Behavioural Modelling of Urban Freight Transportation: Activity and Inter-Arrival Duration Models Estimated Using GPS Data

by

Bryce Winiata Sharman

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy

Graduate Department of Civil Engineering
University of Toronto

© Copyright by Bryce Winiata Sharman 2014
Behavioural Modelling of Urban Commercial Transportation: Activity and Inter-Arrival Duration Models Estimated Using GPS Data

Bryce Winiata Sharman

Doctor of Philosophy

Graduate Department of Civil Engineering
University of Toronto

2014

This dissertation details the development of two behavioural models of urban commercial (freight and services) transportation: activity duration and inter-arrival duration models. These models were designed to comprise two components within the proposed FRELODE (FREight LOgistics DEcisions) modelling framework; FRELODE accepts a list of shipments or service trips by carrier and models how carriers execute urban shipments, creating a series of resulting vehicle movements (trips and tours). My Ph.D. research produced significant progress in four areas. First, the activity and inter-arrival duration models were estimated using a three-month passively-collected GPS dataset from truck-mounted GPS-equipped engine-on-board recorders. To the author’s knowledge, this research represents the first attempt in the literature to estimate components of a commercial-vehicle travel demand model using passively-collected GPS data as the primary data source. Exploring new data sources, such as GPS, is important because little data describing urban commercial vehicle movements are currently available. New data processing techniques were developed to convert the longitudinal GPS data into a travel diary suitable for transportation model estimation. Second, two hazard models of commercial activity duration were estimated. Third, models of inter-arrival duration were estimated using the longitudinal GPS dataset to model the number of days between repeated visits to the same destination. To the author’s knowledge, these represent the first estimated models of inter-arrival duration for commercial transportation. The fourth area of significant process described in this dissertation is the development of the FRELODE modelling framework. FRELODE considers a multiple day time period as commercial establishments do not necessarily operate using consistent schedules, and also accounts for carriers delivering multiple shipments in a single tour. In conclusion, passively-collected GPS data were found to hold promise as a complement to existing data sources of commercial vehicle travel. Also, it is expected that these estimated models, included within the FRELODE framework, will form components within a larger proposed agent-based microsimulation commercial vehicle modelling framework that is currently under development.
Acknowledgements

As I write these acknowledgements, it is rewarding to reflect back about all of the people who have helped make this Ph.D. a most rewarding and fulfilling experience.

First, I would like to thank my supervisor, Prof. Matthew Roorda. Where can I start? I am extremely grateful for his constant support, his guidance and his friendship over these last six years. Also, his consistent advice and his prompt and thorough revisions greatly improved the quality of my research. It is such a pleasure to work under the guidance of a supervisor who is so dedicated and who takes so much delight in helping his students succeed.

I would like to acknowledge all of the other students who have helped me through this program, for all our enlightening discussions and their friendship. Special thanks go to Andrew Wong, Karen Woo, Jiang Hao, Bilal Farooq, Rinaldo Cavalcante, Marcus Williams, Michael Hain, Josée Dumont, Kasra Rezaee, Glareh Amirjamshidi, Chris Bachmann, Keith Cochrane, Aarshabh Misra, Erin Toop and Siva Srikukenthiran. There are many other students who I have not mentioned by name, but you have also contributed so much to my experiences here.

Much thanks goes to Turnpike Global Technologies who provided the data for this research. All of this research would have been impossible without you. I would also like to thank the financial support of the National Sciences and Engineering Research Council of Canada (NSERC) and the Transportation Association of Canada. Your financial assistance helped immeasurably.

To my parents, I’d like to thank you both for your love, support, your understanding and also for lending me your ears at times when I needed someone to talk to.

Finally, I would like to thank by beloved Maira. You left your job and family in Brazil so that we could be together and encouraged me throughout this Ph.D., in good times and in harder times. What a ride this has been. It would have been very hard to complete this Ph.D. without your love and support. As I finish this thesis and then get prepared for our wedding, I am thinking so much of our wonderful time together. I hope that when this is over that we’ll be able to do so much more together. Thank you so very much my love.
Four chapters in this thesis have been reproduced (with modifications) from previously published material. These include:


# Table of Contents

**ACKNOWLEDGEMENTS** .................................................................................................................................................. III

**TABLE OF CONTENTS** .................................................................................................................................................. V

**LIST OF TABLES** ........................................................................................................................................................ IX

**LIST OF FIGURES** ........................................................................................................................................................ X

**CHAPTER 1: INTRODUCTION** ........................................................................................................................................ 1

  1.1 **IMPORTANCE AND EXTERNALITIES OF FREIGHT TRANSPORTATION** ................................................................. 1
  1.2 **IMPORTANCE OF FREIGHT DEMAND MODELLING** .................................................................................................. 4
  1.3 **IMPORTANCE OF HIGH-QUALITY DATA** .................................................................................................................. 6
    1.3.1 *Traditional data sources used to estimate commercial vehicle travel demand models* ............................................. 7
    1.3.2 *Passively-collected GPS data collection for commercial vehicle travel* ................................................................. 8
    1.3.3 *A case for passive GPS data for urban data collection* .......................................................................................... 10
  1.4 **RESEARCH OBJECTIVES** ........................................................................................................................................... 11
  1.5 **THESIS ORGANIZATION** ............................................................................................................................................ 12

**CHAPTER 2: LITERATURE REVIEW** ............................................................................................................................. 14

  2.1 **LITERATURE REVIEW OF FREIGHT TRAVEL DEMAND MODELS** ............................................................................ 14
    2.1.1 *Growth factor methods* ........................................................................................................................................ 14
    2.1.2 *Traditional “four-step” models* ........................................................................................................................... 15
    2.1.3 *Discussion of aggregate four-step models* ........................................................................................................... 20
    2.1.4 *Supply chain (logistics) models* .......................................................................................................................... 23
    2.1.5 *Urban logistics models* .......................................................................................................................................... 27
    2.1.6 *Tour-based models* .................................................................................................................................................. 29
    2.1.7 *Hybrid models* ......................................................................................................................................................... 35
    2.1.8 *Other freight models* ............................................................................................................................................. 38
  2.2 **LITERATURE REVIEW ON GPS DATA COLLECTION EFFORTS** .................................................................................. 40
    2.2.1 *GPS Studies of travel variability* ............................................................................................................................ 41
    2.2.2 *GPS studies examining underreporting in travel surveys* ........................................................................................ 42
    2.2.3 *GPS studies of commercial vehicles* .................................................................................................................... 43
    2.2.4 *GPS studies focussing on data processing algorithms and techniques* ............................................................. 44
  2.3 **SUMMARY** ................................................................................................................................................................. 46
CHAPTER 3: THE FRELODE (FREIGHT LOGISTICS DECISIONS) FRAMEWORK ............................................. 48

3.1 REGIONAL AGENT-BASED MICROSIMULATION CONCEPTUAL FRAMEWORK FOR FREIGHT ............................................. 48
3.2 FRELODE – ESTIMATING URBAN LOGISTICS DECISIONS USING PASSIVELY-COLLECTED GPS DATA ............................................. 51
   3.2.1 Shipment scheduling ........................................................................................................ 53
   3.2.2 Vehicle type selection ..................................................................................................... 54
   3.2.3 Tour formation ................................................................................................................ 54
   3.2.4 Tour start time ................................................................................................................ 55
   3.2.5 Vehicle trip ..................................................................................................................... 56
   3.2.6 Activity duration ............................................................................................................. 56
   3.2.7 Implementation within larger agent based framework ..................................................... 56
3.3 ADVANTAGES OF THE FRELODE FRAMEWORK COMPARED WITH EXISTING COMMERCIAL TRAVEL DEMAND MODELS .......... 56
3.4 SUMMARY .............................................................................................................................. 58

CHAPTER 4: DATA SOURCES USED FOR MODEL ESTIMATION .......................................................... 59

4.1 DESCRIPTION OF THE GPS DATA ........................................................................................ 59
4.2 SECONDARY DATA SOURCES ............................................................................................. 63
   4.2.1 CanMap * RouteLogistics street map ................................................................................ 64
   4.2.2 Transportation Tomorrow Survey .................................................................................... 64
   4.2.3 Teranet’s Ontario Parcel Database .................................................................................... 66
   4.2.4 InfoCanada Database of Companies ............................................................................... 67
4.3 SUMMARY .............................................................................................................................. 68

CHAPTER 5: PROCESSING THE GPS DATA ...................................................................................... 69

5.1 INTRODUCTION TO GPS DATA PROCESSING ...................................................................... 69
5.2 OUTPUT VARIABLES PROVIDED BY THE DATA PROCESSING ............................................. 70
5.3 PREPROCESSING GPS DATA ................................................................................................ 72
5.4 TRIP AND TRIP-END IDENTIFICATION ................................................................................. 73
   5.4.1 Literature review of trip-end detection in GPS studies ....................................................... 75
   5.4.2 Adapted algorithm to identify false-positive and missed trip ends ..................................... 78
   5.4.3 Check for missed trip ends ............................................................................................... 80
   5.4.4 Moving departure and imputed trip ends ......................................................................... 81
   5.4.5 Summary data describing data cleaning results ................................................................. 81
7.5 DESCRIPTION OF SELECTED DISCRETE CHOICE MODELS OF INTER-ARRIVAL DURATION ................................................................. 147
  7.5.1 Random intercept models ................................................................................................................................. 147
  7.5.2 Random intercept ordered probit model .................................................................................................................. 149
  7.5.3 Random intercept multinomial logit model .............................................................................................................. 150
7.6 ESTIMATION OF RANDOM INTERCEPT MODELS .............................................................................................................. 151
  7.6.1 Maximum likelihood estimation ............................................................................................................................... 151
  7.6.2 Bayesian estimation ................................................................................................................................................. 152
  7.6.3 Comparison of maximum-likelihood and MCMC estimation techniques ............................................................... 155
7.7 FIRST MODELLING ATTEMPT – MAXIMUM LIKELIHOOD ESTIMATION OF MODELS WITH NO FIXED PARAMETERS ........... 156
7.8 COMPARING MODEL GOODNESS OF FIT FOR MODELS ESTIMATED USING THE MCMC SOLVER ............................ 159
7.9 ESTIMATION OF RANDOM INTERCEPT MODELS – MCMC SOLVER ...................................................................................... 160
  7.9.1 Random intercept ordered probit model .................................................................................................................... 161
  7.9.2 Random intercept multinomial logit model ................................................................................................................ 164
  7.9.3 Market segmentation model ........................................................................................................................................ 167
7.10 SUMMARY .................................................................................................................................................................. 170

CHAPTER 8: CONCLUSIONS AND FUTURE WORK ............................................................................................................................. 174
8.1 SUMMARY OF CONTRIBUTIONS ........................................................................................................................................ 174
8.2 LIMITATIONS AND FUTURE WORK ............................................................................................................................... 178
  8.2.1 Future work to improve the estimated activity and inter-arrival duration models .................................................. 178
  8.2.2 Future work required to complete FRELODE ........................................................................................................... 179
  8.2.3 Future work required to implement FRELODE into the larger agent-based commercial vehicle framework ............... 179
  8.2.4 Larger scale efforts with respect to estimation of agent-based travel demand models using passively-collected GPS data ............................................................................................................................................... 180

REFERENCES ......................................................................................................................................................................... 182
List of Tables

Table 4.1: Summary of Unprocessed GPS Data ................................................................. 61
Table 4.2: Available Data Fields in Turnpike GPS Dataset .................................................. 62
Table 4.3: Sample of Raw GPS Data Provided by XRS ....................................................... 65
Table 5.1: Trip Output Table ............................................................................................ 71
Table 5.2: Trip-End Output Table ...................................................................................... 71
Table 5.3: Tour Output Table ............................................................................................ 72
Table 5.4: Destination Output Table .................................................................................. 72
Table 5.5: Trip GPS Points Table ....................................................................................... 72
Table 5.6: Trip-End Cleaning Summary ............................................................................. 82
Table 5.7: Test of Criteria Combinations for Depot Identification ...................................... 99
Table 5.8: Standard Industry Classification Covariates Linked from InfoCanada Database of Companies ......................................................... 105
Table 5.9: Example InfoCanada Database Entries ................................................................ 106
Table 5.10: Sample of Processed Database Ready to Link with Property Parcel Information .................................................... 106
Table 6.1: Summary Statistics of Observed Commercial Activity Durations ......................... 117
Table 6.2: Description of Explanatory Variables for Activity Duration Models ..................... 121
Table 6.3: Log-Likelihood of No Covariate Parametric Models with Different Baseline Hazards .......................................................................................... 125
Table 6.4: Number of “Failures” in Each Time Period of Non-Parametric Model ..................... 128
Table 6.5 Model Parameters and Significance for Parametric and Non-Parametric Models, With and Without Linked Establishment Data. ................................................................ 130
Table 7.1: Explanatory Variables for Inter-Arrival Duration Models ..................................... 146
Table 7.2: Comparison of Log-likelihood for Ordered Probit Models with No Fixed Covariates .................................................................................. 157
Table 7.3: Comparison of DIC for Ordered Probit Models .................................................. 163
Table 7.4: Estimated Parameters for the Random Intercept Ordered Probit Models ................. 163
Table 7.5: Comparison of DIC for Multilevel Multinomial Logit Models ............................... 165
Table 7.6: Estimated Parameters for Random Intercept Multinomial Logit Model for Segment 1 (Frequently Visited Destinations) ....................................................................... 166
Table 7.7: Estimated Parameters for the Random Intercept Multinomial Logit Model for Segment 2 (Regularly-Scheduled Destinations) ..................................................................... 167
Table 7.8: Estimated Parameters for the Random Intercept Multinomial Logit Model for Segment 3 (Unscheduled Destinations) ........................................................................ 168
Table 7.9: Estimated Parameters for the Random Intercept Multinomial Logit Model for the Full Dataset .................................................................................. 169
Table 7.10: Estimated Parameters for the Market Segmentation Model .................................. 169
Table 8.1: Anticipated Areas of Improved Policy Response of FRELODE Compared with “Four-Step” Transportation Models ........................................................................... 177
List of Figures

Figure 1.1: Canadian Greenhouse Gas Emissions for Light-Duty and Heavy-Duty Trucks Between 1990 and 2009 .... 3
Figure 1.2: Vehicle Kilometers Travelled for Single-Unit Trucks and Tractors in Canada and Government Expenditures on Roads in Canada .................................................................................................................. 3
Figure 2.1: Commodity Flow "Four-Step" Model Framework .................................................................................. 16
Figure 2.2: Truck Trip "Four-Step" Model Framework .......................................................................................... 18
Figure 2.3: Truck vs. Commodity Flows for a Single Delivery Tour ....................................................................... 22
Figure 2.4: Overall Tour-Based Microsimulation Framework of Calgary Model ....................................................... 31
Figure 2.5: Structure of Ohio Disaggregate Commercial Model ............................................................................... 33
Figure 2.6: Overview of Integrated Model of Fischer et al. (2005) ......................................................................... 36
Figure 2.7: Hybrid Model Structure of Outwater et al. (2012) ............................................................................. 39
Figure 3.1: High-Level Overview of Proposed Agent-Based Microsimulation Conceptual Framework for Freight.... 49
Figure 3.2: High-Level FRELODE Framework ..................................................................................................... 53
Figure 4.1: Study Boundary of GPS Dataset ........................................................................................................ 60
Figure 4.2: Schematic Showing Trips Included in GPS Dataset .............................................................................. 61
Figure 4.3: DMTI Canmap® RouteLogistics Sample ............................................................................................. 66
Figure 4.4: Teranet Ontario Parcel Database Sample ............................................................................................ 67
Figure 5.1: Example False-Positive Stops Recorded on Freeways ........................................................................... 79
Figure 5.2: Example False-Positive Stops on Arterial Roads .................................................................................... 80
Figure 5.3: Analysis Regions .................................................................................................................................. 80
Figure 5.4: External and Internal Validation Measures for the UPGMA and Ward’s Hierarchical Agglomerative Clustering Methods ......................................................................................................................... 90
Figure 5.5: Phase 1 Clustering Results in Downtown Toronto ..................................................................................... 91
Figure 5.6: Schematic Showing Issues in Selecting Appropriate Distance Threshold .............................................. 93
Figure 5.7: Sample Clustering Results using Two-Step Clustering Methods .......................................................... 95
Figure 5.8: Schematic Showing Two Processed Tours ........................................................................................... 100
Figure 5.9 Histogram of the Number of Establishments from the InfoCanada Database of Companies on Each Visited Property Parcel and for Every Observed Activity Duration ................................................................. 106
Figure 6.1: Plots of Observed Survival, Cumulative Hazard and Hazard Distributions ......................................... 119
Figure 6.2 Cox-Snell Residuals for the Parametric Hazards Models ........................................................................ 134
Figure 7.1: Observed Inter-Arrival Durations for the Entire Dataset ......................................................................... 139
Figure 7.2: Number of Visits to each Destination by a Single Carrier ........................................................................ 140
Figure 7.3: Hour-by-Hour Breakdown of Observed Inter-Arrival Durations ............................................................ 144
Figure 7.4: Histograms of Inter-Arrival Durations for the Full Dataset and Individual Segments .......................... 145
Chapter 1: Introduction

1.1 Importance and Externalities of Freight Transportation

Freight transportation is a key component of modern life. It is the link between different stages of a supply chain, which converts products that people use in their everyday life from raw materials to final completed goods sent either to a retail establishment or directly to a consumer. It is a link that allows people to eat foods and consume products that originate from all corners of the world and to export goods created in one location to customers outside of that region.

Freight transportation is a crucial component of the Canadian economy. In 2011, Canada imported a total value of U.S. $281 billion of goods from the United States, while exporting a total value of U.S. $317 billion (Bureau of Transportation Statistics, 2012). This large quantity of trade is crucial to the economic wellbeing of the Canadian economy. According to Invest in Ontario (2012), CAD$1.2 trillion worth of goods are transported annually on Ontario highways. At a local level, according to the Ontario Commercial Vehicle Survey (CVS) an estimated 144,794 single-unit and 344,778 multiple unit truck trips occur in the Greater Toronto and Hamilton Area (GTHA) every week (HDR-iTrans, 2011). This number is likely underestimated due to the low number of survey stations in the GTHA for CVS (discussed further in Section 1.3). The growth in freight transportation has also significantly outpaced growth in passenger transportation. According to the North American Transportation Statistics Database (NATS, 2012), domestic Canadian passenger travel in personal vehicles grew from 466 to 494.7 billion km between 1995 and 2009 (a growth of 6.2% over this period). Meanwhile, road freight activities grew from 71.5 to 118.9 billion metric ton-kilometers (a growth of over 66%) during the same period. This trend is expected to continue.

While freight transportation brings many benefits and allows our modern lifestyle, freight transportation also causes externalities such as congestion, noise, pollution (including greenhouse gas emissions), pavement damage, road fatalities and serious collisions, security issues, and risks associated with hazardous materials movement. These externalities must be managed by government planners to reduce their impacts on the environment and citizens while supporting a well-functioning freight transportation system.

Freight distribution is a major contributor to traffic congestion, especially near major freight generators such as ports, airports, rail yards and warehousing centres (Giulano, Gordon, Pan, Park & Wang, 2010;...
Outwater, Islam & Spear (2005) found that between 5% (for Detroit) and 18% (Sacramento) of Total Vehicle Miles Travelled (VMT) could be attributable to commercial vehicles. Since commercial vehicles are often larger than personal vehicles, each vehicle can have a larger effect on traffic congestion than a passenger vehicle. Hence commercial vehicle traffic should not be ignored when analyzing congestion.

Freight travel is a large source of pollution, including particulate matter from diesel engines and greenhouse gas emissions (GHGs). The previously discussed growth in the economy and freight travel has also meant a large increase in freight transportation related vehicle emissions. In 2005, 28% of all U.S. greenhouse gas emissions were caused by the transportation sector. A total of 19.1% of transportation sector GHG emissions were attributed to freight trucks while a further 28.1% of these emissions were attributed to light trucks, some of which are used for urban freight movement. Also, emissions by freight trucks have increased by 69% since 1990. This is much higher than the increase of emissions by other transportation modes (Davies & Facanha, 2007).

Road freight emissions have followed a similar trend in Canada. In 2009, 27.5% of Canada’s total greenhouse gas emissions were attributed to the transportation sector. Greenhouse gas emissions attributed to light-duty trucks and heavy-duty vehicles were 6.54% and 6.26% of total Canadian emissions during that year (Environment Canada, 2011, p. 3). In 2008, transportation emissions due to light and heavy trucks increased by 110% and 68% respectively compared with 1990 values. Figure 1.1 shows the increase in GHG emissions for light and heavy duty trucks over the last 20 years in Canada.

Besides congestion and emissions, another externality of increasing freight transportation is increased damage to the road infrastructure. Heavy trucks cause disproportionate damage to roads compared with other types of vehicles. To provide an example, the Equivalent Single Axle Load (ESAL) method predicts pavement damage by converting the damage caused by individual vehicles to the damage caused by an “equivalent” load, usually taken as the 18,000 single axle loading. The equivalence ratio for a 2,000 lb single axle is 0.0003 while it is 7.9 for a 30,000 lb single axle. To put this in perspective, the road damage caused by a single 18,000 axle (the reference value) is equivalent to approximately 3,333 2000-lb axle loadings from passenger vehicles (Pavement Interactive, 2008).
Figure 1.1: Canadian Greenhouse Gas Emissions for Light-Duty and Heavy-Duty Trucks Between 1990 and 2009 (data from Environment Canada, 2011)

Figure 1.2 shows how single-unit trucks and tractors are travelling increased distances on Canadian roads and also how government expenditures on roads are increasing in Canada. This increased distance by heavy vehicles means that roads become damaged more quickly. This is reflected in the cost of road expenses, which has steadily been increasing in Canada from $13.8 billion in 2002/03 to $29 billion in 2010/2011 (Transport Canada, 2012).

Figure 1.2: Vehicle Kilometers Travelled for Single-Unit Trucks and Tractors in Canada (data from NATS, 2011) and Government Expenditures on Roads in Canada (data from Transport Canada, 2012)
The freight transportation system is also changing rapidly as shippers, carriers and receivers respond to technological improvements in transportation, communications and information technology, and to changing customer demands. Various factors for these changes include: (Hensher & Figliozzi, 2007)

1. Improvements in transportation, communications and information technologies have radically altered supply chains
2. Increasing use of pull (order-driven) logistics
3. Advent of just-in-time and lean production techniques, where firms minimize inventory costs by using an increased number of small shipments
4. Higher demand for short delivery times, which is causing an increased preference for fast transportation modes of smaller shipment sizes

According to Turnquist (2008), firms are changing their supply chain strategies to reduce their inventory requirements to save warehouse space and costs. As part of reducing inventory costs and to respond to an increasing use of pull logistics, firms are using faster delivery of products using smaller and more frequent shipments. An extreme example is the Just-In-Time business process that aims to minimize inventory costs by having materials arrive on site just before they are needed. Changing business strategies is not only changing the shipment sizes and frequencies, but also has a significant effect on location and supplier decisions. Carriers have also improved the management of their operations by optimizing routes.

Policy makers and planners need tools that can assist them in making high-quality decisions with regards to infrastructure planning and policies that meet the needs of all users of the transportation system, both passenger and freight, while managing externalities. Freight travel demand models are one such tool that allows planners and policy makers to predict future freight transportation patterns. To provide high-quality results, such models should be capable of capturing changing freight transportation patterns and also account for effects of infrastructure and policy changes.

1.2 Importance of Freight Demand Modelling

Freight travel demand modelling has received less attention than passenger travel modelling. Forecasting future commercial travel demand, however, remains an important activity for governments due to the need to prioritize infrastructure improvements in a budget constrained environment and to formulate policies that encourage efficient commercial vehicle travel while minimizing externalities such as congestion and pollution.
Southworth, Meyer & Bronzini (2008) surveyed multiple state transportation departments to identify freight forecasting tool requirements. The following list of decision-making issues considered by public officials at metropolitan and local levels is adapted from this paper. Addressing all of these concerns is likely infeasible for a single freight model, however, ideally freight models would be predictive enough to address at least some of these issues.

**Infrastructure**
- How to locate and remove bottlenecks, especially for road freight
- How to jointly use limited rights of way
- How to address separate rights-of-way and lanes
- How can the public sector invest in system technologies that could be used by the private sector to more efficiently move trucks through the system
- How to alleviate congestion issues that commonly occur at or near freight terminals, such as large airports, seaports and rail terminals

**Transportation Pricing**
- How to estimate potential revenue generation for tolled facilities
- How to use pricing, and dynamic pricing in particular, to affect the level of congestion of a particular facility

**Land Use**
- How to estimate transportation effects of different freight land uses

**Economic development and transport infrastructure funding**
- How to quantify the relationship of freight and logistics to economic development
- How to arrange public-private financing of transportation infrastructure
- How to allocate costs equitably between public and private institutions

**Impacts**
- How to estimate environmental, public health and quality of life impacts of freight flows

According to Gong & Guo (2011), the majority of freight models used in practice are derived from the “four-step” modelling procedure used to model passenger flows. Different types of “four-step” models are described in more detail in Section 2.1.2. Multiple authors (e.g. Holguin-Veras & Thorson, 2000; Hensher & Figliozi, 2007; Samimi, Mohammadian & Kawamura, 2010) have argued that the four-step modelling procedure is not suitable for freight transportation. Deficiencies of using four-step models to
forecast commercial vehicle travel are detailed in Section 2.1.3, but particular deficiencies are summarized below.

1. Decisions about freight travel can be made by multiple firms. Different actors (e.g. shippers, carriers, receivers and freight-forwarders) may be responsible for different aspects of shipment planning (Boerkamps, van Binsbergen & Bovy, 2000; Liedtke, 2009; Roorda, Cavalcante, McCabe & Kwan, 2010; de Jong, Vierth, Tavasszy & Ben-Akiva, 2012).

2. Four-step models use a single model structure for movement of all commodities in spite of a wide variety between the logistics associated with different commodities, for example agricultural goods compared with computer chips (Wang & Holguin-Veras, 2008).

3. Shipments are not independent from one another as firms optimize their supply chains through supplier selection, shipment frequency, use of consolidation and distribution centres and consolidation of multiple shipments on a single vehicle (Hensher & Puckett, 2005). For example, surveys conducted in Denver and Calgary reported an average of 4.97 and 5 stops per tour, respectively (Holguin-Veras & Patil, 2005; Hunt & Stefan, 2007).

4. Services are not included in commodity-based four-step models, which estimate commercial travel based on commodity flows. Calgary surveys revealed that approximately 45% of all reported business stops were made to provide a service and not to deliver or pickup goods (Hunt & Stefan, 2007).

Freight travel, like passenger travel, is a derived demand. Hence, to accurately understand freight travel, modellers must understand what drives freight movements. This involves an understanding of commodity flows between shippers and receivers and also an understanding of the roles of shippers, receivers, carriers and third/fourth party logistics firms in deciding shipment schedules, sizes, modes and routes.

Recently, a number of commercial vehicle travel demand models have been proposed that aim to produce better predictions for freight flows. These models are discussed in greater detail in Chapter 2. There is still a need, however, for additional research to improve the predictive abilities of these models to better address the decision making issues raised earlier in this section while still remaining tractable and estimable given data availability.

1.3 Importance of High-Quality Data

To create high-quality freight transportation models, analysts require high-quality data. A lack of data availability is recognized as a major impediment to the development of high-quality agent-based
Section 1.3.1 of this introduction provides a brief overview of data sources that are traditionally used to estimate commercial vehicle travel demand models. Section 1.3.2 then introduces passively-collected GPS data, a data source that is currently becoming more widely available, while Section 1.3.3 argues that passively-collected GPS data are a suitable data source to estimate components of an urban travel demand forecasting model.

1.3.1 Traditional data sources used to estimate commercial vehicle travel demand models

The following description provides a brief overview of existing data sources that can be used to estimate commercial vehicle travel models.

Economic data can be used to estimate input/output models, which provide a matrix representation of the interconnectedness of industries, households and governments in an area through the total dollar value of shipments between industries.

Another potential data source is from publicly available surveys on freight shipments. Unfortunately there are no Canadian surveys that are equivalent to the U.S. Commodity Flow Survey (CFS). The CFS is a shipper-based survey and is conducted every five years as part of the Economic Census in the United States and is the primary source of national and state-level data on domestic freight shipments in selected industries (Bureau of Transportation Statistics, 2010). Even if CFS data were available in Canada, it would not provide sufficient resolution for modelling purposes in spite of its wealth of information. According to Turnquist (2008), standard publicly available datasets (such as the CFS) are “woefully inadequate” for understanding freight flows and to create many freight models. They provide a “broad brush picture” to show what happened in the survey year but they provide little basis to model why those freight flows occurred. This opinion is shared in Samimi et al. (2010), who state that these data sources are “limited only to highly aggregated information that is not sufficient for the behavioral model development.”

Roadside intercept surveys can also be used to collect freight data. One such example is the Ontario Commercial Vehicle Survey (CVS), whose objective is to obtain information about freight flows on the provincial highway system. Drivers are asked questions related to the origin, destination, routes used, goods carried, vehicle dimensions and shipment weights. This survey is conducted at truck inspection stations, rest areas, road maintenance yards and at border crossings. These surveys are expected to be of use for large-scale modelling efforts but also lack the required resolution for regional and urban...
commercial vehicle models. Also, vehicles with a Gross Vehicle Weight of below 4500 kg are not intercepted, hence the CVS cannot provide information about the travel behaviour of lighter vehicles (HDR-iTrans, 2011).

Regional and urban commercial vehicle models are often based on custom commercial establishment surveys, such as those conducted in the Region of Peel, Region of Durham and the Greater Toronto and Hamilton Area (GTHA) in Southern Ontario (Roorda et al., 2008; HDR-iTrans, 2010; Roorda, Rashidi, Bachmann & Rudra, 2013). The Region of Peel survey asked for information about the firm and about all inbound and outbound shipments on the survey day. Drivers were also asked to complete a form describing their travel activities (McCabe, 2007). This is a lot of information and it is burdensome upon responding firms. Some respondents may also consider this information to be proprietary as it allows identification of their suppliers and customers, including the values and sizes of shipments with these partners. Hence survey response rates are low. Estimated response rates for this and for similar surveys vary between 5% and 25% (McCabe, Kwan & Roorda, 2013). Those authors also found evidence of respondent fatigue as comparisons with GPS records from sample drivers indicated that 13% of drivers were found to have truncated their driver forms and also found evidence of unreported stops. Due to the high response burden, survey durations are usually limited to a single day. This is also true for other recent freight surveys such as in Atlanta, Detroit, Greensboro, Alberta and Ohio (Outwater et al., 2005; Cambridge Systematics, 2003; Hunt, Stefan & Brownlee, 2006; Gliebe, Cohen & Hunt, 2007b).

Hence, it can be seen that lack of data is a serious issue for estimating freight models. Publicly available data, such as input/output tables and government commodity flow surveys are useful to provide broad aggregate information about the economy and about freight flows, but by themselves they do not provide sufficiently fine resolution to understand and predict individual freight movements. Establishment travel surveys are a viable means to obtain such data but they are expensive, require large amounts of effort to implement, suffer from low response rates and are burdensome on respondents, and hence are effectively limited to a single survey day.

1.3.2 Passively-collected GPS data collection for commercial vehicle travel

According to Southworth (2002, p. 4-25), “new ways of collecting freight data need to be explored, including less obtrusive data gathering methods focused on administrative records and remote sensing. We cannot, it seems, rely on shipper and carrier surveys alone.” Due to recent technology improvements, other data sources are becoming available that can supplement shipper and carrier
surveys. Data collected from in-vehicle Global Positioning System (GPS) units provide a means of recording vehicle movements over an extended time period.

Many firms in North America have already installed GPS-equipped Engine On-Board Recorders (EOBRs) on their vehicle fleets. These devices are gaining wider acceptance among carriers as more companies recognize the benefits of using this technology. EOBR systems are used to record driver hours of service in order to ensure compliance with hours of service laws, to automate completion of International Fuel Tax Agreement (IFTA) forms and to allow fleet managers to track their vehicles. As an example, XRS (one company that provides fleet management services) provides GPS tracking services for approximately 114,000 trucks across North America. This number is expected to grow further due to the recent U.S. Congress Moving Ahead for Progress in the 21st Century Act (MAP-21) bill (XRS, 2012). The MAP-21 bill will require all commercial motor vehicles involved in interstate commerce to be equipped with an electronic logging device to improve vehicle operator compliance with hours of service regulations (FHWA, 2012).

Some firms and GPS fleet tracking service providers are willing to provide the vehicle histories recorded by these GPS units to researchers, given that confidentiality and privacy concerns are met. This type of GPS data can be called passively-collected GPS data as the GPS vehicle travel history is the only information that is available to researchers. Additional information is not requested.

Passively-collected GPS data offer multiple advantages compared with travel data collected from travel surveys of commercial establishments. Pendyala (2003) mentions that compared with traditional data collection methods, GPS-based travel data sets are better able to: 1) capture short and infrequent trips; 2) provide accurate temporal information; 3) provide travel information over extended time periods, and; 4) offer detailed route choice, travel itinerary and spatial location, while not placing additional burden upon the respondent.

Another advantage of using data from commercial fleet tracking services is that since many firms are already using GPS monitoring devices in their vehicles, this data source is already widely available (at least in North America). Hence, using only passively-collected GPS data provides a significant advantage over survey data in that sufficient GPS data are likely already available for many municipalities in North America. This means that local data are available with which to estimate and calibrate transportation models instead of relying on models estimated in other locations.

There are issues, however, with relying on passively-collected GPS data from truck-based GPS fleet management services that must be noted. These include:
• GPS data do not contain shipment information such as the weight, volume, commodity type or whether the shipment is direct or is one leg in a consolidated shipment.
• No information is available about unmonitored vehicles.
• Due to privacy concerns, information may not be available about the carrier, visited customers or the vehicle type.
• GPS data from truck-mounted electronic logging devices are only available for trucks and not for other modes.

1.3.3 A case for passive GPS data for urban data collection

The previous section outlined advantages and disadvantages of using passively-collected GPS data from in-vehicle GPS tracking devices. Considering these advantages and disadvantages, passively-collected GPS data are a promising data source for estimating urban commercial logistics decisions travel forecasting models for the following three reasons.

The first reason is that there is a serious lack of currently available data describing urban commercial vehicle travel (discussed in Section 1.3.1). Many of the available freight data collection methods, such as road-side interviews at truck weigh stations, are not suitable for urban settings as these data collection methods cannot reasonably achieve sufficient coverage to observe urban freight travel. Most urban data collection efforts involve surveys conducted on local business establishments. These surveys are time consuming, expensive and suffer from a poor response rate. Also, while surveys in individual cities have been conducted, a lack of consistency in survey data between cities makes it difficult to transfer models. Since GPS data are much more widespread, however, similar GPS data can be obtained in many locations allowing easier model estimation and also the ability to transfer models between locations.

The second reason is that the lack of shipment information is less of an issue for urban commercial travel than for long-distance commercial travel. Firms are expected to pay particular attention to the optimization of long-distance components of their supply chain due to higher transportation costs. It is anticipated, however, that for local deliveries firms will be more responsive to their clients and will not necessarily wait for a full vehicle before making a shipment due to lower shipment costs and penalties for non-optimal shipments. This difference between urban and inter-city delivery patterns has been considered by other researches (such as by Gliebe et al., 2007b) and has also been corroborated through a large number of interviews with managers of large distribution centres in the Greater Golden Horseshoe region of Canada, conducted as part of the research described in Roorda et al. (2013).
Finally, only having truck-based information is a serious issue for long-distance travel, especially since mode choice is of large interest to planning officials. It is generally accepted, however, that the majority of freight movement within a city occurs by truck. Hence, only having data on this single mode of travel is not considered to be an issue in this context.

One other limitation of using EOBR GPS data for freight travel model estimation is very important and must be mentioned up front. This data source has a potential for bias since larger carriers who operate in multiple provinces and states are more likely to use or require EOBRs than smaller and local carriers. It is hoped that as the use of GPS recording becomes more prevalent that the potential for bias will be reduced. While there is a real possibility of bias in this single and relatively small dataset, it must be stressed that the goal of this research is not to present the estimated models as “final”, but to present a modelling framework that can be re-estimated in multiple locations with relatively modest effort.

### 1.4 Research Objectives

The Ph.D. research presented in this dissertation is an exploration in using passively-collected GPS data to develop activity-based models of how carriers undertake the deliveries of contracted urban shipments. These models will be included within a larger agent-based commercial vehicle framework that is currently under development at the University of Toronto. To the author’s knowledge, this Ph.D. research represents the first attempt to estimate components of an agent-based commercial transportation model using passively-collected GPS data as the primary data source. This section provides a breakdown of more specific research objectives in order to meet this overall goal.

The first objective was to research and develop data processing algorithms to convert the raw GPS data into a travel diary of trips, trip ends and tours. New data processing techniques were required as information about carriers (such as the identity of the carrier and of visited locations) were suppressed by the data provider to respect carriers’ privacy concerns. Within this context of little information, data processing steps were required to group GPS observations so that the level of analysis could be shifted from individual GPS observations to destinations. There was also a need to link the destinations with publicly available data sources in order to provide local (preferably down to property-parcel level) attributes required to provide land-use context for model estimation. GPS data processing is the focus of Chapter 5 of this dissertation.

The second objective was to estimate two behavioural models of urban commercial transportation using the processed passively-collected GPS data. Due to the different strengths and weakness of GPS data compared with survey data, the design of the component models needed to be adjusted from existing
models found in the literature. The first estimated model was an activity duration model, which is presented in Chapter 6. The second model uses the longitudinal GPS data to estimate a model of inter-arrival durations, defined as the number of days between visits to the same destination by a carrier. This inter-arrival duration model is presented in Chapter 7 of this dissertation.

The final research objective was to design the *FREight Logistics Decisions* (FRELODE) modelling framework. When completed, FRELODE will model how carriers undertake the deliveries of their contracted urban shipments within the larger agent-based commercial vehicle framework that is currently under development at the University of Toronto. The activity duration and inter-arrival duration models were estimated such that they can be used as two of the components within FRELODE. The FRELODE modelling framework is presented in Chapter 3 of this dissertation.

### 1.5 Thesis Organization

This thesis is organized as follows. Chapter 2 provides a literature review that is divided into two parts. The first part provides an overview of freight transportation models from traditional growth factor and “four-step” models to more recently proposed activity based models that address limitations of the traditional freight transportation models. The second part provides a history of GPS data collection efforts with a focus on passive GPS data collection efforts and also of GPS data collection of commercial vehicle travel.

*Chapter 3: “The FRELODE Urban Logistics Modelling Framework”* starts by outlining the complete agent-based microsimulation commercial vehicle modelling framework that was recently proposed by University of Toronto researchers and describes how the proposed FRELODE model fits into this larger framework. This chapter then presents a high-level overview of FRELODE, describing the component models.

Passively-collected GPS vehicle travel histories are the primary data source used to estimate the models in the FRELODE framework. *Chapter 4: “Data Sources Used for Model Estimation”* first describes the GPS dataset and then describes the additional data sources that were linked to the GPS data to provide a land-use context for the observed GPS data. *Chapter 5: “Processing the GPS Data”* describes the techniques that were used to process the GPS data so that it could be used for freight travel demand modelling applications.

The next two chapters present two models comprising the FRELODE modelling framework. *Chapter 6: “Hazard Models of Commercial Activity Duration on Urban Vehicle Tours”* compares two hazard-based
activity duration models that use the passively-collected GPS data to estimate the length of time spent while making a pickup, delivery, service, or other type of activity.

One of the benefits of passively-collected GPS data is that due to the low response burden, the data can span a duration of months, or even years. This benefit is highlighted in Chapter 7: “Multilevel Modelling of Commercial Vehicle Inter-Arrival Duration”, where the passively-collected GPS data are used to estimate a model of the inter-arrival duration between two visits to the same destination, defined as the length of time between two successive visits at a destination from vehicles operated by the same carrier. This type of analysis cannot be conducted using other data sources, such as shipping surveys, due to the one-day survey period usually employed.

Chapter 8 concludes this thesis by describing the research contributions of this Ph.D. research, policy applications of the proposed model FRELODE framework (including the estimated models) and future work.
Chapter 2: Literature Review

2.1 Literature Review of Freight Travel Demand Models

Chapter 1 highlighted the importance of freight demand modelling and also the importance of high-quality freight data collection, both for modelling purposes and also to better understand the freight transportation system. This section provides a literature review that highlights both state-of-practice freight transportation models and also more recent research that improves upon the standard techniques. The discussion in this chapter focuses on identifying strengths and weaknesses in the different modelling approaches.

2.1.1 Growth factor methods

According to Gong & Guo (2011), most ongoing statewide and urban area freight planning processes adapt either the growth factor or “four-step” modelling approaches. Growth factor methods are described in this section while “four-step” methods are discussed in Sections 2.1.2 and 2.1.3.

Growth factor methods build upon historical trends to predict the future. These models are simple to implement and are not data intensive (Pendyala, Shankar & McCullough, 2000). The Quick Response Freight Manual II (QRFM2) lists two types of growth factor methods: growth factors based on historical freight trends and growth factors based on direct economic projections (Cambridge Systematics, 2007). According to the QRFM2, these methods are commonly used to provide rough estimates of statewide or regional growth for different types of freight demand and also to provide an estimate of freight transportation demand within a corridor or at an intermodal facility.

The first method, growth factors based on historical freight trends, is the more commonly used growth factor method. An analyst can either assume that freight growth follows a linear or a compound curve and then derives an equation for the growth rate. This method is considered to be risky, especially for long-range forecasts, as it does not consider any underlying economic mechanisms that change the demand for freight.

The second method, growth factors based on direct economic projections, recognizes that demand for freight transportation is related to underlying economic indices such as employment, population and income. This method requires estimates of freight traffic and also of economic indicators (such as economic or employment indices) for a base year and forecast economic indicators for future years.
Ideally the economic data and forecasts are available for different commodity groups. The growth in freight traffic is calculated from the projected growth of the economic indicator.

While growth factor methods are a simple and inexpensive way to forecast freight, they are most suitable for analyzing incremental changes in freight activity. These methods assume that all relationships between freight and economic activity remain constant and cannot account for changes in manufacturing and service technologies and also how firms manage their supply chains. Also, these methods do not explicitly incorporate important explanatory variables and hence cannot account for changes in infrastructure, land-use patterns or policy applications such as zoning changes, parking fees, parking availability, fuel taxes and road pricing. Also, these models cannot account for productivity improvements, such as the output per ton of commodities, or larger changes in supply-chain structures.

2.1.2 Traditional “four-step” models

The “four-step” travel modelling methodology was developed for passenger travel and divides travel demand into the following four steps: production and attraction, trip distribution, mode split and assignment. There is some debate in the literature about the applicability of four-step models for commercial vehicle applications. Some authors (e.g. Pendyala et al., 2000; De Jong, Gunn & Walker, 2004; De Jong et al.; 2012) argued that four-step modelling approaches can reasonably be applied to model freight transportation in spite of their issues. Other authors focused on the limitations of four-step models and argued that they lack sufficient behavioural representation (e.g. Boerkamps et al., 2000; Hensher & Puckett, 2005; Hensher & Figliozzi, 2007; Hunt & Stefan, 2007; Donnelly, Mayburg, Shen & Leachman, 2008; Samimi et al., 2010). Limitations of four-step transportation models are discussed in Section 2.1.3 and form the basis for more recently proposed models, several of which are discussed later in this literature review.

Although the four-step model for freight transportation is similar to that for passenger travel the individual steps may vary considerably from their passenger counterparts. Two types of “four-step” models were presented in the Quick Response Freight Manual II, prepared by Cambridge Systematics (2007), commodity flow models and truck trip models. Both of these models are aggregate in nature and focus on the flow of vehicles between traffic analysis zones.

Commodity Flow Models

Commodity flow four-step models recognize that freight travel is a derived demand whose main purpose is to transfer goods between suppliers, receivers and finally to the end customer or retailer. Hence accurate models of freight travel must incorporate flows of goods in the economy. The main
steps in a commodity flow four-step model are: production and attraction, distribution, mode split, truck conversion and network assignment. The described steps are summarized in the flow chart shown in Figure 2.1.

