BAYESIAN APPROACHES FOR CALIBRATING THE MIXED C-LOGIT MODEL OF TRUCK ROUTE CHOICE

A Dissertation

by

YI MENG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Xiubin Wang
Committee Members,	Yunlong Zhang
	Tim Lomax
	Luca Quadrifoglio
Head of Department,	Robin Autenrieth

May 2021

Major Subject: Civil Engineering

ABSTRACT

This study examines logit models applied to the truck route choice problem with data from the Dallas metropolitan area. Instead of assuming a fixed coefficient of a variable in the conventional multinomial logit model, the proposed model assumes a certain probability distribution for each coefficient, typically called the mixed C-logit, in an attempt to better reflect the preference heterogeneity. Three Bayesian approaches with different hierarchy levels are introduced and are solved by the mean-field variational inference with the implementation of the block coordinate algorithm. The associated models are tested on two subnetworks in two scenarios, the first of which has toll alternatives while the other does not. It is found that all the three proposed models notably outperform the conventional multinomial logit model, which conforms to the behavior indicated in the simulation test.

Generally, our study finds that travel time is the most significant factor considered in truckers' route choices in both scenarios. The relative importance of attributes in affecting truckers' route choices differs between scenarios. In Scenario 1, travel time dominates other attributes. However, in Scenario 2, with a less dense network than in Scenario 1, it is found that using a route that entirely consists of state or interstate highway segments is as essential as using a route with a short travel time for most drivers. Additionally, the truck drivers' preference for roadway delay and network density is found to vary widely in the numerical test. In contrast, their preference for travel time and roadway designation is relatively consistent.

ACKNOWLEDGEMENTS

Dr. Wang advised me during this study. Thanks to my committee member Dr. Tim Lomax from Texas A&M Transportation Institute, Dr. Yunlong Zhang, and Dr. Luca Quadrifoglio from the Zachry Department of Civil and Environmental Engineering, for their comments, advice, and time.

Special appreciation goes to Dr. David Schrank, TTI senior research scientist, and program manager of the Mobility Division, for his guidance and suggestions. I extend my gratitude to the Mobility Division of Texas A&M Transportation Institute for financial support through my Ph.D. study.

My gratitude goes to Dr. Bill Eisele from the Mobility Division of Texas A&M Transportation Institute for kindly providing the data for this study.

Worthy of special mention is Dr. Bosheng Zhao, for accompanying me during my study, and my colleagues and my friends in the Department of Civil and Environmental Engineering for their help.

Last but not least, my parents always supported me with their unfaltering love over the years.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a graduate research assistantship from Texas A&M Transportation Institute.

This work's contents are solely the responsibility of the authors and do not necessarily represent the official views of the Texas A&M Transportation Institute.

NOMENCLATURE

O-D	Origin-Destination
RRM	Random Regret Minimization
RUM	Random Utility Maximization
IIA	Independence of Irrelevant Alternatives
GEV	Generalized Extreme Value
PSL	Path Size Logit
MXL	Mixed Logit
MCL	Mixed C-logit
MPSL	Mixed Path Size Logit
MNL	Multinomial Logit
MCMC	Markov Chain Monte Carlo
KL	Kullback-Leibler
MFVI	Mean Field Variational Inference
AMXL	Adjusted Mixed Multinomial Logit Model
CF	Commonality Factor
ELBO	Evidence Lower Bound
BMCL	Bayesian Adjusted Mixed Multinomial Logit
HBMCL1	Hierarchical Bayesian Adapted Mixed Multinomial Logit Model with Inverse Wishart Prior

HBMCL2	Hierarchical Bayesian Adapted Mixed Multinomial Logit Model with Huang's Half-t Prior
BFGS	Broyden-Fletcher-Goldfarb-Shanno
RMSE	Root Mean Squared Error
DBSCAN	Density-Based Spatial Clustering of Application with Noise
AADT	Annual Average Daily Traffic
SSD	Sum of Squared Distances
NTTA	North Texas Tollway Authority
FHWA	Federal Highway Administration
LDA	Linear discriminant analysis
MLP	Multi-Layer Perceptron
RF	Random Forest
CART	Classification and Regression Tree
KNN	K-Nearest Neighbors
SVM	Support Vector Machine
NB	Naïve Bayes
VTTS	Value of Travel Time Savings
S/IS	State/Interstate highway

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

1.1. Background

Trucking operations are an essential component of freight transportation. In 2018, trucks carried the largest proportion of shipments that covered less than 1,000 miles. From a supply chain perspective, trucks transport the majority of tonnage and value of the top 10 commodities, ranging from high-valued electronics to low-valued miscellaneous manufactured products (USDOT FAF4, 2019). Additionally, 59.4% of the tonnage and 62.8% of the value of hazardous materials are shipped by truck. Being a dominant mode in multimodal freight transportation that is integrated with rail, water, and airways, truck transportation accounts for more than half of the total intrastate shipments for most states and 79.7% of intrastate shipments (USDOT BTS, 2019a).

