$$

4.2.2.2 Count Model Specification

For this analysis, given the number of vehicles per zone classified by their GVW, four separate models are estimated by each method to examine number of vehicles by class per zone. Following the MNL and the MXL models, CVs are classified as (i) Cars (GVW unknown); (ii) Light Duty Trucks (GVW 1-2); (iii) Medium Duty Trucks (GVW 3-6); and (iv) Heavy Duty Trucks (GVW 7-8). These are the dependent variables.

On the other hand, the independent variables considered, as in the case of MNL and MXL models are: the natural log of the total number of jobs by industry type, the natural log of area in squared kilometers of each zone, the equidistance distance from the centroid of zones to the CBD and whether zones are in proximity to highways. Furthermore, the a priori expectations for the count models developed are the same as the ones mentioned in section 4.2.1. Table 4-6 presents the coding of the variables used in the analysis, together with a summary of the a priori expectations and Table 4-7 provides the descriptive statistics for the variables.

Table 4 - 6: Independent Variables Coding and their Expected Signs

Variable	Prior Expectation
ln(IND$_i$)	Positive correlation with Medium and Heavy Duty Trucks
ln(WHL$_i$)	Positive correlation with Medium and Heavy Duty Trucks
ln(TRA$_i$)	Positive correlation with Heavy Duty Trucks
ln(RET$_i$)	Positive correlation with Cars and Light Duty Trucks
ln(SER$_i$)	Positive correlation with Cars and Light Duty Trucks
ln(AREA$_i$)	Positive correlation with Cars, Light, Medium and Heavy Duty Trucks
HWYPRO$_i$	Positive correlation with Cars, Light, Medium and Heavy Duty Trucks
CBD$_i$	Positive correlation with Cars and Light Duty Trucks

Table 4 - 7: Independent and Dependent Variables Descriptive Statistics

Variables	Minimum	Maximum	Median	Mean	Standard Deviation
Cars	1	149	12	23.05	30.02
Light Duty Trucks	4	1032	45	89.13	150.96
Medium Duty Trucks	0	241	6	15.65	34.55
Heavy Duty Trucks	0	247	4	15.23	36.95
ln(IND$_i$)	0	8.58	4.28	4.29	2.31
ln(WHL$_i$)	0	6.44	0	1.65	2.18
ln(RET$_i$)	0	7.28	4.65	4.32	1.85
ln(TRA$_i$)	0	6.51	0	2.04	2.34
ln(SER$_i$)	4.70	8.97	6.31	6.44	0.87
ln (AREA$_i$)	-0.85	4.71	0.91	0.89	0.92
CBD$_i$	0.23	11.29	1.14	2.02	2.26

N= 62 Zones

CHAPTER 5: RESULTS AND DISCUSSION

5.1 Population Synthesis

For each synthetic population generated, the total number of vehicles at each zone is calculated by summing the number of vehicles owned by each establishment in the zone. Then, the results are validated by comparing the zonal aggregates to the Polk Data. For this research, the R-squared value was used to assess the correlation between the synthesized and the actual number of vehicles per zone. Consequently, for the different PQ cut-off values tested, a set of 10 populations were generated. After calculating the zonal aggregates for each synthesized population, an average was estimated from the 10 different populations and compared to the Polk data. Table 5-1 presents the R-squared achieved, where, PQ greater than or equal to 0.7, is associated with the highest fit with R-squared value of 0.88.

Table 5 - 1: Results of Comparisons between Synthetic and the Actual Number of vehicles, using the R^2

	Number of Establishments	**Percent from Total Population**	**R^2**
PQ ≥ 0.2	5906	59%	0.83
PQ ≥ 0.3	5048	51%	0.86
PQ ≥ 0.4	4452	45%	0.86
PQ ≥ 0.5	4148	42%	0.86
PQ ≥ 0.6	3744	38%	0.87
PQ ≥ 0.7	**3478**	**35%**	**0.88**
PQ ≥ 0.8	3022	30%	0.87
PQ ≥ 1.0	2795	28%	0.87

Hence, a total of 3,478 establishments are determined to engage in commercial vehicle activities, representing 35% of all establishments in the Windsor CMA. Although a slightly larger share was reported for the Greater Toronto Area (43%) (MITL, 2010) and Edmonton (49%), a 35% of establishments engaging in commercial activities for Windsor is an acceptable share given its overall size when compared to mega regions like Toronto and Edmonton. The variations (which is due to the stochastic nature of the CO method) between the total numbers of vehicles estimated from each run were insignificant as discerned by the descriptive statistics shown in Table 5-2. The total number of vehicles can also be found in graphical form in Figure 5-1. The estimated standard deviation, variance and range indicate the results achieved from the 10 runs are consistent with each other.

Table 5 - 2: Total Number of Vehicles Summary Statistics

Summary Statistics	**CO**
Mean	10,005
Median	9,951
Standard Deviation	256.412
Sample Variance	65,749.56
Kurtosis	-0.357
Skewness	0.231
Range	865
Minimum	9,591
Maximum	10,456
Count	10

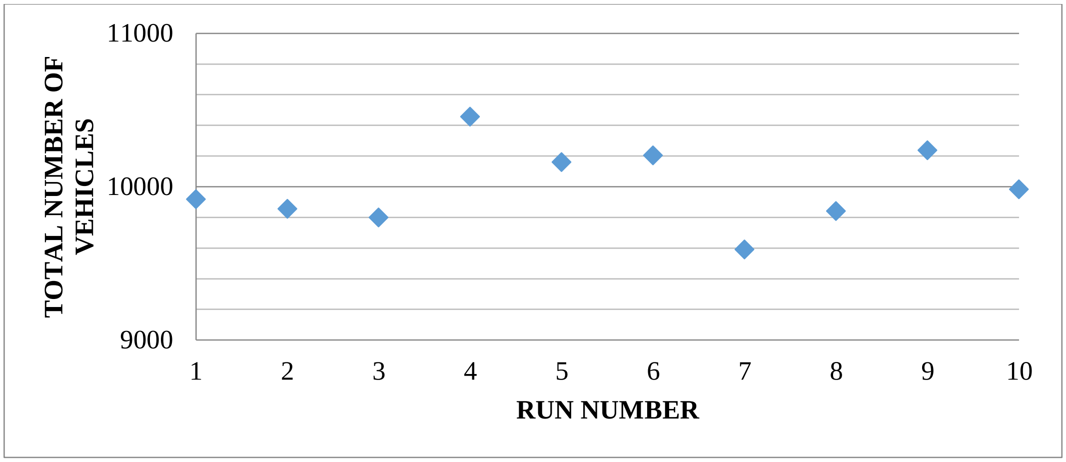


Figure 5 - 1: Total Number of Vehicles Comparison for 10 Establishment Population Synthesized

These establishments are then classified into 5 major industrial groups, that is, retail, basic industry, services, wholesale and transportation accounting for 55%, 19%, 12%, 8% and 6% of establishments that engage in commercial activities, respectively. As mentioned earlier, not all industries will be dependent on the same types of vehicles for their business transportation activities. Therefore, the diversity in the clustering pattern of these industries over space would result in varied spatial distribution of CVs in the various zones across the city.

Figure 5-2 provides a comparison between the synthesized zonal aggregates versus the Polk data, while Figure 5-3 highlights the spatial distribution of these vehicles. The synthesized and the actual number of vehicles per zone have similar spatial distributions. As the trend in Figure 5-2 suggests, the Polk totals are very close to the synthesized totals per zone.

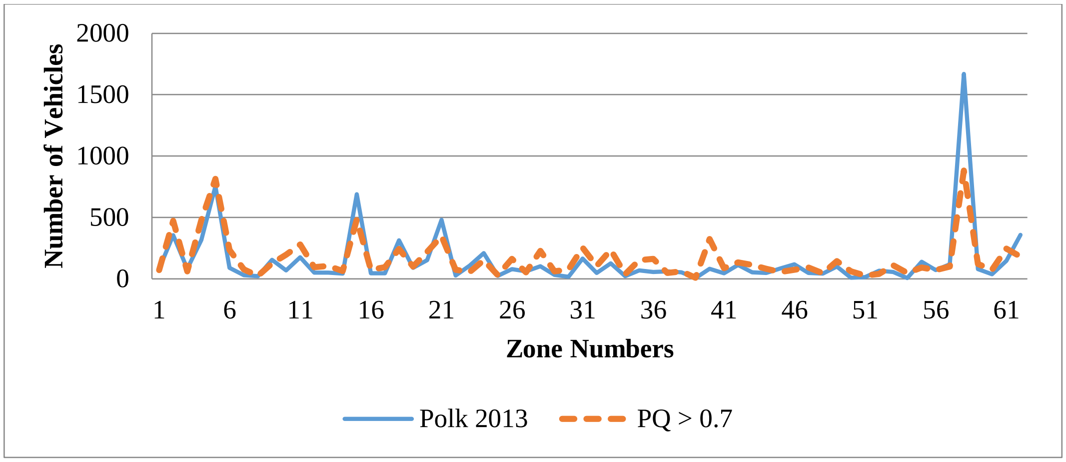


Figure 5 - 2: Comparison of the zonal aggregates of the synthesized population against the Polk Data

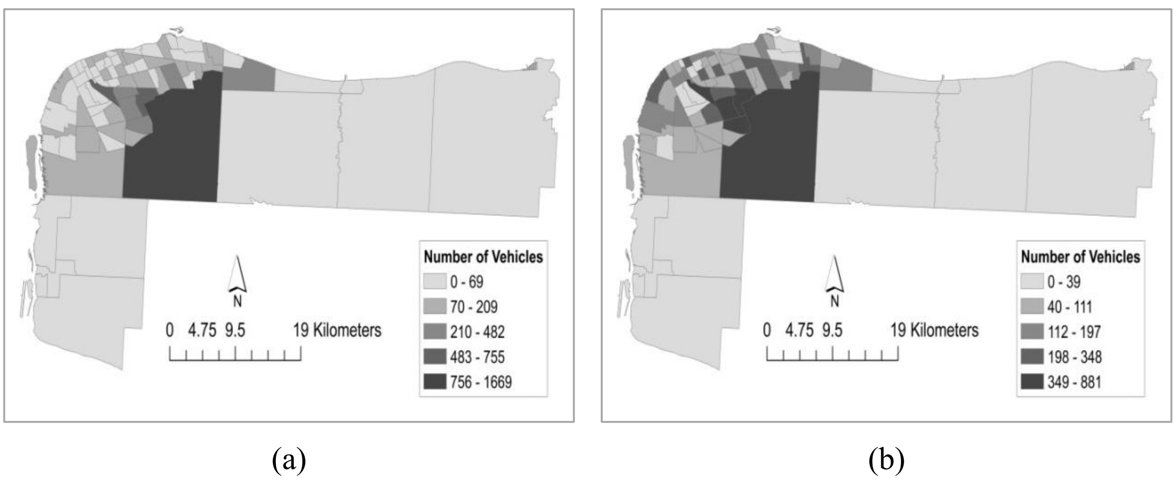


Figure 5 - 3: Spatial distribution of (a) Polk zonal aggregates and (b) Synthesized zonal aggregates in Windsor in 2013

Accordingly, to insure that the utilized Polk data records do not include vehicles that are registered but do not operate from the zone, the following constraints were imposed:

1- If the synthesized aggregates are greater than or equal to the Polk data count in a zone, then the Polk data are used to represent the zonal vehicles in the analysis.

2- If the synthesized aggregates are less than the Polk data count in a zone, then the synthesized aggregates are used to represent the number of vehicles in the zone. Consequently, the registered Polk vehicles in the zone are scaled down proportionally by vehicle class such that the total number of Polk records equals the synthesized aggregate in the zone.

Table 5-3 illustrates the Polk totals versus the synthesized by zones. Out of the 62 zones, 17 zones were found to have zonal totals greater than the synthesized totals. These zones are highlighted and demarked in bold.

After adjusting the Polk data, the total number registered vehicles in the zones comprising the CMA was reduced from 8869 to 7136 vehicles, a 19.5 percent reduction from the original Polk data records. This adjustment eliminates any concerns of vehicles registered to establishments in a census tract but operating from another location, an inherited problem in any acquired Polk dataset. This is the case since the synthesized population is a true representation of the vehicles registered and operated from the same zone. Hence, the Polk data is used to examine the commercial vehicle ownership location to study the spatial prevalence of the various types of commercial vehicles in a given zone.

Table 5 - 3: Synthesized Versus Polk Data Zonal Counts

CTUID	Polk Data	Synthesized	CTUID	Polk Data	Synthesized
Zone 1	**81**	**71**	Zone 32	47	105
Zone 2	355	472	Zone 33	127	237
Zone 3	**82**	**62**	Zone 34	23	38
Zone 4	317	479	Zone 35	69	151
Zone 5	755	814	Zone 36	55	164
Zone 6	90	236	**Zone 37**	**60**	**47**
Zone 7	29	76	Zone 38	53	61
Zone 8	23	25	Zone 39	7	10
Zone 9	**156**	**125**	Zone 40	83	325
Zone 10	68	197	Zone 41	45	86
Zone 11	175	280	Zone 42	112	133
Zone 12	50	94	Zone 43	53	114
Zone 13	50	106	Zone 44	48	83
Zone 14	43	63	**Zone 45**	**85**	**56**
Zone 15	**690**	**483**	**Zone 46**	**118**	**75**
Zone 16	45	75	Zone 47	48	93
Zone 17	46	96	Zone 48	44	46
Zone 18	**313**	**243**	Zone 49	101	144
Zone 19	90	100	Zone 50	10	62
Zone 20	153	220	Zone 51	15	27
Zone 21	**482**	**348**	**Zone 52**	**65**	**39**
Zone 22	26	73	Zone 53	56	111
Zone 23	**108**	**55**	Zone 54	6	47
Zone 24	**209**	**152**	**Zone 55**	**139**	**93**
Zone 25	24	31	**Zone 56**	**72**	**71**
Zone 26	80	163	**Zone 57**	**119**	**99**
Zone 27	**60**	**54**	**Zone 58**	**1669**	**881**
Zone 28	103	229	Zone 59	80	122
Zone 29	32	62	Zone 60	37	73
Zone 30	17	78	Zone 61	146	246
Zone 31	166	255	**Zone 62**	**359**	**180**

5.2 Modeling the Spatial Distribution of Commercial Vehicles

5.2.1 Discrete Choice Models

After adjusting the Polk data, a 40 percent sample was used for the analysis to reduce the computation time required to perform the models estimation. This resulted in a total of 2920 commercial vehicles that were checked and validated to be a good representative sample of the entire CV population.

The estimation results of the Multinomial Logit (MNL) and the Mixed Logit (MXL) models are presented in Table 5-4, with all of the coefficients and their corresponding t-statistics. The results from the four models are fairly consistent across most of the estimated parameters. Both models achieve higher ρ^2 value when the zone-specific dummy variables are introduced. The two MNL models have an adjusted ρ^2 value of 0.181 and 0.203, respectively, while, the models pertaining to MXL have an adjusted ρ^2 value of 0.185 and 0.204, respectively. As far as the achieved ρ^2 values, there is no significant difference between the MNL and MXL models. Therefore, the discussion of the factors explaining the choice probability is focused on the MNL model with zone-specific dummy variables.

