A Generic Form for Capturing Unobserved Heterogeneity in Discrete Choice Modelling: Application to Neighborhood Location Choice

Ali Yazdizadeh

A Thesis In

The Department Of

Geography, Planning and Environment

Presented in Partial Fulfillment of the Requirements

For the Degree of Master of Science (Geography, Urban and Environmental Studies)

at

Concordia University

Montreal, Quebec, Canada

August 2016

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared By:

Ali Yazdizadeh

Entitled: A Generic Form for Capturing Unobserved Heterogeneity in Discrete Choice Modelling: Application to Neighborhood Location Choice

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Geography, Urban and Environmental Studies)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

____Chair
Monica Mulrennan

____Internal Examiner
Craig Townsend

____External Examiner
Ahsan Habib

____Supervisor
Zachary Patterson

____Supervisor
Bilal Farooq

Approved by ____
Chair of Department or Graduate Program Director

____
Dean of Faculty

Date ____

Abstract

A Generic Form for Capturing Unobserved Heterogeneity in Discrete Choice Modelling: Application to Neighborhood Location Choice

Discrete choice models and their strength to predict individual choices mostly depend on the quality of datasets that have been used for model generation. However, even the most comprehensive and detailed datasets are not able to observe all factors pertinent to someone's choice. This issue in the choice modelling literature has been addressed as unobserved heterogeneity, which means that individuals across populations are not affected identically by alternative attributes. Furthermore, such variation in preferences across populations and their sources are not always recognized by researchers.

There are different methods to capture unobserved heterogeneity proposed in the discrete choice literature among which the random parameters approach, also referred to as mixed logit models, the latent class approach and the agent effect approach are the most well know methods. The main contribution of this study is to extend the formulation of LC-MMNL model to capture the agent effect by including a random term in the utility function of the model. Three types of models, Mixed Multinomial Logit (MMNL), Latent Class Mixed Multinomial Logit (LC-MMNL) and Agent Effect Latent Class Mixed Multinomial Logit (AGLC-MMNL) have been generated and the results compared. Considering agent effect simultaneously with other sources of unobserved heterogeneity in a latent class context demonstrates improvement in terms of model fit as well as cross section validation. It enables us to generate a latent class model with a larger number of classes explaining more heterogeneity across the population of a neighborhood location choice study. The AGLC-MMNL model is able to detect four distinct classes of individuals in Montreal, exhibiting different behaviours while facing neighborhood location choices in the context of a Discrete Choice Experiment. The classes of the model are able to explain different behaviours of individuals based on their income level, whether they are transit or car oriented, and the importance of privacy to them.

Dedicated to

Nooshin

Acknowledgements

I would like to gratefully and sincerely thank Prof. Zachary Patterson for his guidance, understanding, patience, and most importantly, his friendship during my graduate studies at Concordia University. For everything you've done for me, Zak, I thank you.

I am also thankful to my co-supervisor Prof. Bilal Farooq, at École Polytechnique de Montréal, without whom this research would have been so lacking. I thank his paramount mentorship in providing a well-rounded experience during my graduate studies.

I would like to thank the Department of Geography, Environment and Urban Planning at Concordia University, especially Prof. Craig Townsend for his beneficial comments on my thesis proposal.

Finally, and most importantly, I would like to thank my wife Nooshin. Her support, encouragement, quiet patience and unwavering love were undeniably the bedrock upon which the past four years of my life have been built. I thank my parents, Mahnaz and Mahdi, for their faith in me and allowing me to be as ambitious as I wanted. It was under their watchful eye that I gained so much drive and an ability to tackle challenges head on. Also, I thank Nooshin's parents, Mahin and Reza. They provided Nooshin and I with unending encouragement and support.

Table of Contents

List of Tables

List of Equations

List of Acronyms

AGLC-MMNL (Agent Effect Latent Class Mixed Multinomial Logit Model)

DCE (Discrete Choice Experiment)

LC (Latent Class)

LC-MMNL (Latent Class Mixed Multinomial Logit Model)

MMNL (Mixed Multinomial Logit Model)

MNL (Multinomial Logit Model)

RP (Revealed Preference)

SP (Stated Preference)

Introduction

Discrete choice models have been widely used by researchers in the past several decades in a variety of disciplines. These models were pioneered in the domain of transportation by the works of McFadden (1977) and Ben-Akiva et al. (1980). More recent developments of this approach have resulted from the work of Hensher and Greene (2013), Bierlaire et al. (2015), Bhat et al. (2016) and Akay (2012), to name but a few.

Discrete choice models firstly were used for analysing the decision making behaviour of travellers by McFadden (1977). However, these models have turned out to be widely useful in economics and other fields of science. Transportation demand modelling and neighbourhood location choice are related fields of study and have major impacts on each other. For example, people usually choose their residential location based on how much it is close to transportation networks or how much a neighbourhood is accessible by car or public transit. Hence, so quickly discrete choice models found their way to residential location choice (McFadden, 1978).

Discrete choice models are econometric tools for predicting and explaining choices between two or more alternatives. For example, while buying a car or house an individual should make a decision to choose a car or house among a set of alternatives. In every discrete choice modelling there are two parties involved. First, the decision maker, who chooses an alternative among a set of alternatives which gives him the highest utility. On the other hand, there is a researcher who does not know anything about the utility of the decision maker. Also, she is not aware about the decision making procedure happened in the mind of the decision maker. She only observes the chosen alternative, some attributes of the alternatives as well as some characteristics of the decision makers. In other words, there are many unobserved attributes and characteristics about which the decision maker is not aware. Discrete choice modeling tries to establish a connection between these observed and unobserved factors to the decision maker's utility.

Discrete choice models commonly assume decision makers have a utility-maximizing behaviour. This assumption, based on Random Utility Maximization (RUM) theory (McFadden, 1977), says that the decision maker chooses the alternative with the highest utility for him

(Train, 2009).

It should be mentioned that the assumption of utility-maximizing decision makers is not the only one choice rule in the literature of choice modelling. There are other choice rules, such as utility satisficing (Simon, 1956; Stüttgen, Boatwright, & Monroe, 2012) and random regret minimising (Thiene, Boeri, & Chorus, 2012). Utility satisficing rule says that the decision maker chooses the alternative which first satisfies some criterion (Stüttgen, Boatwright, & Monroe, 2012). In other words, the decision maker does not choose the alternative with the highest utility, but he chooses the first alternative that is good enough for him.

Random regret minimising rule is another paradigm in discrete choice literature. It assumes that the decision maker not only considers the utility of each alternative, but also he considers the possibility of that the non-chosen alternative has higher utility than the chosen one. In other words, the random regret minimization rule assumes that the decision maker anticipate the possible regret associated with the chosen alternative and consider this in his decision making procedure. (Thiene, Boeri, & Chorus, 2012). While there are other choice rules have been applied in several studies, the current study focuses on utility-maximizing rule, which is the most used in the literature.

As mentioned above, the researchers only observe some attributes of alternatives as well as some characteristics of individuals. Hence, there are many other factors that affect the individuals' choices which are not observed by the researcher and cannot be estimated by the model. However, the initial versions of discreet choice models, called Multinomial Logit (MNL) models, tried to consider the effects of unobserved factors in modelling procedure by including an error term in the mathematical formulation of the model. In any typical discrete choice model, there is a linear utility function that consists of two parts: first, the observed or representative utility, which is the utility measured by the observed attributes of alternatives and characteristics of decision makers. The representative utility can be measured by some parameters estimated statistically. Second, the error term that constitutes aspects of the utility that the researcher cannot observe. In other words, the error term is the difference between the true utility, perceived by decision maker, and the representative utility observed by the researcher. By decomposing the utility function into two parts, i.e. the representative (observed) utility and the error term, the researcher tries to capture unobserved factors, affecting decision maker utility, as well as the observed attributes of alternatives and characteristics of the decision maker.

As mentioned above, discrete choice models are applied on datasets, hence, the quality of dataset can largely affect the model results. It is worth to briefly explain the methods largely used in the literature for gathering data. Discrete choice models are usually estimated on datasets obtained from two different sources: Revealed Preference (RP) and Stated Preference (SP) (Louviere, Hensher, & Swait, 2000).

RP data are obtained by observing actual respondent behaviour in its usual context (Houthakker, 1950; Morikawa, Ben-Akiva, & McFadden, 2002). RP surveys were widely used in early econometrics literature on the topic, as the traditional view was that valid choice data resulted *only* from actual choices having been made (Morikawa et al., 2002).

SP data are used when the available RP data are limited for various reasons, such as introducing a new alternative which is not available to the decision makers in actual context. Also, in some cases, the alternative is available to the decision makers, however, there is not sufficient variability in actual choices and researcher cannot analysis the attributes of interest, in this case, again the SP data are used. In addition, in some cases the researcher wants to estimate the utility of a hypothetical choice which is not available to the decision makers, in this case also she decides to design a SP experiment.

Stated Preferences are elicited through Discrete Choice Experiments (DCEs) (Louviere, Flynn, & Carson, 2010). DCEs, are surveys that aim to simulate choices of individuals by including the attributes of alternatives each individual faced. Researchers ask individuals to choose among a hypothetical set of alternatives (referred as choice tasks) each of which has different levels of attributes characterizing the alternatives. The levels of each attribute of each alternative are defined beforehand based on a study of previous relevant studies, focus groups, piloted versions of the survey, and an experimental design. A simple example of a choice task can be given with three transportation modes e.g. Car, Metro and Bike as alternatives. Each of them has time and cost as varying attributes, with different levels. For example, the travel time of Car, Metro and Bike can be defined as 20, 30, and 45 minutes, respectively. Also, the total travel cost by each alternative can be defined as 8, 3 and 2 CAD, respectively. The respondent is asked to choose an alternative among these three alternatives in each choice task. The researcher can design different choice tasks and ask the respondents to respond to them, repeatedly.

Then, the responses to choice tasks (the choices that respondents make) are analyzed using discrete choice models. There is a large body of literature on the use of stated preference surveys from myriad fields of science including economics (Louviere et al., 2000), marketing (Louviere et al., 2008), environmental valuation (Boxall, Adamowicz, Swait, Williams, & Louviere, 1996), and health economics (de Bekker-Grob, Ryan, & Gerard, 2012) to name but a few.

Through analyzing the respondents' preferences, researchers can reveal details about the effects of different attributes' levels on respondents' choices. In addition, while there are several attributes considered in each choice task, the results can specify which attributes of the alternatives are the most important from individual's point of view. Furthermore, the results of a SP can show how much the different levels of attributes can affect the people's preferences.

Both types of aforementioned surveys, i.e. RP and SP surveys, have their pros and cons. However, regardless of data source, as mentioned earlier, not all factors pertinent to someone's choice can be observed from either stated or revealed preference data. In other words, with both types of data, some of the many factors affecting the likelihood of an alternative are not likely to be observed by the researcher.

For example, we can consider gender as an observed human characteristic that affects survey outcomes. While there are obviously differences between men and women (providing a rationale for using a dummy variable such as 1 for male and 0 otherwise) (Mannering, Shankar, & Bhat, 2016), there is also considerable heterogeneity across individuals of the same gender, including differences in attitudes and life styles that are generally unobservable to the researcher. These differences, i.e. heterogeneities, are not taken into account even when other observable variables have been included in the model. Furthermore, such differences cannot be identified by the SP or RP data.

As mentioned earlier, the classic formulation of discrete choice models, the researcher assumes that these unobserved factors are captured by the error term. However, this approach will lead to large estimated variances for parameters of the model, referred to as biased estimated parameters in the literature (Dugundji & Walker, 2005; Heckman & Singer, 1984; Reader, 1993). In other words, the more unobserved attributes are not represented by observed utility, the more error will be entered into the estimation of the model. Hence, the lack of knowledge about a major part of an individual's characteristics and alternative attributes will affect the model, and generally, the classic models in discrete choice family cannot capture them.

The above-mentioned issue in choice modelling has been addressed as unobserved heterogeneity, which means that individuals across populations are not affected identically by the same alternative attributes. Furthermore, as mentioned above, such variation in preferences across populations and, more importantly their sources, are not always recognized by researchers.

This issue has been documented in a number of studies as: the failure to consider unobserved heterogeneity in choice models, inherently leading to biased and inefficient parameter estimation (Dugundji & Walker, 2005; Heckman & Singer, 1984; Reader, 1993). This can consequently result in incorrect inferences and predictions. There are various approaches in the literature to account for unobserved heterogeneity:

1- Random parameter discrete choice models,
2- Latent class discrete choice models, and
3- Agent effect discrete choice models.

Among the above mentioned approaches the random parameter approach (M. Ben-Akiva, Bolduc, & Walker, 2001; Bhat, 2001; Hensher & Greene, 2003; McFadden & Train, 2000) is the most well-known and widely used. Random parameter discrete choice models, also known as Mixed Multinomial Logit (MMNL) models (McFadden & Train, 2000; Revelt & Train, 1998) or Logit Kernel (M. Ben-Akiva et al., 2001), allow the parameters of a model to vary across observations (M. E. Ben-Akiva & Lerman, 1985) in a sense that a distribution, instead of having unique value, will be estimated for each random parameter in the utility function of model. The most important aspect of an MMNL model is to define an appropriate distribution for the population under study. Usually, the researcher selects a distribution among different available distributions introduced in statistical analysis, based on a prior modelling procedure to find the best distribution fitting the dataset. Hence, some parameters of the MMNL model can be estimated as random values with a user-defined distribution (e.g. Normal, Triangular or Uniform distributions) and then a mean, as well as standard error will be estimated by the model.

Expressed in a different way, MMNL models try to decrease the weight of the error term component in the utility function of discrete choice models by capturing, to some extent, unobserved heterogeneity in the random parameters of the model. However, a major issue related to this approach is that MMNL model assumes a single average behaviour and deviations from it across population. In other words, the MMNL model assumes only one average behaviour for all the individuals in a population. As a result, the behaviour of any individual in the dataset can be estimated by a deviation from this average behaviour - something that is not a valid assumption in many cases. In reality, we have many "average" behaviours in a dataset, as different segments of the population will behave totally differently according to the characteristics of their members. This limitation of MMNL models has resulted in an alternative model, referred to as the Latent Class (LC) model (Boxall & Adamowicz, 2002; Greene & Hensher, 2003) which will be discussed in the following. However, one major issue is the prior assumption about the distribution of the random parameters in the MMNL model that are estimated. The typical approach used in the literature is to generate models with different distributions, such as the normal, triangular and uniform, and test which model fits better and estimates more meaningful parameters. Also, as Green and Hensher (2003) mentioned, for some attributes that are expected *a priori* to produce a negative mean estimate we can use the lognormal distribution. However, choosing the distribution for estimating random parameters requires prior knowledge about the population and a sense about how the real behavior is distributed (Green and Hensher, 2003).

The LC model assumes that there are different segments of a population that behave relatively homogenously within each segment but heterogeneously across (differently from other) segments (Wen, Huang, Fu, & Chou, 2013). While the MMNL model defines random parameters as a continuous probability distribution and assumes an average behaviour for all segments of the population, LC models rest on the assumption that individuals in each segment of the population behave differently and an average behaviour for each segment should be estimated. Hence, the LC models are closer to the reality of the population by considering different average behaviours for different segments of the population.

By means of estimating discrete numbers of classes, LC models provide multiple and totally distinct average behaviours among classes. Each class is associated with a different set of parameters in a “class-segmenting” utility function (Shen, 2009). LC models have been applied in various fields of marketing research, economics, transportation and geography, such as, route choice modelling (Greene & Hensher, 2003), mode choice modelling (Atasoy, Glerum, & Bierlaire, 2011; Gopimatj, 1997a), international air carrier choice (Wen & Lai, 2010), environmental preferences (Morey, Thacher, & Breffle, 2006) and household location choice (Joan L Walker & Li, 2007).