Economic development and population increases have boosted business practices and employment, and freight has shifted from other modes to trucks due to their flexibility. According to the freight analysis framework, it is predicted that in 2035 the truck shipment share will be 66.9%, whereas in 2007, this number was 59.7% (USDOT FHWA, 2007a). In this context, the past decade has seen rapid growth in the truck industry. From 2008 to 2018, the total number of registered trucks has increased by 21.7%, and the vehicle miles traveled have increased by 13.2% (FHWA, 2020). On the one hand, the growth of the truck industry brings remarkable economic benefits, as can be seen from its direct contribution to the gross domestic product of the United States. In 2014, private trucks contributed about $135.7 billion, and for-hire trucks contributed

equally, with $135.4 billion. In 2018, this number increased to $190.9 billion for private trucks and $163.7 billion for for-hire trucks (USDOT BTS, 2019b).

On the other hand, the increased demand for trucks, especially long-haul trucks, exacerbates the congestion of inner-city networks. Since commercial trucks are business oriented, they tend to avoid heavily congested local sites for operational efficiency and productivity, impeding regional economic development (USDOT FHWA, 2008). Nationwide, even though trucks make up only 7% of traffic, they account for 11% of the total annual congestion costs, with a total of $20 million in wasted time and fuel (Schrank et al., 2019). Highway traffic is heavily affected by congestion. Recurring peak-period congestion involved over 10,600 miles of the national highway system (NIS), with an additional 6,700 miles of stop-and-go traffic (USDOT FHWA, 2007b). However, such a problem prohibits the establishment of an efficient and smooth flow of trucks, for which on-time deliveries are of particular importance.

In addition to traffic congestion, the impact of medium- and heavy-duty trucks on the environment and energy is notable. Specifically, medium- and heavy-duty trucks make up 5% of the traffic, whereas they account for over 20% of transportation emissions (EESI, 2015). Moreover, trucks are of vital consideration for highway safety and infrastructure. For instance, trucks that exceed the legal mass limit may potentially lead to an increase in traffic accidents and pavement deterioration (Jacob and Feypell-de La Beaumelle 2010). As operational characteristics distinguish trucks from passenger cars, different regulatory procedures are necessary to plan truck traffic. To facilitate urban transportation planning, numerous studies have concerned various aspects of truck

operations, including mobility, infrastructure, safety, and travel demand (Apronti et al. 2019; Sathaye, Horvath, and Madanat, 2010; X. [Cara] Wang and Zhang, 2017; Z. Wang, Goodchild, and McCormack, 2016). Truck routing behavior, however, has not been fully studied due to modeling difficulty and insufficient data.

1.2. Motivation

Trucking operations exist primarily to serve the business needs of carriers and shippers. These needs have many attributes such as shipping size, pickup and delivery time windows, and others, which have much to do with truck route choice behavior. A truck operation's complexity may also explain the heterogeneity of route choice behaviors of the same individuals under different circumstances. There are two types of shipping options for urban truck freight transportation: truckload shipping and less-than-truckload shipping. In truckload shipping, shipments take the full truckload and are directly transported to the destination without a distribution process. When a shipment takes up space that is less than a full 48-foot or 52-foot standard trailer, it is called a less-than-truckload shipment. Less-than-truckload shipping allows carriers to operate at their maximum capacity by combining shipments and is thus cost efficient and environmentally friendly. Loads are first collected at a terminal and then are distributed by location and transported to regional terminals for delivery. Both types of shipping include short-haul delivery such as within a city and long-haul delivery such as interstate transports.

Trucking route choice is a complicated process for several reasons. First, there is a combination of stakeholders in the decision-making process. The route choice decision

for a truck may be jointly made by the driver, the company manager, and the receiver, compared to the route decision of a passenger car, which is straightforward and purely made by the driver (Jiang and Zhang 2019). Second, many short-haul fleets serve an urban area. The operation of such delivery trucks may need to consider local regulations about road use, delivery time, or parking site restrictions. For example, in New York, trucks and trailers are required to park at specific sites. Trucks are limited to use only certain routes in the city, and some of their deliveries are limited to off-peak hours (NYC DOT - Trucks and Commercial Vehicles). In San Francisco, truck routes are also restricted, and their inner-city trips need to be advised (San Francisco MTA). Additionally, there may be specific time windows for trucks required by the customers.

