Household Vehicle Fleet Decision-Making for an Integrated Land Use, Transportation and Environment Model

by

Jared Duivestein

A thesis submitted in conformity with the requirements
for the degree of Master of Applied Science
Graduate Department of Civil Engineering
University of Toronto

Abstract

Household Vehicle Fleet Decision-Making for an Integrated Land Use, Transportation and Environment Model

Jared Duivestein

Master of Applied Science

Graduate Department of Civil Engineering

University of Toronto

2013

Understanding how households make decisions with regards to their vehicle fleet based on their demographics, socio-economic status and travel patterns is critical for managing the financial, economic, social and environmental health of cities.

Vehicle fleets therefore form a component of the Integrated Land Use, Transportation and Environment (ILUTE) modelling system under development at the University of Toronto. ILUTE is a year-by-year agent-based microsimulation model of demographics, land use and economic patterns, vehicle fleet decisions and travel choices in the Greater Toronto and Hamilton Area.

This thesis extends previous work that modelled the quantity, class and vintage of vehicles in ILUTE households. This revised model offers three key improvements: transaction decisions are made sensitive to travel patterns, fuel costs are better represented, and vehicle purchases are considered in the context of the overall household budgeting. Results are promising, but further model validation is required. Potential extensions of the research are discussed.

Acknowledgements

First and foremost, I owe a great deal of gratitude to Eric Miller for taking me on as a student. He is every bit as awesome of a supervisor as his reputation suggests, and it has been a pleasure to work with him over the last few years.

I would also like to thank Matthew Roorda for volunteering to be the second reader for this thesis, and in many ways the perfect person for this task, given how extensively both his Masters and PhD work are made use of within.

Several other people played a key role in making this work a reality. In particular, James Vaughan, Peter Kucirek and Trajce Nikolov deserve a lot of credit for accommodating my programming (in)capabilities, and helping get things finished on time. Adam Weiss also lent a hand to the programming of a number of data filtering processes.

And what would school be without its social side? I've met a great number of amazing people both inside and outside of the ITS Lab over the last couple of years, who have been the source of sage academic advice and many good times. There's far too many of you to name, but know that you've made Toronto a fun and productive place to live for the last few years.

To my friends in Vancouver, thank you for your ongoing encouragement, even if it's mostly come in the form of nagging about how long it was taking me to finish!

Last but certainly not least, a great big thanks is owed to my family, who have supported me every step of the way over the course of my graduate studies (not to mention all the years prior).

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Twenty-first century policy-makers have the daunting task of attempting to reform much our urban areas away from auto dependency, long after it has been hard-coded into the landscape. Meeting this challenge requires an appropriate set of tools that can fully analyze the relationship between land use, vehicle ownership and travel patterns. The Integrated Land Use, Transportation and Environment (ILUTE) model under development at the University of Toronto is one such tool. ILUTE is a year-by-year agent-based microsimulation model of demographics, land use and economic patterns, vehicle fleet decisions and travel choices in the Greater Toronto and Hamilton Area.

This thesis forms one component of the larger ILUTE project. Specifically, it develops a model for use within ILUTE that represents how households make decisions with regards to their vehicle fleets. Understanding vehicle choices and how it relates to their demographics, socio-economic status, regional land use and economic patterns and travel choices is critical for managing the financial, economic, social and environmental health of cities.

The remainder of this chapter is divided into three sections:

- Section 1.1 provides context with regards to the effects of household vehicle ownership and its links with land use and travel patterns, any why understanding this decision-making process is so important to enable the effective planning and management of cities.

- Section 1.2 details how this thesis can be used as part of ILUTE to improve upon state-of-practice methods of representing vehicle ownership in order to better address these concerns.

- Section 1.3 explains the contents and organization of the rest of the thesis.

1.1 Background and Context

Since its invention in the late 19th century, and especially in the decades proceeding World War II, the automobile has completely revolutionized our way of life. Over the last 65 years, North American cities have radically re-invented their landscapes to accommodate the automobile. Existing roads were widened to accommodate additional traffic, extensive networks of limited-access freeways were run throughout metropolitan areas to speed travel and development bylaws were instituted to require all developments to provide an abundance of free off-street parking for users.

For the first 30 years of the postwar era, the radical expansion in personal mobility afforded by the automobile went hand in hand with an unprecedented level of economic growth and higher standards of living. The automobile became an integral part of the American (and Canadian) Dream, and was recognized as a symbol of social status.

In the 1970's, the automobile hit its first rough patches, brought about by the 1973 and 1979 oil price shocks. However, with a twenty-year span of low oil prices beginning in the early 1980's and lasting to the early years of the new millennium, the predisposition towards auto-oriented development continued unabated throughout much of the continent. With the exception of a few central areas, the GTHA has followed this trend. Data derived from the various Transportation Tomorrow Surveys collected by the Data Management Group at the University of Toronto and presented in Table 1.1 suggest over this two-decade period, the GTHA's already-high motorization rate continued to increase (Data Management Group, 2008).

Table 1.1: GTHA Motorization Levels

Year	**Vehicles per 1000 residents**
1986	509
1991	519
1996	512
2001	536
2006	545

However, over the last five to ten years there have been indications that the dominance of automobility may be on the wane. A number of long-standing problems that trace their roots back to an overreliance of automobile-based transportation have become increasingly apparent and urgent, ranging from the financial to economic, social and environmental.

The remainder of this section provides a brief overview of some of these issues, to illustrate how a better understanding of vehicle fleet decisions by itself or as part of a broader integrated urban model such as ILUTE can help inform decision-making to resolve these challenges.

1.1.1 Direct Health and Safety Impacts

Each and every year vehicle collisions result in death, personal injury and property damage. Despite continuous improvement in road safety, these three types of collisions continue to result in negative social and economic impacts for the persons involved in these collisions as well as society as a whole.

In 2010, vehicle collisions nationwide resulted in 2,227 fatalities and 170,629 injuries (Transport Canada, 2012). In Ontario, the same year saw 885 fatalities, 83,910 injuries and 316,025 property damage collisions (Ministry of Transportation of Ontario, 2013). The corresponding figures for the City of Toronto, in 2011 were 43 fatalities and 18,249 injuries of all types (City of Toronto, 2012).

A 2007 study by Transport Canada found that in 2004, vehicle collisions of all types in Ontario incurred $18 Billion in social costs (Vodden *et al.*, 2007). These costs include social and economic impacts of fatalities and injuries, medical and care costs, emergency response costs and more.

In North America, high levels of auto ownership have gone hand in hand with transportation infrastructure and land use patterns that accommodate them, often to the exclusion of other alternatives. Using data from 448 US counties, Ewing *et al.* (2003) found that traffic fatality rates exhibit a strong correlation with sprawling auto-centric development patterns.

A better understanding of the links between vehicle ownership, land use patterns and transportation can help inform understanding of the secondary impacts of policy and investment decisions, and how these may ultimately influence road safety beyond the immediate impacts of any individual project or study area.

1.1.2 Air Pollution and Secondary Health Effects

In addition to direct health impacts caused by injury and death resulting from collisions, excessive vehicle use has a number of negative health effects via the amount of air pollution it causes. Air pollution Criteria Air Contaminants emitted by vehicles is estimated to cause 440 premature deaths and 1700 hospitalizations annually in the City of Toronto (McKeown, 2007). Mortality costs alone are estimated to be in the range of $2.2 Billion, with further costs imposed by lost economic productivity during the 200,000 restricted activity days created by pollution.

The transport sector accounts for 35% of all greenhouse gases emitted in the City of Toronto, with cars and light trucks accounting for the majority of this number (ICF International, 2007). The auto-dependant nature of the surrounding suburbs suggests that emissions there are even higher. Although greenhouse gases have no direct effect on the health of GTHA residents, policies ranging from the local to international level suggests that the GTHA will need to lower its emissions levels, and that some of this will have to come from reduced levels of vehicle use and ownership.

In what would perhaps be considered a tertiary health impact of auto-centric development, in their book *Urban Sprawl and Public Health*, Frumkin *et al.* (2004), provide extensive documentation of the indirect health costs associated with vehicle dependency. Although the focus of this research was Atlanta, many of the phenomena represented therein apply to the GHTA as well. Studies suggest that improved health resulting from increased use of active transportation modes would save the City of Toronto hundreds of millions in both health and social costs (Toronto Public Health, 2012).

1.1.3 High Costs and Depreciating Investments for Households

Other than real estate, vehicle ownership is often considered to be the largest single-item purchases made by most households. Purchase prices run in the tens of thousands of dollars, even before cost of loan financing is taken into account. Even once purchased, vehicles are expensive to own and operate. The Canadian Automobile Association reports that in 2012, even a basic fuel efficient vehicle such as the *Honda Civic LX* that is driven sparingly (i.e. 12,000 km/year) represents over $4,000 dollars in annual operating and maintenance costs just for insurance, fuel, and licensing and registration (Canadian Automobile Association, 2012). In multi-vehicle households, operation and maintenance costs can account for a substantial proportion of overall household budget.

Both housing and personal vehicles provide day-to-day benefits (or in economics terminology, utility) for their users. This comes in the form of a place to live and a means of access to activity opportunities, respectively. However, real estate generally appreciates in value over time and can be expected to provide a positive rate of return on the investment over the long term, in addition to the day-to-day utility it provides as a form of shelter. On the other hand, the physical vehicle itself is almost always a depreciating asset that will diminish in re-sale value over time. Thus, in addition to the direct costs associated with vehicle ownership, there exist substantial opportunity costs associated with not being able to use that same money for other more productive uses that yield a positive rate of financial return for the household, if only an alternative and more cost-effective means of transportation that provides access to the same activity opportunities were otherwise available.

1.1.4 Economic and Financial Liabilities

In addition to direct user-borne operation and maintenance costs, vehicles generate a large amount of indirect or external costs that are borne by society instead of drivers.

Arguably the most prevalent of these is road congestion, which occurs when roads are used to such a degree that drivers begin to inhibit each other's ability to reach their destination in a timely manner. Overreliance of private vehicles as a primary means of transport for so many people within a major metropolitan area such as the GTHA will result in negative social and economic impacts stemming from congested roads. Metrolinx, the GTHA's regional transportation planning authority estimates that as of 2006, direct costs of congestion to regional residents amounts to $3.3 Billion annually, as well as a further $2.7 Billion in lost GDP output (HDR, 2008). By 2031, these figures are expected to rise to $7.8 Billion and $7.2 Billion, respectively, unless action is taken.

Road congestion decreases the ability of people within the GTHA to access the fullest possible extent of the regional labour market. This results in a sub-optimal allocation of the labour force, which in turn results in decreased economic growth and thus lower take-home wages for workers and lower taxes for governments. Improved understanding of how to facilitate more efficient people and goods movement will help enable the region to increase its economic productivity.

Even as congestion results in calls for expansion of Canada's transportation infrastructure, Canadian municipalities are having trouble paying for the maintenance of the infrastructure they already own. The 2012 *Canadian Infrastructure Report Card* found that 20.6% of roads in Canadian municipalities were in poor condition, and rehabilitation would cost $35.7 Billion (Canadian Infrastructure Report Card, 2012). This figure does not include provincial and federal transportation infrastructure, much of which is also nearing the end of its service life and need of renewal.

One of the economic impacts that has been overlooked until recently is the cost of providing parking for all of the vehicles we own. In his seminal work, *The High Cost of Free Parking*, Donald Shoup estimates that the cost of all parking spaces in the U.S. exceeds the value of all vehicles and may even exceed the value of all roads (Shoup, 2005). Although much of this parking is free to the user, it is not "free" to build, and hidden costs of parking are nonetheless paid for by society in a number of indirect ways.

At a broader level, the auto-centric development patterns that have been built throughout much of the GTHA over the last 50 years rely on cheap energy to survive. Long term trends in energy prices are opaque (at best), but nonetheless, development patterns centered on high vehicle ownership and use and that rely on cheap energy are economically vulnerable to rising prices in the future.

1.2 Research Motivation and Significance

In light of the previously described challenges, modelling household vehicle fleet decisions is important both within its own right and as part of the larger urban process.

Traditionally, transportation demand modelling has been conducted using the Urban Transportation Modelling System (UTMS) or 4-step model. Vehicles in UTMS are generally treated as generic entities, and are often applied as a "fixed" rate that does account for many of the factors that influence how households make decisions about their vehicles.

This thesis will develop a model to explicitly represent the three vehicle-related properties to improve their ability to be represented in travel models. These properties, listed in their order of priority, are:

- The number of vehicles within a household, which will be modelled indirectly based on households deciding to either increase or decrease the size of their vehicle fleets;
- The class of each vehicle within the household, which will be modelled directly; and
- The year of the manufacture of the vehicle within the household, which will be modelled indirectly based on households selecting a vintage (i.e. an age range) which will then be converted to an age, and finally in conjunction with the year being simulated, will be used to assign a model year.

In addition to providing more detail for larger modelling applications, the model will also be able to be used for standalone analysis of vehicle fleet choices. Compared to many existing standalone models, this thesis will improve on them through incorporation of the following:

- The desire to own vehicles as dictated by the travel convenience it provides will be explicitly represented in transaction choices.
- The price of fuel, and sensitivity in vehicle type choices based on fuel costs will be represented as part of the class and model year choices.
- Households will be able to examine the overall cost of vehicle ownership and operating for their household vehicle fleet in the context of their household income level.

In summary, the two main advances that this thesis makes are to:

- Provide a more detailed representation of how households make vehicle fleet decisions compared to many of the existing stand-alone models that have been developed.
- Create a model that can also be seamlessly integrated into a broader urban systems modelling framework to create a higher level of realism than can be offered in state-of-practice transportation planning models.

The following subsections provide examples of several "real world" improvements that this vehicle fleet model will provide both directly and through its use within ILUTE.

1.2.1 Improved Ability to Evaluate Investment and Policy Decisions

Vehicle ownership has often been considered to be one of the strongest indicators of mode choice (Ben-Akiva & Lerman, 1974). However, this is very much a chicken-and-the-egg observation; do households drive because they have a vehicle available, or do they buy a vehicle because it's the most optimal mode to get where they want to go?

Intuitively, we would expect that both of these phenomena are occurring to some degree; households are buying vehicles to make trips that they could not otherwise make very easily, but once they own one, they are also using it for other trips that are feasible (but not necessarily as convenient) by other modes. Given that the fixed costs of owning a vehicle are high but the marginal costs of using it once it is already owned are relatively low, it only makes sense to use it every time it is the most time-efficient and/or pleasant means of travel.

Given that the effects of vehicle ownership and vehicle use/mode choice are so intertwined, holistic transportation planning and policy making should ideally address both of these phenomena simultaneously.

Unfortunately, conventional UTMS-based models possess a shortcoming in terms of vehicle ownership, which can reduce the realism of the model for long-term forecasting as well as impede its ability to provide policy analysis. In typical UTMS-based applications, vehicle ownership is determined by rate-based models that are governed by basic household demographics (e.g. household size and income). A level of vehicle ownership is therefore assigned to the household before any of their travel patterns are modelled. In other words, vehicle ownership levels in UTMS models is *not sensitive to the need to own a vehicle*. Although there are a number of motivations for individuals to purchase personal vehicles (collector/hobbyist, status symbol etc.), providing transportation is the principal motivation for most. In economic parlance, transportation is a derived demand, and exists to provide access to the various activities that individuals seek to engage in. Vehicles are thus simply a means of accessing social and economic opportunities, and in modern North American regions (including the GTHA) they are by far the most efficient means available, and hence form the dominant mode of intra-regional travel (Data Management Group, 2008).

Due to the many concerns associated with automobile dependency discussed above, many regions are seeking to increase realistic travel alternatives for citizens through comprehensive investments in alternative modes, particularly public transit (Metrolinx, 2008). Despite this, public transit ridership is not an end goal; rather it is a means of achieving a multitude of economic, financial, social and environmental goals. Therefore, transportation planning models need to be able to be able to properly represent the effects of different types of investments in terms of their ability to achieve these goals in order to inform sensible and robust decision-making.

The importance of having vehicle ownership sensitive to travel patterns will be illustrated through a hypothetical example. Consider a typical “nuclear family” with two vehicles because each parent needs a vehicle to travel to work. Neither of them travel particularly far, they simply work in areas that are not easily accessible by other modes. Outside of work commuting, all family-related activities such as shopping or transporting children can be managed with a single vehicle. The regional transportation authority happens to be studying the feasibility of constructing a transit line that would stop both near the family’s dwelling as well as the workplace of one of the parents. With the line built, one parent would shift to taking the new transit line, and it is assumed that the model accurately predicts this behaviour. As a result, the model would show a reduction in Vehicle Kilometres Travelled (VKT), which are then multiplied by a mileage rate to reflect the marginal cost of driving. This value is then assumed to represent a cost savings to this family, and thus forms part of the estimated user/societal benefits of the new transit line, which is then used to evaluate whether the line should be built. In practice however, in such a scenario the model grossly underestimates the savings that the family will accrue, because it only accounts for the marginal cost associated with driving to work, which given the short distance travelled, is a relatively minor sum. In reality, vehicle ownership has a number of high fixed or partially-fixed costs (i.e. not solely dependent on use), such as purchase price, insurance, maintenance and depreciation. Eliminating the need for the household to own an additional vehicle eliminates these expenses entirely if they choose to sell it, and represents a level of “money back in the pocket” for the household far in excess of their mileage-related savings (recalling the fixed costs for vehicle ownership discussed in Subsection 1.1.4). Because vehicle ownership is an exogenous input variable in UTMS, it will not allow households to reduce their vehicle ownership expenses, and thus the benefits of the new transit line will be underestimated. This type of situation applies to all projects and policies that aim to not just reduce driving, but rather eliminate the need for vehicle ownership in the first place.

Creating a dynamic relationship between vehicle ownership level and travel patterns will allow for a more comprehensive understanding of the effects of transportation investment decisions.

1.2.2 Improved Basis for Vehicle Emissions Modelling

As touched upon in Subsection 1.1.2 transportation infrastructure investment and policy changes affect not just travel patterns and congestion but also environmental and health outcomes. As such, in considering new investments, it is desirable to understand how changes in the transportation network affect these issues, particularly as it relates to the levels of Criteria Air Contaminants as well as greenhouse gases emitted by the transport sector. Travel demand models can be used as a source of input data into software that models vehicle emissions and pollution dispersion.

To model vehicle emissions with a reasonable level of accuracy, a breakdown of the class of vehicle as well as the model year is required (in addition to having accurately modelled travel patterns). Class and model year information allows the model to properly account for the fuel efficiency of the vehicle as well as the presence of other equipment that influences vehicle emission levels, such as catalytic converters.

Unfortunately, most travel demand models treat all vehicles as a generic entity; there is no differentiation between different classes and ages of vehicles in the model. Instead, after travel patterns are calculated, a vehicle type distribution is applied to the entire regional vehicle fleet. Although the overall fleet composition may be correct "on average", depending on how disaggregate the available input data is that is used to develop and assign fleet type distributions, there is still room for error. For example, different areas within a region may be prone to owning different types of vehicles, and may drive them at different rates, or at different times of the day (i.e. different levels of congestion) and thus a randomly weighted assignment may show vehicles types being driven in the wrong amounts, and consequentially, provide inaccurate forecasts of criteria air contaminant and greenhouse gas impacts.

The vehicle fleet model developed herein explicitly incorporates vehicle class and model year (via vehicle vintage choice) into each and every vehicle in the model. Thus, the disconnect between individual vehicle use and the overall regional vehicle fleet composition is eliminated, as the "correct" (i.e. modelled) class and model year information is directly available for each vehicle.

In addition to providing a better basis for modelling vehicle emissions under present conditions as well as for various infrastructure investments, it also allows more accurate policy sensitive outcomes. For example, in examining an emissions-reduction policy measure in which more fuel efficient vehicles are given government-rebates while less efficient ones are heavily taxed, it would be reasonable to expect that these price signals would have an overall effect on market share. Understanding how this would affect overall transport emissions would be challenging under a traditional travel and emissions modelling framework, since it would generally rely on aggregate level class-choice elasticity measures to predict a shift in vehicle fleet composition, and thus runs into even more of a disconnect between individual vehicle use and overall fleet composition. Conversely, because the vehicle fleet model developed in this thesis directly incorporates the decision-making framework that each individual household uses to make class and model year choices, the emissions-related effects of these policies can be modelled more accurately.

It is worth noting that this thesis does not model which vehicle will be used for which trip in multi-vehicle households, which is also necessary for accurate emissions modelling. However, it does provide information which could be used as a basis to feed such a model.

1.2.3 Improved Representation of Vehicle-Related Costs and Behavioural Responses

The costs of vehicle ownership and operation will be better represented throughout this thesis. For example, continuing with the rebates-and-extra-taxes schemes for different classes of vehicles that was just discussed, it would be possible to predict to what degree consumers would shift towards more fuel efficient vehicles. While this is broadly a positive outcome, it will also be subject to secondary behavioural responses that may have the opposite effect of what was intended. From a purely financial perspective, there is very little difference between fuel becoming cheaper and vehicles becoming more efficient. In other words, once a fuel efficient vehicle is purchased, a household has an incentive to drive more than they would if they had bought a gas-guzzler. It may also have the effect of encouraging them to move to a larger residence further away from their places of work and play. Although the net effect of such a rebate scheme is likely still positive (fuel isn't the only cost of moving further away, travel time also becomes a "cost"), it may be less effective than otherwise anticipated. The vehicle fleet model can be used within ILUTE to model these secondary effects.

More explicit representation of vehicle related costs, particularly in the context of household income can also be used to help create a "transportation affordability" index resulting from different land use patterns and infrastructure investments. For example, housing affordability is typically represented on the basis of the cost of housing being no larger than a certain percentage of household income. A vehicle fleet model in conjunction with the rest of ILUTE could be used to develop more nuanced treatment of affordability that considered both housing and transportation expenses, which would help remove the bias of such indices in favour of areas that have cheaper housing yet are extremely auto dependant.

1.2.4 A Basis to Assess the Effect of Changing Attitudes towards Vehicle Ownership

Over the last several years, there has been an increasing amount of speculation that those born after 1980, known as the Millennial Generation, have substantially different preferences than their parents and grandparents in terms of where and how they seek to live their lives. Briefly, the belief is that Millennials are much more likely to prefer living in a walkable neighborhood close to where they work and can take public transit or ride a bicycle for transportation. In other words, owning a car is not seen as a "goal" or a marker of social status, and if one is bought, it is solely because of the utility it provides as a mode of transportation.

The prevalence of such a change in attitudes is not entirely clear. Although data exists to suggest that this generation is driving less than their predecessors, the degree to which lifestyle preferences as opposed to economic conditions are responsible for this is unclear. A vehicle fleet model within ILUTE could potentially be used to tease-out such generational changes in attitudes. Although attitudes can be collected as part of surveys, they are not something that can be simulated. As such, a direct "lifestyle aspiration" type of variable cannot readily be modelled in ILUTE. However, the transaction choice model developed in this thesis can be combined with an additional survey that collects the same type of data as was used to develop this original model. Once the economic factors that ILUTE *can* explicitly represent are accounted for, it will be possible to assess whether there are any statistically significant inter-generational changes in vehicle-related trends.

A vehicle ownership model within ILUTE is useful not only for assessing whether such a "Millennial factor" exists, but also what the long term effects of such a factor will be on long term trends in land use and transportation patterns within the GTHA.

1.3 Research Scope and Structure

This thesis is grouped into three major components, each of which is detailed below.

The first component is a review of relevant vehicle related works. Chapter 2 is a literature review on the history and research progress to date in vehicle fleet modelling. It includes both aggregate and disaggregate models of ownership, as well as disaggregate models of vehicle transactions, and a number of strategies for modelling the class and age of vehicles. Chapter 3 provides a review of the overall ILUTE modelling system. It then provides details regarding the data and findings of a previous vehicle transaction framework undertaken for ILUTE that forms the starting point for the research in this thesis. Finally, it details several other processes within ILUTE that will or could directly interact with the vehicle transaction model.

The second component documents details the changes and additions made to the existing vehicle transaction model to prepare it for implementation, as well findings from the resultant models. Chapter 4 discusses revisions to the overall transaction model structure (relative to what was presented in Chapter 3), as well as a number of changes to individual model components that are designed to make it able to run in the ILUTE simulation environment after it has been (re)estimated. Chapter 5 then details how the model is made sensitive to household activity scheduling patterns and travel behavior, thereby creating the critical link between vehicle use and ownership. Chapter 6 outlines the specification and estimation of the revised vehicle transaction model that incorporates the improvements developed in Chapters 4 and 5. Finally Chapter 7 discusses the development and implementation of a vehicle fleet initialization model, which is used to provide a starting point for vehicle ownership upon which subsequent annual transaction decisions will be simulated.

The third and final component provides a wrap-up of the research. Chapter 8 provides an extensive list of suggested avenues for both near-term and long-term future research and modelling work to be undertaken to further improve the vehicle fleet model within ILUTE. Chapter 9 concludes the thesis.

Chapter 2

Literature Review

This chapter provides a background on research work relating to vehicle fleet modelling that has been undertaken to date. It is broken down based on the purpose of the model being developed, rather than the mathematical structure of the model.

- Section 2.1 discusses aggregate vehicle ownership models.
- Section 2.2 discusses disaggregate vehicle ownership models.
- Section 2.3 discusses disaggregate vehicle transaction models.
- Section 2.4 discusses vehicle type (i.e. class and age/vintage) models.

2.1 Aggregate Models

The desire to model vehicle ownership began in the late 1950's as the effects of widespread motorization and suburbanization in North America began to take hold. As discrete choice theory had yet to be developed and popularized, these early models sought to represent consumer behaviour at an aggregate level. Given the rapid growth in automobile possession around this time, many early papers developed aggregate models for overall societal motorization rates. Motorization refers to the concept of the overall rate of automobile ownership per capita in a given population group (often a country). These early papers often involved predicting new vehicle sales based on the current vehicle stock and assumed vehicle life spans. Examples of these new-vehicle-sales based models include those by Brems (1956) and Nerlove (1957).

Kreinin (1959) was one of the first to assess used car purchases. Using data from three years of the US Survey of Consumer Finances, he links the propensity of used car purchases to various socioeconomic factors and attitudinal trends, and also notes the behavioural correlations between the number of vehicles owned by a household and tendencies to purchase new or used vehicles.

Beesley & Kain (1964) and Kain & Beesley (1965) and Tanner (1966) argued several competing model specifications for an auto ownership model for Leeds, UK. Specifically, Tanner preferred the use of growth-rate and long-term ownership saturation-level based forecasts for ownership, based on the assumption of UK vehicle ownership rates following trends previously experienced in the more motorized United States. In contrast, Beesley & Kain suggest that vehicle ownership needs to be forecast on the basis on input variables such as income and population density. The principal basis for the disagreement appears to be differences of opinions with regards to whether a more theoretically sound but complex model where input variables themselves must be forecast is better or worse than one that is more simplistic in theory but has less room for forecast error from exogenous input variables.

Button *et al.* (1980) review the work of the UK Regional Highway Traffic Model (RTHM) group and the UK Transport and Road Research Laboratory (TRRL). They conclude that vehicle ownership should be modelled using sigmoid functions with finite saturation levels. They then critique RTHM's work on a number of grounds, particularly their use of the "car purchasing income" causal variable on the grounds that it does not properly account for household spending patterns and changes in price of other products. TRRL work is also criticised for its use of extrapolative power curves, which the authors contend results in inaccurate long term forecasts compared to sigmoid functions.

Dargay (2001) examines the effect of income on auto ownership, and specifically changes in vehicle holdings as a response to changes in income levels. She uses a pseudo-panel methodology on cohort data from the 1970 to 1995 UK Family Expenditure Surveys. After examining three different model specifications, she finds that a semi-log function provides the best model form, and populates it with attributes related to income and other socio-demographic factors. Dargay finds that car ownership levels have has an asymmetric elasticity with respect to income; households are more likely to buy a car when their income rises than they are to dispose of a car when their income decreases by the same amount. This suggests that households quickly become accustomed to having an extra vehicle, and are hesitant to get rid of it even if their financial situation worsens, even if it implies an increasing level of foregone consumption of other goods. Dargay's findings suggest that for disaggregate vehicle transaction modelling, having utility functions that include attributes relating to changes in income may be just as important as attributes relating to total income, and that increases and decreases in income should have different sensitivity parameters applied. Dargay also finds that sensitivity to increases in income declines for households that already have a high level of car ownership. This is likely related to households already having enough cars to maximize their transportation utility at this point, and the motivation for purchasing additional vehicles relates more to collecting different types of vehicles for its own sake, rather than as a means to travel more conveniently.

2.2 Disaggregate Ownership-Based Models

With the advent of discrete choice theory in the 1960's and 1970's, disaggregate models began to be used to estimate vehicle ownership behaviour at the household level; effectively replacing the aggregate models that had been used previously. Disaggregate choice models have a better ability to represent decisions directly in terms of how they are actually made, as well as improve the ability of the models to test the impacts of various scenarios and policies on vehicle ownership.

One of the first examples of a disaggregate vehicle ownership model was developed by Cragg & Uhler (1970). Cragg & Uhler use a multinomial logit model to estimate the number of vehicles a household will own, with the choice set ranging between 0, 1, 2 and 3 or more, using data from the US Survey of Consumer Finances. Variables used in the model incorporated various transformations of income and savings information, as well as information on labour force participation, number of children and location relative to the city centre. Although the model fit is reasonably strong, it is worth noting that it relies entirely on differences in the alternative-specific parameters of non-varying household demographic and socioeconomics variables; nothing in the way of "how useful is owning this many cars to the household" type variables are included.

Ben-Akiva & Lerman (1974) appear to be the first researchers to explicitly model a link between vehicle ownership and travel patterns, thereby breaking reliance on modelling vehicle ownership decisions solely in the context of non-varying household demographics and socioeconomics such as the model developed by Cragg & Uhler. Ben-Akiva & Lerman simultaneously estimate vehicle holdings and commute to work mode choice. This model is conceptually intended to act as the second step in a three step land use and transportation model, with the first step being residential and work location choice and the final step being non-work travel mode choices. The model used data from a 1968 household survey of Washington DC residents, and assumes that one member of the household (the "breadwinner") would commute to work, while the other household member would stay home. Vehicle holdings considered consisted of zero, one and two or more vehicles, while mode choice alternatives consisted of driving or taking transit. These vehicle holding levels and modes were then combined with each other to yield a total of five choice alternatives (since a no-vehicle household cannot drive to work). Six categories of explanatory variables were included in the model, including, transportation level of service (e.g. time, cost for each mode), socioeconomic variables (income, number of licensed drivers, household size), locational attributes (whether the breadwinner works in the central business district), auto ownership costs, housing attributes and spatial attributes (generalized costs for other trips such as shopping). The resultant model is strong at both an overall level of fit as well as for each individual parameter. Direct and cross-elasticity values are also developed both to provide an additional indication of policy sensitivity and intuitive understanding of how changes in independent variables will influence behavioural choices. They expand on this approach further with several different model specifications in Lerman & Ben-Akiva (1976).

Train (1980a) undertakes a similar analysis of mode choice and auto ownership for residents of the San Francisco Bay Area. Train estimates the mode choice and auto ownership models separately, but with mutual conditional probabilities, and is ultimately able to assemble a joint probability model. His model also increases the number of potential modes to seven from Ben-Akiva and Lerman's two, through BART and buses being considered separately, and each with different access modes, as well as a carpool mode. Furthermore, he greatly expands on the number of variables included in the model to improve both overall fit as well as the policy sensitivity that can be tested with it. In particular, his explicit representation of transit access and wait times is an improvement over a single generic out of vehicle transit time variable that most mode choice models used at the time.

Using data from four separate surveys (Boston, the San Francisco Bay area, the Puget Sound area and a nationwide survey from the Netherlands), Bhat & Pulugurta (1998) estimated vehicle holdings using two different model structures, namely a multinomial logit and an ordered logit model. The specific attributes included in the model varied from survey to survey, but generally consisted of socio-economics and income, household location and demographics/family structure. No travel or mode choice related variables were used. After assessing the relative performance of the different model structures, Bhat & Pulugurta conclude that although both perform well, the multinomial logit structure offers better performance, principally due to its ability to have represent alternative-specific effects of exogenous variables.

Chu (2002) creates an ordered probit model of vehicle ownership in the New York area that features a particularly strong use of GIS-based information as part of the choice model. In addition to the usual household demographics and socioeconomic variables, Chu includes a number of accessibility and land use related metrics. An employment accessibility index was calculated using a function of employment levels and travel times to other area (traffic zones) throughout the city. This was calculated for three separate modes that each has its own travel times (walk, transit and drive). The ratio of accessibility levels between drive and all modes put together is then labelled the automobile importance index. The intent of this index is to represent the comparative advantage (or lack thereof) that owning a vehicle would have in terms of its ability for the household member to access work activity opportunities. Variables entitled the “land use entropy index” (based on a function of the percentage of land in the immediate area that is developed, and what it is being used for) and “mixed density index” (a function of nearby population and employment density) were also developed. Model results are strong, and suggest that the land use variables do play a statistically significant role in decision-making, although not to the same degree as demographics and socioeconomics.

In a Canadian context, Potoglou & Kanaroglou (2008) develop a household vehicle ownership model for Hamilton, Ontario. They conduct a survey (CIBER-CARS) that consists of a retrospective survey on vehicle ownership and household composition, and also has a supplementary component on hybrid and alternative fuel vehicles for households that indicate they are considering becoming active in the market for a new vehicle in the near future. Two separate model structures are used for estimation: multinomial logit and ordered logit. Using a GIS database, they were able to generate variables for land use and urban form in addition to the standard household demographics/socioeconomic variables. They use the same calculations for the land use entropy and mixed density indices previously used by Chu, as well as the number of bus stops within walking distance of the household. They also have a dummy variable for whether commute distance is larger than a 6km threshold value. The models both perform well (the multinomial logit more so), and indicate that the land use and urban form variables do have a statistically significant effect on auto ownership levels even after controlling for other factors.

Cao *et al.* (2007) undertake a similar study, but supplement their survey with questions regarding lifestyle preferences and attitudes of respondents. They conducted a household survey in eight neighborhoods throughout Northern California, carefully selected to provide contrasts in neighborhood type, size of the metropolitan area and location within the state. In addition to the standard demographic, household composition and socioeconomic information gathered by the survey, it also featured to large sections relating to respondent attitudes. The first section dealt largely with neighborhood characteristics (both in terms of perceptions of the neighborhood the household actually resided in as well as their “ideal neighbourhood” preferences), while the second section inquired about travel attitudes. Each of these two sections had several dozen statements which the respondent could either agree or disagree with according to a multi-point scale. These responses were then consolidated into attitudinal variables for use in the model. Perceived neighborhood characteristics variables were supplemented with certain “objective” connectivity variables using GIS data after surveys were completed and returned. Two types of models were estimated; an ordered probit model and a static-score model that was limited to households that had moved recently. Two ordered probit models were estimated; one using physical/objective characteristics only and the other also incorporating neighborhood and travel attitudes. The first model shows that although some objective and perceived neighborhood characteristics do have a statistically significant influence on holdings level, these influences are small in magnitude and holdings decisions are largely governed by income and household demographic structure. Once the neighborhood preferences (as opposed to neighborhood characteristics) and travel attitudinal factors are incorporated into the second model, the results indicate that neighborhood characteristics become insignificant and are replaced in importance by the attitudinal factors. This supports the hypothesis that residential location choice and vehicle holdings may be more of a product of self-selection rather than induced by the built environment. However, even with attitudinal variables included, income and household demographics still dominate. In contrast to the ordered probit model, the static-score model suggests that among households that moved, there may be some minor link between objective and perceived neighbourhood characteristics and vehicle holdings. The paper concludes that household demographics and socioeconomics dominate vehicle holdings decisions, and that the build environment may have a more minor effect.

Bhat & Guo (2007) develop a model to estimate vehicle holdings choices in the context of household built environment choice, while explicitly accounting from the presence of potential self-selection in choice of place of residence. Using a joint mixed multinomial logit-ordered response structure, Bhat & Guo develop a model for the San Francisco Bay area that accounts for both residential location choices among several hundred possible zones within a city, as well their subsequent vehicle holdings choices. The residential location choice model runs first and is largely based on household demographics, neighborhood structure (i.e. density, connectivity, access to recreation opportunities etc.) with the only vehicle/travel-related variables being measures of drive time to work (interacted with income level) and measures of local transit and bicycle network accessibility. Based on the residential location choice, vehicle holdings are them modelled, with variables relating to household demographics and socioeconomics, neighborhood urban form and land use patterns, drive-to-work travel time and cost and local accessibility to other transportation modes. Based on the model results, they reach a number of conclusions. Among these are that the built environment affects residential choices as well as vehicle holdings choices, and that policies that alter the built environment should reflect both of these processes when assessing their efficacy. They also note that both the built environment and socioeconomics/household demographics affect vehicle holdings, but the built environment has a smaller impact, and demographics/socioeconomics (especially income) are much more important.

Using data from 25 years' worth of the Statistics Canada Household, Income, Facilities and Equipment survey, Chingcuanco & Miller (2012) develop a meta-model of vehicle holdings. They use a multinomial logit to estimate vehicle holdings for each year over the 25 year time frame. Using the joint context estimation method, they develop three separate types of models (a vary constants model, a varying scales model and a varying constants and scales model). Selected scale parameters and alternative specific constants that vary over time are then regressed against key macroeconomic variables (price of gas and change therein, unemployment rate and change therein, as well as the survey year). The regression model indicates a clear and statistically significant relationship between these macroeconomic variables and the scales and alternative specific constants, suggesting that macroeconomic variables can play a role in household decision-making, even once individual contexts are accounted for (i.e. unemployment rate still matters, even after considering actual household income).

Generally, results from the review of disaggregate models for vehicle ownership level choices suggest the following key insights:

- Multinomial logit models perform better than ordered logit models for vehicle holdings decisions.
- Household demographics and socioeconomics are the dominant factors in ownership decisions.
- Built environment and land use factors also have an effect on vehicle holdings, but they are smaller in magnitude, and partially caused by residential self-selection related to preferences for certain neighborhoods and amounts of travel.
- Estimating ownership decisions in conjunction with actual travel requirements (i.e. how much does owning a vehicle actually improve travel conditions) can improve model performance and policy sensitivity.

2.3 Disaggregate Transaction-Based Models

The initial use of discrete choice models for vehicle ownership almost exclusively modelled vehicle holdings/ownership level (i.e. "how many vehicles does the household own"), per the works discussed in Section 2.2. However, shortly thereafter models that represented vehicle transaction decisions (i.e. "should the household add another vehicle or remove one of their existing ones") began to emerge. The popularity of transaction choice models was largely driven by the belief that they better represented actual human behaviour and household decision-making (de Jong & Kitamura, 2009). Part of this belief stems from the notion that households do not set a certain target number of vehicles to own, but are driven to change their vehicle fleet size by the sentiment/notion that they "don't have enough" or "have too many" vehicles, and that adjustments don't necessarily happen instantaneously due to vehicles being major purchases.

An early version of a vehicle transaction model was developed by de Janosi (1959). While not a full transaction choice set (nor a discrete choice model), de Janosi modelled the probability that a household would purchase a new vehicle using a multivariate linear regression model. He used a dataset of observed transactions (where the dependant variable of transaction decisions was set to 1 if the household did make a purchase and 0 otherwise), and regressed against a series of dummy variables relating to income and finances, marital status, existing vehicle holdings, age and place of residence. Model performance was respectable, but did suffer the conceptual limitations of relying solely on dummy categorical variables rather than linear or non-linear continuous functions, as well as the fact that the model could only represent the decision to buy a vehicle or do nothing, not a full set of transaction choices.

Hocherman *et al.* (1983) provide one of the earliest examples of a transaction-based model of household vehicle fleets using discrete choice theory. Their rationale for using transaction models over an ownership level model is that:

- Representation of costs associated with making a transaction (both financial and in terms of time/effort on the part of the searcher) can be incorporated. Ownership level models generally assume no costs are associated with having to actually acquire a vehicle beyond an implicit assumption regarding price which may be captured either directly as a variable of its own such as in Train (1980) or implicitly through an alternative specific constant.
- Brand or vehicle-type loyalty effects are better accounted for when vehicles are replaced.
- The potential to resell their current vehicle to help purchase a new one could be incorporated.

In their study of household vehicle fleets in Haifa, Israel, Hocherman *et al.* develop a series of models that assess both the type of vehicle that could be purchased as well as whether a vehicle will be purchased in a transaction. The review here will focus predominantly on the transaction choice component of the model, rather than vehicle type choice. They begin with a discussion of the Israeli vehicle market, noting that almost all households in the survey have either no vehicles or one vehicle; multi-vehicle ownership is extremely rare. Given the lack of multi-vehicle households both in Israel (at least at the time this work was conducted) as well as in the data they are using, they elect to limit their model to choices that allow for either no vehicles or one vehicle to be owned by the household. To this end, they develop two separate nested logit models. One is for households that do not own a vehicle; these households are able to choose to either do nothing or buy a vehicle. The buy alternative has a nest underneath it for vehicle type choice that enters the utility function of the buy alternative as a logsum value. The other model is for households that already own a vehicle; these households are able to either do nothing or replace their existing vehicle, with the replace choice also being a nest that incorporates vehicle type choice beneath it.

The transaction choice models for the no-vehicle and one-vehicle households are estimated separately, but have largely the same types of variables (even if their exact specifications and parameter estimates differ). In both cases, the do nothing alternative could be considered the "default" choice with a utility of zero, and the buy/replace alternatives are populated with utility attributes that influence the transaction choice. Both models feature multiple variables relating to household demographics, socioeconomics (especially income) and expected utility to be gained from the particular vehicle being purchased. For the no vehicle households, the commute to work travel times by vehicles versus transit are also considered, as is the distance to the nearest bus stop. The one vehicle household model does not have variables that compare travel by different modes, since it is assumed that nobody would actually return to having no vehicles once they own one (and the model does not allow this). Instead, it does have a number of variables relating to the physical properties and age of the currently owned vehicle, allowing the age and lower performance of the vehicle to help trigger replacement decisions.

Although this paper does break ground in its use of transaction choices as opposed to ownership choices, it does have a number of conceptual challenges for wider applicability, many of which the authors themselves note. Amongst them are:

- It requires separate models for each ownership level (zero vehicles, one vehicle). While this is adequate for the Israeli vehicle market where these two ownership levels account for almost all households, it does not work for more heavily motorized locations where owning, 2, 3 or even 4 or more vehicles is not a rarity. In these situations, the number of models to be developed could become excessive, and it would be easier to develop a single transaction choice model that allowed households to move both up and down.

- Households cannot decrease their vehicle holdings. While the authors note this is almost unheard of in Israel, it is common in other areas, particularly when it comes to a multi-vehicle household reducing their ownership level (but not necessarily to zero).

- They also do not develop a disposal choice model. This is not necessary for their application (since one-vehicle households that replace their vehicles only have one vehicle they could possibly dispose of, the "choice" is self-evident), but this would be required for a higher ownership level household.

Hensher & Le Plastrier (1985) are another early example of progress towards transaction choice modelling, although they do not develop a full model of such. They use data from a five year retrospective survey of four hundred households in Sydney, Australia interviewed in 1980 that asked respondents about their current and past vehicle fleet holdings as well as household and personal demographics. Hensher & Le Plastrier construct a model of vehicle ownership level and vehicle type for each of these vehicles in a base year, and from that year forward, transaction choices (they use the term "fleet adjustments") as well as the vehicle type of each of these transactions (if applicable) are evaluated. Unfortunately, they were unable to actually evaluate a model of transaction choices due to insufficient observations of choices that involve an actual fleet size adjustment. Instead, they represent fleet size adjustments (i.e. transaction choices) through a model of vehicle type choices that applies to all years subsequent to the base year. The only variables related to actual transaction choices in this model are dummy variables that represent what the household did last year; no variables that relate to changing household demographics or socioeconomics or an improved ability to travel are included. The implication is that if having a new vehicle type is found to be desirable for the household, then this implicitly creates a fleet size adjustment. As the authors themselves note, this is not an ideal format, but is rather a step towards developing a fully transaction-based household vehicle fleet model.

On a separate note, in developing all of their models, Hensher & Le Plastrier create separate ownership level model estimations for each year in the retrospective survey (1975 to 1980). They show that these models are statistically different from each other, creating concerns about the temporal transferability of vehicle fleet modelling for even medium-term applications. The degree to which consumer tastes "wander" over time is unknown, as it is not clear whether their models are reflecting actual changes in taste over time or simply having their parameters adjusted in a manner that corresponds to changes in systematic but unobserved (i.e. unmodelled) factors that influence household decision-making with regards to vehicle ownership. This finding has implications for all types of disaggregate vehicle fleet modelling, regardless of whether the process being modelled is ownership level, transaction choice or vehicle type choice.

de Jong & Kitamura (2009) publish the results of work they originally presented in 1992. They do not develop any models themselves, but instead review many of the theoretical foundations behind vehicle transaction models and the advantages it has over ownership level models. Specifically, vehicle transaction models are able to:

- Represent the changes in new product penetration over time, as the existing vehicle fleet is eventually replaced when the household makes transaction that involve purchasing a new vehicle or disposing of an old one. This has applications to policies such as those that encourage the purchase of certain types of vehicles (e.g. more fuel efficient).

- Account for the fact that the vehicle fleet is not usually instantaneously updated as soon as the household conditions change. Not only are there search/transaction costs associated with becoming active in the vehicle fleet market, but the expense of vehicles dictates that households are unlikely to purchase multiple vehicles at once; they must be replaced over time.

- Better represent the models to assess vehicle type choice, because only one vehicle is purchased at a given time. Vehicle type choice can be reasonably assumed to be influenced by the other vehicles already owned by the household. Thus, models that represent both ownership level and vehicle type choice behaviour suffer from the problem that the choice of vehicle type must be modelled in conjunction for all other vehicles if it is to be realistic. Depending on the choice set size for vehicle types (how many classes and years/ages are included), the choice set can be large to begin with. However, the size of the choice set increases exponentially with the number of vehicles owned, and quickly becomes both unreasonably large to model for households with 3 or more vehicles and suffers from violations of the IIA property.

- Factors such as brand loyalty are less able to be represented in static ownership level models.

de Jong & Kitamura finally note that as of 1992, no complete model of vehicle fleet choices using a transaction framework had ever been created. They discuss their plans to create such a model for the Netherlands, although this model was never actually developed.