In LC models each individual is assigned to a “class” according to a logit-like probability function (the class membership function) considering individual characteristics. The term “latent” stems from the fact that the class to which each respondent is assigned is not available to the analyst (Greene & Hensher, 2013). Also, in many cases the number of classes is not known by the analyst. One of the issues regarding latent class models is how to define the number of classes that best represents the true segments across population. In other words, since the research does not know the number of classes, how can she determine the number of classes that should be included in the model? In this case, the best method, widely used by different researchers (Greene & Hensher, 2013; Teichert, Shehu, & von Wartburg, 2008; Joan L Walker & Li, 2007), to determine the number of classes is generating a series of LC models with different number of classes. Then, based on a selection criteria, such as the Bayesian Information Criterion (BIC), the generated models can be compared with each other and the best model selected (Greene & Hensher, 2013).

It should be mentioned that LC models do not allow the parameters of the model to be continuously distributed within each class. Hence, within each class unobserved heterogeneity is not captured, like the MNL model, and only represented by the error term in the utility function of each class. In other words, LC modelling assumes only an average behaviour for all the members of each class without any deviation from it. This limitation has been considered as a drawback in the literature of LC modelling (Greene & Hensher, 2013; Mannering et al., 2016).

The aforementioned limitation of LC models has led researchers to combine the random parameter and latent class approaches to consider the unobserved heterogeneity within each class in the model. Within this approach, also known as the Latent Class Mixed Multinomial Logit (LC-MMNL) model (Greene & Hensher, 2013), first the number of latent classes are specified and then the parameters are allowed to be varied (based on user-defined probability distributions) within each class. The advantage of this combined approach over the LC model is that it not only considers multiple average behaviours across a population, but it also estimates multiple deviations from these average behaviours.

By using LC-MMNL models, the limitations of LC models that estimate one average behaviour within each class is relaxed. Hence, in theory, LC-MMNL models should account for unobserved heterogeneity more efficiently than either LC or random parameters approaches alone. In fact, recently this assumption has been approved by two studies done in transportation context (Bujosa, Riera, & Hicks, 2010; Greene & Hensher, 2013). However, ongoing research is needed to demonstrate the superiority of LC-MMNL models over LC and MMNL models applied on different datasets.

Another source of unobserved heterogeneity in datasets is the correlation across sequential observations (choices) (Mannering et al., 2016). It is worth mentioning that SP datasets can be gathered using two different approaches: times-series and cross-sectional. Time-series data are the sequence of observations of individuals at two or more periods in time. Cross-sectional data, on the other hand are the observations of individuals at the same point in time (Bierlaire, 2014), which means that no sequence of observations for an individual have been gathered. By using time-series data, individual related unobserved factors that are persistent over time would be another source of unobserved heterogeneity. This type of unobserved heterogeneity is known as an “agent” or “panel” effect in the literature (Bierlaire, 2014). Obviously, the MMNL, LC and LC-MMNL models cannot capture this kind of heterogeneity, as they consider choice tasks independently, whether they are faced by the same individual or not. In other word, the above-mentioned models assume that there is no correlation between the choice tasks of the same individual.

In order to accommodate the agent effect in a discrete choice model, many researchers have included a random parameter in every utility function with which an individual is faced. By doing this, the correlation among choice tasks related to a same individual will be captured by this random parameter. However, to my knowledge, no attempt has been made to generate a model that accounts for all aforementioned methods of capturing unobserved heterogeneity. Hence, the intent of this study is to combine the LC-MMNL approach (which itself accounts for unobserved heterogeneity via two other approaches, i.e. random parameter and latent class) and the agent effect approach to allow analysts to make more accurate inferences.

The remaining part of this thesis is organized as follows. In the next section, the formulation of Multinomial Logit, Mixed Multinomial Logit, and Latent Class models will be introduced. Afterwards, the literature on different approaches for capturing the unobserved heterogeneity is reviewed. Finally, the co-authored article, submitted to 96th annual meeting of Transportation Research Board is presented. It should be mentioned that the combined approach in this study requires a new mathematical formulation (of utility function and log-likelihood function of discrete choice models introduced in the literature) [1].

The initial version of discrete choice models, i.e. MNL model, rests on the above-mentioned assumptions about observed utility. The observed utility has a linear formulation and the unobserved utility is represented in the error terms. It is assumed that the error term in the MNL model is distributed independently, identically extreme value (Train, 2009). However, the assumption that the error term for each alternative is independent from the error term of other alternatives is restrictive and is not satisfied in many cases. As in fact, alternatives have some common attributes that cause them to be correlated.

Mathematical Formulation of Discrete Choice Models Accounting for Unobserved Heterogeneity

In this section, I first describe the basic discrete choice model (MNL) that does not account for unobserved heterogeneity across observations, and then continue to describe two other formulations that do. Three discrete choice models, i.e. Multinomial Logit (MNL), LC and LC-

[1] This issue is explained in detail in the co-authored article.

MMNL, will be discussed respectively.

In the discrete choice family of models, it is usually assumed that a sampled individual faces a choice amongst a set of alternatives in each of T choice situations. An individual n is assumed to consider the full set of offered alternatives and to choose the alternative with the highest utility. Denoting X_{nit} as a $K \times 1$ vector of attributes of alternative i in choice situation t and characteristics of individual n and β as a $1 \times K$ vector of estimated parameters associated with X_{nit}, the utility for alternative i may be expressed as (Greene & Hensher, 2003):

$$U_{nit} = \beta X_{nit} + \epsilon_{nit} \tag{1}$$

where U_{nit} is the utility associated with alternative i held by individual n in choice situation t and ϵ_{nit} reflects unobserved influences on utility. The choice probability of alternative i is given as in the Multinomial Logit (MNL) model (Greene & Hensher, 2003):

$$P_{nit} = \frac{\exp(V_{nit})}{\sum_{j \in J} \exp(V_{njt})} \tag{2}$$

where

$$V_{nit} = \beta X_{nit} \tag{3}$$

and j refers to any alternative in choice set J. V_{nit} is the observed utility, also referred to as systematic utility.

In the MNL model, we assume that ϵ_{ni} is independent and identically extreme value distributed (iid). However, in reality, there might be correlations across the alternatives in each choice situation and indeed across alternatives. Hence, we can assume the parameters of the model to be random parameters with a distribution across observations and define the vector of parameters β as (Bierlaire, 2014):

$$\beta' = \bar{\beta}' + \sigma'\xi \tag{4}$$

Where $\bar{\beta}'$ and $\sigma'$are the mean and variance of the β', e.g. $\beta' \sim N\left(\overline{\beta'},\ {\sigma'}^2\right)$, and ξ is a random value with a specified distribution, e.g. $\xi \sim N(\,0,\ 1)$. The utility function of alternative i observed by individual n can be expressed as:

$$U_{ni} = \ \bar{\beta}' X_{ni} + \sigma'\xi X_{ni} + \nu_{ni} \tag{5}$$

where ν_{ni} is random term with zero mean and independent and identically (iid) distribution[2]. The conditional probability of alternative i being chosen by individual n is:

$$P_n(i|\xi) = \frac{\exp(\bar{\beta}' X_{ni} + \sigma' \xi X_{ni})}{\sum_{j \in J} \exp(\bar{\beta}' X_{nj} + \sigma' \xi X_{nj})} \tag{6}$$

and the unconditional probability is demonstrated as:

$$P_n(i) = \int_\xi P_n(i|\xi).\, f(\xi).\, d\xi \tag{7}$$

where $f(\xi)$ is the probability density function of ξ. This choice probability is a weighted average of the MNL model choice probability function evaluated at different values of β, with the weight given by the probability density function $f(\beta|\ \theta\ )$. In most applications, the normal distribution is frequently used for random coefficients. However, other continuous distributions such as the Lognormal, Uniform, Triangular are applicable (Train, 2009).

As shown above, unobserved heterogeneity across observations can be taken into account by the MMNL Model which allows the parameters of the model to vary over observations and estimates them based on an assumed probability distribution (Lee, Fujiwara, Zhang, & Sugie, 2003). Some parameters of the MMNL model can be estimated as random values with a specific mean and a standard error. However, although the MMNL model tries to explicitly consider heterogeneity in the population, the sources of heterogeneity cannot be explained by them. For, as mentioned earlier, in many cases these sources relate to the characteristics of individuals, which cannot be captured by MMNL models.

In LC models, each individual is assigned to a "class" according to a probability function, referred to as a Class Membership Function, which is a logit-like probability function considering individual characteristics. The term "latent" stems from the fact that the class to which each respondent is assigned is not available to the analyst (Greene & Hensher, 2013).

By considering different classes, LC models try to explain taste variation across observations (or individuals) as well as their sensitivity to the variations of alternative variables. In an LC model formulation, the conditional utility function of alternative i for individual n of class s is similar to the MNL model (Greene & Hensher, 2003):

---

[2] In probability theory and statistics, a sequence or other collection of random variables is independent and identically distributed (iid) if each random variable has the same probability distribution as the others and all are mutually independent (Parzen, 1962).

$$U_{nis} = \beta_s X_{ni} + \nu_{nis} \qquad (8)$$

In the latent class model, the probability of alternative i being chosen by individual n of class s is given by:

$$P_n(i) = \sum_{s=1}^{S} P_n(i|s) \times M_{n|s} \qquad (9)$$

where

$$P_n(i|s) = \frac{\exp(\beta_s X_{ni})}{\sum_{j \in C_n} \exp(\beta_s X_{nj})} \qquad (10)$$

$$M_{n|s} = \exp \frac{(\gamma_s Z_n)}{\sum_{s=1}^{S} exp(\gamma_s Z_n)} \qquad (11)$$

Where $\mathrm{Z_n}$ is a vector of class variables consisting of individual socioeconomic characteristics and γ_{s} is a vector of parameters for class s. Equation (9) calculates the probability of alternative i being chosen by individual n of class s, while, Equation (10) is the class membership function of decision maker n in class s.

As mentioned earlier, although LC models offer an alternative perspective to the MMNL, they assume preference homogeneity within each class. Instead of considering latent class model parameters fixed, we can allow them to be random thus allowing for preference heterogeneity within a class (Greene & Hensher, 2013). To accommodate the new layer of heterogeneity, continuous variation of the parameters within classes will be allowed. Hence, the utility function of the LC-MMNL model is (Greene & Hensher, 2013):

$$U_{ni|s} = \beta_s' X_{ni} + \nu_{nis} \qquad (12)$$

Where:

$$\beta_s' = \bar{\beta}_s' + \sigma_s' \xi$$

where $\bar{\beta}_s'$ and $\sigma'$are the mean and variance of the β', e.g. $\beta_s' \sim N(\bar{\beta}_s', \ \sigma_s'^2)$, and ξ is a random value with a specified distribution. The utility function of alternative i observed by individual n can be expressed as:

$$U_{ni|s} = \bar{\beta}_s' X_{ni|s} + \sigma_s' \xi X_{ni|s} + \nu_{ni|s} \qquad (13)$$

where $\nu_{ni|s}$ is random term with zero mean and iid distribution. The conditional probability of alternative i in class s being chosen by individual n is:

$$P_n(i|\xi, s) = \frac{\exp(\bar{\beta}_s' X_{ni|s} + \sigma_s' \xi X_{ni|s})}{\sum_{j \in J} \exp(\bar{\beta}' X_{nj|s} + \sigma' \xi X_{nj|s})} \qquad (14)$$

The class heterogeneity is defined by class membership function like the LC model:

$$M_{n|s} = \exp\frac{(\gamma_s Z_n)}{\sum_{s=1}^{S} exp(\gamma_s Z_n)} \tag{15}$$

The contribution of individual i to the log-likelihood of the model is obtained by integrating the two-layer heterogeneity: within-class heterogeneity and class heterogeneity (Greene & Hensher, 2013). The unconditional probability of alternative i in class s being chosen by individual n in LC-MMNL model is given as follows:

$$P_n(i) = \sum_{s=1}^{S}[M_{n|s} \times \int_{\xi} P_n(i|\xi, s) \,.\, f(\xi) d\xi] \tag{16}$$

Although the MMNL, LC and LC-MMNL account for unobserved heterogeneity across observations, they do not consider the heterogeneity across individuals when using panel (multiple observations from the same respondent) data. While using panel data, each individual faces more than one choice situation at different points of time. This can cause observations from the same individual to be correlated. The source of such correlation is individual related unobserved factors that are persistent across time (Bierlaire, 2014). Hence, addressing the source of this correlation in the LC-MMNL model, first we add the t subscript to the utility function:

$$U_{nti|s} = \beta_s' X_{nti} + \nu_{ntis} \tag{17}$$

here $U_{nti|s}$ is the utility for alternative i being chosen by individual n in class s at time point t.
Afterwards we relax the assumption that ν_{ntis} are independent across t:

$$\nu_{ntis} = \alpha_{ni} + \nu_{ntis}' \tag{18}$$

while ν_{ntis}' is independent across t, α_{ni} capture permanent taste heterogeneity.

Consequently, the conditional probability (on ξ, s and α_n) of alternative i being chosen by individual n in class s at time t is:

$$P_{nt}(i|\xi, s, \alpha_n) = \frac{\exp(\,\bar{\beta}_s' X_{nti|s} + \sigma_s' \xi X_{nti|s} + \alpha_{ni})}{\sum_{j\in J} \exp(\,\bar{\beta}_s' X_{ntj|s} + \sigma' \xi X_{ntj|s} + \alpha_{nj})} \tag{19}$$

where α_n is the vector gathering all parameters α_{ni}. Removing the probability conditions on ξ and s we will have:

$$P_{nt}(i|\alpha_n) = \sum_{s=1}^{S}[M_{n|s} \times \int_{\xi} P_{nt}(i|\xi, s, \alpha_n) \,.\, f(\xi).\, d\xi] \tag{20}$$

Also, the contribution of individual n to the log likelihood is:

$$P_n(i_1, i_2, \dots, i_T|\alpha_n) = \prod_{t=1}^{T_n} P_{nt}(i|\alpha_n) \tag{21}$$

where $T_n \geq 1$ is the number of choice situations in which individual n is engaged. In addition, the choice probability of alternative i unconditional on α_n would be:

$$P_n(i_1, i_2, \dots, i_T) = \int_\alpha \prod_{t=1}^{T_n} P_{nt}(i|\alpha_n)\, f(\alpha) d(\alpha) \tag{22}$$

where $f(\alpha)$ is the probability density function of α. Finally, the unconditional contribution of individual n to the log likelihood is:

$$P_n(i_1, i_2, \dots, i_T) = \int_\alpha \prod_{t=1}^{T_n} [\sum_{s=1}^{S} [M_{n|s} \times \int_\xi P_{nt}(i|\xi, s, \alpha_n)\,.\, f(\xi).\, d\xi]]\, f(\alpha) d(\alpha) \tag{23}$$

One of the issues regarding any random parameter discrete choice model formulation is that solving the integral in unconditional probability functions of such models, such as Equation (7) , (16) and (23), needs simulation (Train, 2009), as such integrals do not have a closed form expression. By simulation, random values (for random parameters) are drawn from an assumed distribution, and then, the integral can be approximated by getting a summation over the values of the function under the integral, while the random parameters of the model have been substituted by the randomly drawn values. Hence, Equation (22) can be approximated by:

$$P_n(i_1, i_2, \dots, i_T) \approx \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_n} [\sum_{s=1}^{S} [M_{n|s} \times \frac{1}{R} \sum_{r=1}^{R} P_{nt}(i|\xi^r, s, \alpha^r)\ ]] \tag{24}$$

where R is the number of draws in simulating procedure.

Also, the overall simulated log-likelihood is given by the following equation:

$$\log L = \sum_{n=1}^{N} \log[\frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_n} [\sum_{s=1}^{S} [M_{n|s} \times \frac{1}{R} \sum_{r=1}^{R} P_{nt}(i|\xi^r, s, \alpha^r)\ ]]] \tag{25}$$

where N is the number of individuals in data set.

Literature Review

This literature review seeks to cover the most prominent studies, mainly conducted in the fields of transportation and market research, which have considered unobserved heterogeneity in discrete choice modelling. Also, the most relevant studies on SP surveys and their application have been reviewed. The reason for reviewing the SP survey literature is that the data used in this study, explained in detail in the co-authored article, has been gathered via the use of a video game SP survey in 2014 (Patterson, Darbani, Rezaei, Zacharias, & Yazdizadeh, 2017).