Based on the estimated coefficients, zones with higher number of basic industry jobs tend to be strongly affiliated with heavy duty trucks, as discerned from the positive and significant parameter of the $ln(IND_i)$ variable. Those zones are also less prone to give rise to the prevalence of cars, light and medium duty trucks as discerned from the parameters of the three interaction terms $ln(IND_i) \times C$, $ln(IND_i) \times M$ and $ln(IND_i) \times H$.

Table 5 - 4: MNL and MXL Model Estimation Results

	MNL Model 1		MNL Model 2		MXL Model 1		MXL Model 2	
Parameter	**Beta**	**t-stat**	**Beta**	**t-stat**	**Beta**	**t-stat**	**Beta**	**t-stat**
$Ln(IND_i)$	0.265	6.36	0.178	3.97	0.293	6.42	0.190	4.07
$Ln(WHL_i)$	0.145	9.56	0.219	11.91	0.184	8.38	0.230	10.53
$Ln(RET_i)$	0.173	9.98	0.127	6.14	0.251	10.29	0.137	6.28
$Ln(TRA_i)$	0.098	3.97	0.160	6.06	0.113	3.65	0.175	5.59
$Ln(SER_i)$	0.168	5.41	0.149	3.39	0.181	5.26	0.148	3.30
$Ln(IND_i) \times C$	-0.349	-7.32	-0.357	-7.36	-0.398	-7.48	-0.375	-7.35
$Ln(IND_i) \times L$	-0.161	-3.75	-0.183	-4.20	-0.196	-4.13	-0.197	-4.34
$Ln(IND_i) \times M$	-0.168	-2.87	-0.203	-3.36	-0.215	-3.33	-0.215	-3.46
$ln(WHL_i) \times M$	0.095	2.09	0.097	2.03	0.250	2.81	0.102	2.06
$ln(TRA_i) \times L$	-0.075	-2.64	-0.054	-1.92	-0.071	-2.27	-0.054	-1.82
$ln(TRA_i) \times M$	-0.131	-2.86	-0.104	-2.28	-0.148	-2.83	-0.106	-2.21
$ln(RET_i) \times H$	-0.136	-2.77	-0.123	-2.7	-0.130	-2.31	-0.118	-2.51
$Ln(AREA_i)$	0.125	5.51	0.116	4.72	0.120	4.76	0.120	4.70
$Ln(AREA_i) \times C$	-0.128	-2.24	-0.172	-2.92	-0.158	-2.48	-0.183	-2.97
$HWYPRO_i$	0.340	5.97	0.267	3.56	0.277	4.53	0.253	3.29
$HWYPRO_i \times C$	-0.556	-4.36	-0.512	-3.92	-0.584	-4.27	-0.523	-3.90
$CBD_i \times C$	0.081	3.46	0.110	4.20	0.088	3.47	0.109	4.08
$ZONE_3$	--	--	0.969	4.10	--	--	1.009	4.18
$ZONE_9$	--	--	-0.688	-4.03	--	--	-0.713	-3.94
$ZONE_{10}$	--	--	-1.046	-4.98	--	--	-1.066	-4.95
$ZONE_{11}$	--	--	-0.398	-2.85	--	--	-0.385	-2.68
$ZONE_{12}$	--	--	-0.534	-2.15	--	--	-0.506	-2.01
$ZONE_{13}$	--	--	-1.250	-5.32	--	--	-1.205	-5.07
$ZONE_{18}$	--	--	-0.298	-2.46	--	--	-0.282	-2.27
$ZONE_{23}$	--	--	-0.988	-4.32	--	--	-0.976	-4.20
$ZONE_{26}$	--	--	0.408	2.02	--	--	0.394	1.91
$ZONE_{27}$	--	--	0.860	3.18	--	--	0.895	3.25
$ZONE_{30}$	--	--	-0.806	-2.07	--	--	-0.838	-2.14
$ZONE_{34}$	--	--	-1.816	-5.20	--	--	-1.776	-5.06
$ZONE_{36}$	--	--	-0.533	-2.27	--	--	-0.523	-2.20
$ZONE_{38}$	--	--	0.478	2.12	--	--	0.466	2.03
$ZONE_{42}$	--	--	-0.819	-4.19	--	--	-0.744	-3.69
$ZONE_{46}$	--	--	0.753	3.55	--	--	0.733	3.39
$ZONE_{50}$	--	--	-1.296	-2.55	--	--	-1.332	-2.60
$ZONE_{52}$	--	--	0.783	2.93	--	--	0.806	2.97
$ZONE_{53}$	--	--	0.444	2.12	--	--	0.421	1.98
$ZONE_{54}$	--	--	-2.036	-3.47	--	--	-2.067	-3.51
$ZONE_{55}$	--	--	0.909	4.72	--	--	0.903	4.60
$ZONE_{61}$	--	--	0.764	4.82	--	--	0.792	4.89
$ZONE_{62}$	--	--	-0.443	-3.13	--	--	-0.435	-2.97
Standard Deviation								
			$Ln(WHL_i)$		0.249	5.7	0.128	2.33
			$Ln(RET_i)$		0.278	8.38	--	--
			$Ln(TRA_i)$		0.147	2.27	0.143	2.58
			$Ln(WHL_i) \times M$		0.500	3.67	--	--
Log Likelihood	-6723.55		-6723.55		-6723.55		-6723.55	
Full Log Likelihood	-5509.13		-5358.31		-5477.286		-5355.27	
ρ^2	0.181		0.203		0.185		0.204	

These results are expected given the nature of these industries (e.g. manufacturing and mining) which relies heavily on larger trucks for their business operation.

In a similar vein, the results from the interaction term $ln(WHLi) \times M$ suggest that zones housing wholesale firms are strongly affiliated with medium duty trucks. As we anticipated, the wholesale industry is likely to own cars and light trucks (as discerned by the positive sign of the $ln(WHL\mathrm{i})$ parameter when compared to the parameters of the interaction terms $ln(WHLi) \times M$ but their dependency on larger trucks is evident in the model. However, the effect of the wholesale jobs vary across the modeled observations given the significance of the standard deviation of the $ln(WHL_i)$ and $ln(WHL_i) \times M$ in the case of the MXL model. A possible explanation for the randomness effect in these parameter could be attributed to the nature of the commercial vehicle ownership for establishments in the wholesale industry.

In the case of the transportation industry, zones housing firms of this type are more prone to give rise to the presence of cars and heavy trucks as can be discerned from the positive and significant parameter of the $ln(TRA_i)$ variable. Also, those zones are less prone to give rise to the prevalence of light and medium duty trucks as discerned from the parameters of the two interaction terms $ln(TRA_i) \times M$ and $ln(TRA_i) \times H$. It is worth noting that the effects of the existence of transportation jobs in the zone vary across the modeled CV observations, as confirmed by the significant standard deviation of the variable in the MXL model.

Zones housing retail trade firms tend to be affiliated with cars, light and medium duty trucks. By comparison, these zones repels heavy duty trucks as can be deduced from the sign of the interaction term $ln(RET_i) \times H$ parameter which is negative and significant relative to the positive and significant $ln(RET_i)$ parameter. Similar to the case of transportation jobs, the impact of the retail jobs on the location choice of commercial vehicles vary across the modeled CV observations as discerned from the standard deviation of the $ln(RET_i)$ variable in the MXL model.

Zones housing services sectors firms are prone to give rise to the presence of all types of vehicles as can be discerned from positive and significant parameter of the $ln(SER_i)$ variable. As expected, larger zones tend to have a positive influence on the presence of commercial vehicles. However, these zones are less prone to give rise to the prevalence of cars as discerned from the parameters of the interaction terms $ln(Area_i) \times C$. While, zones closer to the CBD tend to have a positive influence on the presence of cars, other things being equal.

It is found, zones in proximity to highways tend to be more associated with the presence of heavier vehicles. On the other hand, those zones are less prone to having cars. These results are sensible especially that locations in proximity to highways and interchanges (i.e. highway ramps) are attractive to manufacturing and heavy industry firms that rely heavily on accessibility (Maoh & Kanaroglou, 2009). These firms are more likely to own light, medium and heavy trucks for their goods movement activities.

Finally, a set of 62 location specific variables (i.e. 62 zones comprising firms that engage in commercial activities) were examined. These zone dummies indicate the added utility or disutility associated with the existence of commercial vehicles in particular zones. 23 of the tested location dummies were significant and were added to the final model as illustrated in Table 5-3 and Figure 5-4.

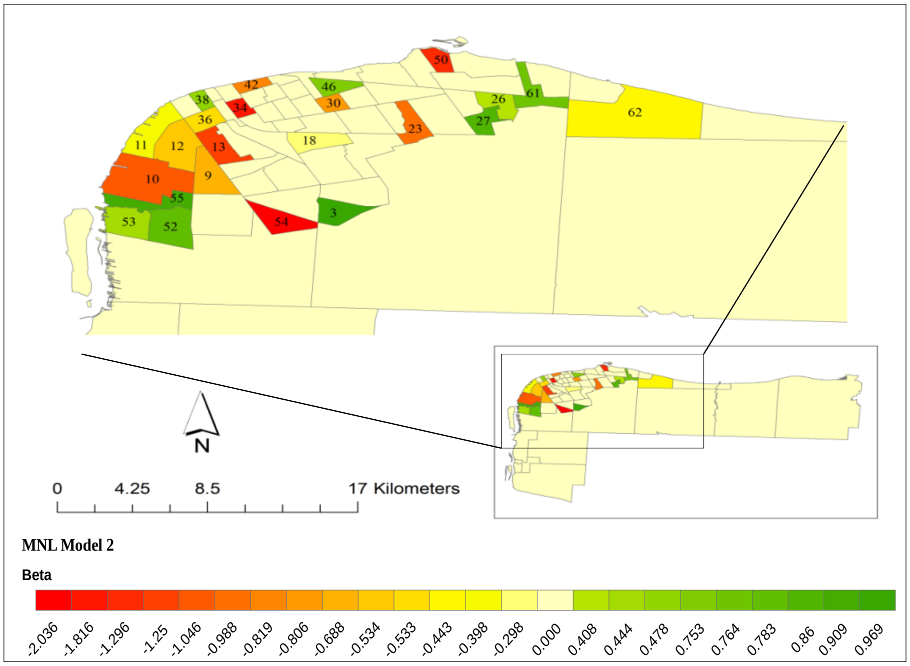


Figure 5 - 4: Spatial Variables Used in MNL Model 2 Specifications

These zones dummies suggests that factors other than those specified in the second MNL and MXL models are important in predicting the probability of finding a given commercial vehicle *c* in one of the 62 zones. Such unobserved factors could be related to land use and zoning by-laws.

To assess the predictive ability of the estimated MNL models (Model 1 and Model 2), the utility of each zone was calculated using the parameters reported in Table 13. At first the utility for each of the 10 alternatives per case is calculated using the estimated parameters. Next, an average utility per zone is calculated based on the zone id, since the choice set formed when estimating the choice model was based on 10 alternative zones as described in section 4.2.1. This resulted in 62 average utilities that were then used to calculate the probability of finding a CV in any of the 62 zones encompassing firms that engage in shipping and receiving goods and/or services. In this case, the sum of the 62 probabilities sum to 1. Using the calculated probabilities, the 2,920 CVs were distributed among the 62 zone and compared to the observed spatial distribution. Figure 5-5 and Figure 5-6 represent the calculated versus the observed CVs per zone for Model 1 and Model 2, respectively.

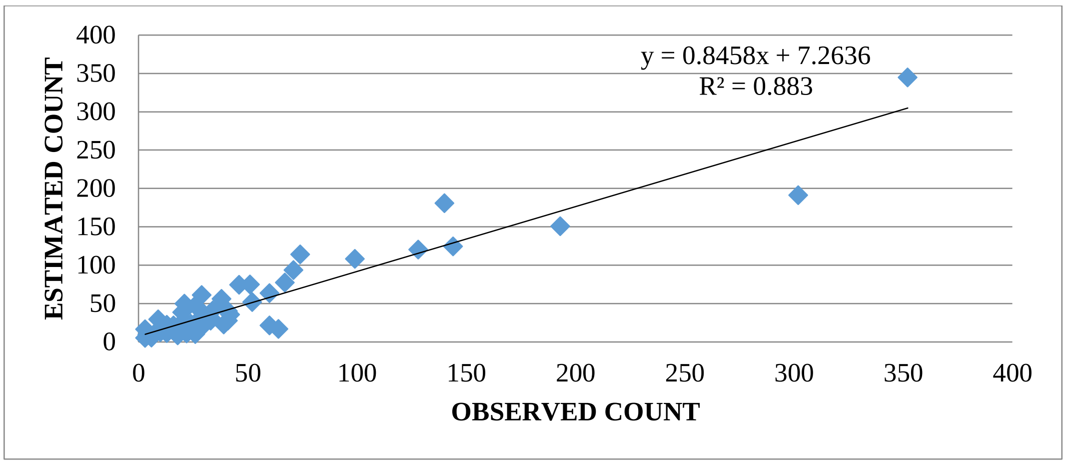


Figure 5 - 5: MNL Model 1 - Observed versus Estimated Numbers Commercial Vehicles

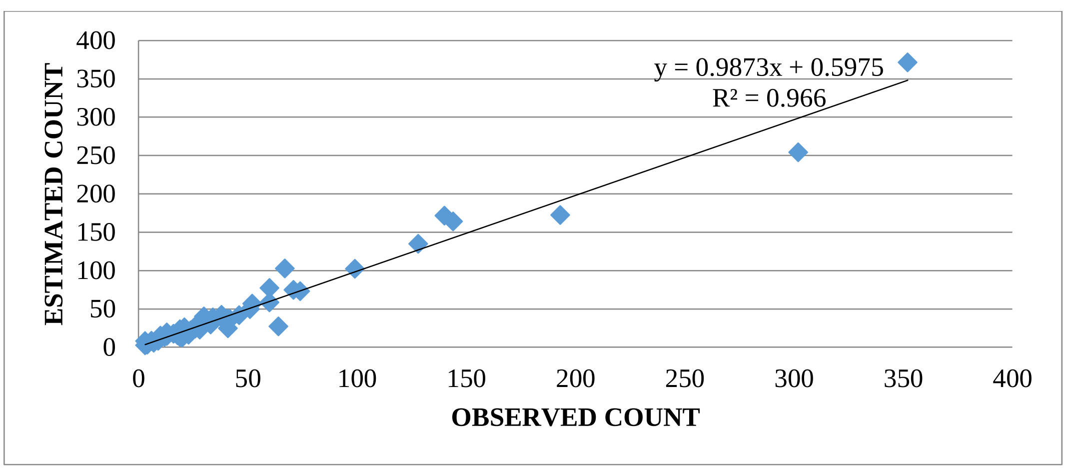


Figure 5 - 6: MNL Model 2 - Observed versus Estimated Number of Commercial Vehicles

As shown, the estimation results suggest that both models are superior in their predictive ability, correctly predicting 88 and 96.6 percent of the modeled CVs. However, a comparison between the two models suggests that other things being equal, the location specific variables contribute improving the prediction of the estimated models.