Roorda *et al.* (2009) take a pre-existing vehicle transaction model for the Toronto area (which will be discussed in detail in Section 3.3) and use an activity-based travel demand model (the TASHA model, which will be discussed in Subsection 3.4.2) to simulate travel patterns for households in the survey data. Specifically, using data from the year 1996 (since this is the year the travel model is calibrated for), Roorda *et al.* simulate the number of vehicle sharing conflicts and overall household travel utility that is incurred at the current ownership level, as well as with one additional and one fewer vehicles. These changes in the number of conflicts and overall household travel utilities are then added onto the same specification of the existing model as "stressors", and the transaction choice model is then re-estimated to incorporate them. Although the overall effect of adding the stressors does appear to be marginal in terms of its ability to improve fit of an (already very strong) choice model, it does greatly increase the sensitivity of the transaction decision to one of its key intuitive attributes, which is the household need for a vehicle for transportation purposes. It thus creates the critical link between vehicle ownership and transportation infrastructure investment decisions that affect traffic flow and mode choice. Several improvements to the transaction choice model are suggested for further improvement, such as better representation of income and vehicle operating costs.

Paleti *et al.* (2011) present another comprehensive transaction-based model of vehicle fleet decisions for use in a microsimulation environment. They conduct a three-part survey on California households similar to the CIBER-CARS survey conducted by Potoglou & Kanaroglou. The first component is a revealed choice of current and recent household vehicle fleet decisions making. This is supplemented by a stated intentions survey of whether the household intends to enter the market for a new vehicle in the near future, and if they indicate they have such an intention, a stated preference survey is administered. The stated preference survey asks respondents to evaluate their hypothetical vehicle choices based on a number of different vehicle related choices including six vehicle class types, seven fuel types (one of the goals of the model is to be able to assess the impacts of incentives to adopt alternative fuel vehicles) and five vintage types, as well as including a no-buy option.

The RP and SP data is then combined to develop a series of models for transaction choice, class choice, fuel type choice, vintage choice and Vehicle Miles Travelled (VMT).

Paleti *et al.* treat vehicle usage differently than Roorda *et al.* Instead of obtaining vehicle travel output from a travel demand model, Paleti *et al.* model VMT directly, largely based on household demographics, socioeconomics and the structure of the exiting vehicle fleet. This is admittedly a more convenient approach than Roorda *et al.*, as it does not rely on a separate external model that needs its own estimation and development, however it both prevents the model from drawing an explicit link to regional transport infrastructure and demand changes, and is not capable of producing the same level of detail in its output (VMT only, rather than also travel utility and scheduling conflicts).

Interestingly, the transaction choice model they use is not a single four-alternative decision structure as outlined by de Jong & Kitamura, but rather two separate binary logit models for a replace choice and a buy choice (both against a "do nothing" alternative). In both these models, transaction choices are largely decided based on household demographics, socioeconomics and the state of the existing vehicle fleet. Travel variables do not enter into the equation, even though the VMT model is intended to be sensitive to ownership level and thus different VMT estimates could hypothetically be developed for different ownership levels (with VMT acting as a proxy variable for overall household travel utility). Instead, the use of the VMT is confined as a means of estimating vehicle operating costs. The exact reason for this is unclear, but may be because they do not feel the VMT estimates produce the sort of information households would use in evaluating fleet size adjustments.

Paleti *et al.* do present an innovative approach to modelling the base year vehicle fleet. Transaction models are intended to fix many of the issues associated with ownership level models, but nonetheless rely on these models for provision of a base year vehicle fleet. Instead of simply developing a static ownership level model, Paleti *et al.* "pre-simulate" past transaction decisions prior to running the full transaction model as part of a microsimulation. Essentially, households start with no vehicles, and then keep deciding to add one (as well as choosing vehicle type) until they reach their desired number of vehicles. This should allow the presence of the first vehicle(s) to influence subsequent choices, as one would expect in real life, rather than all vehicle types being chosen independently, and still avoiding the problems with extremely large choice sets noted by de Jong & Kitamura.

Generally, results from the review of disaggregate choice models for vehicle transactions suggest the following key insights:

- Transaction choice decisions have an improved ability to represent actual household decision-making processes compared to ownership level models.
- Efficient representation of vehicle type choice as a lower level of decision-making is a crucial benefit of transaction modelling that should be taken advantage of.
- The ability of transaction models to explicitly incorporate changes in travel convenience resulting from adding or removing a vehicle to/from the household can improve model fit and explicitly creates a link between transportation infrastructure investment and vehicle holdings.
- Most of the lessons learned from vehicle holdings or ownership level models apply equally to transaction choices.
- Transaction models still require an ownership level model for the base year.
- Temporal transferability is a major concern of vehicle fleet modelling, and long term forecasts should be treated with caution.

Finally, it should be noted that this literature review has been predominantly focused on discrete choice models for transaction choice, because this is the type that has previously been used in ILUTE work on household vehicle fleets and will be used in this thesis. However there are several other types of models that can be applied to transaction decisions. For example, de Jong (1996) provides an example of a transaction choice framework for one vehicle households in the Netherlands that makes use of a duration model. Rashidi *et al.* (2011) develop a joint model of vehicle transaction choices, household residential location choice and job chance choices of individual household members using both Weibull and log-logistic baseline hazard functions.

2.4 Vehicle Type Models

At almost the same time as disaggregate vehicle holdings models began to be developed, the same discrete choice theory was used to develop models of vehicle type choice. The term “vehicle type” is used herein to refer to models that represent choices of either the class of vehicle and/or the vintage/age/model year. These are discussed together as “vehicle type” since many models represent both of these concepts simultaneously. Furthermore, although vehicle type models can often function as independent models, with the exception of the earliest models they are often used as a lower level choice in a vehicle holdings/ownership model or a transaction choice model.

Johnson (1978) was one of the first to develop a series of models relating to vehicle vintage and age. His first model is a simultaneous model of vehicle holdings and vintage; households are modelled as choosing between zero to two vehicles, with all possible combinations of vehicles being new or used (new or used for one vehicle, both new, both used and one of each for two vehicles). A simple multinomial logit model is developed based on household size, age of the head of the household and income level. Several other multinomial logit models using data from various survey years are developed to assess choice of vehicle age for vehicle purchases (with alternative ranging from 1 to 6-or-more years). These models use the exact same attributes to populate the utility functions. All models have reasonably strong levels of fit, but rely on very large alternative specific constants in some cases to make the model fit. This suggests that although the variables included as attributes in the model do matter, there are a number of other unobserved aspects that should be accounted for.

Although not the first models of vehicle class choice, Lave & Train (1979) and Manski & Sherman (1980) present some of the strongest research from among the early years of disaggregate vehicle type choice modelling. Many of the methods they use and issues they either note or address continue to be relevant and applicable to all such models, and their work provides valuable guidelines for development of new models.

Lave & Train develop a model of vehicle class choice for new vehicle purchases using 1976 data from seven US cities. They are the first to develop a class choice model that explicitly incorporates the properties of the vehicle that will be bought, whereas previous models rely on non-varying household demographic and socioeconomic attributes that simply have different alternative specific parameters for each choice alternative. This obviously adds conceptual realism to the model by properly accounting for the benefits of buying a particular vehicle as a component of the decision to buy it, but it also adds some challenges in terms of model estimation, which will be addressed below.

Since the purchase data they use only has information on what vehicles were actually selected by the household, but not those they did *not* select, there is a need to develop a representation of all vehicles in the choice set. To achieve this, all vehicles in the dataset are categorised into their appropriate class category, and then averages are taken of all key physical properties for the vehicles within that class. These average values are assumed to be representative of all vehicles in that class, and thus used to populate the vehicle-property related attributes in their corresponding utility functions for all ten choice alternatives that make up the vehicle market. In other words, the vehicle market is assumed to consist of only ten vehicles (which happen to feature the average values of all properties for the vehicles of each class).

Although this is a reasonable way of representing unobserved choices in the model, it does present difficulties in terms of model estimation. As the vehicle choice set of ten classes is identical for every household, vehicle properties cannot be directly used as attributes in the utility functions because there is no means of identifying *which* of these properties are the ones that are driving decision-making, as they are all collinear. In effect, all of the vehicle properties become bundled into the alternative specific constants for each class.

To circumvent this, Lave & Train create variables that interact household demographics and socioeconomics (which vary from household to household but not from class to class) with the vehicle properties (which vary from class to class but not from household to household). These combined household-and-vehicle-class variables provide a variation in utility levels amongst different vehicles and different households, thereby removing the issues of co-linearity and allowing the impact of the interacted attributes to be identified. Examples of these interacted parameters include functions that relate vehicle cost divided by household income, auto "performance" combined with respondent age or auto weight combined with respondent education levels. Other vehicle-specific or household-specific variables also appear, but use alternative specific parameters, similar to what previous models did. One vehicle property that is not interacted with household properties is fuel efficiency. Instead, it is used in the context of modelling mileage costs; the fuel efficiency (which is constant for each class) is multiplied by the price of gasoline, which varies from city to city, thus creating the required variation in the data that makes identification possible.

Lave & Train use a multinomial logit structure to model the choice decision. The resultant model has a moderate statistical fit, but as the authors themselves note may suffer from violations of the Independence from Irrelevant Alternatives property of the multinomial logit. Specifically, they give the example that if one class of vehicle is removed from circulation, the model would predict that these would-be purchases would distribute themselves proportionality to existing market shares amongst the other nine classes of vehicle still available. In practice, it would be expected that the vehicle classes most similar to the one that had been removed would capture the majority of the new market share.

Manski & Sherman use some of the same approaches as Lave and Train, in the sense that their utility functions for class choice utility functions are largely populated with variables that interact household properties with vehicle properties. However, their model also makes three key changes from that of Lave and Train, one of which results in them actually estimating two separate models; one for one-vehicle households and one for two-vehicle households.

The first major change is their representation of the choice alternatives that households are able to examine. Lave & Train create an "average" vehicle for each class, and all households pick from among these ten. Manski & Sherman use a completely different approach. They use a much larger "real world" vehicle class choice set of around 600 different vehicles. Since a model that featured a choice set of hundreds of different choice alternatives would be computationally burdensome to estimate they instead they create a choice subset for each household, that consists of the vehicle they actually purchased plus twenty-five vehicles randomly selected from among the 600 alternatives. They note that this procedure has been found to product consistent parameter estimates compared to estimating a full choice set while greatly reducing computational burden.

The second major change from Lave and Train relates to the scope of vehicle properties included in the utility functions, rather than model theory or procedure. Lave and Train use relatively few vehicle properties in their model: cost, weight, horsepower (converted into "performance" by dividing it by the weight) and fuel efficiency. In contrast, Manski & Sherman make use of an extensive number of vehicle properties, including number of seats, vehicle weight, luggage space, acceleration, turning radius, braking power, noise level, purchase cost, operating cost, age and whether it is a foreign make. The result is that their models contain 34 and 42 variables for the one and two vehicle households, respectively. These additional properties can serve to improve model fit, but create a large burden for using the model for forecasting purposes, as all of these individual properties themselves must be forecast, potentially introducing more error than it removes through improved model fit.

The third major innovation made by Manski & Sherman is the development of separate models for one vehicle and two vehicle households, as was noted above. The thought is that in multi-vehicle households, households consider the two vehicle types jointly. In other words, they can have vehicles that "specialize" for a particular function that they need a certain type of vehicle for. In contrast, in single vehicle households, the vehicle type purchased must be a "jack of all trades" type of vehicle, unless the household has plans to add to their fleet in the near future. The model for one vehicle households is similar to that developed by Lave and Train, with the exception of the first and second innovations discussed above. In contrast, the second model pairs two different class combinations together as a single choice; utility functions for these two-vehicle choice alternative then incorporate some of the vehicle properties of each vehicle. This approach has the benefit of improving the realism of vehicle type choices among two-vehicle households compared to a model where a household purchasing a second vehicle doesn't "know" about their first vehicle when selecting vehicle type for their second vehicle. However, for ownership levels of more than two vehicles, it can quickly become both econometrically and computationally burdensome to model vehicle type choices simultaneously. Ultimately, both models provide strong fits with most parameters being reasonable in size and magnitude, with a few curious exceptions (such as higher operating costs being a desirable feature for some segments of the population).

Berkovec & Rust (1985) present a similar work to Manski & Sherman in the sense that they have a multi-hundred vehicle choice set of individual makes and models as part of a vehicle type choice model, although they limit their model to one-vehicle households. Unlike Manski & Sherman, they use a nested logit model and group each make and model into different class nests. This allows them to estimate from the full choice set while avoiding the IIA restriction of the multinomial logit. Unfortunately, despite avoiding global IIA violations, there were still unable to produce strong results and suggest further work be undertaken.

Choo & Mokhtarian (2004) use data from a 1998 survey of 1904 households in the San Francisco Bay area to assess the impact of individual personality and attitudes relating to travel and lifestyle as well as the impact of both objective and subjective mobility metrics on vehicle class choice. Note that strictly speaking they do not develop a vehicle type choice model, but rather a model of which class of vehicle is most often driven within the household fleet. However, as the vehicle most often driven is obviously related to the selection of that vehicle for purchase, the findings of this model can reasonably be assumed to apply to a "true" vehicle class choice model as well. Attitudes were collected based a multi-point Likert style scale of agreement or disagreement with a particular statement, and were subsequently consolidated into a number of key attitudinal factors, with the selection of these factors based on the degree to which they correlate with different vehicle class types. Finally, a multinomial logit model of nine vehicle class choices was estimated, with utility function attributes for each choice consisting of a mixture of the attitudinal factors as well as household demographics and socioeconomics. Notably, other than an alternative specific constant, no class-specific factors were included in the model through interaction with household-related variables, because such data was unavailable. Despite this, the resultant model produced a moderate fit, with a number of statistically significant attitudinal factors present. As the authors note, there are some curious findings; for example, pro-high density attitudes correlate with an increased likelihood of owning an SUV, somewhat contracting the belief that higher density urban lifestyles go hand in hand with environmental consciousness and small, fuel efficient vehicles.

Cao *et al.* (2006) extend the above analysis somewhat further, using the same Northern California dataset described in the review of (Cao *et al.*, 2007), with a focus on the choice of heavier vehicles versus conventional passenger cars. They note that descriptive statistics suggest a correlation between smaller vehicles and more traditional urban development patterns (and thus between larger vehicles and suburban sprawl), however seek to investigate to what degree this is a function of self-selection driven by household attitudes versus objective built environment factors. They use the same data on attitudes towards both households' current and preferred neighborhood, as well as travel, and again consolidate these into a number of different attitudinal factors. These are then combined with household demographics and socioeconomics as well as objective measures of neighborhood type and travel patterns. Their model considers four vehicle class choice alternatives (car, van, SUV, pickup) and they estimate six different specifications of nested logit model, with the preferred model having a very strong fit. They conclude by noting that even taking into account both household and attitudinal factors, objective built environment measures (particularly the measure of "outdoor spaciousness") do have an influence on vehicle type choice. The implication is that land use policies that increase density can still have an impact in terms of reducing the size (and thus fuel consumption and pollution) even after secondary behavioural responses (such as some households moving to a different and more suburban neighborhood, and the resultant longer commutes resulting in some households being more likely to buy an SUV) are accounted for.

Potoglou (2008) uses the same CIBER-CARS dataset that was used for the vehicle holdings model developed by Potoglou & Kanaroglou (2008) to develop a model of vehicle type choice. Four vehicle types were represented in the model; namely passenger vehicles, vans, pick-up trucks and sports utility vehicles. Household composition, tenure type, urban form variables (the land use entropy and mixed density indices discussed previously) and work trip mode choice and education level of the respondent are included in this model as attributes. Three econometric model types were estimated, all of which performed relatively similarly. Overall model fit was moderate. Household composition and respondent variables were found to be statistically significant. The urban form measures were found to have a marginal impact on vehicle type choices, with households that live in areas with heterogeneous land uses being less likely to own an SUV. However, unlike Cao *et al.*, no discernible relation was found between commuting distance and the choice to purchase an SUV.

Bhat *et al.* (2009), Eluru *et al.* (2010), Paleti *et al.* (2011) and Vyas *et al.* (2012), all largely written by the same research group, present a number of similar models that feature vehicle types choices modelled in conjunction with other urban-related phenomena using the multiple discrete-continuous extreme value method for model estimation. Several of these papers relate to works developed for

SimAGENT, comprehensive urban modelling system for southern California similar in scope to ILUTE.

Bhat *et al.* model vehicle holdings and type choices in San Francisco in the context of demographics and the built environment, vehicle properties and gasoline prices. Results are generally consistent with the rest of the research already reviewed. In this paper, the built environment is considered an exogenous variable; unlike Choo & Mokhtarian (2004) or Cao *et al.* (2006), no self-selectivity is accounted for. Eluru *et al.* is similar in nature, modelling vehicle fleet size and type choices in conjunction with fleet use and residential location choices. However, it now accounts for the self-selectivity effect between location choice and vehicle ownership by modelling these decisions simultaneously.

The data collection context and transaction choice framework of Paleti *et al.* was discussed previously in Section 2.3. With regards to vehicle type modelling, the models developed for class and vintage are not substantially different than previous papers. However, they also feature possibly the strongest representation of fuel type choice modelling developed to date, with a model of up to seven different fuel choice alternatives considered, based on the results of their SP survey. A similar fuel type model should be seen as a long term goal for ILUTE, and could improve its ability to model vehicle emissions outcomes from policies that encourage the use of alternative fuel vehicles.

Finally, Vyas *et al.* model vehicle type choices (both class and vintage) as well as a means of allocation the primary driver of each vehicle in the fleet.

An aspect of vehicle type choice that has perhaps been understudied is the importance of vehicle origin. The role that vehicle origin plays in household vehicle type choice is unclear, but has implications for the role of vehicle origin in ILUTE. Mannering & Mahmassani (1985) suggest that the same vehicle properties are viewed differently by consumers depending on the whether the vehicle is a foreign or domestic manufacturer. In contrast, Train & Winston (2007) examine vehicle-type choice in the context of vehicle origin, with the aim of assessing the reasons behind the declining market share of domestic vehicles. Although they suffer from econometric challenges in developing the model, they conclude that vehicle manufacturer choice is largely driven by differences in physical vehicle features, rather than different weightings on the importance of features. Given that the work of Mannering & Mahmassani used data collected around the time of the 1979 oil crisis, it is felt that it may not necessarily represent "stable" consumer preferences. For this reason, the findings of Train and Winston will be adopted for the representation of vehicle origins in terms of their influence (or lack thereof) on consumer preferences.

Generally, results from the review of models for vehicle type choice suggest the following key insights:

- The choice set of different vehicle types is best represented by either creating "average" vehicles of each class, or by having a full choice set of makes and models and randomly selecting a smaller number of vehicles as choice alternatives.

- The importance of certain vehicle properties cannot be modelled in isolation, but must instead be interacted with household-related variables.

- Vehicle type decisions in multi-vehicle households should not be modelled in isolation, but rather in the context of the other vehicles already owned by the household. Transaction-based models can simplify this process as multiple vehicle type choices do not need to be estimated simultaneously if only one vehicle is being added to the household fleet at a time.

- Personal attitudes and the build environment both have a discernible impact on the type of vehicle that households purchase, although they play a secondary role to household demographics and socioeconomics.

- Vehicle origin is probably not an important factor in vehicle type choices, beyond certain manufacturers emphasizing certain physical features more than others.

Chapter 3

The Household Vehicle Fleet in ILUTE

This chapter provides the reader with background on ILUTE as a whole, as well as the various existing models within ILUTE that are either directly or indirectly related to the work undertaken in this thesis. The chapter is organized as follows:

- Section 3.1 discusses the overall ILUTE framework, and how the vehicle fleet model fits into the larger picture of the integrated urban modelling system being developed.
- Section 3.2 discusses the data collection effort that is used as the basis for the work undertaken in this thesis.
- Section 3.3 discusses the existing vehicle transaction model that was developed for ILUTE but never implemented.
- Section 3.4 discusses several other components of ILUTE that although strictly speaking are not part of the vehicle fleet model, are indirectly made use of or considered as part of this research.

The information presented herein is not intended to be a comprehensive review of each of these components. Rather it provides enough detail to understand how these other aspects of ILUTE were used or considered in the development of the research documented in this thesis. In all cases, readers seeking further information should refer to the original source.

3.1 The ILUTE Modelling Framework

The Integrated Land Use, Transportation and Environment (ILUTE) model simulates the evolution of urban environments over time (Salvini & Miller, 2005). It aspires to be the "ideal" integrated model that accurately represents all processes experienced in urban areas, and thereby overcomes many of the limitations of current state-of-practice models. To this end, the improvements made by ILUTE can be broadly classified into two categories: better representation of *what* urban processes are represented in the model and better representation of *how* human behaviour drives these processes.

Compared to the traditional 4-step model, ILUTE improves the scope of the phenomena that is modelled by expanding from a travel-only model into one that represents the factors that drive the demand for travel in the first place and how these are in turn themselves influenced by travel. This includes processes such as regional economics, demographic trends, land development and new construction, residential location choices and real estate markets, firmographic processes, vehicle ownership and vehicle emissions models (Salvini & Miller, 2005). Figure 3.1 shows the overall design of ILUTE in terms of how the many separate sub-processes are linked together.

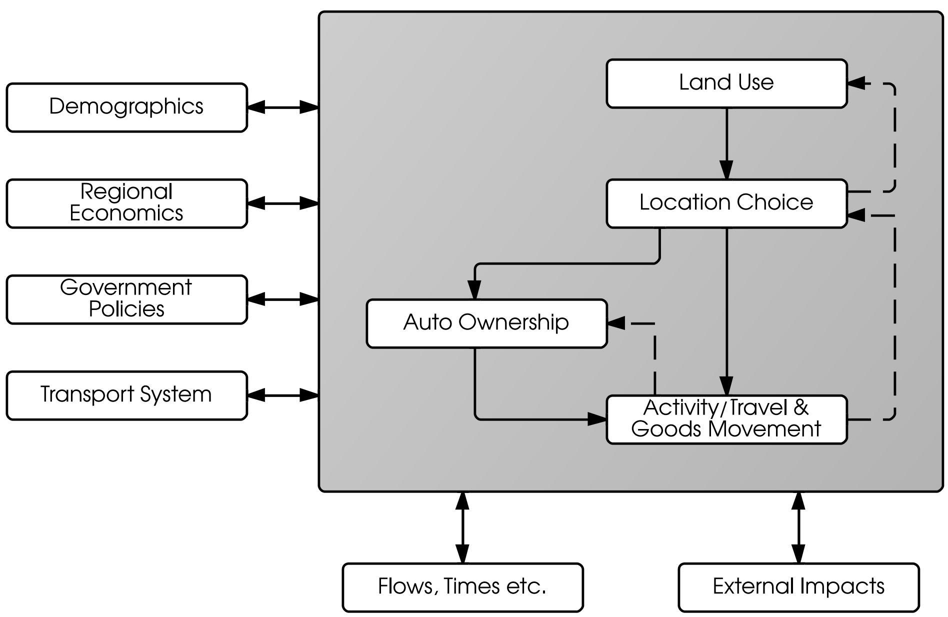

Figure 3.1: Scope of ILUTE modelling system, adapted from Miller *et al.* (1999)

Note that each of the processes represented by these "boxes" will in turn have a separate series of sub-models that can each generate a wide range of output. For example, the auto ownership box, which is essentially what this thesis seeks to create, includes more information that just the number of vehicles a household owns (and is referred to hereafter by the more inclusive term of "household vehicle fleet").

The second major advancement ILUTE offers compared to the state of practice is how these processes are driven by human behaviour. By using a fully agent-based microsimulation environment, ILUTE can represent all of the above process at a fully disaggregate level that allows the mathematical models being implemented to mimic human behaviour as close as possible. This ties in well with the expanded scope of what is modelled in ILUTE. For example, in a traditional four-step model, work trips for commuters are distributed access the region using a gravity model. While this may provide broadly correct aggregate results, it does not actually accurately represent the choice processes that individuals use to make decisions. In contrast, ILUTE is able to model the actual choice of place of employment for an individual based on their individual context (e.g. education level, salary offered, potential travel required etc.), and then based on their choice of employment location, their work trip destination is then fed into the travel model.

The microsimulation aspect of ILUTE also improves forecasting applications by being able to provide simulation output at a regular time step interval (Salvini & Miller, 2005). Between this and its ability to represent 24 hour travel, it means that instead of providing a single peak period of travel for a single year 10-20 years in the future as a forecasting output, ILUTE can provide full 24-hour travel patterns for each and every year.

Due to the theoretical advantages it offers, once implemented, ILUTE should provide more accurate forecasting results, the ability to better use modelling to support policy analysis and a much greater amount of output data to assess the effects of different transport investments.

Full documentation of the ILUTE modelling framework can be found in Salvini (2003).

3.2 Toronto Area Car Ownership Survey

With the household vehicle fleet having been identified as a key mechanism within the Integrated Land Use, Transportation and Environment model, developing a model to account for this behaviour was one of the early priorities of ILUTE.

The first step in developing such a model is data collection. To this end, a survey on household vehicle fleets entitled the Toronto Area Car Ownership Survey (TACOS) was conducted in the spring of 1998 by Roorda (1998).

TACOS was developed as a retrospective survey of vehicle fleet decisions over a nine year time period between 1990 and 1998. Having a repeated survey of the same household on a year-over-year basis was felt to be important to properly account for the temporal dynamics of vehicle fleet decisions. The only two types of surveys that are able to achieve this are panel surveys and retrospective surveys. Since conducting a nine-year panel survey would take nine years, this is obviously infeasible (not to mention expensive). Instead, a retrospective survey was selected. In a retrospective survey, households are contacted for interview and asked to provide details not just for the present day, but dating back for a certain time frame (in this case, nine years). Retrospective surveys have the ability to provide the same information as panel surveys while also collecting data relatively quickly (i.e. in a single interview per household, and over the course of a few months for the entire TACOS dataset). The main risk associated with them is that persons may not remember information correctly from such a long time ago. However, given that vehicles are relatively major purchases for households, it was felt that generally this is the type of purchase that a household *would* remember even if the vehicle had long since been disposed (Roorda *et al.*, 2000).

The TACOS survey can be roughly categorized into two separate components; demographics and vehicles. The demographic information is collected at three separate levels:

- Firstly, it is collected at the Person level. At this level, information is collected on year of birth, sex, education level, place of employment/study for each year between, occupation type(s) between, employment skill level, possession of a driver's license etc. This information is collected for each and every year (i.e. 1990-98) that the Person is within the survey scope.

- Secondly, information is collected about the Decision-Making Unit (DMU). The DMU is the collection of Persons that are considered as a group when making a decision. For example, a family of four living together would be a single DMU whereas a house with four non-related roommates would be four separate DMUs. Information is collected on which Persons in the household belong to which DMU, as well as DMU-as-a-whole information such as Income level, dwelling type and dwelling location (i.e. if the DMU has moved at any point).

- Finally, at the household level. The household refers to the all persons living in the physical structure. They do not make decisions together (because they may be part of different DMUs); but collecting information on other DMUs within the same household provides an efficient means of data collection.

The second component of data collection concerns all of the vehicles owned by members/DMUs of the household over the 1990-1998 timeframe. The vehicle component details on each of the vehicles owned by the population. Vehicle-related questions included the make, model and year of each vehicle, the reasons why that particular type of vehicle was acquired and/or disposed of, why it was bought/disposed of at that particular time, when it was acquired and disposed (if applicable), whether it is a personally owned or company owned vehicle and what type of fuel it uses. To link to the demographic information, which Persons within the household were the primary and secondary owners and drivers of each vehicle were also queried.

Note that for the remainder of this thesis, the term "household" refers to what TACOS calls a DMU. The TACOS concept of households (i.e. a group of DMUs that live in a common dwelling) is not used in this research. The reason for this change is that the term household is a more intuitive concept to understand, and also that the travel model used in this thesis (TASHA) uses the term household to refer to what TACOS refers to as DMUs. Thus, changing the terminology also avoids the confusion of constantly switching between the term "DMU" when referring to TACOS information and "household" when referring to TASHA information.

Roorda then conducted a high level review of findings (mainly using descriptive statistics) to provide preliminary guidelines for model development.

Mohammadian (2002) examined the demographic cross-section captured by the TACOS survey against 1996 Census data. TACOS was compared to the Census on the basis of:

- Tenure type (i.e. owner versus renter)
- Dwelling type
- Income range
- Education level
- Sex
- Household size
- Occupation type

Generally, in all comparisons the TACOS data matched the Census within reasonable limits, and it was concluded that the data represented a relatively unbiased sample of the population.

TACOS forms the primary source of data for both the transaction model developed by Abolfazl Mohammadian (to be discussed in Section 3.3 below) as well as the work undertaken in this thesis.

Full documentation of the TACOS survey can be found in Roorda (1998).

3.3 Vehicle Transaction Model

Subsequent to the collection of the TACOS dataset, an initial vehicle transaction model was developed in 2002 by Abolfazl Mohammadian. This transaction model (referred to hereafter as the "original" model) forms the starting point for the work undertaken in this thesis. Specifically, the work in this thesis acts to update and expand the original model developed by Mohammadian so that it can both make use of and integrate itself within the scope of ILUTE, which has become progressively more defined since the completion of the original work in 2002.

The development of the original vehicle transaction model can roughly be divided up into three distinct tasks:

1. Development of vehicle properties and costs attributes for use in the model
2. Specification of a model structure
3. Model estimation details

Each of these tasks is discussed separately below. Note that this order is not necessarily the same order that these tasks were presented by Mohammadian in his dissertation. Instead, they have been rearranged slightly to present the information in a manner that flows better in the context of what information is actually presented herein, rather than the entirely of Mohammadian's research.

3.3.1 Vehicle Properties and Hedonic Price Model

The first step in the vehicle transaction model developed by Mohammadian was to develop a series of vehicle properties and costs that could be used for modelling purposes. Since vehicle makes were collected in the TACOS survey, this information could be supplemented by looking up vehicles in the *Canadian Vehicle Specification System* (CVSC). Table 3.1 shows a series of directly available and derived vehicle properties that Mohammadian developed for each vehicle in the TACOS database by using the CVSC.

Exploratory data analysis by Mohammadian noted that many of the physical/performance properties of the vehicle (i.e. the first five of the properties listed above) are highly correlated with each other. Developing models with highly co-linear data is likely to increase estimation errors, as statistically it becomes challenging to differentiate which of the two or more highly correlated properties acting as independent variables are responsible for the resultant values of the dependant variable.

Mohammadian therefore used the principal component analysis technique to reduce the correlation between the various vehicle properties. In essence, principal component analysis creates a series of new variables to consolidate the various co-linear variables. These new variables are developed in such a manner that they explain as much of the overall variance as possible, but are not correlated with each

Table 3.1: Vehicle Properties used in Price Model Development

Characteristics	**Unit**
Engine Displacement	Litre
Weight	Tonne
Fuel Intensity	L/100km
Luggage Capacity	m^3
Wheelbase	metre
Age	Vehicle model year versus current year
Class	Subcompact, Compact, Midsize, Large, Wagon, Van, Sport Utility Vehicle, Pickup
Origin	Domestic, Japanese, European
New	If vehicle is a Brand New vintage
Luxury	If the vehicle is a luxury vehicle
Size	If the vehicle is a wagon, van or sport utility vehicle
Cargo	If a vehicle is a cargo-van or a pick-up
Space	$(1 - Cargo) \times LuggageCapacity \div Wheelbase$

other. In this case, Mohammadian consolidated various correlated physical vehicle property variables into two separate factors, which he titled the Vehicle Performance Factor and the Vehicle Space Factor. These are shown in Table 3.2.

Table 3.2: Principal Component Analysis of Vehicle Properties

Variable	**Factor Loading**	
	Vehicle Performance Factor	**Vehicle Space Factor**
Weight	0.339	0
Engine Displacement	0.413	-0.166
Fuel Intensity	0.357	-0.035
Size	-0.049	0.515
Space	-0.157	0.596
% Variance explained	53%	36%

Between them, the two variables represent 89% of the total variance of the properties of vehicles in the TACOS dataset.

With a statistically stronger means of representing physical vehicle properties now available, Mohammadian used this information to develop a means of calculating the expected market price of vehicles. The dependant variables for this task include these two principal component analysis variables as well as several other vehicle properties listed in Table 3.1. As the purchase price of household vehicles was not recorded in TACOS, Mohammadian collected market prices for the corresponding make and model of the vehicle from the *Canadian Red Book Vehicle Valuation Guide*.

Using this data rather than directly reported prices has the benefit of reducing the potential for reporting/memory errors and self-selection bias. With all vehicle properties and market price values available, Mohammadian then estimated the hedonic price model for vehicle value shown in Table 3.3.

Table 3.3: Hedonic Vehicle Price Model (1000's, $1998)

Variable	**Coefficient**	**Standard Error**	**t-statistic**
Subcompact	12.753	0.423	30.135
Compact	13.767	0.381	36.11
Midsize	15.288	0.375	40.718
Large	16.395	0.478	34.285
Station Wagon	13.0141	0.744	17.489
Van	12.897	0.690	18.694
Pickup	11.615	0.613	18.954
Sport-utility	14.974	0.667	22.454
Luxury	26.086	0.745	34.996
New	0.779	0.299	2.604
Natural Logarithm of Vehicle Age	-6.072	0.166	-36.651
Japanese Car	3.746	0.260	14.392
European Car	4.484	0.494	9.068
Vehicle Performance Factor	2.294	0.168	13.628
Vehicle Space Factor	1.038	0.273	3.800
Time[1]	0.514	0.421	12.216
Adjusted-R^2 = 0.82			

[1]Measured in years to/from 1990.

The final step in developing the vehicle property and cost information is the development of values for Operation and Maintenance Costs. Mohammadian developed a model that considered four separate types of costs in the development of O&M expenses. These costs are:

- Maintenance costs
- Fuel costs
- Insurance
- Asset depreciation

Vehicle registration and drivers licensing fees were not included as O&M costs, due to insufficient information at the time the model was being developed. However both registration and licensing costs are relatively minor in magnitude compared to the others, and are not expected to have a substantial impact on vehicle fleet decision making. Based on the 2012 *Driving Costs* guide published by the Canadian Automobile Association, they generally are in the range of 1%-2% of total O&M costs (Canadian Automobile Association, 2012). Furthermore, in the case of a driver's license, many people may possess one even if they don't own a vehicle, but rather for those rare instances where the need to borrow/rent one, or simply use it as a useful form of identification.

The original Operation and Maintenance cost estimates (except Insurance) were calculated by Mohammadian using the *Highway Design and Maintenance Standards* (HDMS) developed by the World Bank. Part of this study involved developing a Vehicle Operating Cost (VOC) model, which can be used to estimate the maintenance, fuel and depreciation costs of a given vehicle. Although there are upwards of seventy different parameters that can be input into the model, most of them were left as their default values (Mohammadian, 2002). Key input assumptions that were made by Mohammadian include:

- Average service life of vehicles are assumed to be 10 years
- 20,000 km of driving annually per vehicle
- 500 hours of driving annually per vehicle

This model was run on all vehicles in the TACOS database in order to obtain the corresponding category of Operating and Maintenance Costs of each one. These O&M costs were then averaged across the twenty-four class and vintage categories supported by the model, in order to obtain standard rates for each potential vehicle type choice that will be estimated and simulated.

Insurance costs were calculated using a sampling of quotes rates from online insurance providers. Insurance rates vary based on characteristics of the driver, the vehicle and travel patterns. As the focus for developing the insurance model was how it varied with vehicle characteristics, a standard person and commute was applied to all quotations (Mohammadian, 2002). Quotations were then obtained for different class and vintage combinations, which were then used to develop a regression model for monthly insurance rates. These were then converted into annual rates for modelling purposes.

The final class-vintage specific Operating and Maintenance Costs developed by Mohammadian are shown in Table 3.4.

3.3.2 Model Structure

With the vehicle properties and costs fully specified, Mohammadian then focused on the assessing the preferred model structure for the vehicle transaction model. Several different modelling strategies and configurations were investigated.

Mohammadian undertook an extensive investigation of the relative benefits of vehicle holdings modelling versus transaction modelling. He ultimately settled on the latter for many of the same reasons that were given in Section 2.3.

In terms of mathematical modelling structure, several different model types were tested. These include nested logit models, random parameters logit (also known as mixed logit) models and Artificial

Table 3.4: Annual Operation and Maintenance Costs ($1998, 1000's)

Class	Vintage	Maintenance	Fuel	Depreciation	Insurance	Total
Subcompact	Brand New	0.090	0.765	2.129	4.003	**6.987**
	Second Hand	1.064	0.777	1.518	3.790	**7.149**
	Used	1.361	0.785	0.833	3.260	**6.239**
	Old	1.610	0.816	0.367	2.490	**5.283**
Compact	Brand New	0.090	0.828	2.117	3.552	**6.587**
	Second Hand	1.038	0.829	1.607	3.340	**6.814**
	Used	1.324	0.833	1.048	2.918	**6.123**
	Old	1.544	0.838	0.432	2.281	**5.095**
Midsize	Brand New	0.090	0.901	3.783	3.783	**8.557**
	Second Hand	1.214	0.921	2.535	3.440	**8.110**
	Used	1.573	0.923	1.431	2.888	**6.815**
	Old	1.876	0.994	0.469	2.182	**5.521**
Large	Brand New	0.090	1.011	3.036	3.600	**7.737**
	Second Hand	1.230	1.105	2.775	3.381	**8.491**
	Used	1.600	1.083	1.762	2.944	**7.389**
	Old	1.917	1.148	0.499	2.318	**5.882**
Station Wagon	Brand New	0.090	0.864	2.080	3.360	**6.394**
	Second Hand	1.081	0.891	1.836	3.163	**6.971**
	Used	1.385	0.898	1.279	2.703	**6.266**
	Old	1.646	0.984	0.391	1.896	**4.917**
Special Purpose Vehicle	Brand New	0.090	1.072	2.218	3.178	**6.558**
	Second Hand	1.150	1.080	1.759	3.499	**7.488**
	Used	1.489	1.119	1.247	3.604	**7.459**
	Old	1.780	1.159	0.483	2.274	**5.696**
Van	Brand New	0.090	1.074	2.368	3.295	**6.827**
	Second Hand	1.118	1.082	1.847	3.081	**7.128**
	Used	1.431	1.087	1.830	2.654	**7.002**
	Old	1.698	1.930	0.510	1.930	**6.068**

Neural Networks. Ultimately, the nested logit model structure shown in Figure 3.2 was decided on.

The model structure developed by Mohammadian has an upper level transaction choice set, and low level choice sets for the type of vehicle that is purchased and/or disposed, depending on the transaction choice. Only one transaction decision may be made per year.

Note that the model has been revised to use the word "Replace" instead of "Trade" as Trade was felt to convey a particular means of disposal (i.e. trading the vehicle in to a used-vehicle dealership for credit towards a new purchase). Conversely, Replace is more linguistically neutral as it only implies that a vehicle was removed and another was added to the household fleet, and makes no mention of disposal method (e.g. used-car dealership, private sale, scrapyard etc.).

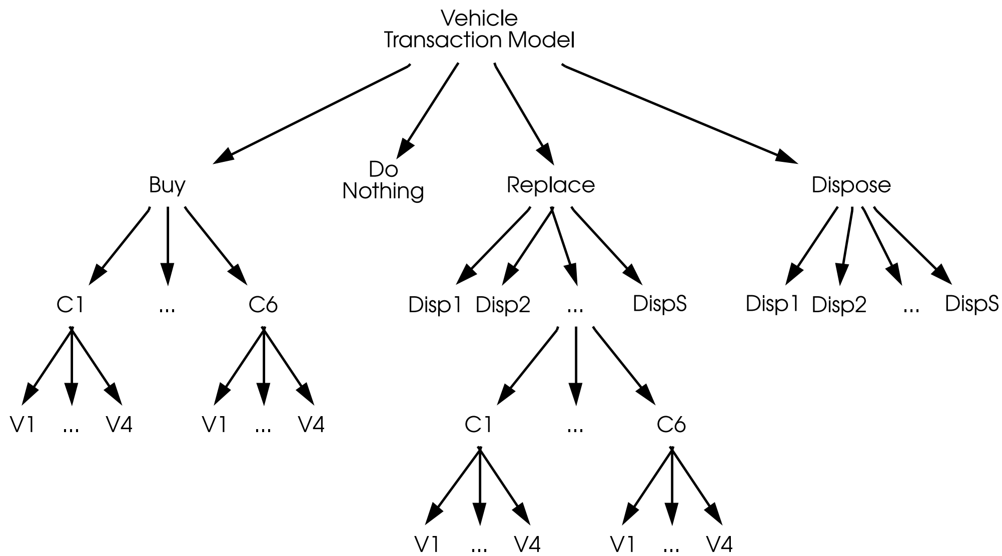

Figure 3.2: Transaction Model Conceptual Choice Structure, adapted from Mohammadian (2002)

3.3.3 Model Estimation Results

The model shown in Figure 3.2 was estimated in several steps, to avoid the difficulties associated with estimating a single model of this large scope all at once.

Mohammadian first estimated a two-level nested logit model for the class and vintage choices. The results are presented in Table 3.5.

Table 3.5: Class and Vintage Choice Model for Buy and Replace Transactions

Explanatory Variable	**Alternatives**	**Coefficient**	**t-statistic**
Vintage Choice			
Brand New vehicle (vintage I) constant	V1	7.347	4.184
Second hand vehicle (vintage II) constant	V2	10.424	6.492
Used vehicle (vintage III) constant	V3	11.601	10.296
Natural logarithm of (market price divided by (income - ownership & operating cost of current household fleet))	All	-0.975	-1.931
Class-Vintage average ownership cost	V1, V2, V3	-1.251	-12.127
Average market price of household fleet	V1, V2	0.101	4.183
Natural logarithm of average age of people in HH	V2	0.51	2.027
Average age of household fleet	V1	-0.102	-3.483
Average age of household fleet	V4	0.067	1.625
Average length of ownership in household fleet	V3, V4	-0.148	-3.475
Natural logarithm of driver's age	V1	1.409	4.324
No. of age code III (used) vehicles in household fleet	V4	-1.058	-2.605

continued on next page...

Table 3.5: Class and Vintage Choice Model for Buy and Replace Transactions (continued)

Explanatory Variable	**Alternatives**	**Coefficient**	**t-statistic**
No. of age code IV (old) vehicles in household fleet	V3	0.383	2.047
No. of people with elementary level of education in HH	V2	0.369	2.305
No. of people with B.Sc. degree in HH	V1	0.411	3.339
No. of people with graduate degree in HH	V4	-0.841	-1.765
Trade transaction dummy	V1	0.473	2.041
No. of employment type 3 (health and medicine) in HH	V2	0.882	3.271
Owner's highest completed level of education	V3	-0.246	-2.411
Class Choice			
Sub-compact (class 1) constant	C1	1.785	2.414
Compact (class 2) constant	C2	2.113	2.983
Mid-size (class 3) constant	C3	0.682	1.002
Large (class 4) constant	C4	-0.236	-0.339
Special-purpose (class 5) constant	C5	3.825	2.875
Market price divided by natural logarithm of income	C1, C2, C3, C4, C6	-0.112	-1.864
Vehicle performance factor	All	0.603	4.53
Vehicle space factor	C5	2.221	4.784
Vehicle space factor	C6	7.583	7.813
Driver has skill level 1 (manager)	C3, C4, C5	0.67	2.499
Driver has skill level 2 (professional)	C1	0.546	2.459
No. of class1 (sub-compact) vehicles in household fleet	C1	0.49	2.667
No. of class3 (mid-size) vehicles in household fleet	C3, C4	0.69	3.683
No. of class4 (large) vehicles in household fleet	C4	1.417	4.709
No. of class6 (special purpose) vehicles in household fleet	C5	0.699	2.957
Driver is male	C4, C5, C6	0.82	4.071
Average weight in fleet (metric ton)	C1, C2	-0.519	-2.986
Driver's highest completed level of education	C5	-0.263	-2.375
Natural logarithm of average age of people in HH	C5	-1.226	-4.02
No. of children divided by no. of people in HH	C6	1.78	2.625
Inclusive Value Parameters			
Inclusive Value for Sub-compact	C1	0.657	7.359
Inclusive Value for Compact	C2	0.621	7.902
Inclusive Value for Mid-size	C3	0.832	9.149
Inclusive Value for Large	C4	0.587	6.093
Inclusive Value for Special-purpose vehicle	C5	1	Fixed
Inclusive Value for Van	C6	0.813	8.765
Log-likelihood at zero: -1897.298			
Log-likelihood at constants: -1789.717			
Adjusted log-likelihood ratio ($\bar{\rho}^2$): 0.263			

Next, he estimated two vehicle disposal choice models, one for Replace transactions and one for Dispose transactions, as shown in Tables 3.6 and 3.7, respectively. Both of these models are multinomial logit models, where the choice set size is equal to the number of vehicles in the household fleet. Note that for Replace transactions the inclusive value of the Class and Vintage choice enters into the disposal choice utility functions, allowing the type of vehicle the household will add to their fleet to influence the one they choose to remove.

Table 3.6: Vehicle Disposal Choice Model for Replace Transactions

Explanatory Variable	**Alternatives**	**Coefficient**	**t-statistic**
Small car constant	C1, C2, C3	0.483	1.551
Van constant	C6	-1.025	-1.128
Natural logarithm of market price of car divided by (Vehicle age - 1)	All	-0.299	-1.646
Luggage capacity divided by wheelbase of vehicle	C1, C2, C3, C4, C6	2.299	1.915
Class-Vintage choice Inclusive Value	All	0.244	1.894
Log-likelihood at zero: -924.814			
Log-likelihood at convergence: -96.852			
Adjusted log-likelihood ratio ($\bar{\rho}^2$): 0.892			

Table 3.7: Vehicle Disposal Choice Model for Dispose Transactions

Explanatory Variable	**Alternatives**	**Coefficient**	**t-statistic**
Second-hand car (1-2 years old) constant	All	3.406	1.309
Used car (3-7 years old) constant	All	4.085	2.012
Old car (8+ years old)	All	5.006	1.894
Natural logarithm of market price of car divided by (Vehicle age - 1)	All	-1.457	-1.826
Luggage capacity divided by wheelbase of vehicle	C1, C2, C3	4.152	1.589
Log-likelihood at zero: -336.874			
Log-likelihood at convergence: -34.952			
Adjusted log-likelihood ratio ($\bar{\rho}^2$): 0.890			

The final step in the model is to estimate the transaction choice decision. This model was estimated as a multinomial logit with four choices (although the Replace and Dispose choices will obviously not apply to a household with no vehicles). The resultant model is shown in Table 3.8. Note that the logsum values of expected utility from the Class-Vintage Choice (for Buy transactions) and Disposal Choice (from their appropriate respective models) influence transaction choices.

