Commuting in Canada's Global Cities

by

Yan Carrière-Swallow

An essay submitted to the Department of Economics
in partial fulfillment of the requirements for
the degree of Master of Arts

Queen's University
Kingston, Ontario, Canada

August 2008

Abstract

This paper examines the commuting behaviour of Canadians living in the country's global city regions of Montreal, Toronto and Vancouver. Individual preferences for transportation mode choice are estimated using micro data from the Census of Population. The choice probability of commuting by automobile is estimated as a function of socioeconomic characteristics and distance from work using McFadden's MNL model. An urban equilibrium approach is adopted to avoid model misspecification due to omitted variables. Estimates from the micro model are then used to predict mode-shares in aggregated census tracts using average characteristics and a spatial variable constructed using GIS methods. Finally, prediction residuals are examined spatially to examine the model's accuracy across urban zones.

Acknowledgments

I thank my supervisor John Hartwick for the discussions which generated a number of the ideas presented in this paper.

Data collection was aided by the always-helpful Jeff Moon, head of the Social Science Data Centre. Long hours in the GIS lab for the completion of the spatial analysis were facilitated by Gerry Barber of the Geography Department. He kindly provided me with a crash course in ArcGIS which greatly improved the quality of this paper. Andrejs Skaburskis of the School of Urban and Regional Planning shared interesting comments about urban job concentration and data considerations. Helpful econometric insights were provided by James MacKinnon.

I thank Garrett Johnson, PhD candidate at Northwestern University, Nicolás Della Penna, and the lovely Karla Montero for their many helpful comments on an early draft.

Errors are of my own doing.

Contents

List of Figures

List of Tables

1 Introduction

Canada's global city regions will be the country's engines of growth in the 21st century economy. Between them, Montreal, Vancouver and Toronto were the destination of 68.9% of the 1.1 million immigrants who arrived in Canada between 2001 and 2006 (Statistics Canada 2008). The competitiveness of these cities relative to their global counterparts will determine whether the country can continue to attract talented individuals who can drive growth (Courchene 2004). The ability of a city's urban infrastructure to efficiently transport inhabitants to their place of work is an important factor of their perceived liveability (Conference Board 2007), and Canadian infrastructure is increasingly strained by the popularity of the automobile.

With oil prices soaring and a growing imperative to reduce emissions in response to climate change, policy makers at all levels of government are scrambling to shift Canadian commuters out of single-occupancy vehicles and into more environmentally-friendly modes of transport. In BC's lower mainland, a new tax on gasoline aims to discourage driving while a multi-billion dollar rapid transit expansion seeks to improve service to Richmond. In Toronto, an ambitious $6-billion light-rail project awaits federal and provincial funding to improve transit service outside the urban core. In Montreal, a recent expansion to the *métro* has brought rapid transit to the populous suburb of Laval and a billion-dollar bidding process is underway to upgrade the metro's cars.

Economists and civil engineers developed the analytical tools for modeling transit demand several decades ago, but the analysis of commuting patterns in Canada's major cities remains confined to descriptive statistics due to a lack of available data. If expensive transit expansions and controversial government incentive schemes are to have the desired impact of reducing our commuters' dependency on the single-occupancy vehicle, we must first answer the question: Why do people drive to work?

In this paper I analyze the worker's choice of transport mode - or, simply, *mode* *choice* - by focusing on the socio-economic characteristics of commuters in Canada's global city regions. The paper is organized as follows: Section 2 provides a brief overview of the literature surrounding the evolution of urban travel demand models and their applications. Section 3 presents descriptive statistics of the commuting situation in each city. Section 4 develops a behavioural discrete choice model suited to estimation using micro-level data. Section 5 presents the simplifying assumptions required to make the model operational with available data, and discusses estimation results. In section 6, GIS methods are used to compute approximate commuting distances for census tracts, and present an aggregation procedure which allows for the prediction of mode share values. Mapping the accuracy of these predictions across the spatial dimensions of the cities allows for inferences about the importance of infrastructure and the performance of the model's simplifying assumptions.

2 Literature Review

2.1 Towards a behavioural model for mode choice

The question of how to estimate the demand for urban travel has been addressed in a rich literature spanning a number of academic disciplines. Through the 1960s, a procedure for analyzing discrete choice decisions had not yet been formalized. Empirical studies of travel demand were limited to the analysis of continuous variables using aggregate data. Transportation engineers used a gravity model approach to estimate traffic volumes between zones as a function of their population (Meyer and Straszheim 1970). This model was successful in predicting aggregate flows, but was of little use for analyzing other dimensions of interest to the researcher and policy maker, including mode selection and trip generation (McFadden 2000). While a model for aggregate travel demand provided a starting point for the planning of major infrastructure projects, a comprehensive model capable of addressing a wider range of policy questions must begin by explaining the decisions of individual commuters.

During the early 1970s, Daniel McFadden laid the foundations for a behavioral theory of travel demand modeling using tools originally developed by mathematical psychologists such as Thurston (1927) and Luce (1959). His contribution to the field of travel demand modeling and more generally to discrete choice analysis over the following thirty years has provided a framework for a wide range of research and earned him the Nobel Prize in Economics in 2000.

A series of articles lead to the presentation of his seminal model of travel demand in a book coauthored with Tom Domencich (Domencich and McFadden 1975). Their behavioural model is derived from microeconomic foundations in consumer theory, and is developed for application to the discrete - and inherently non-marginal - context of the mode choice decision. Many aspects of their methodology form the toolbox used in subsequent research: Decisions are modeled within a random utility framework to capture heterogeneity in tastes across individuals. The individual's utility function is assumed to be separable in order to allow for the analysis of individual consumer decisions in isolation. They demonstrate that using the multinomial logit (MNL) specification is valid so long as restrictions on the distribution of random utility components can be imposed. Such a model could be readily estimated using maximum likelihood even given the limited computing power of the day. They recommend the use of individual survey data for parameter estimation due to the bias introduced by aggregation procedures in non-linear models.

Subsequently, McFadden (1978) generalized his theory by introducing a class of models whose random utility components follow an extreme value distribution, and showed that models belonging to this class that meet a small number of restrictions are consistent with a random utility maximization framework. He then showed that the standard MNL model is part of this Generalized Extreme Value (GEV) family, such that it too is consistent with utility maximization and can be used to model rational individual decision making.

Progress in the development of travel demand models since the 1970s has been limited to refinements of the original framework. The original GEV models have been expanded to incorporate nested decisions, and more recently individual preference heterogeneity through the use of random coefficients. McFadden and Train (2000) present the useful result that the estimated choice probabilities from any MNL specification compatible with random utility maximization can be approximated by a corresponding MNL model with random coefficients, which they refer to as the mixed-MNL model. They propose two methods for estimating such a model which are readily available in modern statistical packages: Maximum Simulated Likelihood, and Simulated Moments. Preliminary research indicates that these less restrictive formulations provide improvements in performance over the basic MNL specification.

2.2 Data and Applications

Revealed preferences data on real choice behaviour has been most commonly used in transport-related applications of the MNL model. This type of data provides researchers with information about the chosen mode, often including wait, travel and walk times, as well as a few socioeconomic characteristics, but not about the alternatives which were not selected. The conventional method used to redress this selection bias is to infer the values for the alternatives using network data on routes, speed limits and congestion. The expense associated with collecting revealed preference data through surveys has limited datasets to relatively small sample sizes and to narrow geographic regions.

Individual trip surveys administered in the San Francisco Bay area over the past 30 years have provided researchers with a rich source of data for estimating demand curves for transit demand at the individual level. One early survey was conducted by McFadden's team of researchers at U.C. Berkeley to forecast the patronage of the planned Bay Area Rapid-Transit project. They collected information suited to the specification of the behavioural MNL model to predict mode shares and estimate the value of time spent on each mode, as well as work time, walk time and wait time.

When ridership data for the completed project became available several years later, the team's mode share predictions were found to be remarkably accurate (McFadden 1978).

The San Fransisco data has also been used in the urban planning literature. Cervero and Kockelman (1997) use the MNL model to examine the effectiveness of urban land use policies on reducing vehicle trip rates and motorized travel distances. They find that while the built environment does not seem to affect work-related travel, non-work trips are significantly affected.

Civil engineers originally expressed skepticism about the validity of disaggregated models, but have since adopted the MNL model and made contributions to its improvement. Using the result shown in McFadden and Train (2000), Bhat (2000) estimated a mixed-MNL model of mode choice that incorporates heterogeneity in individual utility functions. He bemoans scholars' tendency to ignore the possibility of individual taste variations caused by omitted variables, and allows all parameters to vary over individuals. Using time-series survey data on a sample of commuters in the San Francisco Bay region, he employed a Monte Carlo simulation to estimate a Random-Coefficients Logit model using a simulated log-likelihood technique. He then compared his results to the standard MNL model and found that there were significant differences in the estimated sensitivity to level-of-service variables. He was forced to impose a probability distribution on the random coefficients, and calls for further research to make the estimation procedure even more non-parametric.

Since the 1980s, another source of data which has become popular is the survey of stated preferences, commonly used in psychology and market research. Surveys are given to a sample of individuals in which they must select their preferred mode of travel given hypothetical situations about alternatives including prices, wait times, system efficiency, and comfort. The collection of such data is relatively inexpensive, but leads to the question of whether a stated choice actually translates into real outcomes. Latent variable models which allow for the integration of revealed and stated preference data were developed by McFadden, Ben-Akiva and Morikawa (Ben-Akiva *et al.* 1994), and their framework has led to a number of applied studies.

Due to the cost of surveying, empirical applications of McFadden's framework has largely been limited to studies from limited surveys. Small and Verhoef (2007) give examples showing that the method has been used by transit agencies in North America and Europe to carry out feasibility studies, but these are seldom published in the public domain.

2.3 Canadian studies

No survey data of the style collected in the San Francisco Bay exists for Canadian cities, and this has restricted the use of MNL estimation for urban travel demand analysis. Wilson *et al.* (1990) examine data from the Canadian Travel Survey, constructed by Statistics Canada in 1985, to analyze inter-city travel demand in a disaggregate MNL framework. The authors conclude that the data will not offer a good base for disaggregated analysis unless higher resolution data is focused on major city-pairs.

Most Canadian research on urban transportation utilizes Census data due to its accessibility. Danyluk and Ley (2007) used time-series Census data to examine the link between gentrification and commuting mode choice. Their analysis is limited to the calculation of correlations. They did find some evidence that gentrifiers - skilled professionals living in formerly low-income urban neighborhoods - have an increased tendency to walk or bicycle to work, even when controlling for distance to downtown. Some heterogeneity was found across cities, probably due to infrastructure differences or climate. In all cities, gentrifiers are found to use less public transit, and in Toronto, where gentrifiers are more evenly spread out over the CMA, they actually have a greater tendency to drive to work.

A Statistics Canada report by Heinsz & LaRochelle-Côté (2005) examined commuting behaviour using data from the 2001 Census of Population. They tabulate various dimensions of the socioeconomic characteristics of commuters, but do not make statistical inferences about their relative significance. Many of their observations, although intuitive, provide motivation for the selection of relevant variables in section 5.1.2. Since they have access to uncensored micro-level information, Statistics Canada are in a position to estimate a disaggregated model without making the simplifying assumptions required in this paper. So far, econometric estimation of urban travel demand has not been attempted by the national agency.