Further, both models are tested for their ability to predict the spatial distribution of each vehicle class. Table 5-5 illustrates the estimation results, where the correlations between the observed and the estimated number of commercial vehicles by a given class reveal that both models are capable of providing good predictions. Similarly, the estimated results illustrate that the predictive ability of the MNL model improves when the location specific variables are added. Figures that represent the calculated versus the observed CVs by class per zone for MNL Model 1 and MNL Model 2 are shown in Appendix C.

Table 5 - 5: Results of Comparisons between Estimated and the Actual Number of Vehicles by Class

	MNL Model 1	**MNL Model 2**
Cars	65%	77%
Light Duty Trucks	89%	97%
Medium Duty Trucks	86%	92%
Heavy Duty Trucks	75%	83%

5.2.2 Discrete Choice Models with Spatial Effects

The results from the spatial logit models are given in Table 5-6. For both models (MNL and MXL), the spatial parameter was found to be positive and significant as expected.

Table 5 - 6: MNL 3 and MXL 3 Model Estimation Results

	MNL Model 3		MXL Model 3	
Parameter	**Beta**	**t-stat**	**Beta**	**t-stat**
Ln(IND$_i$)	0.305	7.43	0.321	7.45
Ln(WHL$_i$)	0.118	7.59	0.136	7.04
Ln(RET$_i$)	0.165	9.51	0.232	10.03
Ln(TRA$_i$)	0.109	4.44	0.108	4.24
Ln(SER$_i$)	0.106	3.31	0.118	3.47
Ln(IND$_i$) × C	-0.383	-8.36	-0.414	-8.46
Ln(IND$_i$) × L	-0.223	-5.12	-0.237	-5.2
Ln(IND$_i$) × M	-0.187	-3.20	-0.225	-3.57
ln(WHL$_i$) × M	0.129	2.86	0.310	3.49
ln(TRA$_i$) × L	-0.089	-3.16	-0.085	-2.9
ln(TRA$_i$) × M	-0.128	-2.80	-0.140	-2.83
ln(RET$_i$) × H	-0.102	-2.05	-0.103	-1.9
Ln(AREA$_i$)	--	--	--	--
Ln(AREA$_i$) × C	--	--	--	--
HWYPRO$_i$	0.260	4.47	0.216	3.56
HWYPRO$_i$ × C	-0.473	-3.79	-0.496	-3.8
CBD$_i$ × C	0.090	3.82	0.095	3.83
S$_i$	0.00054	9.75	0.00053	8.91
Standard Deviation				
Ln(WHL$_i$)			0.169	4.06
Ln(RET$_i$)			0.258	8.17
Ln(TRA$_i$)			--	--
Ln(WHL$_i$) × M			0.540	4.16
Log Likelihood	-6723.55		-6723.55	
Full Log Likelihood	-5475.87		−5451.99	
ρ^2	0.186		0.189	

Also, the ρ^2 increased to 0.186 and 0.189 indicating that both model are an improvement over MNL Model 1 and MXL Model 1, respectively. In addition, the results from the estimated models are fairly consistent across most of the estimated parameters. Only variables $Ln(AREA_i)$ and $Ln(AREA_i) \times C$ were found to be insignificant in both models with the addition of the spatial variable. As such, these were dropped from the final specification of the models presented in Table 5-6.

Figure 5-7 represents the calculated versus the observed CVs per zone for MNL Model 3. As shown, the predictive ability of MNL Model 3 was calculated to be 90.95 percent. Also, the model's ability to predict the spatial distribution for each vehicle class was tested. Table 5-7 illustrates the estimation results. Similar to MNL 1 and MNL2, the results achieved are indicative of the models ability to provide good predictions. Figures that represent the calculated versus the observed CVs by class per zone for Model 3 are shown in Appendix C.

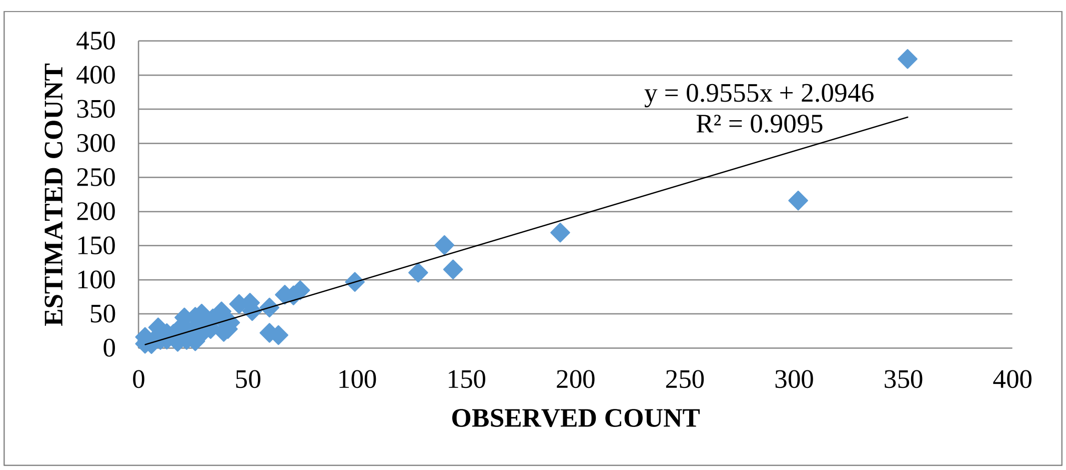


Figure 5 - 7: MNL Model 3 - Observed versus Estimated Number of Commercial Vehicles

Table 5 - 7: Results of Comparisons between Estimated and the Actual Number of Vehicles by Class

	MNL Model 3
Cars	68%
Light Duty Trucks	89%
Medium Duty Trucks	86%
Heavy Duty Trucks	80%

5.2.3 Count Models

Table 5-8 and 5-9 represent statistically significant parameters of explanatory variables for the Poisson and NB regression models, respectively. Separate models are estimated to examine commercial vehicle ownership by class, including, Cars, Light, Medium and Heavy Duty Trucks. The NB regression models show a dispersion factor different from 0 which implies overdispersion in the data; hence the Poisson models would not be appropriate to model vehicle ownership (Table 5-8). Indeed, this in line with the results achieved from the *Moran's I* statistics, where the Poisson Regression is suitable only if the number of vehicles in zone *i* is independent of the number of vehicles in any other neighbouring zone (i.e. no clustering).

The four NB models have an adjusted ρ^2 value of 0.100, 0.153, 0.188 and 0.114, respectively. The significance of the constants in the NB models, namely Model 1, Model 2 and Model 4, in Table 5-9 suggests that factors other than those specified in the model are important in explaining the process of CVs ownership. Such unobserved factors could be related to the ownership costs including the expenditures of purchasing, maintaining and operating the vehicle. These constant could be also capturing information that relates to the behavior of the business establishments.

Table 5 - 8: Poisson Regression Results of Commercial Vehicle by Class

	Model 1		Model 2		Model 3		Model 4	
Parameter	**Beta**	**P-value**	**Beta**	**P-value**	**Beta**	**P-value**	**Beta**	**P-value**
Constant	1.178	0.000	1.988	0.000	0.156	0.306	0.624	0.000
ln(IND$_i$)	-	-	0.087	0.000	-	-	-	-
ln(WHL$_i$)	-	-	0.176	0.000	0.433	0.000	0.201	0.000
ln(TRA$_i$)	0.127	0.000	0.075	0.000	-	-	0.255	0.000
ln(RET$_i$)	0.236	0.000	0.208	0.000	0.16	0.000	-	-
ln(SER$_i$)	-	-	-	-	-	-	-	-
ln(Area$_i$)	-	-	0.149	0.000	-	-	0.341	0.000
HWYPRO$_i$	-	-	-	-	0.510	0.000	-	-
CBD$_i$	0.109	0.000	-	-	-	-	-	-
Number of Observations	62		62		62		62	
Log Likelihood	2643.253		16645.235		1654.560		1662.062	
Full Log Likelihood	-358.910		-575.904		-222.626		-355.726	
ρ^2	0.474		0.779		0.702		0.612	

Note: Model 1 (Cars), Model 2 (Light Trucks), Model 3 (Medium Trucks) and Model 4 (Heavy Trucks)

Table 5 - 9: Negative Binomial Regression Results of Commercial Vehicle by Class

	Model 1		Model 2		Model 3		Model 4	
Parameter	**Beta**	**P-value**	**Beta**	**P-value**	**Beta**	**P-value**	**Beta**	**P-value**
Constant	1.500	0.000	2.233	0.000	0.454	0.073	0.617	0.009
ln(IND$_i$)	-	-	0.115	0.001	-	-	-	-
ln(WHL$_i$)	-	-	0.121	0.003	0.361	0.000	0.152	0.042
ln(TRA$_i$)	0.128	0.006	0.077	0.028	-	-	0.203	0.004
ln(RET$_i$)	0.162	0.001	0.155	0.000	0.159	0.004	-	-
ln(SER$_i$)	-	-	-	-	-	-	-	-
ln(Area$_i$)	-	-	0.149	0.044	-	-	0.591	0.002
HWYPRO$_i$	-	-	-	-	0.384	0.039	-	-
CBD$_i$	0.118	0.010	-	-	-	-	-	-
Dispersion	0.364		0.191		0.280		1.037	
Number of Observations	62		62		62		62	
Log Likelihood	2780.767		16943.815		1701.311		1837.025	
Full Log Likelihood	-221.396		-277.324		-175.874		-180.762	
ρ^2	0.100		0.153		0.188		0.114	

Note: Model 1 (Cars), Model 2 (Light Trucks), Model 3 (Medium Trucks) and Model 4 (Heavy Trucks)

Model 1 shows that Transportation and Retail jobs have a positive impact on Cars ownership levels as discerned from the positive and significant parameters of the *ln(TRA_i)* and *ln(RET_i)* variables. Also, the significant parameters of the CBD_i indicates that zones closer to the CBD tend to have a positive influence on the count of cars, other things being equal. Model 2 showed that the Light Duty Trucks class showed significant positive relationship with all job types except services jobs. In addition, it shows that larger zones tend to attract more Light Duty Trucks. Likewise, the Medium Duty Trucks class shows a positive statically significant relationship with Wholesale and Retail Jobs. Furthermore, the magnitude of the estimated parameters for this $HWYPRO_i$ variable suggests that Light Duty Trucks ownership levels are strongly affiliated with zones in close proximity to highways. Lastly, Model 4 indicates that Wholesale and Transportation jobs are prone to give rise to Heavy Duty Trucks ownership. Again *ln($Area_i$)* is significant, suggesting that larger zones tend to also increase Heavy Duty Trucks counts per zone.

To evaluate the predictive ability of each model, the probability of specific zone i having certain number of commercial vehicles was calculated for each class using the parameters in Table 5-9. Then the predict count per zone for each model were compared to the observed zonal counts. Table 5-10 presents the R-squared achieved; ranging approximately from 55 percent to 81 percent, suggesting that all the models are fairly good in their predictive ability. Furthermore, the results achieved from all the four NB regression models were summed and compared to the observed CVs per zone. Figure 5-8 represents the calculated R-squared, which was found to be 81 percent. Similarly, this reveal that overall the four NB models are capable of providing good predictions.

Table 5 - 10: Results of Comparisons between Estimated and the Actual Number of Vehicles

	R^2
Model 1: Cars	55%
Model 2: Light Duty Trucks	81%
Model 3: Medium Duty Trucks	79%
Model 4: Heavy Duty Trucks	81%

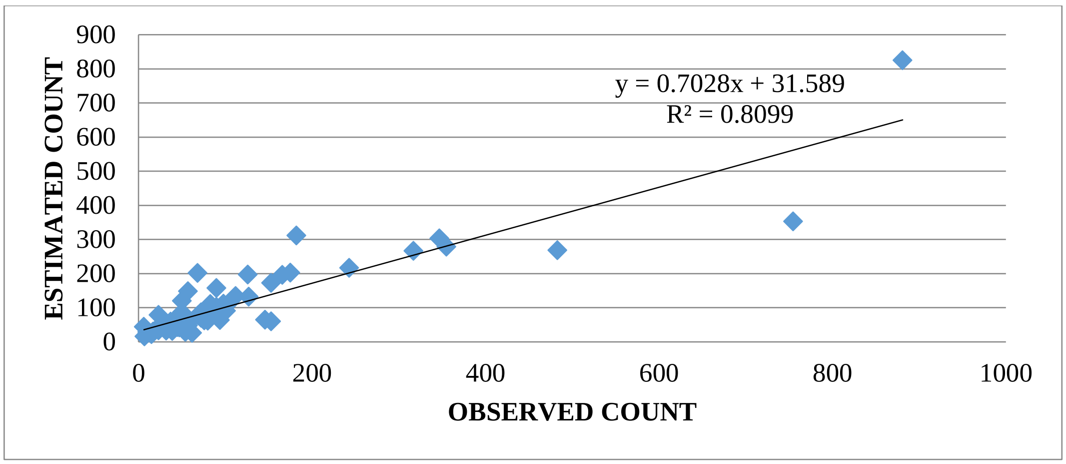


Figure 5 - 8: NB Models: Observed versus Estimated Number of Commercial Vehicles

5.2.4 Discrete Choice Models versus Count Models

To further assess the predictive ability of the two modeling techniques (i.e. discrete choice and count models) the root mean square error (RMSE) was used to measure the differences between values predicted by the models and the actual zonal counts observed. The RMSE can be represented as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \sum (Estiamted - Observed)^2}{N}} \tag{5.1}$$

where *N* is the total number of zones. Table 5-11 and 5-12 provide the estimated *RMSE* for the three MNL models and four NB models, respectively. As far as the achieved results, it is apparent that the three discrete choice models have lower RMSE and therefore are more capable of predicting accurate results.

Table 5 - 11 MNL Models RMSE Results

	MNL Model 1	MNL Model 2	MNL Model 3
Cars	5	4	5
Light Duty Trucks	13	6	14
Medium Duty Trucks	3	2	3
Heavy Duty Trucks	5	4	5
Total	22	12	19

Table 5 - 12: Negative Binomial Models RMSE Results

	RMSE
NB Model 1 - Cars	14
NB Model 2 - Light Duty Trucks	45
NB Model 3 - Medium Duty Trucks	11
NB Model 4 - Heavy Duty Trucks	23
Total	71

CHAPTER 6: CONCLUSIONS

The broad purpose of this thesis was to analyze and model commercial vehicle ownership location in the Windsor CMA. To date, all of the existing efforts have been solely focused on private car ownership modeling. Hence, identifying the factors that explain the prevalence of specific types of commercial vehicles at a certain location in a city will help devise more effective travel demand models. To this end, discrete choice models and count models were applied to investigate the determinants which lead to the spatial prevalence of a specific type of commercial vehicle at its residence zone within an urban area. Furthermore, the problem of data scarcity was tackled by relying on a readily available list of all the registered commercial vehicles in the year 2013 for the Windsor CMA. It should be noted that one of the main reasons for the underdevelopment of commercial vehicle ownership models in the literature is due to the lack of detailed commercial vehicle travel data. Results from the analysis were promising as they helped us understand, model and predict the spatial distribution of commercial vehicles by class and housing establishment location.