![Commodity Flow "Four-Step" Model Framework](image)

**Figure 2.1: Commodity Flow "Four-Step" Model Framework (based on Cambridge Systematics, 2007)**

**Production and Attraction**

This first step estimates the quantity of goods transported from origin zones and the quantity of goods demanded by destination zones within the analysis period. This output value is usually given in tons although intermediate steps may use dollar values, depending on the input data. Many estimated models use economic Input/Output (I/O) tables; examples of such models in a European context were reviewed in de Jong et al. (2004). In such cases, it is required to convert the productions and attractions from dollar values to tonnage.

Another approach was shown in Cambridge Systematics (2007), which illustrated the Indiana Freight Model. Due to large differences in freight travel patterns between different commodity types, independent models were created for each commodity. This model used a zonal-level regression formulation that predicted zonal-level production and attraction for 43 different commodities based on zonal employment, classified by NAICS3 industry categories, and the zone population. Another example
model is from Wisconsin, where separate commodity generation and distribution models were estimated for 25 different commodities (Proussaloglou, Popuri, Tempesta, Kasturirangan & Cipra, 2007).

Depending on the model scope, special generators may be required for large freight facilities such as ports and railway intermodal yards. Production and attraction to external zones must also be included.

Distribution
This step predicts the quantity of goods shipped (in tons) between origin and destination zones during the analysis period. Different methods can be used to estimate the commodity flow distribution. Gravity models are most commonly used while Fratar models are less common (Cambridge Systematics, 2007). When multiregional I/O data are available they can also be used to estimate commodity distribution, such as in the Italian national model (de Jong et al., 2004). Other approaches mentioned in the literature include the fractional split distribution model of Sivakumar & Bhat (2002), which uses a multinomial logit model to determine the fraction of a commodity consumed in each attraction zone that is produced within each production zone. Covariates of this model include zonal size attributes (such as employment by sector) and a measure of the impedance between zones.

Mode Split
Including a modal split model allows forecasting the mode split changes over time. The output from this step is the tons of flows between origins and destinations by mode. Logit models are usually used (Cambridge Systematics, 2007).

Truck Conversion
After the mode split, the unit of interest is still in tons. For road shipments, however, the unit of interest is the number of vehicles. This step converts the commodity tonnage for road shipments into vehicle flows. This step can also be completed for the rail network if railway assignment is also of interest. Cambridge Systematics (2007) provides different examples of the average payload weight carried by truck, or railcar, and typical conversions from annual to daily flows.

Network Assignment
Truck travel is then assigned to the network. Network assignment is usually conducted simultaneously for truck and passenger vehicles using a user-equilibrium multiclass traffic assignment.

Truck Trip Models
Instead of modelling vehicle flows by analyzing shipped commodities, truck trip models directly estimate road vehicle travel. A detailed description of these models is available in Cambridge Systematics (1996).
The basic steps are the same as the standard “four-step” model except that the mode split step is usually omitted as only road travel is usually considered. The mode split model is sometimes replaced by a vehicle size model, which estimates the fraction of trips between zones undertaken using different vehicle classifications. The described steps are summarized in the flow chart shown in Figure 2.2.

![Flow Chart: Truck Trip "Four-Step" Model Framework](based on Cambridge Systematics, 2007)

**Production and Attraction**
This first step estimates the number of truck trips produced by every origin zone and the number of truck trips attracted to every destination zone in the analysis period. These are estimated directly instead of first estimating freight flows, as is done using commodity flow models. Regression models are usually used for these estimations. Trip productions and attractions are usually calculated using regressors such as the number of employees in the zone and establishment floor areas by industry type. Different trip rates are often assumed for different industry sectors. Usually separate regression models are estimated either by vehicle classification or by commodity type. Depending on the model scope, special generators may be required for large freight facilities such as ports and railway intermodal yards. Production and attraction to external zones must also be included.

**Distribution**
A gravity model is usually used to distribute the produced and attracted vehicle trips between zones. If productions and attractions are estimated for different vehicle types in the first step, then different gravity models are usually estimated for each vehicle class since heavier vehicles are usually used for longer trips. Likewise, if productions and attractions are estimated for different commodity groups then a separate distribution model can be estimated for every commodity. The trip distribution models are calibrated using the observed vehicle trip length distribution.
Network Assignment

Truck travel is then assigned to the network, usually as part of an equilibrium multiclass assignment in combination with passenger travel.

Comparison of commodity flow vs. truck trip models

While commodity flow and truck trip models both use the four-step modelling framework, application of these models is markedly different. This section compares the relative advantages and disadvantages of these models. Section 2.1.3 contains a more general discussion about the suitability of the four-step framework for modelling freight flows.

Ideally, freight models should incorporate all of the main dimensions (e.g. weight, volume, value and vehicle trips) of freight into account (Holguin-Veras & Thorson, 2000). According to Holguin-Veras & Thorson, neither commodity flow nor truck trip models are able to capture the full complexity of freight movements because both models only include one dimension of freight demand.

According to Giuliano et al. (2010), it is widely accepted that commodity flow models record truck flows more accurately than truck trip models due to their focus on supply and demand. Hence these models are expected to be more realistic and more robust than truck trip models. Commodity flow models have been developed for many states, see Giuliano et al. for examples, as well as many national and international European models, described in de Jong et al. (2004, 2012).

According to Giuliano et al., commodity based models have not been developed at the metropolitan level primarily due to a lack of data. For example in Canada, Input/Output models are available only at provincial and national levels and hence are not suitable data sources with which to estimate urban transportation models.

Another issue with commodity flow models is that they cannot handle empty trips since these trips do not ship commodities. According to Hogluin-Veras & Thorson (2003), empty trips represented between 30 and 40% of the total trips observed in their case study. This issue of empty trips can be resolved using methods described in Holguin-Veras & Thorson (2003) and in Holguin-Veras & Zorilla (2006).

The focus of truck trip models on vehicle trips instead of commodities is both a strength and a weakness. One main advantage of focusing on vehicle trips is that there is a significant amount of available data such as from traffic counts, screen-line counts and data output from intelligent transportation systems (Holguin-Veras & Thorson, 2000). This is unlike commodity flow models, where
finding accurate commodity flow information is difficult in urban settings. Also, there is no need to separately consider empty trips as these are automatically included.

Weakness of truck trip models include that they do not model underlying economic supply and demand, since they do not consider commodity flows. Also, applications of these models have found that truck generation rates are not easily transferrable between different regions (Giuliano et al., 2010), although other studies have shown that establishment-level models models are transferable (Holguin-Veras et al., 2012). Finally, they are of questionable applicability in situations where multiple freight transportation modes are considered (Holguin-Veras & Thorson, 2000).

2.1.3 Discussion of aggregate four-step models

There has been a large amount of discussion in the literature about the suitability of the passenger-based four-step modelling framework for freight transportation. “The export of modelling ideas from passenger transportation has been called into serious question and as a result we are seeing innovative new research perspectives that seek to understand the complexities of interaction between the key stakeholders in freight distribution... and the way in which the activity-based methods might aid in understanding urban truck flows” (Hensher & Figliozzi, 2007 p. 921).

Issues with this modelling approach include that it cannot account for: 1) the diversity of actors involved in freight shipment decisions, 2) the diversity of freight shipments, 3) that shipment are not planned independently but within a larger supply chain, and 4) that services are not explicitly included in these models. These issues are discussed in this section.

Unlike passenger travel where travel decisions are made by one person, or at least one household, decisions about freight movements can be made by multiple firms. Typical actors that are assumed to play a role in the freight process are shippers, receivers, carriers and third-party logistics firms (3PLs), although any one actor may take multiple roles. Individual actors may be responsible for different aspects of the decision making process. Also, individual actors may not have information about the entire shipment (Boerkamps et al., 2000; Liedtke, 2009; Roorda et al., 2010; de Jong et al., 2012).

One of the issues in freight modelling is the diversity of freight. Expected shipment behaviour is expected to vary drastically between different types of commodities. These diverse commodities encompass a wide spectrum ranging from low value commodities, such as agricultural products, to high-value commodities, such as computer chips that are valued at over one million dollars per ton (Wang & Holguin-Veras, 2008). These differences can be considered within the “four-step” modelling process by
estimating different trip generation, trip distribution and mode choice models for different commodities. The same general process, however, is assumed for all commodities.

The lack of supply chain representation is another common criticism of four-step models. Firms reduce their combined inventory and transportation costs through optimizing their supply chains. Optimization decisions include supplier selection, shipment size and frequency, use of consolidation and distribution centres, and the combination of multiple shipments within a single vehicle.

One result of these optimizations is that the path taken by a shipment may not follow the same pattern as if this same shipment was planned independently of other shipments. For example, shipments may not be sent directly from the origin to the destination. It may be more economical to route shipments through consolidation centers where they can be bundled with other shipments. This bundle could then be transported to a distribution centre where the individual shipments are unpacked. Finally, each individual shipment would then be sent to its final destination.

Melendez & Horowitz (2011) studied responses from the Ontario Commercial Vehicle Survey (CVS) and found that 52% of origins and 54% of destinations were considered to be transshipment points and that in 36% of trips both the origin and destination were at a terminal or a warehouse. These trips would have at least three legs. Manufactured goods are usually shipped to or from transshipment points, as 83% and 72% of these shipments were transshipped at the origin and destination respectively. As would be expected, the use of transshipment points occurs more frequently for longer trips than for shorter trips. (It should be noted here that the CVS is an intercept survey where the survey sites were at truck weigh stations and rest areas located on freeways. Hence this survey is biased towards longer distance trips and heavier vehicles.)

Urban freight is characterized by multi-stop tours. Holguín-Veras & Patil (2005) found that in Denver most vehicles make one tour per day with an average of 4.97 stops per tour. About 35% of trip chains involve 2 stops while some have over 12 stops. Similar findings were also observed in Calgary by Hunt & Stefan (2007), who found an average of approximately five stops per tour from a commercial establishment survey conducted in Calgary Alberta. Figure 2.3 represents how commodity flows and vehicle flows may bear little resemblance to each other in the case of trip chaining. In this figure, a vehicle makes deliveries to five different locations. In a four-step model these would be modelled using five distinct trips, each going from the base to one of the destinations. As is seen here, the prevalence of trip chaining breaks down assumptions of traditional trip distribution models that trips between zones are functions of attributes of the origin and destination zones. Hunt & Stefan (2007) also note
that there is no “primacy” in freight tours. Personal tours are usually comparatively simple (compared with commercial tours) and tend to be organized around work or school. Commercial tours do not have the “skeleton” around which other stops are planned.

![Figure 2.3: Truck vs. Commodity Flows for a Single Delivery Tour (Wang & Holguin-Veras, 2008).](image)

The final issue with many freight transportation models is that they do not include services. According to Hunt & Stefan (2007), Calgary surveys revealed that approximately 45% of intra-urban business stops were made to provide a service. These trips are completely missed in a commodity flow four-step model but may be included in trip based models as these models do not distinguish between trips that deliver goods and services.

The following sections illustrate recently proposed freight models that attempt to alleviate some of these deficiencies. Different models focus on addressing different deficiencies. To the author’s knowledge, none of the proposed models described in this chapter simultaneously addresses all of these issues. Section 2.1.4 discusses models that incorporate supply chains, Section 2.1.5 describes City Logistics models, which include supply chain characteristics into urban freight models, Section 2.1.6 describes Tour-Based models, which do not include supply chains but focus on including service trips and on creating representative tour structures. Hybrid models include elements of supply chain and tour-based models; these are described in Section 2.1.7. Finally, Section 2.1.8 overviews two commercial vehicle travel models that do not fit in the above categories.

---

1 Transportation Research Record: Journal of the Transportation Research Board, No. 2066, Figure 1, p. 2, Washington, D.C., 2008. Reproduced with permission of the Transportation Research Board.
2.1.4 Supply chain (logistics) models

As is discussed in the previous section, “four-step” models do not incorporate supply chain elements. This is acknowledged by many authors to be a weakness in these models as supply chain considerations influence shipment size and frequency, mode choice and the use of consolidation/distribution centres. For example according to Southworth (2002, p. 4-21), “forecasting freight demands is unlikely to be successful unless some understanding of these rapidly evolving supply chain logistics is built into the planning process”. Three different models in this category include a logistics module incorporated in the national freight transport models of Norway and Sweden (de Jong & Ben-Akiva, 2007), the INTERLOG model by Liedtke (2009) and the FAME model of Samimi et al (2010).

Logistics framework for Norway and Sweden freight transportation models

De Jong & Ben Akiva (2007) described a logistics elements module that determines shipment size, mode choice and the use of consolidation/distribution centres. This module forms one component within a freight transportation model that includes:

1. Generation of production-wholesaler-consumption (PWC) matrices that provide aggregate flows of goods by commodity type between producer, wholesaler and attraction zones. These matrices can be generated using spatial input-output models or by spatial computable general equilibrium models.
2. A disaggregation of the zone-to-zone PWC matrices to firm-to-firm flows
3. The logistics elements framework proposed in this paper, which converts the firm-to-firm flows into vehicle flows. The decisions made in this framework include: 1) shipment frequency and size, 2) use and location of consolidation and distribution centres, 3) number of legs in the transportation chain, and 4) mode used in each transportation chain leg.
4. Network assignment

The shipment size and frequency were selected using an inventory logistics model. This model assumes that inventory decisions (with regards to shipment frequencies) are made by the receiver while production inventories are set by the producer. For a given annual flow, \( Q \), between a disaggregate producer and consumer it is not required to model both shipment size and frequency as \( Q = f \times q \) where \( f \) is the shipment size and \( q \) is the order quantity. This shipment size model calculates optimal shipment order quantities considering variables such as: order costs, transportation costs (including consolidation and distribution points), deterioration and damage during transportation, capital costs of goods in transit, inventory costs, capital cost of goods in inventory and stockout costs.
The Transport Logistics model selects the transportation chain. Decisions include whether to use a direct or a consolidated shipment. The model also selects the mode used for each shipment leg and the location of the distribution/consolidation centres (if required). The authors use a random utility model to select between logistics chains. The only variable in the utility function is the total cost of the transportation chain. The total cost of the transportation chain is the sum of the costs for the link plus costs that apply to the entire chain. Costs are reduced for different legs due to consolidation of shipments into the same vehicles or vessels in consolidation centres. To reflect consolidation the transportation cost on a leg is calculated as the shipment’s share of the total cost of operating the vehicle. Hence the vehicle load factors are endogenous in the model. Since shipment costs vary with the behaviour of other shippers, this model is solved using an iterative computation procedure so that every shipper has updated transportation costs. The authors accounted for empty trips by enforcing vehicle balances and generating extra trips that return vehicles to starting zones.

INTERLOG model

The INTERLOG model, presented in Liedtke (2009) is a long distance goods movement model of shipments conducted by independent carriers. This model simulates the auction of transportation contracts and includes feedback regarding shipment-size decisions and tour construction. This is a “bottom-up” model that creates heterogeneous actors using behavioural rules, and simulates their actions and their interaction with other actors using an agent-based simulation. This model does not include local private-fleets, who mainly includes pickup and delivery tours.

Liedtke claims that the advantages of an agent-based approach is that all freight based decisions (such as supplier choice, logistics network design, shipment lot size and mode choice) can be “reviewed by logistics actors to evade the effects of a policy measure.” This model is designed to address questions such as shifting freight from road to rail, off-peak transportation, road pricing, truck bans at certain times of the day, cordon pricing in cities, and subsidies for intermodal terminals.

Steps in the INTERLOG model include firm generation, supplier sourcing and market interaction modules. The firm generation module uses Monte Carlo simulation to generate a scaled-down population of shippers, receivers and carriers. Each firm is characterized by the number of employees, economic activity and location. The supplier sourcing module simulates the demand (in metric tons per year) between actors. Using a “pull” logistics system, this module selects suppliers for every receiver in a manner that enforces total production and consumption by industry sector. Freight generation rates were deduced from production statistics and attraction rates were calculated using input-output matrices of the flows between industry sectors.
The *market interaction* module assigns how goods are divided into individual shipments. Shippers award shipments bundles in contracts that involve pre-defined cargo at certain locations, with constraints such as lot-size, frequency, time windows, weight and compatibility. Carrier dispatching creates daily tours. A carrier sets contract prices from its marginal and full transportation costs with the additional shipments.

Shipment lot sizes are selected by minimizing a total logistics cost function (TLC) to select optimal shipment sizes considering inventory costs, demand of goods (metric tons per time interval), transportation rate, and ordering and handling costs. Due to imperfect information, a shipper may select non-optimal shipment sizes.

The above market interaction and tour-creation models were estimated using constraint-logic programming (CLP) to find feasible solutions for combinatorial problems through search rules. Through analyses of small-scale simulations, the INTERLOG model was found be responsive policy changes. The CLP algorithm used by INTERLOG is capable of formulating and solving practical complex optimization problems. The required computational time, however, increases exponentially with the number of actors and hence this approach is less practical for large-scale simulations due to its complexity.

**FAME model**

The FAME model of Samimi *et al.* (2010) introduces an activity-based freight movement microsimulation framework that incorporates supply chain management concepts and highlights the critical role of individual firms as the primary decision makers. The FAME model was estimated using the Freight Analysis Framework (FAF), national-level input-output tables and also a national online establishment freight survey conducted by researchers at the University of Illinois at Chicago that asked respondents for:

1. **Establishment information:** such as location, number of employees, value of total annual shipments, warehousing, access to intermodal facilities and potential use of different modes
2. **Information on five recent shipments:** including origin, destination, mode, use of consolidation/distribution centres, commodity type, weight, value, cost of shipment, haul time and decision maker

FAME is composed of five modules. The *firm generation module* synthesizes firms including characteristics such as industry type, location, number of employees and floor area. Firms are aggregated by firm type to make a more tractable model and to remove the need for disaggregate data. The *supply-chain replication module* connects firm types, determines the amount of incoming and
outgoing goods between firm types and replicates high-level supply chain strategies such as supplier selection and a dominant form of the supply chain. The *shipment forecasting module* calculates shipment frequencies and sizes. The *logistics planning module* makes decisions such as selection of consolidation and distribution centres, whether to hire a 3PL firm that will arrange the shipment and mode selection. Finally, the *network analysis module* converts shipment tons to vehicle flows, performs traffic assignment and evaluates network performance measures.

**Summary of supply chain models**

The supply chain models that are discussed in this section focus on national or international scales. The Logistics Framework for Norway and Sweden and the INTERLOG model generally assume that firms optimize their supply chains through load consolidation, mode selection and through selection of optimal shipment sizes. These models make use of data sources targeted for larger regions such as input-output tables and the Freight Analysis Framework (FAF) in the United States.

According to Fischer, Outwater, Cheng, Ahanotu & Calix (2005), supply chain models are a significant improvement over commodity-based models as they have a better theoretical foundation and they include how freight moves in a logistics chain, from producers to distribution facilities and finally to consumers. These models represent decisions of shippers, receivers and carriers for the use of intermediate handling facilities, mode selection in each transportation leg, and selection of shipment sizes and frequencies. These models can predict how changes to the transportation system can affect the logistics system, and vice versa. Supply chain models, however, have difficulties for mixed shipments of goods, which Fischer et al. (2005) identifies as a defining characteristic of local pickup and delivery movements. Intermingling between different industries and commodities makes it difficult to distinguish the component logistics chains without large data collection efforts. Also, these models are unable to capture travel for service purposes.

According to Hunt & Stefan (2007) these models are primarily useful in the case where there is a limited range of industries, which may occur in an area dominated by one type of company or economic sector. These models would require, however, substantial data to produce a comprehensive model for a typically diverse North American urban area. Similar arguments are made by Gliebe, Cohen & Hunt (2007a), who claim that supply chain simulations seem “too complex and narrowly focused on specific sub-sections of the full urban economy to be directly applicable in the development of practical tools for urban-level analysis.”
2.1.5 Urban logistics models

Chow, Yang, & Regan (2010) described the GoodTrip model (Boerkamps et al., 2000) and the Tokyo model (Wisetjindawat, Sano, Matsumoto & Raothanachonkun, 2007) as “Urban Logistics” models. These models incorporate logistics behaviour with a focus on urban settings.

**GoodTrip model**

The GoodTrip model of Boerkamps et al. (2000) is a “demand driven, commodity-based freight movement model that incorporates supply chains”. It applies a “four-step” modelling approach to urban supply chains, which positions this model “somewhere in between aggregated four-step models and disaggregated logistics models.” The GoodTrip model is centered on four markets. These markets are modelled sequentially.

The *commodity trade market* determines supply and demand between firms. The commodity trade market creates supply chains starting from raw materials suppliers and ending with the final consumers. Starting with consumer demand, the model calculates the volume of goods demanded of the different goods types in every zone. In between these two categories, goods may flow through different combinations of producers (who create products to sell); trading companies (such as importers and wholesalers; and retail outlets.

The *transport services market* determines the distribution channel, or how goods flow between firms. A distribution channel is defined as a connection between a shipper and receiver. The most basic distribution channel is a direct delivery with one vehicle type. More complex channels consist of different modes, distribution centres and transportation companies, and hence have multiple links. Shipments of different, but compatible goods types can be grouped. Tours are formed for the grouped shipment, a decision that also includes the mode, vehicle type and delivery frequency. The output of this market is the goods movement between locations by mode. Shipment frequency (and hence lot size) is also included in this market.

Finally, the *traffic services market* assigns vehicle traffic to the network and the *infrastructure market* models infrastructure changes. As the infrastructure network develops slowly, this market is not included in framework but is considered to be autonomous.

The GoodTrip model also considers the roles of different actors (shippers, receivers and transporters) when modelling each of the above markets. For example, shippers are assumed to be responsible for
choice of mode, vehicle size and consolidation while receivers are responsible for the volume and demanded goods and shipment size.

The authors claim that this model can evaluate policy and economic effects such as changes in: consumption patterns, supply chain structures (globalization, teleshopping), delivery requirements (smaller and more frequent shipments), distribution patterns (changes in competitive status between modes, tour formation) and environmental improvements (e.g. cleaner engines and transportation system efficiency improvements). Many of these changes cannot be predicted using an aggregate “four-step” modelling system.

Tokyo model
The Tokyo model of Wisetjindawat, Sano & Matsumoto (2006) and Wisetjindawat et al. (2007) is a microsimulation model of urban freight movement that considers the behaviour of different freight agents and their relationship in the supply chain. This model is a modification of the commodity based “four-step” model that considers the behavior of each freight agent individually. The steps in this model include the standard steps in a four-step model: 1) commodity production and consumption, 2) commodity distribution, 3) conversion of commodity flows to truck flows, and 4) traffic assignment. The difference is that these steps are run for individual firms instead of using aggregate, zonal, information.

Before running the modified four-step model, firms (and their attributes such as location, number of employees and floor area) were generated using microsimulation techniques based on aggregate data.

The commodity generation step estimates the amounts of commodity production and consumption of each firm using linear regression based on firm variables such as the floor area, location and number of employees. This is different than the GoodTrip model, which calculates productions by zone.

Instead of the commonly used gravity model the commodity distribution model uses a fractional split distribution method, originally developed by Sivakumar & Bhat (2002), which calculates the fraction of a commodity produced by shipper $i$ that is shipped to receiver $j$. This modelling technique considers the supply chain structure while modelling freight flows. The probability of receiver $j$ selecting shipper $i$ is given as $P_j^k(i) = P_j(C^k) \cdot P_j(z|C^k) \cdot P_j(i|C^k, z), i \in C^k, z$. $P_j(C^k)$ is the probability of the receiver using a distribution channel, $C^k$. The distribution channel models the type of firm from which a receiver purchases their commodities, for example a retailer may choose to purchase goods from raw material suppliers, producers or wholesalers. $P_j(z|C^k)$ is the probability of the receiver selecting zone $z$ given
distribution channel \( C^k \), and \( P_j(i|C^k, z) \) is the probability of selecting shipper \( i \) given zone \( z \) and distribution channel \( C^k \).

\( P_j(C^k) \) is taken directly from survey data, \( P_j(z|C^k) \) is modelled using a spatial mixed logit model where attractiveness indicators include terms such as the number of firms in the zone and the total amount of a commodity produced in that zone. This term includes spatial interactions among decision makers and among alternative zones. \( P_j(i|C^k, z) \) is modelled using an MNL logit model based on the amount of amount of the commodity produced by firms in that zone.

The third component of the model converts commodity flows to truck movements. This component is comprised of three steps: 1) delivery lot size, 2) carrier and vehicle size choice, and 3) vehicle routing. The delivery lot size, \( L_{ij}^k \), is assumed to be related to the distance between the shipper and receiver, \( L_{ij}^k = a + b D_{ij} \). In this model, \( a \) and \( b \) are two parameters that require estimation.

Shippers are assumed to select the carrier and the vehicle size. This step uses a nested logit model to describe the choice process, where the upper-level model is carrier choice (private truck, rental truck, shared truck and hiring a delivery service carrier) and the lower-level model is the vehicle size (small or large truck). For private and rental trucks, the truck will only carry shipments of that shipper while shared and delivery service carriers’ trucks may deliver shipments of multiple shippers at the same time (consolidated shipments). The final step is vehicle routing, which calculates a sequence of customer visits for each truck in a manner that will minimize travel time while being constrained by driver hours and truck carrying capacity. The fourth model component, traffic assignment, is not discussed in these papers.

### 2.1.6 Tour-based models

Four models that explicitly incorporate tours into urban freight microsimulation frameworks are discussed in this section. The first two models, the Calgary model (Hunt & Stefan, 2007) and the Ohio Statewide Model (Gliebe et al., 2007b) are two intra-urban microsimulation freight models. One particular aspect that separates these two models from other models is that they also explicitly consider service trips. According to Hunt & Stefan, Calgary data indicate that approximately 45% of business stops were made to provide a service. Again according to the Calgary data, cars, vans, pick-up trucks and SUVs were used for over 50% of urban commercial vehicle trips. According to Gliebe et al., “... intra-urban commercial movements are characterized by short-distance movements, more service rather than goods delivery, and a much greater emphasis on rapid rather than cost-efficient response.” Hence
both of these papers argue that models that explicitly represent supply chains are less appropriate for urban regions.

The third model reviewed in this section is the tour generation model of Wang & Holguin-Veras (2008). In this model commodity trip productions and attractions are calculated in a similar manner to existing commodity flow models, as are the distribution of commodity flows between O-D pairs. The difference is that instead of a trip assignment that is performed in existing models, this model generates tours that respect the total commodity productions and attractions of every zone. The final reviewed model, described in Wang & Holguin-Veras (2009), uses an entropy-maximization approach to assign truck tours in a manner that minimizes the total impedance within tours.

**Calgary tour-based urban freight model**

The Calgary commercial vehicle model divides commercial vehicle movements into three groups:

1. *External-internal movements* are where at least one of the trip ends occurs outside of the model area.
2. *Fleet-allocator movements* include newspaper delivery, garbage and recycling pickup, and mail and courier services.
3. *Tour-based movements* account for vehicles carrying comparatively small shipments and/or services. According to the Calgary data, tour-based movements account for approximately 70% of the observed urban trips. Tour-based trips are the focus of the model described in Hunt & Stefan (2007).

The trip tables from these three models are combined with those from a household travel model and fed into a network equilibrium trip assignment model. The resulting congested travel times are fed back into these models. This iterative process is continued until a system-wide equilibrium is obtained.

The data source for the Calgary tour-based commercial vehicle model was a set of extensive interviews of over 3000 transportation businesses in the study area. Data of over 64,000 commercial vehicle trips were collected, including: origin, destination, purpose, vehicle fleet and type of commodity. Each surveyed business was requested to provide shipment and vehicle travel data over a 24-hour period.

The structure of the Calgary tour-based commercial vehicle model is shown in Figure 2.4. The *tour generation model* uses exponential regression based on land-use attributes to calculate the number of tours generated per employee during the day. The per-employee tour-generation rate is used to predict the total number of tours produced in a zone. A separate logit model is used to split tour start times into different, aggregate, times of the day.
The vehicle and tour purpose model uses a joint logit model to simultaneously select between three vehicle types (light, medium and heavy) and between three primary tour purposes (goods, services and other), for a total of nine options from this model. The tour start time model prescribes an exact starting time for the tour within the previously determined starting time period. These are sampled from observed start times classified by establishment category and time period.

The next stop purpose, next stop location and stop duration models are used to incrementally add stops to each generated tour. The next stop purpose model is a multinomial logit model that determines whether the next stop is to deliver goods, provide a service or for an “other” stop, such as vehicle refueling or a “return-to-establishment” stop, which ends the tour. Thirteen different models were estimated for different segments, which are combinations of vehicle type, establishment type and tour purpose. If the next stop is not a return-to-establishment stop, the next stop location model is a multinomial logit model that assigns a zone for the subsequent stop in the tour. Again, different models are estimated for the thirteen different segments. The stop duration model calculates a precise stop duration sampled from observed stop durations differentiated by each of the 13 segments.

One issue with the Calgary Model is the number of parameters, with over 400 parameters (calibrated) in the model, which runs the risk of this model being over-fitted to a location. This model has proven to be tractable, however, and has been used in practice in both Calgary and Edmonton.

---

Ohio Statewide Disaggregate Commercial Model

The Ohio Statewide Model of Gliebe et al. (2007b) consists of four components:

1. A household-based person transportation model (PT-SD). This is a tour-based microsimulation model that accounts for personal work-based trips as well as home-based work and non-work travel.
2. The person long-distance travel model (PT-LD) covers personal and service work trips greater than 50 miles in length.
3. An aggregate commercial model, which focuses on long-distance (> 50 miles) commercial vehicle movements (ACM).
4. The Disaggregate Commercial Model (DCM), which models work-establishment based tours that have at least one commercial stop and where no single trip exceeds 50 miles. This model is the focus of the paper of Gliebe et al.

The structure of the DCM model is shown in Figure 2.5. The worker traveller generation model first generates a list of travellers by industry type (industrial, wholesale, retail, transportation handling and service industries) in each traffic analysis zone (TAZ) and then selects a proportion of the travellers in each industry that travel on a single day. The vehicle assignment model then selects a vehicle type (light, medium or heavy) for each single worker who travels on this given day. The starting time assignment model draws a starting time for each tour from an empirical distribution of start times, classified by industry and vehicle types.

The tours for each travelling worker are then grown incrementally. The day is divided into five minute intervals. In each five-minute interval, the worker has the option to stay at the current stop, return to the establishment or head to a new stop (goods, service, meeting, or other). This decision is made in the stop purpose model, which is a single-level multinomial logit model. If the worker completes their stop, the next stop location choice model selects the location of the subsequent stop in the tour. The model estimates the travel time between the two stops and then returns control to the next stop purpose model when the worker arrives at this next stop. This process is continued until the worker completes their tour.
Discrete choice tour generation model given commodity origins and destinations

Wang & Holguin-Veras (2008) proposed a hybrid set of models that consider vehicle trip chaining activities. They proposed a three-stage modelling procedure. The first stage is to estimate the total tons of each commodity produced by and attracted to each zone. This estimation may require an input-output model, although other commodity flow generation techniques can also be used. The second stage is to distribute the tons of commodities into O-D pairs, either using gravity models or another spatial interaction model. These two steps are the same as the first two steps of the commodity flow “four-step” model. The third stage of the model of Wang and Holguin-Veras replaces the trip assignment stage in the “four-step” model with a tour-generation model that satisfies the estimated commodity O-D matrix. A brief description of this model is provided below.

The model of Wang & Holguin-Veras decomposes tour construction into two components: 1) destination choice and 2) the decision to terminate a tour. The destination choice model selects a zone for the next stop. If goods are still available for pickup in the current zone, then the trip to the selected destination is loaded with the average payload value. Otherwise it is an empty trip. After each trip, the binary logit tour termination model tests whether or not the vehicle returns to its base.

---

1 Transportation Research Record: Journal of the Transportation Research Board, No. 2003, Figure 1, p. 19, Washington, D.C., 2007. Reproduced with permission of the Transportation Research Board.
All of the tours in a region are simulated sequentially. The probability of selecting a specific carrier, \( k \), for a given tour is given by \( p(k) = \frac{F_k}{\sum_{i=1}^{K} F_i} \), where \( F_k \) is the total number of trucks operated by carrier \( k \) and \( K \) is the total number of carriers in the freight network. In their model, only a single commodity was transported, all trucks were identical and every carrier was assumed to have enough trucks to satisfy the demand for their shipments.

As the destination choice model must select from a large number of potential destination zones, the authors used stratified importance sampling. Using this technique, the destination alternatives are aggregated to several strata. A desired number of elemental alternatives are sampled from each stratum to form the choice set. A modified multinomial logit model is used to select a destination zone from the stratified choice set. Variables in the destination choice model include: piecewise distance from current location to destination zone, units of cargo available for pickup in the next destination, units of cargo available for delivery in the destination location and the number of locations available in the corresponding stratum. The authors found a non-linear relationship between the current location and prospective destination locations that reveals a smaller negative distance coefficient as the distance increases.

In a test urban network containing 84 nodes and two carriers, the results of this model were found to have reasonable agreement with game theory results based on the same simulated test data, except that the trip length distribution for their model favours longer trips than reported by game theory simulations. The authors proposed that this may be due to their stratification sampling.

**Entropy-based tour formation model**

Wang & Holguin-Veras (2009) proposed an entropy maximization model to assign truck tours. One advantage of this approach is that it only uses aggregate data, reducing data requirements compared with logistics and tour-based models. Besides the smaller data requirements, this type of model requires less computation time and also has fewer behavioural assumptions compared with supply chain and disaggregate tour-based models.

Their entropy maximization formulation contains three states: 1) the micro state refers to an individual tour starting and ending at the home base following a specific tour, \( m \); 2) the meso state is the number of tours following tour pattern, \( m \); and 3) the macro state is the number of total trips produced by and attracted to each zone. Entropy is defined as the number of ways of generating meso states in the network, given the relationship between meso states and macro states as the constraints. Micro states are assumed to be equally probable unless contrary information is available.
The authors estimated two different entropy maximization problems, which varied depending on assumed impedance within a tour. In the first formulation the impedance in a tour was taken as the sum of the travel times in every trip comprising the tour and the sum of all handling times at the stops. In the second formulation two separate impedance functions were written, one for the trip travel times and one for the stop handling times. This second formulation allows firms to respond differently to changes in the two parameters.

The authors compared their two entropy maximization formulations using data from the Travel Behavior Inventory (TBI) survey conducted by the Denver Regional Council of Governments (DRCOG) in the Denver metropolitan area. The authors identified 613 distinct tour routes. 65,385 tours were made on these tour routes per day. Calibrating the expected maximization problems on this dataset took approximately 2 seconds of computer time. The results found a good agreement with observed tours, including distributions of estimated tour times, tour travel times and tour handling times. Once these models have been estimated and calibrated using base-year information, the authors postulate how this entropy maximization approach could be used to estimate future-year tours given future-year trip productions and attractions.

2.1.7 Hybrid models

Using the definition of Outwater (2012), a hybrid model is defined in this thesis as a model that combines aspects of supply-chain and tour-based models. Two different models are discussed in this section, a hybrid model of the Los Angeles region by Fischer et al. (2005) and the FHWA Freight Forecasting Framework regional model that is currently under development (Outwater et al., 2012).

Los Angeles freight framework

Fischer et al. (2005) proposed a hybrid framework that combines elements of supply-chain and tour-based models to create a multimodal freight modelling framework for Los Angeles County. The focus of their research was on domestic freight movements, with an emphasis on truck freight, as freight movements by all modes were projected to increase by 80% between 1995 and 2020.

These authors used a hybrid modelling approach as none of the individual models that they had reviewed met all of their objectives. An overview of this model is presented in Figure 2.6. The logistics chain models the movement of selected commodities throughout the supply chain from suppliers through intermediate handling facilities to as near as possible to the final retail or customer location. Tour-based models were proposed for mixed goods shipments and also for services.
The logistics chain contains three components. First, the economic layer describes economic trade relationship between producers and consumers. This step progresses from identifying: 1) producers and consumers for a commodity, 2) commodity generation and attractions at each producer and consumer, and 3) commodity flow patterns between regions. Next, the logistics layer describes intermediate and final handling steps (e.g. warehouse or manufacturer) for each commodity. Third, the transportation layer describes the modes of transportation, vehicle loading patterns and routes between the steps in the selected supply chain. The logistics chain is most representative for commodities that are shipped using large lot sizes, have little mixing between commodities and also have simple supply chains. The logistics chain was used for some commodities, such as agricultural, mining and petroleum. Other commodities, such as metal and machinery manufacturing, computer and electronic equipment, retail trade and services are treated using a tour-based formation. The implementation of the tour-based model was not detailed in this paper.

Tour-based and supply chain regional freight model for Chicago
Outwater et al. (2012) presented a regional model for the Chicago metropolitan area as part of the Federal Highway Administration (FHWA) Broad Agency Announcement program. This paper presented a proof of concept study that outlines a model that is designed to be sensitive to critical global, regional and facility-based conditions. This model is divided into two parts. The national part of the framework uses a supply chain model for long distance trips while the regional part of the model uses tour-based methods. Figure 2.7 shows the overall model structure.

Figure 2.6: Overview of Integrated Model of Fischer et al. (2005).

1 Transportation Research Record: Journal of the Transportation Research Board, No. 1906, Figure 1, p. 108, Washington, D.C., 2005. Reproduced with permission of the Transportation Research Board.
The national scale part of this model focuses on how firms select suppliers and how suppliers ship goods to their receivers. This national model comprises the following six steps:

1. **Firm synthesis**: synthesizes all firms in the U.S. by industry category and size category. These firms are created at a zonal level.
2. **Supply chain**: connects goods suppliers with receivers
3. **Goods demand**: predicts annual commodity shipments between every shipper and receiver
4. **Business locations**: identifies a more precise location of firms within the previously established larger zones defined in step 1
5. **Distribution channels**: selects between one of four distribution channels (direct shipment, 1 type of intermediate stop, 2 types of intermediate stops, or 3+ types of intermediate stops). A “type of intermediate stop” can refer to an intermodal terminal, warehouse, consolidation centre or distribution centre.
6. **Shipment size and frequency**: selects the shipment size used between two firms
7. **Mode and intermediate transfers**: identifies the mode and routing path (including specific transshipment points if they are used) based on travel time, cost, shipment characteristics and distribution channel characteristics. This model was adapted from de Jong & Ben-Akiva (2007).

The tour-based regional model is used to determine individual vehicle movements required to distribute goods within a region. For long-distance shipments this model considers the shipment legs from the final transfer point to the final delivery point and also from the supplier to the first transfer point. For shipments involving an intermodal handling facility, warehouse or distribution centre, this facility is used as a base for vehicle tours. Given the regional nature of this tour-based model only the truck mode is included. Four main steps are involved:

1. **Vehicle and tour pattern choice model**: This is an MNL model that jointly predicts the vehicle size (small, medium or large) and whether the local shipment is made in a “direct” pattern (the vehicle makes one shipment and then returns to the base for the next shipment) or a “multi-stop” pattern, where a truck makes multiple deliveries and or pickups on a tour.
2. **Number of Tours and Stops**: This step is composed of two parts, the first part is an MNL model that predicts the complexity of undertaken tours (decides between one large tour or multiple smaller tours). The second part uses a hierarchical clustering algorithm to divide the shipments into spatially collocated groups for delivery in a single tour.
3. **Stop sequence and duration**: The authors used a greedy algorithm to sequence stops in a tour. According to Smith (2012) the authors found that a greedy algorithm better reproduced
observed tours from the Texas Commercial Vehicle Survey than results that optimized tour patterns such as Vehicle Routing Problem approaches. Stop duration was calculated using an MNL model.

4. Delivery Time of Day: This final model uses an MNL model to predict the first departure time of a truck leaving its base on the day and the start time for every trip. If a trip is made in an unreasonable time frame, the tour is split (another truck is added).

After these steps the travel times for each trip, stop arrival times and stop departure times can be calculated. Finally, the vehicle tours are converted into aggregate zone trips for traffic assignment.

2.1.8 Other freight models

The final two models examined in this literature review include a model proposed by Giuliano et al. (2010), who present a modelling framework for estimating intra-metropolitan freight flows on the highway network and a freight transhipment model proposed by Chow & Ritchie (2012).

The focus of the Giuliano et al. paper was to address data limitations that inhibit freight modelling efforts by focusing on reliable secondary data sources such as small-area employment data, Commodity Flow Survey (CFS), annual foreign and domestic trade for maritime ports, an import and export database and the IMPLAN database, which provides county-level input/output data within the United States. These data sources are widely available, regularly updated, more easily transferable across regions than other approaches that rely on data only obtainable through metropolitan surveys.

The interregional flows are calculated by estimating five aggregate flows from the CFS and IMPLAN databases, W2LA (imports from outside the U.S. to the LA region), US2LA (imports from the rest of the U.S. to the LA region), LA2W (exports from the LA region to international destinations, LA2US (exports from the LA region to other parts of the U.S.), and LA2LA (remaining commodity supplies, which originate and end within the LA region). Each of these flows are calculated for nine different commodity sectors and are calculated in units of dollars. They then generate the proportions for each sector/flow combination travelling by each mode (air, rail, truck and water). The interregional flows are distributed among a limited number of zones that correspond to entry/exit points to the region. These zones include the two major seaports, the five major airports involved in freight shipping, three major rail yards and 12 highway entry-exit points. For shipments whose final location lay within the LA region, each zone is allocated a proportion of the total shipments using the proportion the attraction and production of that zone divided by the total attraction and production in all zones in the study region.
Regional input-output tables from IMPLAN are combined with local area employment data to estimate intraregional freight trip ends for each traffic analysis zone. Afterwards, the interregional and the intraregional origins and destinations are combined to form O-D pairs. First, the dollar values are converted to tons using dollar/ton factors for each commodity as drawn from the CFS. The commodities
are then converted from annual to daily flows, and then using conversion factors to convert commodity flows to truck trips. The conversion factors are selected to ensure that the total number of observed truck trips is maintained. Trip distribution for intraregional trips is conducted using a gravity model. Separate decay functions are calculated for each of the nine commodity groups. Finally, trip assignments are calculated by adding the intraregional and interregional freight trips to passenger volumes and running a User-Optimal-Strict Network Assignment algorithm, which assumes perfect rationality among travellers, no temporal fluctuations and no modal or link interactions.

Chow & Ritchie (2012) present a freight transhipment assignment problem (FTAP) model, which uses a mathematical programming method that that accepts a commodity flow OD matrix as input and outputs the number of vehicles that use each link in the network. Commercial vehicles are formulated to traverse in cycles, meaning that this model considers tours and empty trips, which can be tracked. The proposed FTAP model is a multimodal model that allows mode changes at transfer points. This problem is linear if congestion effects are not included. Congestion effects can be considered at both links and at transfer nodes. In this case the problem becomes non-linear and is solved using a Frank-Wolfe algorithm. This FTAP model is sensitive to supply-side infrastructure changes, both in transportation networks and also at transhipment facilities. For example, it could be used to predict the changes in commodity flows that would result should the capacity of a road/rail transfer node be increased by 20%.

Chow & Ritchie recognize that determining cost and congestion parameters can be difficult, especially at transfer nodes. To resolve this issue, they used an inverse optimization problem that calibrates the congestion parameters ensuring that the observed commodity flows are the optimal solution to the mathematical programming model. This parameter calibration can be performed using standard nonlinear optimization methods.

### 2.2 Literature Review on GPS Data Collection Efforts

As discussed in Chapter 1, the focus of my Ph.D. research was to estimate activity and inter-arrival duration models of commercial travel using passively-collected GPS data as the primary data source for model estimation. These models are intended for use in the proposed FRELODE urban logistics decisions modelling framework. One of the specific objectives of this research was to develop the new data processing techniques that were required to convert the GPS data into a travel diary from which the arrival and inter-arrival duration models could be estimated. This section provides a review of selected publications that used GPS data to: 1) study travel variability, which is a subject of interest in this research and is discussed in Chapter 7, 2) assess under-reporting in travel surveys, 3) analyze travel
patterns of commercial vehicles, and 4) improved GPS data processing techniques for travel modelling applications.

2.2.1 GPS Studies of travel variability

A number of studies of travel behaviour using data recorded by GPS have been presented in the literature. Certain studies used GPS data to study day-to-day variability of travel behaviour as GPS monitoring is less burdensome upon survey respondents than completing travel surveys over extended time periods.

Examples of GPS studies for this purpose include a GPS-based travel survey conducted in Lexington, Kentucky that set out to explore day-to-day variability in travel behaviour. Variability was assessed for the number of trips, departure times, and travel distances and times. A GPS receiver and Personal Digital Assistant (PDA) were provided to the primary driver in 100 households. These participants were tracked for a target of six days, although some vehicles were tracked for longer (Muthyalagari, Parashar & Pendyala, 2001; Pendyala, 2003). Among the analyses used was an approach presented by Pas (1987), who divided the sum-of-squares term of the least squares regression R-square goodness-of-fit measure into different components that represented within-person and between-person travel variability. In the GPS study, the authors found that for a five-day sample, at least 60% of the travel variability of different measures could be attributed to within-person variability.