3 Commuting in Canada's Glocal Cities

This section describes the commuting infrastructure and characteristics of commuters in Montreal, Vancouver and Toronto. Summary statistics are calculated from the 2001 Census of Population public-use micro-data file for each Census Metropolitan Area (CMA) (Statistics Canada 2006a). This data provides a representative sample of the population without aggregation.

Some resolution is suppressed in order to preserve the anonymity of survey respondents. Commuting distance is reported categorically by 5 kilometre intervals, with the final category labeled "30km and over". I construct a continuous variable using the midpoint of each category and 35km for the open category. No information about residential location is provided beyond the city level, making it impossible to observe neighbourhood characteristics or access to public infrastructure. This data is usually part of the mode-specific variables in the estimation of McFadden's mode choice model. A lack of access to this information will limit the scope of questions this paper can address, and requires a particular framework to avoid misspecification error due to omitted variables.

Income data is censored above at $200,000 and below at -$50,000. The variable includes all sources of reported income, including investments, farm revenues, etc. Capital losses create the possibility of a negative value. An immigrant is defined as a respondent who was not born in Canada and who is either a naturalized citizen, permanent resident, or working visitor in the country. The number of children is censored at a maximum of 2, so the variable is effectively categorical with options "No Children", "One child", and "More than one child". Some variables of interest vary between cities, so descriptive statistics are presented by city to provide the reader with a higher resolution.

3.1 Toronto

Canada's largest Census Metropolitan Area covers an area of nearly 6,000 square kilometres and is home to over 5 million people (Statistics Canada 2006). The 2001 Census public-use micro-data file of individuals residing in the Toronto CMA consists of 125,671 unique observations, of which 50,894 are complete for all variables relevant to the commuting decision.

The city's public transit network, operated by the Toronto Transit Commission, carries nearly 1.5 million passengers on an average work day (TTC 2008). The backbone of the system is a subway/RT network whose four lines cover 68 kilometres and provide a link between suburban commuter trains and the downtown core. The city's downtown core receives frequent service by a network of street-cars who run on tracks in dedicated lanes, and can thus run reliably through peak hour gridlock. A network of GO-trains carries commuters from surrounding suburbs to connecting stations, where the remaining trip to the downtown core can be made via subway or diesel bus. A glance at the crowded park-and-ride lots surrounding GO-train stations along Highway 401 suggest that multi-mode transportation is a popular means of commuting in Toronto, but the Census questionnaire does not provide sufficient information to confirm such an observation. The east-west road corridor is anchored by the heavily congested Highway 401, and the recently constructed toll Highway 407. While heavy congestion occurs on many of Toronto's highways, the city's flat and homogeneous topography does not create major physical bottlenecks.

The average commuter in Toronto travels farther than those in Montreal and

Table 1: Descriptive statistics - Commuters in Toronto CMA

Variable	Mean	Std Dev	Range	
			Min	Max
Age	37.42	12.42	15	85
Income ($)	38,026.75	34,731.32	-49,795	200,000
Distance (km)	12.46	9.94	2.5	35
Married	0.613	-	0	1
Immigrant	0.493	-	0	1
Single-Occupancy Driver	0.647	-	0	1
Children	0.611	0.627	0	2
	Mutually Defined Observations			50,894

Vancouver, an indication of lower population density and job centralization observed in Toronto by Heinsz & LaRochelle-Côté (2005). The city also has the highest proportion of first-generation immigrants. Nearly half of all commuters in the Toronto sample are first-generation immigrants, higher than Vancouver's 41.8%.

The TTC boasts that the city has one of the most heavily used transit systems in North America in per capita terms (TTC 2008). In its recent report on the attractiveness of Canadian cities, the Conference Board of Canada found that Toronto was third in North America in its share of commuters using modes of transportation other than the automobile to get to work (Conference Board 2007). They place Toronto, at 29%, slightly behind Montreal's 30% in their ranking, but do not include carpooling as an alternative mode of transportation. Toronto has the lowest share of single-occupancy drivers in the sample at below 65%.

3.2 Montreal

The CMA of Montreal has a population of 3.6 million residing on a land area of over 4,000 square kilometres. The dataset provides 34,977 complete records for individual commuters in which all relevant variables are mutually defined.

Public transportation is administered by the *Société de Transport de Montréal* (STM) and consists of a multi-modal system including suburban rail lines that lead commuters directly into the downtown core without requiring a transfer. Residents

Table 2: Descriptive statistics - Commuters in Montreal CMA

Variable	Mean	Std Dev	Range	
			Min	Max
Age	37.77	12.32	15	85
Income ($)	33,732.48	29,116.56	-45,370	200,000
Commuting Distance (km)	10.87	8.81	2.5	35
Married	0.462	-	0	1
Immigrant	0.195	-	0	1
Single-Occupancy Driver	0.654	-	0	1
Children	0.594	0.616	0	1
	Mutually Defined Observations			34,977

of the island of Montreal are served by the *Métro* subway network, whose rubber-tired cars run on tracks spanning over 65 kilometres. A recent expansion of the its Orange line brought service to the populous suburb of Laval, giving commuters the opportunity to park-and-ride into downtown in order to avoid traffic and parking difficulties. Since a dual-mode commute does not appear as an option on the Census questionnaire, it is unclear how such users would report their commuting mode.

Montreal's island geography makes the road network prone to bottlenecks on the three principal bridges spanning the St Lawrence River. Travelers on large multi-lane highways such as Highways 20 and 40 face blockages as they join constrained highways running below grade-level into the downtown core such as Highways 10 and 15.

The share of married individuals in the sample is substantially lower than is reported in Montreal and Toronto, and the share of first-generation immigrants is nearly half the share observed in the other two cities. Commuters in the Montreal CMA tend to travel a shorter distance than those in Toronto, and nearly 35% of them do so using a mode of transportation other than the single-occupancy automobile, only slightly lower than the share observed in Toronto.

3.3 Vancouver

The CMA of Vancouver is Canada's third-largest urban region, with a population of over 2 million on a land area of nearly 3,000 square kilometres. The dataset provides

Table 3: Descriptive statistics - Commuters in Vancouver CMA

Variable	Mean	Std Dev	Range	
			Min	Max
Age	38.03	12.67	15	85
Income ($)	35,467.90	31,467.46	-30,000	200,000
Commuting Distance (km)	10.58	8.68	2.5	35
Married	0.619	-	0	1
Immigrant	0.418	-	0	1
Single-Occupancy Driver	0.717	-	0	1
Children	0.558	0.61	0	2
	Mutually Defined Observations			19,002

19,002 individual records for which all variables of interest are mutually defined.

Both public transportation and the network of major roads in the Lower Mainland are administered by the South Coast British Columbia Transportation Authority, better known as Translink, a private contractor regulated by the provincial and municipal governments.

While an above-ground Skytrain provides reliable service to the suburbs of Burnaby, New Westminster and Surrey, a number of important suburbs such as Richmond, Delta and North Vancouver have no access to rapid-transit and face slow diesel-bus commutes. A new rapid-transit line, scheduled to commence operations in 2010, is being constructed to redress the lack of service on the important Richmond-Airport-Downtown corridor. The West Coast Express commuter trains provide service from the outlying eastern suburbs to the downtown core.

The city's coastal geography, which features a number of bays and branches of the Fraser River delta, mean that transportation between communities often requires the use of bridges. While this causes serious congestion problems for automobiles, public transit is not equipped to bypass these bottlenecks and thus does not provide an advantage for commuters coming downtown from outlying communities. The Pacific Gateway project, part of which aims to add road capacity to certain key bridges, has been the subject of heated debate since it was announced in 2006.

The sample is comprised of 41.8% first-generation immigrants, double that found in Montreal but slightly lower than Toronto's share. Vancouver reports the highest share of commuters using single-occupancy vehicles in our study, but the near 30% share of commuters using other modes of transport is still relatively high relative to a larger sample of North American cities (Conference Board 2007). The distribution of income among Vancouver commuters is similar to the rest of our sample, the mean income being between that of Montreal and lower than Toronto.

4 A behavioural model of commuting mode choice

The construction of the model follows the original procedure for discrete choice modeling in urban travel demand analysis developed by McFadden & Domencich (1975). The comprehensive treatments of the subject by Ben-Akiva & Lerman (1985) and Washington *et al.* (2003) include a number of clarifications and improvements of the original methodology. Interested readers are encouraged to consult these texts for a more detailed exposition.

4.1 Defining the partial equilibrium

The choice of commuting mode is linked to all other consumption decisions, but most closely linked to the related decisions of where to live, how much to work, whether to own a car, and at what time to commute. Estimating a general equilibrium model which treats all these decisions as endogenous approaches intractability, and requires a scope of data that is not readily available at sufficient resolution. In applied studies, the choice of equilibrium framework employed is often motivated by the availability of data. To make the problem manageable, it is assumed that an individual factors his decision into smaller, more manageable problems, each of which depends only on characteristics of close substitute goods (Phlips 1974). The choice of commuting mode can be thought of as part of a multi-step decision making process, represented in Figure 1.

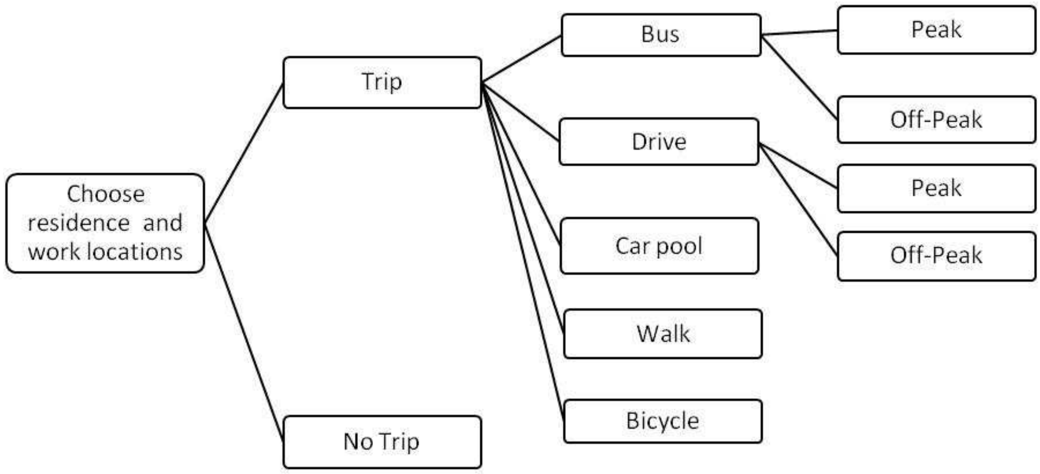

Figure 1: The Commuter's Decision Making Structure

The top branch in the decision tree is assumed to be predetermined in the equilibrium process of residential choice presented in Kraus (2006) and O'Sullivan (2007). The *monocentric city* model makes the simplifying assumption that all jobs are located in one - or a small number of - job centers. Households then choose their residential location to maximize their utility subject to a budget constraint in a trade-off between lower housing prices and higher commuting costs (O'Sullivan 2007, 155). Commuting distance can thus be equated with distance from the downtown core and is predetermined with respect to the mode choice decision. This important assumption allows us to estimate the McFadden model in the absence of data on mode specific characteristics, as will be discussed in section 5.1.1. Urban research confirms that the majority of jobs in Canadian cities are concentrated around the central business district (CBD) and the city's airport (Moos and Skaburskis 2008).