6.1 Summary of Empirical Results

6.1.1 Population Synthesis

This analysis provided the basis to justify the use of zonal data of all the registered commercial vehicles within the study area from sources like Polk to study the spatial prevalence (or assignment) of such vehicles in a given census tract. The main concern is that some CVs that are registered to the establishment in the census tract as given in the Polk data might not be physically located or operating from that location. The hypothesis is that synthesized aggregates are a true representation of the vehicles registered and operated from the same zone. Hence achieving comparable results would provide a basis to use the of Polk data in future research. Using combinatorial optimization techniques, a synthetic population of establishments that engage in commercial activities for the Windsor CMA is created. Then the total number of vehicles owned was assigned to each establishment of the population by linking each establishment directly to an establishment of the micro-sample attained from the survey responses. Using the Polk data as a validation source, comparisons against the synthesized total number of vehicles were made. The Polk data compared well with the synthesized zonal values with a correlation of 0.88. This indicated that the Polk data can be used to examine the spatial distribution of commercial vehicles.

6.1.2 Modeling the Spatial Distribution of Commercial Vehicles

MNL and MXL models were formulated to test the influence of various zonal and vehicle characteristics on the prevalence of commercial vehicles in traffic analysis zones (TAZs). The total number of jobs per zone, zone's land area, whether the zone is in close proximity to a highway and distance to CBD were found to have positive effect on the prevalence of commercial vehicles in specific zones. Moreover, the variability in the spatial distribution of the key types of commercial vehicles was explained based on the presence and dominance of certain industries in these zones. Furthermore, the predictive ability of the estimated MNL models (without and with zonal dummies) was calculated to be 88 and 96.5 percent, respectively. These results indicate that both models are superior in their predictive ability. Predictions pertaining to the MXL model were not conducted due to the complex nature of the calculations of the choice probabilities. Further, to account for the spatial autocorrelation in the modeled data, two separate models were developed (i.e. MNL 3 and MXL 3). For both models, the spatial parameter was found to be positive and significant. Also, the ρ^2 increased to 0.186 and 0.189 indicating that both model are an improvement over MNL Model 1 and MXL Model 1, respectively.

Also, to consider the count nature of the available data, the Poisson and NB regression models predict the number of commercial vehicles per class per census tract. Four separate models were estimated by each method to examine number of vehicles by class per zone. The NB regression models showed a dispersion factor different from 0 which implied overdispersion in the data. As a result the Poisson models deemed to be inappropriate to model commercial vehicle ownership in the case presented here. Variables such as number of jobs per zone, zone's land area, distance to CBD, and whether the zone is in close proximity to a highway were used to explain their effect on commercial vehicle ownership. The results of the models were conclusive, that the number of commercial vehicles per zone is dependent on the following:

- The higher the presence and dominance of certain industries, the higher the CVs count in the zone
- Zones closer to the CBD, suggest higher count of commercial cars in the zones
- Zones in close proximity to highways, suggest higher count of light duty trucks
- Larger the zone, the more light and heavy Duty Trucks counts per zone

Furthermore, the four NB models estimated were able to predict 55 to 81 percent right, showing that all the models are capable of providing good predictions.

6.2 Contributions and Policy Implications

The contribution of the thesis is as follows: 1) it sheds light on commercial vehicle ownership exclusively, something that has not previously been investigated, according to a search of the existing literature; 2) it overcomes data limitations that curtailed the development of predictive commercial vehicle ownership models; and 3) it utilizes different statistical techniques that are suitable to model the commercial vehicle ownership and performs a comparison regarding their predictive abilities.

From a transport policy perspective, modeling and understanding car ownership is essential to both urban transport and land-use planning, since vehicles influence trip generation and traffic on the transportation network. Previous research on urban commercial vehicle movements suggest that commercial vehicle, despite their lower count compared to passenger vehicles, play a more detrimental role when it comes to traffic congestion, energy consumption and delays on urban roads (Hunt and Stefan, 2007). According to Madar (2014), the number of vehicles owned by a particular establishment is an important determinant to predict the number of generated trips. Hence, as in the case of private vehicles, the study shows that the inclusion of commercial vehicle ownership is critical to developing travel demand forecasting models. Therefore, this research offers an in-depth understanding of the factors that affect the prevalence of different commercial vehicle classes in the urban context. The results obtained from the models, provides the basis for developing predictive transport demand models that could be integrated within a more comprehensive land use and transportation simulation model for the study area.

The approach devised in this thesis is also practice ready for two reasons: 1) the modeling effort in this paper makes use of publicly available commercial vehicle registration data and census information that exists for most North American jurisdictions, and 2) the approach can help overcome some of the data limitations that curtailed the development of predictive commercial vehicle ownership models in existing travel demand models. Further, the findings of this thesis have important implication on land use policy and transportation planning decisions, particularly in the Windsor CMA. It allows transportation planners to acknowledge the difference in temporal and geographic distributions of commercial vehicles when compared with private vehicles, to design and provide alternative freight-related and transportation policies that can shape future transportation master plans.

6.3 Study Limitations and Direction for Future Developments

Finally, as mentioned earlier, the models presented here offers a pioneering effort that uses publically available datasets to address an important gap in current research in term of the factors that could help researchers to understand the commercial vehicle ownership process. Limitations of this empirical work relate to the fact that there was no attempt made to explore the covariates of commercial vehicles ownership in the past. A business establishment's decision to purchase commercial vehicles could be also influenced by many relatively complex zonal and transportation characteristics, not included in this study's set of variables. These include things such as ownership costs including the expenditures of purchasing, maintaining and operating the vehicles owned. Furthermore, because the model uses aggregate data, these parameters cannot be generalized to apply at an individual establishment level. Nevertheless, these limitations could be rectified in future research. Further developments of this research could aim to investigate the spatial clustering effect on commercial vehicle ownership levels in the NB count models. Other development is to integrate this modelling system within a travel demand model for the Windsor CMA to have a more robust tool for evaluating impacts of urban freight transport measures and policies.

REFERENCES

Anand, N., Quak, H., van Duin, R. & Tavasszy, L. (2012). City Logistics Modeling Efforts: Trends and Gaps - A Review. *Procedia - Social and Behavioral Sciences*. 39. p.pp. 101–115.

Anon (2013). *City logistics research: a transatlantic perspective ; summary of the first EU-U.S. transportation research symposium*. Conference proceedings / Transportation Research Board 50. Washington, DC: Transportation Research Board.

Anowar, S., Eluru, N. & Miranda-Moreno, L.F. (2014). Alternative Modeling Approaches Used for Examining Automobile Ownership: A Comprehensive Review. *Transport Reviews*. 34 (4). p.pp. 441–473.

Area, H. (2010). *Estimating Urban Commercial Vehicle Movements in the Greater Toronto*. [Online]. Available from: http://sciwebserver.science.mcmaster.ca/~mitl/research/documents/Metrolinx_Report.pdf. [Accessed: 20 April 2016].

Arentze, T., Timmermans, H. & Hofman, F. (2007). Creating Synthetic Household Populations: Problems and Approach: Transportation Research Record: Journal of the Transportation Research Board: Vol 2014, No. *Transportation Research Record: Journal of the Transportation Research Board*. p.pp. 85–91.

Auld, J.A., Mohammadian, A. (Kouros) & Weis, K. (2009). Population Synthesis with Subregion-Level Control Variable Aggregation. *Journal of Transportation Engineering*. 135 (9). p.pp. 632–639.

Auld, J.A., Rashidi, T.H., Mohammadian, A. & Weis, K. (2010). Evaluating transportation impacts of forecast demographic scenarios using population synthesis and data transferability. In: *Proceedings of the 89th Annual Meeting of the Transportation Research Board (DVD), Washington, DC*. [Online]. 2010. Available from: http://131.193.170.173/travelbehavior/Papers/TRB10_Popsyn%20Forecasting%20v2.pdf. [Accessed: 16 February 2016].

Baldwin Hess, D. & Ong, P. (2002). Traditional neighborhoods and automobile ownership. *Transportation Research Record: Journal of the Transportation Research Board*. (1805). p.pp. 35–44.

Beckman, R.J., Baggerly, K.A. & McKay, M.D. (1996). Creating synthetic baseline populations. *Transportation Research Part A: Policy and Practice*. 30 (6). p.pp. 415–429.

Bhat, C., Paleti, R., Pendyala, R., Lorenzini, K. & Konduri, K. (2013). Accommodating Immigration Status and Self-Selection Effects in a Joint Model of Household

Auto Ownership and Residential Location Choice. *Transportation Research Record: Journal of the Transportation Research Board*. 2382. p.pp. 142–150.

Boerkamps, J., van Binsbergen, A. & Bovy, P. (2000). Modeling behavioral aspects of urban freight movement in supply chains. *Transportation Research Record: Journal of the Transportation Research Board*. (1725). p.pp. 17–25.

Central Intelligence Agency (2013). *The World Factbook*. [Online]. 2013. Available from: https://www.cia.gov/library/publications/the-world-factbook/geos/ca.html. [Accessed: 27 April 2016].

Chen, T.D., Wang, Y. & Kockelman, K.M. (2015). Where are the electric vehicles? A spatial model for vehicle-choice count data. *Journal of Transport Geography*. 43. p.pp. 181–188.

Chow, J.Y.J., Yang, C.H. & Regan, A.C. (2010). State-of-the art of freight forecast modeling: lessons learned and the road ahead. *Transportation*. 37 (6). p.pp. 1011–1030.

Chu, Y.-L. (2002). Automobile Ownership Analysis Using Ordered Probit Models. *Transportation Research Record: Journal of the Transportation Research Board*. 1805. p.pp. 60–67.

Dash, S., Vasudevan, V. & Singh, S. (2013). Disaggregate Model for Vehicle Ownership Behavior of Indian Households. *Transportation Research Record: Journal of the Transportation Research Board*. 2394. p.pp. 55–62.

De Jong, G. & Ben-Akiva, M. (2007). A micro-simulation model of shipment size and transport chain choice. *Transportation Research Part B: Methodological*. 41 (9). p.pp. 950–965.

De Jong, G., Gunn, H. & Walker, W. (2004). National and International Freight Transport Models: An Overview and Ideas for Future Development. *Transport Reviews*. 24 (1). p.pp. 103–124.

Eluru, N., Pinjari, A., Guo, J., Sener, I., Srinivasan, S., Copperman, R. & Bhat, C. (2008). Population Updating System Structures and Models Embedded in the Comprehensive Econometric Microsimulator for Urban Systems. *Transportation Research Record: Journal of the Transportation Research Board*. 2076. p.pp. 171–182.

Ferguson, M., Maoh, H., Ryan, J., Kanaroglou, P. & Rashidi, T.H. (2012). Transferability and enhancement of a microsimulation model for estimating urban commercial vehicle movements. *Journal of Transport Geography*. 24. p.pp. 358–369.

Frick, M. & Axhausen, K.W. (2004). *Generating synthetic populations using IPF and monte carlo techniques: Some new results*. In: 2004, Swiss Transport Research Conference.

Gingerich, Kevin, "Modelling Non-residential Real Estate Prices and Land Use Development in Windsor with Potential Impacts from the Windsor-Essex Parkway" (2013). *Electronic Theses and Dissertations.* Paper 4754.

Gómez-Gélvez, J. & Obando, C. (2013). Modeling Car Ownership in Urban Areas of Developing Countries: Case Study of Bogotá, Colombia. *Transportation Research Record: Journal of the Transportation Research Board*. 2394. p.pp. 111–118.

Gonzalez-Feliu, J. & Routhier, J.-L. (2012). Modeling Urban Goods Movement: How to be Oriented with so Many Approaches? *Procedia - Social and Behavioral Sciences*. 39. p.pp. 89–100.

Guo, J. & Bhat, C. (2007). Population Synthesis for Microsimulating Travel Behavior. *Transportation Research Record: Journal of the Transportation Research Board*. 2014. p.pp. 92–101.

Hanly, M. & Dargay, J. (2000). Car ownership in Great Britain: Panel data analysis. *Transportation Research Record: Journal of the Transportation Research Board*. (1718). p.pp. 83–89.

Huang, Z. & Williamson, P. (2001). A comparison of synthetic reconstruction and combinatorial optimisation approaches to the creation of small-area microdata. *Department of Geography, University of Liverpool*. [Online]. Available from: http://pcwww.liv.ac.uk/~william/Microdata/Pop91/Methodology/workingpapers/hw_wp_2001_2.pdf. [Accessed: 17 February 2016].

Hunt, J.D. & Stefan, K.J. (2007). Tour-based microsimulation of urban commercial movements. *Transportation Research Part B: Methodological*. 41 (9). p.pp. 981–1013.

Hunt, J.D., Stefan, K.J., Brownlee, A.T., McMillan, J.D.P., Farhan, A., Tsang, K., Atkins, D. & Ishani, M. (2004). A commercial movement modelling strategy for Alberta's major cities. In: *Proceedings of the 2004 Annual Conference of the Transportation Association of Canada*. [Online]. 2004, Citeseer. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.117.7308&rep=rep1&type=pdf. [Accessed: 19 April 2016].

Hunt, J., Stefan, K. & Brownlee, A. (2006). Establishment-based survey of urban commercial vehicle movements in Alberta, Canada: survey design, implementation, and results. *Transportation Research Record: Journal of the Transportation Research Board*. (1957). p.pp. 75–83.

Jong, G.D., Fox, J., Pieters, M., Daly, A.J. & Smith, R. (2004). A comparison of car ownership models. *Transport Reviews*. 24 (4). p.pp. 379–408.

Joubert, J.W. & Axhausen, K.W. (2009). Inferring commercial vehicle activities in Gauteng, South Africa. *Journal of Transport Geography*. 19 (1). p.pp. 115–124.

Joubert, J.W., Fourie, P.J. & Axhausen, K.W. (2010). *A large scale combined private car and commercial vehicle-based traffic simulation*. [Online]. Available from: http://e-collection.library.ethz.ch/view/eth:68. [Accessed: 20 April 2016].

Karlaftis, M. & Golias, J. (2002). Automobile ownership, households without automobiles, and urban traffic parameters: Are they related? *Transportation Research Record: Journal of the Transportation Research Board*. (1792). p.pp. 29–35.

Klincevicius, M., Morency, C. & Trépanier, M. (2014). Assessing Impact of Carsharing on Household Car Ownership in Montreal, Quebec, Canada. *Transportation Research Record: Journal of the Transportation Research Board*. 2416. p.pp. 48–55.

Liedtke, G. & Schepperle, H. (2004). Segmentation of the transportation market with regard to activity-based freight transport modelling. *International Journal of Logistics*. 7 (3). p.pp. 199–218.