Schönfelder, Axhausen, Antille & Bierlaire (2002) raised concerns about using surveys for long-term travel behaviour studies and questioned if the same support for respondents as was provided in the successful MobiDrive survey could be maintained for larger sample sizes. Their concerns about future long-duration surveys included: 1) the limited pool of respondents willing to undertake long-term travel behaviour studies, 2) non-response, 3) reporting inaccuracy (especially times/durations and distances), 4) fatigue effects in longitudinal surveys, and 5) lack of information about route choice. They used GPS data that had been previously collected by the Rätt Fart (Right Speed) project in Borlänge, Sweden, which was originally intended to test driver response to in-vehicle warnings when exceeding the speed limit. Schönfelder et al. analyzed the Rätt Fart data to infer additional information from the GPS stream, including: inferring the driver of a vehicle within a household through driving characteristics, identifying unique origins and destinations, identifying trip-ends occurring when the vehicle is left running, and identifying the trip purpose using time of day, location and stop duration information.

Stopher, FitzGerald, Zhang & Bretin (2007) and Stopher, Clifford & Montes (2008) reported on a panel GPS survey that was conducted in three phases. Phases 1 and 2 collected data for 28 days from 50 and
46 households respectively. Meanwhile Phase 3 collected data from 36 households for 15 days. Each household member older than 14 years was asked to participate. The authors analyzed non-travel days in an attempt to quantify respondent fatigue as GPS cannot distinguish a genuine non-travel day vs. a day when the participant did not take the GPS device. After participants with an unreasonably high number of non-travel days were excluded, the authors found no evidence of response burden as the mean number of trips per day and the travel time per person per day remained consistent over the observation period. Cumulative measures consistently showed instability for the first 4-6 days before trending to stable values. Also, analysis of the first phase results showed that intra-personal variability accounted for between 40 and 60% of the total variability, even when only looking at weekdays. Implications are that results from one day data sources are unstable.

2.2.2 GPS studies examining underreporting in travel surveys

A number of GPS studies were designed to analyze underreporting in travel surveys. In these studies, a sample of survey respondents was provided with GPS recording units to monitor their travel and compare their observed travel patterns against their responses to self-reported travel surveys. For example, Wolf, Oliveira & Thompson (2003) analyzed differences between the number of trips, total trip distances and travel times between survey and GPS data for the San Diego, Alameda and Sacramento regions of California. In some areas the authors found large discrepancies in vehicle miles travelled reported between the survey and the GPS that would substantially affect models estimated using data collected from that survey.

Similarly, Bricka & Bhat (2006) examined demographics and travel characteristics that affected the likelihood of an individual underreporting trips in a household travel survey and the level of trip underreporting by using a joint binary choice-ordered response model structure. This model was estimated using a GPS-equipped sample of households from the Kansas City Household Travel Survey, conducted in the spring of 2004. The purpose of this model was to: 1) create adjustment factors to control for under-reporting, and 2) help develop better survey techniques.

Stopher, FitzGerald & Xu (2007) assessed the accuracy of the Sydney Household Travel Survey using a sample of participants who also accepted to participate in the GPS study. Each vehicle owned by the household was outfitted with a GPS device and every person 15 years of age or older that was more likely to walk or take public transportation was provided with a wearable GPS device for the same period. This GPS study found that 7.4% of trips in the Sydney Household Travel Survey were not recorded. While this is a noticeable level of under-reporting, it is less than similar U.S. studies that
compared GPS with CATI (computer assisted telephone interviewing) surveys. These authors also found that short trips, trips made after 5 pm, trips for social visits and picking-up/dropping-off a passenger, and trips made by people with high travel loads were more likely to be unreported. They also analyzed differences in start and end times and found a high level of over-reporting of trip distances and durations compared with the GPS results.

### 2.2.3 GPS studies of commercial vehicles

GPS has also been used to monitor commercial vehicle movements. For example Battelle (1999) installed GPS receivers on 140 heavy-duty trucks drawn from a volunteer sample to monitor truck travel patterns in the state of California. Analyses were conducted for different classes of heavy trucks and included speed profiles, trip distances and durations, road use (by road classification type), trip generation (measured by truck starts per day), truck starts by time of day and day of week, idle time, and stop durations.

Logendran & Peterson (2006) reported on a GPS study used to gain a better understanding of the trucking industry and its impact on highways managed by Oregon DOT. This project focused on leveraging technologies to supplement existing data collection methods in order to aid in the development of commodity flow commercial transportation models. These researchers found considerable resistance by trucking firms; only two companies were willing to let researchers monitor their vehicles. Research questions of this study included: 1) the daily number of stops made in a vehicle, 2) expected running time between stops, 3) road segments used by the commercial vehicles, 4) proportion of trucks that followed designated routes, and 5) whether similar daily routes were followed. This analysis was used to forecast future travel and to justify improvements to the highway infrastructure.

McCormack & Hallenbeck (2006) compared the ability of two different electronic data sources to design and test analysis methods examining speed and volume improvements of commercial travel following roadway improvement projects. One technology used electronic transponders while the second technology involved placing GPS units in volunteer trucks. As a comparison of before and after traffic volumes can also be influenced by economic factors, the authors’ analyses included effects on parallel routes and on trip travel times. They found that it is possible to obtain the reliability of truck trips from the electronic data. Advantages of the GPS devices are that they show the actual route taken, however the GPS also required considerable staff effort and recruitment efforts to attract adequate trucking company participation.
The Peel Region Commercial Vehicle Travel Survey collected information from a sample of 597 shipping firms about the establishment, inbound and outbound shipments and driver tours using a mail-out/mail-back survey between September 2006 and May 2007. A sub-sample of firms also participated in a GPS supplement. An analysis between the 37 drivers that returned usable GPS data and completed the driver surveys revealed discrepancies in both the survey and the GPS data. The GPS data missed many short-duration shops (less than five minutes) due to the trip-end identification algorithm used by the GPS recording devices. Both the GPS and paper-pencil surveys suffered from survey truncation. Most drivers, however, had a good match between the GPS and survey results. Stop durations were also found to be accurately recorded by drivers (Kwan, 2007; Roorda et al., 2013).

Greaves & Figliozzi (2008) reported on a 2006 survey conducted in Melbourne, Australia, that used passively-collected GPS data to provide information on urban commercial vehicle tours. GPS data were collected for a one week period from 30 trucks, hence 210 truck-days of data were recorded. This paper presented the rule-based algorithms used to process the raw GPS into meaningful trips and presented results of the pilot study. Tour results indicated extensive trip chaining and a noticeable increase in the average speed outside of peak hours.

2.2.4 GPS studies focussing on data processing algorithms and techniques

A variety of papers have been focused primarily on developing automated processing of the large volumes of data produced by GPS recording devices in order to convert this data into a form suitable for transportation modelling. Stopher et al. (2003) presented an automated processing algorithm to identify trip ends in the GPS data. After trip ends were identified, the GPS stream could then be separated into activities and trips connecting these activities. In this study, manual checking was used to validate the automated procedures. The value of the automated procedures became immediately apparent as one or two days of manual analysis was previously required to process a vehicle-day’s worth of GPS data. Approximately one hour was required to validate an equivalent computer processed GPS data stream.

Srinivasan et al. (2006) described a prototype software package developed to automatically process raw GPS data and output travel diaries suitable for estimating activity-based travel demand models. According to these authors, “it is also evident that the use of passive GPS devices for data collections shifts considerable burden from the respondent to the analyst.” Discussed processing steps included trip-end identification. After trips were separated, the software then computed trip attributes such as
start and end times, trip end locations, trip distance, speed, and also the activity at destination. Activity types were reported in either one of two modes. A “basic-analysis mode” grouped activities into home, work or other based on proximity to respondent identified locations of home and work. In the “enhanced-analysis mode” a multinomial logit model was used to divide other trip activities into shopping, leisure and serve-passenger activities using covariates such as land-use attributes, time of day, stop duration, demographics and the preceding activity.

Du & Aultman-Hall (2007) also examined the trip-end identification problem to identify methods outside of the dwell time that could be used to identify brief stops in trip chains, such as picking up and dropping off passengers. It is difficult to distinguish these stops from stops caused by traffic controls and congestion. The authors developed a rules set that identifies trip ends, both in periods of signal blackout and during GPS reception. This algorithm is described in more detail in Section 5.5 of this dissertation. The authors reported an average accuracy rate (percent of trip ends that are correctly identified) of 94% and an error rate (falsely-identified trip ends) of 9%.

Chung & Shalaby (2005) presented a trip reconstruction software tool to process GPS-based personal travel surveys. This tool used a map-matching algorithm to identify road links and a rules-based algorithm to identify modes (walk, bicycle, bus and car) using the GPS speed and GIS information about the road network and transit network, including accurate locations of transit stops. This research was extended in Tsui & Shalaby (2006), who created two versions of this software package to analyze GPS survey data: a first version was intended for regions where high-quality GIS data are unavailable while the second version was for regions with high-quality GIS data. This research used fuzzy-logic for mode identification, using speed, acceleration and data quality attributes. Researchers also developed algorithms to: 1) complete missing sections of transit trips using the GIS route information, 2) detect underground walking and transit trips, and 3) improve processing of GPS blackout periods that occurred either after an activity or after exiting a subway. This algorithm was tested by 9 volunteers who made a total of 109 trips. These volunteers verified their trips using a prompted-recall survey. The trip-end detection did not miss any trip ends, although some trip-ends were falsely identified due to stops caused by traffic conditions. The mode identification performed well as the mode was correctly identified on 94% of trips when GIS data was available.

Another GPS–based travel survey was conducted in the Netherlands, which collected seven-days of travel data from a large sample of over 1000 participants (Bohte & Maat, 2009). In order to reduce respondent burden, the GPS data were processed automatically and respondents validated the trip diary using a web-based application. A rules-based processing algorithm was developed that included
separating the data stream into trips, determining trip purpose and mode detection. The trip purpose was assigned to home, work or other. Users validated the data processing after the survey was completed and could merge trips, split a trip, adjust travel times and modes, move a destination, add new trips and identify ‘other’ trip purposes during validation. A post-survey evaluation found that 85% of respondents did not find the survey burdensome to complete, showcasing the ability of GPS surveys to reduce respondent burden.

Schuessler & Axhausen (2009) conducted research into post-processing passively-collected GPS data as the sole data source for GPS surveys. Processing steps were as follows. Data cleaning removed erroneous data points and the authors then applied Gauss kernel smoothing to smooth the GPS trail. Trip ends were then detected. Their trip-end detection algorithm is described in more detail in Section 5.4. For mode choice the authors used a fuzzy logic approach that expanded upon the methods presented by Chung & Shalaby and by Tsui & Shalaby. The authors obtained an average of 6.65 travel days collected from 4882 participants from a previously conducted private GPS study without any socio-economic information about the participants. The authors compared the results against the 2005 Swiss Microcensus on Travel Behaviour. The GPS studies revealed participants took more shorter trips per day than the survey data, generally matching an expectation that survey participants miss short trips when responding to surveys. Distance and duration profiles of trips of longer than 20km matched the survey data, validating their trip and activity detection algorithms. The mode detection appeared to produce realistic results.

2.3 Summary

The first part of this chapter presented a literature review of freight travel demand models. This review started with an examination of traditional freight transportation modelling techniques, including Growth Factor Methods and methods based on “four-step” methodologies that were originally developed for passenger travel applications. Limitations of these state of practice models were then discussed.

After the overview of the state of practice models, the attention of this literature review turned to examining recently implemented or proposed modelling frameworks that address some of the issues of the traditional models. It is unlikely that any single model will be able to fully address all of the raised concerns in different regions and on different geographical scales. The following list briefly categorizes the presented models and briefly states their objectives.

1. Supply chain (Logistics) models are national or international models that focus on accurate theoretical representations of how freight moves in a logistics chain. Decisions include
shipment frequencies, transhipment point location, shipment consolidation, mode choice and vehicle routing. It is usually assumed in these models that firms optimize their entire supply chain (inventory plus transportation). Due to their larger geographic scales these models often use multi-regional Input-Output tables to represent the underlying economic conditions.

2. *Urban logistics* models expand commodity-based freight transportation models in order to improve the representation of the supply chain in urban areas. Enhancements of these models include: considering the roles of shippers, carriers and receivers, and also includes supply chain elements such as different distributions channels, locations of distribution and consolidation centres and tour formation.

3. *Tour-based models* focus on generating tours of urban commercial vehicle deliveries. This category varies from the Calgary model, where tours are generated and then assigned stops using a sequential discrete choice tour-generation model, to the methods of Wang & Holguin-Veras (2008, 2009) that create tours that satisfy aggregate commodity flow data or use an entropy maximization model for tour generation.

4. Finally, *hybrid models* combine aspects of the other categories. Often supply chain models are used for long-distance shipments while tour-based models are used for urban deliveries and services.

One study that was not mentioned in this literature review is a regional agent-based microsimulation conceptual framework for freight that is under development by University of Toronto researchers (Roorda *et al.*, 2010). This study is reviewed in more detail in the next chapter.

Following the literature review on commercial transportation models, Section 2.2 of this dissertation outlines a literature review of studies that used GPS data as their primary data source. The purpose of this literature review is to analyze the types of problems that researchers were studying and especially to study the GPS data processing steps used by different authors to convert their raw GPS data into a travel diary of trips and activities that could be used for their respective applications. Chapter 4 of this dissertation describes the GPS data while Chapter 5 presents the data processing steps that were conducted to convert the raw GPS data into travel diaries that were used to estimate the activity and inter-arrival duration models in my Ph.D. research. When applicable, detailed GPS processing is further reviewed in that chapter.
Chapter 3: The FRELODE (FREEight LOgistics DEcisions) Framework

The focus of this chapter is to present a high-level overview of the FRELODE modelling framework for urban logistics decisions. Before this, however, Section 3.1 reviews the agent-based conceptual framework of Roorda et al. (2010), of which FRELODE will form one component when it is implemented. The rest of this chapter then presents a high-level overview of the FRELODE modelling framework.

3.1 Regional Agent-Based Microsimulation Conceptual Framework for Freight

Roorda et al. (2010) presented an agent-based microsimulation framework that explicitly represents the diversity of business establishments and the impact of both long and short term interactions between different business establishments on commercial travel behaviour. This conceptual framework has been developed from the outset so that it can be integrated with the ILUTE (Integrated Land Use Transportation Environment) agent-based microsimulation model (Miller & Salvini, 2001; Salvini & Miller, 2005) of passenger activities and travel in order to create a fully integrated passenger and freight activity-based urban transportation model.

Fundamental to the Roorda et al. framework is the business establishment, which is defined as an organization at a specific location that produces, processes or stores commodities, and/or provides business or logistics services. In general, business establishments are assumed to operate independently. This is a simplification as a firm can contain multiple business establishments that coordinate activities. This simplification is removed for firms that specialize in logistics services as these firms often have multiple business establishments that form a logistics network required to efficiently deliver their contracted shipments.

Another key feature of this framework is the separation of different types of decisions made by business establishments. Figure 3.1 presents a high-level overview of the breakdown of different types of decisions within this framework. The different components are briefly described below.
Figure 3.1: High-Level Overview of the Proposed Regional Agent-Based Microsimulation Conceptual Framework for Freight (Roorda et al., 2010)

The Fundamental Business Decisions component considers long range decisions such as to start a business, change location, decide what commodities and/or services to produce and to terminate the business. This component creates a commercial “synthetic population” of firms operating in the model environment.

The Supply Chain Management Decisions component describes how a firm manages larger aspects of their operational strategy. Example decisions include increasing or decreasing production capabilities, changing resources such as the vehicle fleet or warehouses, establishing long-term relationships with suppliers and receivers and deciding whether or not to use the services of a third party logistics (3PL) firm to manage their supply chain.
Goods and services supply and demand between firms are modelled in the *Market Interaction Decisions* component. In this component, firms sign *commodity contracts* and *business service contracts*. A commodity contract is a contract for the shipment of goods from one business establishment to another. These contracts include: origin and destination establishments, type of commodity, annual quantity of shipped goods (in tons), the price paid by the receiver and the business establishment responsible for arranging the shipment. A business service contract is similar to a commodity contract except that services are provided and not goods. The output of these two components is a list of shipments and service trips between all firms of the synthetic population throughout the study duration.

In the *Logistics Contract Formation Decisions* component, the firm previously designated as being responsible for organizing the shipment selects the firm that will physically deliver the goods, called the *carrier*. Vendors select carriers through an auction process where a vendor advertises a delivery set upon which each carrier can choose to bid, or not. The price that a carrier bids as a delivery price depends on the degree to which this shipment can be bundled into their existing operations. An operational set of models has been estimated for this component (Cavalcante & Roorda, 2013).

The *Logistics Decisions* component is the final step of the entire Roorda *et al.* modelling framework, and focuses on how carriers execute delivery of their contracted shipments. This component accepts a list of shipments and services by carrier and outputs a set of trips and tours used to undertake these shipments and services. Separate logistics decisions models will be specified for long-distance trips and for within-region trips. Long-distance trips include deliveries into and out of the study region, and also includes through trips. Within-region trips include goods and service trips where the shipper and receiver are located within the same urban area. If a long-distance trip uses an intermodal handling facility or consolidation/deconsolidation centres, the local pick-up and/or delivery tour components are also included in the urban logistics decisions component.

The output from the logistics decisions components is a list of vehicle trips and tours by carrier. These vehicle movements are combined with those from other models, such as passenger vehicle trips and tours output from the ILUTE model, into a multiclass user-equilibrium traffic assignment that predicts traffic volumes and travel times throughout the study region.
3.2 FRELODE – Estimating Urban Logistics Decisions Using Passively-Collected GPS Data

FRELODE is proposed to model how carriers execute urban shipments within the logistics decisions component of the larger Roorda et al. framework. FRELODE has been designed to be estimated predominantly using passively-collected GPS data recorded from GPS-equipped truck-mounted Electronic On-Board Recorders (EOBRs) and publicly available secondary data sources.

The primary advantage of designing FRELODE such that it can be estimated predominantly using GPS is that little data documenting urban commercial travel are currently available. As was discussed in Section 1.3.2 of this dissertation, only a handful of surveys of urban commercial travel have been conducted. Hence using GPS data, which are already widely available throughout North America, provides a means of estimating commercial logistics decisions when other data are unavailable. Due to the low number of existing commercial vehicle surveys, this is currently the case in most urban areas. Other advantages of GPS data include that extended observation periods are possible as little burden is imposed upon carriers, that they provide accurate spatial and temporal histories of observed vehicles and that they are well suited to observing trip-chaining behaviour.

Passively-collected GPS data do have serious limitations, however, which affect the design and capabilities of FRELODE. These restrictions include that: passive GPS data do not contain shipment information, information about the carrier and visited customers are often suppressed due to privacy concerns and that GPS data from truck-mounted electronic logging devices are only available for trucks and not for other modes. As no shipment information is available, inventory and handling costs cannot be inferred. FRELODE has been designed to not require this information. If additional survey data become available then these could either be substituted or included with the GPS data to improve the explanatory power of the component models. Models estimated using FRELODE, however, may be easier to use in an operational model as the attributes obtained from survey data would require forecasting.
The input to FRELODE comes from the Logistics Contract Formation component of the Roorda et al. framework. In the Logistics Contract Formation component, carriers are selected by vendors to deliver shipments between different shippers and receivers. The following attributes are anticipated as inputs for each contracted shipment:

- Annual shipment weight to be delivered between the shipper and receiver
- Shipper and receiver locations
- Commodity of goods to be shipped
- Classification of delivery patterns into one of three segments: 1) frequent shipments, 2) regularly-scheduled shipments, and 3) stochastically scheduled shipments. These segments are described in more detail in Section 7.3 of this dissertation.

In FRELODE carriers are assumed to plan their deliveries considering all contracted shipments simultaneously, allowing them to develop trip chains where vehicles visit multiple destinations in a single tour. Figure 3.2 shows the structure of the FRELODE framework, which draws from Roorda et al. (2010) and Outwater et al. (2012). A brief description of each component is presented below. The Activity Duration model and the Inter-Arrival Duration model are the focus of my Ph.D. research and are detailed in Chapters 6 and 7 of this dissertation. While FRELODE has been designed so that it can be estimated primarily using passively-collected GPS data, it is anticipated that the model framework will remain viable if components in the Within-Day Decisions section (see Figure 3.2) are instead estimated using survey data.
3.2.1 Shipment scheduling

FRELODE recognizes that commercial trips are often not made to a destination on a daily basis but may either be scheduled or made upon request and not on a pre-determined timeframe. Shipment scheduling is conducted in a different manner than has been used in other commercial transportation
models. This difference is largely because FRELODE was designed so that it can be estimated primarily using passively-collected GPS data.

Many proposed freight travel demand models, such as de Jong & Ben-Akiva (2007) and Liedtke (2009), assume consistent vehicle schedules with shipment sizes optimized to minimize the combination of transportation and inventory costs. Since the data required to estimate such a model are not available from GPS data a different approach was proposed. Instead of a shipment frequency model, the shipment scheduling component of this framework uses an Inter-Arrival Duration model that can be estimated using the observed number of days between successive visits to the same destination from the GPS data source. An Inter-Arrival Duration model has been estimated and is the focus of Chapter 7 of this thesis.

To calculate the shipment size it is assumed that the receiver processes the delivered goods at a constant rate. Given this assumption, the shipment size can be calculated using the equation

\[ q = Q \times \frac{d}{D} \]

where \( q \) is the size of the current shipment, \( Q \) is the total quantity of goods shipped between the two firms in the study period, \( d \) is the number of days since the last shipment and \( D \) is the total number of days in the study period. Note that it is possible for multiple shipments to be delivered on the same day. In this case the shipment sizes are divided by the number of shipments on that day to form an average shipment size.

3.2.2 Vehicle type selection

After shipments have been assigned to individual days and shipment sizes determined, the remaining models within FRELODE create commercial trips and tours made by carriers within each day. The first step selects a class of vehicle used to deliver shipments to a destination. This component has not yet been estimated as vehicle information was unavailable in the current GPS data source. It is anticipated that a multinomial logit model could be used to select between three categories of truck size (light, medium and heavy). This model could be estimated from other GPS data sources, if vehicle attributes are provided, but would likely be better estimated using survey data containing shipment information. A recently completed commercial establishment transportation survey in the Greater Toronto and Hamilton Area (Roorda, Rashidi, Bachmann & Rudra, 2013) could be used to estimate this model.

3.2.3 Tour formation

The tour formation step groups trip ends visited on the same day by the same carrier using the same vehicle type into vehicle tours. One type of model that could be used for this research is a vehicle
routing problem (VRP). The objective of a VRP is to find the optimal set of tours that serve all of the customers visited on a single day while satisfying a set of constraints. The optimal solution is found by minimizing a cost function (which can include cost, distance and time terms, as desired) given the set of constraints.

The constraints of a tour formation VRP can include the maximum tour duration (in hours) and the maximum vehicle payload. As tour durations are directly observable using the GPS data, this constraint could likely be inferred from observed tour durations. A suitable payload constraint is expected to be less than the actual vehicle payload as using the actual vehicle payload would lead to overly optimized tours compared with observed intra-urban commercial vehicle travel (Smith, 2012). Vehicle payload and shipment sizes cannot be directly inferred from the GPS data. This constraint could be estimated using survey data, if it is available. Otherwise a suitable constraint could likely be inferred from the GPS data by selecting a value that produces representative trip lengths, tour shapes and the number of stops on tours compared with GPS observations. Other tour formation approaches, such as by Outwater et al. (2012), have been proposed in order to produce a more accurate representation of urban commercial tours.

3.2.4 Tour start time

Once the trip ends have been assigned to tours, the Tour Start Time component selects a time of day when the vehicle leaves the depot to commence the tour. This occurs after assigning trip ends to tours because the number of trip ends visited on a tour and the tour total distance travelled likely impact the tour start time. This model has not currently been estimated. As the tour departure time can be observed from the GPS data, GPS data could be used to estimate such a model. Two example models that could be estimated using GPS data are shown below. If survey data are available, then additional attributes describing the shipment, carrier and shippers/receivers could be included.

Two possible approaches for a tour start time model include:

1. Monte-Carlo simulation based on observed tour start time distributions. If required, different distributions can be selected based on observable attributes, such as the number of trip ends and total distance travelled in the tour.
2. Depending on the distribution of the observed start times, a continuous hazard model may also be an appropriate modelling technique. Explanatory variables could then include the travel distance in the tour, the number of trip ends and also attributes of the visited destinations.
3.2.5 Vehicle trip

The travel time for each leg within a tour can be input from a traffic assignment algorithm that calculates travel times given travel demands and the transportation network. Given the trip start time and the trip duration, the arrival time at the next destination on the tour can be calculated.

3.2.6 Activity duration

The activity duration model reflects the amount of time that a vehicle remains at a trip-end, including vehicle loading and unloading, driver breaks or time spent at a service call. The activity duration model determines when a vehicle will leave for the next trip end in the tour. Estimation of an activity duration model was one of the key objectives of my Ph.D. research. A detailed description of this model is presented in Chapter 6 of this dissertation.

3.2.7 Implementation within larger agent-based framework

It must be remembered that the entire FRELODE framework is one component within the larger Roorda et al. agent-based commercial vehicle simulation, which in turn is expected to be operating in parallel with additional agent-based simulation models, such as the ILUTE activity-based passenger travel model.

The output trips and tours from FRELODE are expected to be included within a larger multi-class trip assignment algorithm that simultaneously predicts travel flows for both passenger and commercial vehicle travel on all links in the network.

An iterative solution to the entire multiclass modelling framework will be required so that the computed travel times are compatible not only with the travel times, distances and costs assumed by other components of FRELODE, but also within the larger Roorda et al. framework and parallel passenger models.

3.3 Advantages of the FRELODE Framework Compared with Existing Commercial Travel Demand Models

Section 2.1.3 lists multiple limitations of existing “four-step” travel demand models. These limitations are briefly restated below:

1. Four-step models cannot accommodate that different actors are responsible for different aspects of shipment planning.
2. Four-step models use a single model structure for movement of all commodities in spite of a wide variety between the logistics associated with different commodities.

3. Shipments are not independent from one another as firms optimize their supply chains through supplier selection, shipment frequency, use of consolidation and distribution centres and consolidation of multiple shipments on a single vehicle.

4. Services are not included in commodity-based four-step models.

The Roorda et al. framework is designed from the outset to deal with item 1, the role of different actors in the freight system. This is maintained in the FRELODE framework, which focuses primarily on creating a representation of the operations of carriers given their vehicle fleet, warehouses and their set of contracted shipments.

The FRELODE framework cannot currently address the second item of this list (the different logistics needs of a wide variety of commodities) due to the lack of shipment information from the passive GPS dataset. If shipment or commodity data is made available in addition to the GPS data then separate models can be estimated by industry and by commodity type.

Item three of this list (consideration of supply chains) is partially handled in FRELODE. The inter-arrival duration model can provide a stochastic representation of shipment frequencies and shipment sizes. Shipment frequencies, and hence delivery sizes, are a potential adaptation of supply chains to infrastructure and policy changes and traffic conditions. Likewise, FRELODE is a tour-based system that bundles multiple shipments onto a single vehicle to increase operational efficiencies.

Services can be included in FRELODE without modification as the GPS data do not differentiate between vehicle movements for goods delivery and for services.

Other advantages of the FRELODE framework are that it can conceivably consider the following on commercial travel logistics decisions:

- **Shipment scheduling**: Effects of land-use and travel times on shipment scheduling can be included in the inter-arrival duration model.

- **Road pricing**: Effects of road pricing can conceivably be considered throughout FRELODE, including the: inter-arrival duration, vehicle-type choice, assigning stops to tour and tour start time models. Note that these effects cannot be observed from passively-collected GPS data but could possibly be included if survey data are available.
• **Tour formation:** Changing travel times between destinations influences tour formation. Land-use effects on activity durations can also influence the number of stops that can be visited in a tour. This latter influence is admittedly only considered in an indirect manner given the structure of the FRELODE framework.

• **Road regulations:** The influence of various road policies such as modifications to permitted driver hours of service can be considered by modifying the constraints in the tour formation model.

It should be noted that these policy responses are only preliminary as the component models have not been described at this point in this dissertation. Section 8.1 provides a more detailed examination of policy response given the estimated activity and inter-arrival duration models.

### 3.4 Summary

This chapter presents a high-level overview of the proposed FRELODE urban logistics decisions modelling framework, which is designed to fit within the larger proposed regional agent-based microsimulation conceptual framework of Roorda *et al.* (2010). FRELODE is designed to take a list of shipment and service delivery contracts by carrier from previous models within the Roorda *et al.* microsimulation framework. Model components include shipment scheduling (modelled using an inter-arrival duration model), vehicle type, tour formation, tour start time and activity duration models.

FRELODE focuses primarily on urban commercial vehicle travel and has been designed so that it can be estimated primarily using passively-collected GPS data collected from GPS-equipped Engine On-Board Recorders. These recorders are already used by many carriers to monitor and optimize their operations. This is considered to be of considerable importance as few travel surveys of commercial establishments have been conducted. Passively-collected GPS data are considered to be suitable for urban travel modelling as the lack of available shipment information and lack of data from other modes are likely to be less of a hindrance for urban operations than for long-distance operations. Models of activity duration and inter-arrival duration have been estimated using the passively-collected data and are presented in Chapters 6 and 7 of this dissertation, respectively. The next chapter describes the GPS data source used to estimate these models.
Chapter 4: Data Sources Used for Model Estimation

The purpose of this chapter is to describe the data sources that were used to estimate the activity duration and inter-arrival duration models presented in Chapters 6 and 7 of this dissertation. Section 4.1 describes the GPS data while Section 4.2 describes the secondary sources that were used to provide land use context for the GPS observations.

4.1 Description of the GPS Data

XRS, previously called Turnpike Global Technologies (TGT), offers fleet tracking services. As part of this service they provide RouteTracker™ GPS-equipped engine onboard recorder (EOBR) units to their clients. The RouteTracker™ EOBR records vehicle travel and diagnostic histories and transmits this data to XRS who manages the data stream for their clients. Client benefits of purchasing XRS’s services include monitoring of driver hours of service to ensure compliance with regulations, monitoring of driver behaviour (e.g. idling time and speeds), automatic completion of required International Fuel Tax Agreement (IFTA) forms and fleet tracking.

In 2007, Turnpike Global Technologies provided the GPS travel history for vehicles that operated in the Greater Toronto and Hamilton Area of southern Ontario between April 1, 2007 and June 30, 2007. Figure 4.1 shows the study boundary of the GPS dataset. All GPS points located within the study area for participating vehicles were provided to University of Toronto researchers, as well as all GPS points of inbound and outbound trips into or out of the study area. Information was maintained after the last recorded stop before entering the study area for inbound trips or until the first recorded stop for outbound trips. No information was made available for any trip in which no GPS point was recorded within the study area. Figure 4.2 shows a schematic of GPS points that are included in the dataset.

Table 4.1 shows a summary of the unprocessed GPS data. Examining this table reveals benefits of using passively collected GPS data. Vehicle travel histories of 1618 vehicles, operated by 77 different carriers, were collected over a 91 day period. Hence there were a total of over 147,000 observed vehicle-days, including days where a vehicle did not operate or remained outside of the study region. The dataset included over seven million GPS records. A dataset of this size is virtually impossible to process manually; automated computer algorithms were required. The computer algorithms used for data processing are the focus of Chapter 5 of this dissertation.
According to Kwan (2007), the RouteTracker™ GPS device tracks the location and time of a vehicle every four seconds but only stores this information after the vehicle has travelled a distance of 500 m from the previously recorded location. The distance between recorded GPS points was increased to either one or two miles when a vehicle reached speeds that are only feasible on freeways. This data storage approach is beneficial for XRS and their clients as it reduces data transmission and storage costs. It had, however, large implications on the required data processing.
Table 4.2 shows the relevant information recorded by the RouteTracker™ EOBR for every GPS point. The identity of carriers was not provided in this dataset. Instead, a unique EncryptedCompany_ID was provided to distinguish between vehicles operated by different carriers. This field was crucial for the research as it made it possible to use the carrier as the focus of the model structure instead of having to use individual vehicles as the base unit of reference. This field, however, was only an ID code and it did not reveal the identity of the carriers (which were suppressed by Turnpike Global Technologies to protect confidentiality of their customers). Hence, no attributes of the carriers were available. Vehicles were identified using the EncryptedVehicle_ID field, which was used to follow the travel histories of individual vehicles. No information about the vehicles, for example weight, size or vehicle identification number (VIN), were provided.
The *Latitude* and *Longitude* fields showed the vehicle location. These fields were defined using the WGS84 datum of the Earth’s surface. The *Distance* field recorded the distance travelled by the vehicle since the previously recorded GPS point. This was the actual travelled distance (recorded by the GPS unit) and not a straight-line distance between the two points. This field also doubled as a marker to record vehicle stops and as an error flag, as is described below. The *Odometer* and *ECM speed* fields could not be used as they were not populated for all vehicles.

### Table 4.2: Available Data Fields in Turnpike GPS Dataset

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EncryptedCompany_ID</td>
<td>Unique identifier of every carrier</td>
</tr>
<tr>
<td>EncryptedVehicle_ID</td>
<td>Unique identifier of every observed vehicle</td>
</tr>
<tr>
<td>Point_ID</td>
<td>Unique identifier for every GPS point</td>
</tr>
<tr>
<td>Time stamp</td>
<td>Date and time of the GPS point</td>
</tr>
<tr>
<td>Latitude</td>
<td>Latitude coordinate of vehicle location using the WGS84 datum</td>
</tr>
<tr>
<td>Longitude</td>
<td>Longitude coordinate of vehicle location using the WGS84 datum</td>
</tr>
<tr>
<td>Distance</td>
<td>Road distance travelled in miles since previously recorded GPS point. Also, this field was set to a value of -1 to denote stops recorded by the RouteTracker unit and was set to a value of -2 if errors were encountered.</td>
</tr>
<tr>
<td>Odometer</td>
<td>Total distance travelled by the vehicle in miles (this field was not populated for all vehicles)</td>
</tr>
<tr>
<td>GPS speed</td>
<td>Instantaneous vehicle speed recorded by the GPS unit in units of miles per hour</td>
</tr>
<tr>
<td>ECM speed</td>
<td>Maximum recorded speed since previously recorded GPS point in miles per hour (this field was not populated for all vehicles)</td>
</tr>
</tbody>
</table>

Table 4.3 shows a small sample of the unprocessed GPS data provided by XRS. The *Latitude* and *Longitude* fields of this table have been suppressed to preserve confidentiality of XRS’s customers. Two examples of stops are highlighted in Table 4.3. In this table, the *Distance* field was sometimes assigned a value of “-1”. This was used to denote a stop. Stops were automatically recorded by the RouteTracker unit if the vehicle remained stationary for at least a five-minute interval (Kwan, 2007). Three entries were listed at every stop, which are described as follows:

**Line 1:** Represented the last motion GPS point before the stop.

**Line 2:** The *Latitude* and *Longitude* coordinates identified the actual stop location. The *Timestamp* was set to equal the previous timestamp in order to reflect uncertainty about the actual start time.
of a stop. The Distance field was set as the road distance (in miles) travelled between the previous motion point and the stop location.

**Line 3:** The Latitude, Longitude and Timestamp fields were all set to that of the previous entry. The Distance field was set to “-1” in order to record the stop.

As seen here, the recorded arrival time was the time of the last motion point prior to arriving at the stop. Hence this recorded stop arrival time preceded the actual arrival time. Also, the first record after a stop was located at the following motion GPS record, located at least 500 m from the stop location. The recorded departure time was therefore later than the actual departure time. Hence the recorded stop durations were longer than the actual stop durations.

### 4.2 Secondary Data Sources

The GPS data from the Turnpike RouteTracker units do not contain any behavioural information, such as information about the shipment, the carrier or any land-use information about the visited location. The following additional data sources were interacted with the GPS data to provide land-use context that was used to help process the GPS data and also to provide explanatory variables for the activity duration and inter-arrival duration models. The secondary data sources used in the final estimated models included:

1. CanMap® RouteLogistics street map
2. Transportation Tomorrow Survey
3. Teranet property parcel map
4. InfoCanada database of companies

The use of these data sources for this research is summarized in the following sections.
4.2.1 CanMap® RouteLogistics street map

CanMap® Route Logistics by DMTI Spatial provides a detailed street map throughout Canada. These are provided in shapefiles that can be read using the ArcGIS software package. Figure 4.3 shows a sample of this streetmap within the study area. Besides the physical orientation of a street, the CanMap® Route Logistics provides additional information such as:

1. Street name and starting and ending street numbers
2. City name and first three digits of the postal code
3. Link direction (two or one-way street, including direction)
4. Length
5. Speed limit and uncongested travel time
6. Road hierarchy (e.g. freeway, major road, local road)

The RouteLogistics information was used extensively to process the GPS data as it provided information relating the GPS data to the road network. This data processing is discussed in more detail in Chapter 5 of this thesis.

4.2.2 Transportation Tomorrow Survey

The Transportation Tomorrow Survey (TTS) is a comprehensive telephone travel survey conducted every five years in the Greater Toronto and Hamilton Area that aims to record interviews about the travel history on the survey day of a 5% random sample of all households in the GTHA. Approximately 150,000 surveys were completed in the 2006 survey (Data Management Group, 2013).

In this survey, respondents answered questions about where they live and where they worked or attended school. Hence the TTS provides information about the population and employment in travel analysis zones. This data was used to infer zone-level population and employment densities, which were tested as explanatory variables for the estimated activity and inter-arrival duration models.
### Table 4.3: Sample of Raw GPS Data Provided by XRS

<table>
<thead>
<tr>
<th>Encrypted Vehicle_ID</th>
<th>Point_ID</th>
<th>TimeStamp</th>
<th>Latitude$^1$</th>
<th>Longitude</th>
<th>Distance$^2$</th>
<th>Odometer</th>
<th>GPSSpeed</th>
<th>ECMSpeed</th>
</tr>
</thead>
<tbody>
<tr>
<td>196007</td>
<td>247270808</td>
<td>2007-04-02 00:37:34</td>
<td>---</td>
<td>---</td>
<td>2.019</td>
<td>54386</td>
<td>63</td>
<td>66</td>
</tr>
<tr>
<td>196007</td>
<td>247270809</td>
<td>2007-04-02 00:39:30</td>
<td>---</td>
<td>---</td>
<td>2.037</td>
<td>54388</td>
<td>63</td>
<td>65</td>
</tr>
<tr>
<td>196007</td>
<td>247270810</td>
<td>2007-04-02 00:41:23</td>
<td>---</td>
<td>---</td>
<td>2.044</td>
<td>54390</td>
<td>65</td>
<td>66</td>
</tr>
<tr>
<td>196007</td>
<td>247270811</td>
<td>2007-04-02 00:43:12</td>
<td>---</td>
<td>---</td>
<td>2.019</td>
<td>54392</td>
<td>66</td>
<td>67</td>
</tr>
<tr>
<td>196007</td>
<td>247270812</td>
<td>2007-04-02 00:45:03</td>
<td>---</td>
<td>---</td>
<td>2.027</td>
<td>54394</td>
<td>65</td>
<td>66</td>
</tr>
<tr>
<td>196007</td>
<td>247270813</td>
<td>2007-04-02 00:46:15</td>
<td>---</td>
<td>---</td>
<td>1.172</td>
<td>54396</td>
<td>58</td>
<td>65</td>
</tr>
<tr>
<td>196007</td>
<td>247270814</td>
<td>2007-04-02 00:46:15</td>
<td>---</td>
<td>---</td>
<td>0.462</td>
<td>54396</td>
<td>0</td>
<td>42</td>
</tr>
<tr>
<td>196007</td>
<td>247270815</td>
<td>2007-04-02 00:46:15</td>
<td>---</td>
<td>---</td>
<td>-1</td>
<td>54396</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>196007</td>
<td>247270816</td>
<td>2007-04-02 01:03:48</td>
<td>---</td>
<td>---</td>
<td>1.00</td>
<td>54397</td>
<td>4</td>
<td>60</td>
</tr>
<tr>
<td>196007</td>
<td>247270817</td>
<td>2007-04-02 01:05:41</td>
<td>---</td>
<td>---</td>
<td>2.056</td>
<td>54399</td>
<td>65</td>
<td>66</td>
</tr>
<tr>
<td>196007</td>
<td>247270818</td>
<td>2007-04-02 01:07:31</td>
<td>---</td>
<td>---</td>
<td>2.008</td>
<td>54401</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>196007</td>
<td>247270819</td>
<td>2007-04-02 01:09:23</td>
<td>---</td>
<td>---</td>
<td>2.048</td>
<td>54403</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>196007</td>
<td>247270820</td>
<td>2007-04-02 01:11:15</td>
<td>---</td>
<td>---</td>
<td>2.045</td>
<td>54405</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td>196007</td>
<td>247270822</td>
<td>2007-04-02 01:15:03</td>
<td>---</td>
<td>---</td>
<td>2.022</td>
<td>54409</td>
<td>62</td>
<td>67</td>
</tr>
<tr>
<td>196007</td>
<td>247270823</td>
<td>2007-04-02 01:16:55</td>
<td>---</td>
<td>---</td>
<td>2.046</td>
<td>54411</td>
<td>65</td>
<td>67</td>
</tr>
<tr>
<td>196007</td>
<td>247270824</td>
<td>2007-04-02 01:18:48</td>
<td>---</td>
<td>---</td>
<td>2.068</td>
<td>54413</td>
<td>65</td>
<td>67</td>
</tr>
<tr>
<td>196007</td>
<td>247270825</td>
<td>2007-04-02 01:20:42</td>
<td>---</td>
<td>---</td>
<td>2.076</td>
<td>54415</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>196007</td>
<td>247270826</td>
<td>2007-04-02 01:22:36</td>
<td>---</td>
<td>---</td>
<td>2.078</td>
<td>54417</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>196007</td>
<td>247270827</td>
<td>2007-04-02 01:23:57</td>
<td>---</td>
<td>---</td>
<td>1.432</td>
<td>54419</td>
<td>63</td>
<td>65</td>
</tr>
<tr>
<td>196007</td>
<td>247270828</td>
<td>2007-04-02 01:23:57</td>
<td>---</td>
<td>---</td>
<td>0.759</td>
<td>54419</td>
<td>0</td>
<td>43</td>
</tr>
<tr>
<td>196007</td>
<td>247270829</td>
<td>2007-04-02 01:23:57</td>
<td>---</td>
<td>---</td>
<td>-1</td>
<td>54419</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>196007</td>
<td>247270830</td>
<td>2007-04-02 10:57:55</td>
<td>---</td>
<td>---</td>
<td>1.027</td>
<td>54420</td>
<td>0</td>
<td>43</td>
</tr>
<tr>
<td>196007</td>
<td>247270831</td>
<td>2007-04-02 10:59:54</td>
<td>---</td>
<td>---</td>
<td>2.028</td>
<td>54422</td>
<td>61</td>
<td>66</td>
</tr>
<tr>
<td>196007</td>
<td>247270832</td>
<td>2007-04-02 11:01:46</td>
<td>---</td>
<td>---</td>
<td>2.032</td>
<td>54424</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>196007</td>
<td>247270833</td>
<td>2007-04-02 11:03:37</td>
<td>---</td>
<td>---</td>
<td>2.0169</td>
<td>54426</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>196007</td>
<td>247270834</td>
<td>2007-04-02 11:05:29</td>
<td>---</td>
<td>---</td>
<td>2.038</td>
<td>54428</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>196007</td>
<td>247270835</td>
<td>2007-04-02 11:07:21</td>
<td>---</td>
<td>---</td>
<td>2.041</td>
<td>54430</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>196007</td>
<td>247270836</td>
<td>2007-04-02 11:09:13</td>
<td>---</td>
<td>---</td>
<td>2.038</td>
<td>54432</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>196007</td>
<td>247270837</td>
<td>2007-04-02 11:11:07</td>
<td>---</td>
<td>---</td>
<td>2.07</td>
<td>54434</td>
<td>65</td>
<td>66</td>
</tr>
<tr>
<td>196007</td>
<td>247270838</td>
<td>2007-04-02 11:13:01</td>
<td>---</td>
<td>---</td>
<td>2.082</td>
<td>54436</td>
<td>65</td>
<td>66</td>
</tr>
<tr>
<td>196007</td>
<td>247270839</td>
<td>2007-04-02 11:14:54</td>
<td>---</td>
<td>---</td>
<td>2.055</td>
<td>54439</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td>196007</td>
<td>247270840</td>
<td>2007-04-02 11:16:45</td>
<td>---</td>
<td>---</td>
<td>2.008</td>
<td>54441</td>
<td>65</td>
<td>66</td>
</tr>
<tr>
<td>196007</td>
<td>247270841</td>
<td>2007-04-02 11:18:39</td>
<td>---</td>
<td>---</td>
<td>2.046</td>
<td>54443</td>
<td>64</td>
<td>67</td>
</tr>
<tr>
<td>196007</td>
<td>247270842</td>
<td>2007-04-02 11:20:35</td>
<td>---</td>
<td>---</td>
<td>2.048</td>
<td>54445</td>
<td>63</td>
<td>69</td>
</tr>
</tbody>
</table>

---

1. The Latitude and Longitude fields have been removed to maintain the confidentiality of the GPS data.
2. The Distance field has been truncated. Latitude, Longitude and Distance fields are reported to 17 significant digits.
4.2.3 Teranet’s Ontario Parcel Database

The Ontario Parcel Database is a standardized digital parcel map (map of land plots) that spans the province of Ontario. Information and boundaries were made available on approximately 4.5 million parcels within Ontario (Teranet, 2012). Figure 4.4 provides an example snapshot of the property boundaries taken in the vicinity of the University of Toronto, showing the level of detail recorded by this parcel database. Every property parcel is represented separately. Roads and public spaces are also assigned to their own separate parcels. In this figure, one property boundary (corresponding to part of the University of Toronto) has been selected to show available information, which included the Assessment Roll Number (ARN), PIN (Property Identification Number) and the address (street number, street name and the municipality). It should be noted here that the address information was not available for all parcels. This situation often occurred when multiple addresses were located on the same parcel.
Figure 4.4: Teranet Ontario Parcel Database Sample

The property parcel map was used frequently during data processing. For example property parcels were used to help produce more accurate clusterings of the GPS points. Also, the addresses from the property parcels were linked to the addresses described in the InfoCanada database companies, shown below, to infer a greater understanding of the types of firms that operated out of each parcel. This linkage is further described in Section 5.9.