Table 3.8: Transaction Choice Model

	Do Nothing		**Replace**		**Buy**		**Dispose**	
Explanatory Variable	**Coef.**	**t-stat.**	**Coef.**	**t-stat.**	**Coef.**	**t-stat.**	**Coef.**	**t-stat.**
HH and Fleet Attributes								
Natural logarithm of (total market price ($1000's) of fleet + 1)	0.498	5.504						
Dummy for income > $75000	-0.34	-2.418						
No. of driver's licenses = fleet size	0.384	2.666						
No. of driver's licenses > fleet size					1.03	3.676	-0.869	-2.321
Fleet size			0.301	2.239			0.52	2.435
No. of adults in HH					0.292	2.513		
Changes in HH Attributes								
No. of jobs in HH+			0.317	1.935	0.317	1.935		
HH size+			0.442	1.619				
HH size-							1.608	2.928
Previous Transactions								
Trade Smooth factor (λ = 0.6)			-1.406	-2.566	-2.306	-2.645		
Buy Smooth factor (λ = 0.2)					1.688	1.729	2.397	2.177
Constants and Inclusive Values								
Alternative specific constants			-1.937	-8.386	-6.222	-13.82	-4.038	-8.518
Inclusive value			1	fixed	0.603	19.7	1	fixed
Percent predicted correct		95.60%		59.00%		61.60%		61.70%
Log-likelihood at zero: -5349.7								
Log-likelihood at convergence: -644.5								
Adjusted log-likelihood ratio ($\bar{\rho}^2$): 0.879								

Full documentation of original vehicle transaction model framework can be found in Mohammadian (2002).

3.4 Related Work

Despite not being directly part of the vehicle fleet model, there are a number of other works that influence the development of the vehicle model in order to ensure that the model developed herein will be fully compatible with the rest of ILUTE for simulation purposes. The three most relevant of these other works within ILUTE are discussed below with a brief description given as to what they are as well as how they are used or influence the development of the vehicle fleet model.

3.4.1 Population Synthesis and Demographic Updating

The first step in an agent-based microsimulation environment such as ILUTE is the creation of a synthetic population at the beginning of the simulation period (Pritchard & Miller, 2011). The ILUTE population synthesis procedure is developed for 1986, and as such this is the first year of ILUTE that is simulated. The synthesis procedure is based on Statistics Canada 1986 census information. It makes use of public-use microsamples that contain complete census information for a small sample (usually 1%-4%) of the population as well as the aggregate totals for given geographic areas. These two datasets are combined using an iterative proportional fitting (IPF) procedure that creates a fully synthetic population that matches both datasets as best as possible (Pritchard & Miller, 2011).

The synthesis procedure in ILUTE provides a strong level of detail at the person, family and household levels, both individually and in terms of how they relate to each other. This provides a strong basis to undertake modelling for all of the other subcomponents of ILUTE, as decisions are made at times by individuals and at times by groups of individuals within the household (or a combination of both).

Person-level information generated from the synthesis procedure includes properties such as age, sex, education level, employment status and occupation type. Family-level information generally relates to family structure and summations of the person-level information and details regarding how different persons within the household relate to one another. Finally household-level information contains information on the dwelling (housing type, number of rooms, tenure type etc.) as well as the combinations of individuals and/or families living within it. Note that vehicle ownership level information for 1986 is not included in the synthesis, and will need to be developed separately.

As the synthesis procedure only models the first year in ILUTE (i.e. 1986), a means of representing agent demographics as they change over time is also required. A substantial amount of effort has been put into modelling and validating the change in regional demographics over time, and is detailed by Miller *et al.* (2011)

The key link between the demographics component of ILUTE and the vehicle fleet model is ensuring compatibility of definitions of individual persons and groups of persons. For example, consistent definitions of occupation categories or housing types enable the models estimated from TACOS data to be seamlessly used in ILUTE simulations without any variable definition "translation" errors.

Full documentation of population synthesis procedure and demographic information provided can be found in Pritchard (2008).

3.4.2 Activity-Based Travel Demand Modelling

The Travel/Activity Scheduler for Household Agents (TASHA) is an activity-based travel demand model developed that acts as both a standalone travel model as well as acting as the "T" component in ILUTE (Miller & Roorda, 2003).

Unlike conventional transportation models, TASHA is a fully disaggregated model that explicitly represents both households and all of the persons within them. Despite this, TASHA is able to be run solely using conventional travel survey data such as that provided by the Transportation Tomorrow Survey, with some supplementary schedule resolution rules provided based on household responses to the CHASE survey (Roorda *et al.*, 2008).

TASHA uses the TTS database to develop probabilistic distributions of daily activities that individuals may engage in (e.g. working or shopping). These are known as activity projects, and over a series of steps are combined together to create a daily schedule of activity projects for each person in the model. This daily schedule contains information on what activities the person will participate it, what order they will occur in, for what length of time, and where they will be located. Further details on this process can be found in Miller & Roorda (2003). Collectively, these steps are referred to in the remainder of this thesis as activity generation.

The next main step in the model is travel mode choice. One of the most significant improvements TASHA offers over conventional models is its improved representation of mode choice. Mode choice in TASHA is based on the notion of maximizing overall household travel utility over the course of the entire day. Thus it allows intra-household vehicle allocation to grant the use of vehicles to the person who benefits from it the most, but also allows for passenger modes, ridesharing and trip time adjustments/rescheduling, in addition to the standard suite of non-auto modes. Effectively, TASHA attempts to model household travel mode choice decisions as similarly as possible to how they occur in real life. Appendix A provides details on the actual mode choice calculation procedure used in the software.

Both activity scheduling and mode choice are iterative processes, as trips must be assigned to the regional transportation network, which then are used to iterate schedules and modes until a suitable level of convergence is reached. A flowchart of TASHA's modelling procedure is presented in Figure 3.3.

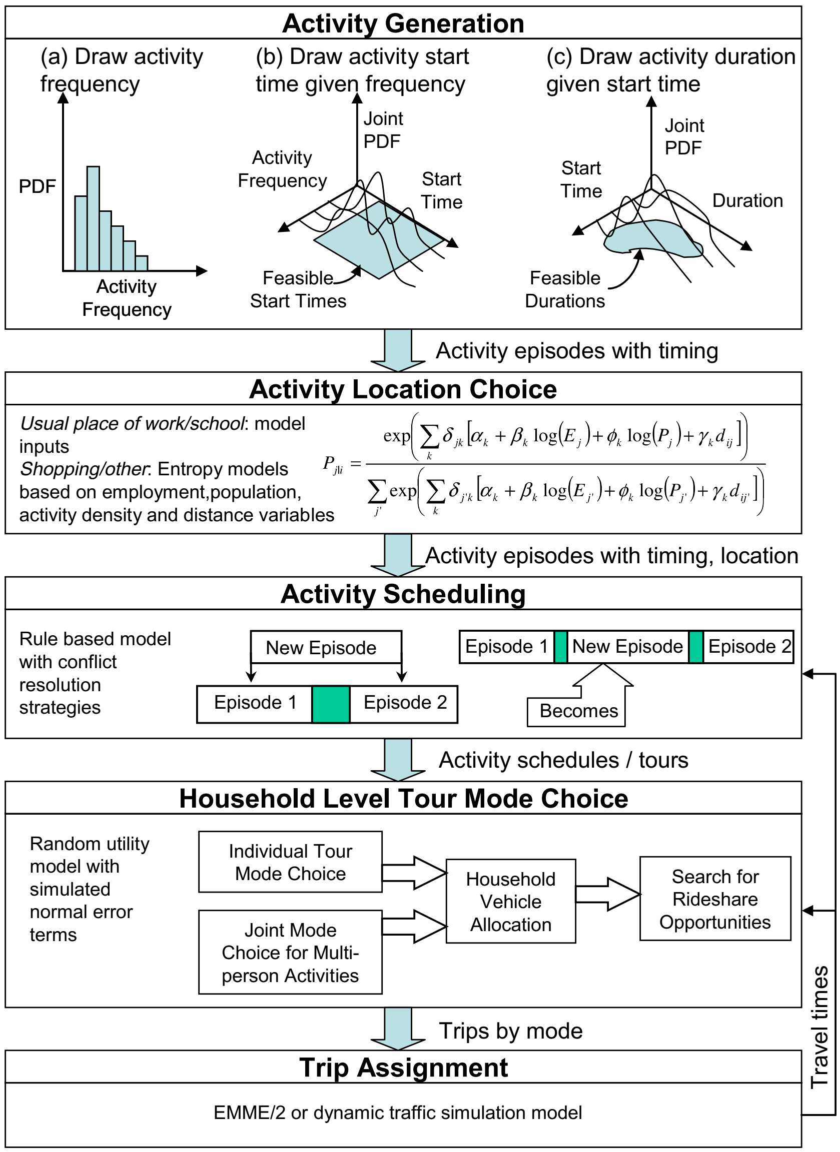

Figure 3.3: TASHA Modelling Procedure, Source: Roorda *et al.* (2008)

Note that in its current incarnation, TASHA treats all vehicles generically (since the TTS also does this). Thus, although it can allocate a vehicle or vehicles to different household members for various tours, there is currently no basis for selecting *which* vehicle is used by *which* household member (Roorda *et al.*, 2009). This is relevant to the application of TASHA (and the vehicle fleet model developed in this thesis) to emissions modelling, which are discussed below.

Similarly to what was done by Roorda *et al.* (2009) and discussed in the literature review, TASHA will be used in this thesis (albeit in an expanded manner) to generate household travel patterns that help inform vehicle transaction choices. This will be discussed further in Chapter 5.

Full documentation of TASHA can be found in Roorda (2005).

3.4.3 Vehicle Emissions Modelling

The final aspect of ILUTE that is considered in this thesis is the vehicle emissions and pollution dispersion model. This model is largely an "output" model in the sense that it does not feedback directly to influence other behavioural processes within ILUTE. As such, unlike the over works reviewed in this section, it is not directly used by the vehicle fleet model developed in this thesis. Rather, the vehicle fleet model was developed with the aim that it would create modelled data that could be used for emissions modelling.

The emissions model was created using Mobile 6.2C, the Canadian edition of the Mobile6.2 software developed by the United States Environmental Protection Agency (Hatzopoulou *et al.*, 2007). Due to the advantages that it offers over traditional 4-step models that were discussed previously, the emissions model makes use of TASHA to supply vehicle travel information.

As noted in Subsection 3.4.2, TASHA treats all vehicles generically; it does not differentiate among them by factors such as class, age or fuel type. However, the emissions model also requires information on vehicle fleets to supplement travel patterns. Currently the model makes use the 28-class vehicle fleet distribution developed by the USEPA. However, the vast majority of vehicles and vehicle emissions in major metropolitan areas fall into the single category of Light Duty Gasoline Vehicles (LDGV). This is therefore the main class of vehicle that was used in the vehicle emissions model, as it comprises the type of vehicle most likely to be accounted for in the TTS, and modelled in TASHA.

As noted by Hatzopoulou *et al.* (2007), the emissions model would benefit greatly from a vehicle fleet model that would interact with TASHA and which together would feed the emissions model with more disaggregate vehicle use information. This need ties in with the discussion in Subsection 1.1.2, which noted that one of the motivations behind this thesis is to provide a basis to significantly improve vehicle emissions modelling.

Documentation of the development of the emissions model was therefore reviewed to understand what output was specifically required from the vehicle fleet model. Specifically, two key inputs from the vehicle fleet model were identified as benefiting the emissions modelling procedure.

The first is improved representation of vehicle types. As noted in the introduction, the vehicle fleet model developed in this thesis will explicitly model vehicle class (modelled directly) and vehicle model year (modelled via the vintage choice). Both of these are useful inputs into the emissions model, and will help it move beyond classifying 90% of vehicles as being simply a LDGV. No representation of fuel type is able to be provided however; all vehicles will have a fuel intensity value assigned to them which can also help inform emissions rates. Not only will be vehicle types improve the overall region-level accuracy of the vehicle fleet, but because each and every household has their own vehicles with their specific classes and model years, disaggregate vehicle information is available for disaggregate travel information, and can be matched with each other, subject to the proper representation of improved within-household representation of vehicle use.

This leads to the second improvement, which is improved representation of within-household vehicle use. Even with the correct number, classes and years of manufacture of all vehicles, as well as correct overall household travel patterns from TASHA, there is still a matter of assigning *which* vehicle is used for *which* trip. As part of the vehicle emission research, a rule-based algorithm was developed that assigns vehicles to particular trips within the household, known as a car allocation model. Essentially, the model assigns a “preferred” vehicle to an individual who will use that vehicle whenever available, and only use others if it is not available (Hatzopoulou, 2008). In instances where there are more drivers than vehicles, individuals who are not assigned a preferred vehicle will choose at random. While this is an adequate procedure, it does not yet have a means of identifying which vehicle is used by which household member in the first place, the assignment of the “preferred vehicle” is still random. Thus, there is still a need to identify who within the household is the primary driver for each vehicle. As shown in Table 3.5 in Section3.3, the class and vintage model developed by Mohammadian makes use of the concept a “primary drive” as well; but it also lacks a model for identifying who this person actually is in a simulation environment. Thus, developing a means of identifying who the primary driver is for each vehicle in the household fleet can be used together with the existing car allocation model to improve representation of which vehicles are used for which trips, and thus overall vehicle emissions modelling.

Finally, the vehicle emissions model will also be indirectly made more accurate by the improved representation of mutual travel behaviour and vehicle fleet size and composition. The integration of these two models (as well as with ILUTE as a whole) should provide more accurate travel behaviour modelling, especially for policy analysis and forecasting purposes.

Full documentation of the vehicle emissions and dispersion model can be found in Hatzopoulou (2008).

Chapter 4

Revisions to Original Vehicle Transaction Model

The model developed by Mohammadian as described in Section 3.3 is used as the foundation for the work undertaken in this thesis. This encompasses changes or additions to model specification, model estimation and simulation methods, model input variables and the scope of what the model seeks to represent. Elements of the Mohammadian model that were not revised in their own right and were otherwise compatible and usable with the revised model remain unaltered.

The additions and changes are discussed in this Chapter and the following one. In this Chapter:

- Section 4.1 discusses revisions to the model decision-making structure.
- Section 4.2 describes changes that were made to vehicle classification, properties and cost information.
- Section 4.3 details how the different vehicle class and vintage combinations that make up the class-vintage choice set of hypothetical vehicle purchases are represented in the model.
- Section 4.4 develops an algorithm that is used assess who within the household is the vehicle's primary driver.
- Section 4.5 reviews a number of smaller revisions that were made to improve various components of the model.

The single largest conceptual improvement the revised model will make is to introduce variables related to how vehicles ownership can facilitate activity participation while minimizing the disutility associated with the travel required to participate in this activity. Essentially, this represents of a quantification of the sentiment of *"How helpful to me is owning a car in terms of allowing me/us to get where I'm/we're going by giving me/us the option to drive?"*. This vehicle use information can also be used to better represent vehicle Operating and Maintenance costs. Due to the level of importance and complexity of the process involved in this addition, it is given its own Chapter, and will thus be discussed in Chapter 5.

4.1 Revisions to Model Structure

A key decision in developing a discrete choice model is selecting the preferred model structure. Although conducting a literature review on existing work in the research area is helpful, a model should nonetheless experiment with several candidate structures to better understand how each of them perform relative to the data being modelled, and then selecting one with both a defensible theoretical basis as well as a strong model fit. This is particularly important when the model is breaking new ground, and there is not a large reservoir of directly relevant literature to draw upon.

Given that the level of detail included in the original ILUTE vehicle transaction model made it one of the most advanced models of its type at the time of it development, Mohammadian undertook a significant amount of experimentation regarding model structure. This included experimentation with nested logit models, mixed logit models and artificial neural networks. Within the nested logit structure, several different specifications were tested, to assess which of class or vintage choice is precedent in vehicle purchase decisions. The conceptual model decision-making that Mohammadian ultimately decided to use is the nested logit structure shown in Figure 3.2.

This model structure has a number of conceptual strengths, including the following:

- The choice to buy a vehicle is sensitive to the expected utility of what vehicle would be bought
- The choice to dispose a vehicle is sensitive to what vehicle would be disposed
- The choice to replace a vehicle is sensitive to both what vehicle would be bought, as well as what vehicle would be disposed.

However, a source of concern was identified with this model. Specifically, stated transaction choice motivations collected in the TACOS survey suggest that in Replace transactions, the triggering event is the need to dispose a vehicle, which then will often influence what vehicle it is replaced with. The current model involves the opposite pattern, where the vehicle being purchased influences the vehicle being disposed.

Therefore, a revised model structure that addresses this concern was developed, as shown in Figure 4.1. The model did not revisit the decision to use a nested logit, or to place the class choice above the vintage choice in terms of nesting levels. Despite the concern discussed above, Mohammadian's rationale for these decisions continues to hold true.

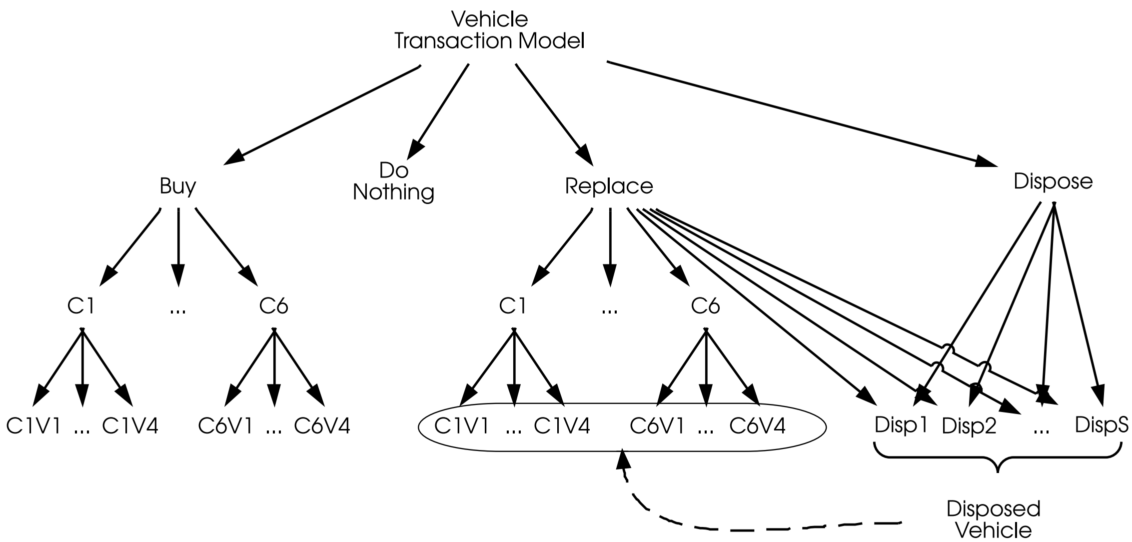

Figure 4.1: Revised Transaction Model Choice Structure

The key aspects to the decision making structure that in the revised model are:

- The Buy transaction choice is sensitive to the expected utility that would come from buying a vehicle through an inclusive value function of the class-vintage choice model.
- The Do Nothing transaction choice has no nests beneath it, and generally just acts as the "default" option, with the other three actions being taken if there is any particular benefit associated with taking one of them.
- The Replace transaction choice is sensitive to the expected utility that would come from buying a vehicle through an inclusive value function of the class-vintage choice model as well as the expected utility that would come from disposing a vehicle through an inclusive value function of the disposal choice model.
- Furthermore, the vehicle being disposed of in the Replace transaction choice can influence the type of vehicle that is purchased, thereby solving the concerns with the original model.
- The Dispose transaction choice is sensitive to the expected utility that would come from disposing a vehicle through an inclusive value function of the disposal choice model.

Given the complexity of the decision-making structure, the model that will be developed is anticipated to be estimated in three separate steps:

- The disposal choice model will be estimated as a multinomial logit model using data from all records of households making a Dispose or Replace decision and therefore selecting a vehicle for disposal. Once the model is estimated, inclusive values of the expected utility of the dispose choice will be calculated for all households. This model is detailed in Section 6.3.

- The class-vintage choice model will be estimated as a nested logit model using data from all records of households making a Buy or Replace transaction, and therefore selecting a vehicle for purchase. The information on which vehicle was disposed of will inform some variables for class-vintage choice in replacement transactions (otherwise the class-vintage models for the Buy and Replace transaction choices will be the same, and can still be estimated together in any case). Once the model is estimated, inclusive values of the expected utility of the class-vintage choice will be calculated for all households. This model is detailed in Section 6.4.

- The transaction choice model will be estimated as a multinomial logit model using data for all households for all years for which they are available. The class-vintage choice model and disposal choice model inclusive values will be used to inform transaction decisions as shown in Figure 4.1. This model is detailed in Section 6.5.

All of the above changes apply to the revised vehicle transaction model structure. However, in addition to the transaction model, this thesis also develops a vehicle fleet synthesis procedure to populate households with vehicles for their first year in the simulation, using an ownership level rather than transaction choice framework. This structure of this model is presented in Figure 4.2, and will be fully detailed in Chapter 7.

Given the number of similar-sounding concepts being discussed, appropriate terminology is required to differentiate between the various models. There following terms will be used throughout this thesis:

- *"Vehicle transaction model"* refers to the entirety of what is shown in Figure 4.1.

- *"Vehicle initialization model"* refers to the entirety what is shown in Figure 4.2.

- *"Transaction choice model"* refers to the top level choice in Figure 4.1 (Buy, Do Nothing etc.).

- *"Ownership level model"* refers to the top level choice in Figure 4.2 (1 Vehicle, 2 Vehicles etc.). This is sometimes alternatively referred to as a "vehicle holdings model" in the literature.

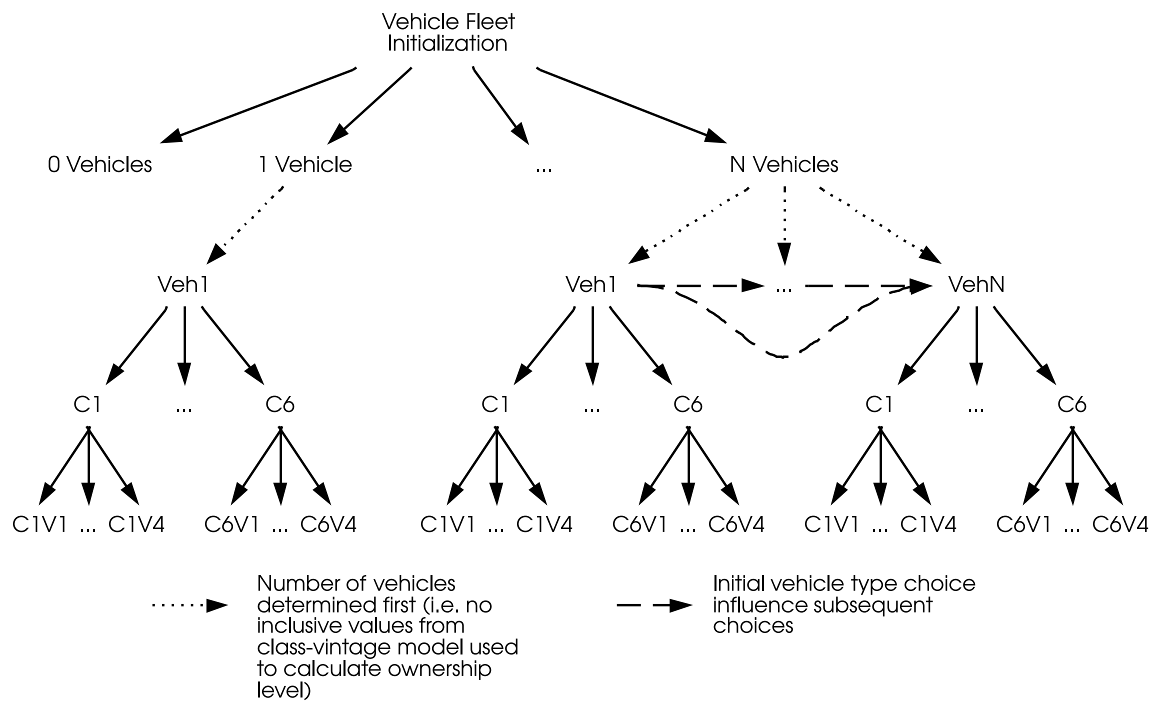

Figure 4.2: Vehicle Initialization Model

- *"Class-vintage model"* or '*'vehicle type model"* refers to the bottom level of the Buy and Replace options in Figure 4.1, as well as the bottom two levels of Figure 4.2 (the class nests of C1, C2 etc. and the class-vintage choices of C1V1, C1V2 etc.).

- *"Disposal choice model"* refers to the dispose choices in Figure 4.1 (Disp 1, Disp 2 etc.).

- *"Vehicle fleet model"* refers to the entirety of the works discussed in this thesis.

4.2 Vehicle Classification, Properties and Price Model

Several aspects of the vehicle classification, property, price and cost definitions and/or models developed by Mohammadian and reviewed in Subsection 3.3.1 have been either altered or moved. These changes are detailed in this section, and are incorporated into all subsequent work undertaken in this thesis.

4.2.1 Vehicle Class Classifications

Identifying and assigning vehicle class can be a challenging task, as the boundaries between different classes are not always entirely black and white. This is doubly true when attempting to consolidate a large number of classes down into a smaller number of ones for more efficient modelling. In developing the original transaction model, Mohammadian worked to consolidate upwards of a dozen different vehicle classes into a more manageable number to be included in the vehicle transaction model.

Several different aspects of the model development appear to have been undertaken at different levels of consolidation. The hedonic price model as given in Table 3.3 implies that cargo vans were still distinct from minivans, and pickup trucks were similarly distinct from sports utility vehicles. In the operation and maintenance cost model shown in Table 3.4, these groups had been consolidated into Van and Special Purpose Vehicle, respectively. The seven consolidated classes given in the O&M cost model are Subcompact, Compact, Midsize, Large, Station Wagon, Special Purpose Vehicle and Van.

Subsequent work in developing the class choice model found that there were an insufficient number of Station Wagons in the data (they make up 72 of the 2228 vehicles, or 3.2% of all vehicles in the TACOS dataset) to develop a robust choice model for this class. Station Wagons were therefore consolidated into several other classes, bringing the number of classes down to the six that were ultimately used in the model. The overall consolidation process undertaken by Mohammadian to bring the number of classes down to six is shown in Table 6.1 of Mohammadian (2002).

In the edition of the TACOS data made available for this thesis, seven different classes were provided, matching the seven that were used in the O&M cost model. For use in this thesis, vehicle class definitions and properties were used as follows:

- Subcompact, Compact and Midsize vehicle classes remain unchanged from the original definition.
- Large vehicles and Station Wagons were combined into a single Large class. As discussed above, Station Wagons make up only a very small proportion of vehicles in TACOS. Furthermore, they have also have suffered from declining popularity since the survey was collected in the 1990's, and are even less common today. Thus, Mohammadian's decision to eliminate them is reasonable. Unfortunately given the seven vehicle classes provided in the dataset, rather than the dozen-plus that were originally consolidated, Station Wagons were not able to be distributed by their respective sizes to different sedan classes. Given that they all had to be placed in a single alternative class, it was decided that Large would be the most appropriate. The Station Wagons therefore have been assigned Large class properties and O&M costs. However, the Large (i.e. original Large plus Station Wagon) class has now been given the "Size" dummy variable to compensate for this.
- Pickup Trucks and Sports Utility Vehicles have previously been consolidated by Mohammadian into a combined Special Purpose Vehicle (SPV) class. However, the hedonic price model shown in Table 3.3 incorporates a "cargo" dummy variable, which influences the Space calculation, which in turn influences both the Vehicle Performance Factor and Vehicle Space Factor, and ultimately

the vehicle price. Mohammadian specifies in his vehicle property definitions that this applies to pickups, but not SUVs. Given that they have both been consolidated into a single class, it either needs to apply for both or neither in order for the model to function. Given that SUVs have removable seating for the middle and back rows, they can easily be converted to provide cargo space if required. Thus, the SPV class as a whole was assigned the cargo space dummy. This class also features the Size dummy variable.

- Cargo Vans and Minivans have also previously been consolidated, this time into a generic Van class. Similarly to the issue surrounding the SPV class, Cargo Vans were assigned the cargo space dummy but minivans weren't. As with SUVs, minivans can remove seating to create cargo space, and thus the Van class as a whole is assigned the dummy variable. This class also features the Size dummy variable.

Table 4.1 summarizes the revised class definitions.

Table 4.1: Changes to Class Category Naming Conventions and Definitions

Original Model Class Name	**Revised Model Class Name**	**Alternative Name**	**Other Properties**
Subcompact	Subcompact	C1	
Compact	Compact	C2	
Midsize	Midsize	C3	
Large, Station Wagon	Large	C4	Size Dummy
Sports Utility Vehicle, Pickup Truck	Special Purpose Vehicle	C5	Size Dummy, Cargo Dummy
Cargo Van, Minivan	Van	C6	Size Dummy, Cargo Dummy

4.2.2 Vehicle Vintage Classifications

The vehicle vintage classifications used by Mohammadian in the original model appear to work very well. However, there was a concern with the naming of the second vintage category as "Second Hand". It was thought that the term "Second Hand" would imply the vehicle had in all cases belonged to another owner previously, whereas in reality, it could still be the original owner, but who has owned the vehicle for 1-2 years. The name was therefore changed to "Nearly New". Table 4.2 summarizes the original and revised naming conventions.

Table 4.2: Changes to Vintage Category Naming Conventions

Vehicle Age Range	Original Model Vintage Name	Revised Model Vintage Name	Alternative Name
-1 to 0 years	Brand New	Brand New	V1
1 to 2 years	Second Hand	Nearly New	V2
3 to 7 years	Used	Used	V3
8+ years	Old	Old	V4

4.2.3 Luxury Vehicles

The hedonic price model developed by Mohammadian and shown in Table 3.3 contains a dummy variable for vehicles designated as being a luxury vehicle, which has the effect of increasing the price of the vehicle by $26,086. In the context of the estimation of the vehicle price model, this is clearly a reasonable course of action, as it identifies a major cost factor for these vehicles, and extracts it from the more tangible physical qualities of the vehicle that also drive purchase price, allowing for an improved overall model fit.

However, how exactly the luxury vehicle variable would be applied in terms of vehicle fleet model development, estimation, and eventually simulation in ILUTE is not obvious. Modelling the fact that a vehicle is a luxury vehicle is not an aim of ILUTE in its own right, based on the intended applications discussed in Chapter 1. Rather, it only matters in terms of how it would affect vehicle ownership levels and class and vintage decisions. The main consumers of luxury vehicles will generally be the wealthiest segment of society, who will tend to have the financial resources to purchase as many vehicles as they require for transportation purposes. It is hard to imagine that there are many household who, for example, would forgo two non-luxury vehicles in favour of one luxury vehicle if the second one was integral to greatly increasing its overall travel and activity participation utility. Thus, the market for purchasing a luxury vehicle given the decision to own a vehicle is likely not competing financially with an alternative action of purchasing an additional vehicle. Rather, the household is likely to already have the financial resources it requires to afford as many vehicles as it required.

Since household income levels are capped at $85,000 (see the discussion in Subsection 4.5.2 below), the very households that are likely to have the extremely high income levels that make them likely to purchase luxury vehicles are the same households that have had their modelled spending power curtailed significantly (i.e. down to $85,000). Thus, removing the type of Veblen Good that these high income households are likely to spend their disposable income on in light of the fact that those same disposable incomes has been reduced may actually have a neutralizing effect, and reduce the level of error incurred by these simplifications.

In summary, the possibility of removing the concept of luxury vehicles from the model would have two clear benefits:

- Knowing that a vehicle is luxury is not important in and of itself; excluding them simply creates a more parsimonious model.
- Removing luxury vehicles may have the effect of helping cancel out other simplifications in the model.

On this basis, the concept of a luxury vehicle is removed from the model; all vehicles will simply be modelled as non-luxury vehicles.

4.2.4 Revisions to Hedonic Price Model

In light of all the changes presented to the vehicle properties and cost models outlined in this section, an updated version of the hedonic price model is presented in Tables 4.3 and 4.4 below. This is the model that will be used for all vehicle price calculations in the model estimation component of the thesis.

Table 4.3: Updated Vehicle Properties used in Price Model Development

Characteristics	**Unit**
Engine Displacement	Litre
Weight	Tonne
Fuel Intensity	L/100km
Luggage Capacity	m^3
Wheelbase	metre
Age	Vehicle model year
Class	Subcompact, Compact, Midsize, Large, Special Purpose Vehicle, Van
Origin	Domestic, Japanese, European
New	If vehicle is a Brand New vintage
Size	If the vehicle is Large, SPV or Van class
Cargo	If a vehicle is a SPV or Van class
Space	$(1 - Cargo) \times LuggageCapacity \div Wheelbase$
Vehicle Performance Factor	Calculated per Table 3.2
Vehicle Space Factor	Calculated per Table 3.2

Further, to provide addition clarity, the means by which the primary vehicle properties are used to calculate derived vehicle properties and ultimately costs is shown in Figure 4.3.

4.2.5 Purchase Prices for Observed Vehicle Transactions

Vehicle prices will be incorporated into the revised model structure as part of the class-vintage selection model. As will be discussed in Section 4.3, a hypothetical vehicle is created for each potential class-

Table 4.4: Updated Hedonic Vehicle Price Model (1000's, \$1998)

Variable	Coefficient	Standard Error	t-statistic
Subcompact	12.753	0.423	30.135
Compact	13.767	0.381	36.11
Midsize	15.288	0.375	40.718
Large	16.395	0.478	34.285
Special Purpose Vehicle	13.295	n/a	n/a
Van	12.897	0.69	18.694
New	0.779	0.299	2.604
Natural Logarithm of Vehicle Age	-6.072	0.166	-36.651
Japanese Car	3.746	0.26	14.392
European Car	4.484	0.494	9.068
Vehicle Performance Factor	2.294	0.168	13.628
Vehicle Space Factor	1.038	0.273	3.8
Time[1]	0.514	0.421	12.216

[1]Measured in years to/from 1990.

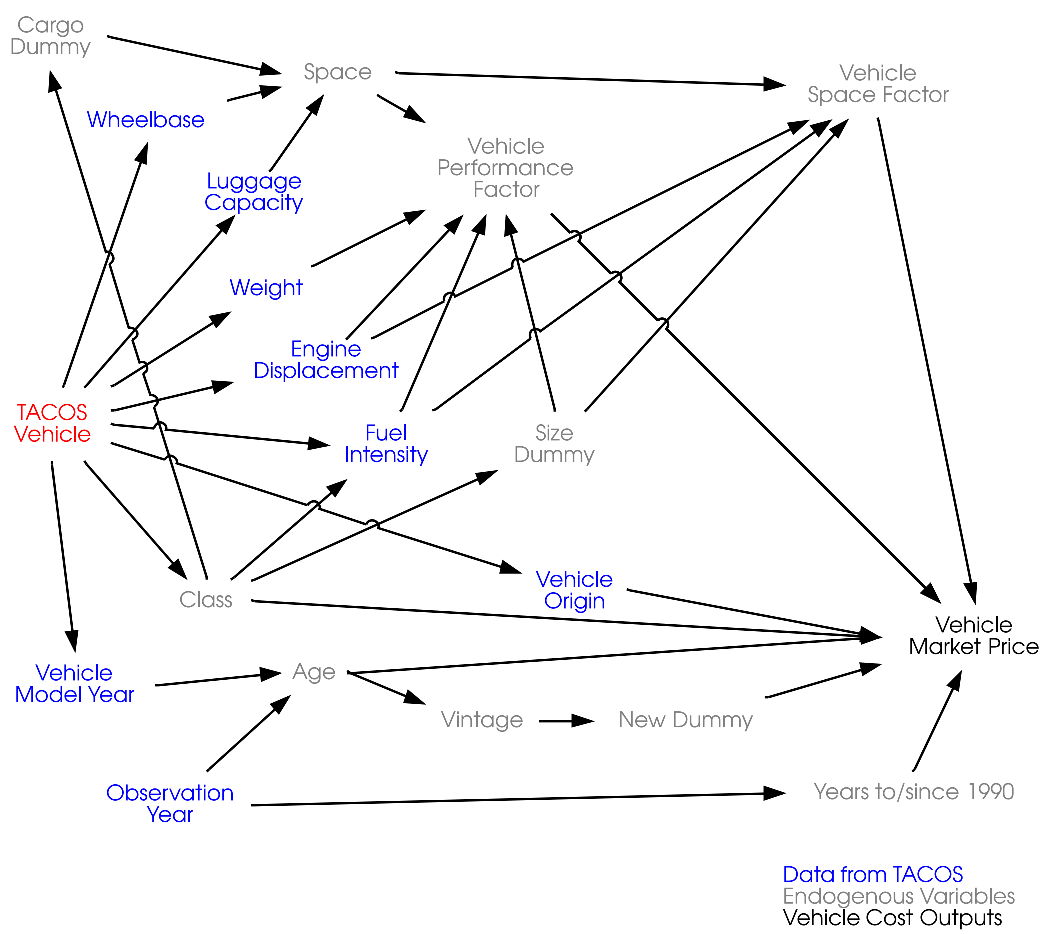

Figure 4.3: Vehicle Properties and Price Costs for Vehicles Recorded in TACOS

vintage choice. For the purposes of class-vintage model estimation based on actual vehicle procurements resulting from buy or replace decisions, this means that there will be one revealed preference selected-for-purchase class and vintage combination, and twenty-three unobserved but (assumed to be) considered-but-not-selected-for-purchase class and vintage combinations. For those twenty-three unselected choice alternatives, obviously the hedonic price model must be used to estimate their costs, as no revealed preference information is available. However, for the vehicle that was selected by the household, market price information was collected from the *Canadian Red Book* for each make and model in the dataset. Thus, there are two potential sources of price information that can be used for the purchase price of the selected vehicle:

- The vehicle market prices as reported in the *Canadian Red Book*.
- Vehicle market price as predicted by the hedonic price model given in Subsection 4.2.4, which is based on the price data in the *Canadian Red Book* being regressed against the physical properties of the vehicle.

Ultimately, the latter alternative was chosen because the other twenty-three class-vintage combinations that were not purchased will have to have their market prices modelled using the hedonic price model in any case, and this will provide an "apples to apples" comparison. Furthermore, once estimation is complete, and simulation-based applications of the model are taking place, all twenty-four alternatives, including the selected one, will have their prices estimated through the model. Thus, using the model also provides an internally consistent source of data to be used for estimation and simulation, and will hopefully make the latter procedure more accurate based on its consistency with the former.

On another note, as vehicles are large capital-intensive purchases for most households, they are not always purchased outright but instead through a financing scheme offered by the dealership. This loan of course results in a higher overall purchase price of the vehicle as a result of interest charges, but allows households to be using their vehicles sooner than if they waited to save up the full purchase cost. Representing these additional financing charges in the model and to whom they would apply would be challenging, and subject to a substantial amount of error regarding who would make use of them. Therefore, they are not directly included, and the capital cost of the vehicle is assumed to be sole direct cost associated with its procurement. However, to implicitly add a degree of realism with respect to the impact of financing costs on vehicle purchase decisions, interest rate values will be incorporated into both the class-vintage and transaction choice model utility functions, as will be discussed in Chapter 6.

4.2.6 Revisions to Operation and Maintenance Costs

The representation of O&M Costs shown previously in Chapter 3 has been modified to work with the revised modelling strategy that forms the main work of this thesis. These changes are detailed below.

Maintenance and Insurance cost values are unchanged, and are summed together to create what will be referred to as O&M Fixed Costs, which will apply to all vehicles regardless of use, on the basis that these are assumed to not be dependent on distance travelled. In practice, some level of variability is inherently included within them, as was discussed in Subsection 3.3.1. Maintenance costs are based on the assumptions made as part of the inputs to the HDMS vehicle operating cost software that incorporates the distance they travel each year as well as the number of hours spent driving. In reality, maintenance costs will have both a fixed component as well as a variable component that increases with vehicle use. Similarly, the insurance costs developed by Mohammadian implicitly assume certain values for annual VKT as well as daily commute distance in order to generate insurance rates.

Obviously treating these costs as being entirely fixed is not ideal given that they do have a variable cost component, but short of recalculating all costs from scratch, this is the most reasonable course of action. The O&M Fixed costs are also assumed to hold constant over time, after accounting for inflation. For example, a Midsize vehicle built in 1980 (and thus of Nearly New vintage) in 1982 will have the same O&M Fixed costs as a 1994-built vehicle will have in the year 1996, once inflation is accounted for.

The O&M Fixed Cost model, as derived from the original model given in Table 3.4 is presented in Table 4.5.

Depreciation costs were also initially included in the O&M Fixed Costs model, but were later removed. Preliminary modelling efforts found that the variables incorporating O&M Fixed Costs had more statistically significant fits with depreciation costs removed. This suggests that most people generally view a vehicle as an investment in the sense that it provides access to activities, but not in the sense that they expect the physical piece of equipment itself to generate a positive rate of financial return. Analogously, households would view purchasing fresh fruit as an investment in nutrition, but are not expecting to be able to leave it in their cupboard it for several years and then sell it at a higher price. In other words, that the fruit is depreciating in re-sale value is not an important part of the decision.

Finally, Mohammadian calculated fuel costs based on the same vehicle use assumptions that were used to generate maintenance costs. This was a reasonable response in the absence of any specific information regarding how much households actually use their vehicles. However, the revised transaction model calculates household VKT information, as will be discussed in Chapter 5. As such, fuel costs can be made sensitive to outputs from TASHA, and thus fuel costs will become a new variable called O&M Variable Costs.

Fuel intensity values are based on class and model year, and can be found in in Appendix B. A given vehicle is assumed to have the same fuel intensity over the entire course of its life. It is recognized that in reality it may decline over time, depending on the level of maintenance put into the vehicle,

Table 4.5: Revised Operation and Maintenance Fixed Costs

Class	Vintage	Maintenance	Insurance	Total
Subcompact	Brand New	0.090	4.003	**4.093**
	Nearly New	1.064	3.790	**4.854**
	Used	1.361	3.260	**4.621**
	Old	1.610	2.490	**4.100**
Compact	Brand New	0.090	3.552	**3.642**
	Nearly New	1.038	3.340	**4.378**
	Used	1.324	2.918	**4.242**
	Old	1.544	2.281	**3.825**
Midsize	Brand New	0.090	3.783	**3.873**
	Nearly New	1.214	3.440	**4.654**
	Used	1.573	2.888	**4.461**
	Old	1.876	2.182	**4.058**
Large	Brand New	0.090	3.600	**3.690**
	Nearly New	1.230	3.381	**4.611**
	Used	1.600	2.944	**4.544**
	Old	1.917	2.318	**4.235**
Special Purpose Vehicle	Brand New	0.090	3.178	**3.268**
	Nearly New	1.150	3.499	**4.649**
	Used	1.489	3.604	**5.093**
	Old	1.780	2.274	**4.054**
Van	Brand New	0.090	3.295	**3.385**
	Nearly New	1.118	3.081	**4.199**
	Used	1.431	2.654	**4.085**
	Old	1.698	1.930	**3.628**

but this is somewhat challenging to reflect, and is not explicitly accounted for. Gasoline prices can also be found in Appendix C. This entire O&M Variable Cost calculation process, including development of VKT information and the adjustment factors applied to the model will be discussed in more detail in Subsection 5.4.3 of Chapter 5.

4.3 Virtual Vehicle Dealership

To purchase a vehicle, consumers must first select which type of vehicle they desire from the stock of candidate vehicles, and then compare the price and properties of each candidate vehicle.

In the vehicle transaction model, this process has been reflected by developing what is essentially a “virtual vehicle dealership”, which creates a choice set of simulated vehicles for purchasers to choose from. For each potential purchase choice, a total of 24 vehicles are created to choose from; one for each class-vintage combination supported by the model.

To create this vehicle choice set, key vehicle properties must be generated, and then used to derive all remaining info about the vehicle. The approach of simulating a "representative vehicle" for each choice in the choice set used here is similar to that developed by Lave & Train (1979), among others. To simulate each class-vintage combination, the following primary seven vehicle properties must be generated:

1. Luggage Capacity (m^3)
2. Wheelbase (m)
3. Engine Displacement (L)
4. Weight (tonnes)
5. Fuel intensity (L/100km)
6. Vehicle age
7. Vehicle origin

Properties 1, 2, 3 and 4 are generated separately for each class, but are considered to be constant over the entire simulation time period. They were calculated based on the average of all vehicles within that class that are listed in the TACOS dataset. The resulting values are given in Table 4.6.

Table 4.6: Simulated Vehicle Attributes by Class

Component	**Subcompact**	**Compact**	**Midsize**	**Large**	**Special Purpose Vehicle**	**Van**
Luggage capacity (m^3)	0.345	0.381	0.445	0.605	2.384	2.161
Wheelbase (m)	2.474	2.604	2.693	2.819	2.819	2.905
Engine displacement (L)	2.030	2.270	3.061	3.842	3.548	3.540
Weight (tonnes)	1.078	1.197	1.385	1.534	1.551	1.640

Property 5, fuel intensity, varies not just by class but also over time. However, it does not depend on vehicle age so much as model year; given that fuel efficiency standards have become more stringent over time. Table 4.7 summarizes the fuel intensity values that are assigned to vehicle based on class and a back-calculation of their model year once vehicle age (Property 6) has been determined and the year being simulated is accounted for. Note that mileage values are the average of "city" and "highway" driving mileage ratings.

In terms of Property 6, vehicle age, the vehicle transaction model framework contains a choice selection process for vehicle vintages, which are groups of vehicle ages. Thus, the vehicle transaction model will assign a vehicle to have a known age *range*, but not an exact age. However, several parts of the modelling process require a specific age in order to operate, including:

Table 4.7: Simulated Vehicle Fuel Intensity (L/100km) by Class and Model Year

Year	Subcompact	Compact	Midsize	Large	Special Purpose Vehicle	Van
1999	7.136	7.650	8.851	8.576	10.721	11.201
1998	7.136	7.650	8.851	8.576	10.721	11.201
1997	6.994	7.749	8.834	9.650	10.810	10.352
1996	7.863	8.150	8.357	9.960	10.047	10.282
1995	7.657	7.767	9.107	9.586	10.585	10.923
1994	7.430	8.213	8.925	9.668	10.221	11.212
1993	7.174	7.980	9.134	8.666	11.099	10.066
1992	7.485	8.175	9.076	9.674	9.889	10.818
1991	7.256	8.191	9.099	9.547	9.939	10.098
1990	7.092	7.792	9.084	9.207	10.937	10.610
1989	7.601	7.953	9.591	9.295	10.475	10.458
1988	7.149	8.125	8.888	9.083	11.881	10.291
1987	7.423	7.822	9.007	9.291	10.262	10.682
1986	7.183	7.395	8.811	9.095	9.700	9.982
1985	6.994	7.080	9.071	10.355	10.560	10.433
1984	7.546	6.595	9.158	9.119	10.551	12.070
1983	7.508	8.020	8.758	11.820	8.532	13.386
1982	7.827	6.662	9.787	8.111	9.370	14.043
1981	8.152	8.401	10.094	10.746	14.701	14.701
1980	8.691	10.454	11.031	12.615	14.293	13.574
1979	10.960	11.823	9.699	13.580	13.884	12.447
1978	10.826	13.210	12.186	14.413	8.111	15.191

- The hedonic price model makes use of the log function of a vehicles age to estimate the market value of the vehicle (which in turn influences whether the household wants to buy the vehicle or not in the first place).
- As described above, the fuel intensity of the vehicle is partially based on model year, which will require a specific age to be known (and again influences the likelihood that the household will buy the vehicle).
- Transaction choice decisions may be influenced by some property relating to the age of vehicles in the household fleet (e.g. age of oldest vehicle, average age of the fleet etc.), which requires specific ages.