Li, J., Walker, J., Srinivasan, S. & Anderson, W. (2010). Modeling Private Car Ownership in China: Investigation of Urban Form Impact Across Megacities. *Transportation Research Record: Journal of the Transportation Research Board*. 2193. p.pp. 76–84.

Madar, G. (2014). *Micro-Data Collection and Development of Trip Generation Models of Commercial Vehicles: An Application for Windsor, Ontario*. [Online]. Available from: http://scholar.uwindsor.ca/etd/5186/. [Accessed: 20 April 2016].

Ma, L. & Srinivasan, S. (2010). Impact of Individuals' Immigrant Status on Household Auto Ownership. *Transportation Research Record: Journal of the Transportation Research Board*. 2156. p.pp. 36–46.

Maoh, H. & Kanaroglou, P. (2006). Geographic clustering of firms and urban form: a multivariate analysis. *Journal of Geographical Systems*. 9 (1). p.pp. 29–52.

Maoh, H. & Kanaroglou, P. (2009). Intrametropolitan Location of Business Establishments. *Transportation Research Record: Journal of the Transportation Research Board*. 2133. p.pp. 33–45.

Martin Frick, I.V.T., Axhausen, K.W. & Zürich, I. (2004). *Generating synthetic populations using IPF and monte carlo techniques: Some new results*. [Online]. Available from: http://matsim.org/uploads/ab225.pdf. [Accessed: 16 February 2016].

McFadden, D. (1978). MODELING THE CHOICE OF RESIDENTIAL LOCATION. *Transportation Research Record*. [Online]. (673). Available from: https://trid.trb.org/view.aspx?id=87722. [Accessed: 20 April 2016].

A. Meyer, D. Meyer, & National Research Council (eds.) (2013). *City logistics research: a transatlantic perspective ; summary of the first EU-U.S. transportation research symposium ; May 30-31, 2013, Washington, DC*. Conference proceedings / Transportation Research Board 50. Washington, DC: Transportation Research Board.

Miller, E.J., Kriger, D.S. & Hunt, J.D. (1998). INTEGRATED URBAN MODELS FOR SIMULATION OF TRANSIT AND LAND-USE POLICIES. *TCRP Web Document*. [Online]. (9). Available from: https://trid.trb.org/view.aspx?id=505877. [Accessed: 19 April 2016].

NCHRP 606 (2008). *National Cooperative Highway Research Progam: Forecasting statewide freight toolkit*. Washington, D.C.: Transportation Research Board.

Pendyala, R., Bhat, C., Goulias, K., Paleti, R., Konduri, K., Sidharthan, R., Hu, H.-H., Huang, G. & Christian, K. (2012). Application of Socioeconomic Model System for Activity-Based Modeling. *Transportation Research Record: Journal of the Transportation Research Board*. 2303. p.pp. 71–80.

Potoglou, D. & Kanaroglou, P.S. (2008a). Disaggregate Demand Analyses for Conventional and Alternative Fueled Automobiles: A Review. *International Journal of Sustainable Transportation*. 2 (4). p.pp. 234–259.

Potoglou, D. & Kanaroglou, P.S. (2008b). Modelling car ownership in urban areas: a case study of Hamilton, Canada. *Journal of Transport Geography*. 16 (1). p.pp. 42–54.

Potoglou, D. & Susilo, Y. (2008). Comparison of Vehicle-Ownership Models. *Transportation Research Record: Journal of the Transportation Research Board*. 2076. p.pp. 97–105.

Prillwitz, J., Harms, S. & Lanzendorf, M. (2006). Impact of life-course events on car ownership. *Transportation Research Record: Journal of the Transportation Research Board*. (1985). p.pp. 71–77.

Regan, A.C. & Garrido, R.A. (2001). Modeling freight demand and shipper behaviour: state of the art, future directions. *Travel Behaviour Research: The Leading Edge. Pergamon, Amsterdam*. p.pp. 185–215.

Roorda, M.J., Cavalcante, R., McCabe, S. & Kwan, H. (2010). A conceptual framework for agent-based modelling of logistics services. *Transportation Research Part E: Logistics and Transportation Review*. 46 (1). p.pp. 18–31.

Ryan, J. & Han, G. (1999). Vehicle-ownership model using family structure and accessibility application to Honolulu, Hawaii. *Transportation Research Record: Journal of the Transportation Research Board*. (1676). p.pp. 1–10.

Ryan, J., Maoh, H. & Kanaroglou, P. (2009). Population synthesis: Comparing the major techniques using a small, complete population of firms. *Geographical Analysis*. 41 (2). p.pp. 181–203.

Samimi, A., Mohammadian, K. & Kawamura, K. (2012). Behavioral freight movement modeling: Methodology and data needs. *Travel Behaviour Research in an Evolving World*. p.p. 147.

Schimek, P. (1996). Household motor vehicle ownership and use: How much does residential density matter? *Transportation Research Record: Journal of the Transportation Research Board*. (1552). p.pp. 120–125.

ter Schure, J., Napolitan, F. & Hutchinson, R. (2012). Cumulative Impacts of Carsharing and Unbundled Parking on Vehicle Ownership and Mode Choice. *Transportation Research Record: Journal of the Transportation Research Board*. 2319. p.pp. 96–104.

Shay, E. & Khattak, A. (2005). Automobile ownership and use in neotraditional and conventional neighborhoods. *Transportation Research Record: Journal of the Transportation Research Board*. (1902). p.pp. 18–25.

Sillaparcharn, P. (2007). Modeling of Vehicle Ownership: Case Study of Thailand. *Transportation Research Record: Journal of the Transportation Research Board*. 2038. p.pp. 98–104.

Simpson, L. & Tranmer, M. (2005). Combining Sample and Census Data in Small Area Estimates: Iterative Proportional Fitting with Standard Software*. *The Professional Geographer*. 57 (2). p.pp. 222–234.

Statistics Canada (2011). *2011 National Household Survey*. [Online]. 2011. Available from: http://dc.chass.utoronto.ca/cgi-bin/census/2011nhs/displayCensus.cgi?year=2011&geo=ct. [Accessed: 20 April 2016].

Stefan, K. & Hunt, J.D. (2004). Review of urban commodity movement demand modelling approaches. In: *Annual Meeting of the Canadian Transportation Research Forum, Calgary, AB*. 2004.

Tavasszy, L. (2007). *Freight demand modeling: tools for public-sector decision making : summary of a conference, September 25-27, 2006, Keck Center of the National Academies, Washington, D.C.* K. L. Hancock, National Research Council (U.S.), & Transportation Research Board (eds.). Washington, D.C.: Transportation Research Board.

Tavasszy, L. a., Smeenk, B. & Ruijgrok, C. j. (1998a). A DSS For Modelling Logistic Chains in Freight Transport Policy Analysis. *International Transactions in Operational Research*. 5 (6). p.pp. 447–459.

Tavasszy, L.A., Smeenk, B. & Ruijgrok, C.J. (1998b). A DSS for modelling logistic chains in freight transport policy analysis. *International Transactions in Operational Research*. 5 (6). p.pp. 447–459.

Thakuriah, P., Menchu, S. & Tang, L. (2010). Car Ownership Among Young Adults: Generational and Period-Specific Perspective. *Transportation Research Record: Journal of the Transportation Research Board*. 2156. p.pp. 1–8.

Train, K.E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Transport Canada, G. of C.T.C.P. (2012). *Transportation and the Economy*. [Online]. 10 July 2012. Available from: https://www.tc.gc.ca/eng/policy/anre-menu-3016.htm. [Accessed: 20 April 2016].

Voas, D. & Williamson, P. (2000). An evaluation of the combinatorial optimisation approach to the creation of synthetic microdata. *International Journal of Population Geography*. 6 (5). p.pp. 349–366.

Williamson, P., Birkin, M. & Rees, P.H. (1998). The estimation of population microdata by using data from small area statistics and samples of anonymised records. *Environment and Planning A*. 30 (5). p.pp. 785–816.

Wilson, A.G. & Pownall, C.E. (1976). A New Representation of the Urban System for Modelling and for the Study of Micro-Level Interdependence. *Area*. 8 (4). p.pp. 246–254.

Wisetjindawat, W. & Sano, K. (2003). A behavioral modeling in micro-simulation for urban freight transportation. *Journal of the Eastern Asia Society for Transportation Studies*. 5. p.pp. 2193–2208.

Wisetjindawat, W., Sano, K. & Matsumoto, S. (2006). Commodity Distribution Model Incorporating Spatial Interactions for Urban Freight Movement. *Transportation Research Record: Journal of the Transportation Research Board*. 1966. p.pp. 41–50.

Wu, G., Yamamoto, T. & Kitamura, R. (1999). Vehicle ownership model that incorporates the causal structure underlying attitudes toward vehicle ownership. *Transportation Research Record: Journal of the Transportation Research Board*. (1676). p.pp. 61–67.

Yang, C., Regan, A.C. & Son, Y.T. (2010). Another view of freight forecasting modeling trends. *KSCE Journal of Civil Engineering*. 14 (2). p.pp. 237–242.

Yang, X.-S. (2010). *Engineering Optimization: An Introduction with Metaheuristic Applications*. John Wiley & Sons.

Zhou, J. & Dai, S. (2012). Urban and Metropolitan Freight Transportation: A Quick Review of Existing Models. *Journal of Transportation Systems Engineering and Information Technology*. 12 (4). p.pp. 106–114.

APPENDICES

APPENDIX A: SIC Detailed Industry Classification

Table A - 1: Two - digits SIC Codes and Detailed Description

SIC_2D	Industry Classification
1	Agricultural Production Crops
7	Agricultural Services
13	Oil And Gas Extraction
14	Mining And Quarrying Of Non-metallic Minerals, Except Fuels
15	Building Construction General Contractors And Operative Builders
16	Heavy Construction
17	Construction Special Trade Contractors
20	Food And Kindred Products
23	Apparel And Other Finished Products Made From Fabrics And Similar Materials
24	Lumber And Wood Products, Except Furniture
25	Furniture And Fixtures
26	Paper And Allied Products
27	Printing Publishing & Allied Industries
28	Chemical Manufacturing
29	Petroleum Refining And Related Industries
30	Rubber And Miscellaneous Plastics Products
31	Leather And Leather Products
32	Stone, Clay, Glass, And Concrete Products
33	Primary Metal Industries
34	Fabricated Metal Products, Except Machinery And Transportation Equipment
35	Industrial And Commercial Machinery And Computer Equipment
36	Electronic And Other Electrical Equipment And Components, Except Computer Equipment
37	Transportation
38	Measuring & Analyzing Instruments Manufacturers
39	Miscellaneous Manufacturing Industries
40	Railroad Transportation

SIC_2D	Industry Classification
41	Local And Suburban Transit And Interurban Highway Passenger Transportation
42	Waste Collection
43	United States Postal Service
44	Retail
45	Transportation by Air
46	Pipelines, Except Natural Gas
47	Transportation Services
48	Trucking
49	Electric, Gas, And Sanitary Services
50	Wholesale Trade-durable Goods
51	Wholesale Trade-non-durable Goods
52	Building Materials, Hardware, Garden Supply, And Mobile Home Dealers
53	General Merchandise Stores
54	Food Stores
55	Automotive Dealers And Gasoline Service Stations
56	Apparel And Accessory Stores
57	Home Furniture, Furnishings, And Equipment Stores
58	Eating And Drinking Places
59	Miscellaneous Retail
60	Depository Institutions
61	Non-depository Credit Institutions
62	Security And Commodity Brokers, Dealers, Exchanges, And Services
63	Insurance Carriers
64	Insurance Agents, Brokers, And Service
65	Real Estate
67	Holding And Other Investment Offices
70	Hotels, Rooming Houses, Camps, And Other Lodging Places
72	Personal Services
73	Business Services
75	Wholesale Trade-durable Goods
76	Miscellaneous Repair Services
78	Motion Pictures
79	Amusement and Recreation Services

SIC_2D	Industry Classification
80	Health Services
81	Legal Services
82	Educational Services
83	Social Services
84	Social non-profit
86	Membership Organizations
87	Engineering, Accounting, Research, Management, And Related Services
89	Miscellaneous Services
91	Executive, Legislative, And General Government, Except Finance
92	Justice, Public Order, And Safety
93	Public Finance, Taxation, And Monetary Policy
95	Admin-Environmental Quality Programs
96	Administration Of Economic Programs
97	National Security and International Affairs
99	Non-classifiable Establishments

APPENDIX B: Spatial Autocorrelation (Moran's I Results)

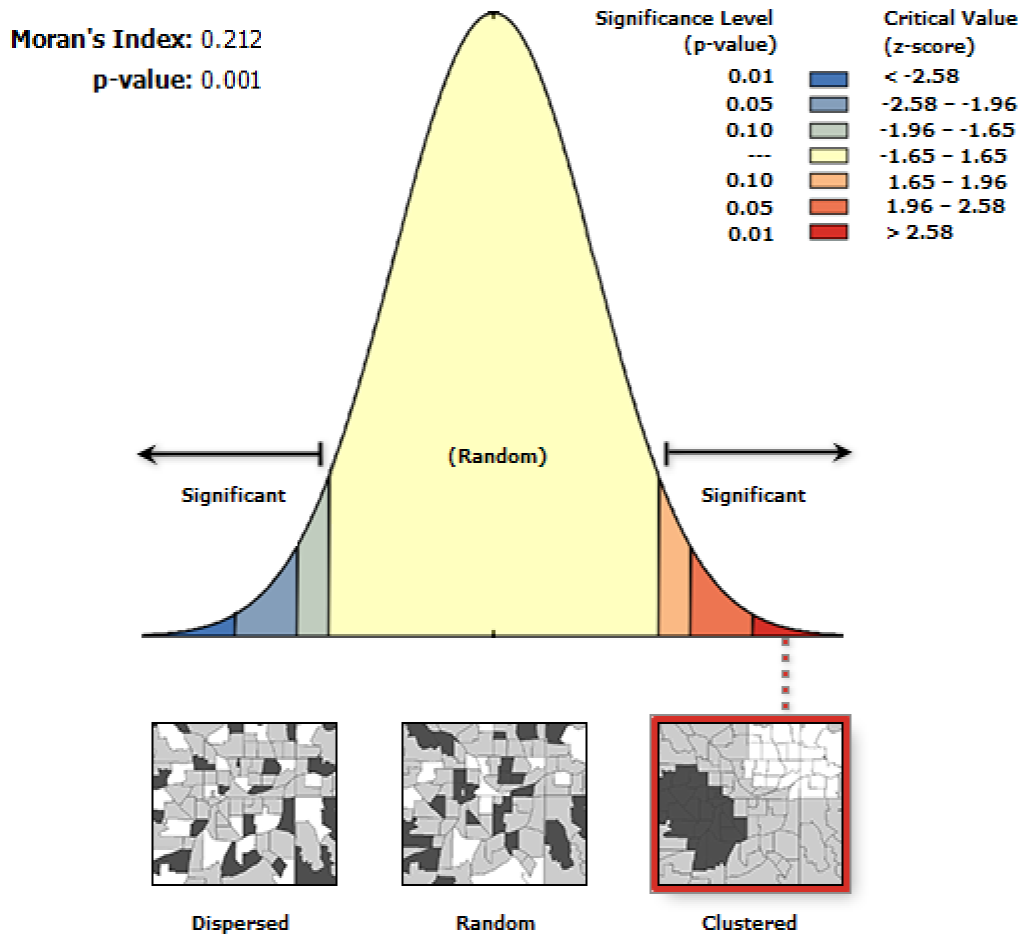


Figure B - 1: GVW 1 Moran's I Results

Given the p-value of 0.001, there is less than 1% likelihood that this clustered pattern could be the result of random chance.