4.2.4 InfoCanada Database of Companies

University of Toronto researchers purchased access to the 2007 InfoCanada database for all firms operating in the Greater Golden Horseshoe region of Southern Ontario. Information was available on approximately 330,000 firms. Attributes for each firm included the business name, street address (including city and postal code), number of employees, sales volume (in $), credit rating, founding year, office size, franchise status, and up to seven Standard Industry Classification (SIC) industry codes that describe a firm’s activities. A description of the activity codes were provided by the United States Department of Labor (2014).

Firms identified in the InfoCanada database of companies were linked to the Teranet property parcels, described previously, by matching the addresses between the two datasets. This linking was done for all
property parcels in the study region prior to the GPS analysis to obtain property parcel level land-use information.

### 4.3 Summary

This chapter provides a brief overview of the GPS data used to estimate the activity and inter arrival duration models estimated in my Ph.D. research. These GPS data were not originally collected for travel survey applications but instead for fleet management. It is important to understand the details of this data source as these details influenced the selected data processing techniques and the types of models that could be estimated using this data.

After describing the GPS data, this chapter then provides an overview of secondary data sources that were used to provide land-use context to the GPS points. These included the CanMap® RouteLogistics digital roadmap, zone-level population and employment data obtained from the Transportation Tomorrow Survey, property parcel boundaries and addresses from Teranet’s Ontario Parcel Database and firm addresses and attributes, such as industry classifications, from the InfoCanada database of companies.

Chapter 5 describes how these raw GPS data were processed to create a travel diary of trips, trip ends, tours and destinations. Chapter 5 then describes how the GPS data were linked with the secondary data sources described in this chapter so that they could be used for model estimation.
Chapter 5: Processing the GPS Data

5.1 Introduction to GPS Data Processing

The GPS data provided by Turnpike Global Technologies could not be used in their raw form to estimate the activity and inter-arrival duration models. These data had to first be processed to transform the raw data into a travel diary consisting of trips, trip ends and tours. Also, since behavioural data were not available from the GPS data, the GPS data were linked with the secondary data sources described in Section 4.2 to provide local land-use context that could be used to aid data processing and also as explanatory variables for model estimation.

Due to the size of the dataset, a completely automated data processing procedure was developed to perform the GPS data processing without the need for any manual intervention or verification. It was decided to use a series of custom written software applications, programmed in C#, to perform the bulk of the analyses as initial testing showed that implementing these routines completely within a GIS environment would cause unacceptably slow processing times. The focus of this chapter is to describe the procedures and algorithms that were used to convert the raw GPS data into a form suitable for estimating the models described in the subsequent chapters of this dissertation.

The desired output format of the processed data was specified at the outset of this research. The targeted output format is described in Section 5.2. Section 5.3 describes preliminary data processing, where the ArcGIS software package was used to relate the GPS data with the Teranet Ontario Property Database and the CanMap Route Logistics street map to provide local context for further data processing. Sections 5.4 through 5.8 describe the processing steps that were automatically performed using the custom C# software. Section 5.4 describes the trip-end identification procedure that tested for falsely-recorded or missed trip ends. Section 5.5 describes a clustering procedure that was used to group trip ends into disaggregate destinations that corresponded to visited physical locations. The clustered GPS trips ends were used for final data processing applications, such as removal of uninteresting trips (Section 5.6), identification of destinations that likely corresponded to a carrier depot (Section 5.7) and tour creation (Section 5.8). Section 5.9 then describes the linking of the GPS data with secondary data in order to infer land-use attributes for model estimation.
5.2 Output Variables Provided by the Data Processing

Before starting programming, a database design study was performed to create the requirements for the software outputs. The study envisioned different types of transportation models and analyses that would be of interest using the processed GPS data and ensured that the proposed database design could be used for these analyses. The final output included different tables for trips, trip ends, destinations, tours and trip GPS points. In the context of this research, these terms are defined below. Tables 5.1 through 5.5 describe the individual fields within these tables.

- **Trips**: A single truck journey starting and ending at different destinations.
- **Trip ends**: A period of time in which a truck is not moving and the driver is performing an activity. A trip end separates two trips.
- **Destinations**: Clustered GPS trip-ends representing visits to the same establishment.
- **Tours**: A group of consecutive trips that start and end at a destination that is likely a depot (base) location.
- **Trip GPS points**: Raw GPS data separated by the trip identifier. This table allows observations of route selection and trip details, but was not used in this research.
### Table 5.1: Trip Output Table

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>tripID</td>
<td>Index of this trip</td>
</tr>
<tr>
<td>departStopID</td>
<td>Index of the trip end from which the truck departed</td>
</tr>
<tr>
<td>arriveStopID</td>
<td>Index of the trip end to which the truck arrived after completing the trip</td>
</tr>
<tr>
<td>tourID</td>
<td>ID of the tour of which this trip formed one component</td>
</tr>
<tr>
<td>distTravelled</td>
<td>Total distance travelled in the trip in miles</td>
</tr>
<tr>
<td>tripDuration</td>
<td>Total elapsed travel time in minutes</td>
</tr>
</tbody>
</table>

### Table 5.2: Trip-End Output Table

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>stopID</td>
<td>Index of the trip end</td>
</tr>
<tr>
<td>stopType</td>
<td>Categorized trip ends based on available GPS points. Options include: Complete, ArriveOnly and DepartOnly. These are explained in Section 5.4</td>
</tr>
<tr>
<td>vehiID</td>
<td>Coded index of the vehicle under observation</td>
</tr>
<tr>
<td>firmID</td>
<td>Coded index of the carrier operating the vehicle</td>
</tr>
<tr>
<td>arriveTripID</td>
<td>Index of the trip inbound towards the trip end.</td>
</tr>
<tr>
<td>departTripID</td>
<td>Index of the trip outbound from the trip end.</td>
</tr>
<tr>
<td>destID</td>
<td>Index of the destination to which the trip end was assigned during clustering (see Section 5.5 for details about the clustering)</td>
</tr>
<tr>
<td>tourID</td>
<td>Index of the tour in which the trip end was observed</td>
</tr>
<tr>
<td>latitude</td>
<td>Latitude coordinate of trip end location</td>
</tr>
<tr>
<td>longitude</td>
<td>Longitude coordinate of trip end location</td>
</tr>
<tr>
<td>timeArr</td>
<td>Arrival day and time of day</td>
</tr>
<tr>
<td>timeDep</td>
<td>Departure day and time of day</td>
</tr>
<tr>
<td>dt</td>
<td>Time interval (in minutes) that vehicle spent at the trip end.</td>
</tr>
<tr>
<td>isImputedStop</td>
<td>Flag describing if the trip end was added in the trip-end detection algorithm. The data describing these trip ends are less accurate. See Section 5.4 for more details about imputed trip ends.</td>
</tr>
<tr>
<td>hasBeenRelocated</td>
<td>Flag describing whether an imputed trip end was moved to match nearby trip ends. See Section 5.4 for more details about moved imputed trip ends.</td>
</tr>
<tr>
<td>numRemainingStopsInTour</td>
<td>Number of trip ends occurring in the same tour after this trip end was completed. Used as an input variable in the activity duration models.</td>
</tr>
<tr>
<td>nDaysLastVisit</td>
<td>Number of days since the previous visit to the same destination. This is the dependent variable for the inter arrival duration models, described in Chapter 7.</td>
</tr>
</tbody>
</table>
### Table 5.3: Tour Output Table

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>tourID</td>
<td>Index of the tour</td>
</tr>
<tr>
<td>distance</td>
<td>Distance travelled on the tour, in miles</td>
</tr>
<tr>
<td>startTime</td>
<td>Date and time of day the truck left the trip-end marking the start of the tour</td>
</tr>
<tr>
<td>endTime</td>
<td>Date and time of day the truck arrived at the trip-end marking the end of the tour</td>
</tr>
<tr>
<td>tourDuration</td>
<td>Elapsed time in minutes between the start and end of the tour</td>
</tr>
<tr>
<td>nStops</td>
<td>Number of trip ends (not including starting and ending trip ends) in the tour</td>
</tr>
<tr>
<td>classification</td>
<td>Flag describing if the tour was closed (started and ended at the same destination) or open (started and ended at different destinations).</td>
</tr>
</tbody>
</table>

### Table 5.4: Destination Output Table

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>destID</td>
<td>Index of the destination</td>
</tr>
<tr>
<td>latitude</td>
<td>Latitude coordinate of the median GPS point in the destination</td>
</tr>
<tr>
<td>longitude</td>
<td>Longitude coordinate of the median GPS point in the destination</td>
</tr>
<tr>
<td>nObjects</td>
<td>Number of trip ends grouped in the destination</td>
</tr>
<tr>
<td>propertyID</td>
<td>ID of the property parcel from the Teranet Property Parcel database.</td>
</tr>
<tr>
<td>isDepot</td>
<td>Flag describing whether the destination likely corresponded to a carrier depot</td>
</tr>
<tr>
<td>isAtKnownLocation</td>
<td>Flag describing whether a destination was at a known location that would likely not correspond to a customer visit. Examples include border crossings, Highway 401 service stations and truck weigh stations.</td>
</tr>
<tr>
<td>isInternal</td>
<td>Flag describing if destination was internal to the study region or not.</td>
</tr>
<tr>
<td>avgStopLength</td>
<td>Average length of trip ends grouped in this destination.</td>
</tr>
</tbody>
</table>

### Table 5.5: Trip GPS Points Table

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ptID</td>
<td>Index of the GPS point</td>
</tr>
<tr>
<td>vehID</td>
<td>Coded index of the vehicle under observation</td>
</tr>
<tr>
<td>firmID</td>
<td>Coded index of the carrier operating the vehicle</td>
</tr>
<tr>
<td>tripID</td>
<td>Index of the trip</td>
</tr>
<tr>
<td>time stamp</td>
<td>Date and time of day of the GPS point</td>
</tr>
<tr>
<td>latitude / longitude</td>
<td>GPS point location</td>
</tr>
<tr>
<td>distance</td>
<td>Road distance travelled in miles since previously recorded GPS point.</td>
</tr>
<tr>
<td>odometer</td>
<td>Distance travelled by the vehicle in miles (where available)</td>
</tr>
<tr>
<td>GPS speed</td>
<td>Instantaneous vehicle speed recorded by the GPS unit in miles per hour</td>
</tr>
<tr>
<td>ECM speed</td>
<td>Maximum recorded speed since previously recorded GPS point in miles per hour</td>
</tr>
</tbody>
</table>

### 5.3 Preprocessing GPS Data

As will become apparent in later sections of this chapter, the data processing required land-use context from the secondary data sources described in Section 4.2. The secondary data were linked to the GPS data using an ArcGIS Python script so that they were available in the custom C# programs described later in this chapter.
1. The GPS points were converted from latitude/longitude coordinates to a UTM projection system. This step was conducted so that distances could be conveniently computed between GPS points and also between GPS points and the road network.

2. A spatial join was conducted on each GPS point to see if that point was located within a 40 m buffer of a freeway centreline. The road network was obtained from the CanMap RouteLogistics electronic map.

3. A spatial join was conducted on all GPS points to find the Planning District according to the 2006 Transportation Tomorrow Survey zone system.

4. A spatial join was conducted to find the Teranet property parcel in which each GPS point was located.

5. Finally, the distance between every GPS-recorded stop and the nearest major road was calculated in meters. The road network was obtained from the CanMap RouteLogistics electronic map. Only distances lower than 50 m were output to save computational processing time.

These attributes were all required to provide land-use context to the automated GPS processing steps described below. Further land-use attributes that were used for transportation model estimation were added after the automated data processing, and are described in more detail in Section 5.9.

The raw GPS data with the attached land-use attributes were then input into the custom C# data processing program. This program first read in the GPS data and then performed basic data cleaning to remove GPS points that had been marked as invalid by the RouteTracker™ GPS units. Invalid GPS points were marked by setting the Distance field to a value of -2 or by setting the latitude and/or longitude values to 0. Following the data cleaning the program could then focus on converting the raw GPS data to a travel diary of trips, trip ends and tours.

### 5.4 Trip and Trip-End Identification

The next processing step was to convert the continuous GPS data stream into distinct trips and trip ends. In this thesis a stop refers to a period of at least five minutes of zero speed, recorded by the RouteTracker™ GPS unit, while a trip end refers to a stop (or a missed stop) that is considered to likely correspond to an activity.

In this step, the GPS points were sorted by company, by vehicle and then by timestamp. The software then examined every GPS point as follows.
• For a motion GPS point, the software predicted whether a trip end was likely missed between this GPS point and the previous motion GPS point using the rules described in Section 5.4.3. Such a case is called a false-negative trip end. In this case the software created an imputed trip end between the two GPS points. If no missed trip end was suspected then the GPS point was added to the current trip.

• For a stop GPS point, the software tested for a false-positive trip end using the rules described in Section 5.4.2. If the stop was not identified as being falsely identified then a trip end was created as described below. Otherwise the stop was removed and subsequent GPS points were appended to the current trip.

When creating trip ends it must be noted that sometimes a vehicle location would “jump” after a stop was recorded, meaning that the vehicle arrived at one location and departed from another location. This “jump” occasionally occurred within the study area and was always assumed for stops located outside of the study area since data were only provided by Turnpike Ltd. after the last stop before entering the study area for an inbound trip and before the first stop located outside the area for an outbound trip. Hence stops located outside of the study region only had departure information if the trip was inbound towards the study region and arrival information for trips that were outbound from the study region.

Three different types of trip ends were created in the software. For arrival-only trip ends, information was only recorded about the vehicle arrival. Hence the arrival time was available but the departure time was not. Similarly, only departure information was available for departure-only trips ends. Both arrival and departure information were maintained for complete trip ends.

Accurate stop locations were only available on stop arrival as the GPS point that marked the departure was not located at the stop but instead at a motion point that was usually at least 500 m removed from the stop location. Hence the stop location was taken from the arrival GPS point. The stop location of departure-only stops was not accurately known.

Separating short-duration trip ends from other stops has proven to be a challenging problem and has been examined in multiple publications. Section 5.4.1 presents a literature review of trip-end detection strategies used in other GPS studies. Sections 5.4.2 and 5.4.3 then describe the rule-based analysis that was used to identify false-positive trip ends and create imputed trip ends where trip ends were likely missed. As no precise location was available for missed trip ends, imputed trips ends were moved in a subsequent processing step to the nearest trip end whose location was accurately recorded. This is
described in Section 5.4.4. Finally, Section 5.4.5 provides some summary statistics about the trip end identification data processing.

5.4.1 Literature review of trip-end detection in GPS studies

The trip-end detection algorithms presented in the literature depended greatly on the GPS receiving equipment available to the researchers. For example, different trip-end detection algorithms were used to process data obtained from vehicle-powered receivers and from portable receivers. Different approaches were also used between researchers that supplied GPS units to survey respondents, and hence could select the GPS unit used, and researchers who used GPS data that was originally intended for other purposes.

Many GPS surveys used vehicle-powered GPS devices. For example, Battelle (1999) used GPS receivers that were connected to the truck on-board computer. These GPS units were calibrated to detect engine ignition. Trip-end identification in this study was performed only through engine-off/engine-on combinations. This approach has downside that it might miss short trip ends where drivers leave the engines running however it is unlikely that it would falsely-identify trip ends, except possibly at locations where trucks may need to make long stops, for example at level rail crossings. These cases can be easily identified using GIS software.

Multiple other studies used GPS receivers that plugged into vehicles for their power source, often using the cigarette lighter. These receivers lose power when the vehicle engine is turned off, which creates a gap in the GPS record. Multiple researchers used a threshold time interval for the power-off state to mark a trip end. These researchers often used additional criteria to detect trip ends when a vehicle engine is not turned off and also to detect trip ends in cases of signal blockage. Suitable trip-end detection thresholds were found to vary based on local traffic conditions. A brief summary of these trip-detection algorithms is provided below.

Schönfelder et al. (2002) identified a trip end in the event of a recording gap (signifying that the vehicle was turned off) and also when the vehicle remained stationary for a period of at least 120 seconds. This study was conducted in the relatively small town of Borlänge, Sweden, and hence it is unlikely that vehicles would remain stopped longer than this period due to traffic conditions.

The computer software described in Srinivasan et al. (2006) used a dwell time threshold for the time that an engine was turned off and a different dwell time threshold that identified trip ends during a period where the vehicle did not move. The threshold values used were not specified in this paper.
In a truck GPS data-collection effort, McCormack & Hallenbeck (2006) used a trip-end identification criterion where the engine was turned off for at least three minutes. The authors also flagged locations to remove trip ends where trucks might be forced to make long stops that do not correspond to activities, such as at drawbridges and railroad crossings. In another GPS study of truck driving patterns, Logendran & Peterson (2006) identified a trip end as any time that a truck does not move for a period of at least five minutes.

In a pilot GPS survey of commercial vehicles in Melbourne, Australia, Greaves & Figliozzi (2008) tested different trip-end identification thresholds to distinguish between genuine trip ends and stops associated with traffic conditions. They found that a threshold of 120 seconds was appropriate for cars while a 240-second threshold was preferable for trucks in their study location.

Du & Aultman-Hall (2007) adapted the previously published trip-end detection method of Stopher et al. (2003). Du & Aultman-Hall used a minimum and a maximum dwell time threshold. No trip end is considered possible below the minimum threshold, and a trip end is considered guaranteed above the maximum threshold. Additional tests were conducted for trip ends between these two thresholds. These tests included: 1) a check for reversed travel direction before and after the potential trip end, and 2) a test if the trip-end was more than a pre-defined distance threshold from road links. They found that for Lexington, KY, a medium-sized city with a population of 250,000 people, suitable minimum and maximum trip-identification thresholds were 40 and 140 seconds respectively. They found that using the two thresholds was also superior to using a single threshold. The authors also tested for trip ends occurring under signal loss conditions. These were identified by assuming that the average speed observed before and after the signal-loss period was maintained throughout the blackout period, and hence any additional time was due to an unobserved trip end.

The study presented in McCormack et al. (2010) is of particular interest for this research as their study also used GPS data collected from GPS vendors that monitored commercial vehicle travel. Their study encountered many of the same difficulties experienced in this research since their commercially-available GPS data also provided infrequent readings and the size of their data source made manual checking of the processed data infeasible. McCormack et al. also had to filter falsely identified trip ends due to traffic congestion and identify missed trip ends that occurred during periods when the GPS signal was lost. They found that a three-minute trip-end identification threshold was suitable for their study area as most signals have a cycle length that is shorter than this time threshold. Also, traffic congestion is not usually sufficient in the central Puget Sound region to cause a truck to remain stationary for this time period. Upon validation, the authors found the presence of many trip ends with a duration of less
than three minutes. These trips were missed by the trip-end identification algorithm. In their case, however, the GPS receivers used by one of their data providers also included engine data. For this data provider McCormack et al. added an additional trip-end detection criterion that recorded a trip end if the vehicle was placed in “park” or if the engine was turned off. To identify trip ends that were missed in periods of signal loss, these authors created an additional trip end if the average speed between any two subsequent GPS points was less than five miles per hour.

A number of other studies were made where at least part of the sample used portable GPS devices in order to monitor non-motorized trips. These devices are battery powered and always remain on. Tsui & Shalaby (2006) performed a GPS study in Toronto and used a trip-end detection threshold of 120 seconds. Those authors reported that this threshold did not miss any activities compared with post-survey logs but that numerous false-positive stops were recorded indicating that this threshold may have been too short for Toronto traffic conditions.

Stopher, Fitzgerald & Xu (2007) provided vehicle-based and portable GPS units depending on the expected travel habits of the individual. Their processing software recorded a trip-end if the GPS unit remained stationary for a duration of at least two minutes. A prompted-recall survey was conducted afterwards allowing survey participants to confirm the processed trips and trip ends.

Bohte & Maat (2009) used a trip end identification threshold of 3 minutes in their study, which was focused on inferring trip mode and trip purpose from GPS data. It should be noted that they also used an internet recall survey after the week-long GPS study to validate the processed trips.

The final study reviewed before determining the trip-end identification algorithm is that of Schuessler & Axhausen (2009), who used a slightly different approach than the preceding studies to automatically process passively-collected GPS data obtained from portable GPS surveying units. First, after removing erroneous data points the authors used Gauss kernel smoothing to reduce random errors in the GPS position. To detect trip ends during periods where the receiver maintains signal reception they identified a trip end if one of two criteria were met. The first criterion was a recorded speed of below 0.1 m/s for a duration of at least 120 seconds. The second trip end criterion searched for GPS bundles that indicated locations with no movement. Schuessler & Axhausen used a threshold of 900 seconds (15 minutes) between consecutive GPS points to identify trip ends that occurred during periods of signal loss. This threshold value is much higher than that used by other authors; it was used to remove false-positive trip-end identification in the case of poor signal reception.
5.4.2 Adapted algorithm to identify false-positive and missed trip ends

The RouteTracker\textsuperscript{TM} units automatically recorded a stop if the vehicle had a period of at least five minutes where a zero maximum speed was recorded. Preliminary testing found that many stops were recorded by RouteTracker\textsuperscript{TM} units that were more likely due to traffic congestion than due to a stop representing an activity. These are referred to as false-positive stops and had to be cleaned from the data before model estimation. The following tests for false-positive stops were conducted in the order shown below.

Many false-positive stops were recorded on freeways. Figure 5.1 shows GPS stops that occurred on freeways in the general vicinity of Mississauga and Pearson airport. GPS stops recorded within a 40 m buffer of a freeway link are shown in this figure using blue circles. Please note that all other GPS stops were suppressed to respect confidentiality of the carriers. All GPS stops located within this 40 m buffer of a freeway link were removed regardless of their duration. The rationale behind removing all stops located on freeways is that these stops were highly unlikely to correspond to activities at nearby commercial establishments, regardless of their duration. The 40 m buffer was chosen by visually examining the GPS data. This buffer width was found to be sufficient to capture the vast majority of GPS points on the freeway but was also small enough that it did not encroach on any properties adjacent to the freeway.

Stops with a duration of less than five minutes were removed to reflect the five-minute time threshold for recording stops by the Turnpike Routetracker unit. This restriction was performed to keep a consistent definition of a trip end as having a duration of at least 5 minutes as this is the criterion used by the RouteTracker\textsuperscript{TM} unit. As is seen in Section 5.4.5, few points were removed using this criterion.

A visual examination of the GPS data also showed the presence of multiple stops on arterial roads that were likely caused by traffic conditions instead visits for commercial activities. Figure 5.2 shows an example of such stops on Eglinton Ave. West between Highway 427 and Kipling Ave. This figure shows GPS stops located within 25 meters of a major road (defined by the CanMap\textsuperscript{®} RouteLogistics street map) superimposed on an aerial view of the region obtained using Bing Maps. All other GPS points were suppressed in order to respect carrier confidentiality. As can be seen in this figure, the vast majority of these GPS stops were not located in the vicinity of any commercial or residential establishments including the GPS points near the Highway 427 off-ramp (to the left of the image), near Martin Grove.
road and near Kipling Ave. Please note that the apartment buildings west of Kipling and south of Eglinton have no parking facilities near Eglinton Ave.

The following test was used to identify false-positive stops near major roads. This test was performed within any planning district from the 2006 TTS survey (also reflected as Regions 0-6 in Figure 5.3) as it was felt that these regions have a higher probability of having sufficient traffic congestion to cause falsely-reported stops on major urban arterial roads. A false-positive stop was recorded in this region if the stop duration was less than 15 minutes and if the stop was located within 20 metres of a major arterial or secondary highway. It is recognized that some of the removed stops likely represent visits for commercial activities; however these thresholds were selected with the prime intention of removing false-positive stops. Section 5.4.5 summarizes the data cleaning results, including how many stops were removed due to each of the following criteria.

Figure 5.1: Example False-Positive Stops Recorded on Freeways
5.4.3 Check for missed trip ends

Recent studies of commercial vehicle travel clearly showed that short trip ends are a common occurrence in urban commercial vehicle travel. For example in the Region of Peel Commercial Travel Survey, Roorda *et al.* (2013) compared driver tours reported from a paper-and-pencil questionnaire and observed by GPS for a single travel day from 37 drivers. This survey also used Turnpike RouteTracker™
GPS units, the same type of units used by vehicles in the present research. In this study, 101 out of the total 435 paper-pencil recorded activities, 23%, had a duration of below 5 minutes and hence were missed by the RouteTracker™ GPS units. Similar behaviour was expected for this study although it could not be quantified due to the different sample. The presence of short stops was also reported by McCormack *et al.* (2010).

While these short trip ends could not be detected from the available GPS data, the software tested for longer missed trip ends occurring between two motion GPS points. Missed trip ends could occur, for example, in the event that a trip end occurred in a location of signal loss. Since no information was available in between two motion GPS points, a similar test for missed trip ends (called an imputed trip end in this thesis) was performed as in McCormack *et al.* (2010). This test automatically created an imputed trip end if the time interval between subsequent motion points was over 30 minutes. It was felt that most trucks could travel a distance of 500 m in 30 minutes, even in severe traffic conditions. An imputed trip end was also created if the average speed between two motion points was less than 5 km/hr and if the time difference between the two points was also greater than five minutes. A trip end was marked with the first GPS point as the arrival stop and the second GPS point as the departure stop. As the location of the trip end is likely not at either of these GPS points, the trip end was flagged as an imputed stop to allow it to be moved to nearby locations.

### 5.4.4 Moving departure and imputed trip ends

After the cleaning process that separated the GPS data stream into separate trips was completed, the locations of departure-only and imputed trip ends were adjusted as these trip ends were created from *Motion* GPS points and not from *Stop* GPS points. Motion GPS points were recorded every 500 m, although the recording distance could be increased to 2 miles if the vehicle reached freeway speeds.

Departure-only and imputed trip ends were moved to the location of their nearest *arrival-only* or *complete* trip end, within a 4 km threshold. The four-kilometer threshold was selected as this corresponded to the upper limit of the distance between two subsequent motion points recorded by the RouteTracker GPS units.

### 5.4.5 Summary data describing data cleaning results

Table 5.6 shows a summary of the total number of GPS points removed from the dataset and also the added number of imputed trip ends.
Table 5.6: Trip-End Cleaning Summary

<table>
<thead>
<tr>
<th>Description</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of stops identified by RouteTracker units</td>
<td>323,799</td>
</tr>
<tr>
<td>False-positive stops within 40 m freeway buffer</td>
<td>-34,389</td>
</tr>
<tr>
<td>False-positive stops below 5 minute duration</td>
<td>-1,154</td>
</tr>
<tr>
<td>False-positive stops 5-30 minutes within 20 m of major road centerline</td>
<td>-39,045</td>
</tr>
<tr>
<td>Removed false-positive stops</td>
<td>-74,588</td>
</tr>
<tr>
<td>Added imputed stops</td>
<td>56,058</td>
</tr>
<tr>
<td><strong>Total trip ends after cleaning</strong></td>
<td>305,269</td>
</tr>
</tbody>
</table>

While the five-minute threshold is known to miss short stops, it can be seen that local traffic conditions regularly caused conditions where a truck remained stationary for at least a five-minute period. This is demonstrated by the 34,389 stops that occurred on freeways alone. While the occasional vehicle was seen to park beside a freeway for extended time periods, this number likely reflects the heavy traffic congestion in the Greater Golden Horseshoe.

A low number of stops under the minimum threshold of five minutes were removed, corresponding to 0.4% of the total stops after cleaning. Hence this criterion is seen to not have a significant impact on analyses conducted using this dataset.

39,045 stops, corresponding to 12% of stops identified by the RouteTracker units, were removed from major arterial roads. These stops were located within 20 m of a major road centerline and also had a duration of between 5 and 15 minutes. While some of these stops were likely for an activity, visual inspection of the data showed that many of these stops could only reasonably be due to signals or traffic congestion. Even after this data cleaning, multiple stops were still found near intersections and on major arterial roads. For example, Figure 5.7 (located in Section 5.5.3) shows many stops located near an intersection.

### 5.5 Clustering Trip-Ends into Destinations

The processing steps described in Sections 5.3 and 5.4 involved converting the GPS points into a list of trips separated by trip ends, which corresponded to activities. At this point all trip ends were independent events. The goal of this section is to describe the clustering technique that was used in this research to group the trip ends into destinations, representing visits to the same establishment. Grouping the trip ends into destinations was useful to analyze travel behaviour over longer time periods. It also allowed for better interpretation of firm behaviour in models estimated using the GPS data as
these models could use the business establishment as the base unit of analyses instead of individually-recorded trip ends.

The concept of grouping GPS-observed trip ends into destinations is not new. Studies using GPS data have grouped GPS trip ends to home and work locations (and sometimes also to shopping and leisure activities) based on their proximity to reported addresses of home and work, and shopping/leisure locations (if available) by survey respondents. One such study was described in Schönfelder et al. (2002).

In the current research no information was provided either about the carrier or their clients and hence addresses were unavailable. Therefore the approach of grouping trip ends within a threshold distance from known locations could not be used. Instead, a statistical clustering approach was used to group the trip ends into destinations.

The organization of this section is as follows. Section 5.5.1 first provides a background of statistical clustering algorithms. This section then introduces different measures of clustering quality including external measures, which compare clustering results against an external validation data source, and internal measures, which examine the clustering quality only using the GPS dataset. These clustering measures formed the basis of comparison between different clustering algorithms.

The research into selecting a suitable clustering algorithm was conducted in two phases. The first phase, described in Section 5.5.2, compared the clustering performance using a test dataset for a single company that also provided driver records of pickups and deliveries that functioned as an external data source to validate the clustering methods. This dataset allowed testing the suitability of different clustering methods and estimation of required parameters for clustering GPS trip ends into destinations. Since driver records are generally not available from third party sources of GPS data, attention was also focused on identifying “internal” measures of cluster quality (that do not require driver records) that provide similar estimates of clustering parameters as external measures. The second phase of the research involved testing the clustering algorithm selected in the first phase using the GPS records of 40 firms.1 This second phase is described in Section 5.5.3.

1 At this point in the study, the data were not yet available from the full 77 firm GPS data. This dataset is a slightly older and smaller dataset provided by Turnpike Global Technologies.
5.5.1 Background on clustering methods

Clustering is a commonly used data mining technique that groups objects into non-predetermined clusters such that objects within a cluster are similar to each other while objects are dissimilar to those in other clusters. Care must be taken when selecting a clustering method as different clustering methods can produce markedly different results. Two different classes of clustering methods were examined in this research; partitioning methods and hierarchical agglomerative methods.

**Partitioning methods**

Partitioning methods divide a database of $n$ objects into $c$ ($\leq n$) clusters where each cluster must contain at least one object and each object must belong to exactly one cluster. The number of clusters must be defined before the clustering analysis. Three commonly used partitioning clustering heuristics include the $k$-means method, the partitioning around medoids (PAM) method and the expected-maximization method. A description of these methods can be found in Gordon (1999) and Han, Lee & Kamber (2009).

One issue with using partitioning methods in this application is that the proper number of clusters was not known a priori. According to Gordon (1999) the most commonly used approach to selecting the number of clusters is to obtain clustering results for a range of clusters, $c$, and then select the number of clusters that produces the best results according to goodness of fit statistics. Another approach is to use a hierarchical agglomerative method to provide an initial estimate of the number of clusters.

**Hierarchical methods**

Whereas partitioning clustering methods divide objects into a predefined number of clusters, hierarchical methods build a hierarchy of clusters. Hierarchical clustering can be divided into agglomerative and divisive methods. Hierarchical agglomerative clustering (HAC) is more common than hierarchical divisive clustering. It operates as follows:

1. Create a distance (or dissimilarity) matrix showing the distance between any two points. In this application the distance is the Euclidean (straight-line) distance between two GPS points.
2. To start this process, every object is assigned to its own cluster.
3. In each step the closest two clusters are merged together. After merging, the distances between all clusters are recalculated. This step is repeated until all objects have been assigned to a single cluster.
4. The hierarchical clustering tree can be cut at different heights (distances) or for a set number of clusters to produce different clustering results. The height at which the hierarchical tree is cut is called the distance threshold as it represents the largest distance within a cluster.
Different HAC methods have been devised, which vary in the manner in which they calculate the distance between two clusters. Two different distance measures considered in this research included the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and Ward’s methods. These methods were recommended by Kaufman & Rousseeuw (1990) compared with other HAC methods.

In the UPGMA method, the distance \( d_u \) between clusters \( R \) and \( Q \) is defined as the average of the distances between all the points in both clusters. In the equation below, \( |R| \) represents the number of objects in cluster \( R \), and \( d(i, j) \) refers to the distance between objects \( i \) and \( j \).

\[
 d_u (R, Q) = \frac{1}{|R||Q|} \sum_{i \in R} \sum_{j \in Q} d(i, j) 
\]

(Ward’s method is intended for interval-scaled measurements and Euclidean distances, both of which apply to the GPS data. Let \( \bar{x}(R) \) be the centroid of cluster \( R \). The distance measured using Ward’s method between two clusters, \( d_w \), is defined as

\[
 d_w (R, Q) = \sqrt{\frac{2|R||Q|}{|R| + |Q|} \left\| \bar{x}(R) - \bar{x}(Q) \right\|^2} 
\]

It is shown by Meilă (2007) that selecting the minimum distance using Ward’s method minimizes the increase in the error sum of squares (ESS), which is defined as the sum of the squared Euclidean distances between the points, \( x_i \), in cluster \( C \) and its centroid.

**Measures for evaluating clustering quality**

Evaluation of clustering methods can be conducted using external tests, which compare clustering results with information not used in constructing the clustering, and internal tests, which compare clustering results using the original data set (Gordon, 1999).

External tests judge how well the produced set of clusters agrees with an externally-provided grouping. One external validation measure shown to be appropriate is the corrected Rand statistic (Gordon, 1999). This index lies between 0 and 1, with 1 meaning that cluster results \( C \) were perfectly matched with the “true” externally-provided clusters \( C’ \). Another external validation statistic is called the VI index, which measures the amount of information lost and/or gained by changing from \( C \) to \( C’ \) (Meilă, 2007). This index lies between 0 and 1, with 0 meaning that clusters are perfectly matched and 1 meaning that clusters are completely dissimilar.
Internal validation judges if the output of a clustering analysis creates compact and isolated clusters. The internal measure used in this application was the average silhouette measure presented by Rousseeuw (1987). This is a measure of how tightly the data within a cluster are grouped. The silhouette of a cluster is defined as

\[ s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \]  

where \( s(i) \) is the silhouette value of object \( i \) within a cluster, \( a(i) \) is the average distance of object \( i \) from all objects in the same cluster, and \( b(i) \) is the average distance from the nearest neighboring cluster (has the lowest distance among all other clusters). The average silhouette of a cluster is the average of the silhouette for every object within the cluster. The silhouette measure varies between -1 and 1, where values close to 1 represent very tight and compact clusters, meaning that \( a(i) \ll b(i) \).

Two other potential internal measures were the G2 and G3 stopping rules presented in Gordon (1999). According to this source, these measures should be used for smaller numbers of clusters as they can display distracting behavior for larger numbers of clusters. Tests confirmed that these measures were not appropriate for the application.

5.5.2 Clustering phase 1: one firm where driver records and GPS data are both available

Second Harvest is a charitable organization located in Toronto, Canada, that collects donated food from restaurants and grocery stores and delivers it to approximately 250 social service programs in the Greater Toronto Area. Second Harvest agreed to install GPS units on six of their vehicles during the period from May 31 to August 22, 2007, and also provided driver pickup and delivery records for the same period. The GPS units placed on the vehicles were the same RouteTracker\textsuperscript{TM} units used in this research. The driver records identified the customer, address, and the delivery/pickup time, thereby providing the information necessary for external testing of clustering methods. The customer addresses from the driver records were geocoded using ESRI’s ArcMap GIS software. The geocoded address was verified using other sources (such as Google Maps or Yahoo Maps) if researchers became suspicious about a particular address (which usually occurred when an address was not linked to any GPS points).
**Linking GPS data with driver records**

If a driver stopped at a destination to pick-up or deliver goods, the locations of the GPS trip end and the geocoded address from the driver records generally did not agree exactly. This occurred for a variety of reasons:

1. Geocoding software interpolates addresses within a street block and hence the geocoded address may not have corresponded to the exact address location.
2. The geocode was assigned along the road while the actual parking or loading location may have been offset from the side of the road.
3. Drivers may have needed to park away from their destination. This can be especially prevalent in dense urban areas, which tend to have fewer parking and loading facilities.
4. The RouteTracker™ unit is understood to usually record locations within an accuracy of 15 meters. Hence GPS errors may cause some discrepancy between two locations.
5. Differences in the arrival time between the driver records and the GPS trip ends could have arisen due to drivers omitting or erroneously recording stops, drivers rounding their arrival time, inaccurately set watch or vehicle clocks, or that drivers recorded the arrival time after walking from the stopped vehicle to the destination. The time recorded by the GPS system is extremely accurate and was not considered to be a major source of error.

Here, the variable $d$ refers to the straight-line, Euclidean, distance between a GPS trip end location and the geocoded destination from the driver record. The variable $t$ refers to the time interval (in minutes) between the trip end time recorded by the GPS unit and the arrival time listed in the driver record.

GPS stops were linked to driver reported destinations if $d$ and $t$ fall within reasonable thresholds. These thresholds were chosen based on extensive visual comparison of both the GPS and driver-reported information. GPS stops were linked to driver-reported destinations if either of the following were true:

\[ t \leq 15 \text{ minutes and } d \leq 500 \text{m}; \text{ or} \]

\[ t \leq 30 \text{ minutes and } d \leq 200 \text{m} \]

**Clustering analysis**

The goal of phase 1 of this research was to test different clustering methods and their respective parameters to identify a preferred method suitable for clustering GPS data when no driver records are available. Therefore, the results of each clustering method were validated using internal and external
validation measures. This analysis was conducted using the open-source software R (R Development Core Team, 2010).

The first clustering analysis test compared the results from the UPGMA and Ward’s HAC methods. Two goals were to: 1) find the method that produced superior clustering results, and 2) test whether the internal and external clustering validation measures agreed when selecting the optimal distance threshold parameter.

Figure 5.4 shows a comparison of external clustering quality measures (VI and corrected Rand indices) and the average silhouette internal quality measure for both UPGMA and Ward’s hierarchical agglomeration clustering techniques using different distance thresholds. This figure shows that Ward’s method was more suitable than UPGMA in this application for the following reasons.

1. The corrected Rand index was consistently higher when comparing the results from Ward’s method against the driver records than for the UPGMA method.
2. The VI index was consistently lower when comparing the results from Ward’s method against the driver records than for the UPGMA method.
3. The corrected Rand index, the VI index and the number of clusters were less sensitive to the distance parameter, allowing more tolerance for a non-ideal distance threshold.

The corrected Rand and VI indices both suggested that 600 m was the ideal distance threshold for Ward’s method. The average silhouette measure agreed with the external validation results by also finding that 600 m was an optimal distance threshold. This distance threshold was optimal for this one firm but may not apply to all firms. When applying this method in practice, it was decided to run the clustering using many distance thresholds and to use an internal distance measure to recommend an appropriate distance threshold for each firm.

Partitioning methods were also tested. In spite of being theoretically preferable, the Expected-Maximization method produced poor results in preliminary testing. This was due to convergence issues that were likely caused by the low number of objects (GPS trip ends) in many clusters.

The k-means and PAM methods were run for between 120 and 145 clusters (Ward’s method with a distance threshold of 600 m produced 139 clusters). The results are not shown here, but the PAM method was found to significantly outperform the k-means method for the average silhouette measure and the corrected Rand and VI indices. The average silhouette measure indicated that 127 is the ideal
number of clusters while the external measures indicated that 132 clusters is preferable. PAM using 127 clusters produced results that were slightly inferior to the clustering found using Ward’s method.

The PAM method also required significantly greater computational effort. Using Ward’s method on a 2.4 GHz Intel Core2 CPU took 17 seconds while the set of PAM clustering runs between 120 and 145 clusters took approximately three hours to complete on the same computer. Hence, due to the computational effort and the better clustering results, this analysis showed that Ward’s method was the preferred clustering method. Figure 5.5 shows clustering results for downtown Toronto using Ward’s method with a 600 m distance threshold.
Figure 5.4: External and Internal Validation Measures for the UPGMA and Ward’s Hierarchical Agglomerative Clustering Methods
5.5.3 Clustering phase 2: GPS data from multiple firms

The goal of clustering was to group GPS recorded trip ends when driver records were unavailable. In the previous section, the results of different clustering methods were evaluated using GPS data from one firm that also provided driver records showing the ‘ground-truth’. In this second phase, the preferred clustering method in the first phase was tested for multiple firms for which only GPS data were available. Opportunities for refinement of the clustering procedure were sought in this phase.

Data

The data used for this research was an older, smaller, dataset that was similar to dataset described in Chapter 4 of this dissertation. GPS records were provided showing the vehicle movements of 40 firms (818 trucks) between October 1\textsuperscript{st} and October 31\textsuperscript{st}, 2006 in the Greater Golden Horseshoe (GGH) region.
in southern Ontario, Canada. Data cleaning and trip-end identification were performed using the same manner described in Sections 5.3 and 5.4.

**Refining clustering results using property boundaries**

Analysis of the GPS data showed that while Ward’s method was very good at identifying clusters in the data, it was very difficult to find a single appropriate distance threshold that applied to the entire geographic domain. Two particular issues were identified:

1. The distance measure between two clusters using Ward’s method increased with the number of objects within each cluster (see Equation 5-2). This caused Ward’s method to produce tight and isolated clusters but made specifying a distance threshold difficult for this application.
2. There was a wide variation in property sizes, ranging from a few square meters to several hectares.

These two issues often appeared at the same locations. For example, the depot of a large trucking firm may cover large areas in order to be able to park their trucking fleet. This same depot would also be the location of many trip ends. Hence large distance thresholds (often in excess of 1.5 km) were required to join all the trip ends occurring in these facilities. These large thresholds, however, would join smaller clusters at great distances that could not conceivably belong to the same destination. Figure 5.6 shows a schematic of some of the issues in selecting an appropriate distance threshold using Ward’s method.