Stated Preference Surveys

As mentioned in the Introduction, there are two methods for gathering data for discrete choice models: revealed preferences data and stated preferences data. The term revealed preference was first used by Samuelson (Samuelson, 1938). He considered individual behaviour as a series of choices. He suggested that by comparing observed behaviour with available alternatives, individual's preferences can be inferred. This theory was subsequently expanded on and eventually resulted in the development of choice models (Abley, 2000). Revealed preference data is gathered either through direct observation, or through surveys asking about actual behaviour. The collected information about actual choices can be used to estimate choice models (Boxall et al., 1996).

Although revealed preference methods have been used widely to identify individual choices in real world, the technique exhibits a number of severe limitations (Abley, 2000):

1- There is a lack of sufficiently large variations in attributes of interest using revealed preference data.
2- There is often considerable correlation between the variables in revealed preference data. This can cause several problems (i.e. it can be difficult to distinguish between the effects of different variables, which make the estimation of the model parameters impossible).
3- Since RP data is based on actual alternatives, the use of these techniques proves difficult for predicting the effect of new products or services.

Due to the problems that researchers have encountered using RP techniques, they have looked for new methods of predicting individual preferences. Their efforts have led to the introduction of SP methods in the field of choice modeling.

Stated preference (SP) techniques are a class of market research tools that allow researchers to identify how consumers value different product or service attributes and choose their preferred alternative among a set of options (Abley, 2000). In an SP survey respondents are asked to rank, rate or choose between different hypothetical product or service scenarios, which consist of different attribute mixes. These choices can be used to infer how respondents value different attributes (Abley, 2000).

SP methods have been widely used by researchers because of the two following reasons: first, RP data cannot exist for alternatives that are currently unavailable alternatives. Second, in some choice sets, there are some existing alternatives without any considerable differences in attribute mixes. As SP methods allow the researchers to include hypothetical alternatives, and ensure sufficient variation in, and orthogonality between, attributes they can be useful for studying such choice sets (Hensher, 1994).

SP surveys can be applied to measure the total economic value; i.e. both non-use value and option value are embodied in SP surveys methods. This characteristic offers consequential potential for estimating the value of future or hypothetical (but realistic) goods and interventions. Stated preference methods show their primacy in providing a more direct route to the valuation of the characteristics or attributes of a good, and of marginal changes in these characteristics, rather than the value of the good as a whole. This can result in consequences of great importance in management, decision making, project evaluation and policy appraisal in which the decisions depends on changes in levels of attributes (Stephens, 2010).

While SP methods enjoy several advantages, the efficacy of their results depends on their ability to present realistic hypothetical alternatives for respondents. This important aspect of SP methods is partly related to the method of presenting alternatives or scenarios, which can be categorized into two major methods: traditional text format and modern visual technologies.

In the traditional text format, respondents only see a written explanation of each alternative and based on its attributes she can choose her preferred choice among a set of alternatives. However, the use of modern visual techniques, such as including video clips or 3D simulation models has been popular among researchers in the context of SP surveys (Appleton & Lovett, 2005; Arnberger & Eder, 2011; Bateman, Day, Jones, & Jude, 2009; Bienabe & Hearne, 2006; R. Laing et al., 2009; R. A. Laing, Davies, Hargreaves, & Scott, 2004; Mostofi-Darbani, Rezaei, Patterson, & Zacharias, 2014; Vriens, Loosschilder, Rosbergen, & Wittink, 1998).

These technologies can be categorized into two major types: 2-Dimension technologies and 3-Dimension ones. There are several examples of using both types in the literature. The following literature describes research using SP surveys as well as other types of surveys using visualization techniques.

2D visual aid techniques usually use images, photos and maps to make surveys more comprehensible. For example, some studies have mentioned that the inclusion of images can help respondents visualize and better understand the attribute levels in choice tasks (Beharry-Borg & Scarpa, 2010; Berninger, Adamowicz, Kneeshaw, & Messier, 2010). In other studies, photographs have been used to present different landscape types to the respondents (Arnberger & Eder, 2011; Molnarova et al., 2012; Yannes, Lownes, Garrick, & Johnston, 2010).

Maps are another kind of 2D visualization technique that can improve the perception of respondents. The contribution of Geographic Information Systems (GIS) to producing more realistic maps has been studied (Bienabe & Hearne, 2006; Yamada & Thill, 2003). Use of “public consultation geographic information systems” to increase public involvement in policymaking procedures as well as the contribution of GIS to promote the goals of nongovernmental organizations, grassroots groups, and community-based organizations has been studied as well (Brown, Weber, & de Bie, 2015; Sieber, 2006).

Using 3D visualization techniques has emerged more recently. For example, Tilahun et al. (2007) have used video-clips from cyclists perspective to show different bicycle facilities to the respondents. Furthermore, 3D modelling software has also been used by different researchers in recent past (R. Laing et al., 2009; R. A. Laing et al., 2004). For example, 3D simulated choice tasks for presenting different housing preferences have been used in the field of urban planning by different studies (Orzechowski, Arentze, Borgers, & Timmermans, 2005). Moreover, virtual reality (VR) technologies have been used to allow the participant an interactive virtual product experience (Tovares, Boatwright, & Cagan, 2014). The use of 3D GIS for presenting the built environment more realistically, by coupling GIS to 3D visualization software, is another kind of modern visual technique for presenting plans and scenarios. Recently, Perdomo et al. (2014) have used 3D traffic simulation software for presenting various characteristics of roundabouts in a SP survey design.

Choice Models Accounting for Unobserved Heterogeneity

The random parameter latent class approach (formally described above) was first applied by Green and Hensher (2003) on three DCE datasets in Australia:

1- A Stated choice experiment for long distance car travel

2- A Stated choice experiment for urban commuting

3- A revealed preference study of long distance non-commuting modal choice

After a detailed comparison based on Bayesian Information Criterion (BIC) index, significant level of parameters, and willingness to pay estimates for the value of total travel time savings and on-time delivery, they concluded that the LC-MMNL model showed superiority over the LC and MMNL models.

Also, the authors stated that several issues should be considered in estimating a mixed logit model. First, selecting the parameters, among all the parameters included in the model, which would be treated as random parameter and their distribution are very important in generating any random parameter discrete choice model. They used Lagrange Multiplier tests (McFadden & Train, 2000) for testing the presence of random components. In addition, they considered Normal, Triangular, Uniform and Lognormal distributions for the distribution of random parameters in their model. Several models were generated based on above-mentioned statistical distributions and their results compared. However, their empirical results showed that all the distributions have at least one major deficiency – typically with respect to sign and length of the tail(s). Hence, they applied truncated or constrained distributions.

Furthermore, as mentioned above, random parameter models need simulation for solving the integral of log-likelihood function. The authors discussed about the number of draws required in the simulation procedure of their MMNL model. They found that as a general rule, the greater the complexity of the model, in terms of the number of random parameters and the treatment of heterogeneity around the mean, correlation of attributes and alternatives, the larger the number of draws is required for having stable parameter estimates. However, they did not suggest a specific number of draws, instead they have stated that the best way to find whether parameter estimates are stabilized is to estimate models over a range of draws (e.g. 25, 50, 100, 250, 500, 1,000 and 2,000 draws).

Considering the latent class approach, LC models have been applied in various contexts of marketing research, economics, transportation and geography arena, for example, route choice models (Greene & Hensher, 2003), mode choice (Atasoy et al., 2011; M. Ben-Akiva & Boccara, 1995; Gopimatj, 1997b; Hosoda, 1999), international air carrier choice (Wen & Lai, 2010), environmental preferences (Morey et al., 2006) and household location choice (Joan L Walker & Li, 2007), to name but a few.

Greene and Hensher (2003) in 2003 applied MMNL and LC models on a Stated Preference (SP) dataset of long distance travel by three road types (2-lane, 4-lane without a median and 4-lane with a median) by car in New Zealand. A triangular distribution was selected for the random parameters in mixed logit model. The authors generated a series of LC models with 2, 3, 4 and 5 classes, and among them the model with three latent classes demonstrated the best fit. However, they did not explain the class membership function for each class. Afterwards, they compared the results of the LC model with a MMNL and MNL model.

For comparing the LC model with MMNL, they argued that since the MMNL and LC models don't have the same structure, comparing them based on a likelihood ratio test is not appropriate. In addition, the authors mentioned that comparing the absolute values of the parameters across models is not informative because of scale differences between the parameters estimated in two models. Their comparison, of LC and MMNL models, is based on willingness to pay indicators and elasticities. Comparing the implied Values of Travel Time Savings (VTTS) across the three latent classes shows a considerable difference across the classes. Overall, the LC model revealed three apparent segments for the mean estimates of VTTS. This interesting result demonstrates that the classes can mainly be distinguished by the values of VTTS for each class. In other words, there is a great variation across individuals with respect to VTTS.

Regarding choice elasticities for travel time and travel cost, they observed substantial differences between MMNL and LC models, especially for travel time. The behavioural response to changes in travel time in LC model is far less than the one in MMNL. They also plotted choice probabilities for each respondent's choice set in the MMNL and LC models. Based on these graphs they concluded that there is a relatively weak relationship between the predicted choice probabilities under the MMNL model and the LC model. In other words, they concluded that each model is representing choice responses quite differently for the majority of the sample. Finally, they have stated that each of the two models has its own merits (Greene & Hensher, 2003):

> "*The latent class model has the virtue of being a semi parametric specification, which frees the analyst from possibly strong or unwarranted distributional assumptions about individual heterogeneity. The mixed logit model, while fully parametric, is sufficiently flexible that it provides the modeler a tremendous range within which to specify individual, unobserved heterogeneity. To some extent, this flexibility offsets the specificity of the distributional assumptions*."

However, despite the stronger statistical support overall for the LC model, they did not conclud the LC model was superior to the MMNL model.

Walker and Li (2007) employed latent class choice models to represent the effect of "lifestyle" on household location choice. They used the data gathered with an SP survey with five alternatives in each choice task. For the class membership function, they specified a logit equation with socio-economic characteristics of individuals as explanatory variables. They used various individual characteristics to identify different classes across the population, such as the information on household structure, employment, the age of the head of household, and resources available to the household (measured in terms of income).

For determining the number of classes, the authors estimated a series of models with different numbers of classes. They used BIC, AIC, and rho-bar-squared statistics to compare different models. These statistics that weigh model fit against model parsimony or number of parameters, are:

1- The AIC (Akaike Information Criterion), that is equal to -2*(LL(β)-K). Where K is the number of parameters estimated by the model and LL(β) is the final log-likelihood value (Joan Leslie Walker, 2001).

2- The rho-bar-squared, that is so similar to AIC and is equal to 1-(LL(β)-K)/LL(0) where LL(0) is the log-likelihood of the initial model with no parameters.
3- The BIC (Bayesian Information Criterion), that is widely used for comparing the latent class models and is equal to (-2*LL(β) + K/LN(N))/N where N is the number of respondent (not the number of choice situations).

The authors generated latent class models with 2, 3, and 4 classes as well as an MNL model. Comparing the results of the latent class models and the MNL model showed the superiority of the latent class models, based on all the statistics mentioned above (Joan L Walker & Li, 2007). However, among the latent class models, the results showed that the model with two classes had the largest BIC, while 4-class model had the largest AIC value. Finally, the authors selected the latent class model with three classes as it revealed great advantages in terms of interpretation of estimated parameters in class membership functions and class-specific choice models.

Also, the authors applied another method for comparing the different classes in the latent class model, referred to as the “importance rating” of variables for each class. In this method, the ten most important variables were determined by taking the difference between the highest and lowest value of each variable, as observed in the dataset, and multiplying this difference by the estimated coefficient for that variable (Joan L Walker & Li, 2007). Afterwards, the variables were ranked based on this rating that demonstrates the potential impact of the variable on the utility function. Then, based on the most important variables in each class, the authors made inferences on the lifestyle of individuals in each class. For example, they concluded that Class 1 was oriented towards a suburban, auto-oriented lifestyle because variables “larger residence”, “off-street parking” and “single house on a lot” were the most important ones in Class1. In addition, Class 2 is transit-oriented lifestyle, which is indicated by the presence of travel time to work by transit as the most important variable.

Furthermore, the results of the class membership function were used to examine whether the socio-economic characteristics were good predictors for the latent lifestyle classes or not. For example, they observed that, based on the estimated coefficients in the class membership function, households in Class 1 were affluent, more established and professional families which is consistent with the most important variables in the utility function of this class. In other words, they expected that affluent, more established and professional families were more suburban and auto-oriented. The overall results of Walker and Li's research show the potential of latent class models unobserved heterogeneity across population.

Teichert et.al. (2008) used latent class models to identify segments across airline passengers. They also divided passengers into two categories: Business and Economy passengers and made two logit models on each category. Afterwards, they compared the results of these models with the latent class models.

Based on BIC statistic, the authors selected the latent class model with five classes among a series of models with 2,3,4,5 and 6 classes. Based on statistical and contextual considerations, they described each segment and interpreted it from an airline's marketing perspective. They selected the first three coefficients with the highest t-value in each segment. Based on these coefficients each segment was explained and characterized. For example, as the punctuality, flexibility and schedule are the most significant attributes in the first class, this class was labeled as "Efficiency/Punctuality".

After labeling the classes, the authors argued that using coefficients in the utility function of each class does not directly deliver information about the composition of the class. Hence, they used two procedures for gaining class based insight: First, they assigned each customer to the class in which he/she had the highest membership probability. Secondly, they used socio-demographics of passengers in each to generate class-based demographic and attitudinal profiles. Based on the profiling process were able to describe the individuals in each class based on their demographic characteristics. For example, passengers in first class were identified as mainly frequent flyers who fly several times per week and were predominantly male. Such inferences depict the strength of the LC model in capturing different behaviours and segments across population. In addition, the results of latent class model in their study demonstrated that the attitudes and socio demographic profiles led to a specific set of preferences in each class.

Shen (2009) investigated the difference between the LC and MMNL models for predicting transportation mode choice based on two SP surveys in Japan. Regarding the MMNL model, all time-associated attributes were defined as random parameters and a Normal distribution for them was selected. However, the results showed a positive value for VTTS, which seems counter-intuitive at first blush. However, the author justified it based on the results from other similar studies such as (Redmond & Mokhtarian, 2001). Redmond and Mokhtarian found that in some cases, an individual is drawn toward longer time associated attributes due to some individual specific reasons, and in such cases the positive sign of time-associated attributes could be possible. This result is very important, as it shows the strength of the LC models to identify behaviours that are surprising at first, but in fact, reflect actual behaviour of groups of individuals in the population that researcher is not aware of.

Shen (2009) also mentioned that three segments of low, medium and high VTTS can be found in the LC model. Moreover, the author compared the choice elasticities of different attributes of LC and MMNL models and concluded that the elasticity of in-vehicle time and travel cost are apparently different between the MMNL and the LC models.

For determining the number of classes, Shen used the AIC and CAIC (Consistent AIC) criteria. The CIC formulation in his study was defined as: -2[LL(β) $-S.K_s + (S-1).K_c$]. Where K_s is the number of parameters in the utility functions of choice model, K_c is the total number of parameters in the class membership function, N is the number of observations in the sample and S is the number of classes in LC model. This formula was suggested by Louviere et al. (Louviere et al., 2000) and is a bit different from the one used by Walker and Li (2007), as mentioned above. Also, the CAIC, which is a variant of the AIC is defined by: 2LL(β) $-$ $[S.K_s + (S-1).K_c - 1][Ln(2N) + 1]$.

Hess et al. (2012) have applied a LC structure to examine whether different underlying choice paradigms work well on a given data set. In their modelling approach, individual classes behaved based on different underlying paradigms. Hence, the difference across classes arose not just from the use of different parameters, but also from different underlying decision paradigms. The results of their study showed the potential of the latent class framework to investigate the source of heterogeneity within a population caused by different decision protocols, as well as taste variation.

Recently, Bujosa et al. (2010) have applied a latent class random parameter multinomial logit model for capturing unobserved heterogeneity in recreational demand. They used Revealed Preference data for recreational trips in Spain. They incorporated different site attributes (such as picnic areas or hiking trails, kilometers of roads, distance to coast, etc.), travel cost and environmental attributes in the utility function of alternatives. In addition, they gathered visitor socio-demographic data, such as income, age and place of residence, and used them in the class membership function of the LC-MMNL model. For determining the number of classes they used BIC for comparing LC-MMNL models with different number of classes.