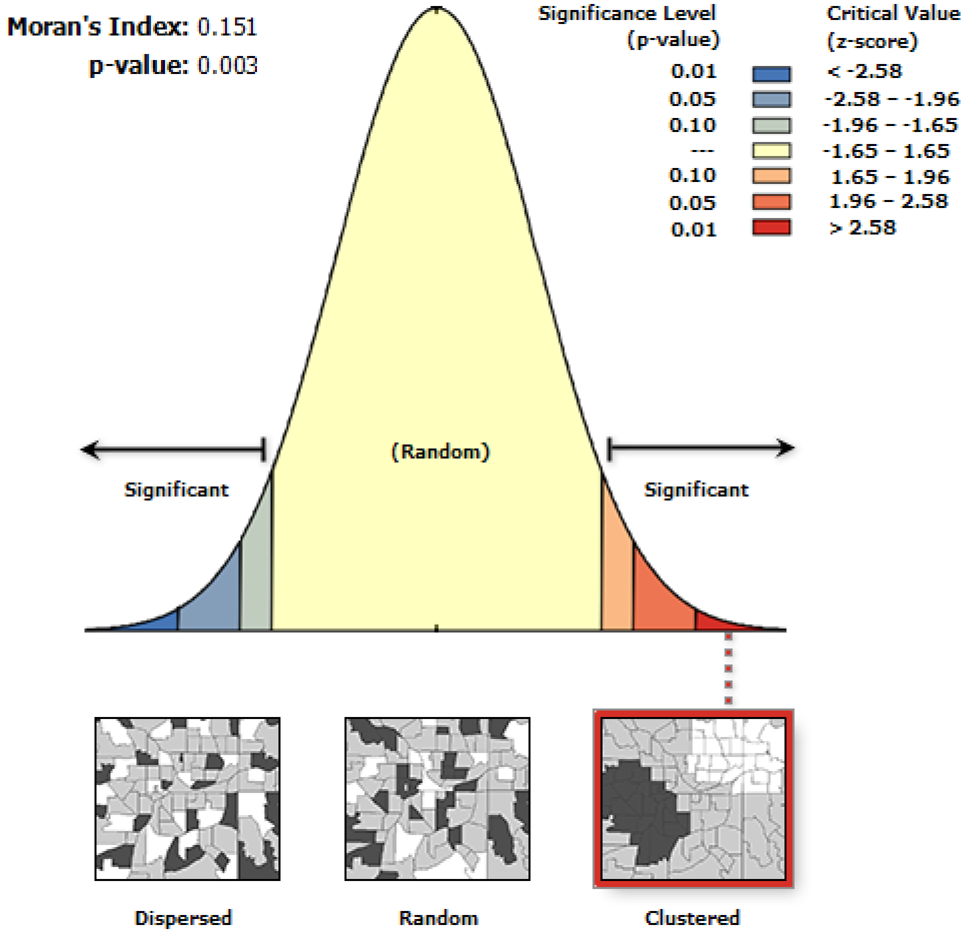


Figure B - 2: GVW 2 Moran's I Results

Given the p-value of 0.003, there is less than 1% likelihood that this clustered pattern could be the result of random chance.

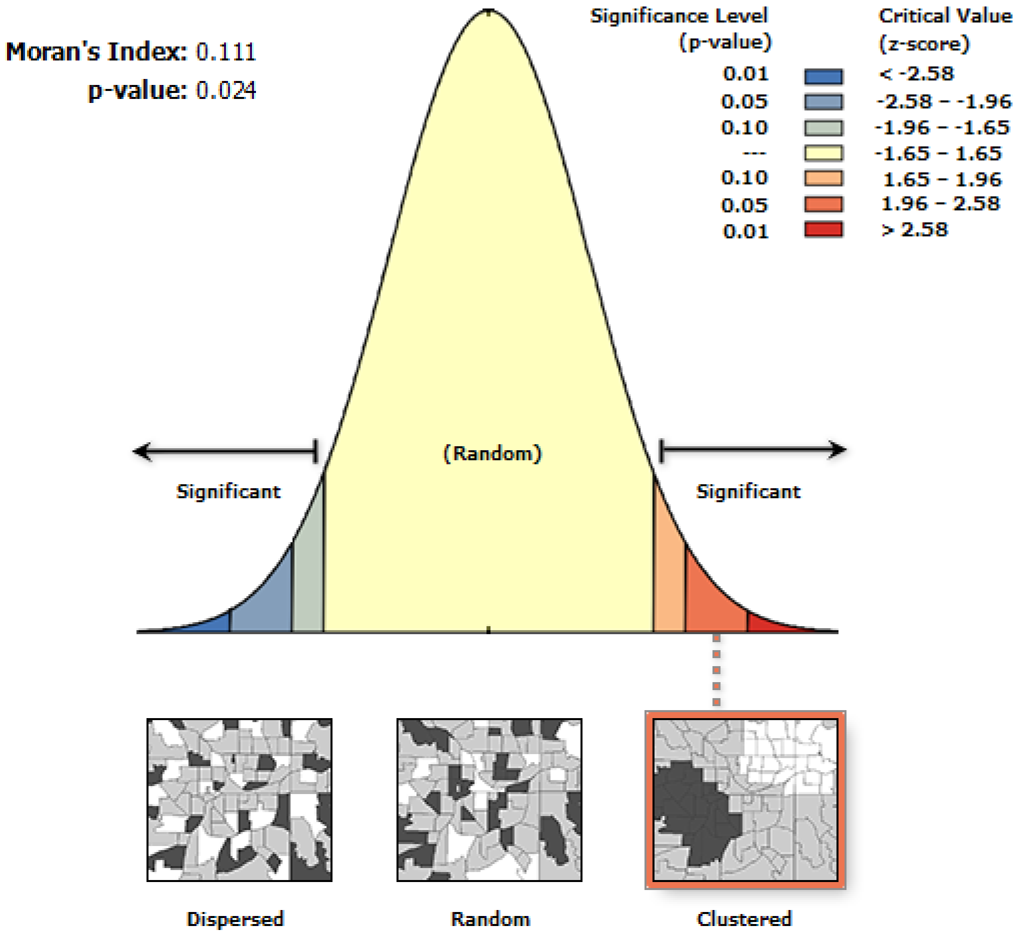


Figure B - 3: GVW 3 Moran's I Results

Given the p-value of 0.024, there is less than 5% likelihood that this clustered pattern could be the result of random chance.

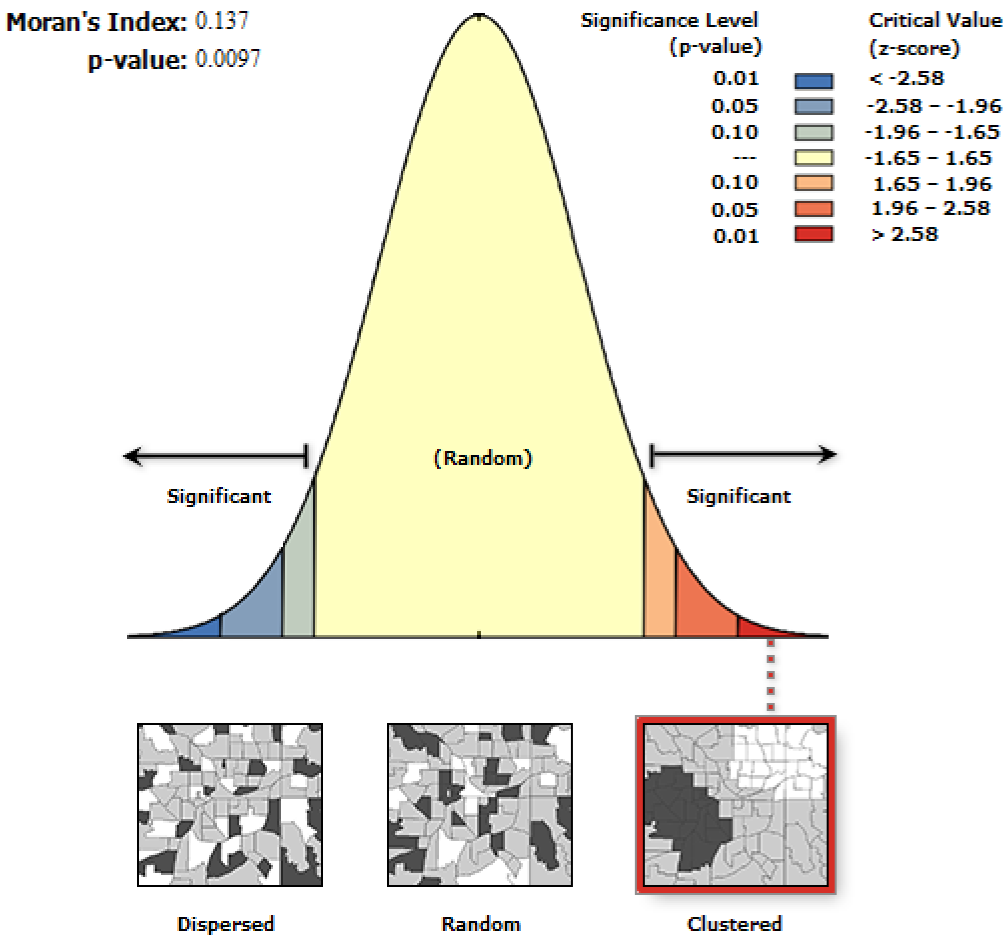


Figure B - 4: GVW 4 Moran's I Results

Given the p-value of 0.0097, there is less than 1% likelihood that this clustered pattern could be the result of random chance.

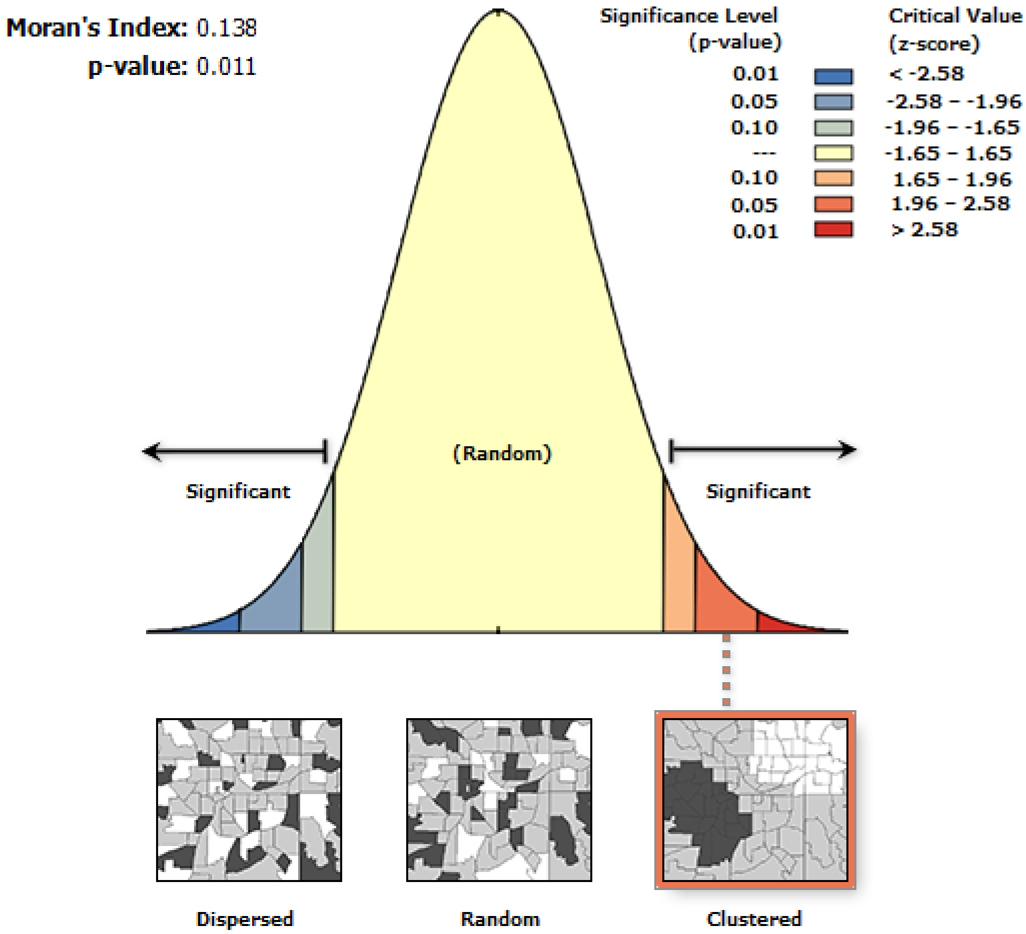


Figure B - 5: GVW 5 Moran's I Results

Given the p-value of 0.011, there is less than 5% likelihood that this clustered pattern could be the result of random chance.

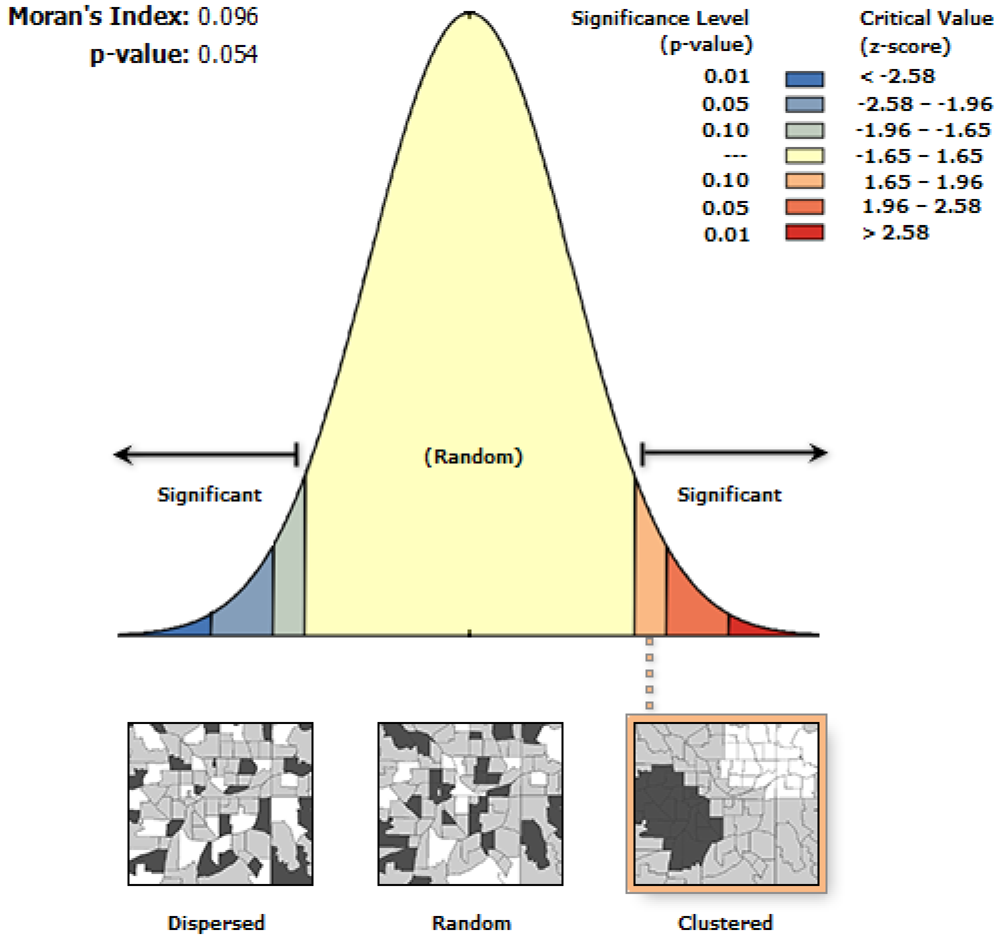


Figure B - 6: GVW 6 Moran's I Results

Given the p-value of 0.054, there is less than 10% likelihood that this clustered pattern could be the result of random chance.