Under the original framework, the model requires exact age but generates vintages. Thus, it requires more detailed endogenous information than it generates, which renders the model unable to be used in year-over-year simulations. To correct this, two potential solutions were examined:

1. Re-specify the model to be a class-age model, where each individual year is represented as a possible purchase choice.
2. Randomly generate a vehicle age within each vintage category.

The effectiveness of both of these options depends in part on how large of an available age range is considered desirable to allow households to choose from when purchasing a vehicle. A decision was made that there would be a total of fourteen vehicle model years available to choose from; ranging from vehicles as new as the *following* model year, and as old as up to 12 years *prior* to the current model year. Few households purchase vehicles that are older than 12 years old at time of purchase, as at this point most vehicles are already nearing the end of their useful lives.

Under the first alternative, each of these fourteen ages must be modelled for six separate class categories, resulting in eighty-four separate class-age categories that must be modelled. The chances of developing a well-fitting model for such a large number of alternatives in the choice set are likely to be quite small. Furthermore, some of the variables such as O&M Fixed costs are already categorized by vintage, and would have to apply equally to all vehicles with their ages in that vintage category. It is also debatable how much difference a single model year makes in purchase decisions; particularly for older vehicles (i.e. is 5 years old *really* that much better than 6 years old?).

Instead, it was decided that using the second alternative would be preferable. Using TACOS as a data source, the age of each vehicle at time of acquisition was calculated for each vehicle. These were then combined into a frequency table, and sorted into their respective vintage categories. The number of observations of a particular age was divided by the number of observations of its corresponding vintage to develop a conditional probability of vehicle age selection. The resultant values are shown in Table 4.8.

Table 4.8: Vehicle Age Assignment

Vintage Category	**Age**	**Age Probability (given Vintage)**
Brand New	-1	17.1%
	0	82.9%
Nearly New	1	55.7%
	2	44.3%
Used	3	27.9%
	4	25.0%
	5	15.7%
	6	14.6%
	7	16.8%
Old	8	22.6%
	9	23.7%
	10	16.1%
	11	16.7%
	12	21.0%

Using this method, the following procedure is to be used for year-over-year simulations:

1. Randomly generate ages for each candidate vehicle for class-vintage choice model.
2. Based on age and current simulation year, back-calculate model year to calculate fixed and variable operation and maintenance costs as well as the market price.
3. The chosen vehicle (if any) will have its model year recorded in the ILUTE database.
4. At the beginning of each simulation year, the age of all vehicles in all household fleets are updated to reflect the passage of an additional year since their model year, their vintage classification is re-determined, and their market price and operation and maintenance fixed costs are in turn updated to reflect this.

A similar but simpler process was used for the assignment of Property 7, vehicle manufacturer origins (i.e. Domestic, Japanese or European). Knowing the vehicle origin is not considered to be important for the type of analysis ILUTE seeks to model; it is simply relevant in terms of how it affects the market price of the vehicle as can be seen in Table 4.4. Similar to the issues faced by vehicle ages, the means of determining vehicle origin for all non-observed vehicles was not detailed by Mohammadian. A simple probability-based model is generated, where each vehicle in the virtual vehicle dealership will be randomly assigned a probability according to Table 4.9. These probability values were developed based on the proportions of the origins of each vehicle in TACOS.

Table 4.9: Vehicle Origin Assignment

Origin Category	**Origin Probability**
Domestic	70%
Japanese	25%
European	5%

It is recognized that this is a simplistic and potentially error-prone manner of assigning vehicle origins, and implemented for convenience. For example, market-shares from different manufacturers will change over time, or some origins will be featured disproportionately often for certain vehicle classes. If this is a source of concern for the ILUTE user, more detailed trends (both historic and forecast) on manufacturer market-share can be input into ILUTE as an exogenous variable and used to assign vehicle origin and therefore vehicle market price. However, as noted by Train & Winston (2007), manufacturer market shares appear to be driven by vehicle design and performance rather than the origin itself, suggesting that beyond its effect on overall vehicle purchase price, this variable is not particularly important in and of itself.

Once all seven vehicles properties have been assigned, other properties derived from them that were previously outlined in Subsection 3.3.1 can be calculated.

Figure 4.4 shows how this calculation process procedure for simulated vehicles from the virtual vehicle dealership, where generally follows the same process as Subsection 4.2.4, but involves additional steps as a result of the simulated nature of the vehicle properties (rather than them already being available in the dataset). All of this information can then be used to inform the utility equations for the class-vintage model.

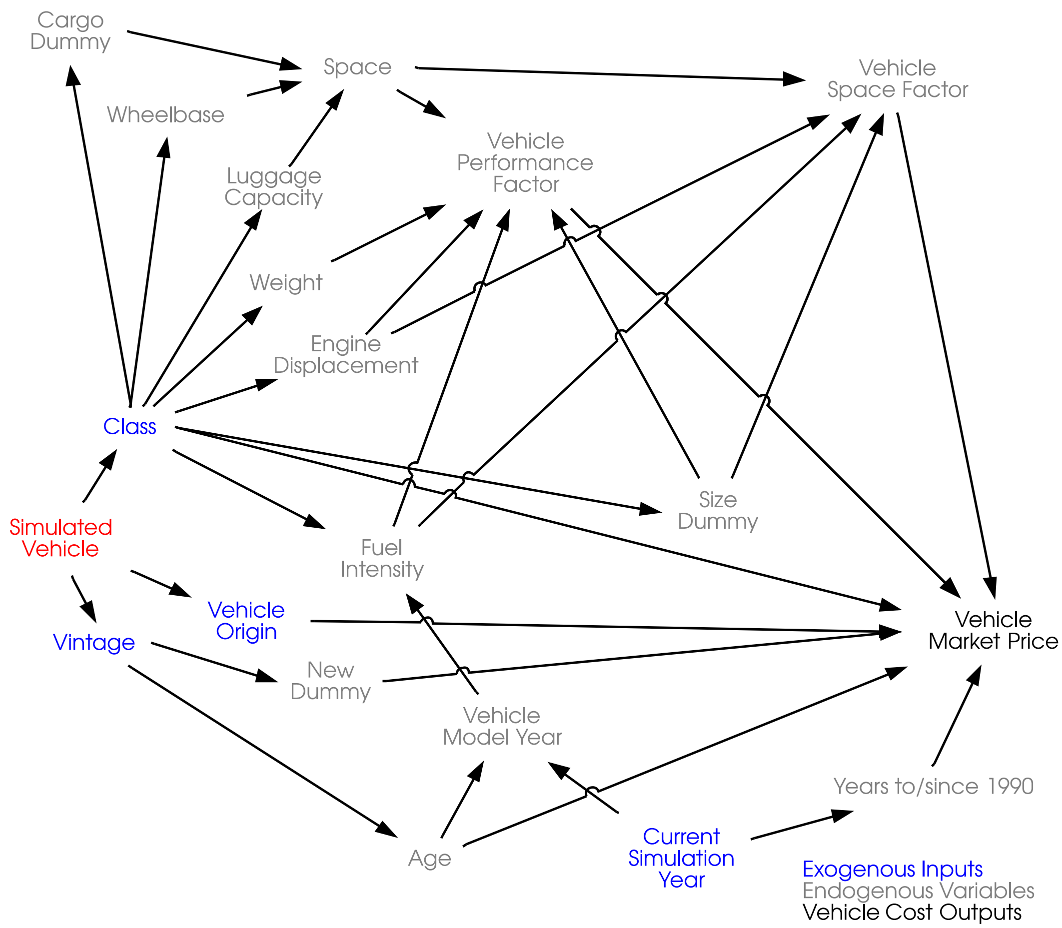

Figure 4.4: Vehicle Properties and Market Price for Simulated Vehicles

The virtual vehicle dealership is used in three separate applications. First, it was used to help estimate the class-vintage choice model. TACOS is a revealed preference survey, and therefore only contains data for the vehicles that households *did* buy; but not those that were considered but not purchased. Thus, in developing a model for class-vintage choices within the transaction choice framework, vehicles-not-bought-but-considered need to be simulated for each choice decision. Thus, the actual vehicle properties for the vehicle class-vintage that was purchased was used in conjunction with simulated values for the other 23 combinations to create a full choice set. The only exception to this is market price, where even the observed vehicle had its price modelled, as was discussed in Subsection 4.2.5. Only households that actually bought or replaced a vehicle in this year were included in this model estimation, which is discussed in further detail in Section 6.4.

The second application builds off the first application and is used as part of the transaction choice model estimation procedure. The class-vintage choice model is estimated solely with observations for households that actually made such a selection (either buying or replacing a car) for the particular year where that transaction occurred. In contrast, the overall transaction choice model includes all households for all years that they are in the dataset, as it must include households that chose to dispose of a car and those that chose to make no changes to their vehicle fleet at all (i.e. the Do Nothing option). An implicit assumption of the model is that each year, households reconsider their vehicle fleet in terms of both number of vehicles and types of vehicles, actively examine vehicles available on the market, and only then decide on a course of action in terms of transaction choice. In other words, households "doing nothing" is assumed to be the result of a rational decision-making process where they were aware of what vehicles they *could have* bought if they decided to buy one, rather than solely inertia (i.e. nothing triggered the need to even think about their vehicle fleet, thus no consideration of it was undertaken in the first place, and they have no idea what is available for purchase). To develop the full model structure, it is assumed that those households that did not take any action nonetheless considered all possible applicable actions (Buy, Dispose, Replace, Do Nothing), and simply chose to Do Nothing. Similarly, those that disposed of a vehicle nonetheless considered buying/replacing one, but chose not to. Thus, the full class-vintage choice set of 24 (instead of just 23) vehicles was required to be simulated, as there is no observed information whatsoever of what vehicles households considered but didn't purchase.

The third and final application is for actual implementation and use of the model for simulation once it has been estimated. As will be discussed in Chapter 8, there are still some advancements in other areas of ILUTE that must be made before the transaction model can actually be run as part of a fully integrated model. However, once these advancements are achieved, the transaction model will be required to simulate vehicle data for millions of households each year over the course of several decades (either the 1986-2006 validation period or later periods as part of further validation or even forecasting). in these cases, the whole 24 vehicle class-vintage choice sets must be simulated and new external input data must be developed for certain properties. For example fuel intensity values from 1999 onward must be collected as they are not included in TACOS data, and using ILUTE as a forecasting tool would require that predictions be made for fuel intensity on a year by year basis up until the end of the forecasting horizon year.

Obviously a choice set of 24 vehicles based on various class-vintage combinations is not the extent of the choice process in the real world, where consumers will tend to compare carefully between different makes and models *within* the various class and vintage categories, or will exhibit brand/model loyalty. While the transaction model used in ILUTE undoubtedly lacks detail and realism in this regard, this is not considered to be a source of concern. As was discussed in Section 1.2, the intent of the vehicle fleet model is to help inform a larger urban simulation model, not to help marketers sell particular makes of vehicles. Thus, while detailed attributes such as the brand preferences, interior finishes, colour etc. are important to real-life people, they are not particularly important for what ILUTE seeks to accomplish.

It should be noted that the virtual vehicle dealership will always have all 24 categories of vehicles available for purchase. Vehicle choices are assumed to be abundantly available; in other words, the dealership will never "run out" of a particular vehicle type. Similarly, when vehicles are disposed of (which implies they could be sold second hand to someone), there is no process by which someone else is identified as the buyer. Unlike the ILUTE housing market, which features a complex market-clearing process, vehicles in ILUTE are assumed to simply "appear" and "disappear" into thin air, so to speak. Aside from having no data with which to develop an actual vehicle market, it does not seem necessary in the first place. Housing choices inherently involve the use of land, which is finite and non-moveable. There are only so many square kilometers of developable land in Toronto and more cannot magically be made. A household can also not simply pick up its physical dwelling and the land it sits on and place it down somewhere else. In contrast, if a particular vehicle type is in exceptionally high demand, more of them can be manufactured. Further, if a household wants to dispose a vehicle, they have options ranging from finding a buyer in their area, finding a buyer outside the GTHA, selling it to a second-hand dealership, or just sending it to a scrap-yard and removing it from the regional vehicle fleet altogether. Since vehicles can just "appear" and "disappear" into/from the GTHA at will, there is no need for a vehicle market-clearing process, and thus one is not included.

All tables in this section are summarized in Appendix B for reader convenience.

4.4 Within-Household Vehicle Assignment

The original transaction model contains a number of variables in the class and vintage choice models that relate to socio-economic and demographic characteristics of the owner and primary driver of the vehicle.

The TACOS data identifies which person(s) within the households are the owner and primary driver (depending on whether they are the same person), and thus information about this person(s) can be incorporated into the data for the model estimation process, as was done in the original model and will be done in the revised model. However, no framework was developed to determine how to assign who within the ILUTE household the owner and primary driver is/are. Thus, although the originalmodel can be estimated using exogenously-provided owner/driver information from TACOS, it is impossible to simulate, since owner/driver information would have to be endogenously provided, with currently no means of providing. This provides two alternatives for developing a revised model:

- Remove all owner and/or primary driver specific information from the model specification, and only use household-specific data for socio-economic variables.
- Develop a means of identifying who the vehicle owner and/or primary driver is within the household.

Given that person-specific demographic and socio-economic variables were found to be statistically significant for several components of the original class and vintage models, it was desired to maintain this flexibility, and thus some means of determining who these individuals are for each vehicle. Furthermore, and possibly more importantly, as was noted in Subsection 3.4.3, this "primary driver" assignment is also useful for improving the vehicle emissions model.

An initial review of ownership and primary driver designations noted two key issues:

- The class and vintage models developed by Mohammadian make much more use of the primary driver characteristics than those of the owner, and thus knowing who will be the primary driver is substantially more important.
- For 84% of vehicles in the TACOS database, the owner and the primary driver the same person.

Therefore, in order to simplify the process, the use of the vehicle owner as a model input is discarded, and the primary driver will be the only person-vehicle mapping procedure to be developed. A search for existing literature in this area was conducted for ideas on how to implement such a procedure. Hensher *et al.* (2008), Golob *et al.* (1996) and Vyas *et al.* (2012) provide the most relevant work.

Hensher *et al.* use a stated preference dataset of groups where two individuals in a joint decision-making unit decide how to purchase a new vehicle, where each individual indicated their preference, and completed several sequential rounds of preferences until reaching an agreement (i.e. equilibrium). A series of mixed multinomial logit models are developed from the data to assess the influence of how priorities change over the course of the negotiations, the relative power of each individual in negotiating with each other, and also a group equilibrium model to represent the household's ultimate choice. Findings suggest an improvement over previous single-iteration stated preference experiments due to better representation of multi-round negotiations and negotiation power differentials.

Unfortunately, no comparable stated preference survey was undertaken as part of TACOS, and therefore the approach detailed by Hensher *et al.* cannot be applied here. Additionally, if such a procedure was used in the ILUTE model, it would be further complicated by the varying numbers of household decision-making agents involved in the negotiation process and the number of different alternatives being arrived at (i.e. there are multiple vehicles to assign, not just one). The approach of Golob *et al.* is similar, and also requires a stated preference survey.

Vyas *et al.* (2012) probably provide the most promising approach; as they rely solely on revealed preference data, and are able to match drivers to vehicles based on both driver and vehicle characteristics, and match them according to a utility maximizing framework that accounts for all potential such combinations. However, developing such a model proved to be computationally challenging.

Instead a simple rule-based algorithm was developed, as follows:

1. Vehicles are ranked based on their current market price (per the hedonic price model).

2. All household members are screened for possession of a driver's license. If they have one, they are added to a list of potential primary drivers, if they do not possess one, they are discarded.

3. Of the potential primary drivers, they are then ranked by employment status (full time, part time, not employed).

4. If there are ties within any category, age is used as a tie-breaker as follows:

   (a) Priority based on decreasing age from 65 to 20; reflecting who the "breadwinners" are most likely to be, and therefore most likely to own a vehicle.

   (b) Priority based on increasing age from the 66 upwards; reflecting that many seniors may still own use vehicles, but that ownership becomes less likely as they get older.

   (c) Priority based on decreasing age from 19 to 16; reflecting that teens may either buy their own vehicle and/or in some cases be given one by relatives.

5. Vehicles are then matched to primary drivers based on their respective rankings. If there are more total vehicles than total primary drivers, once each driver has a vehicle, the primary driver list repeats itself.

This procedure has not yet been validated, but such an endeavour could be undertaken using the same data used for model estimation, including the TASHA household activity schedule information discussed in Chapter 5. At present, the largest source of concern relates to how vehicles are re-assigned within the household when considering new purchase decisions. In the Buy decision in particular, the vehicle will go to the next most highly ranked member of the household. However, they may not necessarily be the one driving it. For example, consider a nuclear family of two parents and one child, with two vehicles and looking to buy a third so their teenage child can have their own. Depending on the quality of the parents' vehicles, they may either buy an additional vehicle for their child, or simply buy an additional vehicle for themselves, with the child getting one of the parents' vehicles as a hand-me-down. In both cases, the child is modelled as being the primary driver of the new vehicle in terms of the demographic and socioeconomic variables that influence the class and vintage choice. However, in the latter situation, if they plan to do a hand-me down, in reality it would be the characteristics of the parent rather than the child that would ideally be used to select the new vehicle. The fact that there is a hand-me-down taking place does get reflected in the post-transaction re-ranking procedure, but at this point the new vehicle has already been purchased, so the ultimate primary driver of the newly-acquired vehicle may not necessarily be the person whose characteristics informed its selection. The effect of this situation on the accuracy of the model is not yet known, but is not expected to be major.

The procedure is nonetheless considered to be adequate for the short-term to estimate a model and enable ILUTE to run properly. However, for ILUTE simulation purposes that require robust emissions modelling output, a stronger behavioural choice model could be developed in the long term if desired, such as the method used by Vyas *et al.* (2012), as it addresses the types of concerns outlined above.

This primary-driver assignment algorithm ties in nicely with the "preferred vehicle" algorithm developed by Hatzopoulou and reviewed in Subsection 3.4.3 and used to inform vehicle emissions modelling. Rather than the preferred vehicle being randomly selected on a first-come-first-serve basis with random assignment, drivers will now have a designated vehicle that will act as their default choice. This will help by providing more accurate information on which vehicles are being used when, and thus what level of emissions they are generating, one of the major goals of developing the vehicle transaction model.

Finally, it is assumed that in the simulation environment, when a person leaves the household, the vehicle(s) for which they are the primary driver leave with them.

4.5 Other Changes to Improve Estimation and Simulation

In addition to the four large changes discussed in Sections 4.1 through 4.4, as well as the development of a travel sensitive parameters to input into the model that will be discussed in Chapter 5, there are a number of smaller changes that were made to the model. These changes range from the changes to the model scope and structure, as well as information on assumptions being key variables and/or how their definitions have been altered. Each of these changes is documented below.

4.5.1 Company-Owned Vehicles

The TACOS survey collected information on whether vehicles were personally owned, or whether they belonged to the employer of one of the members of the household. According to TACOS data, only 127 out of the 2228 vehicles in the database are company owned; a rate of approximately 5.7%.

The previous model developed by Mohammadian excluded these vehicles from the choice set model estimation on the basis that the vehicle-type selection process (i.e. class and vintage) governing these vehicles would be different than that of vehicles purchased by the households themselves. Although this line of thinking is reasonable, it nonetheless triggered two sources of concern:

- Predicting the correct number of vehicles is the highest modelling priority, and the having the right number of vehicles with incorrect classes and vintages is less of a concern than having the wrong number of vehicles altogether, even if the classes and vintages of the remaining ones are more correct. It seems unlikely that a separate company-vehicle ownership model would be able to be reliably estimated with only 127 vehicles in the dataset. Even if this was possible, there would still need to be a link between the vehicle transaction model and a firmographic model so as to assign which employees get company-owned vehicles. Given the challenges associated with just developing a "regular" vehicle choice model (as will discussed in Chapter 6), creating a separate company-vehicle model seems likely to create more error than it will solve. Furthermore, in the intervening time between the implementation of this model and a firmographic based vehicle ownership model, vehicle ownership would be under-predicted for certain ILUTE households.

- There is also some question as to how different vehicle types would actually be if their purchase decisions were made by companies rather than individuals. Obviously the decision structure would be different (i.e. *"why was this particular vehicle type selected?"*), but not necessarily the outcome (i.e. *"what vehicle was actually selected?"*). A cursory assessment of the company-owned vehicles listed in TACOS shows that a disproportionate amount of company-owned vehicles consist of either mid-to-high-end midsize/large sedans or basic SPVs or Vans. In conjunction with the stated reasons listed into TACOS as to why the particular vehicle type was purchased, it appears that owners of company-owned vehicles are either white-collar workers who are given a nice personal vehicle as an employment benefit, or blue-collar trades/labour workers who are given a vehicle with cargo space which they use to take their equipment or personnel from site to site. In both these cases, employment type, education level and income variables can be set up within the class and vintage choice selection model to account for these patterns, and reduce the error associated with the vehicle types companies provide for their employees.

Based on these considerations, it was decided that the benefits of including company-owned vehicles into the overall model and ignoring the fact that they are company-owned outweigh the negatives. Thus, they will be included in the model and treated the same as household-owned vehicles.

4.5.2 Household Income

As one would expect, income level is an extremely strong indicator of vehicle fleet decisions, including both the number of vehicles purchased as well as their classes and vintages. Almost all models reviewed in the literature review conducted in Chapter 2 included income information as an attribute in the choice modelling process, unless such information was not available in the data set.

TACOS collected income information on what income range the household belongs to, rather than the exact household income (in other words, it collected categorical incomes, rather than numeric incomes). Such a strategy is likely to increase response rates (it may be seen as a slightly less personal/private question) and reduce respondent error (i.e. it's easier for the respondent to remember income within a range rather than the exact number). However, for modelling purposes, having numbers, rather than categories is preferable, as it allows for income to enter utility functions as a linear (or non-linear) attribute rather than just a series of dummy variables. Furthermore, numeric income values provide more flexibility to properly integrate with other parts of ILUTE, as it would obviously be difficult to conduct simulations if each different module within ILUTE had its own uniquely defined categorical incomes ranges that do not match up with each other. Furthermore, given the long time periods that ILUTE is anticipated to model, having numeric values provides a better means of incorporating the effects of inflation and changes in income. For example, the population synthesis developed by Pritchard (2008) has different ranges and a different constant dollar year (see Appendix A.1 of that document) which would be hard to use as a simulation basis for the vehicle transaction model unless both are converted to numeric values.

Based on the above considerations, converting from categorical to numeric incomes was deemed necessary in order to properly incorporate income information into the transaction model estimation as well as provide compatibility for future simulation endeavours. Table 4.10 shows the conversion values that were employed.

Table 4.10: Household Categorical to Numeric Income Conversion Assumptions

TACOS Income Category	**Assumed Numeric Income For Modelling**
$0 - $14,999	$10,000
$15,000 - $29,999	$22,500
$30,000 - $44,999	$37,500
$45,000 - $59,999	$52,500
$60,000 - $75,000	$67,500
Over $75,000	$85,000
Don't know	Excluded from data set
Refusal	Excluded from data set

Generally, TACOS categorical incomes were converted to numeric incomes that are simply the midpoint of the income range in question. The two exceptions are the highest and lowest income categories, each of which is detailed below:

- In the lowest category of $0 - $14,999, a value of $10,000 was used instead of the midpoint value of $7,500. The later value was believed to be unreasonably low, on the basis that there are likely very few households who are earning less than $7,500 annually. Even if only a single person in the household was employed, at the 1998 minimum wage of $5.40/hr, it would only take him/her 8 months to earn $7,500. Therefore it is assumed that in reality most of the households within this category would be clustered closer to the upper end of this range, and hence a value of $10,000 was used.

- The highest-level category does not have an upper bound, and therefore a midpoint value cannot be calculated. A value of $85,000 was selected on the basis that the median level of household income for this category is probably not enormously higher than the lower bound of $75,000. Obviously, there is likely to be a number of exceptionally wealthy households in the dataset that earn well above the $85,000 level, and would probably skew the mean value. However, the high incomes of these few individuals do not make everyone else in this income category richer, and thus a value that was thought to be closer to the median value is preferred. It is further worth noting that for extremely wealthy households, the motivations and decision-structure behind ownership decisions may have very little to do with the fact that they are used for transportation than that they are being collected as a Veblen Good. If a household has far more vehicles than they can possibly drive at any one time, the actual importance of vehicle ownership levels with regards to what ILUTE seeks to model declines (e.g. a no-car household adding three cars represents a major effect on their transportation and activity patterns, a household that already has ten cars adding three more is unlikely to result in any major changes). As was mentioned in Subsection 4.2.3, removing the luxury vehicle designation and capping the highest incomes levels are complementary measures that may help reduce errors caused by each other's simplicity.

A total of 75 households out of 935 (or 8.0%) were missing income information (i.e. either "Don't Know" or "Refusal"), and were therefore excluded from the dataset.

Finally, TACOS collected incomes for the year when the survey was taking place, which most likely means 1997 or anticipated 1998 income levels. Collecting income information for eight or nine separate years is likely to decrease response rates, and may become increasingly subject to memory error as the respondent attempts to remember incomes from almost a decade prior. A multi-year panel survey would likely provide better results, but could suffer from high respondent attrition rates and the fact that it would take close to a decade to collect the equivalent dataset, whereas TACOS was collected in a matter of months.

Nonetheless, the single year of income information poses a problem for a transaction model with multiple years of transaction date available. Given the limits of the available data, there were several potential courses of action:

1. The vehicle transaction model can avoid incorporating income altogether. However, this would remove one of the strongest explanatory variables for the types of behaviour that the model seeks to predict. Removing income would likely introduce more error into the model than it would remove if not-100%-accurate income levels were used, so this course of action is not recommended. Furthermore, it would also remove the policy-sensitivity of the model to changes in income level.

2. Use only one year of data in the model, which would be the year than income level was actually collected for (i.e. the most recent full year - 1997). This would have the effect of removing 80%-90% of the observations in the data set, as only data for a single year of transactions could be modelled. The lack of observations would likely result in poor fitting models and less statistically significant parameters. Furthermore, macro-economic factors which vary from year to year that are hypothesized to influence decision making could not be included, since there would be only one year in the model, and thus there would be no variation in their values. Thus, this procedure may end up causing more error than it solves.

3. Assume that the 1997 or 1998 level of income (if adjusted for inflation) is what the household earns for all years that it is in the TACOS dataset. This effectively assumes no wage growth in real dollar terms; everyone's wages would simply increase/decrease with inflation. This is admittedly a gross simplification of the real world, and introduces some sources of error, but does also have four key mitigating factors. These are:

    (a) The early 1990's were a recessionary period in Ontario. During periods of recession, job security will quickly become a higher priority for many workers than wage increases, and as a result, wages may have been somewhat more stagnant for the first half of the ILUTE dataset, thus reducing some error that would otherwise be incurred.

    (b) There's a strong likelihood that many people/households had an increase in income, but simply moved solely within an income category, rather than between them. In practice, even if incomes were collected each year, many households would likely end up being in the same category year after year, and thus their converted numeric income would be unchanging in any case.

    (c) The biggest sensitivity and source of error is the number of employed workers in the household, and how that changes over the 1990-1998 period. A change in household employment levels is likely to have a much larger effect on changes in income levels than wage growth.

    (d) Unlike the first option, this procedure will still provide (hopefully reasonably correct) income information to the model, and therefore improve its robustness. Unlike the second option, this procedure will also make use of the full nine-year dataset, and improve overall model fit.

Although there are clear risks associated with assuming one year of household income as being representative of a longer eight to nine year period, the overall benefits it provides are believed to outweigh the errors it introduces. It was therefore selected as the preferred option and therefore this is the procedure that is used in the development of the vehicle transaction model. Nonetheless, care should be taken in interpreting the impacts of *changing* income levels of vehicle fleet choices. As Dargay (2001) notes, increases in incomes induce a higher level of increase in vehicle ownership than decreases in income do towards decreases in vehicle ownership. Because the data assume a constant income level for all households for all years, they cannot measure the effect of increases or decreases in income, only (what is assumed to be) the absolute value. Any models developed from this data will likely work well for households/time periods where incomes are reasonably static or have slow-but-steady changes, but may suffer from errors in cases where incomes are rapidly rising or (especially) rapidly decreasing.

4.5.3 Motorcycles

The original model elected to eliminate motorcycles, although did not specify the rationale for this decision (Mohammadian, 2002). However, the decision to exclusion of motorcycles was maintained in the revised model, for several reasons:

- Much of the additional data collected for the other vehicle classes (e.g. design attributes of the vehicle used to generate the price model) was not collected for motorcycles, and would require a substantial amount of manual data recording. Given the available timeline, this was not felt to be the most effective area to focus research efforts.
- Motorcycles represent a very small portion of the overall vehicle fleet within the GTHA, and developing a statistically significant model from TACOS may prove challenging, and it could end up simply being merged with other classes, similar to what happened with Station Wagons.
- Further, given the climate in the GTHA, motorcycles are extremely unlikely to be used year-round by many riders. As such, motorcycle owners must generally have other means of travel available to them for the months where the weather becomes too severe for all but the most dedicated riders to operate their bikes.

Travel models tend to try and represent an average autumn day (this is when the TTS survey is collected); there is often no representation of seasonality of travel patterns over the course of a year in the first place. This does introduce some error into the model in terms of market-share of "fair-weather" modes such as motorcycles and bicycles which are likely to vary throughout the year, and could affect model outcomes for policy-based measures seeking to examine these areas.

ILUTE does have the capability of modelling on both a monthly and annual basis, and thus if a month-by-month travel model was desired, this would be possible (although it would need a year-round source of travel data to develop it). If such a model were created, motorcycles (and bicycles) may become more significant modes in the summer months. In this case, a motorcycle-ownership model could potentially be developed if additional data was collected.

Chapter 5

Incorporating Household Travel Patterns

Arguably the single most significant change made from the original vehicle transaction model is to incorporate actual household travel behaviour into the decision-making process. In other words, *vehicle fleet decision-making will be a function of how useful the vehicle is in helping households productively live their lives.* This process will also allow vehicle class and vintage (and thus indirectly, transaction choices) to be made sensitive to fuel prices and vehicle usage levels.

Unfortunately, TACOS did not collect information on household travel patterns. Although having this information would have been helpful, given that it is a retrospective survey, obtaining daily travel information from households for a date up to nine years in the past is clearly infeasible. Fortunately however, it did collect information on people's place of work/school over the course of that time period.

Given that actual travel data are not available, the next best step is to simulate synthetic travel patterns for each household. With a robust travel model, this method should still provide reasonably accurate results.

Fortunately, such a model is readily available, namely the Travel/Activity Scheduler for Household Agents (TASHA). A high-level review of what TASHA does and how it works has previously been provided in Subsection 3.4.2, and will not be repeated here. Complete documentation can also be found in Roorda (2005).

The literature review on disaggregate ownership and transaction models in Section 2.2 and 2.3 notes that Roorda *et al.* (2009) is arguably one of the most comprehensive representations of this behaviour. Given that Roorda et al. have already fed the same vehicle fleet dataset (TACOS) in the same travel model (TASHA) to help improve the same vehicle transaction model (Mohammadian's), there is little need to "re-invent the wheel" here. Roorda *et al.* (2009) collected two types of variables from TASHA that they then used to inform the transaction choice model. This same procedure will largely be repeated, albeit with some refinements with regards to both the type and time period of the information collected. The variables Roorda *et al.* collected from TASHA are:

- Change in the number of scheduling conflicts resulting from adding or removing a vehicle from the household fleet
- Change in overall household travel utility resulting from adding or removing a vehicle from the household fleet

Both of these variables will be collected here as well, although the household travel utility measure will involve two sub-variables, rather than a single overall value. Additionally, some expansion of modelling scope is also required:

- Change in Vehicle Kilometres Travelled (VKT) output will also be collected to help make choices sensitive to O&M Variable Costs (i.e. a function of VKT, fuel intensity and gas prices).
- Roorda *et al.* (2009) also only make use of 1996 transactions, whereas this model will make use of the full 1990-1998 time period collected in TACOS.

The remainder of this Chapter is broken down into four sections:

- Section 5.1 discusses the process of converting TACOS data into a TASHA-readable format;
- Section 5.2 discusses how TASHA was used to provide these households with travel demand;
- Section 5.3 discusses how the supply of transport (i.e. infrastructure) is represented in the model; and
- Section 5.4 discusses what information was extracted from the model once the simulation was complete and how it is intended to be used.

5.1 Conversion of TACOS Data to TASHA-Readable Format

TASHA was developed to use the information provided by the Transportation Tomorrow Survey as its primary source of data, since that survey is the most extensive repository of travel information in the GTHA. TTS collects information on households, each of the people within the household, and each of the trips made over the course of the day by each of those people. The TTS information required to run TASHA comes from the first two categories; households and persons.

Helpfully, TACOS collected an extensive amount of data from survey respondents regarding both households and persons within the household. However, although it collected the same types of information, actual definitions/survey response options between TTS and TACOS vary. For example, the type of occupation that an individual has is collected in both these surveys, but the categories of occupation they can choose from are different. Thus, in order to be able to model the travel patterns of TACOS households using TASHA, the TACOS household and person information must be "translated" into the definitions required by TASHA. The remainder of this section details what information is required to run TASHA as well as how it was derived from TACOS data.

Table 5.1 details the information required for TASHA Household input file.

Table 5.1: TASHA Household Level Information

Category	**Definition**	**TACOS Source Data**
HouseholdID	Household-specific ID number	TACOS DMU ID (which is being referred to as the "household" throughout this thesis), with year added in front so each household appears each year it is in the dataset.
ZoneCol	Traffic Analysis Zone (TAZ) where the household resides	1996 TAZ Info available in TACOS.
ExpansionFactor (optional)	Data expansion factor	Assumed not to be required for TACOS, as discussed in Section 3.2.
DwellingTypeCol	Physical dwelling type the household resides in	Mapped from TACOS DwellingType data. See Table 5.2 below.
PeopleCol	Number of people in the household	Summation of TACOS Persons in the TACOS DMU.
CarsCol	Number of vehicles in the household	Summation of TACOS Vehicles in the TACOS DMU.

Table 5.2 shows the data mapping process required for Dwelling-Type.

Similarly to Table 5.1, Table 5.3 details the corresponding information required for TASHA Person input file.

Tables 5.4, 5.5 and 5.6 show the data mapping process for employment status, occupation type

Table 5.2: TACOS TO TASHA Dwelling-Type Map

TACOS DwellingTypeID	**TASHA DwellingType**
SFH Single Detached House	House
SFH Semi Detached House	House
Townhouse/Rowhouse	Townhouse
Apt. in a house	Apartment
Apt.(condo) in bdg <5 storeys	Apartment
Apt.(condo) in bdg ≥ 5 storeys	Apartment
Other	Unknown
Don't know	Unknown
Refusal	Unknown
Not Entered	Unknown

Table 5.3: TASHA Person Level Information

Category	**Definition**	**TACOS Source Data**
HouseholdID	Household-specific ID number (same as above)	TACOS DMU ID, with year added in front so each household appears each year it is in the dataset.
PersonID	Person-specific ID number	TACOS Person ID, with year added in front so each person appears each year it is in the dataset.
Age	Self-explanatory	Directly available from TACOS Person data.
Gender	Self-explanatory	Directly available from TACOS Person data.
DriversLicense	Whether person has a driver's license	Directly available from TACOS Person data.
TransitPass (optional)	Whether person has a transit pass	Not available in TACOS, not used.
EmploymentStatus	Self-explanatory	Mapped from TACOS Dwelling-Type data. See Table 5.4 below.
Occupation	Type of employment	Mapped from TACOS Dwelling-Type data. See Table 5.5 below.
FreeParking (optional)	Whether a person has free parking at their place of work.	Not available in TACOS, not used.
Student	Self-explanatory	Mapped from TACOS Dwelling-Type data. See Table 5.6 below.
EmploymetZone StudentZone	TAZ of employment location TAZ of school location	Combined derivation from TACOS Person data. See discussion below.

and student status, respectively. Note that the definitions for EmploymentStatusID were not collected directly in TACOS, but rather subsequently derived by Mohammadian.

Table 5.4: TACOS TO TASHA Employment-Status Ma

TACOS FullTimePartTimeID	**TASHA Employment Status**
Full-time	Full Time
Part-time	Part Time
Don't know	Not Employed
Refusal	Not Employed

Table 5.5: TACOS TO TASHA Occupation-Type Map

TACOS EmploymentTypeID	**TASHA Occupation**
Business, Finance, Administration, Clerical	Office
Science, Math, Engineering and Related	Professional
Health and Medicine	Professional
Social/Gov't Service, Religion, Education, Law	Office
Art, Recreation, Culture, Sport	Office
Sales & Service (inc restaurant,insurance,travel)	Retail
Processing, Manufacturing and Machining	Manufacturing
Trades,Transport Equipment (e.g.carpenter,driver)	Manufacturing
Primary Industries	Manufacturing
Don't know	Unknown
Refusal	Unknown

Table 5.6: TACOS TO TASHA Student-Status Map

TACOS EmploymentStatusID	**TASHA StudentStatus**
Employed (ft or pt)	Not a Student
Student/Preschool	Full Time
Not Employed	Not a Student
Retired	Not a Student
Homemaker	Not a Student
Volunteer	Not a Student
Employed & Student	Part Time
Younger than 16	Full Time
Not a current member	Unknown
Don't know	Unknown
Refusal	Unknown

Finally, the TACOS data contains what it refers to as JobZone, which would suggest it applies solely to a place of employment. However, the data also shows several instances of people who are full time students (but unemployed) as having a JobZone, which implies it was also used as a destination zone for students as well. Since only one JobZone TAZ is available for each person for each year, a method of assigning TAZ information for work and/or school locations was required, particularly in cases where someone is both employed and a student.

- If the person is under 18, then the JobZone TAZ is always assigned to School Zone, since it is assumed they would still be in secondary school, even though they may also have a part-time job.
- If the person has TACOS Employment Status "Employed (ft or pt)" or "Employed & Student" from Table 5.6, then the JobZone TAZ is assigned to Employment Zone. In the case of "Employed & Student", it is assumed that of the two possible locations for response, the respondent would have provided the place of employment.
- If a person is over 18 and has a TASHA StudentStatus of "Full Time", then the JobZone TAZ is assigned to School Zone.

With the above procedures completed, TACOS household and person information can be understood by TASHA and used to model demand.

5.2 Transportation Demand Representation in TASHA

The procedure outlined in Section 5.1 results in the conversion of TACOS information into the "language" of TASHA, and thus it can be used for demand modelling.

Given that the use of TASHA for TACOS households essentially amounts to creating "synthetic data" for TACOS households, it is important to ensure that the outputs are as accurate as possible (i.e. the most similar to what TACOS households would have reported as being their actual travel behaviour, had this question been asked).

As discussed in Subsection 3.4.2, TASHA models activity schedules according to probabilistic distributions, which could result in some households being assigned a particularly "extreme" activity schedule. Similarly, mode choice decisions in TASHA are modelled using a probit model, which also has a level of randomness imposed by its assumption of normally distributed error terms in its utility functions, potentially causing different results from simulation to simulation.

The model was therefore specified to run twenty separate iterations of the 24-hour activity scheduler. Each of these twenty iterations in turn has five iterations of the mode choice model applied to it, creating a total of 100 observations of household travel patterns for each household for each year they are in the dataset. These 100 observations are then averaged to create the final data to be used in the vehicle transaction model estimation, as discussed in Chapter 6. This entire process is repeated three times; one for each of the three vehicle ownership levels (current, plus one, minus one) being assessed.

Finally, it is noted that the particular TASHA model used for this thesis was estimated and calibrated using 1996 Transportation Tomorrow Survey data. It therefore assumes that how household and individual make travel behaviour decisions will remain constant over the nine year (1990-1998) period, even if actual behaviour itself changes. This is the same assumption made it any type of travel forecast for future years, and is commonly referred to as temporal transferability. Although there is some level of risk in this, it is felt to be relatively minor, given:

- That there is only a maximum of seven years difference between the TASHA calibration year (1996) and the farthest-away year being modelled (1990); this is much less of a concern than long-term 30+ year forecasts that are often used in planning studies.
- TASHA, being a "next generation" activity-based model, has a more accurate representation of behaviour, and will hopefully be subject to less error in its forecasts.
- TASHA also uses hard-coded employment/school trip distribution information from external sources. Given that this information is available for TACOS households for all years in the data (per what was discussed in Section 5.1), it means that the distribution of these trips will generally be correct at a disaggregate level.

5.3 Transportation Network Supply Representation in EMME

In addition to the process described above where TACOS data is converted to a TASHA-readable format to generate transport *demand*, a representation of transport *supply* is also required. For this purpose, the regional EMME model was used. EMME is a commercial four-step travel demand forecasting software used in many locales worldwide, including the GTHA.

With both travel demand and travel supply modelled, an iterative process takes place where demand is assigned to the available supply, resulting in travel times and costs, which are then fed back into the next iteration of demand choices. This process repeats until a point of convergence is reached.

In the context of TACOS, the data spans the years 1990 to 1998, and thus these are the years that must have the travel patterns modelled using TASHA. Unfortunately, the only readily available historic network for this time period is the 1996 EMME network, which was available as an "already assigned" set of travel time and cost information for each mode. Specifically, the following information was available:

- 1996 EMME road network for AM, off-peak and PM periods; and
- 1996 EMME transit network for AM and PM periods only, and missing GO Transit.

Use of this EMME network does have the potential to introduce several sources of error, specifically:

- The network is obviously not sensitive to changes in transportation infrastructure (i.e. supply) over the nine year period being modelled; what exists in 1996 is assumed to have existed from 1990 onwards. Thus, in the years 1990-1995, there may be road network capacity or transit service being modelled that did not yet exist. Similarly, in 1997-98, network capacity that was added is not actually reflected in the model.

- Since "pre-assigned" 1996 network information was used, overall network demand over the nine year period is also assumed to be static. Specifically, changing travel conditions are a result of population growth, modal shifts, economic performance etc. are not reflected here; 1996 trip times and costs are assumed to be accurate for all years.

- Variable vehicle operating costs are assumed to be a generic value for all vehicles; although TACOS has vehicle-specific fuel efficiencies, there is no straight-forward way to incorporate this into the travel cost information.

- GO Transit is not present whatsoever in this 1996 network. Although GO Transit represents a very small amount of all trips taken in the region, there is a concern that it provides a specific type of service that could influence the level of vehicle ownership in certain types of households. For example, a suburban family may have both adults working; one in downtown Toronto and one in an auto-dependant location. With GO Transit, one adult can drop-off/pick-up the other and then drive to/from work themselves. If the activity patterns suggest that the family only requires one car for evenings/weekends, then there is no travel-related reason for this family to own a second car. However, without GO Transit present, the transit-to-work alternative for a long distance trip to downtown would be unreasonably long (since it would all be local transit), and thus the utility associated with buying an additional vehicle would be very high, and could possibly influence such a choice. Given that TASHA would show a strong utility for buying an additional vehicle, but the revealed-preference TACOS data would not show such a transaction (since the family is in reality making use of GO Transit), it may result in under-prediction of people's desire for travel utility for making transaction and ownership choices. It is assumed/hoped that such a scenario is infrequent enough to not have an excessively negative impact on the quality of the model estimations.

Using "already assigned" information also means that the additional TACOS households being modelled do not affect overall network performance, but this is not considered to be especially problematic as it only amounts to several hundred households being added to a model of several million, and is thus negligible. Despite these concerns, the network was still the best information available, and was thus made use of.

5.4 Travel Sensitivities Extracted From TASHA

Three quantitative values were extracted from TASHA to provide for each household in order to assess the impacts of a change in the composition of the household vehicle fleet. These are:

- The increase/decrease in the number of daily vehicle-use scheduling conflicts within the household from adding/disposing a vehicle. This is discussed in Subsection 5.4.1.
- The increase/decrease in net household travel utility from adding/disposing a vehicle (consisting of two sub-variables). This is discussed in Subsection 5.4.2.
- Total household vehicle kilometers travelled for current holdings and with one more/less vehicle. This is discussed in Subsection 5.4.3.

This entire process is repeated three times; one with the household's actual level of vehicle ownership, one with one vehicle removed from the household, and one with an additional vehicle added. The differences between the values on the above metrics are calculated to assess the travel impacts implied by certain vehicle fleet choices.

5.4.1 Vehicle-Use Scheduling Conflicts

This output measures the number of vehicle-request scheduling conflicts for the current vehicle ownership level, as well as if one vehicle was added/removed from the household. Vehicle scheduling conflicts are intended to serve as a proxy variable for two phenomena:

- The inconvenience of individuals having to plan or negotiate activities with other household members in advance because of the need to share vehicles. A lack of vehicles removes the ability to spontaneously make vehicle-based trips.
- In order to maximize the utility of overall household travel, TASHA will make individual agents change their trip start/end times by up to 15 minutes in order to have a vehicle available to drive. However, as the TASHA scheduler is purely probabilistic; it does not explicitly account for the utility of leaving/arriving at specific times. Thus, a lack of vehicles may force individuals to travel earlier or later than they actually want to, creating an indirect but unaccounted for travel disutility.