They tried different distributional forms, such as Normal, Lognormal and Uniform distributions and applied the Normal distribution for random parameters in their model. The results of their study demonstrated that both LC and LC-MMNL models can reveal two behavioural classes in the dataset. They concluded that the LC and LC-MMNL models outperformed the MMNL in terms of goodness of fit and in-sample validation. Finally, they stated that LC-MMNL model shows superiority over all models investigated in their study.

Greene and Hensher (2013) also applied a random parameter latent class structure for capturing unobserved heterogeneity in choice modelling. They used SP data gathered from freight transporters in Sydney. Each respondent chose between two alternative freight trips characterized by different levels of travel time and travel cost. They argued that a large number of classes will cause estimated coefficients to be imprecise. In addition, they stated that generating latent class models with large numbers of classes can result in large estimated standard errors for the coefficients of the model.

With respect to labeling of classes, they believe that it is hard to put a name on classes, for classes themselves are indeterminate. Instead, they used numbering for naming the classes of their model, such as class 1, class 2, and etc. The results of their model showed the superiority of LC-MMNL model over LC and MMNL models based on BIC index. Another issue mentioned in their study was that while all the coefficients in MMNL model were statistically significant, this was not the case for LC and LC-MMNL models, as the results showed that some coefficients of the LC and LC-MMNL models are significant only in one class.

Moreover, they observed that only one parameter could be considered as a random parameter. In addition, this random parameter was significant only in one class. When comparing the estimated coefficients between two classes, they found that individuals in class 1 focused on trade-offs between total time and total cost, while the respondents in class 2 focused on the trade-off between on time delivery total cost.

Finally, they concluded that there is a need for applying LC-MMNL model structure on different data sets to ensure the superiority of this model over LC or MLM model, regardless of data set characteristics. The authors mentioned that in their current study they suspected the gain from inclusion of random parameters in LC model was data-set specific.

In this section, I reviewed the most dominant studies on MMNL, LC and LC-MMNL models. However, to my knowledge no study has been done for capturing the agent effect in LC-MMNL model context by using the formulation in Equation (17) to (25). So, the aim of this study is to capture the agent effect in latent class mixed multinomial logit context by implementing the mathematical formulation discussed in above mentioned equations on a SP dataset.

Manuscript: A Generic Form for Capturing Unobserved Heterogeneity in Discrete Choice Modelling: Application to Neighborhood Location Choice

Context

This section includes a co-authored paper that builds on a combined approach accounted for unobserved heterogeneity in discrete choice models. The paper extends the methods used in discrete choice family of models for capturing the agent effect as well as other source of heterogeneity that are controlled for by mixed logit and latent class models.

The method, proposed in this paper can be used by researchers to generate more precise predicts of individuals' behaviours while using SP and RP surveys. Also, the method can identify different segments across population and explain the behaviours and preferences of individuals who have high membership probability in each segment.

In the following manuscript, submitted to Transportation Research Board's 96th annual meeting, I had the role of lead author. Bilal Farooq and Zachary Patterson, the paper's second and thirds authors, formulated the research question and analysis methods. Also, Ali Rezaei, the forth author, at the first stages of the study, analysed the models specification.

The findings indicate that considering the generic form proposed by this study for analyzing heterogeneity across population has great potential to detect unobserved heterogeneity due to individual's characteristics as well as agent effect. Also, the method is able to identify latent behaviours, such as considering house price and shopping time as positive utility, which cannot be captured by previous models in the literature and used in this study.

The results of this study also can be considered by researchers designing SP surveys with the aim of generating latent class models on the results of the survey. As mentioned in the conclusion of the following paper, researcher should consider more detailed attributes and variables in experimental design step of a SP survey. Failing to define and include such detailed attributes in SP surveys, can cause latent class models not be able fully identify the behaviours of different segments of the population.

Finally, the agent effect mixed multinomial logit model in this study shows superiority over previous models in the literature. Capturing temporal heterogeneity (agent effect) as well as behavioural heterogeneity with the formulation proposed in this study can be considered as a contribution to the literature.

The paper appears below as it was submitted for consideration to be presented at Transportation Research Board's 96th annual meeting on 1st August, 2016.

Abstract

This article extends the latent class mixed multinomial logit model to accommodate agent effect through the use of random parameters. Three types of models, Mixed Multinomial Logit (MMNL), Latent Class Mixed Multinomial Logit (LC-MMNL) and Agent Effect Latent Class Mixed Multinomial Logit (AGLC-MMNL) have been generated and the results compared. Considering agent effect simultaneously with other sources of unobserved heterogeneity in a latent class context demonstrates improvement in terms of model fit as well as cross section validation. It enables us to generate a latent class model with more number of classes explaining more heterogeneity across population in a neighborhood location choice study. The AGLC-MMNL model is able to detect four distinct classes of individuals in Montreal, exhibiting different behaviours while facing neighborhood location choice resulting from a stated preference survey. Not only different behavioural classes have been identified by the model, but also some unexpected behaviours, such as preferring more expensive houses or considering shopping time as positive utility, in one class observed, which have been explained in detailed. The classes of the model are able to explain different behaviours of individuals based on their income level, whether they are transit or car oriented, and the importance of privacy in their point of view.

Key words: Agent effect mixed latent multinomial logit model, Unobserved heterogeneity, Stated preference survey, Neighborhood location choice

Introduction

Discrete choice models and their strength to predict individual's choices mostly depends on the quality of datasets which have been used for model generation. However, even the most comprehensive and detailed datasets are not able to observe all factors pertinent to someone's choice. This issue in the choice modelling literature has been addressed as unobserved heterogeneity, which means that individuals across populations are not affected identically by alternative attributes. Furthermore, such variation in preferences across populations and their sources are not always recognized by researchers.

One of the issues documented in a number of studies is the failure to consider unobserved heterogeneity in choice models, inherently leading to biased and inefficient parameter estimation (Heckman & Singer, 1984; Reader, 1993). This can consequently result in incorrect inferences and predictions. There are various approaches in the literature to account for unobserved heterogeneity, among which the random-parameter approach is the most well-known and widely used. Random parameter discrete choice models, also known as Mixed Multinomial Logit (MMNL) models (McFadden & Train, 2000) or Logit Kernel (M. Ben-Akiva et al., 2001), allow the parameters of the model to vary across observations (M. E. Ben-Akiva & Lerman, 1985). Some parameters of the MMNL model can be estimated as random values with a user-defined distribution (e.g. Normal, Triangular or Uniform distributions) and then a mean, as well as standard error are estimated for them. Although MMNL models try to explicitly consider heterogeneity in the population, they assume a single average behaviour and deviations from it across population. In other words, although sources of heterogeneity can be explained by MMNL models by considering the characteristics of individuals, they force the model to estimate only an average value, for the population, of a random variable and then the variance around it. This limitation of MMNL models has resulted in an alternative model, referred to as Latent Class (LC) model in the literature (Greene & Hensher, 2003). LC models relax this constraint in the sense that there are multiple average behaviours. Different from the MMNL model, that defines random parameters as a continuous probability distribution, LC models provide a discrete number of classes that account for unobserved heterogeneity across a population. By means of estimating discrete number of classes, LC models provide multiple and totally distinct average behaviours among individuals.

Besides the two mentioned approaches for accommodating unobserved heterogeneity, recently a combination of these two approaches has also been adopted in the literature referred to as the random parameter latent class model. Within this approach, also known as the Latent Class Mixed Multinomial Logit (LC-MMNL) model (Hensher & Greene, 2003), first the number of latent classes are specified and then the parameters are allowed to be varied (based on user-defined probability distributions) within each class. The advantage of this combined approach over LC model is that it not only considers multiple average behaviours across a population, but it also estimates multiple deviations from these average behaviours.

Another source of unobserved heterogeneity in datasets is the correlation across sequential observations (choices) (Mannering et al., 2016) in time series data. By using such type of data, individual related unobserved factors that are persistent over time would be another source of unobserved heterogeneity. This type of unobserved heterogeneity is known as an “agent” or “panel” effect in the literature (Bierlaire, 2014).

While the MMNL, LC or LC-MMNL approaches account for unobserved heterogeneity due to individual related unobserved characteristics across observations, the agent effect approach tries to capture the unobserved heterogeneity due to individual related unobserved factors across individuals. The intent of this study is to combine the LC-MMNL approach and agent effect approach to allow analysts to make more accurate inferences. Our paper begins with a quick review of the literature on random parameters and latent class discrete choice models. The paper then moves on discussion the methodology and formulation of the model for combining random parameter latent class and agent effect approach. Afterwards, the dataset on which the models have been generated is explained. The paper concludes with the results of the generated models and discussing them.

Literature Review

MMNL and LC models have been applied in various contexts such as route choice models (Greene & Hensher, 2003), mode choice(Atasoy et al., 2011; Hosoda, 1999), environmental preferences (Morey et al., 2006) and household location choice (Joan L Walker & Li, 2007), to name but a few. Greene and Hensher (Greene & Hensher, 2003) applied MMNL and LC models on a Stated Preference (SP) dataset in which respondents choose among three road types (2-

lane, 4-lane without a median and 4-lane with a median) for long distance travels by car in New Zealand. A triangular distribution has been selected for the random parameters in MMNL. The authors generated a series of LC models with 2, 3, 4 and 5 classes, and among them the three-class model demonstrated the best fit. They mentioned that since the mixed logit and latent class models are not nested the comparison on a likelihood ratio test is not appropriate. In addition, the authors mentioned that comparing the absolute parameter values across models is not informative because of scale differences. Comparing the implied Values of Travel Time Savings (VTTS) across the three latent classes showed high degree of difference among the classes. Overall, they concluded that the LC model has revealed three apparent segments for the mean estimates of VTTS.

Walker and Li (2007) employed LC models to represent the effect of lifestyle on household location choice. They used a SP dataset with five alternatives in each choice task. In order to determine the number of classes, the authors generated LC models with 2, 3, and 4 classes as well as a model without segmentation. Comparing the results of the latent class models and the model without latent segmentation showed the superiority of the latent class models, based on BIC index (Joan L Walker & Li, 2007).

Shen (2009) compared LC and the MMNL models for transport mode choice based on two SP surveys in Japan. With respect to LC model, three segments of low, medium and high VOTS have been found in each LC model. Also, the choice elasticities, due to in-vehicle time and travel cost, was apparently different between the MMNL and the LCM. Another finding of this study was identifying a positive value for VTTS, which seems counter-intuitive at the first sight. However, the author has explained and justified it based on the results from other similar studies. As Redmond and Mokhtarian (Redmond & Mokhtarian, 2001) stated, in some cases an individual is more toward the longer commute time as it offers some benefits (such as a transition between home and work). Hence, time-associated attributes should not be considered as an unequivocal source of disutility in discrete choice models.

Recently, Bujosa (2010) applied LC-MMNL model accounted for heterogeneity in recreational demand. They used Revealed Preference (RP) data for recreational trips in Spain. The results demonstrated the heterogeneity among individuals' preferences for environmental attributes related to socioeconomic characteristics, specifically the LC-MMNL model revealed two behavioural classes in the data set. Furthermore, they compared the LC-MMNL model with a MMNL and LC model and concluded that, in general, the LC-MMNL model shows superiority over all models, based on goodness-of-fit and in-sample validation.

Greene and Hensher (2013) also applied random parameter latent class structure for capturing heterogeneity in choice modelling. They used panel data resulted from a SP survey in Sydney among freight transporters. Each respondent chose among two alternative freight trips characterized by different levels of travel time and travel cost. The results showed the superiority of LC-MMNL model over LC and MMNL model based on the BIC index and parameter estimation.

As mentioned earlier, using time-series data can cause another source of heterogeneity referred to as agent or panel effect. It is worth mentioning that SP and RP datasets can be gathered via two different approaches: times-series and cross-sectionally. Time-series data are the sequence of observations of individuals at two or more time points. Cross-sectional data, on the other hand are the observations of individuals at the same point in time (Bierlaire, 2014), which means that no sequence of observations for an individual have been gathered. Also, it should be mentioned that, capturing the unobserved heterogeneity due to panel effect can be done via two methods: first, by allowing random coefficients of variables in utility functions to be correlated amongst the observations common to each respondent. Second, by including a random term in utility functions and allowing it to be correlated among all observations common to each individual. To our knowledge, only Hensher and Greene (2013) considered panel effect in their LC-MMNL model via the first above mentioned method.

While there is a large body of literature on MMNL as well as LC models, the combination of them, i.e. LC-MMNL models, is still considered as new field of research. Hence, there is a need for applying LC-MMNL model structure on different data sets to ensure that the superiority of this model (over LC and MMNL models) is not data set specific. Also, to our knowledge no study has been done yet that accounts for agent effect via the above mentioned approach. So, the aim of this study is to capture the agent effect in latent class mixed multinomial logit model via the second approach, which needs new formulation of LC-MMNL model, that will be explained in the next section.

Methodology

In the discrete choice family of models, it is usually assumed that a sampled individual faces a choice amongst a set of alternatives in each of T choice situations. An individual n is assumed to consider the full set of offered alternatives and to choose the alternative with the highest utility. Denoting X_{ni} as a $K \times 1$ vector of attributes of alternative i in choice situation t and β_s as a $1 \times K$ vector of estimated parameters in class s associated with X_{nit}, the utility for alternative i in class s may be expressed as in the LC-MMNL(Bierlaire, 2013):

$$U_{ni|s} = \beta_s X_{ni} + \nu_{ni|s} \quad (1)$$

Where:

$$\beta_s = \bar{\beta}_s + \sigma_s \xi \quad (2)$$

where $\bar{\beta}_s$ and $\sigma_s$are the mean and variance of the β_s, i.e. $\beta_s \sim N\left(\bar{\beta}_s,\ \sigma_s\right)$, $\nu_{ni|s}$ is random term with zero mean and iid distribution, and ξ is a random value with a specified distribution. Hence, the utility function of alternative i observed by individual n can be expressed as:

$$U_{ni|s} = \bar{\beta}_s' X_{ni|s} + \sigma_s' \xi X_{ni|s} + \nu_{ni|s} \quad (3)$$

The conditional probability of alternative i in class s being chosen by individual n is:

$$P_n(i|\xi, s) = \frac{\exp(\bar{\beta}_s' X_{ni|s} + \sigma_s' \xi X_{ni|s})}{\sum_{j \in J} \exp(\bar{\beta}' X_{nj|s} + \sigma' \xi X_{nj|s})} \quad (4)$$

The class heterogeneity is defined by class membership function:

$$M_{n|s} = \exp \frac{(\gamma_s Z_n)}{\sum_{s=1}^{S} exp(\gamma_s Z_n)} \quad (5)$$

Where Z_n is a vector of class variables consisting of individual socioeconomic characteristics and γ_s is a vector of parameters for class s. The contribution of individual i to the log-likelihood of the model is obtained by integrating the two-layer heterogeneity: within class heterogeneity and the class heterogeneity (Greene & Hensher, 2013). The unconditional probability of alternative i in class s being chosen by individual n in LC-MMNL model is given as follows:

$$P_n(i) = \sum_{s=1}^{S} [M_{n|s} \times \int_{\xi} P_n(i|\xi, s) \,.\, f(\xi) .\, d\xi] \quad (6)$$

Although the LC-MMNL accounted for unobserved heterogeneity across observations, it does not consider the heterogeneity across individuals when using panel (multiple observations from the same respondent) data. While using panel data, each individual faces more than one choice situation at different points of time. This can cause observations from the same individual to be correlated. The source of such correlation is individual related unobserved factors that are persistent across time (Bierlaire, 2014). Hence, addressing the source of this correlation in the LC-MMNL model, first we add subscript t to the utility function:

$$U_{nti|s} = \beta_s X_{nti} + \nu_{ntis} \quad (7)$$

Here $U_{nti|s}$ is the utility for alternative i being chosen by individual n in class s at time point t. Afterwards we relax the assumption that ν_{ntis} are independent across t:

$$\nu_{ntis} = \alpha_{ni} + \nu'_{ntis} \quad (8)$$

while ν'_{ntis} is independent across t, α_{ni} capture permanent taste heterogeneity.