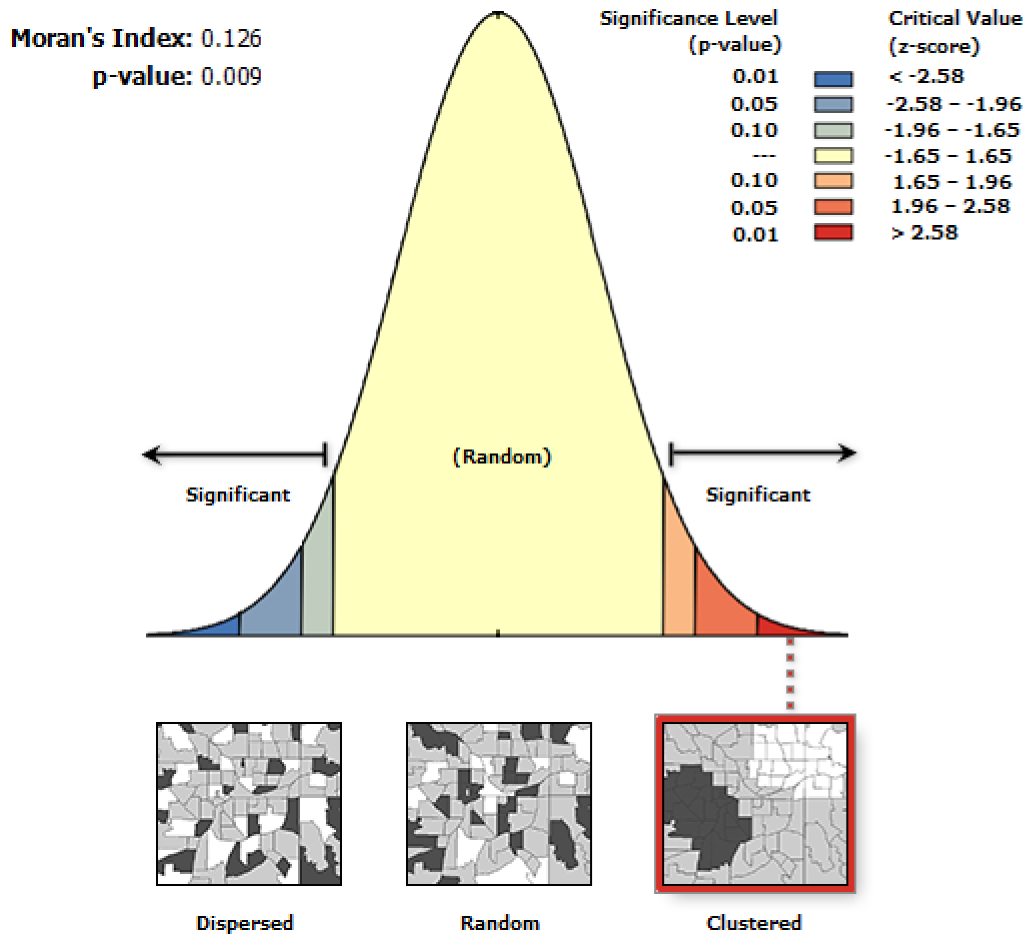


Figure B - 7: GVW 1 Moran's I Results

Given the p-value of 0.009, there is less than 1% likelihood that this clustered pattern could be the result of random chance.

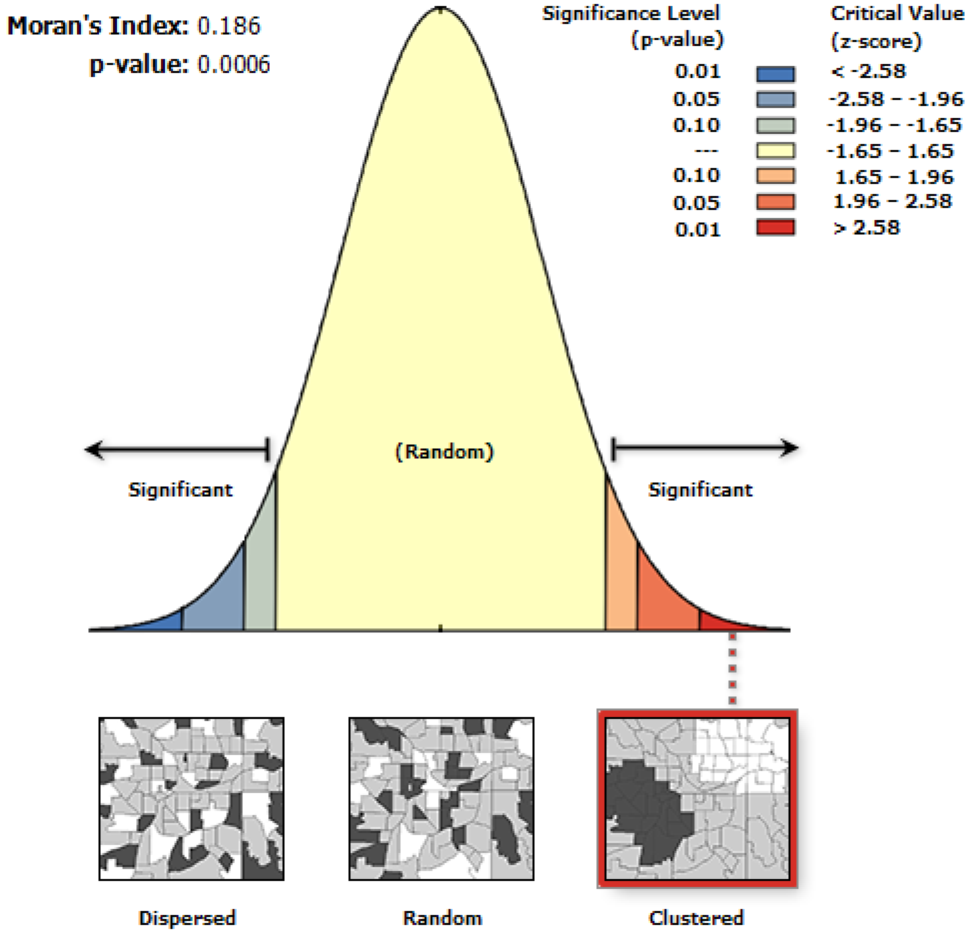


Figure B - 8: GVW 8 Moran's I Results

Given the p-value of 0.0006, there is less than 1% likelihood that this clustered pattern could be the result of random chance.

APPENDIX C: MNL Models - Observed versus Estimated Numbers Vehicles by Class

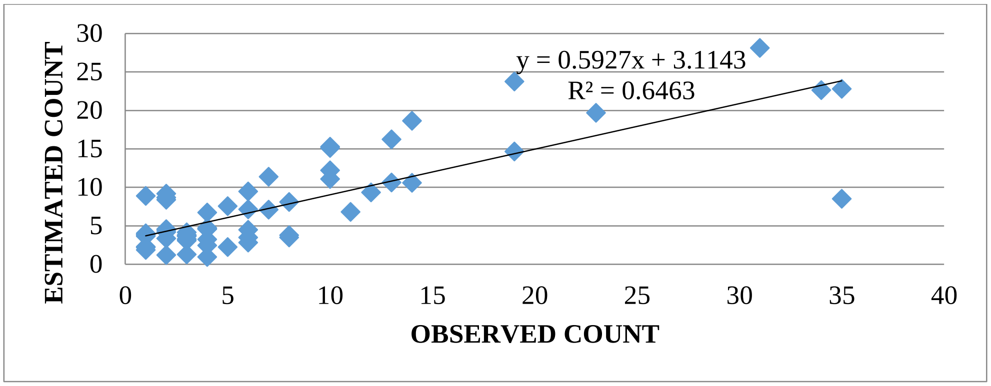


Figure C - 1: MNL Model 1 - Observed versus Estimated Numbers of Cars

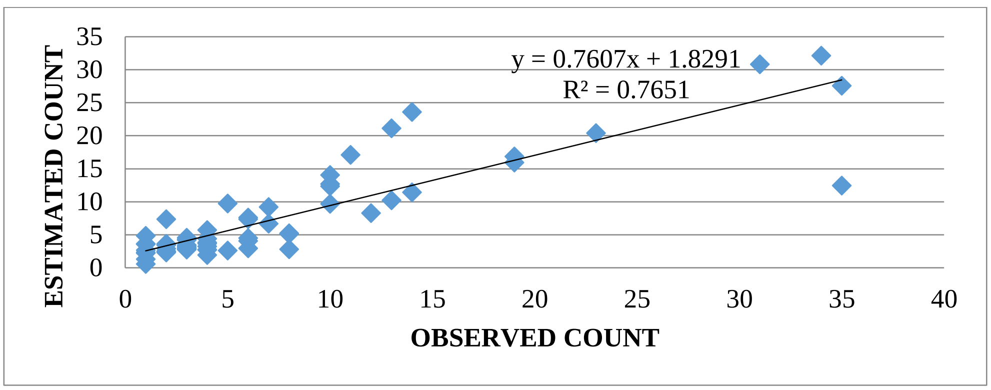


Figure C - 2: MNL Model 2 - Observed versus Estimated Numbers of Cars

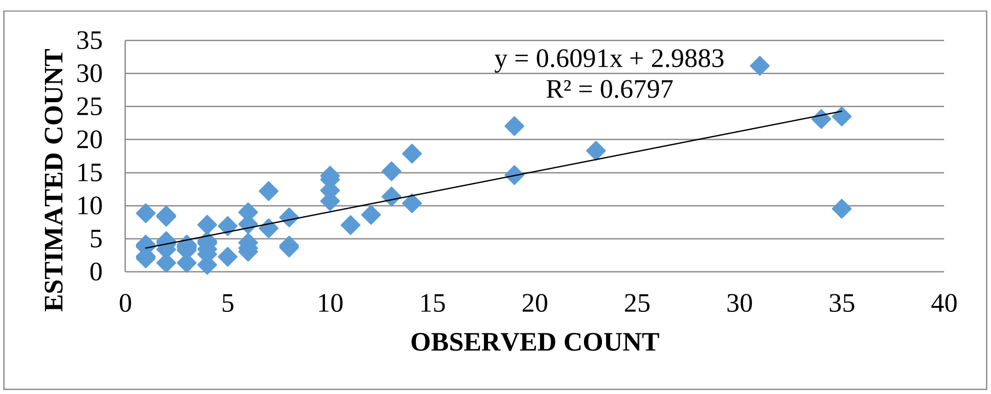


Figure C - 3: MNL Model 3 - Observed versus Estimated Numbers of Cars

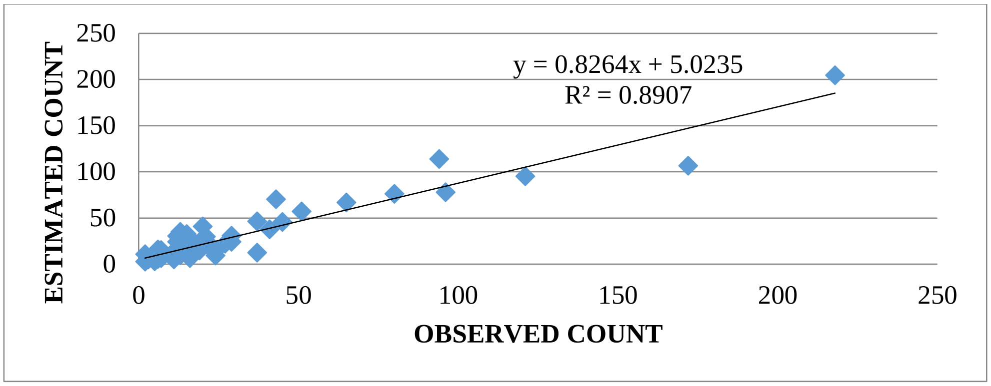


Figure C - 4: MNL Model 1 - Observed versus Estimated Numbers of Light Duty Trucks

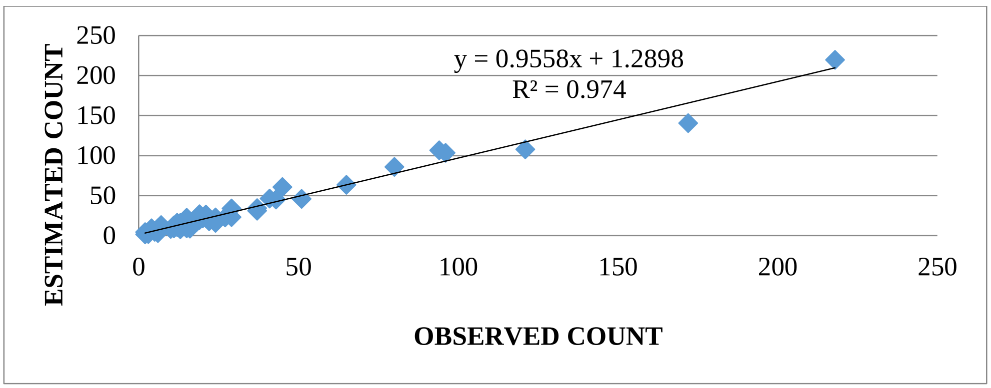


Figure C - 5: MNL Model 2 - Observed versus Estimated Numbers of Light Duty Trucks