While some studies have attempted to endogenize the trip frequency and trip scheduling decisions (Small 1982), such an approach requires data of a very high resolution - ideally at the individual-trip level - which is usually limited to small surveys. None dealing with Canadian cities is publicly available. The partial equilibrium approach assumes that labour market outcomes have been made and that the mode-choice decision will not affect an individual's decision of whether to work, or how much. This restriction may be interpreted as an upper bound on the ratio of the cost of commuting to the wage, such that an individual's demand for commuter trips is perfectly inelastic within a reasonable range of prices. Finally, I take the decision of when to commute to be exogenously determined[1].

4.2 The Random Utility Model

Each individual has the same choice set of available transport modes for their daily commute, which is denoted as $J = \{bus, drive, walk, bike, carpool\}$. Each individual i has an ordinal utility function conditional on the choice of transport mode $j \in J$:

$$U_{ij} = U_i(\boldsymbol{\gamma}_i, \boldsymbol{\phi}_j | q_j = 1);$$

where $\boldsymbol{\gamma}_i$ is a vector of relevant socio-economic characteristics of individual i, and $\boldsymbol{\phi}_j$ is a vector of relevant characteristics of mode j, to be specified in section 5.1.1. Consumers select their optimal mode choice q_j^* by maximizing utility subject to their budget constraint. Since the mode-choice problem is discrete rather than continuous, possible choices must be restricted to binary, mutually exhaustive values

$$\forall j \in J: \; q_j = \begin{cases} 1 & \text{if mode } j \text{ is chosen} \\ 0 & \text{otherwise.} \end{cases}$$

Since only one mode is chosen,

$$q_i q_j = 0; \quad \forall i \neq j, i \in J.$$

These constraints on the discrete model are equivalent to imposing a corner solution on the consumer's maximization problem subject to a bounded continuous choice variable, and such a problem cannot be solved by differentiation (Ben-Akiva and Lerman 1985, 44). Instead, the solution is characterized as a system of inequalities. The individual evaluates the utility $U_{ij}(\boldsymbol{\gamma}_i, \boldsymbol{\phi}_j)$ for all $j \in J$ available within their budget constraint and chooses mode a when it offers a higher level of utility than all other

[1] Working hours are usually determined by the employer.

available modes. The indirect utility function V_{ia} corresponding to mode choice $a \in J$ is defined as a function of income I_i:

$$V_{ia}(\boldsymbol{\gamma}_i, \boldsymbol{\phi}_j, I_i) = U_{ia} \geq \max_J\{U_{ij}\} \text{ s.t. } \sum_J p_j q_j \leq I_i. \tag{1}$$

where p_j is the financial price of commuting using mode j. The utility function used thus far is deterministic by construction, since it is assumed that each individual has access to all information relevant to their decision. Some of these preference variables are not measurable by the researcher, however, and this unobserved heterogeneity among individuals is incorporated by treating the utility function as a random variable. The simple random utility specification proposed by Manski (1977) is adopted:

$$\tilde{V}_{ij} = V_{ij} + \varepsilon_{ij}; \tag{2}$$

where the random component ε_{ij} has $E[\varepsilon_{ij}] = 0$, such that the expectation of the random utility term $\tilde{V}_{ij}$ is equal to the systematic term V_{ij}. Parameters of the indirect utility function are then estimated within a random utility maximization (RUM) framework. When mode a is observed to have been chosen by an individual i, the following choice-probability relationship holds:

$$\begin{aligned} P_a(i) &= \Pr(a|\boldsymbol{\gamma}_i, \boldsymbol{\phi}) \\ &= \Pr(\tilde{V_{ia}} > \tilde{V_{ij}}; \forall j \neq a, j \in J) \\ &= \Pr(V_{ia} + \varepsilon_{ia} > V_{ij} + \varepsilon_{ij}; \forall j \neq a, j \in J) \\ &= \Pr(\varepsilon_{ij} - \varepsilon_{ia} < V_{ia} - V_{ij}; \forall j \neq a, j \in J) \\ &= F(V_{ia} - \max\{V_{ij}, j \neq a\}). \end{aligned} \tag{3}$$

The final line specifies the general utility-difference choice model, where $F()$ is the joint cumulative distribution function of the difference $\varepsilon_{ij} - \varepsilon_{ia}$. Since the V_{ia} terms are non-random, all that is required to complete the model specification is an assumption about the distribution of the difference in the random utility components.

The choice of $F()$ is restricted to functions which bound the choice probability between 0 and 1. Three commonly used functional forms that satisfy the unit condition are the cumulative normal, yielding the *probit* model, the logistic distribution, yielding the *logit* model, and the Cauchy distribution, yielding the *arctan* model. Each of these distributions yields very similar predictions except for values very close to 0 and 1. The Cauchy distribution has thick tails and is ill behaved for values on the tails of the distribution. Both the logistic and normal distributions yield very similar results, but the cumulative normal cannot be expressed in closed-form and thus requires solving an integral numerically in order to make predictions. Consequently, the logistic distribution is adopted to ease further manipulations.

A specification must be chosen for the distribution of the individual random utility components ε_{ij} that will yield a logistic joint cumulative distribution. Domencich and McFadden (1975, 63) present a proof which shows that if random variables ε_{ij} are independent and follow an extreme value distribution[2], then it follows that an individual's choice probability for a given outcome has the desired logistic distribution:

$$\begin{aligned} P_a(i) &= \Pr(V_{ia} + \varepsilon_{ia} > V_{ij} + \varepsilon_{ij}, \forall j \in M) \\ &= \frac{e^{\mu V_{ia} - \alpha_a}}{\sum e^{\mu V_{ij} - \alpha_j}} \\ &= \frac{e^{V_{ia}}}{\sum e^{V_{ij}}}. \end{aligned} \tag{4}$$

The Weibull distribution has variance parameter μ and scale parameter α_j. Since these parameters cannot be identified when utility is linear in its parameters, the usual practice is to consider them as mode-specific characteristics by normalizing to one and zero, respectively (Domencich and McFadden 1975, 65). This procedure involves no loss of generality and implicitly considers the parameters embedded in the parameters defining V_{ij}.

The specification given in equation (4) is known as the multinomial logit (MNL)

[2]Two types of extreme value distributions are more commonly known as the Gumbel and Weibull distributions.

model, and is general to any number of discrete alternatives in the set J. The consistency of the model depends on the assumption of the *independence of irrelevant alternatives* (IIA), which states that the odds ratio between any two modes $\frac{P_a}{P_b}$ is independent of all other alternatives. This implies that the choice probabilities for all modes have uniform cross-elasticity of demand with respect to an attribute of a new alternative.

The worrisome implications of this assumption was first pointed out by (Debreu 1960). Consider a mode-choice situation where a new alternatives is introduced with unobserved characteristics similar to those of an existing alternative, but unlike those of other modes[3]. For instance, a light-rail line is introduced and competes with an articulated bus service, both of which are subject to unmeasured cramped conditions at peak hours. According to the IIA assumption, both car and bus service see a uniform decrease in their choice probabilities. It is more likely that the bus service experiences a larger decrease in mode share than the automobile, since the new rail service is viewed by consumers as being very similar to the bus (Kanafani 1983).

The estimation of a MNL model should be preceded by a simple test to validate the IIA assumption, as proposed in McFadden (2000). In this paper I am interested in explaining an individual's choice of commuting in a single-occupancy vehicle versus the choice of using any other mode. Rather than considering all modes as separate possible outcomes, and estimating each of their choice probabilities, the general MNL model in equation (4) is restricted to the following binomial specification:

$$\begin{aligned} P_a(i) &= \Pr(\varepsilon_{ia} - \varepsilon_{ij} > V_{ia} - V_{ij}) \\ &= \frac{e^{V_{ia}}}{e^{V_{ia} + e^{V_{ij}}}} \\ &= \frac{1}{1 + e^{V_{ia} - V_{ij}}}; \end{aligned} \tag{5}$$

where mode a is driving in a single-occupancy vehicle, and mode j is the best mode

[3]Debreu's example was even more stark than the one I present here: that of a red bus and a blue bus.

for individual i in the set of alternatives $A = \{bus, walk, bike, carpool\} \subset J$. The IIA assumption is not required in this binomial specification, since the probability of a mode being chosen depends only on the difference in utility between it and the next best mode. The introduction of a new mode will not affect the choice probability if it possesses inferior characteristics to an existing alternative (Washington, Karlaftis and Mannering 2003, 274).

5 Econometric Specification & Estimation

5.1 Specifying the Utility Function

A generic functional form for the mode-conditional utility function has been used so far. In order to facilitate estimation, I assume that indirect utility is linear in its parameters[4]:

$$V_{ij} = V(\boldsymbol{\gamma}_i, \boldsymbol{\phi}_j, I_i) = \boldsymbol{\beta}_1 \boldsymbol{\phi}_j + \boldsymbol{\beta}_{2j} \boldsymbol{\gamma}_i; \tag{6}$$

where $\boldsymbol{\beta}_1$ is a vector of parameters with the same dimensions as the vector of mode characteristics $\boldsymbol{\phi}_j$, and $\boldsymbol{\beta}_{2j}$ is a vector of parameters specific to mode j, with same dimensions as the vector of individual characteristics $\boldsymbol{\gamma}_i$. While parameters for mode characteristics are restricted to equality between modes, parameters for socioeconomic characteristics are allowed to vary by mode. While the additively separable form may seem restrictive, note that the function must be linear in the parameters to be estimated, but need not be linear in its explanatory variables. Useful transformations will thus be possible. The following sections will specify the components included in vectors $\boldsymbol{\phi}_j$ and $\boldsymbol{\gamma}_i$.

5.1.1 Commuting distance as a proxy for mode characteristics

In McFadden's original model, the vector of mode characteristics $\boldsymbol{\phi}_j$ includes the individual's travel time and financial cost incurred by using each of the available

[4]This is the most common parametric specification used in the literature (Small and Verhoef 2007).

modes. Since the location of individual households is not reported in public use census data, values for these characteristics are not available. Estimating the model without these variables, and without accounting for differences in mode service levels, would create inconsistent results due to omitted variables. This lack of data has precluded Canadian applications of the McFadden model.

Recall the *monocentric city* assumption adopted in section 4.1, which states that all jobs are located in the CBD. It is thus possible to construct a proxy variable to capture the difference in infrastructure availability between modes without knowing geographic location. The 2001 Census asks respondents to provide the distance they commute in kilometers. The assumption about job concentrations implies that commuting distance should be equivalent to distance from the urban core. Distance from the CBD, in turn, is a proxy for service quality between modes. In order to effectively use the commuting distance variable, it is necessary to guess the functional form of the relationship between distance to CBD and difference in service quality across modes.

The usual assumption made in urban economics models is that individuals face transport costs that are linear in distance (O'Sullivan 2007). A more careful consideration suggests that the relationship is unlikely to be linear and with a zero intercept. While the cost of fuel used while driving is roughly linear in travel distance, the financial cost of travel on public transit is often invariant to distance. Demand estimates from Small & Verhoef (2007) suggest that time spent waiting for transit is perceived to be costlier than time spent riding the bus, such that time costs for bus travel are concave in distance commuted. Conversely, increased wait times for transit in outlying areas imply a convex time cost-distance relationship. For drivers, road congestion near downtown makes travel near the urban core costlier than the the same distance traveled in the outer perimeters of town, but the functional form depends on the structure of the road network and traffic volumes.