Consequently, the conditional probability (on ξ, s and α_n) of alternative i being chosen by individual n in class s at time t is:

$$P_{nt}(i|\xi, s, \alpha_n) = \frac{\exp(\bar{\beta}'_s X_{nti|s} + \sigma'_s \xi X_{nti|s} + \alpha_{ni})}{\sum_{j\in J} \exp(\bar{\beta}'_s X_{ntj|s} + \sigma' \xi X_{ntj|s} + \alpha_{nj})} \quad (9)$$

where α_n is the vector gathering all parameters α_{ni}. Removing the probability conditions on ξ and s we will have:

$$P_{nt}(i|\alpha_n) = \sum_{s=1}^{S} [M_{n|s} \times \int_\xi P_{nt}(i|\xi, s, \alpha_n) \,.\, f(\xi).\, d\xi] \quad (10)$$

Also, the contribution of individual n to the log likelihood is:

$$P_n(i_1, i_2, \dots, i_T|\alpha_n) = \prod_{t=1}^{T_n} P_{nt}(i|\alpha_n) \quad (11)$$

where $T_n \geq 1$ is the number of choice situations in which individual n is engaged. In addition, the choice probability of alternative i unconditional on α_n would be:

$$P_n(i_1, i_2, \dots, i_T) = \int_\alpha \prod_{t=1}^{T_n} P_{nt}(i|\alpha_n)\, f(\alpha) d(\alpha) \quad (12)$$

where $f(\alpha)$ is the probability density function of α. Finally, the unconditional contribution of individual n to the log likelihood is:

$$P_n(i_1, i_2, \dots, i_T) = \int_\alpha \prod_{t=1}^{T_n} [\sum_{s=1}^{S} [M_{n|s} \times \int_\xi P_{nt}(i|\xi, s, \alpha_n) \,.\, f(\xi).\, d\xi]]\, f(\alpha) d(\alpha) \quad (13)$$

Solving the above equation for large choice sets needs simulation by generating draws $\xi^1, \xi^2, \dots, \xi^R$ from $f(\xi)$ for each random coefficient in each class utility function, and $\alpha^1, \alpha^2, \dots\dots \alpha^R$ from $f(\alpha)$ for capturing agent effect that is the same in all choice situations with which an individual will be faced.

Equation (13) can be approximated by:

$$P_n(i_1, i_2, \dots, i_T) \approx \frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T_n}[\sum_{s=1}^{S}[M_{n|s} \times \frac{1}{R}\sum_{r=1}^{R} P_{nt}(i|\xi^r, s, \alpha^r)\ ]] \tag{14}$$

Also, the overall simulated log-likelihood is given by the following equation:

$$\log L = \sum_{n=1}^{N}\log[\frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T_n}[\sum_{s=1}^{S}[M_{n|s} \times \frac{1}{R}\sum_{r=1}^{R} P_{nt}(i|\xi^r, s, \alpha^r)\ ]]] \tag{15}$$

where N is the number of individuals in data set.

Dataset

With the aim of comparing the MMNL, LC-MMNL and Agent Effect Latent Class Mixed Multinomial Logit (AGLC-MMNL) a DCE dataset collected for a neighborhood location choice study in Montreal, Canada (Patterson et al., 2017) was used. The survey was developed and analyzed in a manner conforming with the standards in the field (Louviere et al., 2000). As such, survey development included secondary research and focus groups to establish attributes (and their ranges), as well as the piloting of the surveys.

The survey employed for data gathering was a multimedia version administered through a virtual reality platform developed with Unity (Patterson et al., 2017), where respondents received supplementary textual information as they navigated a residential street in a virtual environment.

Regarding the survey structure, each choice task had two alternative neighborhoods that were characterized by the following attributes: dwelling type (apartment, detached houses, townhouses, and triplexes); front yard depth (6 feet, 9 feet), space between buildings (no space, 20 feet); average home value (low - 20% below base price; medium and high - 20% base price); travel time to work by car (20, 35, 50 minutes); travel time to work by transit (5% below, 30% above travel time to work by car); and finally, travel time to nearby shops on foot (5, 15, 25 minutes).

The socio-economic characteristics of respondents were also gathered. Respondent were asked about their income level. Afterwards, respondents were put into four categories based on their income: low−income (annual income below 20,000 CAD), medium−income (annual income above 20,000 and below 90,000 CAD), high−income (annual income above 90,000 CAD), and non-represented income. We also asked respondent whether they were Francophone or Anglophone as well as their employment status: full time employment, part time employment, unemployed, retired, student and student with part time job. Another characteristic asked from respondents was the type of their current residence: apartment, townhouse, triplex and detached house. Having driver license or not and whether the respondents reside in urban core were another characteristics were also used.

Survey attributes and respondent characteristics have been summarized in TABLE1. Also, TABLE2, presented below, summarizes the characteristics of the sample used to estimated discrete choice models. The surveys were administered at coffee shops in June 2013 and also in February 2014. The sample available for estimation contains 1,104 observations collected from 184 respondents.

Results

Sequence of models was generated to reach the preferred MMNL, LC-MMNL and AGLC-MMNL models. First, regarding the identification of random taste variation, several MMNL models with Normal, Triangular and Uniform distribution were generated and the MMNL model with Normal distribution demonstrates the best overall fit.

Regarding LC-MMNL and AGLC-MMNL models, series of models with two, three, four and five classes were generated. We used three criteria in this step to select among models with different numbers of latent classes: first, the model should be identified based on the smallest singular value of the second derivatives matrix of log-likelihood function (M. E. Ben-Akiva & Lerman, 1985; Bierlaire, 2015). Second, the model should not have any estimated parameters, in the utility functions that are insignificant in all of the classes. Third, among the models that satisfied the first and second criteria, the model with the smallest Bayesian Information Criterion (BIC) was selected. We have used the BIC index with the following formula (Greene & Hensher, 2010, 2013):

$$\text{BIC} = \frac{\left[-2Ln(L)+\frac{K}{Ln(N)}\right]}{N} \tag{16}$$

TABLE 1 Summary of Survey Attributes and Respondents Characteristics

Attributes	Levels
Dwelling type in neighborhood	**a)** Single detached Houses **b)** 2-storey Townhouses **c)** Triplexes **d)** 3-Storey Apartments (6 or 8 units)
Space between buildings	**a)** No space
	b) 20 feet
Front yard depth	**a)** 9 feet (specific to Triplex dwellings)
	b) 6 feet deep (for all dwelling types except triplexes)
	c) 25 feet deep
Travel time to work by car	**a)** 20 minutes
	b) 35 minutes
	c) 50 minutes
Travel time to work by public transit	**a)** 18 minutes **b)** 25 minutes (when travel time to work by car was 20 minutes)
	a) 30 minutes **b)** 45 minutes (when travel time to work by car was 35 minutes)
	a) 50 minutes **b)** 65 minutes (when travel time to work by car was 50 minutes)
Travel time to nearby shops on foot	**a)** 5 minutes
	b) 15 minutes
	c) 25 minutes
Average home value	**a)** % 20 below base price
	b) Base price
	c) % 20 above base price
Socio-economic Characteristics	
High Income	**a)** Above 90,000 CAD
Medium Income	**a)** Above 20,000 and below 90,000 CAD
Low Income	**a)** Below 20,000 CAD
Non-represented Income	**a)** Individuals who did not state their income level
Employment status	**a)** Full time **b)** Part time c) Unemployed **d)** Student **e)** Student with part time job
Language	**a)** French **b)** English
Driver	**a)** Having driver license **b)** Others
Urban Core Residency	**a)** Reside in Urban Core **b)** Others
Current Residency	**a)** Apartment **b)** Townhouse **c)** Triplex **d)** Detached house

TABLE 2 Descriptive Statistics of the entire sample

Characteristics		Share
Gender	Female	41%
	Male	59%
Age	34 and under	34%
	35-44	25%
	45-54	23%
	55-64	16%
	65 and above	1%
Employment status	Full-time	71%
	Part-time	10%
	Student and employed	10%
	Student	4%
	Unemployed	4%
Current dwelling type	Apartment	33%
	Townhouse	17%
	Triplex	6%
	Detached house	44%
Urban Core Residency	Reside in Urban Core	19%
	Others	81%
	High Income	32%
	Medium Income	54%
	Low Income	6%
	Non-represented Income	8%
Language	French	56%
	English	44%
Driver	Driver	59%
	Others	41%

where N is the number of observations in the dataset, K the number of parameters and L the final likelihood of the model. There are other fit measures in the literature for evaluating the discrete choice models, such as $\overline{\mathrm{R}}^2$, $\mathrm{Pseudo}-\overline{\mathrm{R}}^2$ or AIC with the following formulations (Greene & Hensher, 2010):

$$\overline{\mathrm{R}}^2 = 1 - \frac{Ln(L)}{Ln(L_{null})} \tag{17}$$

$$\mathrm{Pseudo} - \overline{\mathrm{R}}^2 = 1 - \frac{Ln(L) - k}{Ln(L_{null})} \tag{18}$$

$$\mathrm{AIC} \; = \frac{[-2Ln(L) + 2k]}{N} \tag{19}$$

Where $\mathrm{L_{null}}$ is the likelihood of an initial model, i.e. a constant only model (Greene & Hensher, 2010). However, as Greene and Hensher (2010) have discussed $\overline{\mathrm{R}}^2$ and $\mathrm{Pseudo} - \overline{\mathrm{R}}^2$ mainly compare the improvement in the final log likelihood of the model to a constant only model (Although $\mathrm{Pseudo}-\overline{\mathrm{R}}^2$ decreases when the number of parameters increase). Considering the fact that the log likelihood function, itself, measures the improvement of model fit whenever a variable is added to the model, we can conclude that these statistics, i.e. $\overline{\mathrm{R}}^2$ and $\mathrm{Pseudo}-\overline{\mathrm{R}}^2$, only do a transformation between zero and one. Obviously, this transformation is worth noting as it only strict the measures to be between zero and one. In other words, the final log likelihood of models can be compared directly to each other, without any need for such a transformation. In addition, even when a model predicts the data perfectly, neither $\overline{\mathrm{R}}^2$ nor $\mathrm{Pseudo}-\overline{\mathrm{R}}^2$ can achieve one (Greene & Hensher, 2010). But, AIC and BIC only consider the final log likelihood of the model as well as the number of observations and the number of parameters (Greene & Hensher, 2010).

It should be mentioned that the larger the number of parameters and observations, the better fit of model will results. Hence, by considering the number of parameters and observations, these indices reward a model with a lower number of parameters and observations (Greene & Hensher, 2010). However, we couldn't find any preference toward BIC or AIC in the literature. Some studies (Greene & Hensher, 2010) have used them simultaneously, and some of them (Greene & Hensher, 2013) have used only BIC or AIC . In the current study, we have used BIC for comparing the models with different number of classes and parameters.

The results of different LC-MMNL and AGLC-MMNL models are shown in TABLE3[1]. In the series of LC-MMNL models, the model with two classes was identifiable and also seemed to be a better model in terms of estimated values of parameters and their level of significance. LC-MMNL models with three, four and five classes not only had a considerable number of parameters that were highly insignificant in all classes (significant levels less than %20), but also were unidentifiable. We tried to improve these models by removing the highly insignificant variables from their utility function. However, the resulting models neither improved (in terms of level of significant of parameters) nor turned out to be identifiable. The reasons for this

[1] The complete results of all generated models in TABLE 3 have been presented in Appendix A.

observation are unclear other than there might be correlation amongst the observations related to each individual which cause the estimated parameters of these models, i.e. LC-MMNL models with more than two classes, to be insignificant.

Among AGLC-MMNL models, the ones with two, four and five classes were identifiable. However, based on the BIC value and the significant level of estimated parameters, the model with four classes demonstrates the best model fit. Hence, for the aim of comparison, we selected LC-MMNL model with two classes and AGLC-MMNL model with four classes.

It should be mentioned that in addition to testing the number of classes, different random distributions, i.e. Normal, Lognormal (for home value and travel time attributes), Triangular and Uniform were tested on LC-MMNL and AGLC-MMNL model. However, as with the MMNL model, the Normal distribution resulted in the best model fit.

The results of estimated models are shown in TABLE 4. Comparing the three models based on their BIC statistics suggest that the AGLC-MMNL model is a statistically significant improvement over the others, given the additional 29 parameters, compared to the LC-MMNL model and 42 extra parameters over MMNL model.

As it is clear in TABLE 4, there are some coefficients of the models that have been removed from one or more classes. The reason for removing such coefficients is that in the modelling procedure, they turned out to be highly insignificant or including them caused model fit to get worse. Hence, we decided not to include these coefficients in the model.

In the AGLC-MMNL model, permanent taste heterogeneity was captured by including a random error parameter (Sigma), which is the same in all utility functions of an individual in all choice situations. The estimated value for Sigma in AGLC-MMNL model is significant at 62% level (TABLE 4). The rather low significant level of Sigma can be explained by the lack of repetition (the number of choice situations each individual faced) available in the dataset. However, we believe that its inclusion helps the AGLC-MMNL model to provide better parameter estimation and model fit, as it has captured the permanent heterogeneity in the model, which is not accounted for by the LC-MMNL model.

TABLE 3. Results of LC-MMNL and AGLC-MMNL models with different number of classes

Model	Number of Classes	Final log-likelihood	Number of parameters	BIC	Identifiable/Unidentifiable
LC-MMNL	2	-550.378	24	1.0002	Identifiable
	3	-514.919	41	0.9381	Unidentifiable
	4	-501.218	53	0.9149	Unidentifiable
	5	-492.128	61	0.8994	Unidentifiable
AGLC-MMNL	2	-538.686	25	0.9791	Identifiable
	3	-497.582	42	0.9068	Unidentifiable
	4	- 499.197	54	0.9113	Identifiable
	5	-502.495	62	0.9183	Identifiable

As it has been shown in TABLE 3, the probability of the first and second class in LC-MMNL model are 0.2424 and 0.7575 respectively. In AGLC-MMNL model the probability of class 1 is close to the one of class 1 in LC-MMNL model and then the class 2 in LC-MMNL model has divided into the class 2, 3 and 4 in AGLC-MMNL model. This shows that by incorporating agent effect we were able to explain more heterogeneity across population

Turning next to the detailed results, the MMNL estimates show the expected negative marginal utilities for cost of house and all the travel times. The coefficients for dwelling type are pointed out relative to the category of single detached houses. For townhouses, triplexes and apartments the coefficients bear negative sign which bring to mind respondents would prefer single detached houses to these dwelling types. The result obtained for space is positive as we expected and highly significant.

With respect to front yard depth, the MMNL model was not able to estimate a coefficient with an acceptable level of significance, so we removed it from MMNL model. These results are consistent with what was found in Patterson et al. (32).