With reference to the figures in Appendix A, conflicts are calculated at the time of the second pass of the mode choice model. The net decrease/increase in vehicle-request conflicts that would result from adding/removing a vehicle from the household is then calculated, and will appear in the transaction choice model for the Buy (BUY_DELTA_NUM_CONF) and Dispose (DISP_DELTA_NUM_CONF) alternatives, respectively. Figure 5.1 shows graphically how these values are obtained:

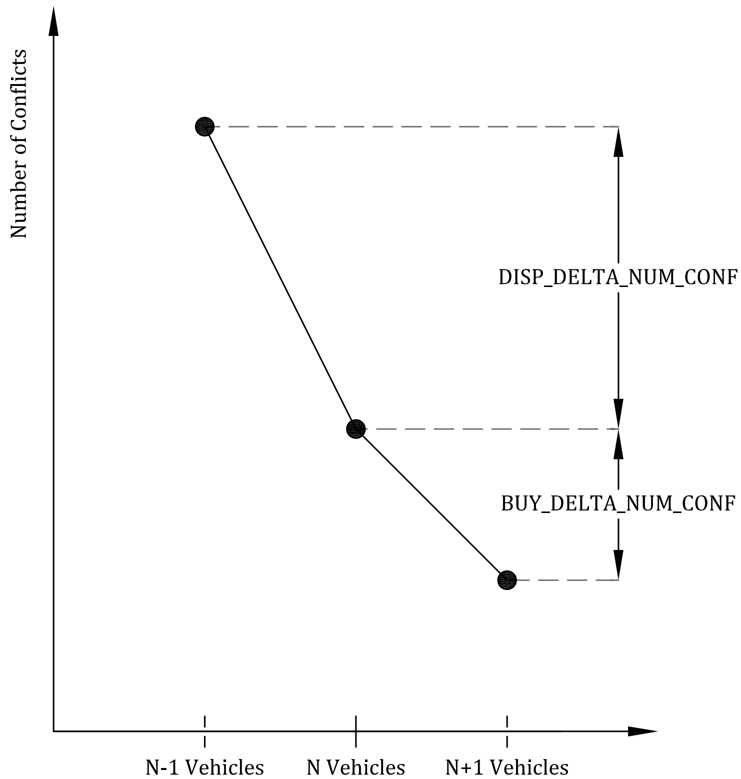

Figure 5.1: Change in Household Vehicle Scheduling Conflicts

There are several key points to note about this process:

- BUY_DELTA_NUM_CONF and DISP_DELTA_NUM_CONF will *generally* be smaller than or equal to zero and greater than or equal to zero, respectively. However, due to the randomness in both the activity generation and mode choice models, it is mathematically possible for adding a car to increase household conflicts or removing a car to decrease conflicts. The intent of simulating travel patterns one-hundred different times is to avoid allowing these "extreme" situations have any major effect on the data.

- An exception to the above rule is no-vehicle households that Buy a vehicle and one-vehicle households that Dispose of a vehicle, which will generally cause conflicts to *increase* and *decrease*, respectively. Although in this situation the model will have the opposite directional effect as it does for all other vehicle ownership levels, it is nonetheless retained. The rationale is that when a no-vehicle household buys a vehicle, it actually *does* increase the intra-household negotiation requirements and time-of-travel compromises that motivate the scheduling-conflict proxy variable compared to not owning a vehicle at all, and thus it is still a valid observation. Obviously, owning a vehicle can still bring net travel benefits, but this already separately accounted for, per Subsection 5.4.2 below. A similar-but-opposite effect applies for one-vehicle households disposing of their vehicle.

- Finally, note that because one hundred different household travel pattern iterations are simulated and then averaged, the change in the number of conflicts for each household used for model estimation will often be a non-integer value, which is not realistic (e.g. one cannot have "three-tenths of a schedule conflict"). For model application purposes, it is up to the user to decide how many iterations of activity scheduling and mode choice they want to use, based on trade-offs between randomness and data storage requirements and processing time.

5.4.2 Overall Household Travel Utility

This output consists of two different measures of overall household travel utility, and calculates both of these for the current vehicle ownership level, as well as if one vehicle was added/removed from the household. Essentially, it quantifies the question *"how much would buying/selling a vehicle impact our household in terms of its ability to go about participating in our daily activities?"*. Deviating from Roorda *et al.* (2009), two separate utilities values are calculated for each ownership level. These are:

- After the second pass of the mode choice model, where vehicle fleet size constraints are applied and household modes have been assigned for each user, with the exception of passenger modes; and
- After the fourth pass of the mode choice model, when passenger mode has also been assigned.

The "Passenger" mode refers to one household picking up and/or dropping off another household at a separate destination on their way to/from their own <u>different</u> destination. When two household members are travelling together to a common destination, the second passenger is using the "Rideshare" mode. Figure 5.2 shows how the household travel utility varies according to what pass the mode choice model is at:

Once this is done for all three ownership levels, the differences in household travel utility resulting from adding/removing a vehicle from the household fleet can be calculated. This creates four different

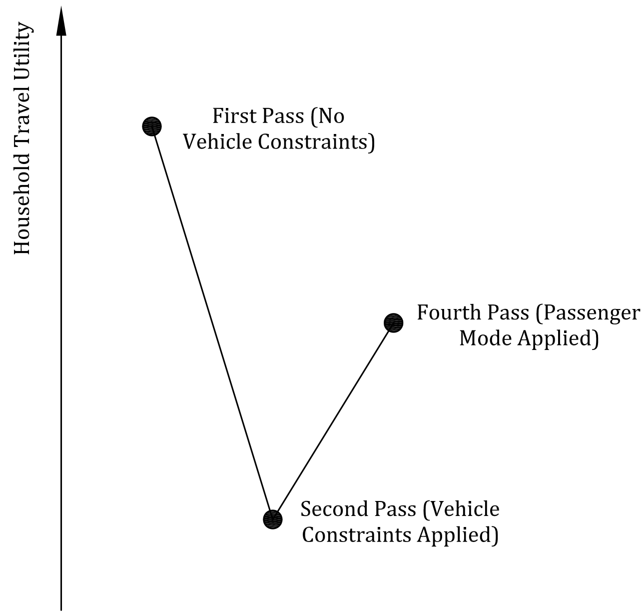

Figure 5.2: Multi-Pass Mode Choice Model Utilities

variables, which are labelled as follows:

- BUY_DELTA_HH_TRAV_UTIL
- BUY_DELTA_HH_PS_UTIL
- DISP_DELTA_HH_TRAV_UTIL
- DISP_DELTA_HH_PS_UTIL

The specific definitions of what each of these variables measures is shown in Figure 5.3 below. There procedure also has several key points to note:

- The intent of calculating two separate delta values (as opposed to just the delta between the first pass and the fourth pass) is that the passenger mode utility recovery may be perceived differently than the initial drop.
- The output from this data is measured in “utils”, which are not a real unit of measurement per se. As such, it is sensitive to changes in the specification of the mode choice model that would alter the scale of the resultant values (e.g. changing the mode choice model structure, what units input variables are measured in, how the alternative specific constants are specified etc.). The data was generated based on the version of TASHA that was most current as of July 2013, and should be adequate for future use of this model and others that are relatively similarly specified. However,

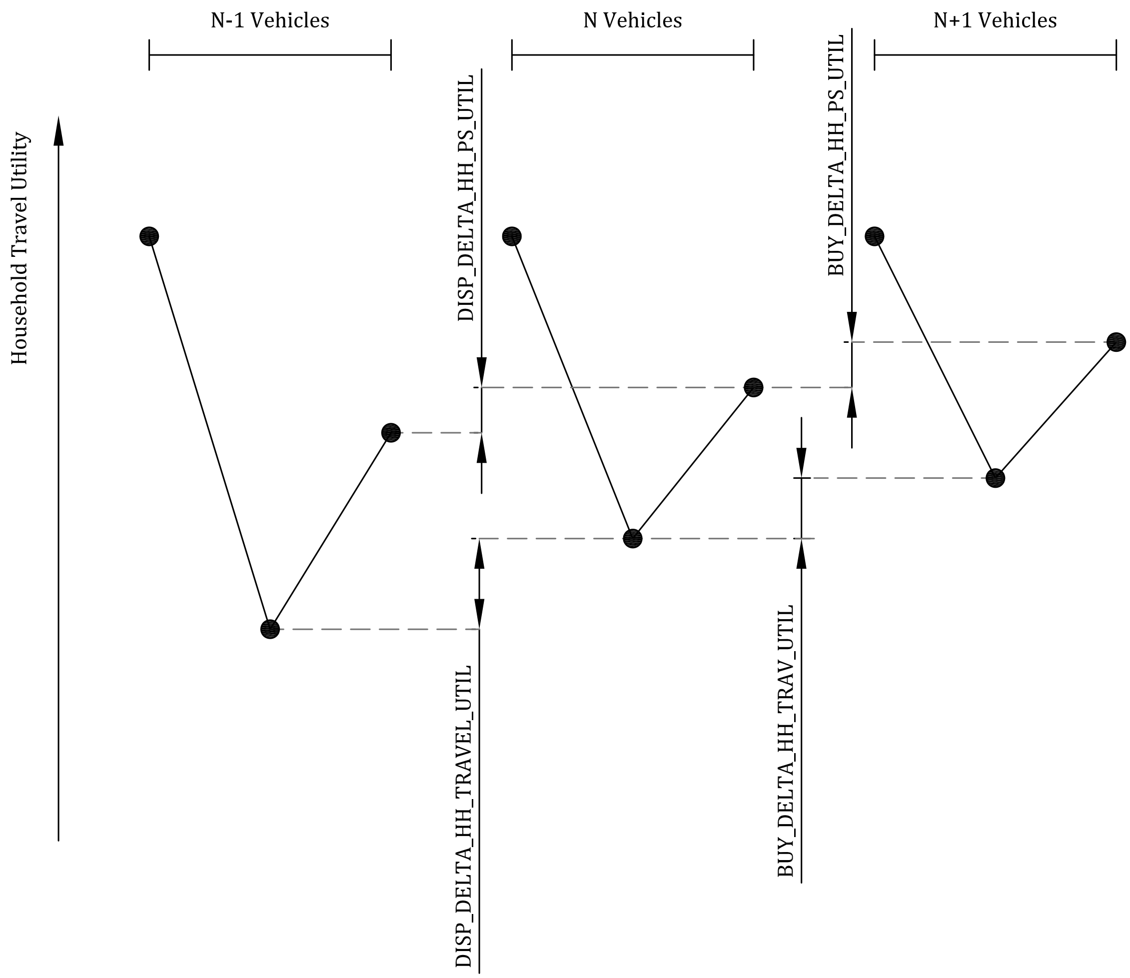

Figure 5.3: Change in Household Travel Utility

all major changes to the mode choice model should involve a corresponding re-estimation of the transaction choice model.

- Unlike the schedule conflict measure, conceptually a zero-vehicle household that buys a vehicle should experience an increase in travel utility resulting from buying the vehicle (or at the very least have the same level of utility, if the extra vehicle is not actually useful). A one-household vehicle that disposes a vehicle is similarly conceptually guaranteed to either have no change (if the vehicle is not actually ever used) or decrease in household travel utility.

- However, as with the schedule conflict measure, there is still a mathematical possibility that adding a vehicle could decrease household travel utility or removing one could increase household utility, due to the randomness of each of these process. This is again treated by averaging a total of 100 different iterations of the household's travel patterns.

5.4.3 Operation and Maintenance Costs Based on Distance Travelled

The third and final quantitative output from TASHA is total household Vehicle Kilometers Travelled (VKT). Under the original model, operating and ownership costs were generally static; they only varied based on the class and vintage of the vehicle. Thus, vehicle operating costs were not sensitive to how much the vehicle is actually used. Per the discussion in Subsection 4.2.6, the revised model seeks to provide a usage-sensitive variable cost component to the Operation and Maintenance Costs that reflects fuel use. This will allow the model to explicitly account for the effect of gasoline prices on the behaviour of households with regards to their vehicle fleet.

TASHA uses the regional EMME network model to generate cost and travel time information for trips as was discussed above in Section 5.4. Once trips are assigned, VKT distances can then be extracted from the simulation. The variable collected here is total household VKT summed across all trips made by all vehicles. Up to three modifications are then made to the VKT value, as follows:

- The VKT values extracted from TASHA are straight-line distances between centroids, not actual travel distance. Since the GTHA is roughly-speaking more of a grid-system of roads and highways, the VKT value is multiplied by **1.4** to get a more realistic value of actual distance travelled.
- TASHA calculates household VKT for a single day. The "actual VKT driven" value is in turn multiplied by **340** to convert this value from an "average weekday" into annual travel.
- A household is assigned a bare minimum of 12,500 kilometres of total household VKT if they own a car. Thus, if the calculated value is lower than 12,500 kilometres it will be raised to this level. This step helps compensate for the fact that the AM transit network was used to model midday transit trips, which may overestimate midday transit mode share since the AM peak will provide higher levels of service (and thus underestimate VKT). Furthermore, some households may use vehicles more on the weekend, where trip patterns are less structured/routine and transit service is inferior.

The above procedure produces the "real" total of total household VKT that will be used for analysis. However, since TASHA treats all vehicles generically and there is no model in use for intra-household vehicle allocation, there is currently no basis for breaking down distance travelled by each vehicle based on model output. Furthermore, when a vehicle is added to or removed from the household fleet, the total household VKT may also change as household members change modes and travel patterns. A separate post-TASHA procedure was developed to provide vehicle-specific VKT values, while also accounting for the change in total driving undertaken by the household. Figures 5.4 and 5.5 show how this is accounted for in the Disposal Choice Model (see Section 6.3) and Class-Vintage Model (see Section 6.4), respectively.

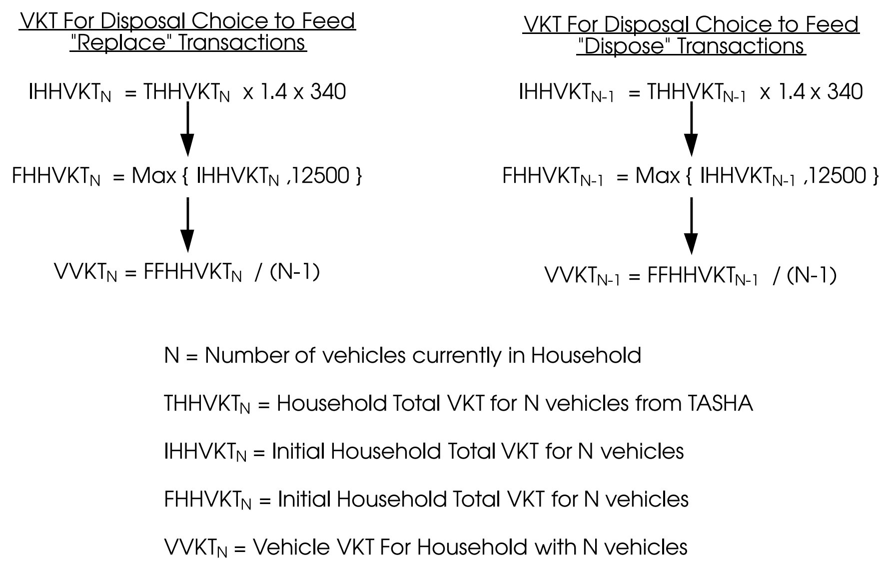

Figure 5.4: Derivation Procedure for Vehicle-Specific VKT in the Disposal Choice Model

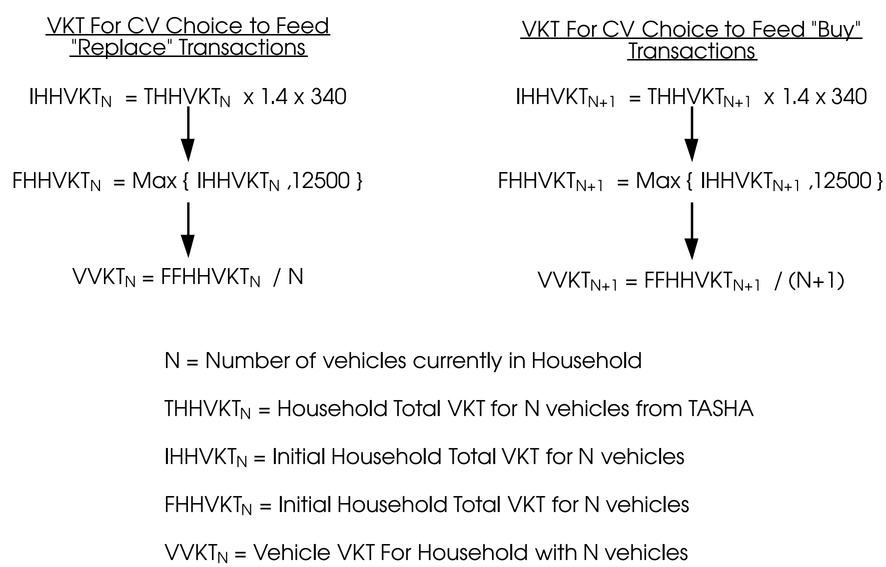

Figure 5.5: Derivation Procedure for Vehicle-Specific VKT in the Class-Vintage Model

Once the vehicles driven by each individual vehicle (or hypothetical vehicle) owned by the household are calculated, the variable costs associated with use of each vehicle can in turn be calculated with the following information:

- The vehicle fuel intensity (L/100km); based on the average of city and highway values. For vehicles in the TACOS database, it is based on their listed values. For hypothetical/simulated vehicles that make use of the virtual vehicle dealership, it is based on of the procedure developed in Section 4.3.

- Vehicle VKT (km) is per what was shown in Figures 5.4 and 5.5.
- Average Toronto area gasoline prices (¢/L), as reported by Statistics Canada. Gas prices used in the model can be found in Appendix C.
- Two adjustment factors are applied. The first multiplies everything by **1.2** in order to account for vehicles not being able to achieve advertised fuel efficiencies. The second factor is solely for unit conversion purposes; it divides everything by **10,000,000** (by 100 to convert VKT to hundreds of km to match fuel intensity units of measure; by 100 to convert gas prices to dollars; and by 1000 to convert dollars to thousands of dollars to match the units used in the O&M Fixed Cost model as well as household incomes).

With these factors in place, the O&M Variable Cost for a vehicle i can be calculated as:

$$OMVAR_i = Fuel\ Intensity\ (L/100km)_i \times VVKT_N(km) \times gas\ price\ (cents/L) \times 1.2 \div 10,000,000$$

For any one vehicle, this procedure is almost identical for all of the four calculation procedures shown in Figures 5.4 and 5.5; only the VVKT value must be adjusted to reflect the number of vehicles in the household (i.e. N, $N-1$, $N+1$).

With this calculations completed, each vehicle now has a usage-dependant Operation and Maintenance Variable Costs for the Disposal Choice model (one each to feed the Dispose and Replace transaction choices) and the Class-Vintage Model (one each to feed the Buy and Replace transaction choices) associated with it that can be used in conjunction with the Operation and Maintenance Fixed Costs described in Subsection 4.2.6 to inform vehicle disposal choices as well as vehicle class and vintage choices, and ultimately transaction choices as well.

Chapter 6

Specification and Estimation of a Revised Vehicle Transaction Model

As detailed in Section 4.1, the revised transaction model will be estimated using a three-stage process, due to the size and complexity of the model, and the difficulties that would be involved in programming and running the estimation as a single procedure. This chapter provides details on the model estimation technique used, general guidelines for the development of the model, and details on estimating each of the three models that make up the overall vehicle transaction model. These topics are broken into subsections as follows:

- The econometric theory that provides the basis for model specification and estimation, detailed in Section 6.1;
- Guidelines for model specifications, detailed in Section 6.2;
- The Vehicle Disposal Choice model, detailed in Section 6.3;
- The Class and Vintage choice model, detailed in Section 6.4; and
- The Transaction Choice Model, detailed in Section 6.5.

Details for all models estimated in this chapter can be found in Appendix D.

6.1 Estimation Technique

This section presents a brief overview of the econometric modelling technique used to estimate the models in this chapter. It focuses largely on the output equations that the analyst will directly work with, rather than the behind-the-scenes computational procedure. The discussion herein in largely adapted from Ben-Akiva & Lerman (1985) and Koppelman & Bhat (2006).

The mathematical models to represent household decision-making behaviour is based on individual (or household, in this case) choice theory. The decision rule for households making decisions that will be employed in this thesis is the concept of utility. Utility is intended to represent how "useful" or "desirable" a particular choice or product is to the decision-maker. Given a selection of potential choices (known as the choice-set) consumers are assumed to select the choice alternative that provides them with the highest level of utility.

In most transportation-related applications, discrete choice theory is used for modelling purposes. The term discrete choice implies that consumers must pick between several discrete alternatives (i.e. one or the other), rather than a continuous spectrum. For example, an individual cannot simultaneously partially travel on two different routes at the same time; they must pick only one.

Discrete choices models can be developed using probabilistic choice theory, where the probability of an individual choosing each alternative in a choice set is assigned a probability. However, representing human behaviour mathematically in terms how decision-makers perceive the utility of different alternatives is challenging. The most common approach for mathematical modelling, and the one that will be used herein is known as Random Utility Maximization (RUM). The operating assumption of RUM is that decision-makers know the true utility of the alternatives that they are choosing between. On the other hand, the analyst can observe some aspects of the utility that decision-makers consider, but there are still other aspects that the decision-maker considers but that the analyst cannot observe. For a person n, the mathematical utility of choice alternative i can therefore be represented as:

$$U_{in} = V_{in} + \epsilon_{in}$$

Where:

- U is the "true" utility that the decision-maker observes
- V is known as the systematic utility, which can be represented mathematically by the analyst
- ϵ is the random error term that represents the unobserved utility that the analyst cannot account for.

The systematic utility can be further assumed to take the form of the following linear-in-parameters function:

$$V_{in} = \beta' x_{in} = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

Here, the x values represent the k attributes that the analyst models the decision-maker as being sensitive to, and the β parameters represent the decision-makers sensitivity to each of those attributes. The linear combination of all of these parameters and attributes quantify the systematic utility for the choice i for the decision-maker n.

The error term ϵ is assumed to take the form of a particular probability distribution. Conceptually, the normal distribution is the strongest candidate, and the resultant models are known as probit models. Unfortunately the normal distribution does not yield a closed-form solution, which makes it challenging to use in practice. Instead, the similarly-shaped Gumbel (or Type I Extreme Value) distribution is used. Models that use this distribution are known as logit models.

With the model specified, β parameters can be estimated using a maximum likelihood search technique. The vector of coefficients of β parameters that results in the log-likelihood value being at its maximum (or least negative) point are then selected for the model. Once the β parameters have coefficients applied to them, then under the logit model the probability of an individual n choosing alternative i can be expressed as:

$$P_n(i) = \frac{e^{V_{in}}}{\sum_{j \epsilon C_n} e^{V_{jn}}}$$

Where C_n is the set of all possible choice alternatives for individual n. With three or more alternatives in the choice set, this is known or the multinomial logit model. As noted in Section 4.1, the multinomial logit will be used for the Disposal Choice, Transaction Choice and Ownership Level models. As would be expected, the probabilities for each alternative in the choice set sum to 1.

$$\sum_{i \epsilon C_n} P_n(i) = 1$$

The multinomial logit features an assumption known as Independence from Irrelevant Alternatives (IIA). In a nutshell, the IIA property will result in inaccurate models in instances where certain alternatives in the choice set are similar to each other compared to other alternatives. The “red bus/blue bus” scenario is the most famous example of this phenomenon. The nested logit allows the IIA property to be circumvented, by grouping the similar alternatives into a single “nest”, so that some of their error term will be common. This is the type of structure that will be used for the Class and Vintage choice model.

The nested logit estimation procedure uses a similar but more complicated maximum likelihood estimation procedure as the MNL. The main additional output that the nested logit has is the logsum parameter θ. The probability of a decision-maker choosing alternative i from within the nest k can be written as:

$$P_n(i|k) = \frac{e^{\left(\frac{V_{in}}{\theta_k}\right)}}{\sum_{j \epsilon C_{nk}} e^{\left(\frac{V_{jn}}{\theta_k}\right)}}$$

Here C_{nk} represents the choice set for person n within the nest k.

The lower-level nests must be solved before the upper-level nests. An output from the lower-level nest is the logsum value, which represents the expected utility for the decision-maker of choosing an alternative within that nest. The logsum (Γ) for the nest k can be calculated as follows:

$$\Gamma_k = \ln\left[\sum_{i \epsilon C_{kn}} e^{\left(\frac{V_i}{\theta_k}\right)}\right]$$

With the logsum value calculated, the probabilities for the upper-level nest k can be calculated as follows:

$$P_n(k) = \frac{e^{(V_{kn} + \theta_k \Gamma_k)}}{\sum_{l \epsilon C_{mn}} e^{(V_{ln} + \theta_l \Gamma_l)}}$$

Note how the logsum value that represents the expected utility from the lower-level choices within nest k influence the likelihood that the decision-maker will choose nest k, and ultimately alternative i within nest k. The probabilities for these two levels can then be multiplied by one another to obtain an overall lower-level probability:

$$P_n(i) = (P_n(i|k))(P_n(k))$$

Since the Class-Vintage Choice model is meant to feed the Transaction Choice model (for the Buy and Replace alternatives), but is estimated separately, logsum values for the m upper level nests can be calculated and used to feed transaction choice t as follows:

$$\Gamma_t = \ln\left[\sum_{i \epsilon C_{mn}} e^{\left(\frac{V_{kn} + \theta_k \Gamma_k}{\theta_m}\right)}\right]$$

A similar process can also be used for the Disposal Choice model to feed the Replace and Dispose alternatives in the Transaction Choice model.

Several statistical tests are available for assessing the overall goodness-of-fit of discrete choice models. The most commonly used is the ρ^2 value. It can be represented as follows:

$$\rho^2 = 1 - \frac{\mathcal{L}\left(\hat{\beta}\right)}{\mathcal{L}\left(0\right)}$$

Where:

- $\mathcal{L}(\hat{\beta})$ is the maximum log-likelihood value of the model (i.e. the one for which all the β parameters have been estimated).
- $\mathcal{L}(0)$ is the log-likelihood when all parameters are set to zero.

The ρ^2 measure is similar in nature to the R^2 measure for regression models. However, while the R^2 value has a direct meaning in terms of goodness of fit (the percentage of the variation in the dependant variable explained by the independent variables), the ρ^2 value has no such direct predictive-accuracy meaning. Although the ρ^2 value will theoretically vary between 0 and 1, the nature of the log-likelihood function makes interpreting its exact meaning difficult. However, generally ρ^2 values in the range of 0.20 to 0.30 represent the lower bound of what would be considered a "good" model.

Similar to the adjusted-R^2, the adjusted-ρ^2 (denoted $\bar{\rho}^2$) measure is often used, especially for large models with many parameters. It is similar to the regular ρ^2, except that it applies a "penalty" of sorts as the number of β parameters (K) in the model increase.

$$\bar{\rho}^2 = 1 - \frac{\mathcal{L}\left(\hat{\beta} - K\right)}{\mathcal{L}\left(0\right)}$$

For individual parameters, the common asymptotic t-statistic test will be used. Additional background on discrete choice modelling can be found in Ben-Akiva & Lerman (1985) or Koppelman & Bhat (2006). All models are estimated with the BIOGEME software package for discrete choice model estimation developed by Bierlaire (2003).

6.2 Model Specification Guidelines

A set of overall goals for the models was developed to act as a guideline for specification and subsequent assessment of the quality of the estimation results. These goals are:

- All variables that are to be endogenously produced within ILUTE/TASHA must either have common definitions or be suitably generic such that sharing simulation data between the different modules within the overall ILUTE modelling framework is straightforward and represents "apples-to-apples" information.

- Variables should ideally have a sound theoretical basis behind why they appear in the utility functions. They should either have an obvious causal relation with the choice being modelled, or a strongly correlated relation which can be defensibly argued will not change over time.

- When exogenous variables are added to model specifications and are already present in other parts of ILUTE (e.g. macroeconomic variables), they should use the exact same data as that already used in ILUTE.

- When ILUTE is eventually used for forecasting, all exogenous variables will need to in turn be forecast in order to run the simulation. When specifying exogenous variables, care should be taken to balance the need for highly-specific variables that improve model fit with the increasing uncertainty in forecasting such highly-specific information for input.

- Variables that are not necessarily statistically significant may nonetheless be important to include in the model if they relate to likely policy-related applications. Being able to model and understand the fact that a policy may have a very marginal and/or scattered effect on overall behavioral choice is a valuable potential application for ILUTE.

- The original transaction model will be used as a guide to examine potential utility function specifications, but will not be strictly adhered to.

- All attributes that incorporate vehicle fleet costs will do so in a manner that presents them relative to the overall budget. Vehicles are sufficiently large purchases that they should not simply be modelled using just benefits provided and costs incurred, because eventually costs reach a hard constraint imposed by household income level, regardless of how much benefit an additional vehicle would provide.

In addition to the above goals, each of the three stages of model development includes a detailed set of definitions of the attributes and the justification for their inclusion discussed in their respected section (in addition to the estimation results).

Note that models presented in this chapter represent the "final" models that were judged to be the overall "best" model that showed the strongest statistical fits possible while also incorporating all of the above-described specification goals. In practice, anywhere from 10-60 different specifications were attempted for each of the three models, many of which incorporated other variables (or alternative specifications for the same variables) that were found to produce inferior results and were discarded.

6.3 Disposal Choice Model

The disposal choice model is used to determine which vehicle in the household vehicle fleet would be disposed given that transaction choice that involved disposal of a vehicle (i.e. Replace or Dispose) was selected by the household. As shown in Figure 6.1, it uses a simple multinomial logit model, but with a variably-sized choice set that corresponds to the number of vehicles in the household fleet.

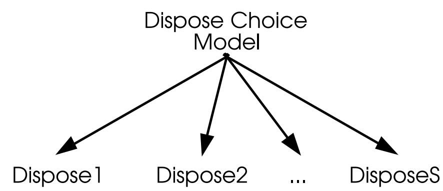

Figure 6.1: Disposal Choice Model Structure

Vehicle holdings for households in TACOS at the time of disposal ranged from one to seven vehicles. As discussed previously in Subsection 3.3.2, the model framework only allows up to one vehicle to be added and/or subtracted from a household in a given year. In the few instances in the data where two or more different vehicles were disposed of in the same year, one of them was randomly selected as the one that was being disposed of, and the others that were disposed of in real-life were assumed to remain part of the household fleet.

Table 6.1 shows the specification of the disposal choice model. This model does not feature any alternative-specific constants because they do not have any obvious meaning in this context. For example, "Alternative 1" does not have any consistent meaning across different households, it is simply the first vehicle listed in for that household in the dataset.

Table 6.1: Disposal Choice Model Variable Definitions

Parameter Name	**Applies to**	**Meaning**	**Justification**
B_HHOMFIX_INC	All	The sum of O&M Fixed Costs for all vehicles in the household except the candidate vehicle that would be disposed of. All of the above is then divided by household income.	See Subsection 4.2.6 and the explanation below for further details. A positive parameter sign is expected.

continued on next page…

Table 6.1: Disposal Choice Model Variable Definitions (continued)

Parameter Name	Applies to	Meaning	Justification
B_HHOMVAR_INC	All	The sum of O&M Variable Costs for all vehicles in the household, less the candidate vehicle that would be disposed of. All of the above is then divided by household income.	See Subsection 5.4.3 and the explanation below for further details. A positive parameter sign is expected.
B_LRGPL	Classes 4,5,6	A dummy variable if the vehicle in question is in the Large class, or larger (i.e. SPV, Van).	There appears to be a general trend towards being more likely to dispose of larger vehicles, all else being equal. A positive parameter sign is expected.
B_USED	All	Vehicle vintage classification is "Used", meaning between 3 and 7 years old.	Households will have a general tendency to dispose of or replace older vehicles, such as those that fall in to the "Used" vintage. A positive parameter sign is expected.
B_OLD	All	Vehicle vintage classification is "Old", meaning 8 years or older	The reasoning is similar to B_USED, except more pronounced as the "Old" vintage represents even older vehicles. A positive parameter sign is expected, and it should be larger in magnitude than B_USED.
B_LUGWB	Classes 1,2,3	Vehicle luggage capacity (m^3) divided by the wheelbase (m).	For smaller vehicles, the ratio of the luggage capacity to the wheelbase appears to play an important role in disposal choice. This variable essentially assesses whether vehicles have lots of storage capacity relative to their overall size. The expected parameter sign is not intuitively obvious here.

As touched on in the table, both B_HHOMFIX_INC and B_HHOMVAR_INC represent the total vehicle O&M Fixed and Variable expenses that the household experience should they dispose the vehicle being considered in the choice alternative; not the savings from the vehicle actually being disposed. Mathematically, this is calculated as follows:

$$HHOMFIX\_INC = \frac{\sum_{i=1}^{S} OMFIX_i - OMFIX_j}{Household\ Income}$$

$$HHOMVAR\_INC = \frac{\sum_{i=1}^{S} OMVAR_i - OMVAR_j}{Household\ Income}$$

Where:

- S = the total number of vehicles in the household fleet
- j = the vehicle currently being considered as the disposal choice
- $OMFIX_{\mathrm{i}}$ = the O&M Fixed cost for vehicle i, as given in Table 4.5.
- $OMVAR_{\mathrm{i}}$ = the O&M Variable cost for vehicle i, as given in Subsection 5.4.3.

The model procedure makes use of all observations of vehicle disposal choices available in TACOS, meaning it includes the vehicles that were disposed of as part of both Dispose and Replace transactions. After cleaning up the TACOS data to remove any instances of households with missing information, a total of 623 vehicle disposal observations across all survey years remained. Use specification detailed above, the model was run in BIOGEME. Table 6.2 summarizes the resultant model.

Table 6.2: Disposal Choice Model Estimation Results

Parameter Name	**Applies to**	**Coefficient**	**Standard Error**	**t-statistic**	**p-value**
B_HHOMFIX_INC	All	22.7	11.8	1.93	0.05
B_HHOMVAR_INC	All	6.75	4.77	1.42	0.16
B_LRGPL	Classes 4,5,6	3.86	0.353	10.93	0.00
B_USED	All	0.945	0.258	3.66	0.00
B_OLD	All	1.55	0.282	5.49	0.00
B_LUGWB	Classes 1,2,3	28.2	2.3	12.28	0.00

Initial Log-likelihood: -1212.302
Final Log-likelihood: -286.852
ρ^2: 0.763
$\bar{\rho}^2$: 0.758

The model appears to perform extremely well; the overall fit as measured by the adjusted rho-squared value is very high, and all parameter are of the correct sign and of reasonable magnitude. Further, all of the parameters except B_OMVAR_INC are statistically significant to the 95% confidence level, and even this particular parameter is still has a reasonably high level of significance.

Key findings from the model estimation indicate that:

- Dollar for dollar, fixed operating and maintenance expenses (B_OMFIX_INC) have an approximately three times stronger influence on disposal choices that variable (i.e. fuel) expenses (B_OMVAR_INC). Two potential explanations are hypothesized for this behaviour. The first is that fixed costs include vehicle maintenance, which is not a regular expense such as fuel (i.e. fill up the gas tank every few weeks). Rather, it arrives in large semi-random increments whenever a component of the vehicle requires repair. Simply being presented up-front with a large repair cost estimate may be the immediate mental trigger that precipitates the decision to dispose of or replace the vehicle. This explanation is supported by the TACOS survey, which asked survey respondents why they disposed of their vehicles. Major repair/maintenance expenses were very frequently cited motivation for disposal or replacement. The second explanation is that during the time period for which TACOS collected data, fuel was relatively inexpensive and stable, and therefore there are few examples of households disposing of a vehicle because the fuel costs were too high. Had the TACOS data been collected 10 years later for the 2000-2008 period, fuel costs may have played a more important role. The consequences of rapidly increasing fuel costs on the temporal transferability of the model are discussed further in Subsection 8.6.1.

- As expected, increasing age is correlated to increasing likelihood of disposal. Compared to a base case of being either Brand New or Nearly New, being of the "Used" vintage has a statistically significant effect on the likelihood of a particular vehicle being disposed. As expected, vehicles that are even older (i.e. the "Old" vintage) have an even larger chance of being selected for disposal. This is again expected to be tied somewhat to the cost of repairs discussed above, but can also reflect the fact that older vehicles will not have all the latest-and-greatest features that many consumers want.

- Potential vehicle resale price does not appear to have a strong effect on disposal choice. A series of models that included the market price of the vehicle at the time of disposal were specified and estimated, but none found this variable to have any significant effect on behaviour. The reasons for this are unclear, but may be relate to the fact that vehicles that would achieve a higher resale price are generally newer, and thus may be recent acquisitions which the household has no intentions of disposing since they only recently acquired it. In other words, the vehicles that the household could sell for the highest price are the ones that are the "best", and thus the ones they want to keep. Few respondents to the TACOS survey indicates that selling price was a major motivation in disposal choices, which further confirms its lack of influence on decision-making.

The strong performance of the model on both a statistical and intuitive level creates a high degree of confidence that it will produce good results in a simulation environment, and it was therefore selected as the model for implementation in ILUTE.

Subsequent to model estimation, the logsum values of expected utilities that would result from disposing a vehicle were calculated for all households in the TACOS dataset (i.e. not just those that actually disposed a vehicle for a particular year, but all households for all years that they are in the dataset). These values are to be used to inform the transaction choice model, a process that will be described in Section 6.5.

6.4 Class and Vintage Choice Model

The class and vintage choice model is used to determine which vehicle type the household would add to their fleet should a transaction choice that involved addition of a vehicle (i.e. Buy or Replace) was selected by the household. As shown in Figure 6.2, it uses a two-level nested logit structure, with six vehicle class nests each containing four vintage categories, for a total of twenty-four vehicles in the choice set.

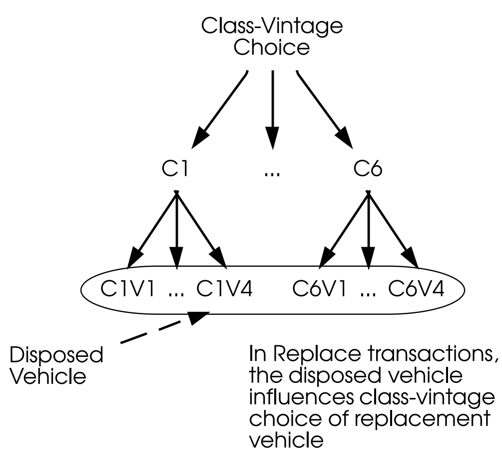

Figure 6.2: Class-Vintage Choice Model Structure

Similarly to the disposal choice model, a maximum of one vehicle is allowed to be purchased each year, and in cases where two or more vehicles were purchased by households in the same year, one of them was randomly selected as being the purchase that was ultimately made. Table 6.3 details the model specification of the final model.

Table 6.3: Class and Vintage Choice Model Variable Definitions

Parameter Name	**Applies to**	**Meaning**	**Justification**
B_V1	V1 for all classes	Alternative Specific Constant for the vintage choice alternative. The same ASC applies to a given vintage for all class types.	Standard component of all discrete choice models. Note that B_V1 will be set to 0 for estimation purposes.
B_V2	V2 for all classes	As above.	As above.
B_V3	V3 for all classes	As above.	As above.
B_V4	V4 for all classes	As above.	As above.
B_HHOMFixOvrHHInc	All[1]	Total household O&M Fixed Costs, including the new vehicle to be bought, divided by household income. If a vehicle is also being disposed (i.e. for a Replace transaction), than the O&M Fixed Cost of the vehicle being disposed is subtracted from the total.	See Subsection 4.2.6 and the explanation below for further details. A negative parameter sign is expected.
B_HHOMVarOvrHHInc	All[1]	Total household O&M Variable Costs, including the new vehicle to be bought, divided by household income. If a vehicle is also being disposed (i.e. for a Replace transaction), than the O&M Variable Cost of the vehicle being disposed is subtracted from the total.	See Subsection 5.4.3 and the explanation below for further details. A negative parameter sign is expected.
B_ChlOvPpl	C6 for all vintages	Number of children in the household divided by the total number of people in the household.	Household with a higher proportion of children were found to be more likely to own Vans (likely minivans) due to their usefulness in shuttling larger numbers of children around. A positive parameter sign is expected.

continued on next page...

Table 6.3: Class and Vintage Choice Model Variable Definitions (continued)

Parameter Name	**Applies to**	**Meaning**	**Justification**
B_DrvMale	C5 for all vintages	Dummy variable if the primary driver (as determined by the procedure outlined in Section 4.4) is male.	Exploratory data analysis indicated that males are more likely to own SPV's (presumably pickups) than females. Whether this is due to the nature of certain employment types such as trades and construction that often require pick-up trucks being disproportionately populated by males, or simply a difference in inherently preferences is not clear, but nonetheless the correlation exists. A positive parameter sign is expected.
B_DrvMgrPro	V1 for all classes	Dummy variable if the primary driver (as determined by the procedure outlined in Section 4.4) has either Manager or Professional employment skills level.	Vehicles whose primary driver is either a Manager or Professional were found to be more likely to purchase Brand New vehicles, even once income level is accounted for. A positive parameter sign is expected.
B_HHVehAvgAge_V1	V1 for all classes	Average age of all vehicles in the current household fleet.	Households that already have newer vehicles are more likely to prefer to purchase newer vehicles. A negative parameter sign is expected.
B_HHVehAvgAge_V4	V4 for all classes	Average age of all vehicles in the current household fleet.	Households that already have older vehicles are more likely to prefer to purchase older vehicles. A positive parameter sign is expected.

continued on next page...

Table 6.3: Class and Vintage Choice Model Variable Definitions (continued)

Parameter Name	Applies to	Meaning	Justification
B_IsClsDisp	All[2]	A dummy variable for whether the class of vehicle is the same as the class of vehicle that is being disposed of in a replace transaction.	The TACOS data reveals that the one of the main reasons for disposing or a certain vehicle is that it is simply too old, and one of the main reasons for buying a particular new one is because they liked the old one they had to get rid of. In this context, it means that many households like the vehicle they already have; they just need a newer one. Thus, there will be a tendency that all else being equal, households will be disproportionately more likely to replace a disposed vehicle with one of the same class. This variable represents an attempt to account for vehicle type choices being partially dependant on the existing household fleet, but avoiding the mathematical complexities associated with a full model of such a phenomenon. A positive parameter sign is expected.
B_PPOvrHHInc	All	Vehicle purchase price (as calculated by the hedonic price model) divided by household income.	Households will consider the purchase price of the candidate vehicle as part of their decision-making, and will consider it in the context of their household income level, which represents their ability to afford it. A negative parameter sign is expected.

continued on next page...

Table 6.3: Class and Vintage Choice Model Variable Definitions (continued)

Parameter Name	**Applies to**	**Meaning**	**Justification**
B_LnPPHHInt	All	Natural logarithm of the vehicle purchase price (as calculated by the hedonic price model) divided by household income and divided by the interest rate expressed in decimal form).	Similar to the above parameter, except this one takes into account the effect of loan costs in order to buy the vehicle. Because loan expenses are not necessarily linear with interest rates, a logarithmic transformation is applied. Thus, increased interest rates will decrease the magnitude of this attribute for a given purchase price and income level. A positive parameter sign is expected, since this would mean that the purchase is "less good" as interest rates rise.
B_VPF	All	Vehicle Performance Factor for the vehicle as specified in Table 3.2 and calculated using either real/observed vehicle properties from TACOS or simulated properties for "virtual vehicles" as shown in Figure 4.4.	Together with the VSF, the VPF is a combination of many of the physical features of a vehicle that households will consider when making a purchase decision. The expected parameter sign is unclear because the VPF has no direct physical meaning.
B_VSF	All	Vehicle Space Factor for the vehicle as specified in Table 3.2 and calculated using either real/observed vehicle properties from TACOS or simulated properties for "virtual vehicles" as shown in Figure 4.4.	Together with the VPF, the VSF is a combination of many of the physical features of a vehicle that households will consider when making a purchase decision. The expected parameter sign is unclear because the VPF has no direct physical meaning.
SUBCOMPACT	C1 nest	Nest co-efficient for the class nest, which contains the four vintage choices for the particular class.	Standard component of a nested logit model.
COMPACT	C2 nest	As above.	As above.
MIDSIZE	C3 nest	As above.	As above.
LARGE	C4 nest	As above.	As above.
SPV	C5 nest	As above.	As above.

continued on next page...

Table 6.3: Class and Vintage Choice Model Variable Definitions (continued)

Parameter Name	Applies to	Meaning	Justification
VAN	C6 nest	As above.	As above.

[1]Different attribute levels for Buy and Replace transaction CV choices.
[2]For Replace transactions only.

As touched on in the table, both B_HHOMFixOvrHHInc and B_HHOMFixOvrHHInc represent the total vehicle O&M fixed and variable expenses that the household experience should they acquire the vehicle being considered in the choice alternative (and in the case of the Replace alternative, also taking into account the savings from the vehicle that is being disposed of). The intent is that adding an expensive-to-operate-and-maintain vehicle to a household with no vehicles represents less of a financial; burden than adding the same vehicle to a household that already has several cars they must pay for. In other words, vehicle choices are considered in the context of how they affect overall household transportation spending, rather than just as an isolated vehicle. The calculation method for the O&M costs of class-vintage alternatives in the Class-Vintage Model that will feed the Buy and Replace alternatives in the Transaction Choice Model, respectively, are given below.

$$HHOMFixOvrHHInc\ (for\ Buy) = \frac{\sum_{i=1}^{S} OMFIX_i + OMFIX_j}{Household\ Income}$$

$$HHOMVarOvrHHInc\ (for\ Buy) = \frac{\sum_{i=1}^{S} OMVAR_i + OMVAR_j}{Household\ Income}$$

$$HHOMFixVarOvrHHInc\ (for\ Replace) = \frac{\sum_{i=1}^{S} OMFIX_i + OMFIX_j - OMFIX_k}{Household\ Income}$$

$$HHOMVarOvrHHInc\ (for\ Replace) = \frac{\sum_{i=1}^{S} OMVAR_i + OMVAR_j - OMVAR_k}{Household\ Income}$$

Where:

- j = the class and vintage of the choice alternative being considered
- k = the vehicle that has been selected by the household for disposal, according to the Disposal Choice Model

- $OMFIX_{\mathrm{i}}$ = the O&M Fixed cost for vehicle i, as given in Table 4.5.
- $OMVAR_{\mathrm{i}}$ = the O&M Variable cost for vehicle i, as given in Subsection 5.4.3.

Per the discussion in Section 4.2, all vehicles that were actually purchased had their "true" vehicle properties used, with the exception of the purchase price which is calculated using the hedonic price model (and had the motivation for this presented in Subsection 4.2.5). All twenty-three other class-vintage combinations make use of the virtual vehicle dealership developed in Section 4.2.

The model estimation makes use of all observations of vehicle procurement choices available in TACOS, meaning it includes the vehicles that were acquired as part of both Buy and Replace transactions. After cleaning up the TACOS data to remove any instances of households with missing information, a total of 998 vehicle type purchase observations remained for all years in the dataset. Use specification detailed above, the model was run in BIOGEME. Table 6.4 summarizes the resultant model.

Even with considerable effort, and the estimation of dozens of different models, a well-fitted model for the class and vintage choice proved to be elusive. The overall fit as measured by the adjusted rho-squared is generally quite poor, and below what is often considered to represent a "good" fit (in the range of 0.20 to 0.30).