While the functional form for the relationship between cost and distance for each mode is difficult to establish, the explanatory variable of interest for the utility difference model is the divergence between the cost of driving and other modes, whose functional form is even more difficult to hypothesize. Since we do not have compelling evidence in defense of a specific functional form for the relationship between the cost difference and commuting distance, second and third-order transformations of the distance term are included in an attempt to capture any potential non-linearity.

There is debate about how to incorporate interactions between income and costs into the utility specification. To capture the idea that high-income earners will be less influenced by differences in financial cost between modes, authors include a term equal to the financial cost divided by the wage. The recognition that high-income earners tend to have higher time costs prompt the inclusion of a term equal to the product of wage and travel time (Train and McFadden 1978). Unfortunately, data limitations do not permit the inclusion of such terms into the model.

5.1.2 Socioeconomic characteristics

The vector of personal characteristics $\boldsymbol{\gamma}_i$ includes variables which have been identified in previous research as possible indicators of a propensity towards a particular mode. An overview of descriptive tabulations of variables in relation to mode choice in Canadian cities is provided in Heinsz & Larochelle-Côté (2005).

Personal income provides a proxy for an individual's valuation of time, where higher income earners are expected to favor a faster mode of transportation than low income earners. It also provides a measure of access to more expensive modes through the budget constraint. Marital status and the number of children in the individual's family are thought to influence an individual's neighborhood choice, and thus their access to transit. Married couples with children tend to reside in prefer quieter areas, which have inferior access to transit even when distance from the CBD is controlled for. They are also likely to have a higher valuation for time than singles, so are expected to favor a faster mode.

Age and gender may affect an individual's preference for perceived security and convenience. Immigrant status is thought to affect an individual's unobserved socialized preferences towards different modes. For instance, immigrants arrive in Canada from countries in which public transportation is pervasive throughout society, and thus perceive the use of buses differently than a native Canadian socialized in a car-centric society.

Car ownership is equivalent to a cost associated with automobile commuting, where lease payments are equivalent to the cost of a public transit pass, and is thus considered a mode-specific variable. These costs are not included directly as determinants of mode choice since (i) data of a sufficient resolution is not available to implement a satisfactory treatment, and (ii) they are very similar across households and cities within Canada and thus are unlikely to add much explanatory power. Car ownership has been modeled as an exogenous determining factor in the mode choice decision, as well as in a structured logit framework on sequential choice (Train 1980). As discussed in Train's paper, the inclusion of such a variable in regression analysis is problematic because it is likely to be endogenous to the mode choice decision - individuals wanting to commute by car must first purchase a vehicle.

City fixed effects are included to account for infrastructure differences between the cities, and city-level omitted variables such as climate.

5.2 Econometric Specification

Substituting the additive utility function given in equation (6) into the general binomial logit model specified in equation (5), the econometric model to be estimated is

$$\begin{aligned} P_a(i) &= \frac{1}{1+\exp[(\boldsymbol{\beta}_1\boldsymbol{\phi}_a + \boldsymbol{\beta}_{2a}\boldsymbol{\gamma}_i) - (\boldsymbol{\beta}_1\boldsymbol{\phi}_j + \boldsymbol{\beta}_{2j}\boldsymbol{\gamma}_i)]} \\ &= \frac{1}{1+\exp[\boldsymbol{\beta}_1(\boldsymbol{\phi}_a - \boldsymbol{\phi}_j) + (\boldsymbol{\beta}_{2a} - \boldsymbol{\beta}_{2j})\boldsymbol{\gamma}_i]} \\ &= \frac{1}{1+\exp[\boldsymbol{\beta}(\boldsymbol{\phi}_a - \boldsymbol{\phi}_j) + \boldsymbol{\delta}\boldsymbol{\gamma}_i]}. \end{aligned} \tag{7}$$

Rewriting the utility difference makes the interpretation of the parameters being estimated explicit. Each parameter in the vector $\boldsymbol{\beta}$ measures the effect of the difference in a mode characteristic on the probability of choosing mode a over j. Specifically, the parameter measures the propensity to choose single-occupancy driving over alternative modes as the commuting distance is increased and service quality affected. The vector of parameters $\boldsymbol{\delta}$ estimates the diverging effects of a socio-economic characteristic on the probability of choosing mode a over j. The parameters are estimated by maximizing the following log-likelihood function over the sample of size I:

$$\ell(\boldsymbol{\delta}, \boldsymbol{\beta}) = \sum_{i=1}^{I} [q_{ai} \log P_a(i) + (1 - q_{ai}) \log(1 - P_a(i))]; \tag{8}$$

where $P_a(i)$ is given by equation (7), and q_{ai} is the indicator variable for individual i making choice a.

5.3 Specification Tests for Parameter Consistency Across Cities

The specification of the model given in equation (7) assumes that explanatory variables have the same marginal effect across city groups on the choice probability. Since the model depends on an abstraction regarding urban form and commuting distance, it is of interest to relax the specification so that parameters can vary between cities.

The hypothesis that all parameter estimates are equal across city groups is tested against the alternative that they are different. The procedure, outlined in Ben-Akiva and Lerman (1985, 195), is a simple application of a likelihood-ratio test for market segmentation. Equation (7) is estimated by maximum likelihood over the whole dataset as the restricted model[5], and a competing unrestricted model is estimated separately for each city samples using the same specification, but without the inclusion of city dummy variables. The specification of the unrestricted model is:

$$P_a(i) = \frac{1}{1 + \exp\{\sum_c i_c(\boldsymbol{\beta}(\boldsymbol{\phi}_a - \boldsymbol{\phi}_j) + \boldsymbol{\delta}\boldsymbol{\gamma}_i)\}}; \tag{9}$$

[5]A full table of estimates is provided in Appendix A.4

Table 4: Likelihood Ratio Test for Parameter Equality Between City Groups

	Model	Log-Likelihood	df
H_0:	Pooled	-60,126.60	11
	Montreal	-19,763.04	9
H_1:	Vancouver	-10,419.67	9
	Toronto	-29,761.98	9

where i_c is an indicator - or dummy - variable equal to one when an observation belongs to city c and equal to zero otherwise. The models are estimated by maximizing the log-likelihood function given in equation (8), and the results presented in Table 4.

The likelihood-ratio test statistic is:

$$t = -2[\ell_R(\boldsymbol{\delta}_R, \boldsymbol{\beta}_R) - \sum_c \ell_{Uc}(\boldsymbol{\delta}_U, \boldsymbol{\beta}_U)] = 363.81;$$

where ℓ_R is the log-likelihood of the model estimated on the full sample of size I, and ℓ_{Uc} the log-likelihood for the model estimated on the sub-sample in city $c = \{Mon, Van, Tor\}$. The test statistic t is asymptotically distributed as $\chi^2(16)$, where the degrees of freedom correspond to the number of restrictions imposed. The null hypothesis that coefficients are equal across city groups is rejected at the 0.01 confidence level.

Differences in road and transit infrastructure will cause commuting distance to affect the mode-choice decision differently between cities, but an argument for city heterogeneity in the effects of socio-economic characteristics is not compelling. A third model is introduced in which only the distance commuted parameters are allowed to vary between cities. The distance variables are interacted with each city dummy variable to obtain the model:

$$P_a(i) = \frac{1}{1 + \exp\{[(\boldsymbol{\phi}_a - \boldsymbol{\phi}_j) + \sum_c [i_c] \boldsymbol{\beta}\boldsymbol{\delta}\boldsymbol{\gamma}_i]\}}.$$

Since it is not possible to identify both of the parameters in the interaction terms, the choice probability is written with a new parameter:

$$P_a(i) = \frac{1}{1 + \exp\{[(\boldsymbol{\phi}_a - \boldsymbol{\phi}_j) + \sum_c \boldsymbol{\beta}_c \boldsymbol{\gamma}_i]\}}; \tag{10}$$

where $\boldsymbol{\beta}_c$ is a vector of city-specific parameters equal to the interaction term $i_c\boldsymbol{\delta}$. A likelihood-ratio test is then carried out to test the hypothesis that model 1, the restricted model, is actually nested in model 3, the unrestricted model. The maximized value of the log-likelihood function for the unrestricted model was -60,041.35, and a test statistic is computed:

$$t = -2[\ell_R(\boldsymbol{\delta}_R, \boldsymbol{\beta}_R) - \ell_U(\boldsymbol{\delta}_U, \boldsymbol{\beta}_U)] = 170.50.$$

The likelihood-ratio test statistic t is asymptotically $\chi^2(4)$ distributed. The null hypothesis is rejected at the 0.01 level, such that Model 3 is preferred to Model 1.

Models 2 and 3 were tested against each other using a similar procedure, and the maximized log-likelihood values and appropriate degrees of freedom reported previously allow the reader to confirm that Model 2 is in fact preferred to Model 3. The evidence from Bhat (2000) suggests that unobserved individual heterogeneity is important, such that relaxing the restrictions on model parameters always improves the explanatory power of the model. However, there is no reason to suppose that socioeconomic characteristics vary by city rather than by any other category, such as income group or ethnic origin. Given the lack of theoretical motivation for the range of flexibility allowed in Model 2, the specification of Model 3 given in equation (10) is adopted.

5.4 Model Estimation

As described in section 4.2, a property of the MNL model is that the random disturbance terms are drawn from independent extreme value distributions. This ensures that the choice probability follows the logistic distribution. Estimating the model when this property is violated leads to inconsistent choice probabilities and inconsistent parameter estimates. If the use of commuting distance as a proxy for mode service levels is inappropriate or inadequate, then the omitted mode characteristic variables will cause heteroskedasticity due to model misspecification, and inconsistent estimates (Washington et al. 2003, 276).

It is desirable for standard errors to be robust to weaknesses in the monocentric city assumption, even if it is valid. White standard errors are used to control for heteroskedasticity of unknown form. Furthermore, clustered standard errors allow for the correlation of random disturbances within city groups. Under this procedure, the variance-covariance matrix is in the form of three boxes of non-zero elements centered on the principal diagonal, while covariance across city groups is assumed to be zero.

While the sign of logit estimation coefficients indicate the direction of their effect, interpretation of their magnitudes relative to the choice probabilities requires a transformation. Marginal effects are computed numerically by varying the independent variable of interest slightly around its mean value and measuring the perturbation in the predicted probability of choosing the single-occupancy automobile.

While the third-order transformation of commuting distance was found to be insignificant, the second and first-order terms were important, suggesting a quadratic approximation is a better description of the data than a simple linear relationship. To demonstrate the difference in magnitude between the effect of distance commuted between cities, the joint marginal effects are computed for the first and second-order terms. An increase of 5 kilometres in the distance between an individual's home and place of work increases the probability of selecting the automobile as the commuting mode by 6.5% in Toronto, 7.9% in Montreal, and by 9.1% in Vancouver.