TABLE 4 Estimated MMNL, LC-MMNL and AGLC-MMNL models

Attribute		MMNL	LC-MMNL		AGLC-MMNL			
			Class 1	Class 2	Class 1	Class 2	Class 3	Class 4
ASC		-0.0895 (-0.94)	-33.9 (-3.09)	-0.029 (-0.12)	13.9 (6.84)	18.7 (5.73)	-39.2 (-3.40)	-51.9 (-2.52)
Dwelling type	Detached houses	fixed	fixed	fixed	fixed	fixed	fixed	fixed
	Townhouse	-0.429 (-2.08)	-	-1.12 (-1.12)	-70.0 (-6.79)	-13.4 (-3.50)	76.5 (3.84)	4.73 (2.17)
	Triplex	-1.06 (-3.61)	-18.2 (-5.99)	-3.44 (-0.80)	-51.7 (-6.43)	-48.2 (-4.94)	17.2 (5.70)	-68.5 (-2.71)
	Apartment	-1.01 (-4.55)	-	-2.35 (-1.80)	15.0 (5.62)	-86.7 (-5.26)	-0.974 (-0.38)	-
Space between buildings (in feet)		0.0332 (3.09)	-	0.0638 (1.43)	-	2.93 (5.50)	-	-
Front yard depth (in feet)		-	-9.73 (-3.96)	0.0362 (0.64)	-0.710 (-6.08)	-0.062 (-0.62)	2.10 (3.35)	-0.286 (-1.92)
Travel time to nearby shops on foot (in minutes)		-0.0502 (-5.40)	4.62 (4.28)	-0.117 (-1.03)	2.70 (6.73)	-3.18 (-5.44)	-3.99 (-3.50)	-5.04 (-2.44)
Average home value (in thousands CDN)		-0.0062 (-4.92)	-3.22 (-3.22)	-0.005 (-0.79)	0.167 (6.43)	-0.129 (-4.18)	-5.22 (-3.50)	-0.064 (-2.39)
Travel time to work by car (in minutes)		-0.0474 (3.16)	-40.3 (-4.22)	- 0.025 (-0.79)	-10.7 (-6.68)	-1.51 (-4.95)	-	-
Travel time to work by transit (in minutes)		-0.0315 (-3.14)	-5.55 (-4.37)	-0.043 (-1.17)	-2.74 (-6.47)	-0.452 (-5.49)	-5.09 (-3.51)	-1.69 (-2.9)
Variance of random parameters								
Space		0.0419 (2.05)	-	0.114 (1.05)	-	0.794 (6.48)	-	-
Apartment		0.826 (1.01)	-	0.774 (0.89)	25.9 (6.19)	-1.07 (-0.99)	0.0915 (0.35)	-
Triplex		-	-	-2.86 (-1.48)	-0.160 (-0.70)	11.4 (6.63)	0.114 (0.85)	35.2 (2.11)
Travel time to work by car (in minutes)		0.0942 (3.93)	-14.7 (-4.33)	-0.175 (-0.78)	-1.24 (-5.38)	3.24 (5.09)	-	-
Sigma (agent effect parameter)		-	-		-0.371 (-0.87)			
Class Membership Attributes								
High Income		-	-0.347 (-0.80)	-	0.272 (0.67)	-	-	-
Medium Income		-	-	-	-	-	-	-0.307 (-0.82)
Low Income		-	-	-	-	-0.283 (-0.79)	-0.283 (-0.79)	-
Zero or Non-represented Income		-	-	-	-0.961 (-1.31)	-	-	-
Employment Status (Full time =1, other = 0)		-	-	-	-0.238 (-0.82)	-	-	-
Language (French=1, English=0)		-	-0.810 (-3.15)	-	-0.379 (-2.21)	-	-	-
Driver (Having driver license=1, other =0)		-	-	-	-	-	-1.21 (-4.61)	-1.21 (-4.61)
Urban Core Residency(Reside in Urban Core =1, other = 0)		-	-	-	-	-0.882 (-3.14)	-	-0.882 (-3.14)
Current Residency (Apartment =1, other = 0)		-	0.544 (1.70)	-	-	-	0.232 (0.76)	-
Class membership probability		-	0.2424	0.7575	0.2272	0.3793	0.2296	0.1637
Number of parameters		12	25		54			
Final Log-likelihood		-596.68	-550.38		-499.19			
BIC		1.0825	1.0003		0.9113			

* The three models are generated using Pythonbiogeme Software (Bierlaire, 2003).

** The number of draws for estimating all random parameters was set to 10 draws.

***t ratios in parentheses

In the estimates for the LC-MMNL model, a drop in the significance level of some attributes in one class is observed. Hensher and Greene (3) and Bujosa (29) also found a number of statistically insignificant parameter estimates in one class. Indeed, the estimates for "Space" and "Front-yard depth" in class 2 is only significant around the 50% level, with the 65% level applying for the estimate of the variance of "Space". This result could in part be explained by the fact that these attributes lack statistical merit for one class. Furthermore, identifying the source of low significance levels in latent class models, Hess et al. (Hess, Ben-Akiva, Gopinath, & Walker, 2011) has mentioned that the drop in the significant level of some attributes could be partly due to respondents who are largely indifferent between the alternatives in the SP choice data. However, finding the exact source of the drop in the significance level of some attributes of the current dataset requires more research, which is beyond the scope of current study. Also, providing another reason, given the complicated and highly nonlinear form of the LC-MMNL model, in order to estimate coefficients with high level of significance we need enough observations and heterogeneity for each segment of the model which was not always the case in the dataset used.

The estimates for Travel time to work by car and transit are negative in both classes of LC-MMNL. However, the walking time to nearby shops in class 1 bears a positive sign. At first glance this result seems surprising, however considering the dissimilarity between purposes of shopping trips and work or educational trips, we can think of satisfactory explanations strongly support this positive sign for walking time to nearby shops. In fact, as Davis et al. (Davis & Hodges, 2012) have mentioned, it is well-known that shopping is a complex human behaviour driven by a broad spectrum of consumer needs which result in different shopping motivations to fulfill them. Research addressing some of these motivations has mentioned diversion from daily routine, self-gratification and physical activity (Tauber, 1972), to kill time and to enjoy shopping as a social event (Buttle & Coates, 1984), adventure shopping, social shopping and gratification shopping (Arnold & Reynolds, 2003) , to name but a few. In fact, some shoppers do shopping as it helps them to relax, de-stress, or have a break from routine. So, its sign is perhaps not counter-intuitive if we assume some shoppers prefer longer walking time for shopping to make the most of their leisure time.

Regarding average home value in LC-MMNL, the coefficients in both classes have the expected negative sign. The different values of price coefficient between two classes (-3.22 vs -0.005 for class 1 and class 2, respectively) support the presence of different attitudes toward it in the data set.

In the LC-MMNL model, the front yard depth in class 1 and 2 have different sign and also different levels of significance (t-ratios are -3.96 vs. 0.64 for class 1 and 2 respectively), as mentioned above. This suggests that not only this attribute has more statistical merit for latent class 1 (up to a class membership probability), but also a different sign for this attribute in two latent classes show totally different points of view across the population.

In the AGLC-MMNL model, a drop in the significance level of some coefficients is also observed, like the LC-MMNL model. The estimates for travel time to work by car and transit bear negative sign as expected. However, like the LC-MMNL model, the estimate for walking time to nearby shops in class 1 has positive sign. It should be mentioned that class 1 and class 2 in the LC-MMNL and AGLC-MMNL models are not equivalent in any sense, and should not be directly compared per se.

Regarding the estimates for average home value in the AGLC-MMNL model, while this coefficient has the expected negative sign in classes 2, 3 and 4, the corresponding estimate for class 1 has a positive sign. It should be mentioned that during the model generating step, we assumed the average home value attribute as a random parameter with lognormal distribution, to force estimate to be negative. However, the resulting model showed weak results regarding goodness of fit and significance level of estimated parameters, i.e. the model was not identifiable and the estimate for home value and some other parameters, like travel time by transit, were totally insignificant. Hence, we concluded there must have been some reasons why these individuals chose the alternative with the higher price in a choice task. While numerous studies in economic theory of behaviour assume price as an indicator of monetary sacrifice, there is a large body of literature in marketing research that concentrates on price as an indicator of product quality (Völckner & Hofmann, 2007). Indeed, numerous studies state a positive and significant relationship between price and customer perceived quality

(Hinterhuber, 2015). Thus, it seems possible that some decision makers found higher prices more satisfying, for they perceived it as an increases in product quality (and vice versa) that result in positive price elasticity of demand.

In the field of choice modelling, Völckner and Sattler (2005) carried out a study to estimate separate roles of price, i.e. price as an indicator of money sacrifice and price as an indicator of product quality, by using latent class models and a hierarchical Bayes procedure. Their results could predict the positive effect of price in the latent class model context. Hence, the positive sign for value of home in class 1 of AGLC-MMNL model, can be explained by the fact that some individuals, who have high membership probability in class 1, may consider home value as an indicator of its quality and are toward choosing alternatives with higher prices in a choice set.

Also, we can provide another complementary explanation for the positive sign of home value coefficient in class 1. First, it is a well-known phenomenon that wealthy people tend to pay more for a product (Horowitz & McConnell, 2003). Second, considering the fact that in the probability membership function of class 1 the High Income attribute has a positive sign, which implies individuals with high income have higher membership probability in this class, support the reason why a positive sign for home values has been found in this class. In other words, individuals with higher income may be willing to pay more for a product and prefer the more expensive alternative, and they definitely consider the relation between price and quality when making such a decision.

There are some other results that demonstrate the relation between the estimated parameters in the class-membership and utility function of each class. First, class 2 is the only class in which the Space between buildings parameter is significant. Regarding the utility function of this class, we can identify a clear preference toward detached houses comparing other classes. So, not surprisingly, if only individuals who exhibit a definite preference toward detached houses concern about the space between buildings. Second, in class-membership function of class 3 and 4 Driver attribute indicate that individuals without driver license have higher probability in this class. Considering that in these classes Travel time to work by car is not significant we can conclude that individuals without driver license are indifferent to Travel time to work by car.

Regarding dwelling type in the AGLC-MMNL model, the coefficient estimates for triplex and townhouse have negative signs in class 1 while the estimate for apartment in this class has a positive sign. This result is different from the estimates for apartment in MMNL and LC-MMNL model. However, it is not counter-intuitive if we assume some people prefer luxury apartments (or penthouses) in high-rise buildings rather than detached houses. This variation among the estimates for dwelling type is also observable in class 3 and 4, where townhouse and triplex have positive sign, implying their preferred utility over detached house. This observation shows the ability of latent class models to identify unobserved heterogeneity among individuals in a population.

Regarding the semantic meaning of the classes, it is not always possible to put a name on the classes, as these classes are themselves indeterminate (Greene & Hensher, 2013) and explained by their probability function. However, we can use the estimated parameters of the model to find which part of the population has higher class-membership probability in each class. Considering this approach, in class 1 of AGLC-MMNL model, the High Income parameter shows a positive sign and Language pose a negative sign in class-membership function. Also, regarding the utility function of Class 1, as mentioned above Average home value has a positive sign. In addition, Apartment shows a fairly positive sign relative to detached house, which means apartments and detached houses are more preferable in this class. Besides, Travel Time to Work by Car shows a negative sign and its absolute value is higher than the absolute value of Travel Time by Transit. Hence, putting all the things together, we can say that class 1 represents (up to a probability) effluent, car-oriented, luxury apartment and detached house loving Anglophone Montrealers for whom the affordability is not a problem and the house, a status symbol.

Considering the same approach, in class 2, Urban Core Residency and Low Income parameters have a negative sign and Front Yard Depth poses positive sign which can be interpreted as middle class and upper middle class non-urban core residents who highly value privacy have higher class-membership probability in this class.

With regard to class 3, Current Residency has a positive sign and Driver has been estimated as a negative value. Also in this class Travel time to work by transit have a negative sign while Travel time to work by car was totally insignificant. Hence, we can assume that transit-oriented, middle class and upper middle class apartment dwellers have higher probability in this class.

With respect to class 4, Medium Income, Driver and Urban Core Residency have negative sign. Also, like class 3, in the utility function of this class, Travel time to work by transit has a negative sign and Travel time to work by car was not significant at all. So, we can say transit-oriented non-urban core residents whose income level is not in the medium income range have higher membership probability in this class.

Validation

One of the most important goals of estimating a choice model is to predict the future behaviour of population. Hence, predictive performance of a choice model is highly important while comparing different model structures. One of the methods to assess the predictive performance of a model is cross-validation. Cross validation consists of partitioning the data into two datasets: a training dataset and validation dataset. For assessing the model's fit on the validation dataset, a performance measure, usually based on predicted choice probabilities, is used.

The estimation dataset contains 1,104 observations. The MMNL, LC-MMNL and AGLC-MMNL models are then applied to the validation dataset (containing 220 out-of-sample observations) to simulate the choice for each observation. The results of the simulated choices have been summarized in a confusion matrix, demonstrated in TABLE 5. The results of the confusion matrix show that the specification of AGLC-MMNL model is better over the MMNL and LC-MMNL models. The overall rate of the true predicted choices is, 60.0 %, 67.3% and 71.4%, for MMNL, LC-MMNL and AGLC-MMNL models, respectively.

TABLE5. Confusion Matrix of Simulated Choice under MMNL, LC-MMNL, and AGLC-MMNL Models

Model	Confusion Matrix				Overall True prediction rate
	Alternatives		Simulated Choice		
			Alt.1	Alt.2	
MMNL	Actual	Alt.1	69 (61.1%)	44 (38.8%)	60.0%
		Alt.2	44 (41.1%)	63 (58.9%)	
LC-MMNL		Alt.1	77 (66.4%)	39 (33.6%)	65.9%
		Alt.2	31 (34.6%)	68 (65.4%)	
AGLC-MMNL		Alt.1	82 (71.9%)	32 (28.1%)	71.4%
		Alt.2	31 (29.2%)	75 (70.1%)	

Conclusion

This article extends the latent class mixed multinomial logit model to accommodate agent effect through the use of random parameters. Considering agent effect simultaneously with other sources of unobserved heterogeneity in a latent class context demonstrates improvement in terms of model fit as well as cross section validation. The results of confusion matrix for the models show the superiority of AGLC-MMNL model over other two models in simulating and predicting the choices of individuals.

Also, the results of this study shows the strength of latent class models in identifying different behaviours across population, especially in complex choice situations, like neighborhood location choice, while a broad spectrum of variables drive the individuals preferences. Some attributes such as dwelling type were estimated with varying values and signs across classes, which present differing viewpoints of individuals' toward type of dwelling. Also, generating models with more latent classes enabled us to predict some unexpected (but realistic) behaviours, i.e. preferring more expensive homes or considering shopping time as positive utility, which were unrecognizable by MMNL model.

Another finding is that, while using latent class models, the variables being included in the SP surveys should be clearly defined to represent distinctive tastes across population. For example, while there are broader types of apartments in the housing markets, from low-price studios to expensive and luxury penthouses, we had considered only one variable representing apartment in our SP survey. As mentioned above, the results of AGLC-MMNL model showed that effluent individuals in our dataset perceived Apartments almost as preferable as detached houses. It is hypothesized that the apartment considered by that class of individuals were luxury and large apartments or penthouses in high-rise buildings, and seemingly dissimilar to the definition of apartment from our point of view. Hence, it is important to include more precisely and clearly defined variables in SP surveys on which a latent class model will be applied.

Finally, due to the fact that in latent class models the log likelihood will increase when the number of classes increases (Greene & Hensher, 2013), it is not possible to exactly determine the contribution of agent effect to the log likelihood of the LC-MMNL model, as the LC-MMNL models in this study have different numbers of classes. However, it is noticeable that considering the agent effect in this study can help the LC-MMNL model to be identifiable with higher number of classes. An ongoing research question will be to test this model on additional datasets to ensure that these results are not dataset specific.

Conclusion

This thesis set out to extend the latent class multinomial mixed logit model by incorporating the agent effect as a random parameter in the utility function of each alternative faced by an individual. The influence of agent effect on discrete choice models has been studied in the literature by the means of two methods: the first method is widely used when there are random coefficients in the model to capture taste heterogeneity, e.g. mixed multinomial logit models. The draws (from a specified statistical distribution) for simulating and solving the log-likelihood integral will be the same for a random coefficient in all utility functions with which an individual is faced. In other words, in the first method, the agent effect is estimated by the same parameters accounting for heterogeneity due to taste variation. The second method considers a random parameter in the utility function of each alternative with which a given individual is faced.

The main contribution of this study is to extend the formulation of LC-MMNL model to capture the agent effect by implementing the second method. The other sources of unobserved heterogeneity are captured by employing a latent class, as well as a random parameters (mixed multinomial logit model) approach. The results showed that while the LC-MMNL model was able to capture two classes of individuals across the population, the AGLC-MMNL model was able to identify four classes across the same population that can be interpreted as a strength of the agent effect approach in capturing unobserved heterogeneity.

Comparing the results of AGLC-MMNL and LC-MMNL model showed that the former is better in terms of model fit, based on the BIC index. In addition, the AGLC-MMNL model in this study revealed some surprising behaviour among individuals in the population. It should be mentioned that similar unexpected behaviour has also been observed in the literature and can be explained by the results of studies in the relevant field of science. For example, in class 1 of the AGLC-MMNL model, the estimated coefficient for House Price was positive, while we expected it to be negative. The positive sign for price can be explained by the fact that some individuals, likely affluent ones, perceive price as an indicator of quality, and hence, related a positive utility to the price while choosing an alternative. This explanation has long roots in economic theories that observe price not only as a monetary sacrifice, but also as an indicator of product quality.