The clustering phase 2 exercise revealed that a single distance threshold could not be appropriately applied throughout the study region. Also, cases 1 and 2 of Figure 5.6 were difficult to differentiate using the GPS data alone. This issue was resolved by using property parcels from the Teranet Ontario property parcel database to aid the clustering process. The following two-step clustering process was proposed.
In the first step, GPS trip ends were clustered using Ward’s method using distance thresholds between 400 m and 700 m. The average silhouette measure was used to find the optimal distance threshold within this range for each firm. A maximum distance threshold of 700 m was chosen as GPS points that are over 700 m apart were unlikely to be visits to the same destination, except on large properties. As shown in Figure 5.6, erring on the side of a low distance threshold would lead to correct results for Case 2 and Case 3 scenarios, but could lead to multiple clusters erroneously being identified on large property parcels.

In the second step, the geometric median GPS point was identified for all clusters. This was the GPS point with the lowest sum-of-squares distances to other GPS points within the same cluster. Clusters whose geometric median fell within the same parcel boundary were merged into a single cluster.

Figure 5.6: Schematic Showing Issues in Selecting Appropriate Distance Threshold
Visual examination of the clustering using this two-step approach showed promising results. Figure 5.7 shows examples of clusters identified using this process. Figure 5.7a shows a location with one large building having parking locations or docking bays on either side of the building. When using a small distance threshold (e.g. 700 m) the points on either side of the building were grouped into separate clusters. Using this two-step approach, however, the two separate clusters created in the first step were joined into a single cluster. Figure 5.7b shows two small nearby clusters that were unlikely visiting the same destination and that were correctly not joined together.

Figure 5.7c shows a significantly more complex area, where two destinations with large numbers of points are located near each other. The two-step clustering approach tested here accurately created large clusters in the two large properties. In one property, the GPS trip ends span approximately an 800 m by 600 m area while the other nearby property extends over 450 m in one direction. These properties are significantly larger than the distance separating them.

5.5.4 Implementation (scaling) issues

Computational issues were found when clustering trip ends for firms possessing large numbers vehicles and making many trips in the region. This was not surprising because Ward’s method has an order of at $n^2$ time complexity (Maimon & Rokach, 2005).

In the dataset described in Chapter 4, over 35,000 trip ends were observed for four firms, and just over 53,000 trip ends were recorded for the firm with the most observed trip ends. Clustering the trip ends from these firms was very time consuming. To reduce the computational burden, the entire domain was subdivided into smaller regions, shown in Figure 5.3. These regions were carefully selected so that they were separated by geographical features (such as rivers, wide freeways or wide areas of undeveloped land) so that clusters did not traverse region boundaries. This technique reduced the clustering time as clustering many smaller regions is much faster than clustering a large region due to the $O(n^2)$ time complexity of the Ward’s method cluster algorithm.

Even with the separation into different regions the clustering was still a time-consuming process that took three days of computer time. It is anticipated that more efficient programming, more powerful computers and a finer division of clustering regions could still reduce the clustering time to manageable levels for datasets that are larger than the dataset analyzed in this study.
Figure 5.7: Sample Clustering Results using Two-Step Clustering Methods
(note that roads and property boundaries have been modified so that physical locations are not revealed)
Destination-level attributes were required to infer land-use data about the visited location. A representative point was chosen for each cluster by finding the geometric median GPS point, which was found by calculating the average distance between every trip end and the other trip ends in the same cluster. The geometric median trip end is the trip end with the smallest average distance to all other trip ends in the cluster. Note that moved imputed and departure-only stops were not included for this analysis.

5.5.5 Parcel mapping

Manual testing during clustering validation showed that, on occasion, it was required to manually modify a property parcel identifier to that of a neighbouring parcel effectively creating a single, larger, property parcel. Visual inspection showed that this modification could be required for the following two reasons.

1. Cases were observed where businesses operated out of multiple adjacent property parcels that acted as a single unit.
2. Due to GPS error, GPS trip ends located near the boundary of a parcel may be inaccurately located on a neighbouring parcel. This situation usually occurs when parking lots are located between buildings, as reflections of the GPS signal off the nearby buildings causes a multipath GPS error.

In this research, identification of parcel ids that required mapping onto neighbouring parcels was manually performed by examining locations where multiple clusters appeared in close proximity and by judging whether these clusters likely represented visits to the same establishment. In this case the property id of one property parcel was manually adjusted to match its neighbouring parcel, thereby having the two property parcels act as one larger property parcel.

5.6 Removal of ‘Uninteresting’ Trips

After clustering, the next processing step was removing short and uninteresting trips. Greaves & Figliozzi (2008) discovered many short trips in their GPS dataset, some of which were genuine trips while others were clearly cases where a truck was moving within the holding yard. They flagged short trips using an automated process and then performed manual checking to confirm if the trips were genuine or not.

Within-yard trips are defined as uninteresting trips in this research project as they did not correspond to a trip in the typical transportation planning context. Perhaps such trips occurred to move a truck from a
parking spot to a loading dock. Observed short ‘trips’ were also found when trucks were turned off and on without moving. While the reason for these events was not known, it was speculated that this could have been required, among other reasons, to keep the temperature within a refrigerated container within prescribed limits.

The previous clustering step was found to be very useful to identify within-yard trips. In this research, an uninteresting trip was defined as a trip that began and ended within the same destination. Using this simple but effective definition, 16,875 within-destination trips were removed out of a total of 199,085 observed trips.

5.7 Depot Detection

Due to privacy issues the identities of the carriers were not provided for this research. Hence carrier depot locations were not known. As the travel behaviour at depots was expected to be fundamentally different compared with other destinations, separating the observations that likely occurred at depots was considered to be an important preprocessing step before estimating behavioural models of urban freight transportation.

An investigation was conducted to test the ability of different rules to identify destinations that likely corresponded with depot locations. Attributes available in the processed GPS data include the number of trip ends and the dwell times at a destination. In general, it was expected that more trip ends would be located at depots than at other destinations and also that longer trip ends would be observed at depots as trucks were parked between tours. The criteria below were tested for their suitability to identify depots:

1. Fraction of trip ends occurring at the destination compared with the total number of observed trip ends made by the carrier within the study area.
2. The average dwell time of trip ends at the destination.
3. Maximum observed dwell time at the destination.
4. The average of the longest observed trip end dwell times observed on each day.

The following tests were used in this research to judge the effectiveness of the above criteria (and also to select suitable threshold values) to identify likely depot locations.

1. Number of observed depots within the study area
2. Percentage of tours with a multiple trip ends at the same destination
3. Maximum observed tour duration
4. Maximum number of trip ends on a tour
The tours were created as described in Section 5.8. Due to data limitations, depots outside of the study area could not be identified as the dwell times could not be observed.

The above tests were chosen for the following reasons. The number of observed depots was considered important as carriers were expected to have 0, 1, 2 or (maybe) 3 depots in the study area. We considered it unlikely that more than a handful of carriers would have four or more depots within the Greater Toronto and Hamilton area. The tours with repeated destinations test was used because while it was expected that while individual tours may visit the same destination more than once, a large percentage of tours with repeat destinations would be an indication that a depot may have been missed. Finally, large maximum observed tour durations and the maximum number of trip ends on a tour were also indicators of unidentified depots.

Seventeen different combinations of the above criteria were tested. The combinations of criteria and respective values for criterion thresholds are shown in Table 5.7. Preliminary analyses showed that the average trip end duration was found to be a poor predictor for depot identification as if there were many visits to the depot then this reduced the average dwell time even if long overnight trip ends were present. The maximum trip-end duration was also found to be poor predictor for depot identification as multiple destinations were found to have at least one long trip end. The average maximum trip end duration in a day was preferred. After extensive evaluation that included analyzing aggregate tests for the different criteria combinations, including an extensive visual analysis, the following criteria were selected.

1. If 500 destinations or fewer were observed for the carrier, at least 4% of trip-ends observed within the study area had to be located within this destination. This percentage was reduced to 2% if over 500 destinations were observed within the study area as it is difficult to meet a 4% criteria in one destination when that many other destinations are present.

2. The average longest trip end in a day at the destination was greater than six hours (360 minutes).
Table 5.7: Test of Criteria Combinations for Depot Identification

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of trip ends recorded by a carrier within study area made at the destination</td>
<td>5</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>10</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>2,4</td>
<td></td>
</tr>
<tr>
<td>Average dwell time of trip ends at the destination (minutes)</td>
<td>240</td>
<td>240</td>
<td>240</td>
<td>360</td>
<td>360</td>
<td>360</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum observed dwell time at the destination (minutes)</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average of daily longest observed trip end dwell times</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>360</td>
</tr>
</tbody>
</table>

* 2,4: 2% of trip ends if less than 500 total destinations within study area, 4% otherwise
5.8 Tour Creation

Tours were assumed to start and end at destinations that were previously defined as a depot. Destinations were also started or ended at trip ends located outside of the study area. Figure 5.8 shows a schematic showing two sample tours that start and end at the depot location. The tour to the left of the depot stays within the study area and is recorded as a single closed tour. The tour located below the depot leaves the study area and hence is recorded as two separate tours as the trip between the two destinations located outside of the study area is not recorded in the dataset. The two identified tours in this case are identified as open tours as they do not start and end at the same destination.

![Figure 5.8: Schematic Showing Two Processed Tours](image)

5.9 Inferring Land-Use Attributes

The data processing steps described between Sections 5.3 and 5.8 of this dissertation provide an overview of the data analyses that were conducted using a custom C# program. The outputs of this program included a list of trip ends, trips, tours and destinations. The next step of the data processing procedure was to link the GPS data with the secondary data sources described in Section 4.2 to infer
suitable land-use attributes that could be used as covariates for the activity and inter-arrival duration models described later in this dissertation.

In passenger travel, multiple researchers have addressed a lack of trip purpose information by combining the GPS survey with an activity diary completed by respondents either during or after the survey period. For example, this approach was used in Pendyala (2003) and Stopher et al. (2007). Other researchers used data that were originally collected by others and hence they could not ask respondents for trip information. In these cases, trip purposes were inferred from the GPS data and household data collected in the survey. For example, Schönfelder et al. (2002) were provided with participants’ socio-demographic information, including: home address, the location and time commitments of work, school, fixed leisure activities and main shopping locations. These researchers assigned a purpose to different activities by testing if the trip end fell within a threshold distance of the locations. Different threshold values were used for different types of activities. For example a larger threshold was used for shopping activities compared with home and work activities due to the greater flexibility of the destination and parking space choice.

Bohte & Maat (2009) used a hybrid approach where any trip ends located within 100 m of a participant’s home or work locations were assigned to those purposes. Other points of interest in the study area were also defined. A trip purpose was assigned to a known point of interest if the trip end was located within a 50 m threshold of that location. After the survey, participants were asked to enter the activity of trip ends that did not fall within a threshold distance of home, work or a known point of interest.

Srinivasan et al. (2006) used either a “Basic” or an “Enhanced” mode to infer the activity type at the trip destination. In Basic mode, trips were classified to “home”, “work” and “other” activities based on if trip end was within a specified threshold distance to home or work locations, which are known for survey respondents. In Enhanced mode, a multinomial logit model was used to further classify “other” activities into more detailed categories, such as: grocery shopping, other shopping, personal services, social, eating out and serve-passenger trip ends. This model was estimated using the Laredo household travel survey and land-use data. The authors also had property parcel-level land-use information that categorized property parcels into land use categories. Explanatory variables in this activity selection model included the parcel-level land-use attributes, trip and activity durations, trip origin, activity start time and demographic information. More information on this model is available in Srinivasan et al. (2005).
Similar approaches could not be used in this research. Due to privacy concerns no information was provided about the carriers or customers visited by the carriers. Hence activity information could not be obtained directly from this data source. Land-use attributes were instead inferred by linking the GPS data to a zone obtained from the Transportation Tomorrow Survey or to a property parcel obtained from the Teranet database. The destination was considered as a point location having the x-y coordinates of the geometric median GPS point in the cluster. This point location was matched with polygon shape objects from the secondary data sources using a spatial join operation in ArcGIS. Using this operation, a match was made between the destination and the polygon shape if the median point of the cluster was located inside of the polygon shape. Attributes of the zone or property parcels could then be connected with the destination.

Zone-level attributes, such as obtained from the Transportation Tomorrow Survey, are useful but more refined information was desired to estimate disaggregate agent-based transportation models. In this research, addresses of property parcels were matched with addresses of firms listed in the InfoCanada database of companies. This approach gave a strong indication of the types of firms located on a property parcel but provided no information about property parcels that do not contain business establishments.

Destinations were assigned to a property parcel in the Teranet property parcel database using a spatial join performed by the ArcGIS software package. Many, but not all, of the property parcels included the address as a property-parcel attribute. In general it was discovered that addresses were included in the Teranet database unless multiple addresses existed on the same property parcel, such as a parcel containing street numbers 450-490. In this case the address field was left blank.

The address matching was not performed by hand but was performed using a custom software package written using Visual Basic for Applications in Microsoft Access. This software automatically matched addresses if the street number, road names and directions (e.g. North or South) were matched. Due to the different map sources, the software package was programmed to recognize similarities such as “Drive” vs. “Dr”, “Avenue” vs. “Ave”, and “North” vs. “N”. A final step matching the cities was performed manually as the different datasets often used different city names for the same location. For example a street address in Etobicoke and Toronto could be the same due to amalgamation.

Of the 7024 property parcels visited at least once during the study period, 1808 parcels included an address matched to at least one enterprise in the InfoCanada database. There are several reasons why matches could not always be made. These included: 1) the parcel attributes did not include the address
as an attribute, 2) the trip end occurred either on a parcel that corresponded to a road or another public space, and 3) the trip end was not located at a commercial establishment.

It is possible that many firms were located in the same property parcel. For example in a shopping centre possibly over 100 stores can share the same address. No attempt was made to find the visited firm on a destination parcel. Instead, the next step of the analysis involved a series of queries performed using Microsoft Access that aggregated attributes of all establishments located on a property parcel to develop property-parcel level covariates for the models described in the next two chapters.

The number of employees and the total sales volume covariates were found by summing these values for all firms located on the property parcel. A number of firms covariate was also obtained by counting the number of business establishments located on the property parcel. Finally the Standard Industrial Classification (SIC) codes of the firms operating on the parcel were treated by using a dummy variable that was 1 if any firm operating on the property had a SIC code that fits within the category. The industry classification dummy variables created in this research are shown in Table 5.8. The SIC codes referenced here were obtained from the United States Department of Labor (2014).

Tables 5.9 and 5.10 show a schematic of the data processing. In this example there are two properties; property parcel 1 corresponds to an address of 40 Main St. E, Toronto while property parcel 2 corresponds to an address of 42 Main St. E, Toronto. Table 5.9 shows sample entries from the InfoCanada database of companies. After processing, the data are converted to a form similar to that shown in Table 5.10. This is the form of the data that are linked with the GPS information. It can readily be seen from this data source that no information is retained that can be traced back and be used to infer businesses visited by a particular carrier. Hence confidentiality is maintained.

A total of 42,746 trip-ends were recorded inside the study area, did not fall within a 40 m buffer of freeway centrelines and that included attributes from the linked InfoCanada Database. This was reduced to 32,889 observations after the trip ends at depots were removed.

While property parcel information were definitely more detailed than zonal-level data, it must be remembered that these were attributes of the property parcel and not of the visited company, which was uncertain in the case where many establishments operated out of a single property parcel. As can be seen from the example in Tables 5.9 and 5.10, a property parcel containing multiple business establishments would include SIC dummy variables from the SIC classification of all firms operating out of that property parcel. This could lead to many SIC dummy variables being set by firms that were not actually visited by the carrier.
Figure 5.9 shows a histogram of the number of firms listed in the InfoCanada database of companies within each property parcel that was visited by at least one of the carriers during the three-month observation period. As is seen in this figure, 791 (44%) of visited property parcels only contained a single establishment while 1385 (77%) of visited parcels contained fewer than five establishments. 30 parcels, however, contained over 50 establishments while one property parcel contained over 350 establishments. It is highly unlikely that the SIC dummy variables described previously would add property parcel information in these particular destinations. Still, as the vast majority of property parcels only contain at most a small handful of firms, it was felt that this approach still provided a useful way to provide a local land-use context for every destination. Figure 5.9 also shows that a similar distribution is observed when each property parcel is weighted by number of activity duration observations, showing that property parcels with many establishments are not visited much more often compared with parcels with fewer establishments.

5.10 Summary

This chapter describes how the raw GPS data provided by XRS were converted into a diary of trips, trip ends, tours and destinations suitable for estimating the disaggregate travel demand models. This chapter then shows how the destinations were linked with land-use information obtained from the Transportation Tomorrow Survey and also with firms listed in the InfoCanada database of companies. The linked land-use attributes could then be used as covariates when estimating component models of the FRELODE modelling framework. After the data processing steps described in this chapter, the information was ready to input into the first disaggregate transportation model, a model of activity duration.
Table 5.8: Standard Industry Classification Covariates Linked from InfoCanada Database of Companies

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>δ parcel has division A</td>
<td>SIC Division A: Agriculture, Forestry and Fishing</td>
</tr>
<tr>
<td></td>
<td>(two digit SIC code: 01-09)</td>
</tr>
<tr>
<td>δ parcel has division B</td>
<td>SIC Division B: Mining</td>
</tr>
<tr>
<td></td>
<td>(two digit SIC code: 10-14)</td>
</tr>
<tr>
<td>δ parcel has division C</td>
<td>SIC Division C: Construction</td>
</tr>
<tr>
<td></td>
<td>(two digit SIC code: 15-17)</td>
</tr>
<tr>
<td>δ parcel has division D</td>
<td>SIC Division D: Manufacturing</td>
</tr>
<tr>
<td></td>
<td>(two digit SIC code: 20-39)</td>
</tr>
<tr>
<td>δ parcel has division E</td>
<td>SIC Division E: Transportation, Communications, Electric, Gas, and Sanitary Services</td>
</tr>
<tr>
<td></td>
<td>(two digit SIC code: 40-49)</td>
</tr>
<tr>
<td>δ parcel has motor freight transportation</td>
<td>Motor Freight Transportation and Warehousing</td>
</tr>
<tr>
<td></td>
<td>(two digit SIC code: 42)</td>
</tr>
<tr>
<td>δ parcel has motor freight transportation service</td>
<td>Motor Freight Transportation and Warehousing</td>
</tr>
<tr>
<td></td>
<td>(three digit SIC code: 473, 474, 478)</td>
</tr>
<tr>
<td>δ parcel has wholesale</td>
<td>SIC Division F: Wholesale Trade</td>
</tr>
<tr>
<td></td>
<td>(two digit SIC code: 50, 51)</td>
</tr>
<tr>
<td>δ parcel has durable wholesale</td>
<td>Durable Goods Wholesale</td>
</tr>
<tr>
<td></td>
<td>(two digit SIC code: 50)</td>
</tr>
<tr>
<td>δ parcel has non-durable wholesale</td>
<td>Non-Durable Goods Wholesale</td>
</tr>
<tr>
<td></td>
<td>(two digit SIC code: 51)</td>
</tr>
<tr>
<td>δ parcel has retail</td>
<td>Retail</td>
</tr>
<tr>
<td></td>
<td>(two digit SIC codes: 52-54, 56,57,59)</td>
</tr>
<tr>
<td></td>
<td>(three digit SIC code: 554)</td>
</tr>
<tr>
<td>δ parcel has retail dep</td>
<td>Retail: Department Store</td>
</tr>
<tr>
<td></td>
<td>(three digit SIC code: 531)</td>
</tr>
<tr>
<td>δ parcel has retail groc</td>
<td>Retail: Grocery Store</td>
</tr>
<tr>
<td></td>
<td>(three digit SIC code: 541)</td>
</tr>
<tr>
<td>δ parcel has retail small food</td>
<td>Retail: Non-Grocery Food Store</td>
</tr>
<tr>
<td></td>
<td>(three digit SIC code: 542-546, 549)</td>
</tr>
<tr>
<td>δ parcel has food retail</td>
<td>Retail: Food Store</td>
</tr>
<tr>
<td></td>
<td>(two digit SIC code: 54)</td>
</tr>
<tr>
<td>δ parcel has retail gasoline station</td>
<td>Retail: Gasoline Service Station</td>
</tr>
<tr>
<td></td>
<td>(three digit SIC code: 554)</td>
</tr>
<tr>
<td>δ parcel has retail eating establishment</td>
<td>Retail: Eating and Drinking Establishment</td>
</tr>
<tr>
<td></td>
<td>(two digit SIC code: 58)</td>
</tr>
<tr>
<td>δ parcel has non-food retail</td>
<td>Retail: Non-Food, Non-Auto and Non-Eating/Drinking Establishments</td>
</tr>
<tr>
<td></td>
<td>(two digit SIC code: 52, 53, 56, 57, 59)</td>
</tr>
<tr>
<td>δ parcel has division H</td>
<td>SIC Division H: Finance, Insurance and Real Estate</td>
</tr>
<tr>
<td></td>
<td>(two digit SIC code: 60-67)</td>
</tr>
<tr>
<td>δ parcel has division I</td>
<td>SIC Division I: Services</td>
</tr>
<tr>
<td></td>
<td>(two digit SIC code: 70-89)</td>
</tr>
<tr>
<td>δ parcel has division J</td>
<td>SIC Division J: Public Administration</td>
</tr>
<tr>
<td></td>
<td>(two digit SIC code: 91-99)</td>
</tr>
<tr>
<td>δ parcel has a service enterprise</td>
<td>The property parcel contains at least one firm in finance, insurance, real estate, services and public administration.</td>
</tr>
<tr>
<td></td>
<td>(SIC Divisions: H, I and J)</td>
</tr>
</tbody>
</table>
Table 5.9: Example InfoCanada Database Entries

<table>
<thead>
<tr>
<th>Firm Name</th>
<th>Firm Address</th>
<th>SIC</th>
<th>Num Employees</th>
<th>Sales Value (Million $)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firm 1</td>
<td>40 Main St E, Toronto</td>
<td>208434</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>Firm 2</td>
<td>42 Main St E, Toronto</td>
<td>580345</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Firm 3</td>
<td>42 Main St E, Toronto</td>
<td>541104</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>Firm 4</td>
<td>42 Main St E, Toronto</td>
<td>563839</td>
<td>6</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 5.10: Sample of Processed Database Ready to Link with Property Parcel Information

<table>
<thead>
<tr>
<th>Property ID</th>
<th>$\delta$ parcel has a manufacturing enterprise</th>
<th>$\delta$ parcel has retail groc</th>
<th>$\delta$ parcel has non-food retail</th>
<th>$\delta$ parcel has retail eating establishment</th>
<th>Num Employees</th>
<th>Sales Value (Million $)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>36</td>
<td>135</td>
</tr>
</tbody>
</table>

Figure 5.9 Histogram of the Number of Establishments from the InfoCanada Database of Companies on Each Visited Property Parcel and for Every Observed Activity Duration
Chapter 6: Hazard Models of Commercial Activity Duration on Urban Vehicle Tours

6.1 Introduction

Commercial activity duration is considered in this thesis to be the length of time that a commercial vehicle remains stationary to pickup or deliver goods, to provide a service or for other activities such as breaks and vehicle refueling. One component of the FRELODE modelling framework proposed in this dissertation is an activity duration model. Obtaining a better understanding of activity duration is important at a local level to establish parking requirements and to predict emissions due to vehicle idling. Commercial activity duration also has a critical impact on tour formation as it impacts the number of activities that can be visited on a single tour.

Very little research on commercial vehicle activity durations is presented in the literature. Three freight transportation models that include disaggregate commercial activity durations are the Calgary model (Hunt & Stefan, 2007), the Ohio Statewide Modeling Project (Gliebe et al., 2007b) and the regional model for the Chicago metropolitan area (Outwater et al., 2012). In the Calgary model, Monte-Carlo simulation was used to select activity durations based on observed distributions for different combinations of tour purpose (goods delivery, service or other), business type (industrial, wholesale, retail, transportation or service firm) and vehicle class (small four-tire vehicles, single-unit trucks with six tires and multiple-unit trucks with more than six tires). A different approach is used in the Ohio Statewide Modeling Project, which discretizes an activity into five-minute intervals. In each time interval a vehicle has the option of staying at the current activity, ending the current activity and moving to a new activity, or ending an activity and returning to the depot. This decision is simulated using a multinomial logit model. In the regional model for the Chicago metropolitan area, the activity duration was modelled using a multinomial logit model based on the size and commodity of the shipments.

This chapter describes the estimation of different activity duration models from the processed passively-collected GPS data described in the previous two chapters. To the author’s knowledge, these are the first commercial activity duration models estimated using passively-collected GPS data. Please note that as is discussed in Section 1.3.3, GPS data provided by fleet tracking service companies (which was the case in this research) are expected to be more representative of larger carriers that operate between provinces and states. Hence these models can only be considered to be representative of urban...
pickup/delivery tours made by larger carriers. The activity duration travel behavior of small local carriers and small businesses with private fleets cannot be inferred from this data and should be estimated from a different data source.

Hazard models are often used to model the time until the occurrence of an event. Advantages of hazard models include the ability to handle non-normal and non-symmetrical distributions of a residual term (compared with ordinary least squares regression) and the ability to account for censoring, which is where an event occurs when it is not under observation (Cleves, Gould, Gutierrez & Marchenko, 2008, p. 2). In this research, an event is defined as when a vehicle leaves an activity.

This chapter is structured as follows. Section 6.2 presents an overview of hazard analyses and hazard models in order to explain available methods and to provide a rationale for the modelling decisions that were made. Following this background describing hazard models, Section 6.3 then presents an overview of the observed activity durations and the covariates that were tested in all models. Section 6.4 then presents and justifies the two different types of hazard models used for this research and then provides more details describing the mathematical formulation of the selected models. After the models are described, Section 6.5 presents the final estimated models and recommends a model selection for use in FRELODE.

6.2 Overview of Hazard Modelling

6.2.1 Description of hazard rate

Let $T$ be a non-negative random variable denoting the time until an event. The cumulative density function is defined as $F(t) = \Pr(T < t)$, which is the probability of an event having occurred before time $t$. The probability density function, $f(t)$, is the first derivative of the cumulative density function, $f(t) = dF(t)/dt$, and describes the probability of activity durations.

The probability density function can be used to represent and model activity durations. This would be the approach used by, for example, an OLS regression model where the dependent variable is the activity duration. Another approach is to consider the hazard. The hazard is “the (limiting) probability that the event occurs in a given interval, conditional upon the subject having survived to the beginning of that interval, divided by the width of the interval” (Cleves, Gould & Gutierrez, 2004, p. 7). This can be expressed mathematically as shown in equation 6-1.
Note that given a probability density function, a hazard function can be calculated using the formula:

\[ h(t) = \frac{f(t)}{S(t)}, \]

where \( S(t) \) is called the survivor function. The survivor function describes the probability that an event has not occurred by (has survived beyond) time \( t \), and is described by:

\[ S(t) = P(T \geq t) = 1 - F(t) \]

Hence the hazard and probability density are related; choosing either way of analyzing the duration of an event does not change the mathematical description of the problem (Kiefer, 1988).

The cumulative hazard function measures the total amount of risk accumulated until time \( t \), and is defined as:

\[ H(t) = \int_0^t h(u)du \]

The cumulative hazard is related to the survival function as \( H(t) = -\ln(S(t)) \). Note that unlike the cumulative density function, \( F(t) \), the cumulative hazard, \( H(t) \), is not a probability; it can exceed a value of one.

A hazard is a rate (has units of \( 1/\text{time} \)). Assuming a constant hazard rate, the hazard can be interpreted as the expected number of times an event will occur in a unit of time if an event is not terminal (meaning that if an event occurs for an observation that it can occur again for the same observation). Hence a constant hazard of 0.2/minute means that the expected activity duration time would be \( 1/0.2 = 5 \) minutes (assuming non-“fatal” events). The cumulative hazard is the integral of the rates, and shows the expected number of times an event would occur before time \( t \), again assuming non-“fatal” events (Cleves et al, 2008, pp. 13-15). As the cumulative hazard increases, the probability of survival decreases.

### 6.2.2 Proportional and accelerated hazard models

In a duration analysis it is often desired to not only understand the hazard, or density, distribution of the time until an event, but also to model the effect of covariates. A general form of such a model is
presented in equation 6-5. In this equation, \( x_j \) is the vector of the covariates included in the model and \( \beta_x \) is the vector of parameters.

\[
h_j(t) = g(t, x_j, \beta_x)
\]  
(6-5)

This generalized form is often simplified into one of two model structures, the proportional or the accelerated hazard models. In both of these simplifications to equation 6-5 the parameters, \( \beta_x \), can be given a partial-derivative interpretation, similar to that of a linear regression model. This interpretation of parameters is not necessarily true in a general model structure (Kiefer, 1988).

**Proportional hazard models**

In the proportional hazard (PH) model, the hazard function is modelled using a baseline hazard function, \( h_{o(t)} \), which is multiplied by a non-negative function \( g(x_j, \beta_x) \) related to the explanatory variables for an observation.

\[
h(t_j) = h_{o(t)} g(x_j, \beta_x)
\]  
(6-6)

The effects of how the elapsed time influences the hazard is included in the baseline hazard function, \( h_{o(t)} \). The hazard for an individual observation is predicted by multiplying the baseline hazard by a function that models the effects of the coefficients, \( g(x_j, \beta_x) \). It is common practice to normalize the effect of the covariates so that \( g(x_j, \beta_x) = 1 \) for the mean value. Hence \( h_{o(t)} \) is a baseline hazard, corresponding to \( g(x_j, \beta_x) = 1 \).

For mathematical convenience, \( g(x_j, \beta_x) \) is often specified as \( g(x_j, \beta_x) = \exp(x_j \beta_x) \) since the non-negativity requirement of the hazard creates no restrictions on \( \beta_x \), estimation is straightforward and the results can be easily interpreted (Kiefer, 1988). After this assumption the proportional hazards model is written as:

\[
h(t_j) = h_{o(t)} \exp(x_j \beta_x)
\]  
(6-7)

**Accelerated hazard models**

Accelerated failure time (AFT) hazard models use the explanatory variables to rescale time:

\[
\tau_j = \exp(-x_j \beta_x) t_j
\]  
(6-8)

In this equation, \( \tau_j \) (the baseline function) follows a prescribed distribution while \( t_j \) is the time to failure of observation \( j \). \( t_j \) is seen to be the base distribution divided by the factor \( \exp(-x_j \beta_x) \). The selected
mathematical form of the covariates, \( \exp(-x_i \beta_x) \), offers the same advantages as for the proportional hazards model. To interpret the effect of covariates using an accelerated hazards model, if \( \exp(-x_i \beta_x) = 1 \) then time passes at its normal rate, if \( \exp(-x_i \beta_x) > 1 \) then time passes more quickly for the subject (failure will occur sooner), and if \( \exp(-x_i \beta_x) < 1 \) then time passes more slowly for the subject (failure will occur later).

This accelerated failure time hazards model can be rewritten as: \( \ln(t_j) = x_i \beta_x + \ln(\tau_j) \). \( \ln(\tau_j) \) is the residual term with density distribution \( f() \). Different baseline hazard distributions can be assumed for the baseline hazard distribution. For example, if \( \tau_j \) is assumed to follow a log-normal distribution then \( \ln(\tau_j) \) follows a normal distribution (Cleves et al., 2008, p. 232).

6.2.3 Modelling the baseline hazard distribution

Consider the proportional hazards model shown in equation 6-7, which is rewritten here:

\[
  h(t_j) = h_o(t_j) \exp(x_i \beta_x) .
\]

In parametric models, a continuous distribution of the baseline hazard function is assumed and the parameters of this function are estimated alongside the parameters for the covariates, \( \beta_x \). Examples of parametric models in the transportation literature include Kim & Mannering (1997) and Lee & Timmermans (2007). The proportional hazards model restricts the form of the baseline hazard (Kiefer, 1998). According to Cleves et al. (2008, p. 236), not all distributions can be interpreted using a proportional hazards approach. While these models could be written in a form as shown in equation 6-7, the models would not consist of a simple function of the regression coefficients and would not be easy to interpret.

Commonly assumed baseline hazard distributions include: exponential, Weibull, Gompertz, log-normal, log-logistic and gamma distributions. The exponential distribution assumes a constant hazard rate, meaning that the failure rate (after the effect of observed covariates) is independent of time. Such a model can be described as memoryless since the conditional probability of failure is unaffected by the event duration. The Weibull and Gompertz models allow monotonically increasing or decreasing hazard rates with time. The log-normal and log-logistic functions both allow the hazard rate to increase and then decrease in time. The gamma distribution is a flexible hazard function that has the exponential, Weibull and log-normal distributions as special cases (Cleves et al., 2008, p. 269). Of these, only the exponential, Weibull and Gompertz models can conveniently be written using a proportional hazards approach (Cleves et al., 2008, p. 236).
One issue with parametric models is inconsistent estimates of the baseline hazard and covariate parameters if the assumed parametric baseline distribution is incorrect (Bhat, 1996; Meyer, 1990). Hence a reasonable baseline hazard distribution must be used. The rest of this section contains an overview of two other approaches to treat baseline hazard distributions in a proportional hazards model. These are the Cox regression model, which conditions out the baseline hazard during parameter estimation, and non-parametric models, which discretize the baseline hazard into discrete intervals (Han & Hausman, 1990; Meyer, 1990). Both of these estimation techniques allow more flexibility of baseline hazards and hence remove concerns of biased parameter estimates.

Cox regression can be used to estimate the covariate parameters $\beta_x$ in a proportional hazards model without assuming a distribution of the baseline hazard function. In this regression model, all $N$ observations are ordered by their event durations from shortest to longest. The conditional probability of observation $j$ failing, given the surviving observations that could fail at the time that observation $j$ fails, is given by:

\[
P_j = \frac{h_j(t)}{\sum_{i=j}^{N} h_i(t)} = \frac{h_{o(t)} \exp(x_1 \beta_x)}{\sum_{i=1}^{N} h_{o(t)} \exp(x_1 \beta_x)} \tag{6-9}
\]

Note that the base hazard term, $h_{o(t)}$, is fixed for a given value of time, and hence this term cancels out of equation 6-9. The probability that failures occurred in the observed order is given by:

\[
P_{\text{observed order}} = P_1 \times P_2 \times \ldots \times P_n = \frac{\exp(x_1 \beta_x)}{\sum_{i=1}^{N} \exp(x_1 \beta_x)} \times \frac{\exp(x_2 \beta_x)}{\sum_{i=2}^{N} \exp(x_1 \beta_x)} \times \ldots \times \frac{\exp(x_N \beta_x)}{\sum_{i=N}^{N} \exp(x_1 \beta_x)} \tag{6-10}
\]

The vector of parameters, $\beta_x$, is estimated to maximize this equation. While the baseline hazard duration is not analyzed during model estimation, it can be calculated afterwards given the covariate parameters. An example use of Cox regression hazard duration models in the transportation literature is in Yee & Niemeier (2000).

In 1990, both Han & Hausman and Meyer proposed variations of the Cox regression model. Instead of cancelling out the baseline hazard, as is done in the Cox regression model, these two methods discretize the failure times into discrete periods and then estimate a non-parametric distribution of this baseline hazard. A brief summary of the model presented in Meyer (1990) is shown below.

From equation 6-4 and the relationship that $H(t) = -\ln(S(t))$ it can be shown that the probability of survival until time $t + 1$, given that it has survived until time $t$, can be written as a function of the hazard
(shown in equation 6-11). Further assuming a proportional hazards model assumption (shown in equation 6-7), equation 6-11 can be divided into the baseline hazard and the effect of the covariates, as is shown in equation 6-12.

\[
P(T_i \geq t + 1 \mid T_i \geq t) = \exp \left[ - \int_t^{t+1} h_i(u) \, du \right] \tag{6-11}
\]

\[
P(T_i \geq t + 1 \mid T_i \geq t) = \exp \left[ -\exp(\mathbf{x}_i \beta_x) \cdot \int_t^{t+1} h_o(u) \, du \right] \tag{6-12}
\]

Note that covariates can be a function of time, \( \mathbf{x}_i = \mathbf{x}_i(t) \). Assuming that all explanatory variables \( (\mathbf{x}_i) \) are constant between times \( t \) and \( t + 1 \), equation 6-12 can be simplified to:

\[
P(T_i \geq t + 1 \mid T_i \geq t) = \exp\left[ -\exp(\mathbf{x}_i \beta_x + \gamma(t)) \right], \tag{6-13}
\]

where \( \gamma(t) = \ln \left[ \int_t^{t+1} h_o(u) \, du \right] \). \hspace{1cm} \tag{6-14}

When estimating non-parametric equations, the distribution of the baseline hazard function, \( \gamma(t) \), is estimated concurrently with the vector of the covariate parameters, \( \beta_x \).

6.2.4 Censoring

Censoring is when an event occurs while the subject is not under observation. The implications of censoring depend on the context. *Right censoring* occurs when the event has not occurred by the end of an observation period. One advantage of hazard models compared with, say, regression models is that hazard models can accommodate right censoring without difficulty. The entries can be used for estimation using the information that the subject is known to have “survived” until the end of the observation period.

Left censoring occurs when the subject failed before the observation period. This could occur when the observation period starts after the onset of risk of the event occurring. If an event occurs before the observation period occurs then the subject is never observed and hence never enters the dataset. This can occur in the current project if an activity lasts for fewer than five minutes. In such a case no record of the stop is observed. Without any information being recorded, these stops cannot be included in the dataset. Hence the activity durations predicted in this model are conditioned on an activity lasting at least five minutes.
6.2.5 Heterogeneity

Hazard models make an implicit assumption that the survivor function is homogeneous over the sample. Heterogeneity arises when different individuals within a population have potentially different distributions of the dependent variable after controlling for the effects of observable variables. Not accounting for heterogeneity can potentially lead to significantly misleading inferences about the hazard. As is discussed in Kiefer (1988), not including heterogeneity causes a downward bias of baseline hazard estimates with time because the more frail individuals in the population tend to fail earlier leaving a more robust population as time progresses. The population hazard is observed to diminish over time even though the hazard for individuals stays the same.

According to Keifer, the effect of not including heterogeneity on the estimates of covariate parameters is less clear, but some evidence suggests that parameters are biased towards zero if heterogeneity is not included. In general, not including heterogeneity can be considered to have a similar effect to that of missing an important covariate.

Heterogeneity is usually included by specifying a distribution of the unobserved heterogeneity across individuals within a population. As described by Cleves et al. (2008), in the case of heterogeneity the hazard for an individual $j$ given the “frailty” of that individual, $\alpha_j$, can be written as the hazard multiplied by the frailty term:

$$h(t_j|x_j, \alpha_j) = \alpha_j h(t_j|x_j)$$  \hspace{1cm} (6-15)

Heterogeneity across the population is commonly assumed to follow an unobserved distribution. The hazard for the population is then integrated over this distribution, as is shown in equation 6-16.

$$h(t_j|x_j) = \int_0^\infty h(t_j|x_j) g(\alpha) d\alpha$$  \hspace{1cm} (6-16)

In the previous equation, $g(\alpha)$ is the assumed heterogeneity distribution. In theory any continuous distribution can be used if it only supports positive numbers and has a mean of one. There is no statistical guideline for selection of an appropriate distribution. Gamma and inverse-Gaussian distributions are often assumed since they allow for a closed-form solution of the maximum-likelihood function (Meyer, 1990). This is called parametric heterogeneity in this thesis. Examples of transportation research that assumed parametric heterogeneity include Keifer (1988), Han & Hausman (1990) and Hensher & Mannering (1994).
Another approach is to use a non-parametric distribution of the unobserved heterogeneity. This approach was used by Bhat (1996), and makes no assumption of the shape of the heterogeneity distribution. There has been some discussion in the literature on the validity of assuming (prescribing) a particular distribution of the unobserved heterogeneity function. For example, Keifer stated that “estimates are typically more sensitive to specification of the survivor function for each individual or group than to that of the mixing distribution.” Meyer (1990) found the computation of the discrete distribution of unobserved heterogeneity to be difficult and claims that the choice of heterogeneity distribution may be unimportant when the baseline hazard is non-parametrically estimated. Other authors, such as Bhat, have reported improved model fit when a non-parametric distribution was estimated for the hazard distribution.

While in principle it is possible to include heterogeneity in a Cox regression hazards model, it requires the presence of a similar number of integrals to the number of observations in the sample. Given that approximately 6000 samples are available in the present dataset (see Section 6.4.1 for details on the data), including heterogeneity in a Cox regression model can be considered to be intractable for the activity duration model present in this dissertation.

### 6.3 Data and Description of Observed Hazard and Covariates

The GPS data and the data analysis steps used to convert the raw data into a diary of trips, tours and activities were described in Chapters 4 and 5 of this thesis, respectively. Data processing steps included the removal of likely false-positive GPS-recorded trip ends, construction of trips and tours, clustering of GPS trip ends into destinations, identification of destinations that are likely depots, and linking the destinations with disaggregate commercial enterprise information obtained from the InfoCanada database of companies through linking the address of the property parcel with the address of the listed firms in this database. The output of this analysis was aggregated into property-parcel industry classifications and total levels of employment and sales volumes in order to respect privacy concerns of the observed carriers.

It should be noted here that the activity durations observed by the RouteTracker GPS system exceeded actual activity durations, as is discussed in Section 4.1. The observed activity durations were not altered prior to model estimation as no consistent and justifiable alteration procedure was evident. Hence the activity durations should be adjusted downwards on implementation of the full FRELODE framework to account for longer observed activity durations from the data source.
As was mentioned at the end of Chapter 5, 42,746 trip ends were recorded in the database. This was reduced to 32,889 trip ends after observations at likely depot locations were removed. The focus of the models described in this chapter was activity durations during urban commercial vehicle tours. It was expected that activity durations after inter-urban trips would be markedly different than during an urban tour. Hence, only trip ends located within Durham Regional Municipality, York Regional Municipality, Toronto Division, Peel Regional Municipality, Halton Regional Municipality and Hamilton Division (as defined by the Canadian Census Division) were maintained. Trip ends were also removed if they were within a tour that had a duration of over 24 hours. Also, the activity duration of the final trip end in a tour was removed as this was expected to correspond to a depot or to some other kind of base. Finally, due to the data linkage of the GPS data with the InfoCanada database, GPS points were only maintained if they were a member of a destination that was mapped to a property parcel containing at least one enterprise in the InfoCanada database of companies. Of the 7024 property parcels visited at least once during the study period, 1808 parcels included an address that matched with at least one enterprise in the InfoCanada database of companies. Explanations for this comparatively low number are presented in Section 5.9. After the additional criteria, a total of 6,182 GPS trip ends from 1,391 separate destinations were used to estimate the activity duration models.

6.3.1 Observed hazard distribution

Table 6.1 shows a summary of the observed GPS activity durations. The mean duration is seen to be 70 minutes, however there is a large variation as the standard deviation is over 90 minutes. The distribution is seen to be right-skewed as the median (45.37 minutes) is less than the mean. There is also a long right hand tail with a small number of activities having very long durations.

As was discussed in Section 6.2.3, the selection of how the baseline hazard distribution is modelled depends in large part on the observed hazard distribution of the sample. The first part of this research analyzed overall characteristics of the observed GPS commercial activity data to first obtain a high-level overview of the activity durations, and then to plot the observed hazard distribution.

Figures 6.1a, 6.1b and 6.1c show the survival, cumulative hazard and the hazard functions of the observed “time at risk” of commercial activity durations, respectively. As the observation period starts at five minutes an activity is only at risk of ending after this five minute threshold. Hence the time at risk is merely the observed activity duration minus the five minute observation start threshold.
Table 6.1: Summary Statistics of Observed Commercial Activity Durations (All Times are in Minutes)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Observations</td>
<td>6183</td>
</tr>
<tr>
<td>Mean</td>
<td>69.97</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>90.71</td>
</tr>
<tr>
<td>Minimum</td>
<td>5</td>
</tr>
<tr>
<td>Percentile Observations:</td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>8.65</td>
</tr>
<tr>
<td>5%</td>
<td>15.41</td>
</tr>
<tr>
<td>10%</td>
<td>18.11</td>
</tr>
<tr>
<td>25%</td>
<td>26.47</td>
</tr>
<tr>
<td>50%</td>
<td>45.37</td>
</tr>
<tr>
<td>75%</td>
<td>79.48</td>
</tr>
<tr>
<td>90%</td>
<td>129.50</td>
</tr>
<tr>
<td>95%</td>
<td>193.47</td>
</tr>
<tr>
<td>99%</td>
<td>547.11</td>
</tr>
<tr>
<td>Maximum</td>
<td>1167.58</td>
</tr>
</tbody>
</table>

As shown in equation 6-3, the survival function is defined as $S(t) = 1 - F(t) = P(T \geq t)$. The survivor function was estimated using Kaplan-Meier estimation. To use this estimation technique, the data must first be sorted by failure time, $t_j$. For each failure time, the following are recorded:

- $n_j$ – number of activities at risk at time $t_j$ (still surviving at time $t_j$ and that can fail at this time)
- $d_j$ – number of failures (defined as where the vehicle leaves the activity) recorded at time $t_j$

Note that if an activity duration is censored after time $t_j$ this activity is removed from the “at risk” set for subsequent failure times. The survival distribution is estimated after every failure time using the following equation (Cleves et al., 2008, p. 93):

$$S(t) = \prod_{j \mid t_j \leq t} \left(\frac{n_j - d_j}{n_j}\right)$$  \hspace{1cm} (6-17)

Each term in this equation is the ratio of the number of surviving observations divided by the number of observations at risk at time $t$. The product is over all failure times less than or equal to time $t$.