Despite this, all lower-level parameters are of the correct/anticipated sign, and most of them show strong levels of statistical significance. The Gumbel scale parameters for the nests are all larger than one, implying that the logsum values are all between 0 and 1, which is consistent with theory. Furthermore, the t-statistics for the scale parameters are all insignificant against a value of 1. Since the Gumbel scale parameters being equal to 1 would imply that the model could be simplified to a single-level multinomial logit, the insignificance of these parameters against a value of 1 implies that grouping the vehicles by class nests was a valid means of structuring the model.

Additional findings from the model include:

- Dollar for dollar, O&M Variable Costs appear to be about six times more important to potential buyers than O&M Fixed Costs.
- Both O&M Fixed and Variable costs (B_HHOMFixOvrHHInc and B_HHOMVarOvrHHInc, respectively) are substantially more important than the main variable for purchase cost, B_PPOvrHHInc. However, it should be noted that the effects of vehicle purchase price (and finance costs) are also incorporated into the B_LnPPHHInt function, and thus purchase costs will influence decision-making more than a direct examination of parameter values would suggests. Nonetheless, it is expected that dollar for dollar the two O&M variables would still be weighted much higher than purchase

Table 6.4: Class-Vintage Choice Model Estimation Results

Parameter Name	**Applies to**	**Coefficient**	**Standard Error**	**t-statistic**	**p-value**
B_V1	V1 for all classes	0	fixed		
B_V2	V2 for all classes	-0.616	0.266	-2.32	0.02
B_V3	V3 for all classes	-0.257	0.185	-1.38	0.17
B_V4	V4 for all classes	-0.888	0.43	-2.07	0.04
B_HHOMFixOvrHHInc	All[1]	-8.91	6.07	-1.47	0.14
B_HHOMVarOvrHHInc	All[1]	-52.3	13.3	-3.93	0.00
B_ChlOvPpl	C6 for all vintages	1.65	0.433	3.82	0.00
B_DrvMale	C5 for all vintages	0.544	0.179	3.05	0.00
B_DrvMgrPro	V1 for all classes	0.339	0.174	1.95	0.05
B_HHVehAvgAge_V1	V1 for all classes	-0.0352	0.018	-1.96	0.05
B_HHVehAvgAge_V4	V4 for all classes	0.0481	0.026	1.85	0.06
B_IsClsDisp	All[2]	0.979	0.0955	10.25	0.00
B_LnPPHHInt	All	0.688	0.356	1.93	0.05
B_PPOvrHHInc	All	-0.989	0.466	-2.12	0.03
B_VPF	All	-0.192	0.14	-1.37	0.17
B_VSF	All	-0.242	0.87	-0.28	0.78
Gumbel Scale Parameter[3]					
SUBCOMPACT	C1 nest	1.14	0.448	0.34[4]	0.73
COMPACT	C2 nest	1.27	0.402	0.57[4]	0.57
MIDSIZE	C3 nest	1.08	0.505	0.16[4]	0.87
LARGE	C4 nest	1.34	0.464	0.47[4]	0.64
SPV	C5 nest	1.23	0.711	0.51[4]	0.61
VAN	C6 nest	1.99	0.934	0.29[4]	0.29

Initial Log-likelihood: -3171.698
Log-likelihood at constants: -2930.125
ρ^2: 0.076
$\bar{\rho}^2$: 0.070

[1]Different attribute levels for Buy and Replace transaction CV choices.
[2]For Replace transactions only.
[3]BIOGEME outputs the Gumbel scale parameter; the inverse of the logsum parameter.
[4]t-test against 1.

costs, for the simple reason that vehicle purchase is a one-time capital cost whereas O&M costs are incurred every year, so households would perceive these costs in the context of the expected amount of time they would own the vehicle rather than a single years' worth.

- As was anticipated, all else being equal, replacement vehicles are more likely to be the same class that the household disposed of.
- Demographic factors such as B_ChlOvPpl, B_DrvMale and B_DrvMgrPro are also functioning as anticipated.

The strong performance of individual parameters is therefore somewhat at odds with the poor overall fit of the model. It suggests that although all parameters included in the model are important, there are other issues that are also important that have not been accounted for.

A review was also undertaken of TACOS households' stated reasons for purchasing their particular type of vehicle for clues. Several of the motivations given by survey respondents match those already included in the model. Paraphrased examples of some of these responses are:

- *"This vehicle was a good price and/or affordable"* (accounted for in B_PPOvrHHInc and B_LnPPHHInt)
- *"I wanted a vehicle that was roomy/big"* (accounted for in B_VPF, B_VSF and logusum co-efficents)
- *"I already own this vehicle, I just needed a new one"* (accounted for in B_IsClsDisp)
- *"It gets good mileage"* (accounted for in B_HHOMVarOvrHHInc)
- *"This vehicle is reliable and easy to maintain"* (accounted for in B_HHOMFixOvrHHInc)

Note that mileage-related motivations were cited more frequency frequently that maintenance-cost related motivations, which lends credibility to its higher parameter estimate.

However, there were also a number of stated reasons survey respondents gave for purchasing a particular class and vintage that are not able to be represented in either this choices model or any model at all. Examples of these types of responses (again, paraphrased) include:

- *"A friend was selling it"*
- *"My friends and/or family members have the exact same one"*
- *"I always buy cars from [manufacturer]"*
- *"It looks fast and I liked it"*
- *"The sales guy did a good job"*
- *"I don't know"*

It is clear that the decision-making process for vehicle type choice is incredibly heterogeneous, and that no single model is easily able to represent the wide range of general priorities that individuals have when they make such decisions, nor the circumstances that surround each individual purchase.

Based on the literature review on vehicle type choice undertaken in Section 2.4, there a number key types of attributes that other models have used that have not been included here.

Models such as the ones developed by Choo & Mokhtarian (2004) and Cao *et al.* (2006) make use of attitudinal factors of drivers and households. Although their results are convincing, they are not applicable to this work. Firstly, because TACOS data did not collect attitudinal factors, and more critically, because it is not possible to simulate agents' attitudes, and thus even if the model could be estimated it could not be run in ILUTE.

Another strategy used in models such as those developed by Cao *et al.* (2006) and Potoglou (2008) is to use a highly spatially disaggregate variables such as the provision of transit service and land use mix that are calculated with GIS software. This has been found to offer some improvement in model performance (although not nearly enough to result in substantially stronger model performance). In ILUTE, dwellings are located on the basis of Census Tract and/or Traffic Analysis Zone (TAZ), rather than highly disaggregate parcel-specific locations. Without this parcel level information, using transit-proximity measures is risky, as "average distance" from the centroid of a census tract or TAZ introduces a level of aggregation bias into the data. Furthermore, most of the literature only discusses collection of this information for the purpose of estimating a model. In the case of ILUTE where the long term goal is year-by-year simulation, developing this process for several decades' worth of simulations could easily run into data difficulties, particularly in newer exurban and car dependent areas where transit services and land use patterns could be expected to change substantially over time.

Finally, Manski & Sherman (1980) note the interdependence of the choice of different vehicle types in a multi-vehicle household, suggesting that the classes of the vehicles already in the household may influence class choice of new vehicles. Unfortunately, despite modelling every conceivable influence in terms of owning certain classes making other classes (including vehicles of the same class) more/less likely to be purchased, no statistically significant correlations were able to be detected.

Therefore, in spite of its shortcomings, the model shown in Table 6.4 is recommended for use in ILUTE because it is unlikely that a substantially better performing model could be developed without considerable long term effort. Even then the gains may be minor, as although the models in the papers cited above did perform better, the difference is relatively marginal.

Similarly to the disposal choice model, after model estimation, the logsum values of expected utilities that would result from purchasing a vehicle were calculated for all households in the TACOS. Two different logsum values are calculated for each household; one for the Buy alternative and one for the Replace alternative. The reason is that some variables such as B_IsClsDisp only apply to the Replace alternative, and some attributes will have different values for Buy and Replace and different (e.g. total household expenditures on O&M Fixed and Variable costs). These values are also used to inform the transaction choice model, a process that will be described in Section 6.5.

6.5 Transaction Choice Model

The third and final model to be developed as part of the vehicle transaction model is the transaction choice model. The transaction choice model is used to determine what action that household will take each year with regards to their vehicle fleet. Possible choices consist of Buying a new vehicle, Doing Nothing, Replacing a vehicle or Disposing of a vehicle already in their fleet.

The transaction choice model was estimated last because it uses information about the expected utility to be gained from the particular vehicle type the household would acquire if it was to acquire a vehicle, as well as the expected utility to be gained from disposing of a particular vehicle if it was to dispose of a vehicle in order to influence the transaction choice.

The transaction choice model is a multinomial logit model featuring the four choice alternatives described above and shown in Figure 6.3. Similar to the disposal choice model, the choice set is dynamic; households with no vehicles are limited to the Buy and Do Nothing choices, while households with at least one vehicle will have access to all four choice alternatives. Only one transaction choice can be made each year, and a maximum of one vehicle can be added or removed.

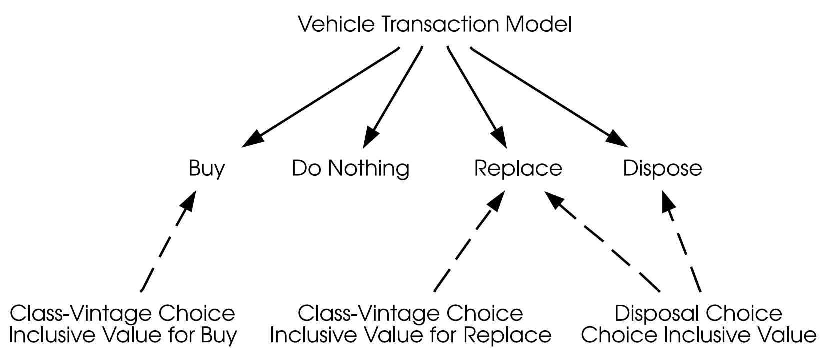

Figure 6.3: Transaction Choice Model Structure

Table 6.5 details the model specification.

Table 6.5: Transaction Choice Model Variable Definitions

Parameter Name	**Applies to**	**Meaning**	**Justification**
B_BUY	Buy	Alternative Specific Constant for the choice alternative.	Standard component of all discrete choice models. Note that B_BUY will be set to 0 for estimation purposes. The expected parameter signs are not intuitively obvious here.
B_REPL	Replace	As above.	As above.
B_DN	DoNothing	As above.	As above.
B_DISP	Dispose	As above.	As above.
B_BUY_DELTA_NUM_CONF	Buy	The change in the number of vehicle-request scheduling conflicts in a household after the second pass of modal assignment that would result from adding an additional vehicle to the household fleet.	See Subsection 5.4.1 for further details. A negative parameter sign is expected.
B_BUY_DELTA_HH_TRAV_UTIL	Buy	The change in the overall household travel utility prior to passenger assignment (i.e. after the second pass of modal assignment) that would result from adding an additional vehicle to the household fleet.	See Subsection 5.4.2 for further details. A negative parameter sign is expected.
B_BUY_DELTA_HH_PS_UTIL	Buy	The change in the overall household travel utility after passenger assignment (i.e. after the fourth pass of modal assignment) that would result from adding an additional vehicle to the household fleet.	See Subsection 5.4.2 for further details. A positive parameter sign is expected.

continued on next page...

Table 6.5: Transaction Choice Model Variable Definitions (continued)

Parameter Name	**Applies to**	**Meaning**	**Justification**
B_HHINC	Buy	Dummy variable for the number of persons in the household has increased since the previous simulation year.	If there are more people in the household, there are likely to be more trips being made, and thus potentially more of a need for an additional vehicle. A positive parameter sign is expected.
B_JBINC_BUY	Buy	Dummy variable for the number of employed persons in the household has increased since the previous simulation year.	Even more so than a change in the number of people in the household, the change in the number jobs in the household may affect the number of vehicles required on two levels. Firstly, commuting is often a time-inflexible trip, and the number of people vehicles for commuting may govern total ownership requirements. Secondly, because more employment within the household means more income which can be used to purchase and operate vehicles. A positive parameter sign is expected.
B_JBINC_REPL	Replace	As above.	As above.
B_DLGTFS	Buy	A dummy variable for when the number of drivers licenses in the household is greater than the vehicle fleet size.	With more potential drivers than available vehicles, there is likely to be additional pressure to purchase more vehicles, even beyond the number of conflicts calculated in B_BUY_DELTA_NUM_CONF. For example, persons within the household may want their "own" vehicle, even if they could share. A positive parameter sign is expected.

continued on next page...

Table 6.5: Transaction Choice Model Variable Definitions (continued)

Parameter Name	Applies to	Meaning	Justification
B_NUMCHLD	Buy	The number of children (defined as being younger than 18) in the household.	Children may encourage additional vehicle purchases as they can create more distributed trip patterns for a family that necessitate another vehicle in order to complete all daily activities. For example, a parent that used to be able to bus to work but now has to drop their kids off at school, pick them up from day-care by a certain time and run them to and from swimming/music etc. lessons may find that a vehicle is the only reasonable way of doing this. A positive parameter sign is expected.
B_INTRT_BUY	Buy	Sensitivity of buy and replacement choices to interest rates.	Lower interest rates make it cheaper for households to take out a loan to buy a new vehicle. This variable is intended to reflect the effects of the cost of borrowing on the likelihood of purchasing a vehicle. Negative parameter signs are expected.
B_INTRT_REPL	Replace	As above.	As above.

continued on next page...

Table 6.5: Transaction Choice Model Variable Definitions (continued)

Parameter Name	Applies to	Meaning	Justification
B_UNEMPRT_BUY	Buy	Sensitivity of buy and replacement choices to unemployment rates.	This variable is intended to act as a proxy for economic uncertainty. Although actual employment and income within the household is explicitly modelled, this variable is intended to represent the presence of the sort of economic malaise that may suggest that household members are at a risk of losing the jobs they currently have, and thus are avoiding the financial risks associated with making a major purchase such as a vehicle. Negative parameter signs are expected.
B_UNEMPRT_REPL	Replace	As above.	As above.
B_ZERO_VEH	Buy	Dummy variable for a household that owners no vehicles.	If households have no vehicles, they are more likely to purchase at least one to have it at their disposal. A positive parameter sign is expected.
B_CVINCVAL	Buy, Replace	The logsum value for the expected utility of the vehicle selected for purchase from the class-vintage choice model.	The expected utility gain that arises from buying or replacing a vehicle stems in part from the expected utility provided by whichever vehicle is selected for purchase, as calculated using the model developed in Section 6.4. Note that different attribute values will be calculated for the Buy and Replace class-vintage alternatives. A positive parameter sign is expected.

continued on next page...

Table 6.5: Transaction Choice Model Variable Definitions (continued)

Parameter Name	Applies to	Meaning	Justification
B_AGEAVG	Replace	The average age of all vehicles in the household fleet.	A lower average age would suggest that there is less of an immediate need to replace any vehicles. A positive parameter sign is expected.
B_DINCVAL	Replace, Dispose	The logsum value for the expected utility resulting from disposing an existing vehicle in the household fleet from the disposal choice model.	The expected utility gain that arises from replacing or disposing a vehicle stems in part from the expected utility provided by disposing whichever particular vehicle is selected for disposal, as calculated using the model developed in Section 6.3. A positive parameter sign is expected.
B_NUM_VEH	Do Nothing	The number of vehicles owned by the household.	The more vehicles a household has, the less likelihood that will choose a Do Nothing action, as there is more likely to be a need to replace (or dispose of) vehicles. A negative parameter sign is expected.
B_OLD_NN	Do Nothing	A dummy variable for whether the oldest vehicle in the household is of a Nearly New vintage or newer.	When the entire household vehicle fleet is at most two years old, then they are less likely to need to make any changes to their fleet (especially replacing or disposing of a vehicle, since all vehicles would have been acquired only very recently). A positive parameter sign is expected.

continued on next page...

Table 6.5: Transaction Choice Model Variable Definitions (continued)

Parameter Name	**Applies to**	**Meaning**	**Justification**
B_LNYRSTRAN	Do Nothing	The natural logarithm of (number of years since the last non-Do Nothing + 1)	Households have been observed to have a "cool down" period between transactions that involve making a change to their vehicle fleet (i.e. everything other than "Do Nothing"). It may take them a number of years to either assemble appropriate financial resources to purchase a new vehicle, or simply have enough stress and aggravation from not owning enough vehicles, or even to conclude that they have too many vehicles, before they make a change.
B_DISP_DELTA_NUM_CONF	Dispose	The change in the number of vehicle-request scheduling conflicts in a household after the second pass of modal assignment that would result from removing an existing vehicle from the household fleet.	See Subsection 5.4.1 for further details. A negative parameter sign is expected.
B_DISP_DELTA_HH_TRAV_UTIL	Dispose	The change in the overall household travel utility prior to passenger assignment (i.e. after the second pass of modal assignment) that would result from removing an existing vehicle from the household fleet.	See Subsection 5.4.2 for further details. A negative parameter sign is expected.

continued on next page...

Table 6.5: Transaction Choice Model Variable Definitions (continued)

Parameter Name	**Applies to**	**Meaning**	**Justification**
B_DISP_DELTA_HH_PS_UTIL	Dispose	The change in the overall household travel utility after passenger assignment (i.e. after the fourth pass of modal assignment) that would result from removing an existing vehicle from the household fleet.	See Subsection 5.4.2 for further details. A positive sign is expected.
B_HHDEC	Dispose	Dummy variable for the number of persons in the household has decreased since the previous simulation year.	If there are less people in the household, there are likely to be less trips being made, and thus potentially less of a need for vehicle, leading to one of them being disposed of. A positive parameter sign is expected.
B_JBDEC	Dispose	Dummy variable for the number of employed persons in the household has decreased since the previous simulation year.	Even more so than a change in the number of people in the household, the change in the number jobs in the household may affect the number of vehicles required on two levels. Firstly, commuting is often a time-inflexible trip, and the number of people vehicles for commuting may govern total ownership requirements. If there are less people commuting, then there may be less of a need for that addition vehicle which only gets used for commuting. Secondly, if there are less employed individuals within the household, then there is less income with which to operate and maintain vehicles. A positive parameter sign is expected.

continued on next page...

Table 6.5: Transaction Choice Model Variable Definitions (continued)

Parameter Name	**Applies to**	**Meaning**	**Justification**
B_DLLTFS	Dispose	A dummy variable for when the number of drivers licenses in the household is less than the vehicle fleet size	In this situation, the household would have more vehicles than they can use at any one time, and could dispose of a vehicle with no negative transportation-related consequences or effects on the desire s of household members to own their own vehicle (ignoring the possibility that the same household member may have different vehicles for different transportation tasks). A positive parameter sign is expected.
B_CHDEC	Dispose	Dummy variable for the number of children in the household has decreased since the previous simulation year.	Households where a child leaves the household (e.g. moves out) are likely to dispose of a vehicle as they may find they are less in need of one.

The model generally aims to make use of all households in TACOS for all years that they were available, although two years are ultimately excluded:

- 1990 transactions are excluded because certain variables used in the model (i.e. dummy variables for number of people/jobs in the household increasing/decreasing) are relative to the previous year. Since there is no information for 1989, these values cannot be calculated for 1990. Thus, the 1991 transactions are the earliest to appear in the model, since they make use of the 1990 data.

- 1998 is excluded altogether from the transaction choices (although was included for disposal and class and vintage choices models). The reason for this is that the TACOS survey was collected in early 1998, and households making transactions later in the year are not accounted for. Included the data would have therefore resulting in a model that predicted higher levels of "Do Nothing" than what is actually occurring.

After cleaning up the TACOS data to remove any instances of households with missing information, a total of 4164 observations of transaction choices remained. Use specification detailed above, the model was run in BIOGEME. Table 6.6 summarizes the resultant model.

Table 6.6: Transaction Choice Model Estimation Results

Parameter Name	**Applies to**	**Coefficient**	**Standard Error**	**t-statistic**	**p-value**
B_BUY	Buy	0			Fixed
B_REPL	Replace	2.85	0.676	4.21	0.00
B_DN	Do Nothing	0.178	0.845	0.21	0.83
B_DISP	Dispose	-1.85	0.679	-2.72	0.01
B_BUY_DELTA_NUM_CONF	Buy	0.226	0.0835	2.71	0.01
B_BUY_DELTA_HH_TRAV_UTIL	Buy	-0.193	0.0826	-2.33	0.02
B_BUY_DELTA_HH_PS_UTIL	Buy	-0.333	0.222	-1.5	0.13
B_HHINC	Buy	0.321	0.201	1.59	0.11
B_JBINC_BUY	Buy	0.242	0.187	1.29	0.20
B_DLGTFS	Buy	1.83	0.19	9.66	0.00
B_NUMCHLD	Buy	0.139	0.0599	2.32	0.02
B_INTRT_BUY	Buy	-0.0758	0.0363	-2.09	0.04
B_UNEMPRT_BUY	Buy	-0.171	0.0703	-2.43	0.02
B_ZERO_VEH	Buy	1.48	0.284	5.2	0.00
B_CVINCVAL	Buy, Replace	0.151	0.0369	4.09	0.00
B_JBINC_REPL	Replace	0.292	0.181	1.62	0.11
B_INTRT_REPL	Replace	-0.0387	0.0339	-1.14	0.25
B_UNEMPT_REPL	Replace	-0.0761	0.0646	-1.18	0.24
B_AGEAVG	Replace	0.0853	0.0154	5.55	0.00
B_DINCVAL	Replace, Dispose	0.00224	0.00511	0.44	0.66
B_NUM_VEH	Do Nothing	-0.487	0.0763	-6.38	0.00
B_OLD_NN	Do Nothing	0.31	0.174	1.78	0.08
B_LNYRSTRAN	Do Nothing	-0.263	0.0856	-3.08	0.00
B_DISP_DELTA_NUM_CONF	Dispose	-0.0525	0.169	-0.31	0.76
B_DISP_DELTA_HH_TRAV_UTIL	Dispose	0.0765	0.17	0.45	0.65
B_DISP_DELTA_HH_PS_UTIL	Dispose	0.332	0.325	1.02	0.31
B_HHDEC	Dispose	1.07	0.547	1.95	0.05
B_JBDEC	Dispose	0.889	0.39	2.28	0.02
B_DLLTFS	Dispose	1.21	0.33	3.66	0.00
B_CHDEC	Dispose	-0.783	0.736	-1.06	0.29

Initial Log-likelihood: -5105.029
Final Log-likelihood: -2522.544
ρ^2: 0.506
$\bar{\rho}^2$: 0.500

The results from model estimation suggest that the transaction choice model has a strong overall fit, based on its $\bar{\rho}^2$ value. Furthermore, all parameters (except one) are of the expected sign and of reasonable magnitude, and most are statistically significant at a 95% confidence level. Examples of the types of behaviour represented in the model include:

- The sole parameter that does not have the expected sign is B_DISP_DELTA_NUM_CONF, which is positive instead of negative. It was anticipated that this sign would be negative because disposing a vehicle would generally increase household sharing conflicts, and in order for this to be "bad", a negative parameter would be required in order to decrease the utility of the Dispose alternative. However, as was touched upon in Subsection 5.4.1, the conflict parameters will have opposite effects depending on how many vehicles are in the house. In this case, what may be happening is that one-vehicle households that dispose of their vehicles are ultimately reducing their vehicle sharing conflicts (since there is no vehicle to fight for) which is a "good" thing. If the influence of these one-vehicle households outweighs that of the multi-vehicle households that share vehicles (wherein reducing vehicle ownership would increase conflicts) then a positive parameter sign makes sense. Thus, although a positive parameter sign was unexpected, it is not unreasonable, and was therefore retained in the final model.

- In the Buy alternative, the partial household travel utility recovery created by the passenger mode does appear to be statistically different from the utility drop experienced by the number of car constraints. The opposite phenomenon (in terms of both signs and relative magnitude) is observed for the Dispose alternative's corresponding parameters, but they are insufficiently statistically significant to make any strong conclusions.

- In any case, the net travel utility gained by Buying a vehicle was found to have a strong effect on transaction decisions, but the opposite effect does not occur for Disposing a vehicle. This suggests that having separate parameters for the Buy and Dispose alternatives provides better results than having a single parameter for both. It suggests once households get accustomed to having an extra vehicle they do not want to part with it, even if they don't really need it very much. This correlates well with TACOS survey responses, where *"because we needed one"* was the most commonly cited motivation for buying an additional vehicle, whereas *"we didn't really need it/we had too many"* was almost never cited for disposing of one.

- The importance of the logsum value of expected utility to be gained from purchasing a particular class and vintage of vehicle, as represented by the B_CVINCVAL parameter is significant, but of moderate impact. This suggests that the benefits and drawbacks brought about by a particular type of vehicle do play a role in deciding whether or not to Buy an additional vehicle, but not necessarily a major one. Note that the logsum values of class and vintage choice were largely negative, but generally of low magnitude. The exception to this is in households where purchasing an additional vehicle would result in transportation-related expenses consuming an excessive amount of their overall household budget. In this case, individual vehicle utilities are quickly dominated by the various ownership and operation-related cost parameters, and the resulting logsum is large in magnitude but negative. Therefore, the B_CVINCVAL parameter is effective at helping "prevent" households without enough money from purchasing a vehicle, but otherwise suggests the decision to Buy or Replace a vehicle is largely independent of what type of vehicle is actually purchased.

- The importance of the logsum value of expected utility to be gained from disposing of a particular vehicle in the household fleet, as represented by the B_DINCVAL parameter is the correct sign, but insignificant in both magnitude and statistical significance. Although this suggests that the vehicle being disposed has very little to do with the choice to dispose a vehicle, in practice this is not entirely true. Rather, Dispose and Replace decisions are often triggered by a specific repair/maintenance cost event, and which cannot be directly represented in the disposal choice logsum value. Thus although in the "real world" the utility of disposing of a particular vehicle will be extremely high, which should be reflect in the logsum, the model cannot directly represented this, and hence is the parameter is essentially irrelevant.

- Households are more likely to buy a vehicle if an additional person is employed than they are to dispose of one if there are less people employed, which again demonstrates how households quickly get used to owning a certain number of vehicles, and are hesitant to give them up in the face of falling income, as previously noted by Dargay (2001).

- The relative numbers of vehicles and drivers licenses in the households indicate the desire to own one's own vehicle can affect transaction choices.

- Results show that households are hesitant to make vehicle purchases in times of either high unemployment or high interest rates. Furthermore, the alternative-specific parameters are higher for Buy than Replace. This suggests that despite their hesitancy, households may acquiesce to a new purchase if they must replace a vehicle which is worn out altogether, and still need the same number of vehicles.

- The presence (or lack thereof) of children in the household plays a significant role in transaction choices.

One variable that Mohammadian used that is not included here is a smoothing parameter on certain transaction choices. A similar set of parameters were developed for this transaction model, but were not found to be statistically significant. The reasons for this are not clear, but may be related to the fact that other variables which have been included in the model indirectly represent something similar to what the smoothing factors did previously.

Given the strengths of the model in terms of both its statistical fits as well as its intuitively obvious findings, it is considered appropriate for use in ILUTE.

With the disposal choice, class and vintage choice and transaction choice models having been estimated, it is possible to model how a household's vehicle fleet will change over time (provided input data is available to feed the model), if the vehicle fleet is already known for the initial year being modelled. A vehicle fleet initialization model was not developed either as part of the population synthesis procedure developed by Pritchard and summarized in Section 3.4.1, nor by Mohammadian as part of the work described in Section 3.3. Such a model is necessary; otherwise, households would all start with no vehicles when they first enter into the model, both at the "start" of ILUTE in 1986 as well as all subsequent new immigrating households. This would cause ongoing inaccuracies in ILUTE's ability to model both travel and land use patterns. Thus, such a model must be developed, and Chapter 7 will detail this process.

Chapter 7

Development of a Vehicle Fleet Initialization Model

Chapters 3 to 6 have dealt exclusively with the implementation of a vehicle transaction model for ILUTE. Transaction models, by definition, change vehicle ownership levels *relative to the current ownership level.* As discussed in Section 2.3, this is considered to be an approach that is more reflective of a real-life decision-making. However, it also requires an initial vehicle fleet starting point for each household upon which all future household ownership decisions will be based. Without this initial fleet vehicle, the simulation would begin with all households in the region starting with no vehicles whatsoever in their first year in the model. This applies to both households synthesized as part of the 1986 ILUTE Base Year (since the population synthesis procedure does not include vehicle ownership, as noted in Subsection 3.4.1), as well as all households formed during the 20-year simulation period (e.g. resulting from immigration). Such a situation would be problematic on several levels:

- Since households can add a maximum of one vehicle per year, it would take several years for vehicle usage to climb to accurate levels. In the intervening time, travel patterns would be extremely unrealistic for each household. On an aggregate level, this means that region-wide motorization will always be lower than it should be, since new households will constantly be "catching up" to the level of ownership that they "should" have.
- Representations of fleet composition in terms of class and model year would show vehicles as being purchased in later years than they would otherwise have been. As external macro-economic factors as well as year-sensitive vehicle properties such as fuel efficiency will influence class and vintage

choice, this will result in inaccurate properties for the overall GTHA vehicle fleet.

- Because of the inter-dependent nature of the land use, vehicle ownership and daily transportation choices within ILUTE, there is a risk that systematic errors in the initial years with regards to vehicle fleet will propagate errors in both the transportation and land-use components of the model, which will in turn affect vehicle fleet choice, and so on.

As such, a robust vehicle fleet initialization is required to provide a solid foundation upon which the transaction model can be run. Regrettably, there is no known available source of data for 1986 that contains all of the information required to develop a robust initialization model. As a result, the TACOS dataset, despite running from 1990-1998, was used to develop such a model. This implicitly assumes that with the use of 1986-specific exogenous inputs, the model is otherwise temporally transferable. This is essentially the same assumption being made for the transaction model as a whole that is intended to run from at least 1987-2006, and possibly beyond once ILUTE reaches a stage where it can begin having its forecasting accuracy validated.

The structure of the initialization model is relatively straightforward; it consists of a two-step process wherein the number of vehicles in the household is determined, and then a class and vintage is selected for each vehicle. In the case of multi-vehicle households, information on the class and vintage selection of their first vehicle can be used to inform the class and vintage selection process of subsequent vehicles. Figure 7.1 shows the basic algorithm structure for this procedure.

The remainder of Chapter 7 details the development of each of the steps in the initialization model. Specifically:

- Section 7.1 justifies the vehicle ownership level model that was selected, based on the datasets considered for its development.
- Section 7.2 then discusses the class and vintage selection procedure.
- Finally, Section 7.3 discusses potential means of validating the model.

7.1 Vehicle Ownership-Level Model

Three steps were involved in development of the ownership level model. The first was to select the preferred model structure to be used. The second was to determine the best data source available for use in estimating the model. Finally, the third step was to specify the model variables and actually estimate the model.

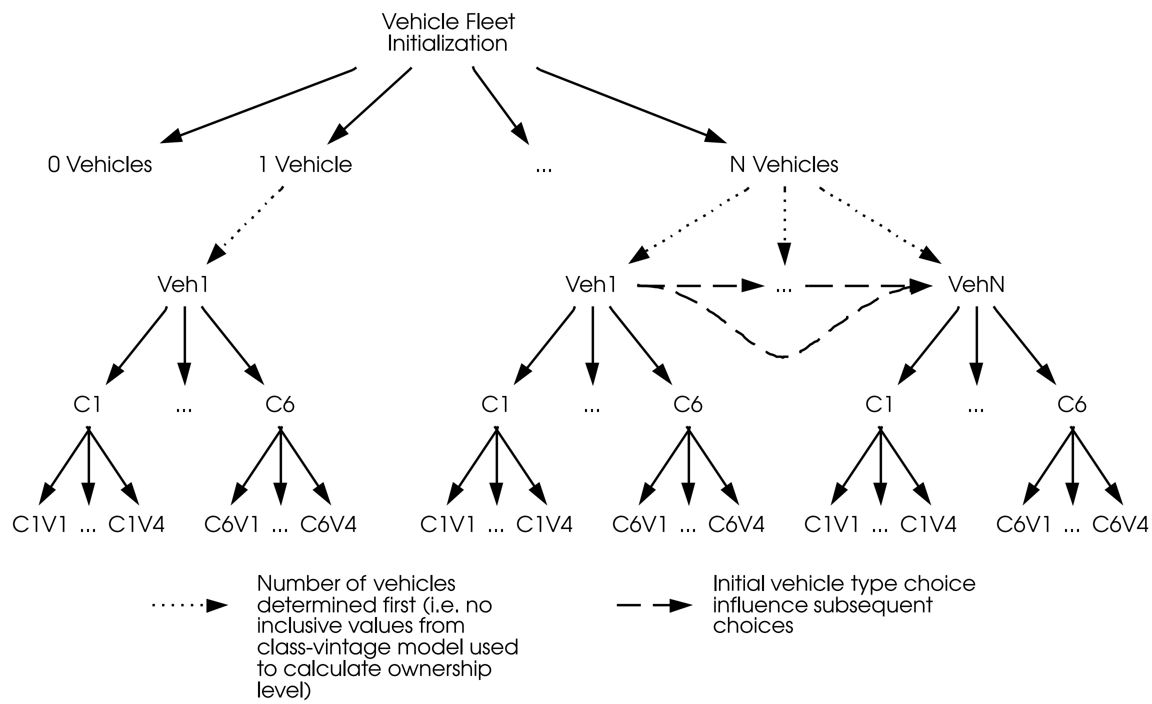

Figure 7.1: Vehicle Initialization Model

7.1.1 Model Structure

Section 2.2 provides an extensive review of vehicle ownership models that have been previously developed. Based on the general consensus from existing literature on the subject, and the paper by Bhat & Pulugurta (1998) in particular, a decision was made to employ the multinomial logit Model (MNL) for this application. The MNL is believed to generally offer better performance than alternative model structures such as the ordered logit as a result of its ability to allow for alternative specific parameters.

The choice set for the vehicle ownership model was specified to range from between zero to four vehicles. Observed vehicle ownership in the TACOS and TTS databases reaches as high as seven and several dozen vehicles, respectively, although these levels are very uncommon. It is unlikely that there are many households for whom more than four vehicles are actually required for transportation purposes on a daily basis and thus the vehicles are being purchased for these reasons. Beyond the four-vehicle level, it is assumed that ownership is more so motivated by individuals collecting vehicles as a hobby. Unsurprisingly, the TACOS data indicates that the high ownership level households are strongly correlated with the highest income category. In these very-high ownership level households, the motivations and decision-making processes that individuals use to make these purchase decisions are likely to be substantially different than those purchasing vehicles as a means of transport, and thus it would be challenging to model those decisions at all. Furthermore, the primary purpose of the vehicle fleet model is to improve the accuracy of the interaction of transportation and land-use patterns in ILUTE, and vehicles sitting in a garage as part of a collection are not particularly relevant to this aim. Figure 7.2 shows the proposed model structure.

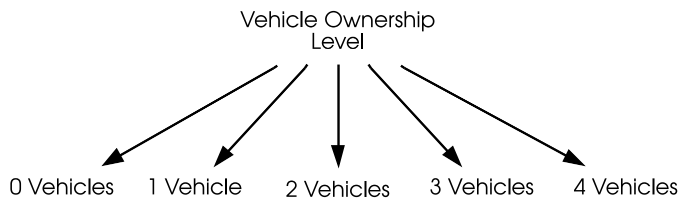

Figure 7.2: Ownership Level Model

7.1.2 Data Source

There is no single obvious source of data with which to develop the ownership level model. Instead, there are several sources of data featuring a number of pros and cons that were carefully considered before selecting a preferred data source. The three sources of data considered were:

- The Statistics Canada 1986 Household Income, Facilities and Equipment survey (HIFE).
- The 1986 Transportation Tomorrow Survey (TTS) collected by the Data Management Group at the University of Toronto.
- The 1990-1998 TACOS dataset used for the vehicle transaction model.

Note that 1986 data is selected for the HIFE and TTS surveys because it is the ILUTE Base Year, and will thus be the single year that has far more households requiring fleet initialization versus all other years. However, this same model will still be able to function for the entire ILUTE simulation period. The benefits and drawbacks of each data source are assessed in Table 7.1.

Table 7.1: Comparison of Potential Vehicle Ownership Level Data Sources

Data Source	1986 HIFE	1986 TTS	TACOS
Advantages	• Data collected in most important year requiring initialization (1986). • Very detailed demographic information. • Data definitions are very consistent with population synthesis procedure, which also uses Statistics Canada datasets.	• Data collected in most important year requiring initialization (1986). • Allows for an accurate geographic overlap with that are being modelled in ILUTE.	• Dataset and definitions are consistent with the transaction model. • Multi-year dataset may help reduce temporal transferability error. This will assist with accurately modelling initial vehicle ownership for in-migrating households or individuals over longer time frames than just 1986. • The same dataset can be used to develop a class and vintage model (set Section 7.2 below). • Allows for an accurate geographic overlap with the area being modelled in ILUTE.

continued on next page...

Table 7.1: Comparison of Potential Vehicle Ownership Level Data Sources (continued)

Data Source	**1986 HIFE**	**1986 TTS**	**TACOS**
Disadvantages	• Geographic area op data collection does not match with ILUTE modelling area (it can only be constrained to Ontario urban centres with populations of over 500,000 people). • Highly geographically aggregate data does dot allow location-specific dummy variables to act as proxy variables for auto-dependency in travel patterns. Location-specific dummies can be useful in the absence of actual trip information.	• Does not contain income information, which is a key variable in ownership level. • Using TTS as a primary estimation dataset removes its ability to be a completely independent data source for validating the Base Year simulation output.	• Data not collected in lost important year requiring initialization (1986).

Ultimately the TACOS dataset was selected as being the preferred source for model implementation. The TTS was the second strongest candidate, but it was recognized that it would also provide a very valuable source of entirely independent information with which to validate the model outputs; something that could not be accomplished if it was used in the model estimation.

As TACOS is a retrospective survey, it contains information on households for as many as nine separate years (1990-1998). All households were included for all the years that they were recorded in the survey, and thus an individual household may appear as many as nine separate times. Factors such as the number of workers or dwelling type are changing over the course of the survey period, and vehicle ownership levels should be responsive to this. All household changes (e.g. moving, increase in jobs or household size, changes in vehicle ownership level) are assumed to occur at the very start of the year. Or in other words, the data for a particular year is made after all changes for that particular year have occurred. Thus, if over the course of a given year a household moves, gains an additional member and buys a car, the data for that year is structured for the point in time after which all of these events have taken place.

7.1.3 Model Specification and Estimation

The third and final step in developing the ownership level model was to specify the individual utility functions of each choice alternative and estimate the model itself. Given that this model must run subsequent to the 1986 population synthesis procedure, but prior to any travel modelling, using travel utility information to aid in the estimation of ownership levels is not possible as this information does not yet exist. Thus, the model must rely solely on socio-economic and location-specific factors that are present in both the TACOS dataset and in the synthetic population developed by Pritchard (2008).

A number of vehicle ownership model papers were reviewed to develop a sense of the types of demographic attributes that would help inform the model. Types of variables used were relatively consistent between different types of models, and generally fell into several main categories:

1. Number of people in the house, categorized by age or employment status
2. Dwelling type
3. Provision of transit services
4. Income level
5. Development patterns (e.g. urban, suburban) and/or land use mix

Categories 1, 2 and 4 are all available in the TACOS dataset, and can be included in the model (although the reader is advised to recall the discussion related to income levels in Subsection 4.5.2). Categories 3 and 5 proved to be more challenging to incorporate.

Category 3, the provision of transit service, contained variables often relied on detailed household location information and proximity to transit information calculated with GIS software. This incurs the same problems with regards to spatially disaggregate data previously discussed in Section 6.4.

Category 5 contained a mix of specific measures, ranging from GIS based land-use mix calculations (often focusing on the mix of population and employment levels within a certain catchment from the residence) to dummy variables for locations broadly deemed to be “urban” or “suburban”. Similar to the GIS-related concerns for Category 3, creating a precise definition of various urban forms with which to categorize variables can be challenging and labour intensive for even a single year. Having to do this for a large number of years one at a time would be a very slow process since ILUTE will change the land use patterns each simulation year.

In the interests of examining whether some form of geographic dummy variable(s) was appropriate, the level of household vehicle ownership by each region within the GTHA was calculated using the 1986

TTS data. Figure 7.3 shows the results.

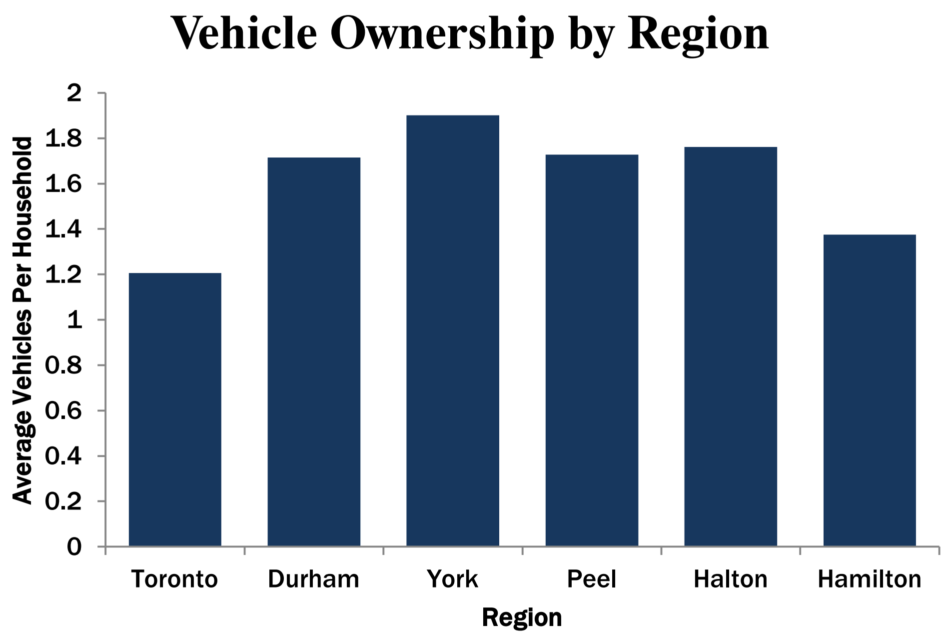

Figure 7.3: Vehicle Ownership Rates by Region

As can be seen, there is a substantial geographic variation in vehicle ownership levels throughout the GTHA, with the lowest levels being in Metropolitan Toronto (now the City of Toronto) and Hamilton, and the highest levels being found in the Region of York. Dummy variables were therefore created for Toronto, Hamilton and York. The regions of Durham, Peel and Halton all had relatively consistent ownership rates of around 1.7 vehicles per household, which acts as the "default" rate.

The use of geographic dummy variables is an acknowledged risk because it essentially acts as a proxy variable that may imply several different hidden behavioural patterns:

- The urban form of the areas of the region that predate widespread motorization (i.e. predominantly the central areas of Toronto and Hamilton), which are more conducive to walking and have higher levels of mixed use activity. If the suburban regions begin to urbanize, the urban form differences would minimize, and thus the magnitude of the parameters of the dummy variables would have to be re-estimated to reflect this.

- Transit service provided by the TTC in 1986 was substantially higher than that provided in many of the suburbs, which can have an effect on the number of vehicles households require. Note that even the proximity-to-transit variables used in Category 3 often do not fully represent the impact of transit service, as they typically represent the distance to *a* transit service from a house (generally the morning point of origin) as a measure of potential transit use, regardless of how useful/useless said transit service actually is in getting the members of the household to their actual destinations. If transit service becomes proportionately more useful in the suburbs (or is subject to major cutbacks in Toronto), then this could have an impact on the relative level of vehicle ownership. As with the urban form concerns, the model may have to eventually be re-estimated to reflect this.

- Household structures and may be different between the different regions. In particular, the suburban regions would be expected to have a higher proportion of households consisting of the traditional family structure, and fewer singles or childless couples who would generally tend to have lower vehicle ownership levels. This can be accounted for in the ownership level estimation through appropriate demographic variables (see category 1).

- A level of self-selection may exist in the form people who want a more suburban, car-oriented lifestyle deliberately moving to the parts of the GTHA that allow for this. In this case, if attitude discrepancies remain reasonably constant between the Regions, then the parameter estimates could arguably be reasonable over the long term if these individuals intend to hang on to their higher levels of automobile ownership over the long term regardless of changes to the transit service or urban form of certain parts of the city.

The use of geographic-dummy variables is clearly fraught with a number of risks related to temporal transferability of the resulting model. However, it was judged that the benefits of including them would hopefully outweigh the negatives. Furthermore, two separate "safety-checks" can be applied in the estimation:

- The TACOS dataset runs from 1990-1998, and is thus based on a nine-year spread of observations rather than a single-year estimations, which will hopefully minimize the impact of any single event or trend that was only present for a short period.

- The assigned vehicles for the 1986 synthetic population can be independently validated against the 1986 TTS dataset to obtain a sense of their overall level of accuracy. Note that inaccurate results do not necessarily imply a poor level of transferability between time periods, as the synthetic population itself could be subject to error. Validation efforts are discussed in detail in Section 7.3.

Finally, it should be noted that there were many instances of households relocating to dwellings in different parts of the region over the course of the survey. These movements were carefully tracked, and location specific dummy variables were only applied in the years in which the household was actually living in those particular locations.

Table 7.2 below summarizes the variables that are used to populate the utility functions for each ownership level.

Table 7.2: Initialization Model Ownership Level Variable Definitions

Parameter Name	Applies to	Meaning	Justification
B_NOVEH	0 Vehicles	Alternative Specific Constant for the choice alternative.	Standard component of all discrete choice models. Note that B_NOCAR will be set to 0 for estimation purposes.
B_ONEVEH	1 Vehicle	As above.	As above.
B_TWOVEH	2 Vehicles	As above.	As above.
B_THREEVEH	3 Vehicles	As above.	As above.
B_FOURVEH	4 Vehicles	As above.	As above.
B_WKPER_1	1 Vehicle	Alternative-specific parameters for the number of workers (both full-time and part-time) in the household.	Workers may need a vehicle for commuting purposes, and if all household members work similar hours, this figure may represent the minimum number of vehicles required based on who must commute by vehicle. Either positive or negative parameter signs are expected for the lower ownership levels, and positive ones for higher ownership levels.
B_WKPER_2	2 Vehicles	As above.	As above.
B_WKPER_3	3 Vehicles	As above.	As above.
B_WKPER_4	4 Vehicles	As above.	As above.
B_NOWKPER_2	2 Vehicles	Alternative-specific parameters for the number of workers (both full-time and part-time) non-workers.	Non-workers are less likely to make regular travel since they dont commute (unless they are students). Nonetheless, non-workers will still require some degree of mobility, which may include a vehicle, and the likelihood of this increases with increasing numbers of non-workers. Either positive or negative parameter signs are expected for the lower ownership levels, and positive ones for higher ownership levels.
B_NOWKPER_3	3 Vehicles	As above.	As above.
B_NOWKPER_4	4 Vehicles	As above.	As above.

continued on next page...