Limitations

One of the most important issues in generating discrete choice models is the data used for modelling. Gathering precise and detailed data is time consuming and expensive, and needs a great deal of research to identify the best variables to include in discrete choice experiments. The data used in this study was not specifically designed for using latent class models. Hence, the variables in the models were not defined in detail to reflect all taste variation across population. For example, the variable Apartment in this model had not been defined precisely. As we know there are several types of apartments in housing markets, from low-price studios to expensive and luxury penthouses. However, in our dataset all types of apartment have been defined under the label of Apartment. This limitation in the dataset can be partly captured by the AGLC-MMNL model, however, it is obvious that if were to provide the model with more detailed attributes, the model results would be more accurate and efficient.

The small size of the dataset in this study. While a wide range of alternative attributes were included in the model, the sample size was not large enough to provide sufficient variation across this population. This limitation resulted in low significant levels for some estimated coefficients in the model.

The other limitation is that while there were other characteristics of individuals in our dataset, like gender or age, these attributes did not turn out to be significant in any of the models in this study. Many studies have showed that there are significant differences between individuals of different age and gender. However, for reasons unclear for us, these attributes were not significant in our model.

Another limitation of the current thesis is the correlation of systematic part of the utility with the error term, referred to as “endogeneity” in the literature. As mentioned by Guevara and Ben-Akiva (2012), endogeneity in discrete choice models can be caused by errors in variables or omission of variables that are correlated with the observed ones, especially in choice of residential location. There are methods for controlling endogeneity in the literature that can be incorporated in the modelling procedures of the current study.

Also, as it is obvious from TABLE5, the size of coefficients has changed in different classes of LC-MMNL and AGLC-MMNL model. Although such changes were observable in other studies such az the one done by Greene and Hensher (2013), this fact also should be investigated and explained in the future works.

Although, considering the agent effect in this study helps up to identify more classes across the population, it is not possible to exactly determine the contribution of the agent effect on the log likelihood of the LC-MMNL model, as the LC-MMNL models in this study have different numbers of classes, and as mentioned in the literature, the greater the number of classes, the better is models fit.

Future work

The agent effect latent class mixed multinomial logit model can be applied on different datasets to identify different classes of individuals across populations and to analyse their behaviour. However, the strength of this model highly depends on the quality and size of the datasets. Hence, for future work, before any application of this model, one should be aware about the experimental design procedure for gathering data from a population. In other words, while aiming to use latent class, random parameter and agent effect approaches all in the same structure, there should be enough variation in the dataset. Identifying all segments of the population and designing an experiment to cover all the segments and include detailed attributes could help the model to efficiently predict and estimate the individual behaviour.

Furthermore, as we guess that the results of AGLC-MMNL model in this study are dataset specific, there is a need to implement the AGLC-MMNL model on other datasets from different fields of study, like route choice or destination choice, to find out whether the model results are dataset specific or not.

Another issues, mentioned in the previous section, such as the endogeneity and the large changes in the scale of estimated parameters should be investigated in the future and find answers for them.

Finally, as mentioned above, due to the different number of classes, it is hard to specify the contribution of the agent effect to the log-likelihood of the AGLC-MMNL model. Hence, further research is needed to develop statistical tools that can determine the true contribution of agent effect in AGLC-MMNL model.

References

Abley, J. (2000). *Stated preference techniques and consumer decision making: new challenges to old assumptions*: Cranfield School of Management.

Akay, A. (2012). Finite-sample comparison of alternative methods for estimating dynamic panel data models. *Journal of applied Econometrics, 27*(7), 1189-1204.

Appleton, K., & Lovett, A. (2005). GIS-based visualisation of development proposals: reactions from planning and related professionals. *Computers, Environment and Urban Systems, 29*(3), 321-339.

Arnberger, A., & Eder, R. (2011). Exploring the heterogeneity of cultural landscape preferences: A visual-based latent class approach. *Landscape Research, 36*(1), 19-40.

Arnold, M. J., & Reynolds, K. E. (2003). Hedonic shopping motivations. *Journal of retailing, 79*(2), 77-95.

Atasoy, B., Glerum, A., & Bierlaire, M. (2011). *Mode choice with attitudinal latent class: A Swiss case-study.* Paper presented at the 2nd International Choice Modelling Conference, Leeds.

Bateman, I. J., Day, B. H., Jones, A. P., & Jude, S. (2009). Reducing gain–loss asymmetry: a virtual reality choice experiment valuing land use change. *Journal of environmental economics and management, 58*(1), 106-118.

Beharry-Borg, N., & Scarpa, R. (2010). Valuing quality changes in Caribbean coastal waters for heterogeneous beach visitors. *Ecological economics, 69*(5), 1124-1139.

Ben-Akiva, M., & Boccara, B. (1995). Discrete choice models with latent choice sets. *International journal of Research in Marketing, 12*(1), 9-24.

Ben-Akiva, M., Bolduc, D., & Walker, J. (2001). Specification, identification and estimation of the logit kernel (or continuous mixed logit) model. *Department of Civil Engineering Manuscript, MIT*.

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand* (Vol. 9): MIT press.

Berninger, K., Adamowicz, W., Kneeshaw, D., & Messier, C. (2010). Sustainable forest management preferences of interest groups in three regions with different levels of industrial forestry: An exploratory attribute-based choice experiment. *Environmental management, 46*(1), 117-133.

Bhat, C. R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B: Methodological, 35*(7), 677-693.

Bhat, C. R., Astroza, S., & Bhat, A. C. (2016). On allowing a general form for unobserved heterogeneity in the multiple discrete–continuous probit model: Formulation and application to tourism travel. *Transportation Research Part B: Methodological, 86*, 223-249.

Bienabe, E., & Hearne, R. R. (2006). Public preferences for biodiversity conservation and scenic beauty within a framework of environmental services payments. *Forest Policy and Economics, 9*(4), 335-348.

Bierlaire, M. (2003). *BIOGEME: A free package for the estimation of discrete choice models.* Paper presented at the Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.

Bierlaire, M. (2013). *Mixtures and latent variables in discrete choice models: an introduction.* Paper presented at the International Symposium on Transportation and Traffic Theory.

Bierlaire, M. (2014). Discrete panel data. *Mathematical modeling of behaviour-EPFL course*.

Bierlaire, M. (2015). BisonBiogeme 2.4: syntax of the modeling.

Boxall, P. C., & Adamowicz, W. L. (2002). Understanding heterogeneous preferences in random utility models: a latent class approach. *Environmental and resource economics, 23*(4), 421-446.

Boxall, P. C., Adamowicz, W. L., Swait, J., Williams, M., & Louviere, J. (1996). A comparison of stated preference methods for environmental valuation. *Ecological economics, 18*(3), 243-253.

Brown, G., Weber, D., & de Bie, K. (2015). Is PPGIS good enough? An empirical evaluation of the quality of PPGIS crowd-sourced spatial data for conservation planning. *Land Use Policy, 43*, 228-238.

Bujosa, A., Riera, A., & Hicks, R. L. (2010). Combining discrete and continuous representations of preference heterogeneity: a latent class approach. *Environmental and Resource Economics, 47*(4), 477-493.

Buttle, F., & Coates, M. (1984). Shopping motives. *Service Industries Journal, 4*(1), 71-81.

Davis, L., & Hodges, N. (2012). Consumer shopping value: An investigation of shopping trip value, in-store shopping value and retail format. *Journal of retailing and consumer services, 19*(2), 229-239.

de Bekker-Grob, E. W., Ryan, M., & Gerard, K. (2012). Discrete choice experiments in health economics: a review of the literature. *Health economics, 21*(2), 145-172.

Dugundji, E., & Walker, J. (2005). Discrete choice with social and spatial network interdependencies: an empirical example using mixed generalized extreme value models with field and panel effects. *Transportation Research Record: Journal of the Transportation Research Board*(1921), 70-78.

Gopimatj, D. A. (1997a). Modeling heterogeneity in discrete choice processes: Application to travel demand. *Transportation Research Part A, 1*(31), 86.

Gopimatj, D. A. (1997b). Modeling heterogeneity in discrete choice processes: Application to travel demand. *Transportation Research Part A, 31*(1).

Greene, W. H., & Hensher, D. A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological, 37*(8), 681-698.

Greene, W. H., & Hensher, D. A. (2010). *Modeling ordered choices: A primer*: Cambridge University Press.

Greene, W. H., & Hensher, D. A. (2013). Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model. *Applied Economics, 45*(14), 1897-1902.

Guevara, Cristian Angelo, and Moshe E. Ben-Akiva. "Change of scale and forecasting with the control-function method in logit models." Transportation Science 46, no. 3 (2012): 425-437.

Heckman, J., & Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica: Journal of the Econometric Society*, 271-320.

Hensher, D. A. (1994). Stated preference analysis of travel choices: the state of practice. *Transportation, 21*(2), 107-133.

Hensher, D. A., & Greene, W. H. (2003). The mixed logit model: the state of practice. *Transportation, 30*(2), 133-176.

Hess, S., Ben-Akiva, M., Gopinath, D., & Walker, J. (2011). Advantages of latent class over continuous mixture of logit models. *Institute for Transport Studies, University of Leeds. Working paper*.

Hess, S., Stathopoulos, A., & Daly, A. (2012). Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies. *Transportation, 39*(3), 565-591.

Hinterhuber, A. (2015). Violations of rational choice principles in pricing decisions. *Industrial Marketing Management, 47*, 65-74.

Horowitz, J. K., & McConnell, K. E. (2003). Willingness to accept, willingness to pay and the income effect. *Journal of Economic Behaviour & Organization, 51*(4), 537-545.

Hosoda, T. (1999). *Incorporating unobservable heterogeneity in discrete choice model: mode choice model for shopping trips.* Massachusetts Institute of Technology.

Houthakker, H. S. (1950). Revealed preference and the utility function. *Economica, 17*(66), 159-174.

Laing, R., Davies, A. M., Miller, D., Conniff, A., Scott, S., & Morrice, J. (2009). The application of visual environmental economics in the study of public preference and urban greenspace. *Environment and Planning B: Planning and Design, 36*(2), 355-375.

Laing, R. A., Davies, A.-M., Hargreaves, A., & Scott, S. (2004). The application of 3D modelling techniques in built environment evaluation. *Studies in Regional and Urban Planning, Issue 10*.

Lee, B. J., Fujiwara, A., Zhang, J., & Sugie, Y. (2003). *Analysis of mode choice behaviours based on latent class models.* Paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne.

Louviere, J. J., Flynn, T. N., & Carson, R. T. (2010). Discrete choice experiments are not conjoint analysis. Journal of Choice Modelling, 3(3), 57-72.

Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated choice methods: analysis and applications*: Cambridge University Press.

Louviere, J. J., Street, D., Burgess, L., Wasi, N., Islam, T., & Marley, A. A. (2008). Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information. *Journal of choice modelling, 1*(1), 128-164.

Mannering, F. L., Shankar, V., & Bhat, C. R. (2016). Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic methods in accident research, 11*, 1-16.

McFadden, D. (1977). QUANTITATIVE METHODS TOR ANALYZING TRAVEL BEHAVIOUR OF INDIVIDUALS: SOME RECENT DEVELOPMENTS.

McFadden, Daniel. "Modeling the choice of residential location." Transportation Research Record 673 (1978).

McFadden, D. (1980). Econometric models for probabilistic choice among products. *Journal of Business*, S13-S29.

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of applied Econometrics, 15*(5), 447-470.

Molnarova, K., Sklenicka, P., Stiborek, J., Svobodova, K., Salek, M., & Brabec, E. (2012). Visual preferences for wind turbines: Location, numbers and respondent characteristics. *Applied Energy, 92*(0), 269-278. doi: http://dx.doi.org/10.1016/j.apenergy.2011.11.001

Morey, E., Thacher, J., & Breffle, W. (2006). Using angler characteristics and attitudinal data to identify environmental preference classes: a latent-class model. *Environmental and Resource Economics, 34*(1), 91-115.

Morikawa, T., Ben-Akiva, M., & McFadden, D. (2002). Discrete choice models incorporating revealed preferences and psychometric data. *Advances in Econometrics, 16*, 29-56.

Mostofi-Darbani, J., Rezaei, A., Patterson, Z., & Zacharias, J. (2014). *Video Game vs. Traditional Text-Only SP Survey of Neighborhood Choice.* Paper presented at the Transportation Research Board 93rd Annual Meeting.

Orzechowski, M., Arentze, T., Borgers, A., & Timmermans, H. (2005). Alternate methods of conjoint analysis for estimating housing preference functions: Effects of presentation style. *Journal of housing and the built environment, 20*(4), 349-362.

Parzen, E. (1962). On estimation of a probability density function and mode. *The annals of mathematical statistics, 33*(3), 1065-1076.

Patterson, Z., Darbani, J. M., Rezaei, A., Zacharias, J., & Yazdizadeh, A. (2017). Comparing text-only and virtual reality discrete choice experiments of neighbourhood choice. *Landscape and Urban Planning, 157*, 63-74.

Perdomo, M., Rezaei, A., Patterson, Z., Saunier, N., & Miranda-Moreno, L. F. (2014). Pedestrian preferences with respect to roundabouts—A video-based stated preference survey. *Accident Analysis & Prevention, 70*, 84-91.

Reader, S. (1993). Unobserved heterogeneity in dynamic discrete choice models. *Environment and Planning A, 25*(4), 495-519.

Redmond, L. S., & Mokhtarian, P. L. (2001). The positive utility of the commute: modeling ideal commute time and relative desired commute amount. *Transportation, 28*(2), 179-205.

Revelt, D., & Train, K. (1998). Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of economics and statistics, 80*(4), 647-657.

Shen, J. (2009). Latent class model or mixed logit model? A comparison by transport mode choice data. *Applied Economics, 41*(22), 2915-2924.

Sieber, R. (2006). Public participation geographic information systems: A literature review and framework. *Annals of the Association of American Geographers, 96*(3), 491-507.

Simon, H. A. (1956). Rational choice and the structure of the environment. Psychological review, 63(2), 129.

Stephens, M. (2010). Review of Stated Preference and Willingness to Pay Methods. London: Accent, in association with RAND Europe,.

Stüttgen, P., Boatwright, P., & Monroe, R. T. (2012). A satisficing choice model. *Marketing Science, 31*(6), 878-899.

Tauber, E. M. (1972). Why do people shop? *The Journal of Marketing*, 46-49.

Teichert, T., Shehu, E., & von Wartburg, I. (2008). Customer segmentation revisited: The case of the airline industry. *Transportation Research Part A: Policy and Practice, 42*(1), 227-242.

Thiene, M., Boeri, M., & Chorus, C. G. (2012). Random regret minimization: exploration of a new choice model for environmental and resource economics. *Environmental and Resource Economics, 51*(3), 413-429.

Tilahun, N. Y., Levinson, D. M., & Krizek, K. J. (2007). Trails, lanes, or traffic: Valuing bicycle facilities with an adaptive stated preference survey. *Transportation Research Part A: Policy and Practice, 41*(4), 287-301.

Tinguely, L., Danalet, A., de Lapparent, M., & Bierlaire, M. (2015). *Destination Choice Model including a panel effect using WiFi localization in a pedestrian facility.* Paper presented at the Proceedings of the 15th Swiss Transport Research Conference.

Tovares, N., Boatwright, P., & Cagan, J. (2014). Experiential conjoint analysis: An experience-based method for eliciting, capturing, and modeling consumer preference. *Journal of Mechanical Design, Transactions of the ASME, 136*(10).

Train, K. E. (2009). *Discrete choice methods with simulation*: Cambridge university press.

Völckner, F., & Hofmann, J. (2007). The price-perceived quality relationship: A meta-analytic review and assessment of its determinants. *Marketing Letters, 18*(3), 181-196.

Völckner, F., & Sattler, H. (2005). Separating negative and positive effects of price with choice-based conjoint analyses. *Marketing ZFP, 27*(JRM 1), 5-13.