Figure 6.1b presents the cumulative hazard distribution of the observed commercial activity durations. The cumulative hazard was estimated using the Nelson –Aalen estimator, which is shown in equation 6-18. The hazard at each failure time $j$ is estimated as the number of failures divided by the number of observations at risk of failure at that time. The summation is over all failure times less than or equal to time $t$. 


Figure 6.1c presents the hazard distribution of the observed data. The observed hazard represents the derivative of the estimated cumulative hazard, \( \hat{H}(t) = \frac{d(\tilde{H}(t))}{dt} \). As the observed derivatives can fluctuate wildly, the estimated hazards were smoothed using a Gaussian kernel with bandwidth 7 before plotting in Figure 6.1c. This bandwidth was selected as it was found to be the lowest bandwidth (least smoothing) that removed the excessive perturbations using a visual inspection of the produced hazard plots.

Analysing figure 6.1c, the hazard increases until a peak value occurring at a duration of approximately 30 minutes and then decreases until an activity duration of approximately 500 minutes, after which it increases again. Up until the time of 500 minutes the observed hazard has a single mode and varies reasonably steadily with time.
Figure 6.1: Plots of Observed Survival, Cumulative Hazard and Hazard Distributions
6.3.2 Model covariates

Table 6.2 describes the explanatory variables used for modelling. Explanatory variables included interactions between the time of day and an indicator that shows whether the activity was located in a dense area. The rationale was that activity durations are expected to differ by time of day between central regions and more suburban regions that experience different traffic patterns and parking restrictions, and that in general were expected to have more loading infrastructure available. A dense region was classified, as in Newman & Kenworthy (2006), as a region with a gross combined population and employment density of over 3500 per square kilometer. Population and employment density information were gathered from the 2006 Transportation Tomorrow Survey. Time of day variables only considered weekdays. Weekends were treated using separate explanatory variables.

Explanatory variables also included attributes of the trips (distances of the inbound and outbound trips to and from the trip end), and the number of trip ends (activities) in the tour. Activity durations were expected to be longer as inbound and outbound trip distances increased, anticipating larger shipment sizes with increased travel distances. Shorter activity durations were anticipated as the number of activities on a tour increased. This is because having additional stops on a tour would impose a constraint on the duration of time that could be spent at any one activity. Other available covariates were obtained from the linkage between the property parcel address and the address of firms listed in the InfoCanada database of companies, which provided an insight into the type of industries operating on each property parcel. Please see Section 5.9 for more details.
Table 6.2 Description of Explanatory Variables for Activity Duration Models

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_{\text{morning} - \text{dense}}$</td>
<td>=1 if arrival time is between 05:00 &amp; 10:59 and is in a dense region, 0 otherwise</td>
<td>0.023</td>
</tr>
<tr>
<td>$\delta_{\text{noon} - \text{dense}}$</td>
<td>=1 if arrival time is between 11:00 &amp; 12:59 and is in a dense region, 0 otherwise</td>
<td>0.045</td>
</tr>
<tr>
<td>$\delta_{\text{early aft} - \text{dense}}$</td>
<td>=1 if arrival time is between 13:00 &amp; 14:59 and is in a dense region, 0 otherwise</td>
<td>0.057</td>
</tr>
<tr>
<td>$\delta_{\text{peak aft} - \text{dense}}$</td>
<td>=1 if arrival time is between 15:00 &amp; 17:59 and is in a dense region, 0 otherwise</td>
<td>0.063</td>
</tr>
<tr>
<td>$\delta_{\text{evening} - \text{dense}}$</td>
<td>=1 if arrival time is between 18:00 &amp; 20:59 and is in a dense region, 0 otherwise</td>
<td>0.036</td>
</tr>
<tr>
<td>$\delta_{\text{late eve} - \text{dense}}$</td>
<td>=1 if arrival time is between 21:00 &amp; 23:59 and is in a dense region, 0 otherwise</td>
<td>0.019</td>
</tr>
<tr>
<td>$\delta_{\text{weekend} - \text{dense}}$</td>
<td>=1 if arrival time is on a Saturday or Sunday and is in a dense region, 0 otherwise</td>
<td>0.022</td>
</tr>
<tr>
<td>$\delta_{\text{morning} - \text{not dense}}$</td>
<td>=1 if arrival time is between 05:00 &amp; 10:59 and is in a non-dense region, 0 otherwise</td>
<td>0.076</td>
</tr>
<tr>
<td>$\delta_{\text{noon} - \text{not dense}}$</td>
<td>=1 if arrival time is between 11:00 &amp; 12:59 and is in a non-dense region, 0 otherwise</td>
<td>0.103</td>
</tr>
<tr>
<td>$\delta_{\text{early aft} - \text{not dense}}$</td>
<td>=1 if arrival time is between 13:00 &amp; 14:59 and is in a non-dense region, 0 otherwise</td>
<td>0.139</td>
</tr>
<tr>
<td>$\delta_{\text{peak aft} - \text{not dense}}$</td>
<td>=1 if arrival time is between 15:00 &amp; 17:59 and is in a non-dense region, 0 otherwise</td>
<td>0.167</td>
</tr>
<tr>
<td>$\delta_{\text{evening} - \text{not dense}}$</td>
<td>=1 if arrival time is between 18:00 &amp; 20:59 and is in a non-dense region, 0 otherwise</td>
<td>0.096</td>
</tr>
<tr>
<td>$\delta_{\text{late eve} - \text{not dense}}$</td>
<td>=1 if arrival time is between 21:00 &amp; 23:59 and is in a non-dense region, 0 otherwise</td>
<td>0.047</td>
</tr>
<tr>
<td>$\delta_{\text{right} - \text{not dense}}$</td>
<td>=1 if arrival time is between 00:00 &amp; 04:59 and is in a non-dense region, 0 otherwise</td>
<td>0.041</td>
</tr>
<tr>
<td>$\delta_{\text{weekend} - \text{not dense}}$</td>
<td>=1 if arrival time is on Saturday or Sunday and is in a non-dense region, 0 otherwise</td>
<td>0.044</td>
</tr>
<tr>
<td>artripdist_sqrtntripends_sqrt</td>
<td>Square root of the distance travelled on the inbound trip(km)</td>
<td>3.806</td>
</tr>
<tr>
<td>Delaware_tripdist_sqrtntripends_sqrt</td>
<td>Square root of the distance travelled on the outbound trip(km)</td>
<td>3.443</td>
</tr>
<tr>
<td>ntripends_sqrt</td>
<td>Square root of the number of trip ends visited on the tour</td>
<td>2.134</td>
</tr>
<tr>
<td>totalsalesvol-ln$^8$</td>
<td>Natural logarithm of the total value ($) of all firms located in a property parcel</td>
<td>16.69</td>
</tr>
<tr>
<td>$\delta_{\text{parcels has manufacturing}}$</td>
<td>The property parcel contains at least one manufacturing firm. (SIC Division D)</td>
<td>0.314</td>
</tr>
<tr>
<td>$\delta_{\text{parcels has freight service}}$</td>
<td>The property parcel contains at least one firm engaged as a freight forwarder, broker or agent. (two digit SIC code: 47)</td>
<td>0.060</td>
</tr>
<tr>
<td>$\delta_{\text{parcels has durable wholesale}}$</td>
<td>The property parcel contains at least one durable goods wholesale establishment. (two digit SIC code: 50)</td>
<td>0.274</td>
</tr>
<tr>
<td>$\delta_{\text{parcels has non-dur. wholesale}}$</td>
<td>The property parcel contains at least one non-durable goods wholesale establishment. (two digit SIC code: 51)</td>
<td>0.246</td>
</tr>
<tr>
<td>$\delta_{\text{parcels has retail, dep. store}}$</td>
<td>The property parcel contains at least one department store. (four digit SIC code: 5311)</td>
<td>0.141</td>
</tr>
<tr>
<td>$\delta_{\text{parcels has retail, groc}}$</td>
<td>The property parcel contains at least one grocery store. (four digit SIC code: 5411)</td>
<td>0.188</td>
</tr>
<tr>
<td>$\delta_{\text{parcels has retail, gas}}$</td>
<td>The property parcel contains at least one gasoline service station. (four digit SIC code: 5541)</td>
<td>0.060</td>
</tr>
<tr>
<td>$\delta_{\text{parcels has retail, eating}}$</td>
<td>The property parcel contains at least one eating/drinking establishment. (two digit SIC code: 58)</td>
<td>0.256</td>
</tr>
<tr>
<td>$\delta_{\text{parcels has retail, gen}}$</td>
<td>The property parcel contains at least one retail establishment from a category that is not included above.</td>
<td>0.441</td>
</tr>
<tr>
<td>$\delta_{\text{parcels has a service enterprise}}$</td>
<td>The property parcel contains at least one firm in finance, insurance, real estate, services and public administration. (SIC Divisions H, I and J)</td>
<td>0.421</td>
</tr>
</tbody>
</table>

$^8$ Note the number of observations in the dataset was reduced from 6182 to 5637 if the variable “totalsalesvol-ln” was included in a model many firms in the database were recorded as having an annual sales value of $0.$
6.4 Methods:

6.4.1 Overview of modelling efforts used in this research

In hazard modelling, one of the first decisions is the selection of how to model the baseline hazard distribution. Three different approaches were considered in this research, parametric hazard modelling, the semi-parametric Cox regression model, and non-parametric estimation of the baseline hazard distribution.

Advantages of parametric models are that they are efficient, meaning that more precise parameter estimates can be obtained from a parametric model for a given dataset. For example, a continuous hazard distribution is estimated for a parametric model while information is lost in the non-parametric models of Han & Hausman (1990) and Meyer (1990) by discretizing the baseline hazard distribution into intervals. Cox regression is likewise less efficient as only the order of failures is used in model estimation; failure times are not considered. The primary disadvantage of parametric hazard models is that there is a risk of biased estimates of covariates and of the baseline hazard function if the form of the assumed baseline hazard function is a poor representation of the hazard distribution (Bhat, 1996; Meyer, 1990). Parametric models are also unable to accurately model cases with complex hazard distributions, including bimodal distributions.

In the Cox regression model, the baseline hazard is removed by performing a series of comparisons at each observed failure. Advantages of this model include that it can be used to estimate the effects of covariates regardless of the distribution of the baseline hazard. Using this model the baseline hazard distribution is not estimated alongside covariate parameters, but it can be estimated afterwards. Disadvantages of this model include that it is less efficient than parametric models, tied failure times cause difficulties due to the reliance on ordering observations by failure time, and that including heterogeneity requires solving a similar number of integrals as the number of observations (Han & Hausman, 1990).

The final model considered in this research was non-parametric estimation of the baseline hazard distribution. Non-parametric models maintain the benefits of the Cox regression model in that they can handle a wide variety of baseline hazard distributions. Unlike the Cox regression model, however, these methods can also account for unobserved heterogeneity, either using a parametric or a non-parametric heterogeneity distribution. Bhat (1996) recommends at least testing a non-parametric baseline hazards model to confirm that estimated parameters are unbiased. Disadvantages of non-parametric baseline
hazard distributions include that they are less efficient than parametric models with a suitable baseline hazard distribution due to the discretized baseline hazard that considers the hazard in each interval to be constant.

Figure 6.1c shows the hazard distribution of the observed commercial activity durations. In this figure, the hazard is seen to rise to a peak at an activity duration of around 30 minutes and then falls until an activity duration of approximately 500 minutes. As can be seen in Table 6.1, this 500 minute threshold is slightly below the 99th percentile of the observed activity durations. The baseline hazard up until this 500 minute threshold is consistent with the log-normal, log-logistic and gamma parametric hazard distributions.

Given that the covariates for this research are based from passively-collected GPS data combined with secondary data sources, the available covariates are not expected to provide a complete description of the activity duration. Hence having a model that can consider unobserved heterogeneity was considered to be of large importance for this research. Cox regression was therefore not considered for this research due to the difficulty of modelling unobserved heterogeneity using this modelling approach.

Given that the shape of the observed hazard distribution could reasonably be matched by commonly-used baseline hazard distributions and given other advantages of parametric hazard models (namely the ability to consider unobserved heterogeneity and the increased efficiency of these models), the decision was made to initially test parametric models. It was also decided to follow the advice of Bhat (1996) and also test a non-parametric approach in order to confirm whether the parameters estimated using the parametric hazard distribution model were biased. The non-parametric model of Meyer (1990) was selected for this research. Advantages of this model compared with previous research included: 1) parameters are easily interpreted; 2) probabilities of surviving each period are constrained to lie between 0 and 1; and 3) unobserved heterogeneity can be accommodated.

It was decided to assume a parametric distribution of the unobserved heterogeneity for both the parametric and non-parametric hazard models. The major advantage of assuming either the gamma or inverse Gaussian probability distributions is that these distributions allow a closed-form log-likelihood equation, simplifying model estimation. The suitability of parametric and non-parametric distributions of the unobserved heterogeneity has been debated in the literature. In this research it was decided to use the findings of Kiefer (1988) as a guide, who suggested that “the specification of the mixing distribution is not so important as long as the individual distribution is correctly specified (or not too badly misspecified).” Meyer (1990) also reported that computation of the discrete unobserved
heterogeneity distribution is often difficult in practice and that the selection of heterogeneity distribution may be unimportant if a non-parametric baseline is used.

Two different model specifications were estimated for both the parametric and the non-parametric hazard models. The first specification was estimated without including covariates from the linked InfoCanada database. Covariates used in this model included the time of day interacted with dense vs. non-dense regions, and trip and tour attributes. These were all directly observed from the travel diaries created by the processed GPS data. This specification could be appropriate if firm information is not available. The second specification included InfoCanada establishment information and was intended for use in FRELODE. The other purpose for estimating two separate model specifications, one with and one without the parcel-level commercial establishment information, was to separately test the benefit of adding these covariates and to determine if they improved the fit of the estimated models.

6.4.2 Parametric models

Modelling the baseline hazard

Due to the very long tail in the observed activity duration distribution it was decided to censor the data at a duration of 300 minutes (shown in the vertical lines in Figure 6.1). This was important as it was felt that not censoring these outliers would impact the quality of the estimated models. Figure 6.1c shows that the hazard could not have been accurately modelled using any of the previously discussed baseline hazard distributions if activity durations of over 500 minutes were maintained as the hazard became unstable and increased (causing a bi-modal distribution) beyond this point. Table 6.1 shows that the selected 300 minute censoring threshold was well over the 95th percentile of observed activity durations. The data of activities with a duration of longer than 300 minutes were not thrown away (which would bias the data) but instead were recorded as having survived up until a time of 300 minutes, with no information being available beyond this point.

The event time in a parametric hazards model must start at 0 at the onset of risk. Since the minimum activity duration maintained in the processed GPS data was five minutes, the onset of risk occurs at this time. All activity durations were therefore reduced by five minutes to reflect the period in which the activities were not at risk of ending. When using a parametric hazards model in an operational model, this five-minute time interval would need to be added to calculated activity durations.
Initial hazard models were estimated with no covariates in order to select a suitable baseline hazard distribution for the parametric models. Table 6.3 shows the log-likelihood of models without covariates that were estimated using five different baseline hazard distributions: Exponential, Weibull, log-normal, log-logistic and gamma distributions. The separate columns in this table show the results estimated for the uncensored and censored datasets.

### Table 6.3: Log-Likelihood of no Covariate Parametric Models with Different Baseline Hazards

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Log Likelihood</th>
<th>Log Likelihood (censored at $t = 300$ minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>-9071.76</td>
<td>-8819.79</td>
</tr>
<tr>
<td>Weibull</td>
<td>-9071.07</td>
<td>-8751.37</td>
</tr>
<tr>
<td>Log-normal</td>
<td>-8361.42</td>
<td>-8292.46</td>
</tr>
<tr>
<td>Log-logistic</td>
<td><strong>-8336.24</strong></td>
<td><strong>-8284.40</strong></td>
</tr>
<tr>
<td>Gamma</td>
<td>-8358.55</td>
<td>-8292.40</td>
</tr>
</tbody>
</table>

As is seen in Table 6.3, the log-logistic distribution is seen to provide the best fit for the observed activity durations as measured using the model log-likelihood, both for the censored and the uncensored data. The log-logistic model distribution is consistent with the observed hazard distribution (Figure 6.1c) which rises to a peak at a duration of around 30 minutes and then drops up until a duration of 500 minutes.

The log-logistic baseline hazard distribution was selected based on the results of these initial models that did not use any covariates. As the log-logistic baseline distribution does not have a convenient proportional hazards representation (Cleves et al., 2008), an accelerated failure time parametric hazards model was used. The detailed formulation of this model is presented next.

**Accelerated failure time parametric hazard models**

As was described in Section 6.2.2, accelerated hazard models use the covariates to rescale a baseline time distribution. The mathematical formulation of this model is $\tau_j = \exp(-x_j\beta_x) t_j$, where $\tau_j$ is the baseline hazard function (which follows a prescribed distribution) and $t_j$ is the time to failure of observation $j$. This accelerated failure time hazards model can be rewritten as: $\ln(t_j) = x_j\beta_x + \ln(\tau_j)$.

In a log-logistic distribution, $\tau_j$ is distributed log-logistic with parameters $(\beta_0, \gamma)$. For this model, the previous equation can be rewritten as $\ln(t_j) = x_j\beta_x + u_j$, where $u_j$ follows a logistic distribution with mean 0 and standard deviation $\pi\gamma/\sqrt{3}$.
Hazard models are usually estimated using maximum likelihood techniques. In the presence of right censoring the likelihood function to be maximized is shown in equation 6-19. In this equation, \( N \) is the number of observations, \( t_j \) is the time that the observation \( j \) either failed or was censored, \( \beta_x \) are the parameters for the vector of explanatory variables \( x_j \) and \( \Theta \) are the parameters related to the assumed base hazard distribution. \( d_j \) is a dummy variable that is set to 1 if observation \( j \) is right censored and is set to 0 otherwise.

\[
L = \prod_{j=1}^{N} \left\{ h(t_j \mid x_j \beta_x, \Theta) \right\}^{d_j} S(t_j \mid x_j \beta_x, \Theta) \tag{6-19}
\]

In the case of the log-logistic distribution, the survival and the probability density functions are defined in equations 6-20 and 6-21 respectively (adapted from StataCorp, 2013). In these equations, \( \lambda = \exp(-x_j \beta_x) \). The log-logistic distribution is defined by one baseline parameter, \( \gamma \).

\[
S(t) = \left\{ 1 + (\lambda t)^{1/\gamma} \right\}^{-1} \tag{6-20}
\]

\[
h(t) = \frac{1}{\gamma} \left( \frac{1}{\gamma - 1} \right) \left( 1 + (\lambda t)^{1/\gamma} \right) \tag{6-21}
\]

Equations 6-19 through 6-21 describe the maximum likelihood function in the case of no unobserved heterogeneity. When including heterogeneity, the survivor and hazard functions must be modified, as shown in equations 6-22 and 6-23 respectively. \( \alpha_j \) is an unobserved observation specific effect.

\[
S(t_j \mid x_j, \alpha_j) = \left\{ S(t_j \mid x_j) \right\}^{\alpha_j} \tag{6-22}
\]

\[
h(t_j \mid x_j, \alpha_j) = \alpha_j h(t_j \mid x_j) \tag{6-23}
\]

The unconditional survivor function is obtained by integrating the conditional survivor function over the probability distribution of the unobserved heterogeneity \( \alpha_j \), as is shown in equation 6-24. For identification purposes, the distribution of the unobserved heterogeneity is assumed to have a mean of 1 and variance \( \theta \).

\[
S_{\theta}(t_j \mid x_j) = \int_0^\infty \left\{ S(t_j \mid x_j) \right\}^{\alpha_j} g(\alpha_j) d\alpha_j \tag{6-24}
\]

It can be shown that when \( \alpha \) is assumed to follow a gamma distribution, that equation 6-24 can be simplified to equation 6-25 (Cleves et al., p. 303). The parameter \( \theta \) of the unobserved heterogeneity distribution is estimated alongside the parameters of the baseline hazard and the covariates. Note that closed form solutions of the survivor and hazard functions are also available if an inverse Gaussian
distribution is assumed for the unobserved heterogeneity (Meyer, 1990). Both the gamma and inverse-Gaussian distributions were tested. A slightly better model fit was observed using the gamma distribution, and hence this distribution was used throughout the remainder of this research. The log-likelihood function including heterogeneity is shown in equation 6-26. In this equation, \( \Gamma(\theta) \) is the gamma function.

\[
S_\theta(t_j|x_j) = \left[1 - \theta \ln(S(t_j|x_j))\right]^{-1/\theta} \quad (6-25)
\]

\[
L = \prod_{j=1}^{N} \left(\Gamma(\theta) h(t_j|x_j)\right)^{d_j} \left[1 - \theta \ln(S(t_j|x_j))\right]^{1/\theta} \quad (6-26)
\]

To employ an accelerated parametric hazards model in a microsimulation environment, recall that \( \ln(t_j) = x_j \beta_x + \ln(t_j) \). Hence the first step is to draw a value from the log-logistic base activity duration distribution, \( \tau_j \) (whose parameters are estimated alongside parameters for the covariate variables). This activity duration is then added to the effect of the covariate variables, \( x_j \beta_x \) in order to calculate the logarithm of the activity duration, \( \ln(t_j) \).

### 6.4.3 Non-parametric model

As discussed in section 6.4.2, the non-parametric model of Meyer (1990) was also tested in this research. The log-likelihood function of this model, including a gamma distributed unobserved heterogeneity, is shown in equation 6-27. Each observation \( j = 1, ..., N \) enters a state at time \( t = 0 \) and is observed for \( k_j \) time periods, at which point it either leaves the state (completed duration) or is right-censored. \( c_j \) is a censoring index, which takes a value of 1 for a completed event and 0 otherwise. \( \nu \) is the variance of the gamma mixing distribution (the mean is always normalized to 1). \( x_j \) is the vector of explanatory variables. The vector of parameters, \( \beta_x \), contains the parameters defining the effects of the covariates, \( x_j \), while \( \gamma_t \) is the estimated baseline hazard in time period \( t \).

\[
LL = \sum_{j=1}^{N} \ln \left\{ \frac{1 + \nu \sum_{t=0}^{k_j-1} \exp\left[\gamma_t + x_j^{t} \cdot \beta_x\right]}{1 + \nu \sum_{t=0}^{k_j} \exp[\gamma_t + x_j^{t} \cdot \beta_x]} \right\}^{-1/\nu} - c_j \left\{ 1 + \nu \sum_{t=0}^{k_j} \exp[\gamma_t + x_j^{t} \cdot \beta_x] \right\}^{-1/\nu} \quad (6-27)
\]

The `pgmhaz8` routine in Stata (Jenkins, 2004) was used to estimate this non-parametric model. Fourteen discrete time intervals were selected. Care was taken to ensure that a good number of activities ended in every duration interval to form a good estimate of the base hazard parameter for every interval.
Table 6.4: Number of “Failures” in Each Time Period of Non-Parametric Model

<table>
<thead>
<tr>
<th>Time Interval</th>
<th>Time</th>
<th>Number of Durations in Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>t &lt; 15 min</td>
<td>244</td>
</tr>
<tr>
<td>2</td>
<td>15 min &lt; t &lt; 25 min</td>
<td>1036</td>
</tr>
<tr>
<td>3</td>
<td>25 min &lt; t &lt; 35 min</td>
<td>805</td>
</tr>
<tr>
<td>4</td>
<td>35 min &lt; t &lt; 45 min</td>
<td>665</td>
</tr>
<tr>
<td>5</td>
<td>45 min &lt; t &lt; 55 min</td>
<td>518</td>
</tr>
<tr>
<td>6</td>
<td>55 min &lt; t &lt; 65 min</td>
<td>426</td>
</tr>
<tr>
<td>7</td>
<td>65 min &lt; t &lt; 80 min</td>
<td>502</td>
</tr>
<tr>
<td>8</td>
<td>80 min &lt; t &lt; 95 min</td>
<td>382</td>
</tr>
<tr>
<td>9</td>
<td>95 min &lt; t &lt; 125 min</td>
<td>417</td>
</tr>
<tr>
<td>10</td>
<td>125 min &lt; t &lt; 155 min</td>
<td>209</td>
</tr>
<tr>
<td>11</td>
<td>155 min &lt; t &lt; 185 min</td>
<td>105</td>
</tr>
<tr>
<td>12</td>
<td>185 min &lt; t &lt; 245 min</td>
<td>133</td>
</tr>
<tr>
<td>13</td>
<td>245 min &lt; t &lt; 305 min</td>
<td>74</td>
</tr>
<tr>
<td>14</td>
<td>t &gt; 305 min</td>
<td>122</td>
</tr>
</tbody>
</table>

Non-parametric hazards models are employed as follows. The hazard function for a point can be found by multiplying the base distribution by the effect of the covariate variables, \( g(x_i) = \exp(-x_i\beta_x) \). The hazard distribution can be converted into a probability distribution by remembering that \( h(t) = \frac{f(t)}{s(t)} \) and that \( H(t) = \int_0^t h(u)du = -\ln(s(t)) \). The base activity duration can be simulated either by a purely random or a systematic draw (Halton, Metropolis Hasting, etc.) from the probability distribution.

### 6.5 Results/Empirical models

Table 6.5 shows the estimated parameters for the parametric and non-parametric survival models, the significance of each parameter and the adjusted \( \rho^2 \) goodness of fit measure, defined in equation 6-28.

\[
\text{adjusted } \rho^2 = 1 - \frac{\text{log likelihood at convergence} - \text{no. of parameters}}{\text{log likelihood of base distribution only model}} \quad (6-28)
\]

The use of an accelerated parametric model was mandated by the selection of a log-logistic baseline hazard distribution and the proportional hazards model specification of Meyer (1990) was used for the non-parametric model. One downside of this combination of models is that parameter estimates were not directly comparable between the estimated parametric and non-parametric models due to the different model specifications. For the parametric accelerated failure time model, a positive coefficient
means that this parameter increases the activity duration. For the proportional non-parametric hazards model a positive coefficient means that the parameter increases the hazard, thereby reducing the activity duration. To avoid confusion, ↑ or ↓ arrows have been placed next to a parameter value, with a ↑ arrow meaning that increasing the value of this explanatory variable increases the activity duration and ↓ decreases the activity duration. The significance level is reported as ***, ** and * for 99%, 95% and 85% significance respectively.

Effects of the time of day and of density were similar between the models that did not include the establishment explanatory variables, specification 1, and the models that included these explanatory variables, specification 2. Activities occurring in the morning were longer both in dense and in non-dense regions. In dense regions especially, activities occurring in the early afternoon, peak afternoon and evening were shorter than those at other times. Outside of the morning period, the activity duration was consistently longer in non-dense regions than in dense regions.

In all model specifications, an increase in the length of inbound and outbound trips was found to be positively correlated with increased activity durations. This effect is intuitive as firms are likely to travel further to deliver larger deliveries. An increase in the number of activities in a tour is negatively correlated with the time spent at each activity visited during that tour. This effect is also expected as increasing the number of activities increases the time pressure to complete the tour by the end of the working day.
<table>
<thead>
<tr>
<th>Specification 1: No InfoCanada Covariates</th>
<th>Specification 2: Includes InfoCanada Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parametric</strong></td>
<td><strong>Non-Parametric</strong></td>
</tr>
<tr>
<td>Adj. $\rho$</td>
<td>Adj. $\rho$</td>
</tr>
<tr>
<td>0.061</td>
<td>0.0872</td>
</tr>
<tr>
<td>0.0327</td>
<td>0.0465</td>
</tr>
<tr>
<td><strong>No. Covariate Vars</strong></td>
<td><strong>No. Covariate Vars</strong></td>
</tr>
<tr>
<td>11</td>
<td>24</td>
</tr>
<tr>
<td><strong>No. Base Vars</strong></td>
<td><strong>No. Base Vars</strong></td>
</tr>
<tr>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td><strong>LL. Full Model</strong></td>
<td><strong>LL. Full Model</strong></td>
</tr>
<tr>
<td>-7768.14</td>
<td>-6951.11</td>
</tr>
<tr>
<td>-13098.41</td>
<td>-12900.2</td>
</tr>
<tr>
<td><strong>LL. No covariates</strong></td>
<td><strong>LL. No covariates</strong></td>
</tr>
<tr>
<td>-8284.4</td>
<td>-7644.45</td>
</tr>
<tr>
<td>-13568.56</td>
<td>-13568.56</td>
</tr>
<tr>
<td><strong>Adj. $\rho$</strong></td>
<td><strong>Adj. $\rho$</strong></td>
</tr>
<tr>
<td>0.061</td>
<td>0.0872</td>
</tr>
<tr>
<td>0.0327</td>
<td>0.0465</td>
</tr>
<tr>
<td><strong>Confidence levels:</strong> *** 99%, ** 95%, * 85%, ○ &lt; 90%**</td>
<td><strong>Confidence levels:</strong> *** 99%, ** 95%, * 85%, ○ &lt; 90%**</td>
</tr>
</tbody>
</table>

**Table 6.5 Model Parameters and Significance for Parametric and Non-Parametric Models**, With and Without Linked Establishment Data.
In the specification 2 models, all parameters had the expected effect on activity duration. Activity durations were positively correlated with the sales volume of the firms in the property parcel, which was expected to be correlated with shipment sizes. Longer activities were expected at retailers receiving large shipments, such as department and grocery stores, and wholesalers due to anticipated longer loading and unloading times. Activities at freight service providers, such as shipping agents and freight forwarders, were also longer. Shorter activities were found at manufacturing facilities; this was surprising but was consistent throughout all models. Activities at gasoline service stations, eating establishments and service industries were shorter, as expected.

Table 6.5 shows that all parameters had the same sign between the parametric and the non-parametric models. This was important as it suggested that the parametric hazard distribution did not produce biased parameter estimates, at least to the degree of changing the sign of any estimated parameter. Direct parameter estimates could not be compared for specification 1 (which did not include InfoCanada covariates) as the forward stepping model estimation procedure suggested the inclusion of different covariates for the parametric and non-parametric models. The same covariates were maintained in the parametric and non-parametric models for specification 2 (which included property-level attributes obtained from linking with the InfoCanada database). A comparison of the ratio between the parametric and non-parametric models showed a similar ratio, with the single exception of the covariate $\delta_{\text{night - not dense}}$. The unobserved heterogeneity was not significant for the parametric models (essentially zero in both Specifications 1 and 2) while it was significant (at 99%) for the two non-parametric models.

An analysis of the adequacy of the estimated parametric models can be performed using a Cox-Snell residual plot, which compares the Cox-Snell residuals against an estimate of the cumulative hazard of the Cox-Snell residuals. If the model fits the data then this plot forms a straight line from the origin with a slope of 1 (Klein and Moeschberger, 2003).

Figure 6.2 compares Cox-Snell residuals for the parametric models with: no covariates, specification 1 (no InfoCanada covariates) and specification 2 (includes the InfoCanada covariates). As is seen in this figure, the Cox-Snell residuals of all of these estimated parametric models follow the desired 45° reference line, and hence adequately fit the data. Starting with a model estimated with no covariates, the model fit is shown to be improved with the addition of the directly observed GPS data from specification 1 and then further improved with the additional InfoCanada covariates. Hence, the improved fit of the Cox-Snell residual plot provides further justification for including these covariates in the activity duration model.
Due to its simpler model structure, continuous response, lack of evidence of incorrect signs due to biased parameter estimates and the adequate fit to the data as measured by the Cox-Snell residuals, the parametric model with the InfoCanada covariates is recommended for implementation in FRELODE.

6.6 Summary

Two hazard models of activity duration in urban commercial vehicle tours were estimated and are described in this chapter. These models were estimated entirely using passively-collected GPS data, described in Chapter 4, that were then processed into a travel diary of trips, trips ends (activities) and tours. These GPS data were connected with additional publicly available data sources using the techniques described in Section 5.9 of this thesis. To the author’s knowledge, this is the first activity duration model estimated for commercial vehicle activities from GPS data. Data from 6,182 trip ends were used for model estimation.

This chapter starts with an overview of hazard duration (survival) models, including the definition of the hazard, survival and cumulative hazard functions. Proportional and accelerated hazard models are then presented. One of the first decisions that had to be made when selecting a particular hazard model was the selection of the baseline hazard function. Section 6.2.3 provides an overview of three approaches used in the literature to model the baseline hazard: 1) a parametric hazards model, in which the baseline hazard is modelled using a continuous function; 2) the semi-parametric Cox-regression model; and 3) non-parametric estimation of the baseline hazard distribution. Sections 6.2.4 and 6.2.5 then describe the concepts of censoring and the importance of including unobserved heterogeneity in the hazard model formulation.

Section 6.3 provides an overview of the distribution of the observed activity durations and also of the survival, cumulative hazard and hazard distributions. The hazard was seen to rise to a peak value at a duration of approximately 30 minutes and then decrease until an activity duration of approximately 500 minutes before it increased again. After observing the baseline hazard, it was figured that a parametric hazards model using either a log-normal, log-logistic or gamma baseline hazard could provide an appropriate estimation of the hazard duration. This choice was confirmed by estimating parametric hazards models with no covariates. The log-logistic baseline hazard distribution had the largest log-likelihood, and hence this distribution was selected to model the baseline hazard. Note that the log-logistic baseline hazard duration is only available using the accelerated parametric model. A non-parametric model was also estimated following the advice of Bhat (1996) who advocated at least testing a non-parametric model as there is the potential for bias in parametric model estimation. A gamma
distribution for unobserved heterogeneity was used for both the parametric and non-parametric models as this distribution allowed for a convenient closed-form solution of the unobserved heterogeneity and hence simplified model estimation.

Two specifications were estimated for each of the parametric and non-parametric models. The first specification only used data obtained directly from the processed GPS travel diary while the second specification also contained linked establishment information from the InfoCanada database of companies. These two model specifications were both estimated because: 1) it may not always be feasible to link the GPS data with an InfoCanada/InfoUSA database and hence it was useful to estimate a model without this information, and 2) to test how adding the additional covariates improved the model fit.

Results showed that:

1) Passively-collected GPS data can be used to estimate activity duration models when linked with other data sources providing appropriate explanatory variables.

2) The addition of industry classifications of firms operating out of the property parcels using the linkage with the InfoCanada database of companies improved the model fit, including higher goodness of fit measures and improved observed distribution of the Cox-Snell residuals.

3) A comparison of the estimated parameters of the models estimated with the linked InfoCanada database between the parametric and non-parametric models found little evidence of incorrect signs due to parameter bias in the parametric model.

4) Cox-Snell residual plots showed that the parametric model provided an adequate fit for the data.

For the above reasons, the parametric hazards model that includes the additional property-parcel land-use information provided by the InfoCanada database is recommended for implementation within FRELODE.
Figure 6.2  a) Cox-Snell Residuals for the Parametric Hazards Models.  b) Zoomed-In View of the Same Residual Plot
Chapter 7: Multilevel Modelling of Commercial Vehicle Inter-Arrival Duration

7.1 Introduction

One large advantage with using passively-collected GPS data compared with traditional travel survey data collection techniques is their ability to monitor travel duration over extended time periods. These longitudinal data are particularly valuable to analyze and to estimate the models that describe the shipment scheduling component of FRELODE as the inter-arrival durations can be directly observed from the processed GPS data. *Inter-arrival duration* is defined in this research as the number of days between visits to the same individual destination by the same vehicle fleet operator.

Travel behaviour models developed using survey data collected on a single “typical weekday” make the implicit assumption that travel behaviour is consistent between days. A small number of multiple week studies have been conducted for passenger travel that examined the consistency of travel behaviour between days (Huff & Hanson, 1986; Muthyalagari et al., 2001; Pas, 1987; Pas & Sundar, 1995; Pendyala, 2003; Schlich, Schönfelder, Hanson & Axhausen, 2004; Stopher & Zhang, 2011). These studies have found that, in general, while travel patterns are repeated between days, travel behaviour is not sufficiently repetitious that a single day is an adequate characterization of a person’s routine travel.

Pas (1987) separated the residual term of a linear regression model into components representing between-person and within-person variability. Based on data collected over a five weekday period in the Reading Activity Diary Survey, the within-person variability accounted for almost 50% of the residual. This research was extended in Pas & Sundar (1995), who found that 38% of the total variability in the number of trips and 42% of the variability of the travel time in a three-day travel diary collected in Seattle in 1989 was due to intrapersonal variability. Even higher levels of variability were found in two longitudinal GPS surveys. A 3-5 day sample of a survey from Lexington KY, and 10 weekdays of data from a 28-day survey in Adelaide, Australia, found that over 60% of the total variability could be attributed to intrapersonal variability (Muthyalagari et al., 2001; Pendyala, 2003; Stopher et al., 2007). Hence while the between person variability is important, within-person variability should also be considered. Huff & Hanson (1986) used a different approach. Using data from a 35-day survey conducted in Uppsala they created two measures, the first measuring the repeatability of a person’s travel patterns and the second creating a similarity measure comparing a person’s travel behaviour
between two days. While the repeatability measure found high levels of repetition, the similarity measures found a low average similarity between days for all individuals. Their results showed that while selected behaviours are repetitive, the activities over a single day did not compare well to other days. These conclusions were also found by Schlich et al. (2004), who analyzed the repetition and similarity of leisure activities from the MobiDrive survey, a six-week survey conducted in the cities of Karlsruhe and Halle/Salle in the spring and autumn of 1999.

For commercial vehicle travel, pickups, deliveries and service trips are often not made on a daily basis but are planned over a longer time period. A common assumption among inter-city freight travel demand models is that firms optimize shipment sizes in order to minimize the combination of transportation and inventory costs. These models generally use an extension of the Economic Order Quantity (EOQ) model (e.g. de Jong & Ben-Akiva, 2007; Liedtke, 2012; Combes, 2012). While these models have a solid theoretical underpinning from logistics operations, the required data are difficult to acquire for all companies in a region. For example, the simplified shipment frequency model of de Jong & Ben-Akiva (which does not include transportation costs) still requires parameters such as: unit cost per order, value of transported goods (per tonne), storage costs per unit per year and the discount rate per year. These quantities are not generally available to transportation planners and modellers.

For the intra-urban commercial vehicle movements analyzed in this paper, however, the arguments for EOQ-style models are less compelling because: 1) lower shipment costs (shorter trips) reduce the penalty for non-optimal shipment patterns, 2) increasing use of pull logistics (Simchi-Levi et al., 2003) means that firms must quickly respond to customer demand and cannot wait for regularly-scheduled restocking shipments, 3) firms may choose to operate using a consistent routine (e.g. one shipment per week) to facilitate supply chain administration instead of completely optimizing their supply chain, and 4) EOQ models cannot be used to estimate the frequency of service trips, which are more common in urban areas (Hunt & Stefan, 2007). These arguments may also apply to some longer-distance inter-urban shipments, depending on context.

Shipment frequencies have been rarely modelled in urban freight transportation models. One example is the Tokyo model (Wisetjindawat et al., 2007), which treats shipment frequency as a simple commodity specific function of shipment size, which in turn is a function of shipment distance. Other urban commercial vehicle travel demand models, such as the tour-based models of Hunt & Stefan (2007) and Gliebe et al. (2007b) do not explicitly consider the time between repeated visits to the same destination.
In one the few longitudinal GPS studies of commercial vehicle transportation of which the author is aware, Logendran & Peterson (2006) did a passive GPS study that covered 23 trucks over a two year period with at least a two-week monitoring period per vehicle. One of their study questions was if LTL trucks run similar routes every day. The authors did a manual examination of the routes operated by the 23 recorded vehicles and found that 34.8% of the trucks in the study followed a daily route while 65.2% did not.

In this review it is shown that shipment scheduling is usually estimated using models based on Economic Order Quantity theory, which calculates optimal shipment frequencies and delivery lot sizes considering transportation and inventory costs. Besides the lack of high-quality data required to estimate and forecast shipment frequencies using EOQ style models, these models also cannot account for within-firm variation in shipment schedules (similar to the variation that has been observed in passenger travel) due to the cross-sectional data sources used for model estimation. Please see Section 1.3.1 of this dissertation for an overview of recent surveys of commercial vehicle activities. Variation in travel patterns to individual destinations is expected to be especially pronounced for intra-urban commercial vehicle due to the prevalence of service trips, comparatively low penalties for non-optimal routing and scheduling and the use of pull logistics.

In passenger travel the expectation underlying many models estimated using cross-sectional surveys is that individuals and households have similar travel patterns between weekdays. Variation in travel between days has been demonstrated in multiple studies, as is previously shown in this chapter. The expectation of EOQ-style shipment scheduling models is not consistency between days but instead that destinations are visited using consistent schedules that vary between destinations as influenced by transportation and inventory costs. It is expected that longer duration data sources would provide additional benefits to better understand and quantify this variation.

This research estimates disaggregate statistical models of inter-arrival duration to assess the effect of time, location and land-use attributes using the observed inter-arrival duration intervals recorded by GPS-equipped tracking units. Estimating an inter-arrival duration model instead of a shipment frequency model allows the model to account for within-firm variation in travel patterns. To the authors’ knowledge, this is the first study that models commercial vehicle inter-arrival durations directly using longitudinal GPS data. The intention is to incorporate these inter-arrival duration models to model delivery scheduling in FRELODE.
This chapter is structured as follows: Section 7.2 presents initial analyses that demonstrated a wide variation in observed inter-arrival duration patterns between destinations and then presents the need to classify or cluster destinations into market segments so that the estimated models are representative of individual firms. This section then presents a brief literature review of market segmentation techniques and then describes the market segmentation approach selected for use in this research. Following the breakdown of destinations into different segments, Section 7.3 presents an overview of the observed inter-arrival duration distributions for the entire dataset and also for each segment. The covariates that were tested in all inter-arrival duration models are shown in Section 7.4. Section 7.5 presents the random effects models inter-arrival duration models estimated in this research. As estimation of these models does not simply involve maximizing closed-form likelihood functions, Section 7.6 overviews of different estimation techniques that have been used to estimate random-effects models in other studies. Section 7.7 presents the outcomes of initial models that were estimated without covariates. After an overview of the goodness of fit measures used to compare different model specifications, shown in Section 7.8, Section 7.9 then compares the estimated models before recommending an inter-arrival duration model to be incorporated in FRELODE.

### 7.2 Market Segmentation

The idea of using market segmentation to group destinations was original conceived by the desire to separate a relatively small number of frequently-visited destinations that were seen to dominate the overall dataset. As can be seen in Figure 7.1, the mode of the distribution of the observed inter-arrival durations of the entire dataset is on day 0, meaning that most visits to a destination by a carrier occurred on the same day as the previous visit to that destination. The next most commonly observed inter-arrival duration was 1 day, meaning that successive visits to the same destination occurred on subsequent days. Well over 50% of observed inter-arrival durations either occurred on the same day or on the following day. The probability of visits on a given day decreased as the number of days increased, with slight exceptions for 7 and 14 days, which marks one and two-week intervals respectively. This observed distribution of inter-arrival durations was not intuitive for many businesses.
Figure 7.1. Observed Inter-Arrival Durations for the Entire Dataset

Figure 7.2 shows the number of visits to each destination by a single carrier. The x-axis of this figure is the destination index, where the destinations are sorted in ascending order. The destination with index 1 has the fewest number of visits while the destination with index 2809 has the highest number of visits. The y-axis is the total number of visits to the destination. The horizontal lines in this figure correspond to 1 and 3 visits per week. These correspond to the 78th and the 93.5th percentile of the destinations respectively. Hence, this figure shows that most destinations are visited infrequently but that a small number of destinations are visited regularly. It can be inferred by comparing this graph to the large number of same-day and next-day visits observed in Figure 7.1 that a small number of destinations are the source of the majority of observed same-day and next-day visits, which account for over 50% of observed inter-arrival duration intervals.
Due to small number of destinations that were seen to be dominating the observed inter-arrival durations in the overall dataset, some type of market segmentation is clearly seen to be necessary to separate the analyses of these frequently-visited destinations. In addition, an additional category was added into the market segmentation that further grouped the destinations in terms of whether the destinations were visited by a carrier using fixed schedules or not. The research question here was to study if it is possible to use GPS data to observe whether firms use either push-based or pull-based supply-chain systems based on their longitudinal travel patterns. Estimating separate inter-arrival duration models in each market segment also allows the possibility of exploring different explanatory variables for each segment.