Table 7.2: Initialization Model Ownership Level Variable Definitions (continued)

Parameter Name	Applies to	Meaning	Justification
B_SFH_1	1 Vehicle	Alternative-specific parameters for a dummy variable representing whether the household lives in a Single Family Detached House.	In the very broad sense, single family detached houses can serve as a useful indicator of lower-density and vehicle-dependent neighborhoods. Here it is used as a rough proxy for vehicle-intensive development patterns. Either positive or negative parameter signs are expected for the lower ownership levels, and positive ones for higher ownership levels.
B_SFH_2	2 Vehicles	As above.	As above.
B_SFH_3	3 Vehicles	As above.	As above.
B_SFH_4	4 Vehicles	As above.	As above.
B_TO_1	1 Vehicle	Alternative-specific parameters for a dummy variable representing whether the household is located in Metropolitan Toronto (now City of Toronto).	Much of the inner areas of Metropolitan Toronto predate widespread motorization, and have an urban form more that is more amendable to facilitating a care-free or car-light lifestyle. Additionally, the level of transit service provided by the TTC in 1986 was vastly superior to that of the suburbs, which would on average reduce the need to own a vehicle for persons living in Toronto. Either positive or negative parameter signs are expected for the lower ownership levels, and negative ones for higher ownership levels.

continued on next page...

Table 7.2: Initialization Model Ownership Level Variable Definitions (continued)

Parameter Name	Applies to	Meaning	Justification
B_TO_2	2 Vehicles	Alternative-specific parameters for a dummy variable representing whether the household is located in Metropolitan Toronto (now City of Toronto).	Much of the inner areas of Metropolitan Toronto predate widespread motorization, and have an urban form more that is more amendable to facilitating a care-free or car-light lifestyle. Additionally, the level of transit service provided by the TTC in 1986 was vastly superior to that of the suburbs, which would on average reduce the need to own a vehicle for persons living in Toronto. Either positive or negative parameter signs are expected for the lower ownership levels, and negative ones for higher ownership levels.
B_TO_3	3 Vehicles	As above.	As above.
B_HAM_1	1 Vehicle	Alternative-specific parameters for a dummy variable representing whether the household is located in the City of Hamilton.	Similarly to Toronto, much of the inner areas of Hamilton have development patterns that can serve to reduce vehicle dependency compared to the regional average. Either positive or negative parameter signs are expected for the lower ownership levels, and negative ones for higher ownership levels.
B_HAM_2	2 Vehicles	As above.	As above.
B_YK_1	1 Vehicle	Alternative-specific parameters for a dummy variable representing whether the household is located in the Region of York.	Exploratory data analysis suggested that in 1986, the Region of York was exceptionally auto-dependent; even more so than other suburban regions such as Peel, Durham and Halton. As such a dummy variable was included for York Region households to help reflect this. Negative parameter signs are expected for the lower ownership levels, and positive ones for higher ownership levels.
B_YK_2	2 Vehicles	As above.	As above.

continued on next page...

Table 7.2: Initialization Model Ownership Level Variable Definitions (continued)

Parameter Name	Applies to	Meaning	Justification
B_YK_3	3 Vehicles	As above.	As above.
B_YK_4	4 Vehicles	As above.	As above.
B_LNINC_1	1 Vehicle	Alternative-specific parameters for a natural logarithm of household income level, expressed in 1000s of 1998 dollars + 1. Further information on the treatment of income can be found in Subsection 4.5.2.	Vehicles represent major life expenditures, and additional income provides more financial capacity with which to purchase them and enjoy the mobility benefits they may provide. Positive parameter signs are expected.
B_LNINC_2	2 Vehicles	As above.	As above.
B_LNINC_3	3 Vehicles	As above.	As above.
B_LNINC_4	4 Vehicles	As above.	As above.

Note that these variables represent the structure of the final model that was selected for implementation. All attributes initially had alternative-specific parameters for all four non-Zero Vehicle choice levels, but several of them (e.g. B_NOWKPER_1) were eliminated after initial modelling efforts found that they were of very low statistical significance.

After cleaning up the data to remove instances of missing information, a total of 5829 observations remained for use in the model estimation. Based on the final specification listed above, the model was run in BIOGEME. Table 7.3 summarizes the resultant model. Detailed output from BIOGEME can be found in Appendix D.

Finally it is worth noting that despite TACOS being selected as the preferred source of data, for comparative purposes a similarly-specified model was nonetheless also estimated from the 1986 Statistics Canada HIFE data The HIFE-based model was generally found to have an inferior fit and was subsequently discarded from consideration.

7.2 Vehicle Class and Vintage Assignment

Of the three sources of data considered for use in the ownership level model discussed above (HIFE, TTS and TACOS), only TACOS also includes information on the vehicles themselves. HIFE contains three very broad categories of class type ownership (Motorcycle, Automobiles and Vans/Trucks). As motorcycles have been discarded from the model, this means that only two categories of variables are present in HIFE while six exist in ILUTE; in light of this, HIFE is clearly not an appropriate source of

Table 7.3: Initialization Model Ownership Level Estimation Results

Parameter Name	**Applies to**	**Coefficient**	**Standard Error**	**t-statistic**	**p-value**
B_NOVEH	0 Vehicles	0	Fixed		
B_ONEVEH	1 Vehicle	-0.211	0.115	-1.84	0.07
B_TWOVEH	2 Vehicles	-2.85	0.164	-17.37	0.00
B_THREEVEH	3 Vehicles	-6.69	0.417	-16.04	0.00
B_FOURVEH	4 Vehicles	-13.4	2.17	-6.17	0.00
B_WKADLT_1	1 Vehicle	0.567	0.0441	12.86	0.00
B_WKADLT_2	2 Vehicles	1.29	0.056	23.11	0.00
B_WKADLT_3	3 Vehicles	1.75	0.101	17.24	0.00
B_WKADLT_4	4 Vehicles	2.17	0.201	10.79	0.00
B_NOWKADLT_2	2 Vehicles	0.324	0.0254	12.76	0.00
B_NOWKADLT_3	3 Vehicles	0.515	0.0729	7.06	0.00
B_NOWKADLT_4	4 Vehicles	1.1	0.152	7.21	0.00
B_SFH_1	1 Vehicle	0.728	0.0694	10.49	0.00
B_SFH_2	2 Vehicles	1.55	0.0877	17.69	0.00
B_SFH_3	3 Vehicles	1.69	0.187	9.01	0.00
B_SFH_4	4 Vehicles	0.621	0.423	1.47	0.14
B_TO_1	1 Vehicle	-0.713	0.0799	-8.93	0.00
B_TO_2	2 Vehicles	-1.49	0.096	-15.54	0.00
B_TO_3	3 Vehicles	-1.02	0.175	-5.86	0.00
B_HAM_1	1 Vehicle	-0.468	0.114	-4.11	0.00
B_HAM_2	2 Vehicles	-0.736	0.133	-5.52	0.00
B_YK_1	1 Vehicle	-0.845	0.163	-5.19	0.00
B_YK_2	2 Vehicles	-0.732	0.171	-4.29	0.00
B_YK_3	3 Vehicles	0.201	0.252	0.8	0.42
B_YK_4	4 Vehicles	1.02	0.533	1.92	0.06
B_LNINC_1	1 Vehicle	0.164	0.0268	6.12	0.00
B_LNINC_2	2 Vehicles	0.259	0.0356	7.26	0.00
B_LNINC_3	3 Vehicles	0.427	0.0921	4.64	0.00
B_LNINC_4	4 Vehicles	1.26	0.527	2.39	0.02
B_CHLD_3	3 Vehicles	-0.548	0.0979	-5.59	0.00
B_CHLD_4	4 Vehicles	-1.26	0.272	-4.62	0.00

Initial Log-likelihood: -10236.025
Final Log-likelihood: -7521.380
ρ^2: 0.373
$\bar{\rho}^2$: 0.370

data. The Transportation Tomorrow Survey only collects information on the number of vehicles in the household, and is therefore not useful for a class and vintage model.

In addition to the alternative of simply re-using the TACOS-based class and vintage model developed for the transaction choice model, a search was conducted to determine whether there were any other appropriate sources of data available for use.

The closest possible source of information that was found is the Statistics Canada Travel to Work surveys, which were run between 1976 and 1984 and which contain information on vehicle class and ages. However this data source was discarded on the following grounds:

- The survey runs from 1976 to 1984, and would have the same issues of temporal transferability between the years that data were collected and the 1986 ILUTE Base Year. Furthermore, as a class-vintage model still needs to be applied for all new persons/households in the GTHA over at least the 20 year simulation period, the model would get more out of date over time, while TACOS would get more up to date as the simulation years entered the 1990's.

- The survey focused more on the respondents' most recent commute to work, not on household vehicle ownership. As such, vehicle information was only collected for the vehicle used to make the trip, but not class and model year information for any other vehicles in the household (or even if there were other vehicles in the household). There may also be a bias in commute vehicles being disproportionately more likely to be a certain type of vehicle from within the household fleet (e.g. those that are more fuel efficient, to reduce costs).

- Further to the previous point, much of the additional information that was collected details the commute to work (such as trip distance) and not household demographics and socio-economic factors or the existing vehicle fleet. For example, the only demographic variable included is marital status. In contrast, the travel-based variables are not useful for running the model in ILUTE because travel is modelled subsequent to vehicle ownership, and thus cannot be used until the following year, which can make use of initial year travel patterns. Actual make and model information are not available either, so properties such as fuel efficiency and luggage capacity cannot be determined.

- The categories for both class and vintages vary from year to year, which makes assembling a consistent multi-year dataset challenging. For example, the class variable at times consolidates several classes together in some years that are disaggregated in other years. In one instance, vehicle weight is used instead of class categories altogether. Similarly, vehicle ages are often grouped into several years at once, which do not necessarily match the ILUTE/TACOS definition of vehicle vintage.

- The survey does not contain income information, which has been found to be a strong determinant in class and vintage choice, as per the previously estimated class-vintage model.

Given the problems associated with the only other available source of data, a decision was made to simply re-use the previously estimated class and vintage choice model given in Table 6.4. There are three caveats specific to using the same TACOS-based class-vintage model for each vehicle in the initialization model:

- The IsClassDisp variable (a dummy variable for whether the new candidate replacement vehicle is the same class at the one that is being disposed in replacement transactions) will always be zero for all options, since no vehicles are being replaced in the Base Year model (since there are none to replace). In this regard, it functions similarly to how the class-vintage model is used for the Buy alternative.

- Operating and Maintenance Fixed Costs in multi-vehicle households will not reflect the fact that for the first vehicle assigned the household will not have any other vehicles to influence its selection. In other words, the HHOMFixOvrHHInc value will be too low for the first vehicle(s) having their class and vintage assigned. Thus, there is a risk that the household will over-commit to an excessively expensive to operate and maintain first vehicle, because they do not "realize" that they still have a second vehicle that they must also maintain. To rectify this, it is suggested that second, third or fourth vehicles (depending on modelled ownership level) have an assumed "average" fixed O&M cost that will be used when the class and vintage of the first vehicle is being assigned. Once the first vehicle has been assigned, the "true" fixed O&M costs of this vehicle can be used in conjunction with the average fixed O&M costs of the third and fourth vehicles when candidate class and vintage combinations are being modelled for the second vehicle. This process then continues until all vehicles are assigned. It is suggested that the average fixed O&M cost simply be the mean value for all twenty-four class and vintage combinations in the model. This works out to be 4.177 (thousands) in 1998 dollars, but as cost-to-income ratios are used in the model, this value must be inflated or deflated as required to match the currency year of the household's income (both of which should ideally match the year being simulated).

- Operating and Maintenance Variable Costs, and therefore the HHOMVarOvrHHInc variable will suffer the same problem as the fixed O&M costs described above. Here it is recommended that a total household annual VKT of 20,000 km be assumed, as this is the same assumption used by Mohammadian in developing the original O&M cost models. Further, fuel intensity values be calculated based on the average of all vehicle classes over a period of fourteen model years, beginning the year after the year for which the simulation is running (since model years are usually released prior to that actual year) and running up to 12 years prior. In the case of the 1986 Base Year synthesis, this would mean the applicable model years would be 1974-1987, inclusive.

Figure 7.4 provides a graphical representation of how both the O&M Fixed and Variable costs can be accounted for in the class and vintage assignment process of the Base Year model.

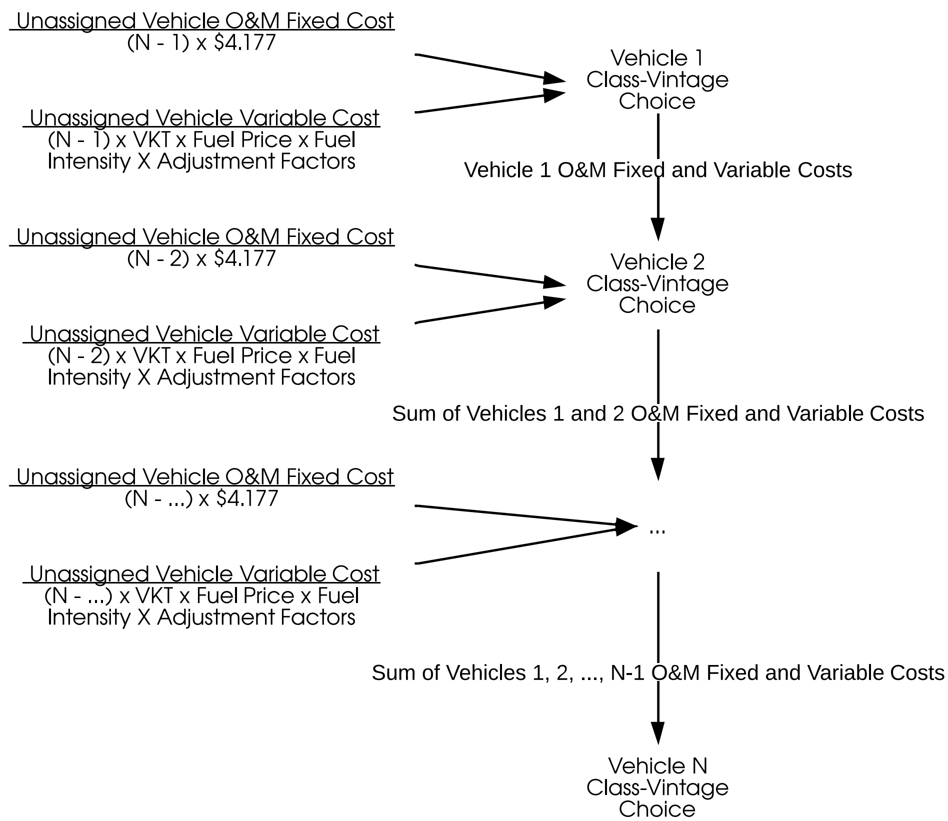

Figure 7.4: O&M Costs for Fleet Initialization

As touched on above, the class-vintage model simulates candidate vehicles as much as 12 years older than the year being simulated. Thus, for the 1986 Base Year currently employed in ILUTE, fuel intensity values by class will be required for as far back as 1974 in order to properly run the class-vintage model. To date, the TACOS dataset has been used to calculate average values for fuel intensity, however for model years prior to 1982, data becomes increasingly sparse (since vehicles would have already been 10 years old at the first survey year in 1990, there are relatively few of them in the database), and thus subject to a higher level of uncertainty in the fuel intensity averages. It is therefore suggested that an external source of data, such as the Red Book, be used to develop average fuel intensity figures by class and model year for between 1974 and 1982. This is the same process discussed previously as being required for all years between 1999 and 2007 in order to run the vehicle model for the full 1986-2006 benchmarking period.

Finally, note that Section 4.4 has already detailed the proposed mechanism by which vehicle use within the household and upon household formation/dispersal is to be conducted.

7.3 Model Validation for ILUTE Base Year

As was mentioned in Section 7.1, model validation using an independent dataset is an important step in ensuring the overall accuracy of the model. At the present time, the vehicle initialization model, similar to the rest of the vehicle transaction model is not yet able to be estimated.

The vehicle ownership level model could be relatively easily validated for the base year of 1986 once the labour force component of ILUTE has been also validated. The ownership level model is sensitive to both employment levels and income, and if these numbers are inaccurate, the corresponding vehicle ownership output will also be inaccurate and not compare favourably against the Transportation Tomorrow Survey data.

Regrettably, the class and vintage model cannot be validated without a large amount of further work and data collection; these requirements are discussed in Section 8.1.

Chapter 8

Future Research

The work undertaken in this thesis has both enabled household vehicle holdings to be modelled in ILUTE, and also improved the realism and policy sensitivity of the model used to generate this information. Despite these improvements, there are several avenues available for further improvement to the model. These range from the validation of the model developed in this thesis, additional exogenous data collection, implementation of additional surveys to assess temporal transferability of model findings, alternations of the model structure to incorporate additional choices, and changes to how the model is used in ILUTE. Each of the suggested improvements is discussed below.

8.1 ILUTE Integration and Validation of Overall Vehicle Fleet Model

At the beginning of the research undertaken in this thesis, a focus was made on programming the original Mohammadian transaction model into ILUTE. However, this effort was abandoned after several issues that were discussed (and resolved) in Chapter 4 became apparent. The revised model developed in this thesis has yet to be programmed into ILUTE, and thus this forms the first priority for implementation.

Some data collection effort for exogenous variables is also required. Specifically, the current fuel intensity dataset (as shown in Table 4.7 in Section 4.3) only contains data for the years from 1978 to 1999. Further collection of data is therefore required for all classes of vehicles in the following years:

- 1974 to 1978, since running the Base Year model for 1986 households will allow them to purchase vehicles up to 12 years old (i.e. 1974 models). Preferably, additional data would also be collected for 1979 through 1982, as the averages calculated from TACOS fleets for these years are slightly sparse, and could benefit from a larger sample size.

- 1999 to 2007, so that the 20-year validation runs can be tested. Note that 2007 data is required as the virtual vehicle dealership makes vehicles available for one year ahead of the model year
- Any model years beyond 2007 depending on the desired forecasting horizon.

Additionally, a work choice location model is required for ILUTE, to create the work trip destination zones required for TASHA to simulate household travel patterns. This model is currently under development, and should be completed in the near future.

Finally, making the vehicle transaction model sensitive to household travel patterns requires that household travel then be modelled each year that ILUTE is being simulated. Although TASHA is capable of generating demand, a network model of transport supply is required in order to finish assigning trips. EMME networks of the GTHA are generally created in five year increments that match the TTS survey collection years, and are known to exist for as far back as 1986. Although the older networks will likely require some level of "dusting off", it should be possible to make use of them for this application, so that each model can represent both its designated year as well as the two years prior and subsequent should be adequate.

Once all of the above work is completed, it will be possible to run the vehicle fleet model in ILUTE, and begin the validation process. As a result if ILUTE's highly integrated nature, validation of the vehicle fleet cannot easily be undertaken in isolation, but must be considered in the context of the validation of the spatial distribution of the demographic model, household and work location choice models and the travel model.

Finding appropriate data with which to validate the vehicle fleet simulations in ILUTE may also prove challenging. In terms of vehicle ownership level, the TTS provides a very valuable data set, and is recommended as the primary resource for validation, as it provides information on not only then number of vehicles in the household, but also many of the demographic and travel-pattern variables that are used to predict transaction choices.

Assessing the accuracy of the class and vintage choices will prove more challenging. As discussed in Section 7.2, there does not appear to be an obvious source of data with which to compare. This is particularly concerning due to the challenges associated with modelling class and vintage choice that were discussed in Section 6.4.

Arguably the only source of disaggregate data that may be available to validate the class and vintage assignments of the 1986 GTHA vehicle fleet is Ministry of Transportation of Ontario vehicle registration data. However, whether such information from as far back as 1986 still actually exists, as well as what format it would be compiled in is unknown. Furthermore, even if it does exist, there are likely to be a number of institutional barriers related to personal privacy that could make obtaining said information challenging.

Finally, the rule-based primary-driver model should also be validated to determine its level of predictive accuracy, and whether a more detailed model should be developed similar to several of those reviewed in the literature.

8.2 Potential Small-scale Refinements

Depending on the findings of the initial validation process with regard to how well the vehicle transaction model performs in a simulation environment, there are several potential easy-to-implement changes that would improve accuracy of exogenous inputs. This could potentially correct some errors in the model without having to re-estimate it altogether. These refinements include:

- Additional data collection for the class-specific properties generated by the virtual vehicle dealership (wheelbase, luggage capacity, engine displacement, weight), to see if these are changing over time, similar to how fuel intensity values are assumed to change over time.

- A review of the accuracy of the hedonic price model for predicting vehicle purchase prices far into the future. The estimation of the model is such that in each year, vehicle prices are increased by \$514 (in \$1998) above what they were the year before; reflecting the fact that vehicles are becoming more expensive over time, even once inflation is accounted for. While this model has a strong fit with the dataset it was regressed against (achieving an adjusted-R^2 value of 0.82), it would be advisable to investigate this price escalation pattern over a 20+ year dataset of vehicle prices to see if it still holds true, as it would imply that vehicles would be over \$10,000 more expensive in 1998 dollar terms at the end of that time frame than they were at the start. If the rate of price increases has diminished since the period modelled in the data collection, than this could lead to the model under-predicting vehicle holdings as the model would be assuming that vehicle purchases would be more expensive than they actually are. If the model appears to be over-predicting purchase costs for longer-term applications, then it could be revised if necessary.

- More detailed representation of person-specific insurance rates and how they vary based on the characteristics of which person within the household the vehicle is registered to, how much it is used etc.

- More detailed representation of insurance and maintenance costs that vary based on actual vehicle use.

8.3 Linking Vehicle Transactions to Place of Residence and Job Choices

The place of residence, place of employment and vehicle fleet decision models have all been developed as separate models within ILUTE, rather than a single "master model" where all these decision processes occur simultaneously. Given that these processes are nonetheless inter-dependant, a framework must be created where each of them take into account the effects it will have on the other.

As a first step, a decision had to be made with regards to the order in which the various choice processes were simulated in ILUTE. It was decided that the vehicle transaction model should be the last major process within the "demographics and land use" side of ILUTE, in the sense that it will come after the demographic updating, residential location choice and employment location choice processes, but prior to the TASHA-based travel modelling for that particular year. The output from TASHA is then used to inform the "land use and vehicles" based choice processes for the following year.

This order of operation decision is based on the assumption that residential and employment location choice will usually be higher on the decision-making hierarchy for most families, and as such, they would purchase the number of vehicles they feel are required for their locations of residence and employment, rather than relocate to an area that matches their current vehicle holdings.

Although this procedure is probably the best available in the current ILUTE framework, it does ignore the fact that households may make place of residence and job location choices that are based in part on how it would affect their vehicle fleet requirements. Literature suggests that there is in fact a discernible location-choice-versus-vehicle-fleet-requirements decision-making process which is not captured by the currently intended simulation order of ILUTE (Eluru *et al.*, 2010). There is potential to improve on this order of operations over the long term to provide a more detailed consideration of the trade-offs between household location, job location, vehicle fleet requirements and commuting, although it would add complexity to both model location-choice specifications as well as increase processing time for simulations.

Briefly, the current residential location choice model within ILUTE develops a list of ten candidate residences for households to rank from when determining whether they wish to move. The household then ranks these choices, although they are not guaranteed to get their top pick, as a market clearing process is run to assign various houses to the various interested parties in a manner that seeks to imitate the real-world real estate market. In order to make place of residence location choice versus vehicle ownership requirements more explicit, household travel patterns and therefore vehicle fleet requirements and costs can be "mentally simulated" by the household for each potential residence, and therefore used to inform preferred housing choices. A similar process could be used for job location choices. While this could potentially produce better and more policy-sensitive results, it would also introduce new computing requirements, as each household that becomes engaged in the market would have to have travel patterns and vehicle ownership simulated for ten different housing locations. Nonetheless, it is likely a worthwhile long-term improvement to ILUTE to help more explicitly tie together its various component models.

8.4 Utility of Activity Participation Opportunities in Transaction Choices

In economics terminology, travel is a derived demand. In other words, it's not something we actively want do to (with some exceptions, such as recreational trips), but rather something we are willing to put up with because the reward for making the trip is the ability to participate in a given activity (e.g. work, education, shopping, socializing etc.). Trips are then made when the utility experienced from participating in the activity is larger than the disutility associated with making the trip, or more simply, when the pros outweigh the cons.

The current TASHA model generates activity schedules from TTS data in a probabilistic manner, and thus there is no explicit representation of the utility gained by participating in these activities. On the other hand, the drawbacks of participating in them (i.e. the travel disutility) are explicit represented.

In the case of work and education activities, their locations are intended to be exogenously input into the model. Thus, the utility brought about by participating in these activities in their particular locations can be explicitly represented in ILUTE as part of the decision to reside/work/attend school that that particular location and the effect of vehicle fleet requirements can be accounted for in those decisions, per what was just discussed in Section 8.3. However, for other trips such as shopping that use a gravity-based model location choice, the model only explicitly incorporates the disutility of travel, but not the benefits of participating in the activity at the particular location. This can in turn create misleading results in terms of how they may influence the decision to buy a vehicle. For example, consider the following response to a household with no vehicles buying a vehicle:

"Now that I have a car, I can now get to my usual shopping mall in a much more efficient manner. If I keep shopping at the same mall, I will now have a 20 minute car trip to the mall, instead of the 40 minute transit trip I used to put up with. On the other hand, now that I have a car, there are additional shopping opportunities that I am able to access. There is another mall that will take me 40 minutes to drive to, but it has better stores than the one I currently shop at. I value the ability to access this other mall even more than I value the time savings that I would experience if I stayed at my current mall, and thus I will now shop at this new one".

Clearly, the greater benefits to the household in terms of lessening their travel disutility would come from simply using their vehicle to access their existing shopping centre, and that these benefits are generally lost when the household decided to relocate their shopping activities. However, far from meaning the vehicle didn't help the household because it did not provide any travel benefits, it means it was even more beneficial than those travel time savings, because it allowed the household to achieve a more optimal participation in shopping activities which they value even more than a travel time decrease.

In the context of the current transaction choice model discussed in Section 6.5, the change in travel disutility associated with a change in the number of vehicles a household possesses influences decision-making. However, the benefits those vehicles provide in terms of ability to access activities is not. This same issue was previously noted by Roorda *et al.* (2009). Thus, in the scenario described above, the model would assume that the vehicle is effectively "useless" because it did not provide any travel improvements, even though in fact it is more useful than those travel improvements would be. That is, the benefits it does provide are not being properly quantified in the transaction choice utility functions.

Based on these considerations, there is a risk that the model will underestimate the benefits of owning a vehicle in suburban areas where vehicle ownership is very beneficial in providing households with the abilities to access activity opportunities. In the event that these activity-participation utilities are eventually explicitly quantified, it is recommended that the transaction choice model be re-estimated so that the utility functions can explicitly account for this.

8.5 Incorporation of Hybrid and Alternative-Fuel Vehicles

A potentially worthwhile scope extension of the vehicle fleet model would be to explicitly represent the presence of hybrid and alternative-fuel vehicles (HAFVs). HAFV's, broadly defined as anything other than conventional gasoline or diesel fueled or hybrid vehicles, could be represented as either a generic designation, or further broken down into hybrid electric vehicles, hydrogen vehicles etc.

Representation of HAFVs in ILUTE could be useful for representing the impacts of congestion on fuel consumption, and in turn being able to produce more accurate emissions analysis. It could also be used to understand how cheaper O&M Variable Costs (i.e. if electric vehicles are much cheaper to re-charge than petroleum-based vehicles are to re-fuel) could have unintended consequences in terms of increasing vehicle use and propagating vehicle-dependant development patterns; both of which would go against the grain of many of the environmental and economic goals that the introduction of HAFVs could otherwise accomplish.

Modelling consumer willingness to purchase non-conventionally fueled vehicles predates the commercial sales of these vehicles by several decades. The earliest studies were conducted in the early 1980's, such as those by Train (1980b), Beggs & Cardell (1980), and Beggs *et al.* (1981). The 1990's also saw a number of fuel choice models developed, particularly in California, including those by Bunch *et al.* (1993), Brownstone *et al.* (1996), Golob *et al.* (1996), and Brownstone *et al.* (2000). Several models have also been developed subsequent to the actual commercialization of hybrid vehicles in the early 2000's, including those by Adler (2003), Potoglou & Kanaroglou (2007), Hess *et al.* (2011) and Vyas *et al.* (2012). All of these models assessed decision-making for vehicle fuel-type choice, and had varying degrees of integration with other vehicle-related decision making (e.g. class of vehicle, transaction choices etc.).

Representation of HAFVs in the revised transaction model was invested through the possibility of incorporating the CIBER-CARS survey conducted by Potoglou & Kanaroglou (2007) of the McMaster University Centre for Spatial Analysis. CIBER-CARS is a similar survey to TACOS that collected revealed vehicle fleet choices for households in the Hamilton area. However, households that indicated that they were considering purchasing a new vehicle in the near future were then given a stated preference survey that assessed how difference incentives could influence the choice to purchase a HAFV.

Despite its similarities to TACOS, the collected data was found to be sufficiently different that a jointly-estimated model was deemed to not be feasible as part of this thesis, and was therefore not included. Nonetheless, this is thought to be a worthwhile future endeavor, and in the event that another TACOS-style survey is undertaken (as will be discussed in Section 8.6 below), information on HAFV ownership could be collected, and a supplemental stated preference survey could also be administered to respondents.

8.6 Conduct of "TACOS II" to Assess Temporal Stability of Behaviour

At the time of this research, the original 1998 TACOS dataset was already 15 years old, with the oldest transaction decisions dating back 23 years to 1990. While this is a valuable historic record of transactions and useful for much of ILUTE's current 20-year validation period, it risks becoming increasingly out of date over time, with the final year of simulation (2006) being eight years removed from the most recent set of transactions recorded in TACOS (1998). The eventual use of ILUTE for forecasting purposes would only further exacerbate this concern.

The age of the survey is of particular concern given the existence of several major socio-economic trends that have taken place since the survey was undertaken, all of which are likely to have some influence of vehicle-related decision-making. While many of these issues can be accounted for as input attributes into the model, there is nonetheless an assumption implicit in the model that the behavioural response framework to changing realities will be consistent, even if the behaviour itself is not. To what degree this holds true in real life is unclear.

It is therefore suggested that a "TACOS II" be undertaken to provide an up-to-date dataset of vehicle ownership decisions that would reflect any of these trends, and can be used to assess the temporal transferability of the current model.

Several of the trends that may influence household vehicle fleet decision making are discussed below, including an examination of how the model accounts (and/or does not account) for them. This list of concerns is by no means meant to be exhaustive; rather it simply presents a number of important trends that have occurred since the collection of the original TACOS survey. A TACOS II survey should largely collect all of the same information of the original TACOS survey and cover the same geographic area, thereby allowing for apples-to-apples comparisons between the two datasets. In addition, it should feature the supplementary SP survey on alternative fuel vehicles discussed above.

8.6.1 Response to Rising Fuel Costs

Arguably the single largest long-term change to affect vehicle use since the TACOS data was collected is the rapid rise in gasoline prices. Prices peaked in the summer of 2008, and declined as a result of the late-2000's recession, but are still well above their late 1990's level, even accounting for inflation. Gasoline prices remained relatively stable over the course of the 1990-1998 timeframe that TACOS collected data for. However, as seen in Figure 8.1, which shows average self-serve gasoline prices for the GTA, fuel costs began to rise sharply thereafter.

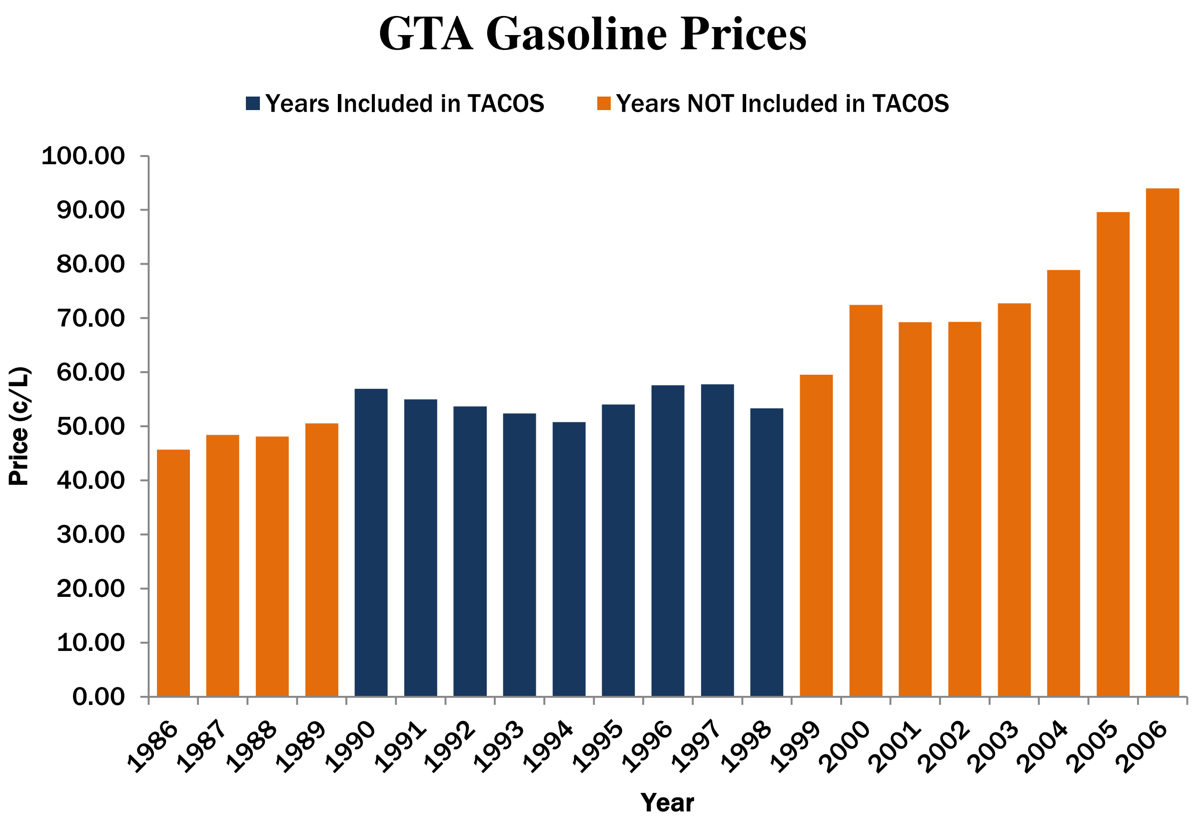

Figure 8.1: Average GTA Gasoline Prices by Year

Unsurprisingly, these rising costs resulted in declining sales of large "gas-guzzlers" such as SUVs, and a shift in the market towards smaller and more fuel efficient vehicles, as well as a general trend towards less driving.

As discussed in Section 6.4, the revised transaction model is able to explicitly incorporate the effect of fuel costs relative to overall household income, and thus how "affordable" a given type of vehicle is relative to how much it needs to be used. The model assumes that the total fuel expense brought about by rising fuel prices is what households take into account when making decisions, rather than the change in fuel prices. While the absolute fuel cost is what ultimately drives household budgets, it is perhaps worth investigation whether perception of fuel costs is more closely driven by changes in prices rather than absolute prices. With rapidly increasing prices, over the short term the household will feel budgetary "pain" as they try to maintain their current lifestyle and activity patterns in the face of it becoming increasingly unaffordable. However, the degree to which households will over the long-term restructure their activity patterns in light of rising fuel costs , and to what level they are elastic in terms of what other purchases they will forego in order to be able to keep owning and driving a vehicle is unclear.

Unfortunately, the effect of changes in fuel prices (either in ¢/L or as a percentage, relative to where it was one/two/three etc. years prior) could not be reliably tested using the TACOS data. The variation in fuel prices over the 1990-1998 timeframe covered in the data was so small that these parameters were found to be insignificant. However, if a new survey was collected that incorporated the 2000-2008 period, then given the rapid change in fuel costs over this time frame, a strong statistical significance may emerge.

8.6.2 Changing Environmental Attitudes

The rise in public concern relating to environmental conservation, particularly relating to anthropogenic climate change may also be causing a shift in societal preferences towards smaller vehicles, above and beyond the financial incentives provided by rising fuel costs. While fuel costs are likely the dominant motivation, it is worth noting that despite the cost of fuel decreasing substantially since its peak in the summer of 2008, vehicle purchasing patterns have still maintained some of the preference for fuel-efficiency. Whether this is due to environmental motivation or simply anticipation of future increases in fuel costs is unclear; likely it is a combination of both.

Unlike fuel costs, quantifying and simulating the effects environment-related attitudes on vehicle purchases would be exceptionally challenging. Although data on environmental preferences could be incorporated as part of a survey and incorporated into a model without significant difficulty, simulating the environmental attitudes of ILUTE agents and how they evolve over time is not possible. As such, the best realistic alternative is probably to simply collect up-to-date data, which would have these attitudes implicitly incorporated in alternative-specific preferences for both vehicle-type choice and transaction choices, and seek to understand how these "non-simulate-able" parameters may change over time.

8.6.3 Young Adults Staying At Home

According to the 2011 Census of Canada, 56.3% of Toronto CMA adults aged 20-29 live with their parents, having either returned after living independently or never left at all (Statistics Canada, 2012). In several of the outer-suburban municipalities (many of which also happen to be the fastest-growing areas in the region) this figure reaches upwards of 75%. Although this trend has generally stabilized since the 2006 Census, it nonetheless likely represents a change since the time period encompassed by TACOS. No Toronto CMA level data is available for pre-2006 years, but on a national level, it has risen from 32.1% in 1991 to 42.3% in 2011. Given that most of the reasons suggested by Statistics Canada for these increases (e.g. cost of housing, cultural preferences, pursuit of education) have arguably become more acute in the Toronto CMA than most other regions in the country over the past two decades, it is not unreasonable to assume that it has changed at least as much as the national average, if not more.

The effect of this “failure-to-launch” or “boomerang kids” trend on vehicle ownership is not entirely clear. On the one hand, it may be expected that adult children living with their parents will have access to their parents vehicles, and vehicle sharing within the household can potentially lower overall vehicle ownership rates than if adult children were living separately. On the other hand, the outer-suburban areas where these trends are most striking also happen to be the most auto-dependant areas of the GTHA, and these adult children may need require their own personal vehicle in order to access jobs and social opportunities, in which case vehicle ownership could become more affordable due to these young adults not needing to pay rent/mortgages on their housing.

8.6.4 The Social Status of the Automobile

There is no doubt that the importance of the automobile in North American culture extends vastly beyond its utility as a means of transportation. Owning one’s first vehicle has long been a symbol of passage into adulthood, and expensive vehicles are a conveyor of social status.

In recent years, there has been a significant amount of speculation in popular media that attitudes towards automobile ownership and the “American Dream” as a whole are changing, particularly among today’s young adults, the “Millennials”. Kalita & Whelan (2011), Weissman (2012) and Briggs (2013) are typical examples of this narrative.

One of the most frequently-cited pieces of evidence cited as evidence of this change in attitudes is a decline in driver’s licence possession rates among US youth. Data from the Transportation Tomorrow Surveys reveal that the same trend is occurring in the GTHA. Table 8.1 shows the percentage of GTHA residents between the ages of 16 and 29 with valid driver’s licenses for the start and end years of ILUTE.

Table 8.1: Youth Driver License Possession Rates by Region

% of Under-30 Persons with Driver’s Licenses			
Year	**1986**	**2006**	**Percentage Point Change**
Toronto	72.5%	64.6%	-7.9%
Durham	81.9%	75.0%	-6.9%
York	80.9%	78.0%	-2.8%
Peel	80.4%	73.8%	-6.6%
Halton	82.2%	80.5%	-1.7%
Hamilton	75.0%	70.4%	-4.6%
Total	75.7%	72.8%	-2.8%

On the other hand, a recent and somewhat more exhaustive study from the University of California Transportation Center suggests that changes in youth travel behaviour are predominantly the result of economic conditions (e.g. high student debt and poor job prospects), and thus that explanations suggesting changes in attitudes and life aspirations may be overstated (Blumenberg *et al.*, 2012).

Still, at the very least, perceptions of declining interest in vehicle ownership has concerned auto manufacturers enough to put a substantial amount of effort into understanding how they can better market their products to Millennials (Chozick, 2012).

Given the lack of clarity and consensus on this issue, a more up to date dataset would certainly prove useful in helping understand to what degree these attitudes exist above and beyond economic and/or environmental concerns. With a TACOS II survey, today's young adults can be compared with the young adults of the 1990's, to assess whether above and beyond the economic factors cited by Blumenberg *et al.* there is still a noticeable shift towards either smaller vehicle and/or less/no vehicles.

Similar to the changes in environmental attitudes discussed in Subsection 8.6.2 above, ILUTE cannot explicitly model how the "life goals" of agents change over time. However, at the very least, a "born after year X" dummy variable could be introduced if it was found to improve model predictive power. Care should be taken with such a change though, as it would implicitly assume that the next generation after Millennials would have many of the same attitudes as they do, which is entirely unclear at the moment without a better understanding of how inter-generational changes in attitudes are formed.

8.7 Effect of Vehicle-Share Services

In recent years, vehicle sharing services have become increasingly popular, and Toronto now features three such services; AutoShare, ZipCar and car2go. AutoShare, founded in 1998, was the first vehicle-share service set up in Toronto (Costain *et al.*, 2012). However, vehicle sharing did not experience significant growth until 2006, when ZipCar entered the Toronto market. At the time of ZipCar's arrival, AutoShare users and vehicle counts stood at 2,500 and 80, respectively. Over a six year period from 2006 to 2012, patronage rose substantially, with (between the two companies) at least 30,000 individuals, households or businesses registered with these services, making use of over 700 vehicles (Flavelle, 2012). Car2go also joined the market in 2012, although it offers a slightly different service model than the other two companies.

Given that the level of use of these services was still fairly minor at the end of ILUTE's current 20-year testing period of 1986-2006 (let alone the TACOS data collection period which ended in 1998), the influence of these services was not considered in the current model.

However, should vehicle sharing continue to grow in popularity at such a high rate, it may begin to have a non-negligible impact on vehicle ownership level for the areas that are covered by such services. AutoShare believes that each of their vehicles removes 8-10 private vehicles from the road (Dunn, 2006). This appears to be a reasonable figure, and is supported by data from other regions. In a North America wide survey of vehicle share users, Martin *et al.* (2010), found that each vehicle-share vehicle decreases overall vehicle ownership by 9 to 13 vehicles, in the form of households either actively disposing a vehicle or not buying a vehicle that they would have otherwise bought. The study further found that average private vehicle holdings of households that use car-sharing services reduced from 0.47 vehicles/household to 0.24 vehicles/household; a reduction largely caused by one-vehicle households being able to dispose of their single vehicle. Martin et al. also note that only 12% of households that owned one or more vehicles prior to joining a vehicle share didn't reduce their vehicle ownership level. Conversely, in a survey of City CarShare users in the San Francisco Bay Area, Cervero *et al.* (2007) found that the vehicle sharing service had not caused them to dispose of already-owned vehicles at a significantly different rate than a non-user control group, but that City CarShare users were less likely to increase their vehicle holdings.

In any case, the notion that vehicle share services have an effect on household vehicle holdings is clear. At the very least, the presence of a vehicle-share service will definitely result in some transactions shifting from "Buy" to "Do Nothing", and (depending on which source is to be believed) also increase the likelihood of "Do Nothing" choices becoming "Dispose" choices instead.

Incorporating vehicle share services into ILUTE could be a worthwhile long-term project to improve spatial validity of household vehicle holdings once the model is used for forecasting purposes. At the very least, it is suggested if an updated TACOS II survey is undertaken and has a sufficiently large sample frame, enquiring about interviewee membership in a vehicle sharing service could provide useful data to model membership and use of said services, and set-up feedback mechanisms between vehicle share membership, household vehicle holdings and travel mode choice.

8.8 Long-Term Trends in Automotive Technology and Ownership

The majority of the recommendations discussed above largely encompass concerns related to improving the accuracy and sensitivity of the model for historical runs, such as the 1986-2006 runs currently being undertaken or recent trends that have occurred between 2006 and the present day (2013). For long-term transportation modelling that is often required for infrastructure investment assessments, some level of "guesstimation" is required with regards to what choices people will be able to make, and how the framework that measured their behavioural responses to those changes may evolve over time.

In the context of vehicles, the underlying fundamentals of ownership and use that govern household behavioural choices have remained largely unchanged for at least the last 60-70 years. Obviously, aspects of vehicle technology (e.g. fuel efficiency improvements and new fuel sources, airbags, new classes of vehicles such as minivans and SUVs being developed etc.) have all changed substantially over this time. However, the model of each household purchasing and using its own vehicle, and needing parking spaces to store it whenever it is not in use has been constant.

Very recently, the vehicle-sharing services discussed in Section 8.7 have begun to detach the concepts of "owning a vehicle" and "ability to use a vehicle whenever one is needed" from each other. While this is a fundamental change, the business model of these services requires a base level of both population density and pedestrian friendliness that is unlikely to be found outside of central cities and possibly suburban town centres in the near or even medium term and thus is still not a realistic alternative to ownership for many GTHA households. However, the advent of driverless vehicles could stand to alter this greatly.

Having long been relegated to the realm of science-fiction literature, driverless vehicles appear likely to begin appearing in the marketplace in the near future. Most notably, the Google Driverless Car has been undergoing testing for several years and is intended to be put to market later this decade. Already, several American states have passed laws allowing autonomous vehicles to operate on their roads.

Certainly, driverless vehicles appear to be a boon for vehicle sharing services, as they remove the need for an urban form where there is a sufficiently large customer base within walking distance of each vehicle; essentially a service could be created that combines the convenience of a taxi with the lower costs of a vehicle share service. Combined with the convenience of being able to be dropped off at a front door, rather than a parking lot, this type of service could greatly increase in popularity and therefore have the effect of reducing vehicle holdings. On the other hand, they may also encourage further spatial dispersion of residences and workplaces and long trip distances by allowing for productive use of in-vehicle travel time, which could have an opposite effect.

A full assessment of potential impacts of driverless vehicles on household vehicle ownership is well beyond the scope of this thesis, and constitutes a thesis unto itself. The intent here is simply to note that it would be an advisable course of research and model development for any use of ILUTE that involves long-term forecasting.

Chapter 9

Conclusion

This thesis assessed how households make decisions regarding their vehicle fleet. Vehicle fleet decisions are inherently intertwined with both land-use and transportation patterns, and must be understood in the context of how it affects, and is in turn affected by, these patterns. Understanding these processes is important because collectively they have a significant effect on both public and private financial resiliency, economic accessibility and development opportunities as well as social equity, public health and environmental impacts.