For long-haul delivery, trucks are required to follow the hours-of-service limit regulations from the Federal Motor Carrier Safety Administration (FMCSA, USDOT). Among the regulations, drivers have to take a mandatory 30-minute break by the eighth hour of driving, and the total driving duration needs to be less than 11 hours, including stops for fuel and breaks. After a 14-hour maximum service time, a mandatory 10 hours off duty is required. The regulation of service time motivates carriers to pay for saving driving time and other operating costs such as fuel and labor costs. This regulation also affects the truckers' planning of trips since rests or pick-up/delivery sites need to be included in their service time. As mentioned earlier, many cities have local regulations on the access of inner-city parking spots for trucks; limitations on parking sites and service time regulations have increased the complexity of planning and operating urban truck fleets.

Moreover, the route choice of drivers may also be affected by cooperation among industries. For example, drivers may use a gas station that provides membership discounts. Consequently, the modeling difficulty of truck route choice attributes all of the aforementioned factors that affect the route choice of commercial carriers and shippers.

A thorough understanding of the truck routing behavior can soundly guide policymakers on budgeting transportation infrastructure and developing strategic regulations and procedures to alleviate traffic congestion and maintain highway system performance. Specifically, with strategic planning, delivery travel time can be reduced and so can the corresponding emissions, potential crashes, and noise. In addition, a proper logistic operation will reduce local congestion and thus encourage economic activities and save energy. Many private enterprises think that participating in planning studies as well as market assessment studies results in sharing proprietary information (Dowling et al., 2014). This misconception limits the amount and quality of data, especially the stated preference survey data available for truck traffic research. It is pointed out that for carriers, understanding truck routing helps them improving operating efficiency. Cooperating with the public sector in the study of route choice behavior benefits an enterprise with saved vehicle operating and human resource costs and increased fleet productivity and on-time deliveries. Moreover, a good knowledge of route choice helps carriers make quick re-routing decisions in emergencies (Luong, Tahlyan, and Pinjari 2018).

However, the lack of understanding of the preference heterogeneity of truckers' route choice has prevented practitioners from understanding truck traffic well. The term "preference heterogeneity" refers to the variation in evaluating observable factors across a group of individuals in different situations. Models such as latent class models and mixed logit models that explicitly consider individual variation of preferences appear to have gained popularity in recent years due to their superior performance in capturing choice makers' decisions compared to traditional models such as multinomial logit models.

The literature that explicitly considers preference heterogeneity has been mainly in the area of studying passenger travel. In theory, the frameworks therein shed light directly on the study of truck travel behavior. Thus far, the literature on truck route choice has mainly resorted to multinomial or binary logit models with various extensions. The data used is typically from two categories: stated preferences or revealed preference. The stated preferences are obtained typically through surveys using hypothetical situations. In contrast, the revealed preferences are obtained today mainly through GPS tracking records, such as the click data recorded from GPS devices with minimal time intervals. In an effort to reveal preference heterogeneity with survey data, two types of models are applied in existing studies, namely, the latent class model and the mixed logit model, to the stated preference surveys. Feng, Arentze, and Timmermans (2013) address heterogeneity with the latent class model, assuming that truck route choice priorities are distinct between subgroups. Rowell, Gagliano, and Goodchild (2014) apply item response theory and latent class analysis, allowing the latent variables

that affect the sensitivity of truck route choice to external factors to be continuous and discrete, respectively. The applications of the mixed logit model are only found in Arentze et al. (2012) and Toledo et al. (2013). Both only consider random error terms associated with certain variables, with road pricing and congestion in the former and tollway usage indicators and toll price in the latter. An application of the ordered probit model is given by Poulopoulou, Spyropoulou, and Antoniou (2015). The correlation of responses from the same individual was captured by random effects included in the model. However, to the best of the author's knowledge, empirical evidence about truck drivers' choice heterogeneity from a revealed preference source (GPS data) remains unveiled.

GPS data from truckers are used to construct route trajectories in this study. It has high precision in vehicle location, and therefore the vehicle trajectories indicated in it are reliable and accurate. However, the GPS data has noise because it records every location of the truck. Some locations recorded are irrelevant to truck driving such as rest and gas station visits. As a result, it requires a great effort in data cleaning and processing, including identification of origin-destination (O-D) pairs, the combination of sub-trips into one complete trip, removing sub-trips from existing trips, and extraction of actual travel time with break time cleared. The data used in this study is from a GPS dataset that contains more than 1.6 billion GPS data points and 29 million trips for the Dallas metropolitan area in Texas. It is observed from this dataset that many trips have intermediate destinations during their main trips, which is mentioned as a sub-trip issue in this study and needs to be addressed with the map-matching process. The map-

matching process refers to matching the cleaned trajectories to the network topology and identifying the driver's trajectories. A series of existing truck route choice studies have adopted a binary choice setting, which avoids the impact of network complexity on map-matching (Arentze et al., 2012; Knorring, He, and Kornhauser 1923; Sharma, Snelder, and van Lint 2019). From a network perspective of truck route choice, however, the map-matching procedure is usually oversimplified. The only study that has included a map-matching process is from Oka et al. (2019), which matched GPS points by calculating a cost that corresponds to distance and which identified the only actual route as the one with the minimal total cost for each pair of O-D. This study provides a map-matching process that addresses the data's sub-trip issue with a clustering method.