The rest of this section describing the market segmentation approach used in this research is structured as follows. Section 7.2.1 provides a brief literature review of market segmentation approaches that have been used by other authors. Section 7.2.2 then describes the market segmentation procedure that was selected for this research.
7.2.1 Literature review on market segmentation approaches

The following studies examined market segmentation in a context of obtaining higher-accuracy mode choice models.

In Ben-Akiva & Lerman (1985, p. 194), survey respondents were segmented into different categories using threshold values of a single variable. A likelihood ratio test was used to compare the segmented model with a more restricted single model for the entire population. Badoe & Miller (1998) extended this approach by sequentially segmenting the population using different attributes to create finer and finer homogenous groups. An existing group was segmented into two finer groups in every step. The result was a hierarchical (tree) structure, with leaves of the tree as mutually exclusive, exhaustive segments.

More recent research has used a clustering approach that simultaneously groups individuals into segments using multiple variables (e.g. Outwater et al., 2003; Beirão & Cabral, 2008). Individuals were classified into deterministic clusters using their socioeconomic attributes.

The above studies are examples of “exogenous segmentation”, in which the segmentation and modelling within each segment are conducted sequentially with no feedback from the modelling process to segment formation. Since there is no feedback to the segmentation process, there is no guarantee that the a priori selected segments maximize homogeneity within segments and heterogeneity between segments (Ishaq, 2011).

Bhat (1997) and Greene & Hensher (2003) used endogenous segmentation, in which the segmentation is performed alongside model estimation. These two papers describe a Latent Class Modelling approach that simultaneously estimates a multinomial logit model to segment individuals and a separate, lower-level, mode-choice model for each segment. Bhat, Frusti, Zhao, Schönfelder & Axhausen (2004) also used an endogenous segmentation approach for an inter-arrival duration model for shopping trips that classified shoppers into erratic and regular shoppers. Ishaq (2011) developed a “Flexible Model Structure” (FMS) that simultaneously grouped individuals into segments, assigned different model structures to each segment and estimated a discrete choice model for mode choice for every model structure.
7.2.2 Adopted segmentation approach for inter arrival duration modelling

The reason for using market segmentation was to separate outlying frequently-visited destinations, and also to test if the GPS data could be used to observe push vs. pull logistics by observing patterns in the longitudinal data history. Hence the exogenous market segmentation approach described in Ben-Akiva & Lerman was selected for this research. Two measures were used to segment the destinations: 1) the average number of visits per week by the carrier, and 2) regularity/stochasticity of shipments to the destination as measured by the coefficient of variation of the observed inter-arrival durations to the destination.

To segment the destinations in a manner that best improved the likelihood of the entire dataset, tests were run as follows for different combinations of the threshold values of the average number of visits per week and the regularity of inter-arrival durations as measured by the coefficient of variance. If one threshold is created for each of these two measures then a total of four possible segments are obtained, creating a “low” and “high” segment for each measure. Tested threshold values included an average of 1, 2 or 3 visits per week and coefficient of variance thresholds of 0.6, 0.7, 0.8, 0.9 and 1.0. An ordered probit model was estimated for every segment for each combination of thresholds. The log-likelihood of each model was then summed to produce a combined log-likelihood for the entire dataset. Ordered discrete models are appropriate for inter-arrival duration modelling as there is a natural scale in the selection of a discrete duration outcome. For these tests the observed arrival distributions were discretized into the same intervals shown in Section 7.3 with the single exception that an additional interval was created for inter-arrival durations of 28 days or greater. In the four segment tests the optimum coefficient of variance thresholds were found to be different for the low and high-frequency destinations. Since there is no justification for this difference it was decided to use a three segment approach. The three market segments selected for this research are as follows:

1. **Frequently-visited destinations** were defined as destinations where the carrier makes an average of at least 2 visits per week. This segment accounts for a small number of observed destinations but accounts for the majority of the observed inter-arrival episodes.

2. **Regularly-scheduled destinations** are destinations that operate using fixed shipment schedules (e.g. weekly deliveries). The criteria used to identify these destinations was that they were visited by the carrier at an average rate of less than two visits per week and the inter-arrival duration coefficient of variance for these visits was 0.7 or below.
3. **Unscheduled destinations** are destinations where the shipments are not made on regular schedules. For example, a shipper may respond to customer orders by sending deliveries at the next available opportunity instead of waiting for the next scheduled shipment. These destinations were identified where the average visiting rate by the carrier was less than two visits per week, and where the inter-arrival duration coefficient of variance for these visits was greater than 0.7.

Figure 7.4 (presented in the next section) shows the observed distribution of inter-arrival durations for the full dataset and also for each of the observed segments. This figure shows that each of the different market segments have a different distribution of observed inter-arrival durations compared with the full dataset and compared with each other. This figure shows the importance of using some kind of market segmentation as the behaviour predicted by analyzing the full dataset is markedly different than that of any of the segments.

When using the model for forecasting applications, neither the average weekly visits by a carrier to a destination, or the coefficient of variance of these visits are directly observable attributes of a firm. Hence a discrete choice model is required to classify destinations into one of three segments. A first model separates the destinations into the segments. A separate inter-arrival duration model is then estimated for each segment.

A preliminary exogenous market segmentation model was developed using the GPS data. A traditional multinomial logit model was used, which was estimated using Stata. One observation was made for every destination that was visited by a carrier at least twice (forming one inter-arrival duration observation). A total of 2,018 destinations were observed. The results from this market segmentation model are presented in Section 7.9.3.

### 7.3 Selection of Inter-Arrival Duration Intervals for this Study

The processed GPS data record the start time of every trip end. The precise duration between repeated arrivals to the same destination is known within an accuracy of seconds. Hence it is possible to use a continuous model, such as ordinary least-squares regression or hazard models, for this purpose. However, a discrete choice approach is preferable to a continuous approach for the following reasons:

1. The inter-arrival duration distribution is not smooth due to large differences in daytime/nighttime stop frequency (for example an inter-arrival duration of 12 hours is far less common than 24 hours (see Figure 7.3) and due to weekly patterns (see Figure 7.4). Using
modelling approaches with continuous dependent variables would hide patterns (e.g. weekly deliveries) in the data due to assumed residual distributions in either OLS regression or hazard statistical models.

2. The exact inter-arrival duration is of less interest than the number of days between repeat visits. The time of day of a delivery is more appropriately determined using vehicle-routing logic.

Intervals were selected so that they would be applicable for all three segments and for the full dataset. After some experimentation, the selected intervals were: 1) same day, 2) following day, 3) 2-3 days apart, 4) 4-6 days apart, 5) 7-14 days apart, and 6) 15+ days apart. Figure 7.4 shows the histogram of the inter-arrival duration for each segment and the full dataset.

![Figure 7.3: Hour-by-Hour Breakdown of Observed Inter-Arrival Durations](image)

### 7.4 Explanatory Variables

The explanatory variables shown in Table 7.1 were tested in all of the models presented in this chapter. If a variable from Table 7.1 has not been included in a model, it means that the parameters for this variable were not statistically significant or that including the variable did not improve the model goodness of fit. The number of employees and sales revenue were not considered since these fields were missing for many firms in the InfoCanada database. Visits on Mondays and Tuesdays were not
included in this table as these are the reference levels for the discrete choice models. These days were found to have similar travel behaviour characteristics during model estimation.

<table>
<thead>
<tr>
<th>Segment</th>
<th>Total Number of Trip Ends</th>
<th>Number of Trip Ends in Segment</th>
<th>Total Number of Destinations</th>
<th>Number of Destinations in Segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Segmentation</td>
<td>27203 (100%)</td>
<td>17887 (65.8%)</td>
<td>2018 (100%)</td>
<td>278 (13.8%)</td>
</tr>
<tr>
<td>Segment 1:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequently</td>
<td>27203 (100%)</td>
<td>17887 (65.8%)</td>
<td>2018 (100%)</td>
<td>278 (13.8%)</td>
</tr>
<tr>
<td>Visited</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Destinations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segment 2:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regularly-Scheduled Destinations</td>
<td>3097 (11.4%)</td>
<td>3097 (11.4%)</td>
<td>592 (29.3%)</td>
<td>592 (29.3%)</td>
</tr>
<tr>
<td>Destinations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segment 3:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unscheduled</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Destinations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 7.4: Histograms of Inter-Arrival Durations for the Full Dataset and Individual Segments
Table 7.1: Explanatory Variables for Inter-Arrival Duration Models (Industry Classifications are from United States Department of Labor (2014))

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st visit is on a Wednesday</td>
<td>The trip end preceding the inter-arrival duration of interest arrived on a Wednesday</td>
<td>0.180</td>
</tr>
<tr>
<td>1st visit is on a Thursday</td>
<td>The trip end preceding the inter-arrival duration of interest arrived on a Thursday</td>
<td>0.193</td>
</tr>
<tr>
<td>1st visit is on a Friday</td>
<td>The trip end preceding the inter-arrival duration of interest arrived on a Friday</td>
<td>0.181</td>
</tr>
<tr>
<td>1st visit is on a Weekend</td>
<td>The trip end preceding the inter-arrival duration of interest arrived on a Saturday or Sunday</td>
<td>0.097</td>
</tr>
<tr>
<td>ln(distance to freeway)</td>
<td>Natural logarithm of the straight-line distance from the destination to the nearest freeway in metres</td>
<td>6.635</td>
</tr>
<tr>
<td>ln(distance to depot)</td>
<td>Natural logarithm of the straight-line distance from the destination to the nearest identified depot of the carrier in km. If no depot was identified for the carrier a distance of 200 km was assumed.</td>
<td>3.043</td>
</tr>
<tr>
<td>parcel has food retail</td>
<td>The property parcel contains at least one food retail establishment. (SIC Major Group 54)</td>
<td>0.272</td>
</tr>
<tr>
<td>parcel has non-food retail</td>
<td>The property parcel contains at least one store non-food retail establishment. (SIC Major Groups 52, 53, 56, 57, 59)</td>
<td>0.368</td>
</tr>
<tr>
<td>parcel has durable wholesaler</td>
<td>The property parcel contains at least one durable goods wholesale establishment. (SIC Major Group 50)</td>
<td>0.257</td>
</tr>
<tr>
<td>parcel has non-durable wholesaler</td>
<td>The property parcel contains at least one non-durable goods wholesale establishment. (SIC Major Group 51)</td>
<td>0.227</td>
</tr>
<tr>
<td>parcel has manufacturing</td>
<td>The property parcel contains at least one manufacturing firm. (SIC Division D)</td>
<td>0.283</td>
</tr>
<tr>
<td>parcel has a service enterprise</td>
<td>The property parcel contains at least one firm in finance, insurance, real estate, services and public administration. (SIC Divisions H, I and J)</td>
<td>0.424</td>
</tr>
<tr>
<td>parcel has freight transportation provider</td>
<td>The property parcel contains at least one firm that is a motor freight carrier or warehouse (SIC Major Group 42)</td>
<td>0.170</td>
</tr>
<tr>
<td>parcel has freight service provider</td>
<td>The property parcel contains at least one firm engaged as a freight forwarder, broker or agent. (SIC major Group 47)</td>
<td>0.061</td>
</tr>
</tbody>
</table>
7.5 Description of Selected Discrete Choice Models of Inter-Arrival Duration

7.5.1 Random intercept models

As can be seen in Figure 7.4, the number of visits differed between destinations, even after the market segmentation. Hence the mean inter-arrival duration (which is inversely proportional to the number of visits to a destination) also differed between destinations. For example, a destination with an average of two visits per week (the upper threshold of the regularly-scheduled and unscheduled market segments) would have a mean inter-arrival duration of 3.5 days. Meanwhile the mean inter-arrival duration would increase to 14 days if an average of one shipment was made every two weeks to a destination. Therefore, there is a correlation between the inter-arrival durations of observations from the same cluster. This situation is common for longitudinal studies (Skrondal & Rabe-Hesketh, 2004, p. 80).

Multilevel random effects models were selected for this application. Multilevel models are appropriate when data have a natural grouping and hence observations are not independent from one another. According to Goulias (2003), multilevel models better reflect the context of decision making, can reduce bias from model misspecification and remove the assumption that observations used to estimate the model parameters are independent by representing correlations in errors within groups. As is discussed in Rabe-Hesketh, Skrondal & Pickles (2005), “random effects models are useful for modeling panel data or grouped cross-sectional data where the responses for the same person or group cannot be assumed to be independent after conditioning on exogenous variables.”

According to Skrondal & Rabe-Hesketh (2004, p. 50), there are two types of random effects, random intercepts and random coefficients. Random intercepts allow different intercepts for different clusters, and can represent unobserved heterogeneity within units on the overall response. Random coefficients, meanwhile, represent unobserved heterogeneity on how explanatory variables affect model response. The main objective of using multilevel models was to account for the fact that different destinations have a different mean inter-arrival duration. This effect is accounted for using a random intercept models. Since random intercept models accounted for the desired effects, and also since they are the simplest form of random effects models, only this type of model was considered for this research. Random coefficient models were not tested in this research but could be included in future work.
This research estimates two-level random-intercept models. The inter-arrival durations are the level 1 observations. The level 2 grouping is the destination since shipping patterns are expected to vary between firms but inter-arrival durations to the same destination are not independent.

The remainder of this section presents the concept of random intercept models using the example of a random intercept linear regression model. Sections 7.5.2 and 7.5.3 present the random intercept models that were used in this research, the random intercept ordered probit model and the random intercept multinomial logit model, respectively.

Ordinary least-squares regression provides a good example to explain the concept of multilevel modelling. A single-level OLS regression model assuming a single explanatory variable is

$$y_i = a + bx_i + \epsilon_i$$  \hspace{1cm} (7-1)

where \(y_i\) is the observed value for observation \(i\), \(a\) is the intercept, \(x_i\) is the observed independent variable for observation \(i\), and \(b\) is the slope of the line. \(\epsilon_i\) is the error term, which is assumed to be normally distributed with mean 0 and variance \(\sigma_\epsilon^2\).

A two-level OLS model for observations \(i\) in groups \(j\), is described as:

$$y_{ij} = a + bx_{ij} + u_j + \epsilon_{ij}$$  \hspace{1cm} (7-2)

Variables \(a\) and \(b\) are fixed parameters. \(u_j\) and \(\epsilon_{ij}\) are random parameters in the model and are assumed to follow a normal distribution with mean 0 and variance \(\sigma_u^2\) and \(\sigma_\epsilon^2\), respectively. \(\sigma_u^2\) is the variance between clusters while \(\sigma_\epsilon^2\) is the variance between observations within the same cluster. Note that \(u_j\) only has a subscript \(j\). Hence this term is fixed for all observations occurring within a group. The \(u_j\) term allows the intercept value to vary for every destination. A variance component model is used in this research to describe a random intercept model with no explanatory variables. Note that the intra-class correlation coefficient, \(\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\epsilon^2}\), is a useful measure that represents the proportion of the between cluster variance, \(\sigma_u^2\), to the total residual variance, \(\sigma_u^2 + \sigma_\epsilon^2\) (Skrondal & Rabe-Hesketh, 2004, p. 51).

As can be seen here, a random intercept multilevel model has the advantage of being able to accommodate groupings in clusters while only requiring estimation of a single additional parameter (at least for an ordinary least squares regression model). Disadvantages of this model structure include: (Skrondal & Rabe-Hesketh, 2004, p. 51 and 83).
the distribution of \( u_j \) and \( \epsilon_{ij} \) are assumed in the model specification

- random intercept models assume independence between the random term, \( u_j \), and the covariates.

- when normal distributions are assumed for \( u_j \) and \( \epsilon_{ij} \) the maximum likelihood estimator cannot be expressed using a closed form solution.

### 7.5.2 Random intercept ordered probit model

Two multilevel models were tested, a random intercept ordered probit model and a random intercept multinomial logit model. The ordered probit approach was considered since there is an inherent scale in the inter-arrival duration intervals.

According to Train (2009, pp. 159-163), in an ordered model the inter-arrival durations are represented using an unobserved variable, \( U \), where higher values of \( U \) represent a longer duration. \( U \) is decomposed into observed and unobserved components, \( U = \beta'x + \epsilon \), where \( \beta' \) is the transpose of the parameter vector and \( x \) is the vector of explanatory variables. The unobserved factors, \( \epsilon \), are assumed to be random. \( U \) is distributed around \( \beta'x \) with the shape of the distribution following the distribution of \( \epsilon \). In an ordered probit model, \( \epsilon \) is distributed standard normal.

Cutoff points, \( k_1, \ldots, k_{t-1} \), where \( t \) is the number of intervals in the response, are estimated alongside the vector of parameters, \( \beta \). The probability of a particular inter-arrival duration is the probability of variable \( U \) falling between different cutoff points. For example, the probability of the next trip end occurring on the same day is given in equation 7-3. In this equation, \( \Phi \) is the standard cumulative normal distribution.

\[
\text{Prob(same day)} = \text{Prob}(U < k_1) = \text{Prob}(\epsilon < k_1 - \beta'x) = \Phi(k_1 - \beta'x)
\] (7-3)

The probability of the next trip end occurring on the following day is given in equation 7-4. Similar equations can be constructed for the other intervals.

\[
\text{Prob(following day)} = \text{Prob}(k_1 < U < k_2) = \text{Prob}(\epsilon < k_2 - \beta'x) - \text{Prob}(\epsilon < k_1 - \beta'x) = \Phi(k_2 - \beta'x) - \Phi(k_1 - \beta'x)
\] (7-4)

In a two-level random intercept ordered probit model, the probability of the next visit occurring on the following day is modified to that shown in equation 7-5 below. In this equation, subscripts \( i \) and \( j \) are added, where the subscript \( i \) refer to individual observation \( i \) while a subscript of \( j \) refers to a group \( j \).
The \( u_j \) term is a second random term that is normally distributed with mean 0 and variance \( \sigma^2_x \), where \( \sigma^2_x \) is the random component of the model. This variable is fixed for a given destination (note the subscript of \( j \) instead of \( ij \) that is used in the other terms). \( \sigma^2_x \) is estimated alongside other model parameters, adding one additional parameter that is estimated compared with a single level ordered probit model.

\[
Prob(\text{following day}) = \text{Prob}(k_1 < U_{ij} < k_2) = \text{Prob}(\epsilon < k_2 - \beta'x_{ij} - u_j) - \text{Prob}(\epsilon < k_1 - \beta'x_{ij} - u_j) = \Phi(k_2 - (\beta'x + u_j)) - \Phi(k_1 - (\beta'x + u_j))
\]

(7-5)

Note that a positive coefficient in the parameter \( \beta_{ij} \) increases the probability of a lower threshold being selected. Hence positive coefficients of \( \beta_{ij} \) refer to shorter inter-arrival durations while negative values of \( \beta_{ij} \) refer to longer inter-arrival durations.

### 7.5.3 Random intercept multinomial logit model

In a discrete choice model, the probability of decision maker \( i \) selecting alternative \( k \) is the probability that the utility of alternative \( k, U_{ik} \), is the greater than that of all other alternatives for decision maker \( i \). Since the researcher does not observe a decision maker’s utility, the utility is split into two components, an observable utility, \( V_{ik} \), and the unobserved aspects of utility, \( \epsilon_{ik} \), where

\[
U_{ik} = V_{ik} + \epsilon_{ik}.
\]

In a logit model, the unobserved utility, \( \epsilon_{ik} \), is assumed follow a Type I extreme value distribution, which is independently, identically distributed (iid). If the estimated model is linear in parameters, then the utility is defined as: \( U_{ik} = \beta'x_{ik} + \epsilon_{ik} \). This leads to the familiar expression for logit models, which shows the probability of decision maker \( i \) selecting outcome \( y \) as alternative \( s \).

\[
Prob(y_i = s) = \frac{\exp(\beta^{(s)}x_i)}{1 + \sum_{k=1}^{t-1} \exp(\beta^{(k)}x_i)}
\]

(7-6)

In this equation, \( x_{ik} \) is the vector of observable variables for individual \( i \) and alternative \( k \) while \( \beta \) is the vector of estimated parameters. The response variable has \( t \) intervals. A set of \( t - 1 \) equations is estimated, which contrast each interval with the reference interval (denoted here with the subscript \( t \)).

The probability of being in the reference interval, \( t \), is obtained by subtraction:

\[
Prob(y_i = t) = 1 - \sum_{k=1}^{t-1} \pi_i^{(k)}.
\]

In the random intercept multinomial logit model, a random intercept term is added to the utility function to represent that different clusters can have different alternative specific constants. The utility function of individual \( i \) in group \( j \) of alternative \( k \) is modified to \( U_{ijk} = V_{ijk} + u_{ij}^{(s)} + \epsilon_{ijk} \). In this
function, \( \epsilon_{ijk} \) remains as a Type I extreme value distribution while the cluster level random effect, \( u^{(s)}_{jk} \), is assumed to be distributed \( N(0, \sigma^2_{u(s)}) \). Here, \( \sigma^2_{u(s)} \) is a symmetric matrix of length \( t-1 \), where the diagonal element \((s,s)\) corresponds to the variance in interval \( s \), and the non-diagonal element \((r,s)\) corresponds to the covariance between intervals \( r \) and \( s \), \( \text{cov}(u^{(s)}_j, u^{(r)}_j) = \sigma_{u(r,s)}, s \neq r \). The symmetric variance matrix \( \sigma^2_{u(s)} \) is estimated alongside the fixed parameters in the model. As six discrete categories are proposed for this model (see Section 7.3), the matrix \( \sigma^2_{u(s)} \) contains 15 random effects parameters that require estimation.

The assumed Type 1 extreme value distribution of the \( \epsilon_{ijk} \) term in the utility function is assumed to be independently and identically distributed (iid). This assumption means that the distribution is assumed to be the same for all the discrete time intervals, and also that there is no correlation between time intervals. It is acknowledged in this research that ordinal responses are not entirely suited to being modelled using a multinomial logit model as there may be correlation between the different intervals. For example, there may be a different relationship between interval 2 (1-day) and interval 3 (2-3 days) than between interval 2 (1 day) and interval 6 (14+ days). Hence while this model considers non-independence of observations at the same destinations, it does assume independence of the time intervals. Given that the objective of using this model is an exploration of estimating a discrete choice model that includes effects of covariates on destinations selecting inter-arrival durations within a specific interval (instead of either increasing or decreasing a single underlying latent variable) this model was considered reasonable for this research.

### 7.6 Estimation of Random Intercept Models

Two different estimation techniques were tested: maximum likelihood and Markov Chain Monte Carlo (MCMC) Bayesian estimation. This section provides a brief overview of the two estimation techniques tested in this research and then provides an overview of comparative advantages and disadvantages of the two techniques for the estimation of random effects models.

#### 7.6.1 Maximum likelihood estimation

In general, the likelihood function of a random effects model does not have a closed form (Skrondal & Rabe-Hesketh, 2004, p. 159). Instead, parameter estimation is usually based on the marginal likelihood (the likelihood of the data conditioned on the unobserved (latent) random-effects variables) integrated over the latent variable distribution. For a two-level model with a single random intercept the likelihood for a given cluster, \( j \) is given by equation 7-7.
In this equation, \( f_{ij}^{(2)}(\theta) \) is the likelihood of level 2 cluster \( j \) given the vector of all parameters, \( \theta \). The random-intercept parameter is assumed to follow a normal distribution, hence \( g(u_j; 0, \sigma^2) \) is the normal distribution of random intercept, \( u_j \), with mean 0 and variance \( \sigma^2 \). \( f_{ij}^{(1)}(\theta | u_j) \) is the conditional likelihood of unit \( i \) within cluster \( j \) given the parameters, \( \theta \), and the random effect, \( u_j \) (Rabe-Hesketh, Skrondal & Pickles, 2005). Given the assumed normal distribution of the latent random intercept, the integral shown in equation 7-7 does not have a closed form solution but instead must be integrated numerically during model estimation.

The GLLAMM software package (Rabe-Hesketh, Skrondal, & Pickles, 2005) was selected to estimate the random-intercept models due to the extensive literature available describing the estimation techniques used in this software (for example Rabe-Hesketh & Skrondal, 2012; Skrondal & Rabe-Hesketh, 2004) and its inclusion in the Stata software package, which was used for the activity duration models described in Chapter 6 of this dissertation.

The integration over the distribution of random effects parameter, \( g(u_j; 0, \sigma^2) \), given parameters \( \theta \) of the random effects model can be calculated using a variety of different techniques. When using the GLLAMM software, adaptive quadrature integration was selected to integrate over the latent variables on the advice of Rabe-Hesketh, Skrondal & Pickles (2005). These authors found that the use of adaptive quadrature provides unbiased estimates, even for large cluster sizes and intraclass correlations (intra-class correlation is defined in Section 7.7 of this dissertation). Quadrature methods approximate the area under a definite integral by using a weighted sum of the integrand at specified points.

The GLLAMM program uses Newton-Raphson to maximize the marginal likelihood with respect to \( \theta \). The Hessian is calculated using numerical differentiation. An iterative solution is required. Every iteration includes: 1) using the Newton-Raphson procedure to update the parameter values, and then 2) updating the location and weights of the quadrature integration points of the random effects parameter (find the random-effects parameter values and their weights that approximate the integral in equation 7-7 (Rabe-Hesketh, Skrondal & Pickles 2005).

### 7.6.2 Bayesian estimation

Let \( \theta \) represent the parameters in a model. In the Bayesian estimation technique, two distributions of \( \theta \) are defined representing the probability that the parameters take a specific value: 1) the prior
distribution, \( k(\theta) \), which represents a researcher’s idea of the distribution of \( \theta \) before any data is collected, and 2) the posterior distribution, \( K(\theta|Y) \), which represents the updated distribution of \( \theta \) given the data, \( Y = \{y_1, ..., y_N\} \) (where \( N \) is the number of data points in the sample and \( y_i \) is the observed choice of observation \( i \)).

The likelihood function of the observed outcomes, \( Y = \{y_1, ..., y_N\} \), given the parameters \( \theta \) is defined as:

\[
L(Y|\theta) = \prod_{n=1}^{N} P(y_n|\theta)
\]  
(7-8)

The posterior distribution, \( K(\theta|Y) \), can be related to the prior distribution, \( k(\theta) \) and the likelihood function, \( L(Y|\theta) \), using Bayes theorem:

\[
K(\theta|Y) = \frac{L(Y|\theta) k(\theta)}{L(Y)}
\]  
(7-9)

As is discussed in Train (2004, p. 286), equation 7-9 can be simplified to 7-10, the term \( L(Y) \) becoming a normalizing constant that ensures that that posterior distribution integrates to 1. Hence, the posterior distribution is proportional to the prior distribution (probability before seeing the sample) multiplied by the likelihood function (probability that a set of parameter values \( \theta \) would produce the observed outcome, \( Y \)).

\[
K(\theta|Y) \propto L(Y|\theta) k(\theta)
\]  
(7-10)

According to the von-Mises theorem, presented on pages 288-291 of Train (2004), as the sample size increases:

- the posterior distribution becomes a normal distribution
- the variance of the posterior distribution becomes the same as the sampling variance of the maximum likelihood estimator,
- the mean of the posterior distribution converges to the maximum of the likelihood function
- the mean of the posterior distribution is asymptotically equivalent to the maximum likelihood estimator.

Hence according to Train (2004, p. 290), “instead of maximizing the likelihood function, the researcher can calculate the mean of the posterior distribution and know that the resulting estimator is as good in classical terms as the maximum likelihood.”

Estimation of the mean of the posterior distribution does not require maximization of a likelihood function, but can instead be estimated by taking draws of the posterior distribution, \( K(\theta|Y) \), and then
averaging the results. The standard deviation of the posterior distribution is estimated from the standard error of the draws.

Two algorithms are often used to take draws from the posterior distribution, Gibbs sampling and Metropolis-Hasting algorithm. These are often referred to as Monte Carlo Markov Chain (MCMC) methods. These methods use the current draw of the parameters, \( m \), as a base from which the next draw is calculated. Hence there is a correlation between successive draws using these two methods. As normal distributions were assumed for all random intercept latent variables, Gibbs method was used to sample from the posterior distribution.

In Gibbs sampling, sampling from a multivariate distribution is performed using a series of draws, where each draw is performed over the conditional density of one parameter while the others are held constant. Consider an example with three parameters, \( \theta_1, \theta_2, \theta_3 \). The joint posterior distribution of these parameters is \( K(\theta_1, \theta_2, \theta_3) \). Gibbs sampling could be performed as follows. Start with any values of the parameters that result in a non-zero posterior distribution. Call these \( \theta_1^0, \theta_2^0, \theta_3^0 \), where the zero superscript refers to the zeroth (starting) iteration. Draw for a value of \( \theta_1 \) using existing values of the other parameters, \( K(\theta_1 \mid \theta_2^0, \theta_3^0) \). Then draw for the second parameter given the current value of the first and third parameters, \( K(\theta_2 \mid \theta_1^1, \theta_3^0) \), and finally for the third parameter given the current values of the other parameters, \( K(\theta_3 \mid \theta_1^1, \theta_2^1) \). This process is repeated over many iterations.

As is apparent from the brief description of Gibbs sampling, the parameters \( \theta \) are adjusted throughout the estimation, starting from the prior distribution. A burn-in procedure is required to ensure that the distribution of the parameters is representative of the posterior distribution before sampling occurs. Use of an insufficient burn-in period may cause biased estimation results.

When using Bayesian estimation techniques it is required to define the prior distribution of the parameters, \( k(\theta) \), which represents the knowledge of these parameters before data. Flat prior distributions can be used to represent no knowledge of the parameters before distribution. Flat prior distributions were created as described in Train (2009, pp. 297-298). A normal distribution with large variance was used for fixed parameters. If a single random-effect variance is estimated, such as for the random intercept ordered probit model, a flat prior distribution for the variance was obtained by assuming an inverted gamma distribution with low degrees of freedom and scale parameters. This distribution is used as the variance cannot have a negative value. The prior distribution of multivariate variance matrices was defined using a Wishart distribution, which is a generalization of the inverted gamma distribution. This distribution was defined to minimize the information in the distribution.
Section 7.6.1 provided a brief overview of using a marginal maximum likelihood solver for model estimation, while this section provided a very brief introduction to MCMC estimation techniques with the intention of providing a brief overview of the topic as well as introducing estimation controls, such as the prior distribution, number of iterations and the number of burn-in iterations. The values used in this research are described in Section 7.9. The next section then provides a brief comparison, outlining advantages and disadvantages of the two estimation techniques for multilevel models.

### 7.6.3 Comparison of maximum-likelihood and MCMC estimation techniques

ML and Bayesian analyses of the same data can produce rather different point and interval estimates (Browne & Draper, 2006). According to Train (2004, p. 291), “When the two estimates are not similar, other grounds must be used to choose between them … since their asymptotic properties are the same.”

Following this discussion, Train then mentions that the posterior mean and the maximum of the likelihood function often differ in practice as the sample size is insufficient for asymptotic conditions to hold. According to Browne & Draper (2006), in the case of multilevel models, there must not only be a sufficient number of observations, but also that there must also be a sufficient number of clusters in the dataset for asymptotic conditions to apply. Browne & Draper tested parameter estimation results for a two-level variance component (no fixed covariate) linear regression model and then for a three-level logistic regression model. They report that for the two-level linear regression variance component model, that both maximum likelihood and Bayesian approaches using a diffuse prior could be made to yield approximately unbiased point estimates of the cluster-level variance when the number of level two clusters was not small (more than a dozen clusters). In the three-level logistic regression model, the maximum likelihood estimation produced biased estimates and had poor coverage of the interval estimates of the random-effects variables. It must be noted however, that these authors tested different numerical integration techniques, called MQL and PQL (marginal and penalized quasi likelihood) that, according to Skrondal and Rabe-Hesketh (2004, p. 214), can sometimes produce severely biased parameter estimates. Hence those poor results may not be applicable to the current research, which used a different numerical integration technique.

The primary advantage of using Bayesian (MCMC) estimation techniques compared with marginal maximum likelihood techniques for estimating multilevel models is that MCMC estimation does not require optimization of a function, which can be difficult numerically (Train, p. 282). Instead, draws can
be made “from practically any posterior distribution, no matter how complex” (Train, 284). Hence MCMC “can be used to estimate complex models for which other methods are either unfeasible or work poorly” (Skrondal & Rabe-Heskesh, p. 213).

Disadvantages of the MCMC method include:

1) It is difficult to judge convergence when using MCMC techniques, hence often a large arbitrary number of iterations is used to ensure convergence.
2) There can be issues when diffuse priors are used as they may cause slow convergence and a posterior distribution that does not integrate to 1.
3) Use of MCMC can make it difficult to identify a lack of identification.
4) Since MCMC methods do not estimate a likelihood function, a researcher cannot use a likelihood based inference of model quality, such as the $\rho^2$ measure used to compare hazard models in Chapter 6 of this thesis. The model comparison measures used in this research for MCMC estimated models are summarized in Section 7.9.

7.7 First Modelling Attempt – Maximum Likelihood Estimation of Models with No Fixed Parameters

In the process to estimate the discrete choice models, the first estimated models only contained the random-error terms (no fixed parameters were estimated). These comparatively simple models were first estimated in order to understand and to test the suitability of different estimation techniques for this application and also to test the benefit of using a random-intercept model compared with traditional ordered-probit and multinomial logit models. Random-intercept ordered probit models were attempted first, followed by random-intercept multinomial logit models.

Table 7.2 shows the results for these first random-effects simulations performed. $\rho$ is the intra-class correlation coefficient, which is a measure of the total residual variance that is explained by the intra-class variance.

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$  \hspace{1cm} 7-11

where $\sigma_u^2$ is the level 2 (cluster level) variance and $\sigma_e^2$ is the level 1 (observation-level) variance. Note that by definition for an ordered probit model, $\sigma_e^2$ is 1. In results table 7.2 (shown below) the variance is provided followed by its standard error, shown in parentheses.
The cluster level correlation is seen to account for between 20% (segment 3) and 55% (segment 2) of the total variation within the models. These differences in the variances is at least somewhat due to the definition of those two segments as the distinguishing criterion between segments 2 and 3 is the coefficient of variance of the within destination inter-arrival durations. Either way, however, the cluster-level variance accounts for a minimum of 20% of the total, and hence is seen to be important.

The cluster level correlation also shows the effect of the market segmentation. The cluster-level correlation was 0.548 for a model on the complete dataset, meaning that over 50% of the total variance could be attributed to intraclass correlation. The intraclass correlations were reduced in the segmented model. Segments 1 and 3 are the classes with the largest numbers of observations. The intraclass correlation of these two classes is 0.307 and 0.205, showing that less of the total variation is explained by intraclass correlation compared with a single estimated model. This shows that the segmentation is successful in grouping similar destinations into each segment.

Table 7.2: Comparison of Log-likelihood for Ordered Probit Models with No Fixed Covariates

<table>
<thead>
<tr>
<th></th>
<th>Segment 1</th>
<th>Segment 2</th>
<th>Segment 3</th>
<th>Sum of Three Segments</th>
<th>Full Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Frequently-Visited</td>
<td>(Regularly-Scheduled</td>
<td>(Unscheduled Destinations)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Destinations)</td>
<td>Destinations)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Observations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 1 units (i)</td>
<td>17,887</td>
<td>3,097</td>
<td>6,219</td>
<td>27,203</td>
<td>27,203</td>
</tr>
<tr>
<td>Level 2 units (j)</td>
<td>278</td>
<td>592</td>
<td>1,148</td>
<td>2,018</td>
<td>2,018</td>
</tr>
<tr>
<td>Log-Likelihood</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-level model</td>
<td>-23,471.5</td>
<td>-4,825.56</td>
<td>-11,100.88</td>
<td>-39,398.0</td>
<td>-44,915.6</td>
</tr>
<tr>
<td>Random-Intercept model</td>
<td>-20,820.6</td>
<td>-4359.77</td>
<td>-10,921.86</td>
<td>-36,102.3</td>
<td>-37,165.1</td>
</tr>
<tr>
<td>Difference</td>
<td>2,650.9</td>
<td>465.8</td>
<td>179.2</td>
<td>3,295.7</td>
<td>7,750.5</td>
</tr>
<tr>
<td>Cluster-level correlation</td>
<td><strong>σ_u^2 (S.E.)</strong></td>
<td><strong>1.2342 (0.123)</strong></td>
<td><strong>0.2573 (0.0268)</strong></td>
<td>---</td>
<td><strong>1.2118 (0.0568)</strong></td>
</tr>
<tr>
<td></td>
<td><strong>0.4431 (.0402)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A random-intercept ordered-probit model has one additional parameter compared with the standard ordered-probit model, which is the cluster-level variance term \( σ_u^2 \). A likelihood ratio test can be used to test a hypothesis that the additional parameter should be zero by comparing the differences between the full model (the random effects model in this case) and the restricted model (the ordered probit model). The test statistic is \( -2 \left( LL(\hat{ϕ}^H) - LL(\hat{ϕ}) \right) \), where \( LL(\hat{ϕ}^H) \) is the log-likelihood of the restricted
model while $LL(\hat{\beta})$ is the log-likelihood of the full (unrestricted) model. This test statistic is distributed chi-squared with degrees of freedom equal to the number of restrictions (Train, 2009, p. 70). The test statistic threshold for a 99.9% confidence and one degree of freedom is 10.83. Hence an improvement of 5.42 in the likelihood means then the null hypothesis can be rejected. As the lowest improvement in log-likelihood from any of the estimated models was 179.2, the null hypothesis that the cluster-level variability is zero can safely be rejected. This is further corroborated as the standard errors of the level 2 variances are less than 10% of the estimated variances in all models (implying a t-statistic of over 10 and hence the variance is significant to at least 99% confidence). The likelihood ratio test can also be used to test to test the effects of the market segregation. In this case the null hypothesis is that the parameters are equal in all three segments. The full model is the model with market segmentation while the restricted model is the single model used for the entire dataset. As each ordered probit model contains five threshold parameters, the degree of restriction is 10 between the full and restricted models. This is increased to 12 for the random-intercept ordered probit model with the additional latent cluster-level variability that is included. The test-statistic thresholds for a chi-square distribution with 10 and 12 degrees of freedom are 29.59 and 32.91 respectively. As the improvement of the log-likelihood of the segment models compared against a single model estimated on the entire dataset was 5517.6 and 1062.8 for the ordered probit and the random intercept ordered probit model respectively, the null hypothesis that the parameters are equal in all segments can be rejected.

The next step was a comparison of the results estimated from a multinominal logit model compared with the random-intercept multinominal logit model. Both of these models were estimated without any fixed covariates in order to start with a comparatively simple model that could be used to test the initial model estimation techniques and to gain basic experience estimate these random-intercept models. Again, the GLLAMM package was used for this model estimation.

As was discussed in Section 7.5.3, the random-intercept multinominal logit model estimated with a full variance matrix has 15 latent variables for the current study with 6 discrete-choice intervals. This can be a difficult model to estimate for a likelihood-based solver, as the marginal likelihood must be integrated over a joint multivariate normal random effects distribution of 15 latent variables.

Initial tests proved that the GLLAMM solver required excessive computational resources to estimate this problem. For example, estimation of a random-intercept multinominal logit model with no fixed parameters and only four quadrature points per latent variable was attempted on segment 2, which is the segment with the lowest number of observations (this can be seen in Table 7.2). This model had not
converged after two-days on a computer with a dual core Intel Core 2 CPU 6600 (2.4 GHz) with 3.25 GB RAM.

Due to the long run times required to estimate the random-intercept multinomial logit model, estimating a high-quality model using either forward or backwards estimation techniques was clearly infeasible with available computer resources. Hence I made the decision to test other estimation techniques for random effects models. The MCMC estimation technique was used for the rest of this research.

7.8 Comparing Model Goodness of Fit for Models Estimated Using the MCMC Solver

According to Spiegelhalter, Best, Carlin & van der Linde (2002), model comparisons are typically defined using a measure of fit, which is composed of a deviance statistic and a measure of complexity. As increasingly complex models will provide a better fit to the data, the models are compared by trading off model fit and complexity.

Two commonly-used model comparisons are the Akaike information criteria (AIC) and the Bayesian information criteria (BIC), defined in equations 7-12 and 7-13 as presented by Train (2009, p. 367). In these equations, $K$ is the number of parameters and $N$ is the sample size.

\[
AIC = -2 LL + 2 K \tag{7-12}
\]

\[
BIC = -2 LL + \log(N) K \tag{7-13}
\]

In these two model comparisons, the deviance statistic is defined as twice the log-likelihood, while the complexity are defined differently, with the BIC assigning a higher complexity value, which favours more succinct models than the AIC.

The BIC measure of model fit is difficult to apply to hierarchical models. In Bayesian analysis, different models can have the same marginal probability distribution, $p(y) = \int_\theta p(y|\theta)p(\theta)d\theta$ but can be considered to have a different number of parameters depending on how the likelihood and the prior distribution are composed. For the number of parameters, it is also unclear if the number of level 1 parameters or level 2 clusters should be used. Hence the BIC approximation is rarely used for multilevel regression modelling (Skrondal & Rabe-Hesketh, 2004, p. 265).
Neither the BIC nor the AIC model fit criteria can be used as is to compare models estimated using MCMC estimation as this estimation technique does not maximize the log-likelihood function. Instead the Deviance Information Criterion (DIC) was used to compare alternative model formulations.

The deviance information criterion (DIC) is defined in equation 7-14, and is described below (Celux, Forbes, Robert & Titterington 2006).

\[
DIC = \bar{D} + P_D
\]  
(7-14)

The term $\bar{D}$ is the expected value of the deviance, $\bar{D} = E^\theta[D(\theta)]$. The deviance is defined as:

\[
D(\theta) = -2\log(P(y|\theta)) + 2\log h(y),
\]

where y are the data, $\theta$ are the unknown parameters, $P(y|\theta)$ is the likelihood function, and $h(y)$ is a standardizing term that is defined by the data alone ($h(y)$ is set to 1 for model comparisons, and hence this second term is removed from the deviance equation. The expected value of the deviance, $\bar{D}$ is a measure of how well the model fits the data. Lower values indicate a better model fit. When estimating a model using the MCMC algorithm, the likelihood function, $P(y|\theta)$, is not available. The advantage of using the DIC is that the expected deviance, $\bar{D}$, can also be estimated from the average of $D(\theta)$ from the samples of $\theta$ drawn during MCMC estimation, and hence is obtainable without defining the likelihood function.

The term $P_D$ is a complexity measure showing the effective number of parameters in a model, and is defined as “the difference between the posterior mean of the deviance and the deviance at the posterior estimates of the parameters of interest” (Spiegelhalter et. al, 2002). This is defined mathematically as $P_D = \bar{D} - D(\bar{\theta})$. In this equation, $D(\bar{\theta})$ is the deviance calculated using the expected value of $\theta$, $\bar{\theta}$, which is also the resulting estimator of $\theta$ using the MCMC estimation technique. Larger values reflect a more complex model.

According to Spiegelhalter et al. (2002), in models with negligible prior information the DIC is approximately equivalent to the AIC. As diffuse priors were used to estimate all models in this research, this is the case in the estimated models. Hence the DIC can be viewed in a similar manner to the Akaike information criterion.

### 7.9 Estimation of Random Intercept Models – MCMC Solver

Model development of both the random-intercept ordered-probit and the random-intercept multinomial logit models was conducted as follows. The models were built using a forward-stepping model building approach where one parameter was added at a time and tested for significance and an improved DIC. If both were not accomplished then the parameter was removed. If a previously-added
parameter was made insignificant then the model was tested without that previously-added parameter and the two models were compared using the DIC to determine which of the two parameters should be kept. During model development, models were estimated using 500 burn-in iterations, and the MCMC algorithm was run for 5000 iterations. After the structure of the final model was developed, a more precise model estimation was performed following an example provided in Train (2009, p. 303). 10,000 burn-in iterations were used to ensure that the joint distribution was properly developed before sampling. Afterwards, 1,000 samples were obtained by sampling from the posterior distribution for a further 10,000 iterations, but only every 10th sample was saved for model estimation to reduce correlation between successive samples. Since no information about the parameters was known prior to data collection, diffuse (flat) priors were used for all models.