All the variables in the vector $\boldsymbol{\gamma}$ are significant determinants of the choice probability. As expected, higher income earners are more likely to drive to work alone. The result supports the intuition that higher income earners (i) are more likely to be able to afford a personal car, or (ii) have a higher time valuation which makes the fastest mode more appealing. As discussed in section 5.1.1, the difference between the

Table 5: Model 3 - MNL Estimates for the Choice of Driving to Work in a Single-Occupancy Automobile

Parameter	Estimate	Robust Std. Err.	p-value	$\frac{dy}{dx}$ (at mean)
Income ($1,000)	0.00998	0.00147	0.000	0.00214
Distance (km):				
Toronto	0.06811	0.00047	0.000	0.01462
Montreal	0.07908	0.00103	0.000	0.01698
Vancouver	0.09676	0.00032	0.000	0.02077
Squared-Distance (km^2)				
Toronto	-0.00142	0.00003	0.000	-0.00031
Montreal	-0.00111	0.00001	0.000	-0.00024
Vancouver	-0.00242	0.00002	0.000	-0.00052
Age	0.02108	0.00216	0.000	0.00453
Children	0.32958	0.03077	0.000	0.07075
Immigrant Status	-0.55096	0.07381	0.000	-0.12065
Marital Status	0.20897	0.09935	0.035	0.04508
Male	0.60597	0.03458	0.000	0.12959
Montreal FE	-0.47555	0.02255	0.000	-0.1047
Toronto FE	-0.27403	0.00318	0.000	-0.05886
Constant	-1.0357	0.0545	0.000	-
			N	104,873
			Clusters	3
			Correctly Predicted	74,277
			Success Rate	70.83%

speed of the car relative to the speed of transit will vary according to distance from downtown. Despite heavy congestion on rush hour roads, it is likely that commuting by car is a faster alternative to transit for most users.

Each additional child[6] increases an individual's probability of commuting by single-occupancy automobile by 7.1%, while marriage increases it by 4.5%. It appears that individuals with children either (i) select quieter or lower density neighborhoods with poorer transit access, even controlling for commuting distance; or (ii) have a stronger preference for the faster mode because they have additional commitments outside their job.

City fixed effects are found to be important and of substantial magnitudes: Montreal commuters are nearly 10.5%, and Toronto commuters nearly 5.9% more likely to

[6]Recall that data on children is censored at a maximum of 2 per individual.

use alternative transit than those in Vancouver. These results indicate the city infrastructure's bias towards the automobile relative to public transit. Despite a temperate climate, the extensive efforts of urban planners to create a mixed land-use policy in the downtown core, and intentional avoidance of a major highway network, Vancouver is found to be most favorable to single-occupancy commuting in the sample.

Interestingly, immigrants are found to be twelve percent less likely to commute using a single-occupancy vehicle than native Canadians, even when income and other factors are controlled for. This provides evidence for the hypothesis that immigrants have been socialized with different preferences towards transportation which either discourage the use of the automobile, or, more likely, normalize the use of public transportation and the sharing of personal space. Another possible explanation is that recent immigrants tend to live in ethnic enclaves, often of higher density than the standard suburb and thus better served by transit.

5.5 Post-estimation tabulations

The effectiveness of the model is examined by calculating its success rate at predicting mode choice outcomes. If an automobile driver is predicted to have a probability of driving higher than 0.5, or a user of another mode is predicted to choose the car with a probability smaller than 0.5, then the individual's choice is considered to have been successfully predicted. According to this procedure, the model successfully predicts over 70% of the 104,873 observations in the sample.

A closer examination of the individuals correctly predicted versus those incorrectly predicted reveals a potential deficiency. The model is far more successful at predicting the choices of car drivers than users of other modes. The "wrong" mode is predicted more prevalently for individuals with lower income, who are younger, who are more likely to be single, have fewer children, and commute a shorter distance. These observations provide preliminary evidence that the model has difficulty predicting the mode choice of people living downtown, although there is insufficient data

Table 6: Comparing Samples of Predicted Values

Variable	Success	Incorrect	Difference
Single-Occupancy Commuters	0.833	0.246	0.587
Income ($)	38,990.21	29,184.45	9,805.76
Distance Commuted	12.16	10.19	1.97
Age	38.34	35.97	2.37
Gender Ratio ($\frac{m}{f}$)	0.84	1.45	-0.61
Children	0.63	0.52	0.11
Marital Status	0.59	0.49	0.1
Sample Size	74,277	30,596	

to verify this claim with precision at this stage of the analysis.

This section has estimated the effect of socioeconomic characteristics and commuting distance on the mode-choice decision, despite data limitations on mode specific characteristics. To avoid model misspecification from omitted variables, a *monocentric city* framework is assumed, and the distance commuted variable is then included to capture service differences between modes. While this method does not allow for the evaluation of a full demand curve from which elasticities for time and quality characteristics can be calculated, it is hoped that this crude approach captures sufficient variation to avoid inconsistency due to omitted variables. If the distance variable does not capture enough variation in the difference between mode characteristics, or it is incorporated with the wrong functional form, then the model is misspecified and MNL estimates are inconsistent (Washington et al. 2003, 276).

To investigate the effectiveness of the model across spatial dimensions, it is necessary to use data aggregated at the census tract level for which geographic boundaries are known.

Table 7: Panel Dimensions

	Toronto	Montreal	Vancouver	Total
1996	803	756	297	1,856
2001	924	846	385	2,155
2006	994	860	408	2,262
Total	2,721	2,462	1,090	6,273

6 Predicting Mode Shares Under Aggregation

6.1 Aggregate Data

This section uses aggregate data from the Canadian Census of Population from the years 1996, 2001 and 2006 - the only years in which information about commuting mode choice was collected. Data is reported at the census tract level. A census tract comprises between 2,500 and 8,000 individuals and have highly variable areas. While data of higher resolution is available at the census dissemination area level, it is subject to stricter confidentiality constraints which result in a large number of missing observations, with a bias towards areas with low population density.

Table 7 reports the number of census tracts available for analysis of the commuting mode choice decision. Variables reported in the census tract level are reported continuously, and descriptive statistics are provided in Table 8. Nominal gross income is reported for the calendar year previous to the census. Nominal income values are converted to real values using the national consumer price index's full basket of goods, with 2001 as the base year for the purpose of consistency with the previous estimation results. The age and gender share variables correspond to commuters only, not to the general census tract population.

Average commuting distance corresponding to the variable available in the microdata files is not provided. I proceed by assuming that jobs are located in one or a small number of centres in each city. A variable is constructed from available data to approximate commuting distance. The Statistics Canada study by Heinsz & LaRochelle-Côté (2005) found that the two most important zones of job concentration

Table 8: Summary statistics - Canadian Census of Population - Census Tracts in Vancouver, Toronto & Montreal 1996-2006

Variable	Mean	Std Dev	Min	Max
Share using car	0.6343	0.1808	0	1
Est Avg Distance (km)	11.56	6.928	0.1049	49.668
Children per family	1.1601	0.2655	0.1304	4
Share Married	0.6128	0.169	0.0769	1
Real Income (2001$)	31,494.64	15,177.26	0	281,573.50
Share of Immigrants	0.3263	0.1872	0.0053	0.8165
Average Age	37.69	4.25	26.28	63.74
Share Male	0.5296	0.0343	0.333	1
		Mutually Defined observations		6,273

in all of Montreal, Vancouver and Toronto are the area around the downtown core, and, more surprisingly, the area surrounding the airport[7], but that there has been trend toward dispersed suburban employment locations. This research rests on the use of a *dualcentric city* approach as an approximation of urban form in order to calculate unobserved commuting distances from geographical points.

The latitude and longitude coordinates of the centroid of each census tract are provided by Statistics Canada, and allow for the calculation of the distance between each tract and the city's job centres[8]. The geographic coordinates used for the CBD were taken from Danyluk and Ley (2007), and correspond to the point of peak land value for each city. Airport coordinates are from the International Aviation Transportation Authority (IATA).

The distances between the census tract centroid and each job centre in its city - the CBD and the airport - are calculated. A line-of-sight distance is computed using the Haversine formula to account for the Earth's curvature and the uneven spacing between meridians, as described in Appendix A.1. The geography of our

[7]Heinsz & LaRochelle-Côté (2005) report the share of total CMA jobs in the zones surrounding the CBD and airport, respectively, as follows: Vancouver - 20.3% and 10.4%; Montreal - 15.8% and 11.5%; Toronto - 16.9% and 14.3%

[8]It should be noted that the centroid of a non-convex shape sometimes lies outside the shape itself, such that the geographic coordinates provided could be improved by obtaining the coordinate of maximum population density within the tract

Table 9: Constructing an Estimate for Commuting Distance

	Toronto	Vancouver	Montreal	Total
Avg Reported Distance	12.46 (9.94)	10.57 (8.68)	10.87 (8.81)	11.59 (9.39)
Approx. Avg. Distance	25.49 (14.15)	21.54 (13.53)	17.83 (11.37)	22.55 (12.59)
Rescaling Factor	0.489	0.491	0.61	

cities often impedes direct travel, however, such that a better measure of commuting distance can be obtained by calculating the distance between residence and workplace over the physical road network. To do this, Navtek's comprehensive database of street networks and the ArcGIS software package were used to compute the shortest driving distance between each census tract and both the airport and CBD, as described in Appendix A.2.

The average commuting distance for each census tract was then computed using a split function in order to avoid overestimating the commuting distance for people living very close to the airport or the CBD. For census tracts within five kilometres of either job centre, the distance to the closest centre was used. For census tracts outside these two zones, a weighted average of the two distances was computed where the weights correspond to the destinations' respective job concentration shares, as reported by Heinsz & LaRochelle-Côté (2005).

To determine whether the average distance computed is indeed representative of the commuting distances reported by individual households in the 2001 Census microdata files, their values were compared. As expected, the distance computed is larger than the reported value due to the simplifying assumption that job locations which ignores the existence of numerous employers in suburban areas. The estimates are rescaled by city using the ratio of means of the average reported commuting distance to the computed average commuting distance, as shown in Table 9. This ensures that the computed commuting distances are of a comparable scale to the actual values, while preserving the heterogeneity between census tracts in their distance from the job centres. Without this rescaling, the magnitude of the coefficients from the micro model would not correspond to the constructed variable, and mode share predictions would be biased.

An important portion of commuting cost is incurred due to the time consumed by the activity. As was discussed in Section 5.1.1, the ideal variables for analysis are the difference between drive, transit, bicycling and walking times across modes for a given census tract's likely commuting route. While location data is omitted from census micro data files to maintain anonymity, computing these variables for aggregated observations using geographic coordinates is feasible. It is currently possible to calculate the driving route that will minimize drive time between two points using rough approximations for peak-period congestion, but data is not yet publicly available for computing the same problem over the transit network. The arrival of online mapping tools such as Google Maps and OpenStreetMaps has already made this data available for certain cities in the United States, and will likely do so for major Canadian cities in the near future.