Findings of this thesis are intended to be a component of the larger Integrated Land Use, Transportation and Environment modelling framework under development at the University of Toronto. ILUTE is an advanced next-generation large-scale urban simulation model that can provide more accurate, conceptually correct and policy-sensitive representations of urban behavioural processes, and can help inform assessments of the issues described above.

This thesis is based on a previous vehicle transaction model developed by Abolfazl Mohammadian. The overall structure of this new model is relatively consistent with what was originally developed, and several subcomponents have been retained for this revised model. Nonetheless, there are three key conceptual improvements made to the original model. These are:

1. The desire to add or remove a vehicle from the household fleet explicitly accounts for the expected increase/decrease in travel convenience that taking such an action would invoke.

2. The impact of fuel prices is more explicitly accounted for; this should help improve representation of changing behaviour in terms of both what and how many vehicles a household will own in the face of rising fuel prices.

3. Purchase price as well as operation and maintenance costs are both represented in the context

of their overall impact on household budgets, and thus the ability of households to afford these expenses. This process can also be extended to help inform housing location choices.

A number of other conceptual changes were also made to the original work to allow it to be used as part of a simulation process.

A total of four separate choice models were estimated: a disposal choice model, a class-vintage choice model, a transaction choice model and a vehicle ownership level model that only applies to households in their first year in the ILUTE "world"to provide a starting vehicle fleet from which they can make transaction choices from that point forward.

Model results were generally found to be well-fitting, with the exception of the class-vintage model which, despite significant efforts, still suffers from a high level of predictive error. A literature review on this type of model suggests that class and vintage choices are challenging to predict, and most comparable models suffer from similar weak performance. A review of the original TACOS data found that stated reasons for purchasing their particular vehicles varies substantially between households and even between vehicles, and thus class and vintage choices have a relatively heterogeneous decision-making framework.

To the best of the author's knowledge, the vehicle fleet model developed herein is believed to be one of the most holistic vehicle models to be developed in terms of its ability to incorporate dynamic feedback loops with household place of residence, place(s) of employment, participation in other activities, and of course travel choices and transportation infrastructure. It also provides an excellent platform with which to develop more accurate assessments of vehicle emissions and pollution exposure.

Despite these advancements, there is still further development required in this area, and a number of potential extensions and applications of the work were also discussed.

The most pressing short-term work required is the implementation of the model into ILUTE, a minor amount of additional data collection for exogenous variables to make it simulate over the 20-year validation period and then validation of the accuracy of the model in terms of its ability to predict behaviour over long periods of time. Subject to the development of other parts of ILUTE, the transaction choice model can then be extended to more explicitly incorporate how add/subtracting a vehicle from the household fleet would be in providing households with the ability to live in certain locations as well as access certain jobs and other activity participation opportunities (i.e. above and beyond just making travel more convenient).

Over the long-term, if the model is to be used for forecasting purposes, a number of potential scope extensions are recommended. A second round of data collection is also recommended to help assess the temporal transferability of the vehicle fleet model, and whether decision-making patterns are changing in ways that the model is unable to account for. Thought should be given to new vehicle technology and business trends such as hybrid and alternative fuel vehicles, vehicle-share services and driverless vehicles, and how they could be represented in ILUTE and what sorts of effects they would have on the rest of the larger urban system.

Bibliography

Adler, Thomas. 2003. Incentives for Alternate Fuel Vehicles : A Large-Scale Stated Preference Experiment. *10th International Conference on Travel Behaviour Research.*

Beesley, M.E., & Kain, J.F. 1964. Urban Form, Car Ownership and Public Policy: An Appraisal of Traffic in Towns. *Urban Studies*, **1**(2), 174–203.

Beggs, S, Cardell, S, & Hausman, J. 1981. Assessing the potential demand for electric cars. *Journal of Econometrics*, **16**, 1–19.

Beggs, Steven D., & Cardell, N. Scott. 1980. Choice of smallest car by multi-vehicle households and the demand for electric vehicles. *Transportation Research Part A: General*, **14A**, 389–404.

Ben-Akiva, Moshe, & Lerman, Steven R. 1974. Some estimation results of a simultaneous model of auto ownership and mode choice to work. *Transportation*, **3**, 357–376.

Ben-Akiva, Moshe, & Lerman, Steven R. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press.

Berkovec, James, & Rust, John. 1985. A nested logit model of automobile holdings for one vehicle households. *Transportation Research Part B: Methodological*, **19**(4), 275–285.

Bhat, Chandra R., & Guo, Jessica Y. 2007. A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B: Methodological*, **41**(5), 506–526.

Bhat, Chandra R, & Pulugurta, Vamsi. 1998. A comparison of two alternative behavioral choice mechanisms for household ato ownership decisions. *Transportation Research Part B: Methodological*, **32**(1), 61–75.

Bhat, Chandra R., Sen, Sudeshna, & Eluru, Naveen. 2009. The impact of demographics, built environment attributes, vehicle characteristics, and gasoline prices on household vehicle holdings and use. *Transportation Research Part B: Methodological*, **43**(1), 1–18.

Bierlaire, Michel. 2003. BIOGEME: a free package for the estimation of discrete choice models. *3rd Swiss Transport Research Conference*.

Blumenberg, Evelyn, Taylor, Brian D., Smart, Michael, Ralph, Kelcie, Wander, Madeline, & Brumbaugh, Stephen. 2012. *Driving Costs - Beyond the price tag: Understanding your vehicles expenses*. Tech. rept. Institute of Transportation Studies, University of California Los Angeles.

Brems, Hans. 1956. Long-run automobile demand. *The Journal of Marketing*, **20**(4), 379–384.

Briggs, James. 2013 (June 19). *Letter from a millennial: We're not going to buy your house*. Baltimore Business Journal.

Brownstone, David, Bunch, David S., Golob, Thomas F., & Ren, Weiping. 1996. A transactions choice model for forecasting demand for alternative-fuel vehicles. *Pages 87–129 of: Research in Transportation Economics*, vol. 4. Elsevier Ltd.

Brownstone, David, Bunch, David S., & Train, Kenneth. 2000. Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research Part B: Methodological*, **34**(5), 315–338.

Bunch, David S., Bradley, Mark, Golob, Thomas F., & Kitamura, Ryuichi. 1993. Demand for clean-fuel vehicles in California: a discrete-choice stated preference pilot project. *Transportation Research Part A: Policy and Practice*, **27A**(3), 237–253.

Button, K.J., Fowkes, A.S., & Pearman, A.D. 1980. Disaggregate and aggregate car ownership forecasting in Great Britain. *Transportation Research Part A: Policy and Practice*, **14A**, 263–273.

Canadian Automobile Association. 2012. *Driving Costs - Beyond the price tag: Understanding your vehicles expenses*. Tech. rept. Canadian Automobile Association.

Canadian Infrastructure Report Card. 2012. *Volume 1: 2012 Municipal Roads and Water Systems*. Tech. rept. Canadian Construction Association, Canadian Public Works Association, Canadian Society for Civil Engineering and the Federation of Canadian Municipalities.

Cao, Xinyu, Mokhtarian, Patricia L., & Handy, Susan L. 2006. Neighborhood design and vehicle type choice: Evidence from Northern California. *Transportation Research Part D: Transport and Environment*, **11**(2), 133–145.

Cao, Xinyu, Mokhtarian, Patricia L, & Handy, Susan L. 2007. Cross-sectional and quasi-panel explorations of the connection between the built environment and auto ownership. *Environment and Planning A*, **39**(4), 830–847.

Cervero, Robert, Golub, Aaron, & Nee, Brendan. 2007. City CarShare: Longer-Term Travel Demand and Car Ownership Impacts. *Transportation Research Record*, **1992**(1), 70–80.

Chingcuanco, Franco, & Miller, Eric J. 2012. A meta-model of vehicle ownership choice parameters. *13th International Conference on Travel Behaviour Research*.

Choo, Sangho, & Mokhtarian, Patricia L. 2004. What type of vehicle do people drive? The role of attitude and lifestyle in influencing vehicle type choice. *Transportation Research Part A: Policy and Practice*, **38**(3), 201–222.

Chozick, Amy. 2012 (March 23). *As Young Lose Interest in Cars, G.M. Turns to MTV for Help*. The New York Times.

Chu, You-Lian. 2002. Automobile ownership analysis using ordered probit models. *Transportation Research Record: Journal of the Transportation Research Board*, **1805**, 60–67.

City of Toronto. 2012. *2011 Collision Summary*. Tech. rept. Traffic Safety Unit, Traffic Management Centre, Transportation Services, City of Toronto.

Costain, Cindy, Ardron, Carolyn, & Habib, Khandker Nurul. 2012. Synopsis of users behaviour of a carsharing program: A case study in Toronto. *Transportation Research Part A: Policy and Practice*, **46**(3), 421–434.

Cragg, John G., & Uhler, Russell S. 1970. The demand for automobiles. *The Canadian Journal of Economics/Revue canadienne d'Economique*, **3**(3), 386–406.

Dargay, Joyce M. 2001. The effect of income on car ownership: evidence of asymmetry. *Transportation Research Part A: Policy and Practice*, **35**, 807–821.

Data Management Group. 2008. *2006, 2001, 1996 and 1986 Travel Survey Summaries for the Greater Toronto and Hamilton Area*. Tech. rept. Department of Civil Engineering, University of Toronto.

de Janosi, Peter E. 1959. Factors influencing the demand for new automobiles. *The Journal of Marketing*, **23**(4), 412–418.

de Jong, Gerard. 1996. A disaggregate model system of vehicle holding duration, type choice and use. *Transportation Research Part B: Methodological*, **30**(4), 263–276.

de Jong, Gerard, & Kitamura, Ryuichi. 2009. A review of household dynamic vehicle ownership models: holdings models versus transactions models. *Transportation*, **36**(Dec.), 733–743.

Dunn, Collin. 2006 (November 16). *The TH Interview: Kevin McLaughlin, AutoShare President*. Treehugger.

Eluru, Naveen, Bhat, Chandra R., Pendyala, Ram M., & Konduri, Karthik C. 2010. A joint flexible econometric model system of household residential location and vehicle fleet composition/usage choices. *Transportation*, **37**(4), 603–626.

Ewing, Reid, Schieber, Richard a, & Zegeer, Charles V. 2003. Urban sprawl as a risk factor in motor vehicle occupant and pedestrian fatalities. *American Journal of Public Health*, **93**(9), 1541–5.

Flavelle, Dana. 2012 (June 7). *Toronto car-sharing market in over-drive*. The Toronto Star.

Frumkin, Howard, Frank, Lawrence, & Jackson, Richard. 2004. *Urban Sprawl and Public Health: Designing, Planning, and Building for Healthy Communities*. Island Press.

Golob, Thomas F., Kim, Seyoung, & Ren, Weiping. 1996. How households use different types of vehicles: A structural driver allocation and usage model. *Transportation Research Part A: Policy and Practice*, **30**(2), 103–118.

Hatzopoulou, Marianne. 2008. *An Integrated Multi-Model Approach for Predicting the Impact of Household Travel on Urban Air Quality and Simulating Population Exposure*. Ph.D. thesis, University of Toronto.

Hatzopoulou, Mirianne, Miller, Eric J, & Santos, Bruno. 2007. Integrating Vehicle Emission Modeling with Activity-Based Travel Demand Modeling: Case Study of the Greater Toronto, Canada, Area. *Transportation Research Record: Journal of the Transportation Research Board*, 29–39.

HDR. 2008. *Costs of Road Congestion in the Greater Toronto and Hamilton Area: Impact and Cost Benefit Analysis of the Metrolinx Draft Regional Transportation Plan: Final Report*. Tech. rept. Greater Toronto Transportation Authority.

Hensher, David A., & Le Plastrier, Vicki. 1985. Towards a dynamic discrete-choice of household automobile and composition model. *Transportation Research Part B: Methodological*, **19B**(6), 481–495.

Hensher, David A., Rose, John M., & Black, Ian. 2008. Interactive Agency Choice in Automobile Purchase Decisions: The Role of Negotiation in Determining Equilibrium Choice Outcomes. *Journal of Transport Economics and Policy*, **42**(2), 269–296.

Hess, Stephane, Fowler, Mark, Adler, Thomas, & Bahreinian, Aniss. 2011. A joint model for vehicle type and fuel type choice: evidence from a cross-nested logit study. *Transportation*, **39**(3), 593–625.

Hocherman, Irit, Prashker, Joseph, & Ben-Akiva, Moshe. 1983. Estimation and use of dynamics transaction models of automobile ownership. *Transportation Research Record: Journal of the Transportation Research Board*, 34–55.

ICF International. 2007. *Greenhouse Gases and Air Pollutants in the City of Toronto Toward a Harmonized Strategy for Reducing Emissions.* Tech. rept. ICF International in collaboration with Toronto Atmospheric Fund and Toronto Environment Office.

Johnson, Terry R. 1978. A cross-section analysis of the demand for new and used automobiles in the United States. *Economic Inquiry*, **XVI**, 531–548.

Kain, John F., & Beesley, ME. 1965. Forecasting car ownership and use. *Urban Studies*, **2**(2), 163–185.

Kalita, S. Mitra, & Whelan, Robbie. 2011 (January 13). *No McMansions for Millennials.* Wall Street Journal Blogs.

Koppelman, Frank S., & Bhat, Chandra. 2006. *Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models.* Tech. rept. Prepared For U.S. Department of Transportation Federal Transit Administration.

Kreinin, Mordechai E. 1959. Analysis of Used Car Purchases. *The Review of Economics and Statistics*, **41**(4), 419–425.

Lave, Charles A., & Train, Kenneth E. 1979. A disaggregate model of auto-type choice. *Transportation Research Part A: General*, **13A**, 1–9.

Lerman, Steven R., & Ben-Akiva, Moshe. 1976. Disaggregate behavior model of automobile ownership. *Transportation Research Record: Journal of the Transportation Research Board*, 34–55.

Mannering, F, & Mahmassani, Hani. 1985. Consumer valuation of foreign and domestic vehicle attributes: econometric analysis and implications for auto demand. *Transportation Research Part A: General*, **19A**(3), 243–251.

Manski, Charles F., & Sherman, Leonard. 1980. An empirical analysis of household choice among motor vehicles. *Transportation Research Part A: General*, **14**(1976), 349–366.

Martin, Elliot, Shaheen, Susan A., & Lidicker, Jeffrey. 2010. Impact of Carsharing on Household Vehicle Holdings. *Transportation Research Record: Journal of the Transportation Research Board*, **2143**(Dec.), 150–158.

McKeown, David. 2007. *Air Pollution Burden of Illness from Traffic in Toronto Problems and Solutions*. Tech. rept. Toronto Public Health, City of Toronto.

Metrolinx. 2008. *The Big Move: Transforming Transportation in the Greater Toronto and Hamilton Area*. Tech. rept. Metrolinx: An Agency of the Government of Ontario.

Miller, Eric J, & Roorda, Matthew J. 2003. A Prototype Model of 24-Hour Household Activity Scheduling for the Toronto Area. *Transportation Research Record: Journal of the Transportation Research Board*, **1831**, 114–121.

Miller, Eric J., Kriger, David S., & Hunt, John D. 1999. *Integrated Urban Models for Simulation of Transit and Land Use Policies: Guidelines for Implementation and Use*. Vol. Report 48. Transit Co-operative Research Program.

Miller, Eric J., Farooq, Bilal, Chingcuanco, Franco, & Wang, David. 2011. Historical Validation of Integrated Transport-Land Use Model System. *Transportation Research Record: Journal of the Transportation Research Board*, **2255**(Dec.), 91–99.

Ministry of Transportation of Ontario. 2013. *Ontario Road Safety Annual Report (ORSAR) 2010*.

Mohammadian, Abolfazl. 2002. *Dynamic Modelling of Household Automobile Transactions Within A Microsimulation Framework*. Ph.D. thesis, University of Toronto.

Nerlove, Marc. 1957. A note on long-run automobile demand. *The Journal of Marketing*, **22**(1), 57–64.

Paleti, Rajesh, Eluru, Naveen, Bhat, Chandra R., Pendyala, Ram M., Adler, Thomas J., & Goulias, Konstadinos G. 2011. Design of Comprehensive Microsimulator of Household Vehicle Fleet Composition, Utilization, and Evolution. *Transportation Research Record: Journal of the Transportation Research Board*, **2254**(Dec.), 44–57.

Potoglou, Dimitris. 2008. Vehicle-type choice and neighbourhood characteristics: An empirical study of Hamilton, Canada. *Transportation Research Part D: Transport and Environment*, **13**(3), 177–186.

Potoglou, Dimitris, & Kanaroglou, Pavlos S. 2007. Household demand and willingness to pay for clean vehicles. *Transportation Research Part D: Transport and Environment*, **12**(4), 264–274.

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

Pritchard, David R. 2008. *Synthesizing Agents and Relationships for Land Use / Transportation Modelling*. M.A.Sc. thesis, University of Toronto.

Pritchard, David R., & Miller, Eric J. 2011. Advances in population synthesis: fitting many attributes per agent and fitting to household and person margins simultaneously. *Transportation*, **39**(3), 685–704.

Rashidi, Taha H., Mohammadian, Abolfazl, & Koppelman, Frank S. 2011. Modeling interdependencies between vehicle transaction, residential relocation and job change. *Transportation*, **38**(6), 909–932.

Roorda, Matthew J. 1998. *Toronto Area Car Ownership Study: A Longitudinal Survey and a Preliminary Analysis of Results*. M.A.Sc. thesis, University of Toronto.

Roorda, Matthew J. 2005. *Activity-Based Modelling of Household Travel*. Ph.D. thesis, University of Toronto.

Roorda, Matthew J, Mohammadian, Abolfazl, & Miller, Eric J. 2000. Toronto area car ownership study: A retrospective interview and its applications. *Transportation Research Record: Journal of the Transportation Research Board*, **1719**, 69–76.

Roorda, Matthew J., Miller, Eric J., & Habib, Khandker M.N. 2008. Validation of TASHA: A 24-h activity scheduling microsimulation model. *Transportation Research Part A: Policy and Practice*, **42**(2), 360–375.

Roorda, Matthew J., Carrasco, Juan A., & Miller, Eric J. 2009. An integrated model of vehicle transactions, activity scheduling and mode choice. *Transportation Research Part B: Methodological*, **43**(2), 217–229.

Salvini, Paul. 2003. *Design and Development of the ILUTE Operational Prototype: A Comprehensive Mi-crosimulation Model of Urban Systems.* Ph.D. thesis, University of Toronto.

Salvini, Paul, & Miller, Eric J. 2005. ILUTE: An Operational Prototype of a Comprehensive Microsimulation Model of Urban Systems. *Networks and Spatial Economics*, **5**(2), 217–234.

Shoup, Donald. 2005. *The High Cost of Free Parking.* APA Planners Press.

Statistics Canada. 2012. *2011 Census of Population: Census in brief: living arrangements of young adults aged 20 to 29.* Tech. rept. Statistics Canada, Government of Canada.

Tanner, J.C. 1966. Comments on Forecasting Car Ownership and Use. *Urban Studies*, **3**(2), 143–146.

Toronto Public Health. 2012. *Road to Health: Improving Walking and Cycling in Toronto.* Tech. rept. Toronto Public Health, City of Toronto.

Train, Kenneth E. 1980a. A structured logit model of auto ownership and mode choice. *The Review of Economic Studies*, **47**(2), 357–370.

Train, Kenneth E. 1980b. The potential market for non-gasoline-powered automobiles. *Transportation Research Part A: General*, **14A**, 405–414.

Train, Kenneth E., & Winston. 2007. Vehicle choice behavior and the declining market share of us automakers. *International Economic Review*, **48**(4), 1469–1496.

Transport Canada. 2012. *Canadian Motor Vehicle Traffic Collision Statistics: 2010.* Tech. rept. Government of Canada.

Vodden, Keith, Smith, Douglas, Eaton, Frank, & Mayhew, Dan. 2007. *Analysis and Estimation of the Social Cost of Motor Vehicle Collisions in Ontario: Final Report.* Tech. rept. Transport Canada.

Vyas, Gaurav, Paleti, Rajesh, Bhat, Chandra R., Goulias, Konstadinos G., Pendyala, Ram M., Hu, Hsi-Hwa, Adler, Thomas J., & Bahreinian, Aniss. 2012. Joint Vehicle Holdings, by Type and Vintage, and Primary Driver Assignment Model with Application for California. *Transportation Research Record: Journal of the Transportation Research Board*, **2302**(Dec.), 74–83.

Weissman, Jordan. 2012 (March 25). *Why Don't Young Americans Buy Cars?* The Atlantic.

Appendix A

TASHA Modal Assignment Procedure

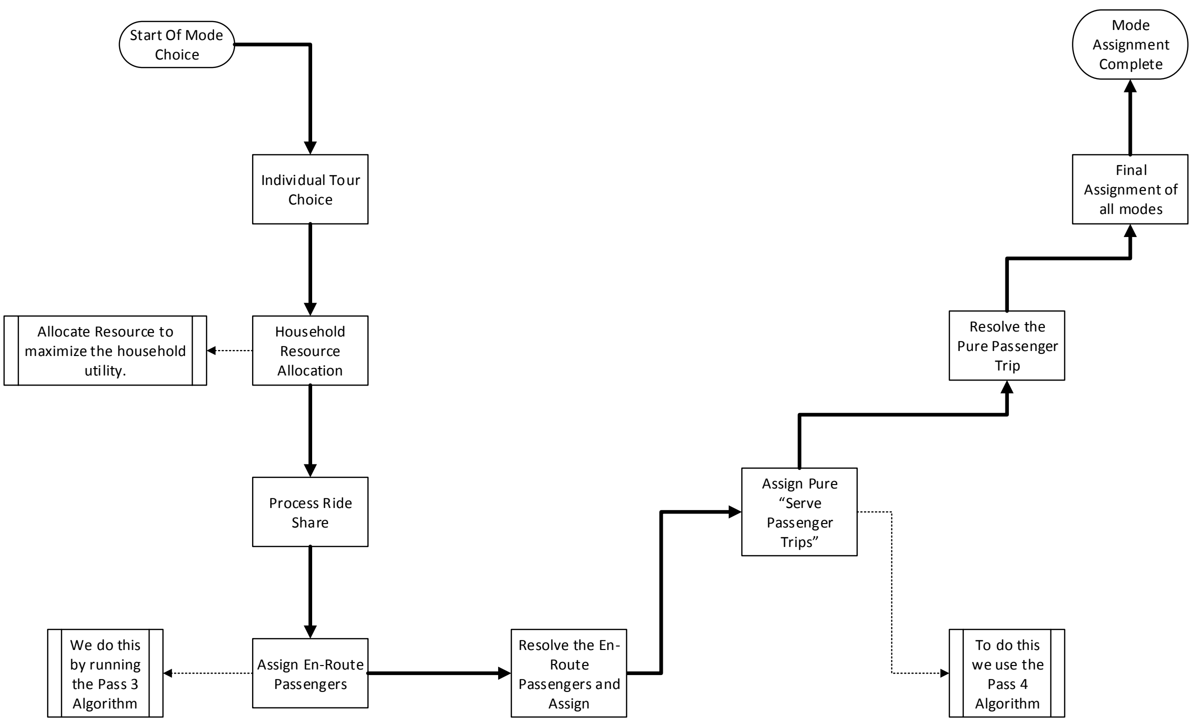

Figure A.1: TASHA Modal Assignment Procedure (1 of 3)

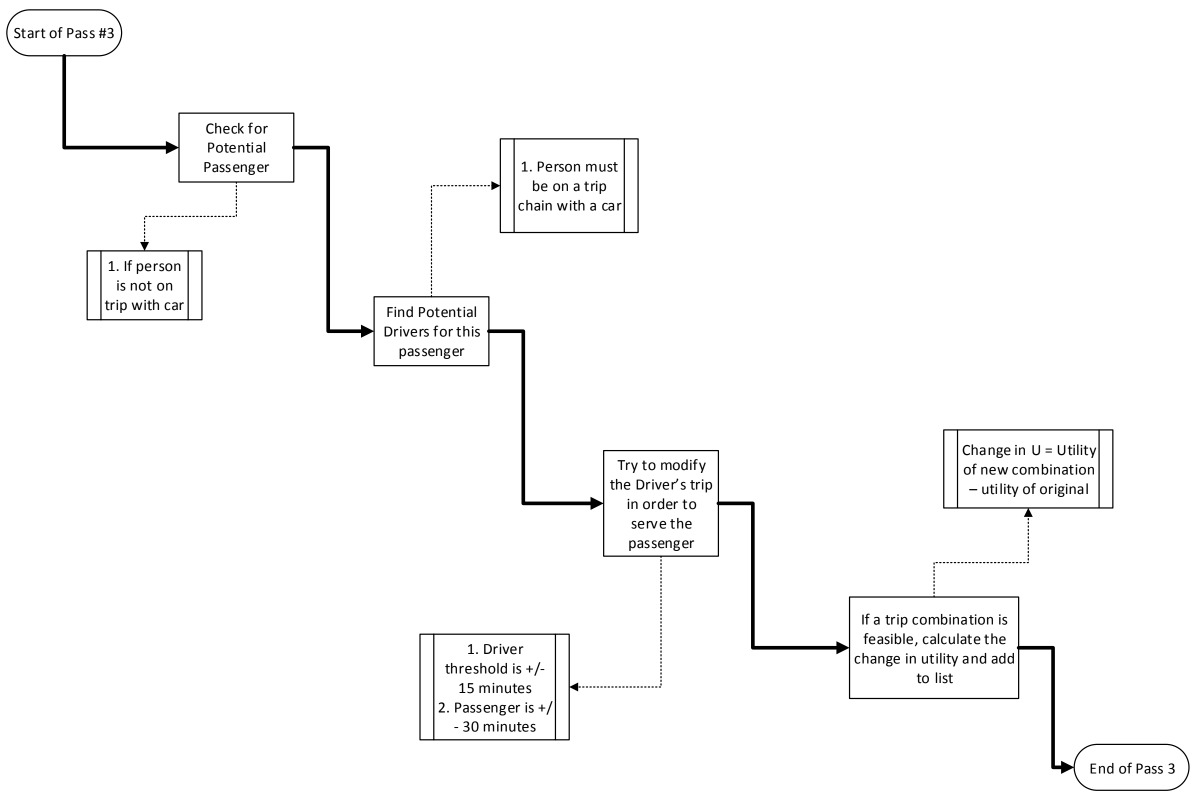

Figure A.2: TASHA Modal Assignment Procedure (2 of 3)

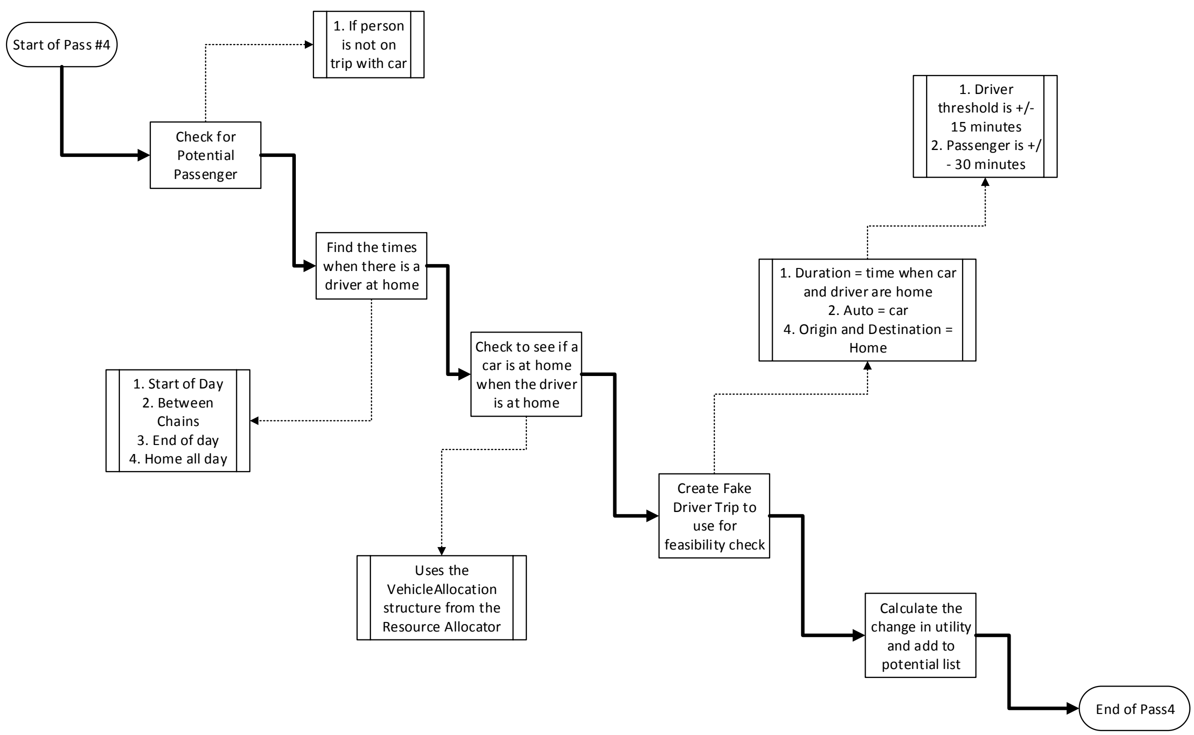

Figure A.3: TASHA Modal Assignment Procedure (3 of 3)

Appendix B

Vehicle Property Generation and Price Model

Table B.1: Simulated Vehicle Attributes by Class

Vehicle Attribute	**Subcompact**	**Compact**	**Midsize**	**Large**	**Special Purpose Vehicle**	**Van**
Luggage capacity (m^3)	0.345	0.381	0.445	0.605	2.384	2.161
Wheelbase (m)	2.474	2.604	2.693	2.819	2.819	2.905
Engine displacement (L)	2.030	2.270	3.061	3.842	3.548	3.540
Weight (tonnes)	1.078	1.197	1.385	1.534	1.551	1.640

Table B.3: Simulated Vehicle Fuel Intensity (L/100km) by Class and Model Year

Year	**Subcompact**	**Compact**	**Midsize**	**Large**	**Special Purpose Vehicle**	**Van**
1999	7.136	7.650	8.851	8.576	10.721	11.201
1998	7.136	7.650	8.851	8.576	10.721	11.201
1997	6.994	7.749	8.834	9.650	10.810	10.352
1996	7.863	8.150	8.357	9.960	10.047	10.282
1995	7.657	7.767	9.107	9.586	10.585	10.923
1994	7.430	8.213	8.925	9.668	10.221	11.212
1993	7.174	7.980	9.134	8.666	11.099	10.066
1992	7.485	8.175	9.076	9.674	9.889	10.818
1991	7.256	8.191	9.099	9.547	9.939	10.098
1990	7.092	7.792	9.084	9.207	10.937	10.610
1989	7.601	7.953	9.591	9.295	10.475	10.458
1988	7.149	8.125	8.888	9.083	11.881	10.291
1987	7.423	7.822	9.007	9.291	10.262	10.682
1986	7.183	7.395	8.811	9.095	9.700	9.982
1985	6.994	7.080	9.071	10.355	10.560	10.433
1984	7.546	6.595	9.158	9.119	10.551	12.070
1983	7.508	8.020	8.758	11.820	8.532	13.386
1982	7.827	6.662	9.787	8.111	9.370	14.043
1981	8.152	8.401	10.094	10.746	14.701	14.701
1980	8.691	10.454	11.031	12.615	14.293	13.574
1979	10.960	11.823	9.699	13.580	13.884	12.447
1978	10.826	13.210	12.186	14.413	8.111	15.191

Table B.4: Vehicle Age Assignment

Vintage Category	**Age**	**Age Probability (given Vintage)**
Brand New	-1	17.1%
	0	82.9%
Nearly New	1	55.7%
	2	44.3%
Used	3	27.9%
	4	25.0%
	5	15.7%
	6	14.6%
	7	16.8%
Old	8	22.6%
	9	23.7%
	10	16.1%
	11	16.7%
	12	21.0%

Table B.5: Vehicle Origin Assignment

Origin Category	**Origin Probability**
Domestic	70%
Japanese	25%
European	5%

Table B.6: Updated Hedonic Vehicle Price Model (1000's, \$1998)

Variable	**Coefficient**	**Standard Error**	**t-statistic**
Subcompact	12.753	0.423	30.135
Compact	13.767	0.381	36.11
Midsize	15.288	0.375	40.718
Large	16.395	0.478	34.285
Special Purpose Vehicle	13.295	n/a	n/a
Van	12.897	0.69	18.694
New	0.779	0.299	2.604
Natural Logarithm of Vehicle Age	-6.072	0.166	-36.651
Japanese Car	3.746	0.26	14.392
European Car	4.484	0.494	9.068
Vehicle Performance Factor	2.294	0.168	13.628
Vehicle Space Factor	1.038	0.273	3.8
Time[1]	0.514	0.421	12.216

[1]Measured in years to/from 1990.

Table B.7: Revised Operation and Maintenance Fixed Costs

Class	Vintage	Maintenance	Insurance	Total
Subcompact	Brand New	0.090	4.003	**4.093**
	Nearly New	1.064	3.790	**4.854**
	Used	1.361	3.260	**4.621**
	Old	1.610	2.490	**4.100**
Compact	Brand New	0.090	3.552	**3.642**
	Nearly New	1.038	3.340	**4.378**
	Used	1.324	2.918	**4.242**
	Old	1.544	2.281	**3.825**
Midsize	Brand New	0.090	3.783	**3.873**
	Nearly New	1.214	3.440	**4.654**
	Used	1.573	2.888	**4.461**
	Old	1.876	2.182	**4.058**
Large	Brand New	0.090	3.600	**3.690**
	Nearly New	1.230	3.381	**4.611**
	Used	1.600	2.944	**4.544**
	Old	1.917	2.318	**4.235**
Special Purpose Vehicle	Brand New	0.090	3.178	**3.268**
	Nearly New	1.150	3.499	**4.649**
	Used	1.489	3.604	**5.093**
	Old	1.780	2.274	**4.054**
Van	Brand New	0.090	3.295	**3.385**
	Nearly New	1.118	3.081	**4.199**
	Used	1.431	2.654	**4.085**
	Old	1.698	1.930	**3.628**

Appendix C

Macroeconomic Variables

Table C.1: Macroeconomic Data

Year	CPI	Interest Rate (%)	Unemployment Rate (%)	Gasoline Price (¢/L)
1986	79.0	9.2	7.0	45.70
1987	82.7	8.4	6.1	48.38
1988	87.6	9.7	5.1	48.08
1989	92.8	12.3	5.0	50.57
1990	96.5	13.0	6.2	56.91
1991	99.2	9.0	9.6	55.00
1992	100.6	6.8	10.8	53.66
1993	102.2	5.1	10.9	52.38
1994	102.6	5.8	9.6	50.77
1995	104.5	7.3	8.8	54.03
1996	107.3	4.5	9.0	57.57
1997	107.9	3.5	8.4	57.78
1998	109.1	5.1	7.2	53.34
1999	112.5	4.9	6.4	59.53
2000	116.2	5.8	5.7	72.43
2001	118.1	4.3	6.3	69.28
2002	122.2	2.7	7.2	69.30
2003	125.4	3.2	6.9	72.73
2004	126.8	2.5	6.8	78.89
2005	129.5	2.9	6.6	89.60
2006	130.8	4.3	6.3	94.01

Appendix D

Model Estimation Details

D.1 Disposal Choice Model

This section provides the estimation details of the Disposal Choice model that was discussed in Section 6.3.

D.1.1 Description

This is a multinomial logit model for household vehicle disposal choices

Model	:	Multinomial Logit
Number of estimated parameters	:	6
Number of observations	:	623
Number of individuals	:	623
Null log-likelihood	:	-1212.302
Cte log-likelihood	:	-390.197
Init log-likelihood	:	-1212.302
Final log-likelihood	:	-286.852
Likelihood ratio test	:	1850.900
Rho-square	:	0.763
Adjusted rho-square	:	0.758
Final gradient norm	:	+6.307e-004
Diagnostic	:	Convergence reached...
Iteration	:	161
Run time	:	00:04
Variance-covariance	:	from analytical hessian

D.1.2 Summary statistics

Summary statistics

Number of observations = 623

$$\begin{aligned}
\mathcal{L}(0) &= -1212.302 \\
\mathcal{L}(c) &= -390.197 \\
\mathcal{L}(\hat{\beta}) &= -286.852 \\
-2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] &= 1850.900 \\
\rho^2 &= 0.763 \\
\bar{\rho}^2 &= 0.758
\end{aligned}$$

D.1.3 Parameters

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	B_LRGPL	3.86	0.324	11.89	0.00
2	B_LUGWB	28.2	2.13	13.26	0.00
3	B_OLD	1.55	0.292	5.31	0.00
4	B_OMFIX_INC	22.7	9.63	2.36	0.02
5	B_OMVAR_INC	6.75	4.92	1.37	0.17
6	B_USED	0.945	0.289	3.26	0.00

D.2 Class and Vintage Choice Model

This section provides the estimation details of the Class and Vintage Choice model that was discussed in Section 6.4.

D.2.1 Description

This is a nested logit model for vehicle type choice. At the top level, a class choice is computed. It includes six possible classes; Subcompact, Compact, Midsize, Large, Special Purpose Vehicle (pick-ups and SUV's) and Van. The lower level of the nest is a vintage choice model, which includes four vintage categories; Brand New (-1-0 y/o), Nearly New (1-2 y/o), Used (3-7 y/o) and Old (8+ y/o).

Model	:	Nested Logit
Number of estimated parameters	:	21
Number of observations	:	998
Number of individuals	:	998
Null log-likelihood	:	-3171.698
Cte log-likelihood	:	-3031.594
Init log-likelihood	:	-3171.698
Final log-likelihood	:	-2930.125
Likelihood ratio test	:	483.145
Rho-square	:	0.076
Adjusted rho-square	:	0.070
Final gradient norm	:	+2.141e-002
Diagnostic	:	Convergence reached...
Iteration	:	39
Run time	:	00:12
Variance-covariance	:	from finite difference hessian

D.2.2 Summary statistics

Summary statistics

Number of observations = 998

$$
\begin{aligned}
\mathcal{L}(0) &= -3171.698 \\
\mathcal{L}(c) &= -3031.594 \\
\mathcal{L}(\hat{\beta}) &= -2930.125 \\
-2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] &= 483.145 \\
\rho^2 &= 0.076 \\
\bar{\rho}^2 &= 0.070
\end{aligned}
$$

D.2.3 Parameters

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	B_ChlOvPpl	1.65	0.717	2.30	0.02
2	B_DrvMale	0.544	0.236	2.30	0.02
3	B_DrvMgrPro	0.340	0.316	1.07	0.28
4	B_HHOMFixOvrHHInc	-8.91	11.3	-0.79	0.43
5	B_HHOMVarOvrHHInc	-52.2	18.7	-2.80	0.01
6	B_HHVehAvgAge_V1	-0.0352	0.0293	-1.20	0.23
7	B_HHVehAvgAge_V4	0.0481	0.0473	1.02	0.31
8	B_IsClsDisp	0.979	0.0971	10.08	0.00
9	B_LnPPHHInt	0.688	0.550	1.25	0.21
10	B_PPOvrHHInc	-0.989	0.821	-1.21	0.23
11	B_V2	-0.616	0.499	-1.24	0.22
12	B_V3	-0.257	0.251	-1.02	0.31
13	B_V4	-0.888	0.808	-1.10	0.27
14	B_VPF	-0.192	0.316	-0.61	0.54
15	B_VSF	-0.242	1.69	-0.14	0.89
16	COMPACT	1.14	0.800	0.17[1]	0.86
17	LARGE	1.27	0.935	0.28[1]	0.78
18	MIDSIZE	1.08	1.07	0.08[1]	0.94
19	SPV	1.34	1.57	0.21[1]	0.83
20	SUBCOMPACT	1.23	0.866	0.26[1]	0.79
21	VAN	1.99	1.95	0.51[1]	0.61

[1] t-test against 1

D.3 Transaction Choice Model

This section provides the estimation details of the Transaction Choice model that was discussed in Section 6.5.

D.3.1 Description

This is a multinomial logit model that assesses household vehicle transaction decisions with four potential alternatives: Buy a vehicle, Replace a vehicle (i.e. both Dispose of one and Buy one), Do Nothing and Dispose of a vehicle.

Model	:	Multinomial Logit
Number of estimated parameters	:	29
Number of observations	:	4164
Number of individuals	:	4164
Null log-likelihood	:	-5105.029
Init log-likelihood	:	-5105.029
Final log-likelihood	:	-2522.544
Likelihood ratio test	:	5164.970
Rho-square	:	0.506
Adjusted rho-square	:	0.500
Final gradient norm	:	+1.662e-002
Diagnostic	:	Convergence reached...
Iteration	:	27
Run time	:	00:10
Variance-covariance	:	from analytical hessian

D.3.2 Summary statistics

Summary statistics

Number of observations = 4164

$$\begin{aligned}
\mathcal{L}(0) &= -5105.029 \\
\mathcal{L}(c) &= ??? \\
\mathcal{L}(\hat{\beta}) &= -2522.544 \\
-2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] &= 5164.970 \\
\rho^2 &= 0.506 \\
\bar{\rho}^2 &= 0.500
\end{aligned}$$

D.3.3 Parameters

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	B_AGEAVG	0.0853	0.0155	5.50	0.00
2	B_BUY_DELTA_HH_RS_UTIL	0.226	0.0814	2.78	0.01
3	B_BUY_DELTA_HH_TRAV_UTIL	-0.193	0.0803	-2.40	0.02
4	B_BUY_DELTA_NUM_CONF	-0.333	0.237	-1.41	0.16
5	B_CHDEC	-0.783	0.750	-1.04	0.30
6	B_CVINCVAL	0.151	0.0380	3.97	0.00
7	B_DINCVAL	0.00224	0.00510	0.44	0.66
8	B_DISP	-1.85	0.681	-2.71	0.01
9	B_DISP_DELTA_HH_RS_UTIL	-0.0525	0.133	-0.40	0.69
10	B_DISP_DELTA_HH_TRAV_UTIL	0.0765	0.129	0.59	0.55
11	B_DISP_DELTA_NUM_CONF	0.332	0.288	1.15	0.25
12	B_DLGTFS	1.83	0.195	9.42	0.00
13	B_DLLTFS	1.21	0.342	3.53	0.00
14	B_DN	2.85	0.680	4.19	0.00
15	B_HHDEC	1.07	0.556	1.92	0.06
16	B_HHINC	0.321	0.201	1.60	0.11
17	B_INTRT_BUY	-0.0758	0.0357	-2.12	0.03
18	B_INTRT_REPL	-0.0387	0.0350	-1.10	0.27
19	B_JBDEC	0.889	0.380	2.34	0.02
20	B_JBINC_BUY	0.242	0.186	1.31	0.19
21	B_JBINC_REPL	0.292	0.182	1.61	0.11
22	B_LNYRSTRAN	-0.263	0.0856	-3.08	0.00
23	B_NUMCHLD	0.139	0.0598	2.32	0.02
24	B_NUM_VEH	-0.487	0.0769	-6.34	0.00
25	B_OLD_NN	0.310	0.176	1.76	0.08
26	B_REPL	0.178	0.855	0.21	0.84
27	B_UNEMPRT_BUY	-0.171	0.0698	-2.44	0.01
28	B_UNEMPRT_REPL	-0.0761	0.0654	-1.16	0.24
29	B_ZERO_VEH	1.48	0.279	5.30	0.00

D.4 Ownership Level Model

This section provides the estimation details of the Ownership Level model that was discussed in Section 7.1.

D.4.1 Description

This is a multinomial logit model used to estimate the number of vehicle a GTHA household will own, ranging from 0 to f.

Model	:	Multinomial Logit
Number of estimated parameters	:	30
Number of observations	:	6360
Number of individuals	:	6360
Null log-likelihood	:	-10236.025
Cte log-likelihood	:	-7521.380
Init log-likelihood	:	-10236.025
Final log-likelihood	:	-6417.772
Likelihood ratio test	:	7636.507
Rho-square	:	0.373
Adjusted rho-square	:	0.370
Final gradient norm	:	+3.752e-002
Diagnostic	:	Convergence reached...
Iteration	:	143
Run time	:	01:14
Variance-covariance	:	from analytical hessian

D.4.2 Summary statistics

Summary statistics

Number of observations = 6360

$$\begin{aligned}
\mathcal{L}(0) &= -10236.025 \\
\mathcal{L}(c) &= -7521.380 \\
\mathcal{L}(\hat{\beta}) &= -6417.772 \\
-2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})] &= 7636.507 \\
\rho^2 &= 0.373 \\
\bar{\rho}^2 &= 0.370
\end{aligned}$$

D.4.3 Parameters

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	ASC_FOURVEH	-13.4	1.21	-11.10	0.00
2	ASC_ONEVEH	-0.211	0.107	-1.97	0.05
3	ASC_THREEVEH	-6.69	0.489	-13.70	0.00
4	ASC_TWOVEH	-2.85	0.162	-17.54	0.00
5	B_CHLD_3	-0.548	0.103	-5.33	0.00
6	B_CHLD_4	-1.26	0.290	-4.33	0.00
7	B_HAM_1	-0.468	0.111	-4.22	0.00
8	B_HAM_2	-0.736	0.130	-5.65	0.00
9	B_LNINC_1	0.164	0.0240	6.84	0.00
10	B_LNINC_2	0.259	0.0371	6.99	0.00
11	B_LNINC_3	0.427	0.120	3.56	0.00
12	B_LNINC_4	1.26	0.305	4.13	0.00
13	B_NOWKADLT_2	0.324	0.0248	13.05	0.00
14	B_NOWKADLT_3	0.515	0.0786	6.56	0.00
15	B_NOWKADLT_4	1.10	0.127	8.62	0.00
16	B_SFH_1	0.728	0.0692	10.52	0.00
17	B_SFH_2	1.55	0.0876	17.72	0.00
18	B_SFH_3	1.69	0.185	9.13	0.00
19	B_SFH_4	0.621	0.399	1.56	0.12
20	B_TO_1	-0.713	0.0799	-8.93	0.00
21	B_TO_2	-1.49	0.0970	-15.38	0.00
22	B_TO_3	-1.02	0.177	-5.79	0.00
23	B_WKADLT_1	0.567	0.0453	12.53	0.00
24	B_WKADLT_2	1.29	0.0572	22.64	0.00
25	B_WKADLT_3	1.75	0.0923	18.91	0.00
26	B_WKADLT_4	2.17	0.193	11.24	0.00
27	B_YK_1	-0.845	0.163	-5.20	0.00
28	B_YK_2	-0.732	0.174	-4.22	0.00
29	B_YK_3	0.201	0.246	0.82	0.41
30	B_YK_4	1.02	0.513	1.99	0.05