7.9.1 Random intercept ordered probit model

Table 7.3 shows the DIC statistic for the standard (single-level) ordered probit model with no explanatory variables (OP), the variance component ordered probit model (OP-VC), which contains the random intercept terms but no fixed explanatory variables, and the complete estimated random intercept ordered probit model (OP-RI) for all three segments and for the complete dataset. Table 7.4 shows the estimated parameters for the random intercept models.

It was shown in Section 7.7 that both the market segmentation and the use of a random-intercept ordered-probit model improved the model fit for an ordered-probit model estimated with no covariates. The estimated models in this section, which included covariates, confirm this remained the case for the models with covariates.

The deviance information criterion, discussed in Section 7.8, was used to compare different models. Using this criterion, the improvement in model fit through segmentation was apparent. For example in the OP model, the DIC for the sum of the three segments was over 10,000 less than that of the model estimated on the full dataset. The effect of segmentation was reduced, although still important, for the OP-RI models as the difference in the DIC between the full and the sum of the three segments was approximately 1,200. This change in improvement also showed how allowing higher-level grouping at the destination level improved the model quality by allowing variation between destinations.

Examining the OP-RI models, most of the improvement in the DIC was due to the random terms rather than due to the fixed covariates. Take the full dataset as an example, a single-level model with the same fixed parameters as the OP-RI model was estimated. The DIC of this model was 88,165. This is much
closer to the DIC of the OP model (89,841) than to the DIC of the OP-RI model (71,858). This showed that adding the random terms improved the fit of the model, and hence these terms should be retained.

The improved homogeneity within segments was also observed in the variance of the random term, $\sigma^2_v$, as is seen in Table 7.4. The variance of the estimated OP-RI model of the full dataset was higher than Segments 1 and 3, which have more observations than Segment 2. Segment 2 is expected to have a higher random variance term since destinations were assigned to this segment if there was little variance in travel behaviour within the destination. The reduction of the variance of the random term was another indication that the segmentation strategy used in this research increased the homogeneity within each segment.

The day of week parameters were important, which was expected since the number of shipments on weekends was markedly lower than on weekdays. For example in the full dataset, 5.9% and 3.9% of shipments were delivered on Saturday and Sunday respectively. This contrasted with weekdays, where 16.4% of shipments were delivered on Mondays, up to a peak of 19.3% of shipments that were delivered on Thursdays.

An increased distance from the destination to the carrier depot and from the destination to the nearest freeway was also found to increase the inter-arrival duration, as was seen from the negative sign on these parameters. In this case the distance could be considered as a proxy variable for transportation costs. The fact that increasing transportation costs decreases the shipping frequency is intuitive.

Few destination SIC values were significant and also improved the model fit. Non-food retail establishments had longer inter-arrival durations, as did manufacturing firms and service firms. Freight service providers had shorter inter-arrival durations, which was expected for these firms.
### Table 7.3: Comparison of DIC for Ordered Probit Models

<table>
<thead>
<tr>
<th></th>
<th>Segment 1 (Frequently-Visited Destinations)</th>
<th>Segment 2 (Regularly-Scheduled Destinations)</th>
<th>Segment 3 (Unscheduled Destinations)</th>
<th>Sum of Three Segments</th>
<th>Full Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single level (OP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance component,</td>
<td>46,953</td>
<td>9,661</td>
<td>22,212</td>
<td>78,826</td>
<td>89,841</td>
</tr>
<tr>
<td>two-level with no</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>covariates (OP-VC)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Intercept,</td>
<td>41,092</td>
<td>8,242</td>
<td>21,570</td>
<td>70,904</td>
<td>72,046</td>
</tr>
<tr>
<td>two-level with</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>covariates (OP-RI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Covariates of random</td>
<td>40,932</td>
<td>8,199</td>
<td>21,542</td>
<td>70,673</td>
<td>71,858</td>
</tr>
<tr>
<td>intercept model without</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>random effects terms</td>
<td>(not estimated)</td>
<td>(not estimated)</td>
<td>(not estimated)</td>
<td>(not estimated)</td>
<td>88,165</td>
</tr>
</tbody>
</table>

### Table 7.4: Estimated Parameters for the Random Intercept Ordered Probit Models

<table>
<thead>
<tr>
<th></th>
<th>Segment 1 (Frequently-Visited Destinations)</th>
<th>Segment 2 (Regularly-Scheduled Destinations)</th>
<th>Segment 3 (Unscheduled Destinations)</th>
<th>Full Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Component</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Const (≤interval 1)</td>
<td>0.477</td>
<td>*</td>
<td>-2.957</td>
<td>***</td>
</tr>
<tr>
<td>Const (≤interval 2)</td>
<td>1.487</td>
<td>***</td>
<td>-2.343</td>
<td>***</td>
</tr>
<tr>
<td>Const (≤interval 3)</td>
<td>2.246</td>
<td>***</td>
<td>-1.557</td>
<td>***</td>
</tr>
<tr>
<td>Const (≤interval 4)</td>
<td>3.005</td>
<td>***</td>
<td>-0.765</td>
<td>***</td>
</tr>
<tr>
<td>Const (≤interval 5)</td>
<td>3.774</td>
<td>***</td>
<td>0.687</td>
<td>***</td>
</tr>
<tr>
<td>δ 1st visit is on a Thursday</td>
<td>-0.228</td>
<td>***</td>
<td>-0.227</td>
<td>***</td>
</tr>
<tr>
<td>δ 1st visit is on a Friday</td>
<td>-0.264</td>
<td>***</td>
<td>-0.14</td>
<td>***</td>
</tr>
<tr>
<td>ln(distance to freeway)</td>
<td>-0.129</td>
<td>**</td>
<td>-0.056</td>
<td>***</td>
</tr>
<tr>
<td>ln(distance to depot)</td>
<td>-0.031</td>
<td>○</td>
<td>-0.119</td>
<td>***</td>
</tr>
<tr>
<td>δ parcel has non-food retail</td>
<td>-0.258</td>
<td>*</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>δ parcel has non-durable wholesaler</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>δ parcel has manufacturing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>δ parcel has freight service provider</td>
<td>0.653</td>
<td>**</td>
<td>0.167</td>
<td>-</td>
</tr>
<tr>
<td>δ parcel has a service enterprise</td>
<td>-0.221</td>
<td>**</td>
<td>**</td>
<td></td>
</tr>
</tbody>
</table>

### Random Components

| σ²_v                      | 0.413                                       | ***                                         | 1.278                                | ***        |

Significance levels: *** 99.9%, ** 99%, * 95%, - 90%, ○ < 90% significance
7.9.2 Random intercept multinomial logit model

In the random-intercept multinomial logit model, interval 1 (same day) was used as the reference interval for segments 1 (frequently-visited destinations) and for the full dataset since most observations fall into this interval. For segments 2 and 3 (regularly-scheduled and unscheduled destinations), interval 5 (7 – 14 days) was chosen as the reference because it contained the largest number of observations and also since it was a convenient interval for a comparison of weekday effects.

The full random variance-covariance matrix was estimated for segment 2, regularly-scheduled destinations. Convergence could not be obtained for the other segments or for the full dataset using the full random variance-covariance matrix. The off-diagonal terms of the random effects variance matrix were set to zero in the models estimated for these segments.

Table 7.5 shows the DIC statistic for the standard (single-level) multinomial logit model with no explanatory variables (MNL), the variance component multinomial logit model (MNL-VC), which contains the random intercept terms but no fixed explanatory variables, and the complete estimated random intercept multinomial logit model (MNL-RI) for all three segments and for the complete dataset. Tables 7.6, 7.7, 7.8 and 7.9 show the estimated parameters for the random intercept models estimated for segments 1, 2 and 3, and the full dataset respectively.

The DIC for the MNL model was the same as the OP model for all of the segments. This was expected since there are no explanatory variables; the thresholds of the OP model function in a similar manner to the constants of the MNL model. The MNL-VC model had an improved DIC compared to the OP-VC model, showing that destinations have different grouping effects for the different intervals.

The model fit for the MNL-RI model was found to be superior to the OP-RI model. For example, the sum of the DIC for the three segments for the MNL-RI model was 67,461 while it was 70,673 for the ordered probit model. This improvement was expected as the MNL model structure allows a covariate to increase or decrease the probability of the specific inter-arrival duration interval instead of creating longer or shorter trip ends in general.

For the MNL-RI models, the sum of the DICs for the models estimated from the three segments was approximately 1400 less than the DIC estimated on the full dataset. This showed the benefit of destination segregation into the behavioural categories.

Most of the improvement in the DIC was due to the random parameters compared with the fixed parameters. For example, the DIC of a single-level multinomial logit model with the same fixed
parameters as the full model had a DIC of 86,318. This was much closer to the DIC of the MNL model with no covariates (89,841) than to the DIC of the MNL-RI model (68,854). This showed that adding the random terms improves the fit of the model, and hence should be retained.

In all of the estimated MNL-RI models, the weekday had an important effect, which was expected due to fewer weekend deliveries. An increased distance from the destination to the carrier depot and to the nearest freeway was also found to lead to longer inter-arrival durations, as shown by the increasing parameters for these covariates for the longer duration intervals.

Few destination-specific SIC variables could be maintained in all of the segments. In segment 1, the standard error for intervals 5 and 6 made it difficult to keep these parameters in this segment. The only variables that were retained in Segment 1 were $\delta_{\text{parcel has food retail}}$, $\delta_{\text{parcel has non-food retail}}$, and $\delta_{\text{parcel has freight service}}$. Retail establishments in this segment are expected to either be large stores or clusters of many stores, such as shopping malls. Retail establishments in this segment were less likely to have multiple deliveries on the same day, which was shown by the positive coefficients on the other inter-arrival duration intervals.

For regularly scheduled but infrequent shipments (Segment 2), very few destination-specific SIC variables could be retained. This shows that the delivery schedules were not determined by the destination industry class and that other factors were more prevalent.

**Table 7.5: Comparison of DIC for Multilevel Multinomial Logit Models**

<table>
<thead>
<tr>
<th></th>
<th>Segment 1 (Frequently-Visited Destinations)</th>
<th>Segment 2 (Regularly-Scheduled Destinations)</th>
<th>Segment 3 (Unscheduled Destinations)</th>
<th>Sum of Three Segments</th>
<th>Full Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single level (MNL)</td>
<td>46,953</td>
<td>9,661</td>
<td>22,212</td>
<td>78,826</td>
<td>89,841</td>
</tr>
<tr>
<td>Variance component, two-level with no covariates (MNL-VC)</td>
<td>40,456</td>
<td>7,751</td>
<td>21,404</td>
<td>63,011</td>
<td>70,948</td>
</tr>
<tr>
<td>Random Intercept, two level with covariates (MNL-RI)</td>
<td>39,428</td>
<td>7,395</td>
<td>20,653</td>
<td>67,476</td>
<td>68,858</td>
</tr>
<tr>
<td>Covariates of random intercept model without random effects terms</td>
<td>(not estimated)</td>
<td>(not estimated)</td>
<td>(not estimated)</td>
<td>(not estimated)</td>
<td>86,318</td>
</tr>
</tbody>
</table>
Table 7.6: Estimated Parameters for Random Intercept Multinomial Logit Model for Segment 1 (Frequently Visited Destinations)

<table>
<thead>
<tr>
<th></th>
<th>Interval 1 (Same day)</th>
<th>Interval 2 (Next day)</th>
<th>Interval 3 (2-3 days)</th>
<th>Interval 4 (4-6 days)</th>
<th>Interval 5 (7-14 days)</th>
<th>Interval 6 (≥14 days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Component</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Const.</td>
<td>-0.675 ***</td>
<td>-2.552 ***</td>
<td>-3.654 ***</td>
<td>-7.333 ***</td>
<td>-10.19 ***</td>
<td></td>
</tr>
<tr>
<td>$\delta$ 1st visit is on a Wednesday</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta$ 1st visit is on a Thursday</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta$ 1st visit is on a Friday</td>
<td>-0.621 ***</td>
<td>0.404 ***</td>
<td>1.537 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta$ 1st visit is on a Weekend</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln (distance to depot)</td>
<td>0.138 ***</td>
<td>0.167 ***</td>
<td>0.072 ○</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln (distance to freeway)</td>
<td></td>
<td>0.19 ***</td>
<td>0.181 **</td>
<td>0.586 ***</td>
<td>0.562 ***</td>
<td></td>
</tr>
<tr>
<td>$\delta$ parcel has food retail</td>
<td>0.463 ***</td>
<td>0.709 ***</td>
<td>0.267 ○</td>
<td></td>
<td></td>
<td>1.789 **</td>
</tr>
<tr>
<td>$\delta$ parcel has non-food retail</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1.171 -</td>
</tr>
<tr>
<td>$\delta$ parcel has freight service</td>
<td></td>
<td>0.358 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Component</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval 1</td>
<td></td>
<td>0.403 ***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval 3</td>
<td></td>
<td></td>
<td>1.383 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval 4</td>
<td></td>
<td></td>
<td></td>
<td>2.847 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.739 ***</td>
<td></td>
</tr>
<tr>
<td>Interval 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10.78 ***</td>
</tr>
</tbody>
</table>

Significance levels: *** 99.9%, ** 99%, * 95%, - 90%, ○ < 90% significance

In segment 3, retail establishments were more likely to have inter-arrival durations of 7-14 days, as was shown by the negative coefficients for all of the other intervals. The distance to the depot was seen to have a particularly strong effect, showing that fewer shipments were made as the distance increases. A lower effect was seen for the distance to the nearest freeway, although increasing this distance did decrease the probability of same-day visits.

Table 7.9 shows the estimated parameters for the random intercept multinomial logit model estimated using the full dataset. Again it was difficult to keep parameters for interval 6 in the model estimated from the full dataset due to the low percentage of observations and the higher standard errors. Similar trends were seen in the covariates compared with the other segments.
Table 7.7: Estimated Parameters for the Random Intercept Multinomial Logit Model for Segment 2 (Regularly-Scheduled Destinations)

<table>
<thead>
<tr>
<th></th>
<th>Interval 1 (Same day)</th>
<th>Interval 2 (Next day)</th>
<th>Interval 3 (2-3 days)</th>
<th>Interval 4 (4-6 days)</th>
<th>Interval 5 (7-14 days)</th>
<th>Interval 6 (≥14 days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Component</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Const.</td>
<td>-4.336</td>
<td>***</td>
<td>-2.749</td>
<td>***</td>
<td>-2.271</td>
<td>***</td>
</tr>
<tr>
<td>δ 1st visit is on a Wednesday</td>
<td>-0.48</td>
<td>*</td>
<td>-3.222</td>
<td>***</td>
<td>0.882</td>
<td>***</td>
</tr>
<tr>
<td>δ 1st visit is on a Thursday</td>
<td>-3.684</td>
<td>***</td>
<td>-0.94</td>
<td>***</td>
<td>0.852</td>
<td>***</td>
</tr>
<tr>
<td>ln (distance to depot)</td>
<td>-0.331</td>
<td>**</td>
<td>0.02</td>
<td>○</td>
<td>1.608</td>
<td>***</td>
</tr>
<tr>
<td>δ parcel has food retail</td>
<td>-0.875</td>
<td>**</td>
<td>-0.398</td>
<td>**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>δ parcel has manufacturing</td>
<td>-0.467</td>
<td>○</td>
<td>-0.002</td>
<td>-0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Component</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval 1</td>
<td>3.211</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval 2</td>
<td>1.881</td>
<td>**</td>
<td>9.568</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval 3</td>
<td>1.567</td>
<td>**</td>
<td>6.853</td>
<td>***</td>
<td>6.261</td>
<td>***</td>
</tr>
<tr>
<td>Interval 4</td>
<td>1.12</td>
<td>**</td>
<td>4.276</td>
<td>***</td>
<td>3.586</td>
<td>***</td>
</tr>
<tr>
<td>Interval 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval 6</td>
<td>-2.872</td>
<td>***</td>
<td>-5.126</td>
<td>***</td>
<td>-4.146</td>
<td>***</td>
</tr>
</tbody>
</table>

Significance levels: *** 99.9%, ** 99%, * 95%, - 90%, ○ < 90% significance

7.9.3 Market segmentation model

The estimated parameters for the preliminary market segmentation model based on the GPS data are shown in Table 7.10. In a similar fashion as the inter-arrival duration models, very few SIC variables were statistically significant. As expected, the distance from the destination to the carrier depot and the distance from the destination to the nearest freeway had an effect on the market segmentation. Manufacturing firms were less likely to be frequently visited destinations. Food retail establishments were more likely to be frequently visited destinations, perhaps due to short shelf life of perishable products. Non-food retail establishments were more like to receive regularly-scheduled shipments.

The log-likelihood for the model is -1888.0 while the null likelihood for the constants only model is -1924.6. $\rho^2_{(0)}$ (the $\rho^2$ compared with the null log-likelihood) is 0.148, however $\rho^2_{(c)}$ (the $\rho^2$ compared with the constants only model) is only 0.019. Hence the covariates provide little explanatory power for this model. This shows that the GPS data and the establishment data from the InfoCanada database provide little information on how to select appropriate segments for the destinations. A survey with better explanatory variables could be augmented with questions about shipment frequencies and patterns to assess a company’s market segment and provide a better platform for modelling.
Table 7.8: Estimated Parameters for the Random Intercept Multinomial Logit Model for Segment 3 (Unscheduled Destinations)

<table>
<thead>
<tr>
<th>Fixed Component</th>
<th>Interval 1 (Same day)</th>
<th>Interval 2 (Next day)</th>
<th>Interval 3 (2-3 days)</th>
<th>Interval 4 (4-6 days)</th>
<th>Interval 5 (7-14 days)</th>
<th>Interval 6 (≥14 days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const.</td>
<td>1.24 ***</td>
<td>-0.631 ***</td>
<td>0.781 ***</td>
<td>-0.426 **</td>
<td>0.378 ***</td>
<td>-0.29 -</td>
</tr>
<tr>
<td>( \delta ) 1st visit is on a Wednesday</td>
<td>-0.349 **</td>
<td>-0.314 *</td>
<td>0.081 **</td>
<td>0.231 **</td>
<td>-0.298 ***</td>
<td>-0.195 *</td>
</tr>
<tr>
<td>( \delta ) 1st visit is on a Thursday</td>
<td>-1.603 ***</td>
<td>-0.426 **</td>
<td>1.046 ***</td>
<td>0.918 ***</td>
<td>-0.109 **</td>
<td>0.834 ***</td>
</tr>
<tr>
<td>( \delta ) 1st visit is on a Friday</td>
<td>-2.152 ***</td>
<td>0.108 *</td>
<td>0.349 **</td>
<td>0.512 ***</td>
<td>0.275 ***</td>
<td>-0.262 *</td>
</tr>
<tr>
<td>( \delta ) 1st visit is on a Weekend</td>
<td>0.694 ***</td>
<td>0.275 *</td>
<td>0.349 **</td>
<td>0.512 ***</td>
<td>0.275 **</td>
<td>0.223 *</td>
</tr>
<tr>
<td>ln (distance to depot)</td>
<td>-0.275 ***</td>
<td>-0.2 ***</td>
<td>-0.114 **</td>
<td>-0.059 -</td>
<td>0.102 *</td>
<td></td>
</tr>
<tr>
<td>ln (distance to freeway)</td>
<td>-0.108 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \delta ) parcel has food retail</td>
<td>-0.521 ***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \delta ) parcel has non-food retail</td>
<td>-0.362 **</td>
<td>-0.531 ***</td>
<td>-0.231 **</td>
<td>-0.298 ***</td>
<td>-0.253 *</td>
<td></td>
</tr>
<tr>
<td>( \delta ) parcel has non-durable wholesale</td>
<td>-0.262 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \delta ) parcel has manufacturing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Random Component

| Interval 1 | 0.738 *** |
| Interval 2 | 0 0.654 *** |
| Interval 3 | 0 0.133 ** |
| Interval 4 | 0 0 0 0.121 * |
| Interval 5 | 0 0 0 0 1.204 *** |

Significance levels: *** 99.9%, ** 99%, * 95%, - 90%, ○ < 90% significance
### Table 7.9: Estimated Parameters for the Random Intercept Multinomial Logit Model for the Full Dataset

<table>
<thead>
<tr>
<th>Interval 1 (Same day)</th>
<th>Interval 2 (Next day)</th>
<th>Interval 3 (2-3 days)</th>
<th>Interval 4 (4-6 days)</th>
<th>Interval 5 (7-14 days)</th>
<th>Interval 6 (≥14 days)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed Component</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Const.</td>
<td>-0.505***</td>
<td>-1.902***</td>
<td>-3.36***</td>
<td>-3.6***</td>
<td>-2.458***</td>
</tr>
<tr>
<td>$\delta$ 1st visit is on a Wednesday</td>
<td>0.06</td>
<td>-0.368***</td>
<td>0.552***</td>
<td>0.166*</td>
<td></td>
</tr>
<tr>
<td>$\delta$ 1st visit is on a Thursday</td>
<td>-1.429***</td>
<td>1.339***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta$ 1st visit is on a Friday</td>
<td>-0.789***</td>
<td>1.644***</td>
<td>0.259***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta$ 1st visit is on a Weekend</td>
<td>0.335***</td>
<td>1.188***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln (distance to depot)</td>
<td>0.097***</td>
<td>0.228***</td>
<td>0.317***</td>
<td>0.434***</td>
<td>0.624***</td>
</tr>
<tr>
<td>ln (distance to freeway)</td>
<td></td>
<td>0.144*</td>
<td>0.143***</td>
<td>0.23***</td>
<td></td>
</tr>
<tr>
<td>$\delta$ parcel has food retail</td>
<td>0.41***</td>
<td>0.479***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta$ parcel has non-food retail</td>
<td>0.13</td>
<td>-0.304**</td>
<td>0.7***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta$ parcel has non-durable wholesale</td>
<td>0.161*</td>
<td>0.166-</td>
<td>0.33**</td>
<td>0.357**</td>
<td>0.325-</td>
</tr>
<tr>
<td><strong>Random Component</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval 2</td>
<td>0.515***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval 3</td>
<td>0</td>
<td>1.151***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval 4</td>
<td>0</td>
<td>0</td>
<td>2.086***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval 5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.341***</td>
<td></td>
</tr>
<tr>
<td>Interval 6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10.109***</td>
</tr>
</tbody>
</table>

Significance levels: *** 99.9%, ** 99%, * 95%, - 90%, ○ < 90% significance

### Table 7.10: Estimated Parameters for the Market Segmentation Model

<table>
<thead>
<tr>
<th></th>
<th>Segment 1 (Frequently visited Destinations)</th>
<th>Segment 2 (Regularly-scheduled Destinations)</th>
<th>Segment 3 (Unscheduled Destinations)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const.</td>
<td>-0.881***</td>
<td>1.153***</td>
<td></td>
</tr>
<tr>
<td>sqrt (distance to depot)</td>
<td>-0.0952***</td>
<td>0.0318**</td>
<td></td>
</tr>
<tr>
<td>sqrt (distance to freeway)</td>
<td>0.00294</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta$ parcel has manufacturing</td>
<td>-0.296*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta$ parcel has food retail</td>
<td>0.259*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta$ parcel has non-food retail</td>
<td>0.433***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta$ parcel has freight transportation provider</td>
<td>0.250○</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Significance levels: *** 99.9%, ** 99%, * 95%, - 90%, ○ < 90% significance
7.10 Summary

This chapter presents a model of inter-arrival duration, defined here as the number of days between visits to the same individual destination by the same vehicle fleet operator. The main contributions of the research to the literature are described below.

To the authors’ knowledge, this is the first study that models commercial vehicle inter-arrival duration directly from longitudinal GPS data. Inter-arrival duration is the observed outcome of commercial vehicle scheduling over time and is inversely related to shipment frequency. There is potential that application of such a model across the full population of commercial vehicle fleets could be conducted in conjunction with a tour based model approach and could provide information to support the development of trip generation estimates, reflecting some of the impact of changes in land-use, the distance between depots and customers and other attributes included in the model on commercial vehicle scheduling. The advantage of developing better models of the underlying process is an improved sensitivity to policy, land use and logistics. This research is a preliminary step toward this goal, recognizing that the range of policy variables and logistics constraints in the models is currently limited by data availability. More advanced models, such as including latent-class market segmentation, more descriptive random-effects models (such random coefficient models) and future models that estimate the data into finer intervals (instead of, say, 7-13 days), would be able to take better advantage of the detailed information available in this dataset. These models are left for future research.

A second innovation of this research is that all of the statistical models are developed using passively-collected data collected from truck-mounted GPS tracking systems that are ubiquitous across North America. It is of considerable importance that the models developed in this research do not rely on survey data. Conducting long-term travel surveys of commercial establishments is challenging due to privacy concerns and the burden placed upon responding firms. Yet without longitudinal data, day-to-day variability in commercial shipping patterns cannot be properly understood. This research has shown that passively-collected GPS data, merged with other secondary sources of information, can provide long-term truck travel history that is needed for modelling inter-arrival durations. Of course, improved information would potentially allow for better models; for example, if information about the carrier, the truck fleet and the shipments/services delivered could be made available alongside the GPS data.

Initial observations of the inter-arrival durations for all destinations showed that examining the data in its entirety provided a poor representation of the inter-arrival duration of individual firms. This poor representation was found to be due to a relatively small number of frequently-visited destinations that
dominated the dataset. Hence a market segmentation approach was proposed to remove these outliers into a separate dataset in order to improve the overall dataset to a better representation of the decisions made by individual firms. I was also interested in asking a research question to observe whether longitudinal GPS data could be used to distinguish between firms that used push vs. pull logistics. The following segments were created.

1. **Frequently-visited destinations** were defined as destinations where the carrier makes an average of at least 2 visits per week. 13.8% of destinations were classified into this category, accounting for 65.8% of observed inter-arrival durations. These destinations were defined as having an average of at least two visits per week by the observed carrier.

2. **Regularly-scheduled destinations** are destinations that operate using fixed shipment schedules (e.g. weekly deliveries) and are a proxy for firms that do not have frequent shipments and that operate using relatively fixed schedules. 29.3% of destinations were classified into this category, accounting for 11.4% of observed inter-arrival durations. The criteria used to identify these destinations was that they were visited by the carrier at an average rate of less than two visits per week and the inter-arrival duration coefficient of variance for these visits was 0.7 or below.

3. **Unscheduled destinations** are destinations where the shipments are not made on regular schedules. 56.9% of destinations were classified into this category, accounting for 22.9% of observed inter-arrival durations. These destinations were identified where the average visiting rate by the carrier was less than two visits per week, and where the inter-arrival duration coefficient of variance for these visits was greater than 0.7.

The thresholds of an average of two shipments per week and the coefficient of variance of 0.7 were selected by estimating ordered probit models of segments separated using different combinations of these two attributes and then selecting the combination of thresholds for which the sum of all log-likelihoods was the maximum. It is accepted that neither of these attributes are immediately observable for a firm, and hence an additional model is required to segment the destinations. It is acknowledged that since the attributes that were used to segment the destinations were not directly observable from the dataset that a latent class model is also an appropriate modelling approach that would assign the destinations into segments and calculate the inter-arrival durations for each segment within the same model. Such a latent class model is expected to be complex due to the selected discrete choice models (discussed starting in Section 7.5). Development of this model is left for future research.
In spite of the availability of exact inter-arrival durations in the processed GPS dataset, it was decided to aggregate the inter-arrival durations into six discrete intervals as the inter-arrival duration distribution is not expected to be smooth due to large differences in daytime/nighttime stop frequency and that using modelling approaches with continuous dependent variables would hide patterns (e.g. weekly deliveries) in the data due to assumed residual distributions in either OLS regression or hazard statistical models. The aggregated inter-arrival duration intervals were: 1) same day, 2) following day, 3) 2-3 days apart, 4) 4-6 days apart, 5) 7-14 days apart, and 6) 15+ days apart.

Discrete choice models of the inter-arrival durations were then estimated. Due to wide variations in the mean inter-arrival durations of different destinations it was decided to use a random intercept model, which includes a latent destination-level variance that allows destinations to have different mean inter-arrival durations. Only random-intercept models were tested as they are (comparatively) simple random-effects models that met the objective of allowing the mean inter-arrival duration to vary by destination. More complex models such as random coefficient models were not tested. Two types of discrete random-intercept models were tested, a random intercept ordered probit model and a random intercept multinomial logit model.

Initial model estimations were performed using a maximum-likelihood estimator that used adaptive quadrature to integrate over the distribution of the unobserved latent destination-level variance. Initial results of the ordered-probit model demonstrated that use of market segmentation and random-intercept models resulted in statistically significant model improvements (using the likelihood-ratio test) compared with the non-segmented and single-level ordered probit models respectively. The maximum likelihood solver, however, was found to require greater computational resources than were available for the random-intercept multinomial logit model. Subsequent model estimation was performed using the MCMC estimation algorithm.

Market segmentation was used as the observed inter-arrival durations from the full dataset were found to produce a poor representation of expected trip generation behaviour of individual firms. The improvement in model fit due to market segmentation was observed for both the random-intercept ordered probit and the random intercept multinomial logit models, as measured by an improved DIC measure of the models estimated from the three segments compared with models estimated using the full dataset. A preliminary market segmentation model using only GPS data was estimated as a placeholder until a more detailed model can be estimated using survey data (a detailed history of inter-arrival durations is likely not needed to estimate such a model. While the covariates of this model did
little to improve the model fit compared with a constants only model, the benefits of market segmentation are still apparent by the improved consistency with the behaviour of individual firms.

The modelling results showed that multilevel models are advantageous over single-level models as they can reflect that different trips to the same destination are not independent. Both multilevel models were found to perform better than the traditional, single level, versions of these models using the DIC measure. The random intercept multinomial logit model was found to have greater predictive power than the random intercept ordered probit model. This shows that many of the covariates increase or decrease the probability of having an inter-arrival duration within a specific interval instead of increasing or decreasing the shipping frequency, overall.

The model parameter estimates indicate that inter-arrival duration is dependent on the weekday of the trip end preceding the arrival. Longer inter-arrival durations are associated with increased distance between a destination and the depot or the nearest freeway.

Few destination parcel-level SIC attributes were found to be significant and to improve the model fit. This is expected because a property parcel can accommodate many different establishments and because firms sharing the same SIC attributes are heterogeneous. Manufacturing, retail and wholesalers of non-durable goods exhibit longer inter-arrival durations, while freight service providers had more frequent trips.
Chapter 8: Conclusions and Future Work

8.1 Summary of Contributions

The focus of my Ph.D. research was to explore using passively-collected GPS data in order to develop activity-based models of commercial travel behaviour that describe how carriers undertake the deliveries of contracted urban shipments and services. To the author’s knowledge, the research described in this thesis represents the first attempt in the literature to estimate components of a commercial-vehicle transportation planning model using passively-collected GPS data as the primary data source.

It is of considerable importance that the models developed in this research do not rely on survey data. It has been repeatedly recognized in the literature that insufficient data are available from existing data sources by which researchers and practitioners can understand, analyze and then predict urban commercial transportation. As discussed in Section 1.3.1, traditional data sources for commercial transportation models (such as input/output tables and government commodity flow surveys) are useful to provide broad aggregate information about freight flows but do not provide sufficiently fine resolution to understand and predict freight movements. Establishment travel surveys are a viable means to obtain such data but they are expensive, require large amounts of effort to implement, suffer from low response rates and are burdensome on respondents. Hence travel surveys are effectively limited to a single survey day. Also, while surveys in individual cities have been conducted, a lack of consistency in survey data between cities makes it difficult to transfer models.

Advantages of using GPS data collected from in-vehicle GPS tracking and fleet monitoring units include the existing widespread deployment of such systems; many carriers already use this technology in their day-to-day operations. As a single example, in the year 2012 XRS provided GPS tracking services for approximately 114,000 trucks across North America (XRS, 2012). Also, passively-collected GPS data collection efforts are less burdensome than surveys on participating firms, and hence can feasibly collect data over long time intervals. Issues with GPS, however, include the lack of available shipment information and that information about the shipper, receiver and carrier are often suppressed in order to respect privacy concerns. Finally, GPS data from truck-mounted electronic logging devices are only available for trucks and not for other modes.
Specific objectives of my Ph.D. research are summarized as follows. The first objective was to research and develop data processing algorithms that could convert the raw GPS data into a travel diary of trips, trip ends and tours. New data processing techniques were required as information (such as the identity of the carrier and of visited locations) were suppressed by the data provider to respect carriers’ privacy concerns. New and innovative data processing methods developed in this research included: 1) clustering methods that grouped observed trip ends into destinations, enabling studies of longitudinal behaviour, 2) a new technique to use the clustered destinations to identify short and uninteresting “trips”, such as when a truck did not leave the depot, and 3) a new linkage allowing the inference of the attributes of firms operating in particular property parcels by linking the property parcel address with addresses of firms listed in the InfoCanada database of companies.

The second objective was to estimate two behavioural models of urban commercial transportation using the processed passively-collected GPS data. The first estimated model was an activity duration model while the second used the longitudinal GPS data to estimate a model of inter-arrival durations, defined as the number of days between visits to the same destination by a carrier. The activity duration and the inter-arrival duration models estimated during this research were both designed to exploit relative advantages of GPS data compared with transportation surveys. These advantages include accurate observations of arrival and departure times and especially the long observation periods that are feasible using passively-collected GPS data.

The activity duration and inter-arrival duration models were estimated such that they could be used as two of the components within the proposed FRELODE conceptual framework of urban logistics decisions, which is the final contribution of my Ph.D. research. FRELODE was designed to predict carrier urban logistics decisions within a larger agent-based microsimulation framework and has been designed to be estimated primarily using passively-collected GPS data.

Compared with traditional “four-step” freight models, FRELODE has been designed to be more behaviourally valid than current state-of-practice models by recognizing: 1) the diversity of actors involved in freight shipment decisions, 2) increased operating efficiencies through scheduling tours instead of considering shipments to be independent, and 3) including services and not just goods deliveries. When completed, this framework has the potential for improved analysis of infrastructure, congestion and freight transportation policies.

The focus of FRELODE is to represent how carriers and receivers plan the physical transportation of existing urban goods and service delivery contracts. Hence the policy applications of this framework are
focused on how shipment scheduling and vehicle routing decisions can be affected by infrastructure, land-use and policy. Table 8.1 presents applications where FRELODE is expected to better represent the adjustments to travel behaviour of carriers compared with traditional “four-step” transportation models.

Policy responses shown in Table 8.1 include:

- **Land-use**: Responses are separated into two rows, the first which analyzes the impact of types of firms operating out of an individual property parcel while the second analyzes the impacts of distances between shippers, receivers and carriers (reflecting urban form).

- **Effect of travel times on commercial transportation behaviour**: As congestion effects and infrastructure changes affect travel times, these models include effects of congestion and infrastructure changes not only on route choice (calculated separately in a multi-class traffic assignment) but also how firms adjust their behaviour to accommodate changing travel conditions.

- **Effect of Driver hours of service laws**: These laws are one example of policies that are in place to increase road safety. The policy responses presented here could be used to analyze the impacts of changing these laws on carrier travel behaviour.

- **Road pricing**: Shows the ability of the models to respond to how firms change behaviour due to road pricing schemes beyond route choice selection.

In Table 8.1 the policy responses of the estimated activity and inter-arrival duration models are marked in black to show that they reflect responses of the estimated models presented in this thesis. As is seen in this table, the estimated activity duration and inter-arrival duration models can include the impacts of property-parcel land use and also distances between firms to predict how firms adjust their activity durations and shipment scheduling. Effects of travel times, driver hours of service laws and road pricing cannot be included in these models as they are currently estimated. It is anticipated that these policy responses can be included in the vehicle type choice, assign trips to tours and the tour start time components within FRELODE. The policy responses of these undefined components are marked in grey as the policy response cannot be ascertained with certainty until these components have been estimated. As a reminder, an overview of the component steps in FRELODE is provided in Figure 3.2, located on page 53.
Table 8.1: Anticipated Areas of Improved Policy Response of FRELODE Compared with “Four-Step” Transportation Models

<table>
<thead>
<tr>
<th>Component</th>
<th>Policy Application</th>
<th>Shipment Scheduling: Inter-Arrival Duration</th>
<th>Vehicle Type Choice</th>
<th>Assign Trip Ends to Tours</th>
<th>Tour Start Time</th>
<th>Activity Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land use: Property-parcel usage</td>
<td>Aggregated attributes of firms operating on property parcel are included in shipment scheduling</td>
<td>Attributes of the carrier, shipper and receiver are expected to influence vehicle type choice</td>
<td>---</td>
<td>Attributes of the carrier, and shippers and receivers visited in a tour could be used to estimate tour start times</td>
<td>Aggregated attributes of firms operating on property parcel are included in the activity duration model</td>
<td></td>
</tr>
<tr>
<td>Land use: Distances between shippers, receivers and carriers</td>
<td>Distances between the destination and carrier depot, and distance from the destination to the nearest freeway are included in this model</td>
<td>Shipment distance could be included in a model of vehicle type choice</td>
<td>Distances between destinations affects tour formation, including tour shapes and the number of trip ends</td>
<td>Shipment distance could be included in a model of tour start time</td>
<td>Inbound and outbound trip distances are included in this model</td>
<td></td>
</tr>
<tr>
<td>Travel times (congestion effects and infrastructure improvements)</td>
<td>---</td>
<td>Travel times could be included in a vehicle type choice model</td>
<td>Travel times are expected to form a key component of the tour formation component</td>
<td>Expected travel times could influence tour start times</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Effect of Driver Hours of Service Laws</td>
<td>---</td>
<td>---</td>
<td>Driver hours of service laws could form a constraint when planning tour formation</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Road Pricing</td>
<td>---</td>
<td>Road tolls may affect vehicle size, depending on how the tolling system is structured</td>
<td>If available, road costs can be included as a cost to be minimized in the tour formation component</td>
<td>Time-dependent road-tolls are expected to influence tour departure times</td>
<td>---</td>
<td></td>
</tr>
</tbody>
</table>
8.2 Limitations and Future Work

This section examines four categories of future work from this research: 1) potential improvements to the estimated activity and inter-arrival duration models, 2) specifying and estimating the remaining components of FRELODE, 3) developing, programming and implementing FRELODE within the larger Roorda et al. agent-based simulation framework, and 4) advancing the use of GPS data for estimation of other models of commercial transportation.

8.2.1 Future work to improve the estimated activity and inter-arrival duration models

While both the activity and inter-arrival duration models presented in this thesis form a significant improvement compared with previous research efforts for commercial vehicle travel, there are further opportunities for improvements. Examples of such improvements are described below:

1. As was discussed in Section 7.5, including random effects terms is beneficial when estimating models from longitudinal datasets as they can account for non-independence of observations. It would be beneficial to test if using a random-effects model structure would also improve the quality of the estimated activity duration models.

2. Only random-intercept models were tested for the inter-arrival duration models. It would be useful to test the benefits of using more complex random-effects models, such as random-coefficients models that can consider the effects of different destinations on estimated model parameters and not only on the intercept term of the model.

3. In the estimated inter-arrival duration models a market segmentation model was specified to divide destinations into one of three market segments: frequently-visited, regularly-scheduled and unscheduled destinations. This market segmentation model was required as the attributes used to segment the destinations (namely the average frequency and the coefficient of variance of the shipments) are not immediately observable. It is anticipated that a more predictive market segmentation model could be estimated using survey data. If such data are unavailable then it would be beneficial to test using a latent class model that would simultaneously assign the destinations into segments and calculate the inter-arrival durations for each segment.

4. Test additional covariates, such as travel times between depots and destinations, to see if they improve model fit, explanatory power and policy response.
5. The estimated models will require validation before being implemented within the FRELODE modelling framework.

8.2.2 Future work required to complete FRELODE

At this point, the remaining FRELODE components, Vehicle Type Selection, Assign Trip Ends to Tours and Tour Start Time, have not yet been completed.

Given the limitations of the current GPS data set, the Vehicle Type Selection component will likely need to be estimated using a different data source. This data source would likely be based on a recently completed commercial vehicle survey that was conducted throughout the Greater Golden Horseshoe region by Roorda et al. (2013).

The Assign Trip Ends to Tours component will likely require some experimentation to produce realistic tours. Required tasks will include determining suitable tour constraints, such as the maximum payload and the maximum tour duration and to identify a suitable cost function. Researchers may choose to use either a distance or a time-based cost function, or a combination of the two into a generalized cost. If multiple terms are required, such as time, distance and transportation costs, a set of parameters will need to be estimated that will relate these additional terms to each other. It is expected that this component could be estimated using GPS data.

The final model that needs estimating is the Tour start time model. It is anticipated that this model can be estimated directly from the GPS data, either by drawing a start time from an observed distribution of times or by estimating a statistical model, such a hazard model.

Also of interest is a further exploration of using the market segmentation approach presented in Chapter 7 into improving a behavioural understanding of the trip generation behaviour of different types of commercial establishments. The market segmentation approach proposed in this research could be used not only for the remaining models within FRELODE, but potentially also for other models within the larger Roorda et al. agent-based commercial vehicle travel modelling framework.

8.2.3 Future work required to implement FRELODE into the larger agent-based commercial vehicle framework

After the individual components have been completed, the FRELODE framework would then need to be developed, programmed and then tested before it could be implemented within the larger Roorda et al. agent-based microsimulation framework.
One particular question during implementation is the selection of feedback loops to ensure the consistency of travel times between all components in the model. For example the travel times assumed in the Assign trips to tours model should match those produced from the network assignment conducted using the output of trips and tours from FRELODE. The exact location of feedback loops is unclear at this point. Previous stages within the Roorda et al. framework, for example the Market Interaction Decisions and the Logistics Contract Formation components will also require updated travel times. Tests of the accuracy and numerical stability of different feedback mechanisms will likely be required when implementing the large-scale framework. Finally, the models in the FRELODE framework will need to be validated and calibrated from traffic counts and other data sources.

8.2.4 Larger scale efforts with respect to estimation of agent-based travel demand models using passively-collected GPS data

Much of the focus of this research has been on how to use passively-collected GPS data from truck-mounted GPS units as the data source to estimate activity and inter-arrival durations models of urban commercial transportation. This research provides a useful first step in order to help future researchers use GPS data to estimate travel demand models. The analysis methods used in this thesis are likely not only applicable to analyze commercial vehicle travel, but likely to passenger travel as well.

Future research is needed to better integrate available GPS data with secondary data sources to improve the explanatory power of the estimated models. While links to the Transportation Tomorrow Survey and the InfoCanada database were used in this research, other data sources can also be included. Available data sources will depend on context. It is worth exploring additional property-parcel level data sources beyond those tested in this research. For example, municipal governments possess significant property parcel information including zoning information, e.g. residential, commercial, industrial (including finer distinctions, as available) and property tax records. Adding such data would allow trip ends to non-commercial properties to be included in the model.

Additionally, improved GPS processing techniques could be used to make better use of available data and to allow processing of larger datasets. Potential improved GPS processing are as follows:

1. In this research a destination was only linked with land-use data if the destination centroid lay within a property parcel containing matched industries from the InfoCanada database. Hence all destinations located on roads and other public properties were removed. Further data processing could examine linking destinations located on public properties to neighbouring
property parcels. This would allow a larger number of GPS points to be retained for model estimation.

2. The clustering algorithm is currently a limiting factor with regards to processing larger datasets due to computational effort required for clustering the GPS points into destinations. Development of more efficient clustering techniques would allow processing of larger datasets.

3. Trip end identification is currently problematic due to the conflicting requirements of identifying short-duration trip ends and high traffic congestion levels that cause large time intervals with no movement within a trip. Further research into data processing may identify improved trip end identification algorithms.

Finally, passively-collected GPS data have limitations that cannot likely be addressed solely by linking the GPS data with secondary data sources. Such limitations include the lack of shipment information that explains why schedules were competed in a certain fashion. Methods to integrate GPS data and survey data would be beneficial to gain the advantages of the longitudinal GPS data combined with the increased explanatory power of survey data.
REFERENCES


StataCorp. (2009). *Stata Statistical Software: Release 11.* College Station, TX: StataCorp.


Stopher, P. R., & Zhang, Y., 2011. Repetitiveness of daily travel. *Transportation Research Record: Journal of the Transportation Research Board,* 2230, 75-84.