6.2 Aggregation procedure

Observations in the data represent averages over the population of the tract N_t. To predict mode shares in census tracts, the behavioural MNL model specified in equation (10) is aggregated using an *average individual* procedure (Ben-Akiva and Lerman 1985, 135). A forecast of the share of total demand for mode a in census tract t can be obtained as an average of the individual choice-probabilities:

$$E[S_a(t)] = \frac{1}{N_t}\sum_{i=1}^{N_t} P_a(i|\boldsymbol{\gamma}_i, \boldsymbol{\phi}). \tag{11}$$

When predicting mode-shares for aggregated areas, however, each individual's full vector of characteristics $\boldsymbol{\gamma}_i$ cannot be observed. The amount of information required to estimate $E[S_a(t)]$ can be reduced - at the expense of efficiency - by creating T representative individuals possessing the average characteristics of the N_t commuters in census tract t. The average individual representing tract t is assumed to have the

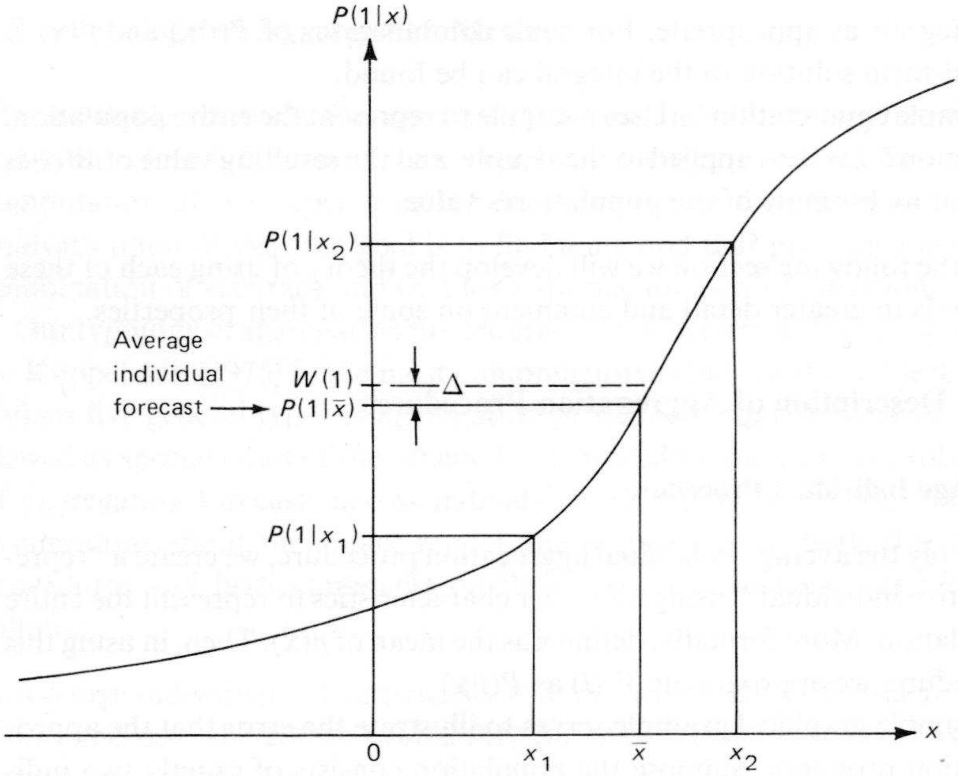

Figure 2: Prediction bias due to averaging

indirect utility specification

$$\bar{V}_{ta} = V_{ia}(\bar{\gamma}_t, \boldsymbol{\phi}_j); \quad \text{where } \bar{\gamma}_t = \frac{1}{N_t}\sum_{i=1}^{N_t} \gamma_i. \tag{12}$$

Since the logit choice probability model is non-linear, predictions using the average of individual probabilities will be biased. For values where the logistic function is convex, the prediction will be biased down. Conversely, for regions where the logistic function is concave, the averaging procedure will tend to be biased up. A graphical representation of this result is reproduced from Ben-Akiva and Lerman (1985, 136) in Figure 6.2.

To illustrate the intuition behind this bias, consider an example of two census tracts. In the first, there are a large group of low-income households as well as a small group of high-income professionals. In the second, there is a homogeneous population of middle-class commuters. The two tracts have the same average income, such that the averaging procedure predicts an identical mode share based on that parameter. However, the true share of automobile use is likely to be higher in the second tract than in the first, since individuals with income far below the sample mean will have much lower choice probability than individuals near the mean. This bias could be reduced by incorporating information about the distribution of the descriptive variables within census tracts.

While Statistics Canada defines census tracts geographically, with an aim of maintaining consistency over time, they do also make some effort to group together individuals of similar socioeconomic status (Moos and Skaburskis 2008). The application of an average individual with census tract level data can thus be thought of as a loose *classification* procedure, in which individual groups are formed on the basis of minimizing within-group variance in a certain important characteristic to reduce the loss of resolution that results from averaging over the group.

If the parameter estimates calculated in the previous section do an adequate job of describing the data, aggregate mode shares can be predicted using the averaged characteristics reported in each tract. The representative utility function given in equation (12) is substituted into equation (11) to specify the expectation of the share of individuals in census tract t using mode a as a function of the MNL estimates:

$$E[S_a(t)] \approx P_a(t|\bar{\boldsymbol{\gamma}}_t, \bar{\boldsymbol{\phi}}, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}) = \frac{1}{1 + exp(-\hat{\boldsymbol{\delta}}\bar{\boldsymbol{\gamma}}_t - \hat{\boldsymbol{\beta}}\bar{\boldsymbol{\phi}})}. \tag{13}$$

Predictions of the expected share of commuters using single-occupancy vehicles are computed using equation (13). They are then compared to their average values to measure the accuracy of predictions. The residual error u_t for tract t is computed as the difference between the predicted and observed mode share:

$$u_t = E[S_a(t)] - S_a(t). \tag{14}$$

Residuals for the entire sample have a mean of 0.05, indicating that the aggregation procedure has biased estimates in favor of single-occupancy driving. This is consistent

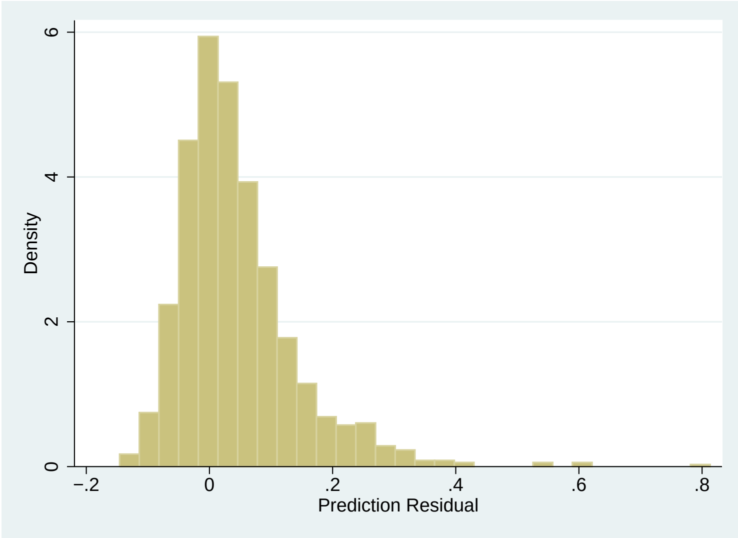

Figure 3: Histogram of Prediction Residuals for Vancouver

with the observation that most of the observed shares of commuter using automobiles lie in the concave portion of the cumulative logistic distribution, with values closer to 1. It may also be an indication that the MNL model has been misspecified: either important variables have been omitted, or the functional form of some variables has been incorrectly modeled. This bias is small and very similar across cities, and residuals for Vancouver have a mean of 0.045. A histogram showing the distribution of residuals for Vancouver is shown in Figure 3.

6.3 Spatial Analysis

Given the simplifying assumption about urban job concentrations, which assumes a homogeneous decrease in service quality with distance from CBD, it is of interest to examine how these residuals vary spatially across the city region. This will provide a pictorial overview of the relative success of the prediction procedure. The residuals were divided by city and classified by quantile groups - colour labels were then assigned to ease their identification when plotted.

- Between the 100 and 90 percentile: Red
- Between the 90 and 75 percentiles: Orange

- Between the 25 and 75 percentile: Green
- Between the 25 and 10 percentiles: Baby blue
- Between the 10 and 0 percentiles: Navy blue

Each census tract was plotted onto a map with its colour representing the relative size of its prediction residual. The map of Vancouver is included in Figure 4[9]. Upon examination of the geographic plot for Vancouver, several interesting observations can be made:

1 Car use is overpredicted for nearly all tracts in the downtown core.

2 Car use is overpredicted for three groupings of tracts lying diagonally to the south-east of the downtown core. These surround the major transport hubs of Metrotown, New Westminster and Surrey, along the city's rapid transit *Skytrain* line.

3 Car use is underpredicted for a large group of tracts located to the south of downtown in the municipality of Richmond.