Vriens, M., Loosschilder, G. H., Rosbergen, E., & Wittink, D. R. (1998). Verbal versus realistic pictorial representations in conjoint analysis with design attributes. *Journal of Product Innovation Management, 15*(5), 455-467.

Walker, J. L. (2001). *Extended discrete choice models: integrated framework, flexible error structures, and latent variables.* Massachusetts Institute of Technology.

Walker, J. L., & Li, J. (2007). Latent lifestyle preferences and household location decisions. *Journal of Geographical Systems, 9*(1), 77-101.

Wen, C.-H., Huang, W.-W., Fu, C., & Chou, P.-Y. (2013). A latent class generalised nested logit model and its application to modelling carrier choice with market segmentation. *Transportmetrica A: Transport Science, 9*(8), 675-694.

Wen, C.-H., & Lai, S.-C. (2010). Latent class models of international air carrier choice. *Transportation Research Part E: Logistics and Transportation Review, 46*(2), 211-221.

Yamada, I., & Thill, J.-C. (2003). Enhancing stated preference surveys in transportation research: the contribution of geographic information systems. *Transportation planning and technology, 26*(5), 377-396.

Yannes, C., Lownes, N., Garrick, N., & Johnston, R. (2010). Operationalizing Place Making in a Choice Experiment Context. *Transportation Research Record: Journal of the Transportation Research Board, 2144*(-1), 121-129. doi: 10.3141/2144-14

Appendix A

The results of LC-MMNL and AGLC-MMNL model with different number of classes.

TABLE A1 Estimated LC-MMNL Model with Three Classes

Attribute		LC-MMNL		
		Class 1	Class 2	Class 3
ASC		-100.0 (0.00)	7.10 (12.48)	-.0719 (-0.39)
Dwelling type	Detached houses	Fixed	Fixed	Fixed
	Townhouse	144.00 (134.27)	2.24 (2.36)	-0.99 (-2.55)
	Triplex	-60.70 (-53.34)	-117.00 (-107.47)	-1.80 (-2.57)
	Apartment	-74.8 (-64.61)	-100 (-39582266.95)	-1.15 (-2.8)
Space between buildings (in feet)		-	6.41 (81.52)	-
Front yard depth (in feet)		-1.72 (-41.15)	0.35 (11.16)	-0.01 (-.70)
Travel time to nearby shops on foot (in minutes)		-7.50 (-109.45)	-3.48 (-76.37)	-0.07 (-4.11)
Average home value (in thousands CDN)		-14.90 (-123.79)	-0.64 (-112.24)	0.01 (2.00)
Travel time to work by car (in minutes)		-10.20 (-99.72)	-4.57 (-45.71)	-
Travel time to work by transit (in minutes)		0.80 (16.32)	0.13 (3.04)	-0.07 (-5.32)
Variance of random parameters				
Space		-	0.00 (-2.12)	-
Apartment		0.17 (39.32)	-0.10 (-2.65)	0.28 (0.20)
Triplex		-0.12 (-69.99)	0.11 (1.47)	1.10 (1.19)
Travel time to work by car (in minutes)		-0.01 (-14.13)	0.01 (0.45)	-
Class Membership Attributes				
High Income		-0.48 (-1.23)	-	-
Low Income		-	-0.14 (-0.38)	-
Zero or Non-represented Income		0.54 (1.21)	-	-
Employment Status (Full time =1, other = 0)		-0.43 (-1.530)	-	-
Language (French=1, English=0)		-0.19 (-1.13)	-	-
Driver (Having driver license=1, other =0)		-	0.04 (0.19)	0.04 (0.19)
Urban Core Residency(Reside in Urban Core =1, other = 0)		-	-12.90 (-8.12)	-
Current Residency (Apartment =1, other = 0)		-	-	-0.48 (-1.92)
Number of parameters		44		
Final Log-likelihood		-543.449		
BIC		0.9901		

TABLE A2 Estimated LC-MMNL Model with Four Classes

Attribute		LC-MMNL			
		Class 1	Class 2	Class 3	Class 4
ASC		17.8 (5.58)	7.94 (5.45)	5.45 (-0.18)	-46.9 (-2.94)
Dwelling type	Detached houses	Fixed	Fixed	Fixed	Fixed
	Townhouse	-71.5 (-5.43)	-2.51 (-1.45)	54.80 (0.12)	-4.00 (-2.35)
	Triplex	-55.10 (-5.32)	-36.7 (-4.95)	2.59 (0.04)	-67.5 (-3.01)
	Apartment	20.6 (5.43)	-85.5 (-6.45)	-29.1 (-0.26)	-
Space between buildings (in feet)		-	4.49 (6.78)	-	-
Front yard depth (in feet)		-1.7 (-5.46)	0.4 (5.14)	2.12 (0.15)	-0.08 (-1.55)
Travel time to nearby shops on foot (in minutes)		4.47 (5.56)	-4.47 (-6.96)	-2.18 (-2.60)	-4.77 (-3.10)
Average home value (in thousands CDN)		0.262 (5.45)	-0.17 (-6.85)	-12.4 (-0.16)	-0.126 (-3.09)
Travel time to work by car (in minutes)		22.00 (5.44)	1.19 (5.25)	-	-
Travel time to work by transit (in minutes)		-2.60 (-5.34)	-0.987 (-6.120)	-4.81 (-0.16)	-1.350 (-2.95)
Variance of random parameters					
Space		-	4.49 (6.78)	-	-
Apartment		23.0 (5.33)	-2.41 (-1.16)	0.218 (0.11)	-
Triplex		0.028 (0.41)	35.5 (6.98)	0.264 (0.05)	43.3 (3.07)
Travel time to work by car (in minutes)		-6.88 (-5.40)	4.43 (6.79)	-	-
Class Membership Attributes					
High Income		0.143 (0.40)	-	-	-
Medium Income		-	-	-	-0.231 (-0.62)
Low Income		-	-0.317 (-0.78)	-0.317 (-0.78)	-
Zero or Non-represented Income		-1.10 (-1.29)	-	-	-
Employment Status (Full time =1, other = 0)		-0.319 (-1.12)	-	-	-
Language (French=1, English=0)		-2.82 (-1.650)	-	-	-
Driver (Having driver license=1, other =0)		-	-	-1.290 (-4.54)	-1.290 (-4.54)
Urban Core Residency(Reside in Urban Core =1, other = 0)		-1.030 (-3.03)	-	-	-1.030 (-3.03)
Current Residency (Apartment =1, other = 0)		-	-	0.138 (0.44)	-
Number of parameters		53			
Final Log-likelihood		-501.218			
BIC		0.9148			

TABLE A3 Estimated LC-MMNL Model with Five Classes

Attribute		**LC-MMNL**				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
ASC		62.2 (3.19)	3.78 (6.16)	-20.3 (-7.59)	-25.9 (-3.36)	-70.1 (-4.79)
Dwelling type	Detached houses	Fixed	Fixed	Fixed	Fixed	Fixed
	Townhouse	-46.4 (-3.56)	12.8 (3.53)	-17.2 (-7.2)	19.5 (4.38)	69.4 (3.89)
	Triplex	-121.0 (-3.18)	-159.0 (-3.42)	-81.4 (-7.17)	-48.5 (-3.94)	-60.3 (-4.75)
	Apartment	-78.9 (-3.12)	-137.0 (-3.21)	-113.0 (-7.25)	-	-
Space between buildings (in feet)		-	8.53 (3.07)	-	-	-
Front yard depth (in feet)		0.364 (2.92)	0.577 (4.32)	-1.77 (-7.21)	26.4 (3.82)	-9.16 (-4.55)
Travel time to nearby shops on foot (in minutes)		7.96 (3.27)	-7.23 (-3.18)	0.70 (7.37)	-39.10 (-3.87)	-4.72 (-4.66)
Average home value (in thousands CDN)		-1.02 (-3.2)	-0.495 (-3.26)	-12.4 (-7.15)	-1.14 (-4.28)	0.179 (-4.28)
Travel time to work by car (in minutes)		29.8 (3.1)	5.68 (3.32)	-	-	-
Travel time to work by transit (in minutes)		1.13 (2.9)	-2.23 (-2.97)	-6.53 (-6.97)	8.97 (3.88)	-18.7 (-4.63)
Variance of random parameters						
Space		-	0.0147 (0.14)	-	-	-
Apartment		0.191 (2.53)	-1.96 (-2.89)	0.028 (43.9)	-	-
Triplex		0.0445 (0.77)	-63.0 (-3.46)	0.0289 (20.47)	0.0325 (9.21)	-98.0 (-4.28)
Travel time to work by car (in minutes)		-0.00801 (-0.99)	0.4 (2.87)	-	-	-
Class Membership Attributes						
High Income		-0.413 (-0.63)	-	-	-	-
Medium Income		-	-	-	0.478 (1.91)	0.478 (1.91)
Low Income		-	0.0571 (0.13)	0.0571 (0.13)	-	-
Zero or Non-represented Income		-0.673 (-0.68)	-	-	-	-
Employment Status (Full time =1, other = 0)		-0.803 (-1.690)	-	-	-	-
Language (French=1, English=0)		-0.292 (-1.05)	-	-	-	-
Driver (Having driver license=1, other =0)		-	-	-1.67 (-7.9)	-1.67 (-7.9)	-1.67 (-7.9)
Urban Core Residency(Reside in Urban Core =1, other = 0)		-	-1.65 (-4.11)	-	-1.65 (-4.11)	-
Current Residency (Apartment =1, other = 0)		-	-	0.616 (1.93)	-	-
Number of parameters		61				
Final Log-likelihood		-492.128				
BIC		0.899				

TABLE A4 Estimated AGLC-MMNL Model with Two Classes

Attribute		LC-MMNL	
		Class 1	Class 2
ASC		-0.123 (-0.72)	-
Dwelling type	Detached houses	Fixed	Fixed
	Townhouse	-	-0.82 (-1.85)
	Triplex	8.59 (4.06)	-2.08 (-2.07)
	Apartment	-	-1.73 (-2.49)
Space between buildings (in feet)		-	0.05 (1.55)
Front yard depth (in feet)		-29.6 (-4.06)	0.0242 (1.14)
Travel time to nearby shops on foot (in minutes)		16.60 (3.96)	-0.09 (-2.30)
Average home value (in thousands CDN)		-8.9 (-4.20)	-0.00384 (-2.21)
Travel time to work by car (in minutes)		71.80 (4.23)	0.02 (1.16)
Travel time to work by transit (in minutes)		-17.20 (-4.01)	-0.04 (-2.03)
Variance of random parameters			
Space		-	-0.0649 (-1.26)
Apartment		-	-0.778 (-0.50)
Triplex		-	-0.553 (-1.30)
Travel time to work by car (in minutes)		0.00047 (0.05)	-0.13 (-1.70)
SIGMA		0.249 (0.20)	
Class Membership Attributes			
High Income		-0.32 (-0.66)	-
Language (French=1, English=0)		-0.89 (-3.48)	-
Current Residency (Apartment =1, other = 0)		0.57 (1.85)	-
Number of parameters		25	
Final Log-likelihood		-548.69	
BIC		0.997	

TABLE A5 Estimated AGLC-MMNL Model with Three Classes

Attribute		LC-MMNL		
		Class 1	Class 2	Class 3
ASC		44.3 (1.36)	2.99 (0.29)	-52.6 (-3.07)
Dwelling type	Detached houses	Fixed	Fixed	Fixed
	Townhouse	-	33.3 (0.41)	145 (3.02)
	Triplex	-42.8 (-1.27)	-	63.8 (3.1)
	Apartment	121 (1.3)	-220 (-0.35)	65.8 (2.71)
Space between buildings (in feet)		-	14.4 (0.36)	-
Front yard depth (in feet)		-2.05 (-1.24)	1.32 (0.35)	-1.8 (-3.62)
Travel time to nearby shops on foot (in minutes)		6.58 (1.26)	-13 (-0.35)	-5.76 (-3.24)
Average home value (in thousands CDN)		0.678 (1.28)	-0.503 (-0.34)	-4.54 (-3.13)
Travel time to work by car (in minutes)		48.3 (1.25)	2.93 (0.36)	
Travel time to work by transit (in minutes)		-7.88 (-1.31)	-1.76 (-0.35)	-5.34 (-3.17)
Variance of random parameters				
Space		-	0.296 (0.29)	-
Apartment		-1.72 (-1.2)	-2.05 (-0.23)	152 (3.26)
Triplex		83.2 (1.33)	-	-0.506 (-1.23)
Travel time to work by car (in minutes)		-19.9 (-1.25)	-	-13.2 (-0.35)
SIGMA		-0.536 (-0.12)		
Class Membership Attributes				
High Income		0.198 (0.58)	-	-
Low Income		-	-0.197 (-0.5)	-0.197 (-0.5)
Zero or Non-represented Income		-1.78 (-2.38)	-	-
Employment Status (Full time =1, other = 0)		-0.207 (-0.76)	-	-
Language (French=1, English=0)		-0.358 (-2.01)	-	-
Driver (Having driver license=1, other =0)		-	-	-1.16 (-4.48)
Urban Core Residency(Reside in Urban Core =1, other = 0)		-	-0.643 (-2.25)	-
Current Residency (Apartment =1, other = 0)		-	-	0.168 (0.53)
Number of parameters		42		
Final Log-likelihood		-497.582		
BIC		0.907		

TABLE A6 Estimated AGLC-MMNL Model with Five Classes

Attribute		**LC-MMNL**				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
ASC		8.28 (1.42)	0.43 (0.38)	-8.78 (-1.64)	-19.02 (-3.47)	-4.03 (-0.11)
Dwelling type	Detached houses	Fixed	Fixed	Fixed	Fixed	Fixed
	Townhouse	-5.88 (-1.24)	-3.93 (-0.73)	11.2 (1.62)	3.69 (2.64)	1.25 (0.03)
	Triplex	-14.4 (-1.62)	-9.52 (-1.00)	1.45 (1.24)	5.82 (2.8)	-5.51 (-0.05)
	Apartment	-10.1 (-1.5)	-7.15 (-0.92)	-5.92 (-1.7)	-	-
Space between buildings (in feet)			0.138 (0.73)			
Front yard depth (in feet)		0.021 (0.33)	0.0327 (0.74)	0.467 (1.53)	0.431 (3.85)	-0.763 (-0.06)
Travel time to nearby shops on foot (in minutes)		1.19 (1.45)	-0.258 (-0.77)	-0.532 (-1.42)	-1.91 (-3.44)	-0.421 (-0.06)
Average home value (in thousands CDN)		-0.15 (-1.44)	-0.00532 (-0.73)	-2.75 (-1.67)	-0.0939 (-3.88)	0.0132 (0.22)
Travel time to work by car (in minutes)		3.92 (1.45)	0.145 (1.03)			
Travel time to work by transit (in minutes)		0.213 (1.2)	0.00108 (0.02)	-0.936 (-1.62)	-0.324 (-3.14)	-1.52 (-0.07)
Variance of random parameters						
Space			-0.0294 (-0.4			
Apartment		0.213 (0.2)	-3.85 (-1.07)	0.482 (0.25)		
Triplex		0.448 (0.32)	2.3 (1.86)	0.103 (0.08)	0.519 (0.45)	-9.34 (-0.06)
Travel time to work by car (in minutes)		0.107 (1.05)	0.272 (0.87)			
SIGMA		-1.49 (-0.42)				
Class Membership Attributes						
High Income		-0.302 (-0.32)	-	-	-	-
Medium Income					0.382 (0.37)	0.382 (0.37)
Low Income			0.288 (0.7)	0.288 (0.7)		
Zero or Non-represented Income		-1.04 (-1.14)				
Employment Status (Full time =1, other = 0)		-1.04 (-1.18)				
Language (French=1, English=0)		-0.276 (-1.12)				
Driver (Having driver license=1, other =0)				-2.33 (-2.95)	-2.33 (-2.95)	-2.33 (-2.95)
Urban Core Residency(Reside in Urban Core =1, other = 0)			-1.6 (-2.15)		-1.6 (-2.15)	
Current Residency (Apartment =1, other = 0)				0.426 (0.96)		
Number of parameters		62				
Final Log-likelihood		-502.495				
BIC		0.918				