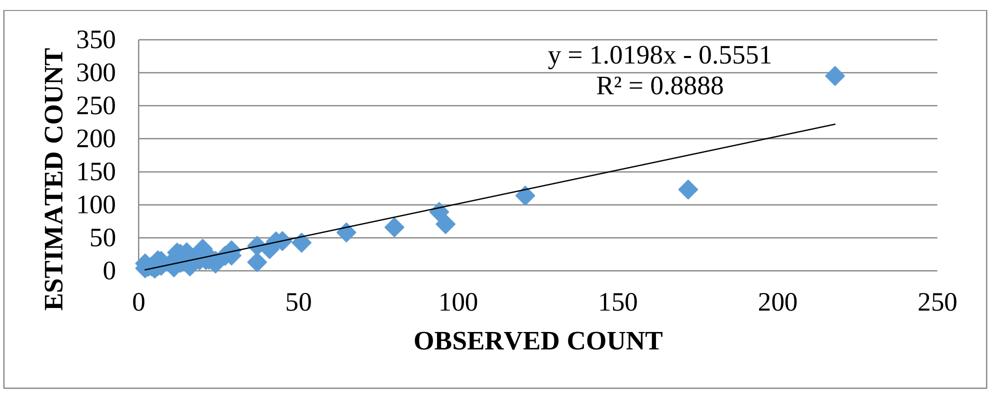


Figure C - 6: MNL Model 3 - Observed versus Estimated Numbers of Light Duty Trucks

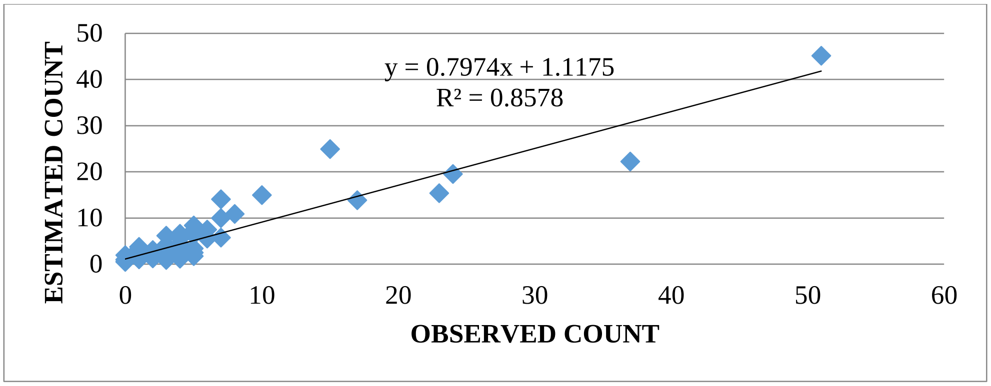


Figure C - 7: MNL Model 1 - Observed versus Estimated Numbers of Medium Duty Trucks

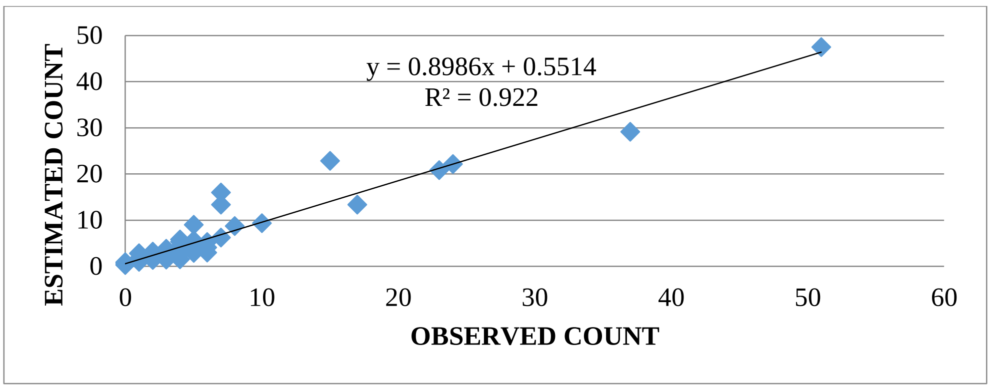


Figure C - 8: MNL Model 2 - Observed versus Estimated Numbers of Medium Duty Trucks

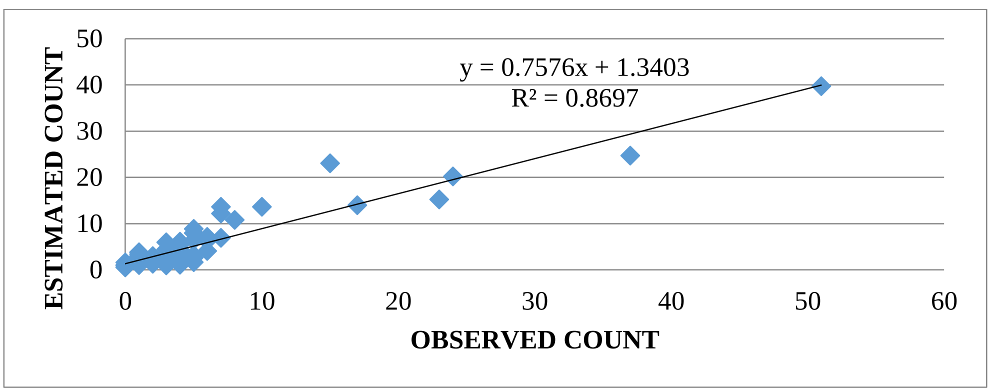


Figure C - 9: MNL Model 3 - Observed versus Estimated Numbers of Medium Duty Trucks

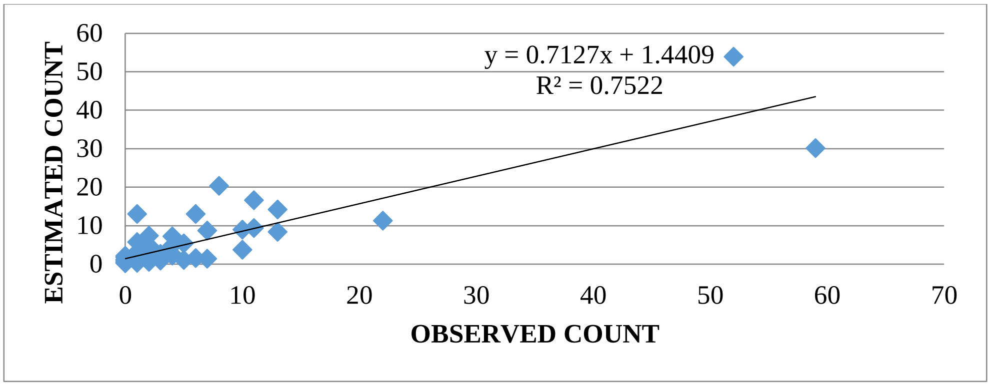


Figure C - 10: MNL Model 1 - Observed versus Estimated Numbers of Heavy Duty Trucks

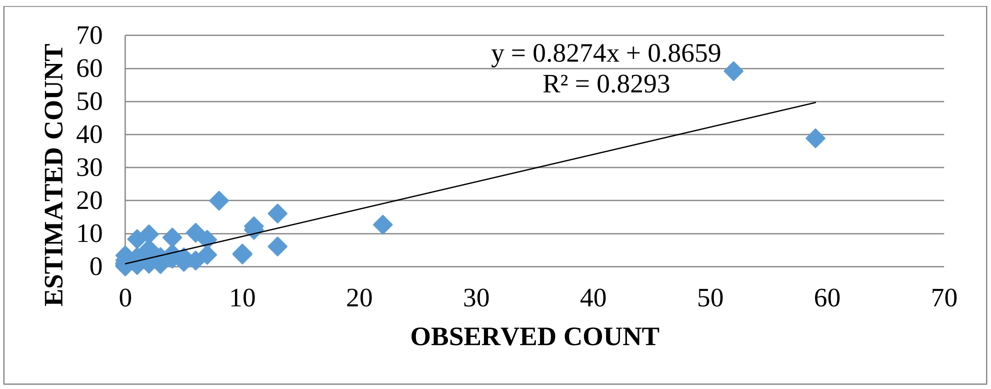


Figure C - 11: MNL Model 2 - Observed versus Estimated Numbers of Heavy Duty Trucks

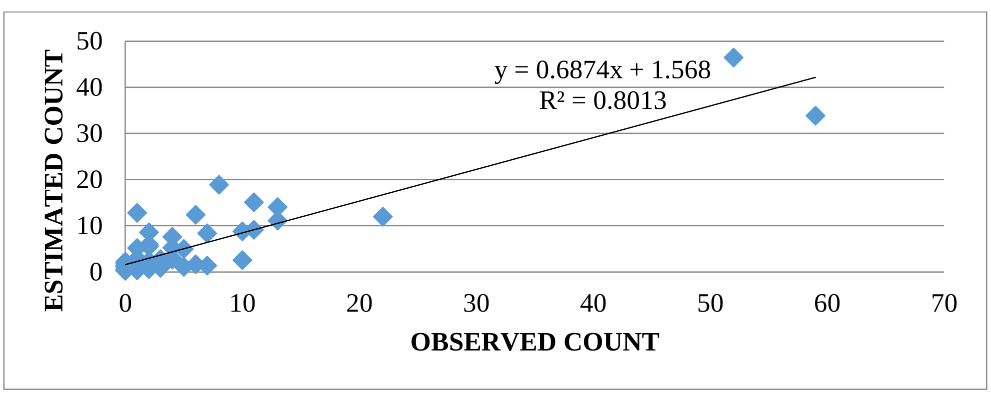


Figure C - 12: MNL Model 3 - Observed versus Estimated Numbers of Heavy Duty Trucks

APPENDIX D: NB Models - Observed versus Estimated Numbers Vehicles by Class

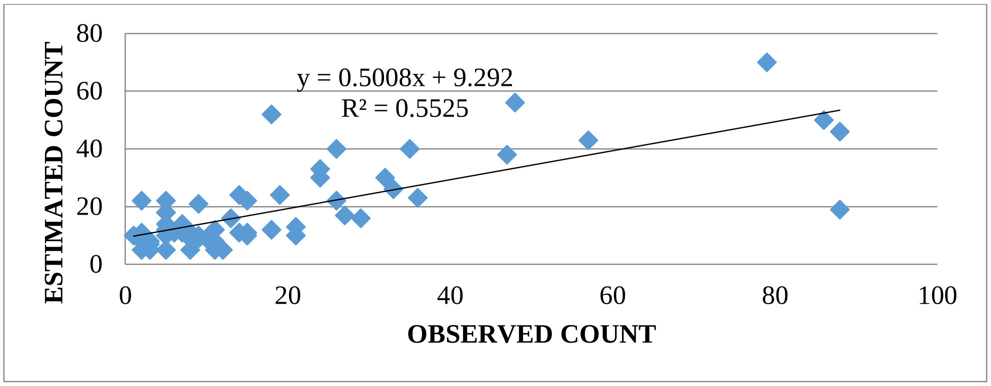


Figure D - 1: NB Model 1 - Observed versus Estimated Numbers of Cars

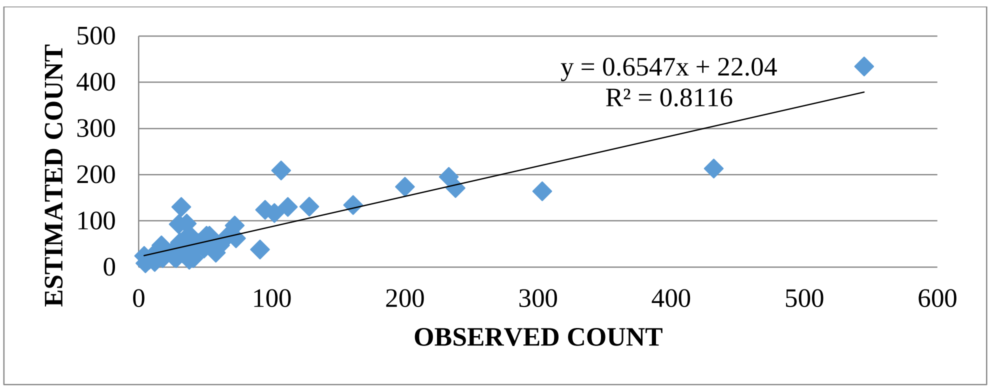


Figure D - 2: NB Model 2 - Observed versus Estimated Numbers of Light Duty Trucks

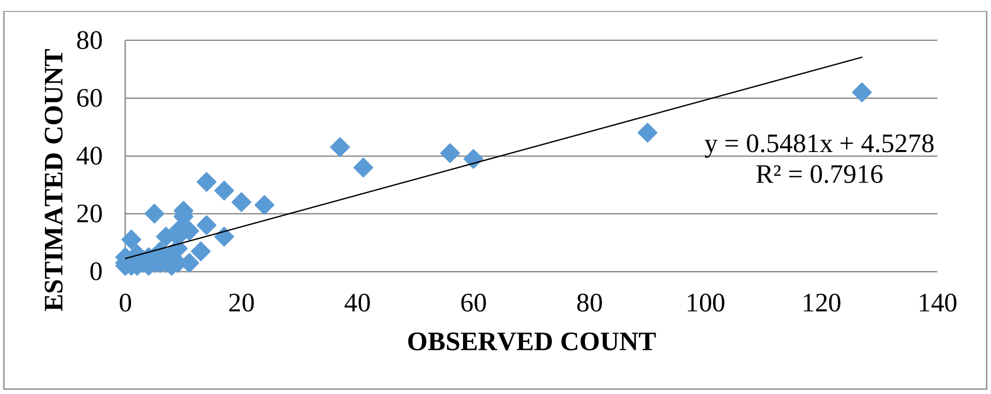


Figure D - 3: NB Model 3 - Observed versus Estimated Numbers of Medium Duty Trucks

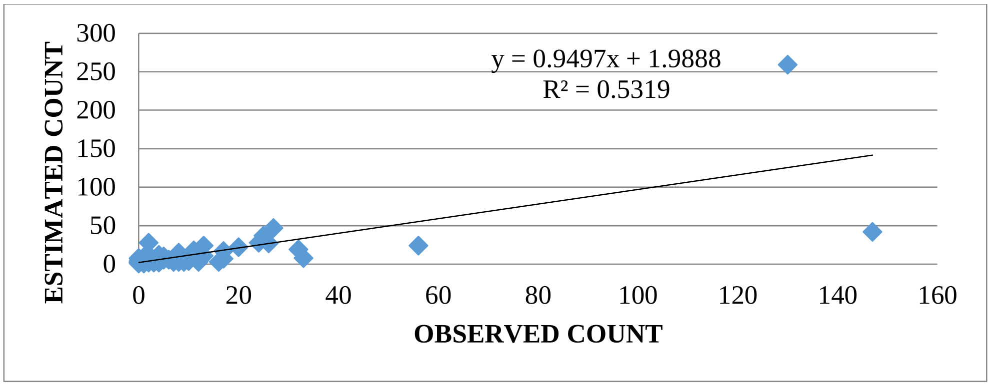


Figure D - 4: NB Model 4- Observed versus Estimated Numbers of Heavy Duty Trucks

VITA AUCTORIS

NAME:	Aya Hagag
PLACE OF BIRTH:	Kalyobiya, Egypt
YEAR OF BIRTH:	1992
EDUCATION:	University of Windsor, BASc in Civil Engineering, Windsor, ON, 2013
	University of Windsor, MASc in Civil Engineering, Windsor, ON, 2016