4 The large majority of tracts in a band of medium density outside the CBD are well predicted.

5 Census tracts near major geographic bottlenecks are well predicted.

[9] Maps for Montreal and Toronto are included in Appendix A.3

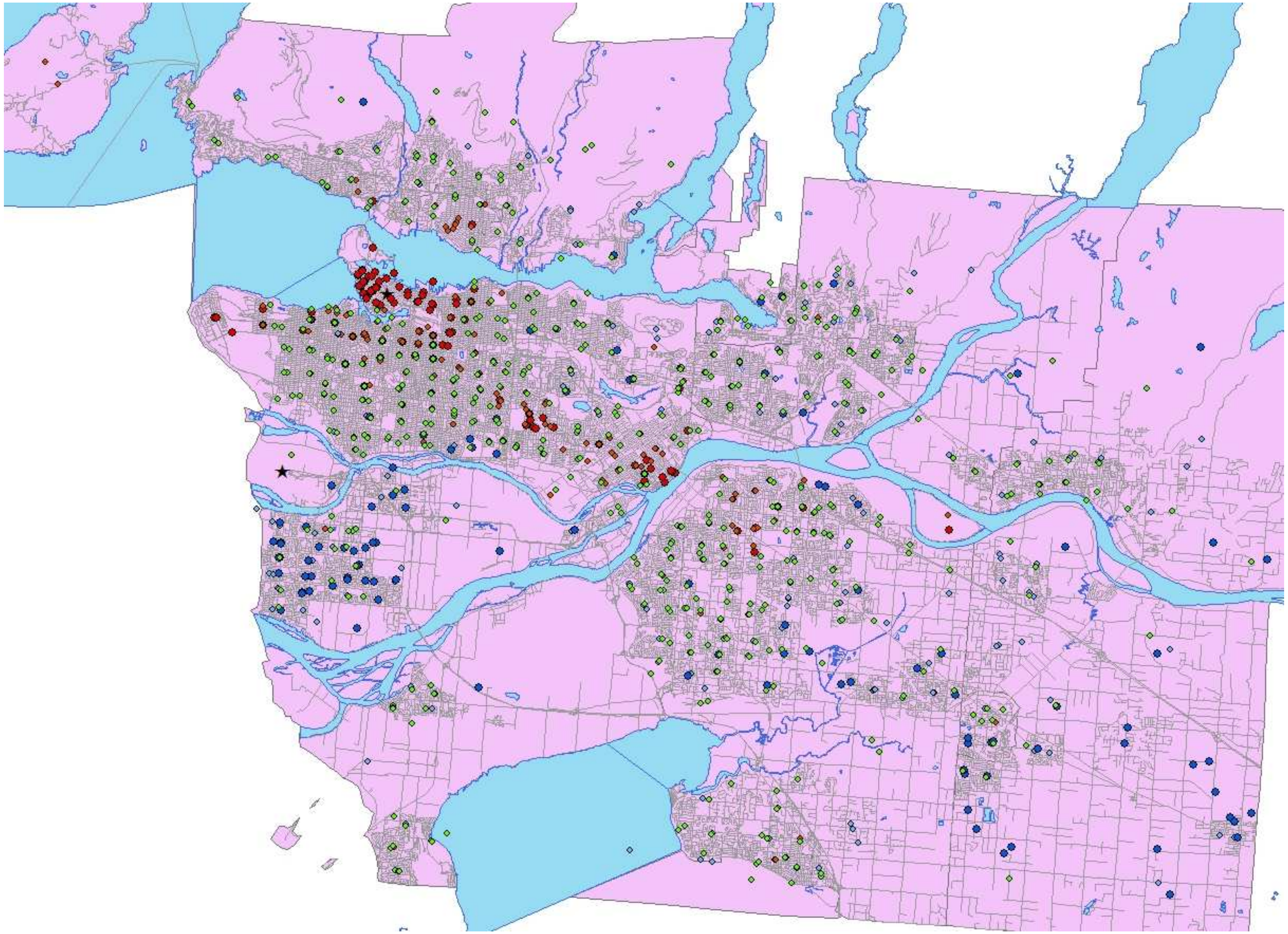

Figure 4: Mode-Share Prediction Residuals - Vancouver CMA

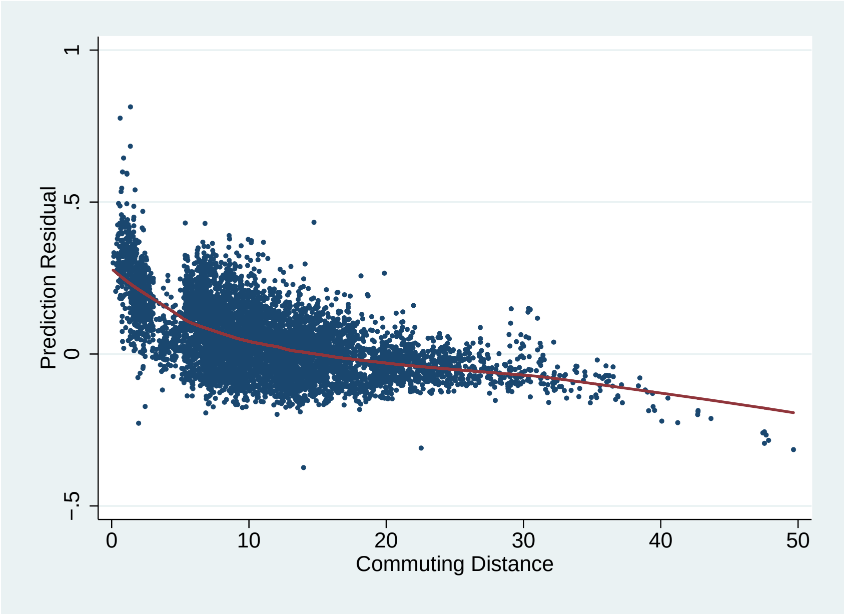

Figure 5: Prediction Residuals vs. Commuting Distance for all cities with Lowess-smoothed trend

The systematic groupings of outliers identified above signal both the success and the shortcoming of the model. The assumption of a binodal city implies that distance from downtown is homogeneously related to service quality. While this may be a satisfactory simplification, it does not capture the effects of rapid-transit lines, which offer increased service quality to neighboring residents despite their distance from the CBD. This unobserved heterogeneity explains the outliers located near Metrotown and New Westminster, and those in areas under-serviced by transit such as Richmond. However, the systematic groupings of outliers are to be expected given the information used and these intuitive results indicate that while not all variability can be explained using available data, the estimations of socioeconomic characteristics do provide sensible predictions for a wide range of observations.

The plots for Montreal and Toronto also show that car use is overpredicted in downtown plots and underpredicted in outlying areas, but the model performs well for a wide belt in the inner suburbs of all the cities. A plot of prediction residuals for census tracts in all cities against their computed commuting distance, provided in Figure 5 suggests that additional non-linearity remains between mode choice and commuting distance despite the inclusion of second and third-order terms in the parametric model. The amounts to misspecification of the functional form of commuting distance in the model, and signals the need for a non-parametric treatment of the variable.

7 Conclusion

A shortage of publicly available data on road congestion, transit network times, and individual residential locations limits the scope of policy questions which can be addressed through econometric analysis of commuting in Canada's global cities. This has precluded the use of McFadden's powerful MNL model for disaggregated commuter travel in Canada. Despite this limitation, this paper has shown that sensible predictions of mode share are possible if the researcher is willing to make a simplifying - and heroic - assumption about urban form.

In Section 4, a modification of McFadden's MNL model is specified by adopting the urban economic framework of an equilibrium city with one or two job centres. This assumption allows information about commuting distance to be related to a measure of service quality for public infrastructure. This allows us to estimate the effects of socioeconomic characteristics on the mode choice decision while controlling for mode specific characteristics which are not publicly available. Many common beliefs about mode preferences are confirmed by the estimates - namely that alternative modes are preferred by individuals with lower income, who are unmarried, are younger and have fewer children. Even when controlling for these and other important factors, immigrants are found to be 12% less likely to commute by single-occupancy vehicle than their counterparts who are born in Canada. This provides evidence for the existance of the North American "car culture" often cited in mass media. Toronto and Montreal are found to have urban infrastructure that favor the use of alterative modes substantially more so than Vancouver.

In Section 6, an aggregation procedure is presented which allows for the prediction of mode shares across census tracts. There are three likely sources of bias in the predictions: (i) the use of an averaging procedure in a non-linear model, (ii) inconsistent MNL estimates due to specification error from omitted variables, and (iii) inconsistent MNL estimates due to misspecification of the functional form for distance commuted. However, the relatively small size of this bias (less than 5%) suggests that the procedure still provides useful results for analysis of commuters in Canada's global cities.

A spatial analysis reveals that while some systematic errors have been made in the specification, the dualcentric city assumption may have captured enough variability in service access for the model to make sensible predictions despite potential inconsistency. It will become possible to resolve the model's shortcomings within the next few years as more network data is made publicly available.

References

Ben-Akiva, Moshe, and Steven Lerman (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand* (Cambridge, MA: MIT Press)

Ben-Akiva, Moshe, M. Bradley, T. Morikawa, J. Benjamin, T. Novak, H. Oppewal, and V. Rao (1994) 'Combining revealed and stated preferences data.' *Marketing Letters* 5(4), 335–350

Bhat, Chandra R. (2000) 'Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling.' *Transportation Science* 34(2), 228–238

Cervero, Robert, and Kara Kockelman (1997) 'Travel demand and the 3ds: Density, diversity, and design.' *Transportation Research D* 2(3), 199–219

Chamberlain, Robert G. (1996) 'What is the best way to calculate the great circle distance (which deliberately ignores elevation differences) between two points?' http://www.census.gov/cgibin/geo/gisfaq?Q5.1 mirrored at http://www.movable-type.co.uk/scripts/gis-faq-5.1.html

Conference Board of Canada (2007) 'City magnets: Benchmarking the attractiveness of canada's cmas'

Courchene, Thomas J. (2004) 'Social policy and the knowledge economy: New century, new paradigm.' *Options Politiques*

Danyluk, Martin, and David Ley (2007) 'Modalities of the new middle class: Ideology and behaviour in the journey to work from gentrified neighbourhoods in canada.' *Urban Studies* 44(11), 2195–2210

Davidson, Russell, and James MacKinnon (2004) *Econometric Theory and Methods* (New York: Oxford University Press)

Debreu, Gerard (1960) 'Review of luce, individual choice behavior.' *American Economic Review* 50, 186–188

Domencich, Thomas, and Daniel McFadden (1975) *Urban Travel Demand* (Amsterdam: North Holland)

Heinsz, Andrew, and Sébastien LaRochelle-Côté (2005) 'Work and commuting in census metropolitan areas 1996-2001.' *Statistics Canada Analytical Paper*

Kanafani, Adib (1983) *Transportation Demand Analysis* (New York: McGraw-Hill)

Kraus, Marvin (2006) In *A Companion to Urban Economics,* ed. Richard J. Arnott and Daniel P. McMillen (Malden, MA: Blackwell) chapter Monocentric Cities, pp. 96–108

Luce, Robert D. (1959) *Individual Choice Behavior: A Theoretical Analysis* (New York: Wiley)

Manski, Charles F. (1977) 'The structure of random utility models.' *Theory and Decision* 8(9), 229–254

McFadden, Daniel (1974) 'The measurement of urban travel demand.' *Journal of Public Economics* (3), 303–328

—— (1978) In *Transportation Economics: Selected Readings,* ed. T. Oum et al (Korea Research Foundation) chapter The Theory and Practice of Disaggregate Demand Forecasting for Various Modes of Urban Transportation

—— (2000) 'Disaggregate behavioral travel demand's rum side.' International Association of Travel Behavior Analysts; Brisbane, Australia

McFadden, Daniel, and Kenneth Train (2000) 'Mixed mnl models for discrete response.' *Journal of Applied Econometrics* (15), 447–470

Meyer, J., and M. Straszheim (1970) (Korea Research Foundation). Reprinted in:

Moos, Markus, and Andrejs Skaburskis (2008) 'The redistribution of residential property values in montreal, toronto, and vancouver: Examining neoclassical and marxist views on changing investment patterns.' *Environment and Planning A.* Advanced online publishing

Oppenheim, Norbert (1995) *Urban Travel Demand Modeling* (New York: John Wiley & Sons)

O'Sullivan, Arthur (2007) *Urban Economics* (New York: McGraw-Hill)

Phlips, L. (1974) *Applied Consumer Analysis* (Amsterdam: North Holland)

Sinnott, R.W. (1984) 'Virtues of the haversine.' *Sky and Telescope* 68(2), 159

Small, Kenneth A. (1982) 'The scheduling of consumer activities: Work trips.' *American Economic Review* 72(3), 467–479

Small, Kenneth, and Erik T. Verhoef (2007) *The Economics of Urban Transportation* (Abingdon, ON: Routledge)

Statistics Canada (2006a) '2001 census of population [canada] public use microdata file (pumf): Individuals file'

—— (2006b) '2006 census of population: Community profiles.' http://www.statcan.ca/english/census/index.cfm

—— (2008) 'Census snapshot - immigration in canada: A portrait of the foreign-born population, 2006 census.' Statistics Canada Catalogue no. 11-008-X

Thurston, L. (1927) 'A law of comparative judgment.' *Psychological Review* 34, 273–286

Toronto Transit Commission (2008) 'Ttc operating statistics 2007.' http://www.toronto.ca/ttc/pdf/operatingstatistics2007.pdf, accessed July 26 2008

Train, Kenneth (1980) 'A structured logit model of auto ownership and mode choice.' *The Review of Economic Studies* 47(2), 357–370

Train, Kenneth, and Daniel McFadden (1978) 'The goods/leisure trade-off and disaggregate work trip mode choice models.' *Transportation Research* 12, 349–353

Washington, Simon P., Matthew G. Karlaftis, and Fred L. Mannering (2003) *Statistical and Econometric Methods for Transportation Data Analysis* (Boca Raton, FL: CRC Press)

Wilson, Frank R., Sundar Damodoran, and J. David Innes (1990) 'Disaggregate mode choice models for intercity passenger travel in canada.' *Canadian Journal of Civil Engineering* 17, 184–191

A Appendix

A.1 Haversine Formula

An accurate calculation of the distance between two sets of geographic coordinates must measure the length of an arc along a spherical surface in order to account for the convergence of the meridians and curvature of the Earth's surface. While a Cartesian approach using Pythagorus' theorem provides a reasonable approximation if the points are very close together, this is unsatisfactory as the CMAs considered in this paper cover spans of over 100kms.

The Haversine Formula was introduced by astronomer Sinnott (1984) as a computationally well-behaved method of completing the task under the approximation that the Earth is spherical, and was recently recommended by Chamberlain (1996) as the preferred method for calculating the great-circle-route distance between points $x(lat_x, long_x)$ and $y(lat_y, long_y)$.

Let

$$\begin{aligned} R &= \text{radius of the Earth at } x \\ &= 6,378km - 21\sin(lat_x) \end{aligned}$$

$$\begin{aligned} \Delta lat &= lat_y - lat_x \\ \Delta long &= long_y - long_x \\ a &= \sin^2(\frac{1}{2}\Delta lat) + [\cos(lat_x)\cos(lat_y)\sin(\frac{1}{2}\Delta long)] \\ c &= 2\sin^{-1}(\min\{1, \sqrt{a}\}). \end{aligned}$$

Then the distance between x and y in kilometres is given by

$$d = R \times c.$$

A.2 GIS Methodology

The ArcGIS 9.2 software suite from ESRI, the industry standard in geographic information systems, was used for spatial analysis, and to generate the variables estimating travel distance in the aggregate data set. The following work-flow was carried out:

1 The Navtek 2006 Canadian road network and administrative boundary data were used as the basis for the construction of the city maps.

2 The North American Datum (1983) geographic coordinate system was used as the spatial reference

3 All data points were projected using a Universal Transverse Mercator transformation specific to the city's zone: Vancouver, UTM zone 11; Montreal, UTM zone 19; Toronto, UTM zone 17. The application of this projection adequately represents the curvature of the Earth in the work area, such that lengths can be accurately computed.

4 The length of each road segment in the Navteq network were computed. The calculation carried out on the projected coordinate system gives a similar result to the Haversine Formula described in Appendix A.1.

5 The set of 6,000 census tract centroids provided by Statistics Canada, and the coordinates for the CBD and airport locations were overlaid onto each city map.

6 A computationally-intensive Origin-Destination cost matrix was solved using the Network Analyst extension to compute the shortest drive distance over the assembled road network for each combination of points. Constraints for one-way streets were enforced, and U-turns were universally allowed. No weighting system was implemented to encourage the use of arterial roads.

7 The distances were merged back into Stata and rescaled as described in Section 6.1, Table 9.

A.3 Figures

The map of prediction residuals for Toronto and Montreal are provided on the following pages. Recall the legend for the colour scheme used in the graphing of census tracts, which refer to positions within the distribution of prediction residuals for the city.

- Between the 100 and 90 percentile: Red
- Between the 90 and 75 percentiles: Orange
- Between the 25 and 75 percentile: Green
- Between the 25 and 10 percentiles: Baby blue
- Between the 10 and 0 percentiles: Navy blue

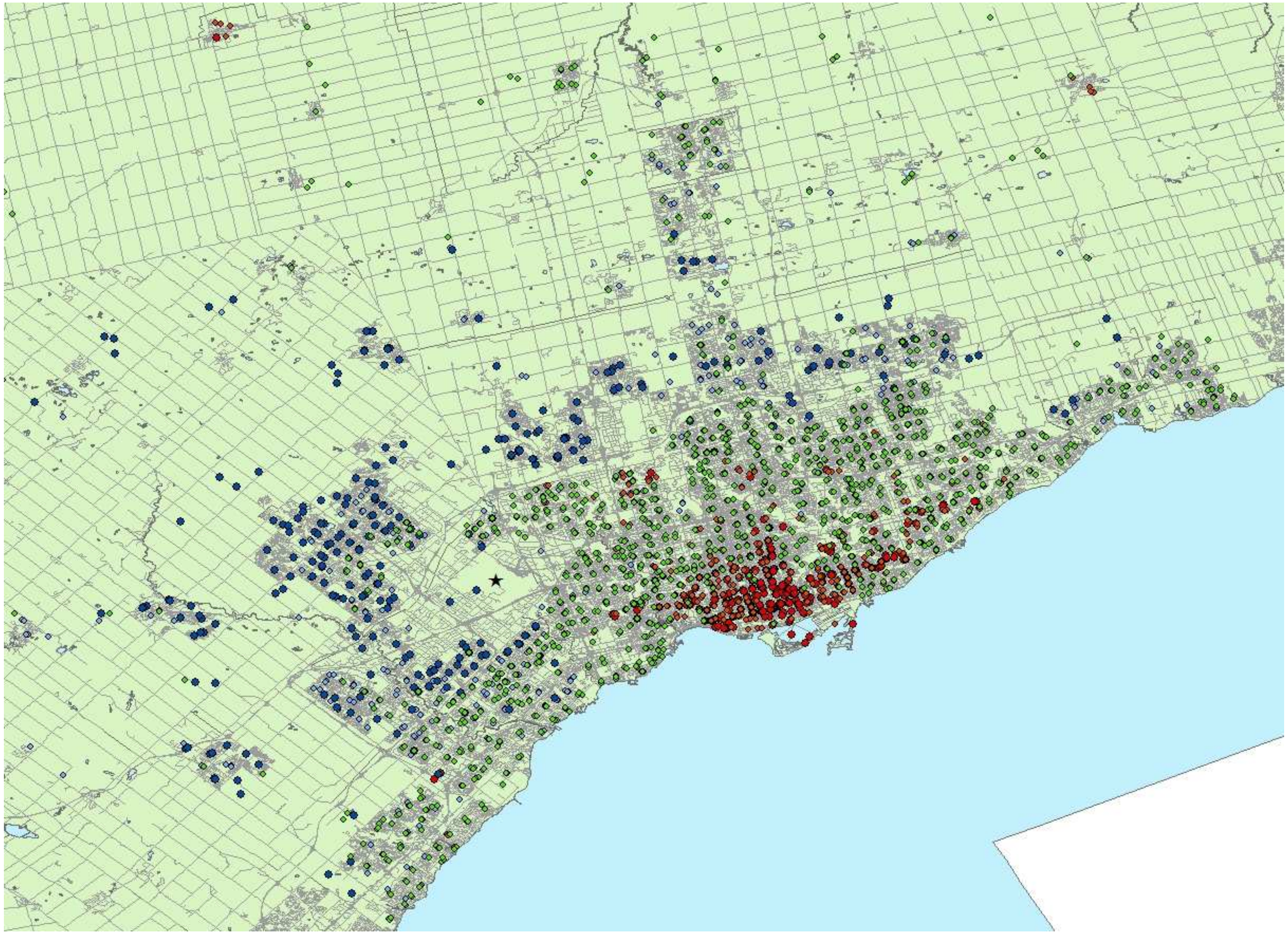

Figure 6: Mode-Share Prediction Residuals - Toronto CMA

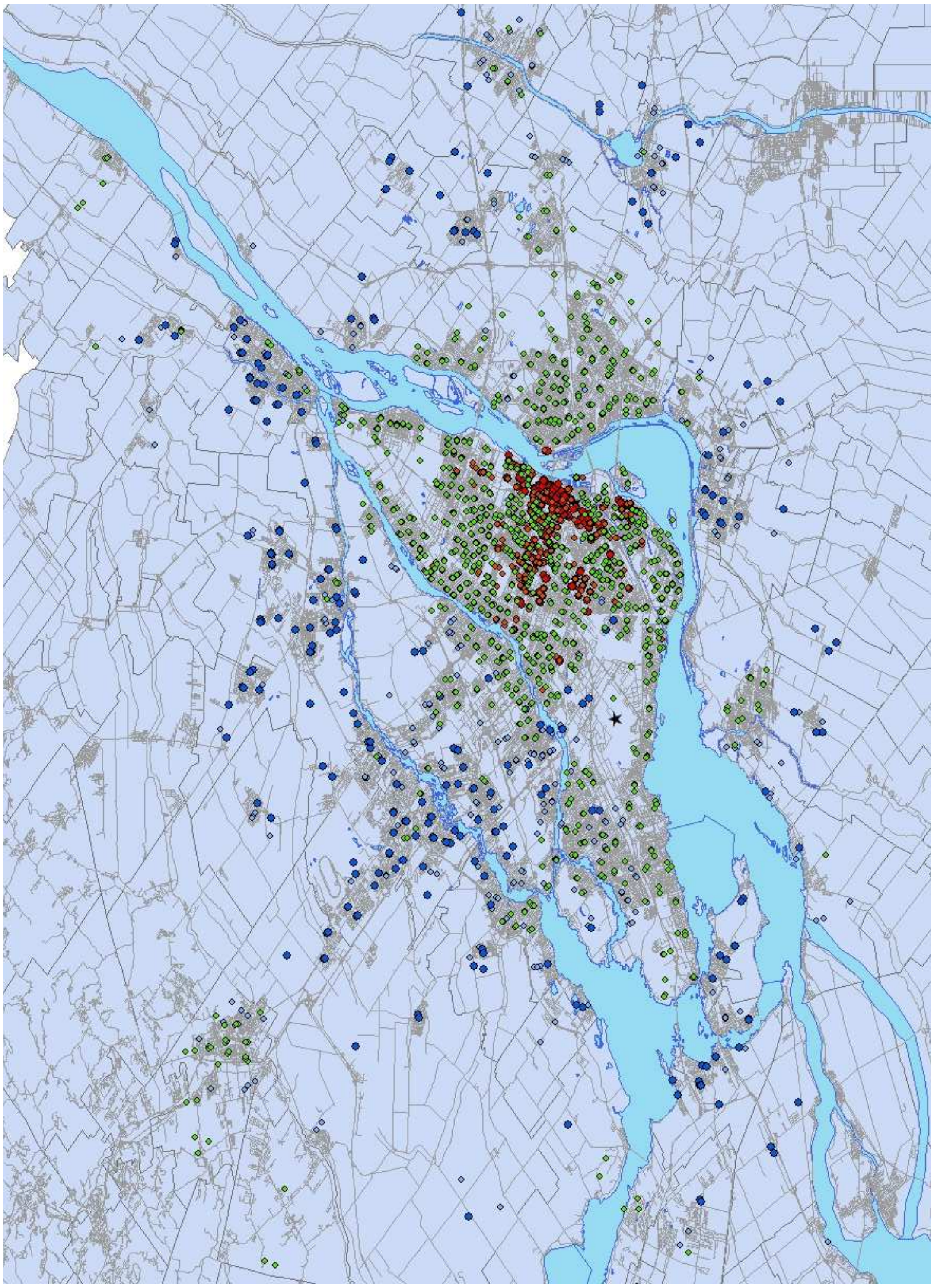

Figure 7: Mode-Share Prediction Residuals - Montreal CMA

A.4 Omitted Estimation Results

Model 1 was presented in section 5.3 but later abandoned in favor of Model 3. The estimation results for this earlier specification are provided for the purpose of comparison. Note that the estimates for parameters associated with the vector $\boldsymbol{\gamma}$ are very similar to those reported in Model 3, as expected.

Table 10: Model 1 - MNL Estimates for the Choice of Driving to Work in a Single-Occupancy Automobile

Reference group: Vancouver, non-immigrant, unmarried women

Parameter	Estimate	Robust Std. Err.	p-value	$\frac{dy}{dx}$ (at mean)
Income ($1,000)	0.00996	0.00037	0.00	0.00214
Distance (km)	0.07876	0.00258	0.00	0.01692
Squared-Distance (km2)	-0.00158	0.00008	0.00	-0.00034
Age	0.02109	0.00073	0.00	0.00453
Children	0.33161	0.01307	0.00	0.07124
Immigrant Status		0.01583	0.00	-0.12195
Marital Status	0.20432	0.0176	0.00	0.04409
Male	0.60574	0.01427	0.00	0.12965
Montreal FE	-0.43493	0.02152	0.00	-0.09566
Toronto FE	-0.39828	0.02009	0.00	-0.08559
Constant	-0.99392	0.03057	0.00	-
			n	104,873
			Correctly Predicted	74,217
			Success Rate	70.77%