In this study, we construct choice sets of truck drivers' routes for their O-D travels by extracting the O-D trajectories from the processed data in the Dallas-Fort Worth area, with unused routes by the truckers excluded for the purpose of modeling. The models are tested on these extracted alternative sets, with individual attributes recorded in the GPS data. These attributes include the instantaneous travel time, the annual truck delay per lane, the roadway designation, the toll as well as the commonality factor (CF), which is described in detail in Section 3.2. A similar way of modeling route choice behavior for passenger cars is available in the studies by Frejinger and Bierlaire (2007); Haghani, Sarvi, and Shahhoseini (2015); Lue and Miller (2019); Tang et al. (2020); Yang, Zhang, and Grembek (2016); and Madadi et al. (2019).

1.3. Problem statement

The study problem is about truck route choice-making and can be summarized as follows. There is a group of truck drivers who each need to travel from an origin to a destination. There are a set of alternative routes for each driver, and each route is characterized by a set of known attributes, including travel time, which is comprised of the travel times on the roadway segments that the trip traverses, tolls,[1] truck delays due to congestion, roadway designations, and the CF. The truck drivers make their route choice based on these known attributes. Specifically, the roadway designation refers to roadway classifications such as arterial, collector or local. In this study, the roadway designation is a binary indicator variable, which takes zero when the alternative does not contain local road segments and one otherwise. The CF is an adjustment factor to the utility function that accounts for network overlapping, and in that, it represents the network density.

This study proposes the utility function and estimates the values (distribution) of the associated coefficient for each variable (i.e., attribute) included. The goal is that the calibrated utility function will be able to predict truck travel choices as reflected in reality. In addition, the calibrated utility model would allow practitioners to evaluate the relative importance of the attributes considered in the truckers' route choice decision-making. In particular, the study problem is modeled with a mixed C-logit model via a Bayesian estimation method, which allows random coefficients in the utility function that conforms to certain distributions. The model is evaluated by the maximum deviation

[1] If an alternative does not contain any toll segment, the toll price is set to zero.

of the predicted probability from the true probability of each alternative route which is used for a specific O-D pair under various choice situations.

1.4. Contributions

Two contributions arise in this study.

First, the study has applied the Bayesian estimation method to calibrate the utility function for the first time for the truck route choice problem. In particular, a mixed C-logit model is proposed, which includes a CF to account for overlapping routes in the network. As described earlier, the coefficients of the explanatory variables in the utility function are generally assumed to be random, whose parameters for the probabilistic distribution is estimated by the proposed Bayesian method. Each random coefficient can be partitioned into a fixed mean term plus a random term which follows the same distribution, but has a mean of zero and the same variance. Then the error component of the mixed logit model can be reformed by combining this random term and the error term in the utility function, which is assumed to be i.i.d. Gumbel distributed. With this new formulation, the mixed logit model can thus be called the error component model. The applied Bayesian approach is capable of addressing the correlation of the variance of the error component in the utility function and consequently accounts for the preference heterogeneity (Hess & Train, 2017). Specifically, three Bayesian approaches with different levels of hierarchy were applied and evaluated in both simulation tests and the numerical test with real-world data.

Second, as stated earlier, many trips have multiple en-route visits between their origin and destination pair. In theory, a route between two visits is considered a trip;

however, the sequential trips in this situation may be somewhat correlated (at least there is such a likelihood of suspicion). In an attempt to separate such trips from those that are defined by directly travel to the destination without en-route visits, we apply a clustering method to identify trips that have common intermediate destinations, which is a minor contribution of this study.

1.5. Overview

This dissertation consists of seven sections. In Section 1, an overview of the study background, motivations, problem statement, and the contributions of the work is presented. In Section 2, the decision rules of route choice, commonly used models for route choice and the variational inference that is used for solving the proposed methods are reviewed in detail. In Section 3, the proposed C-logit model and three Bayesian calibration methods are specified in detail. In Section 4, the proposed model is applied to a three-route network in two simulation tests. The first simulation test compares the proposed model to the existing mixed path size logit, the mixed logit, and the multinomial logit for a detailed evaluation of the model performance in probability prediction and parameter estimation. The second simulation test explores the model performance with different amounts of data input. In Section 5, a comprehensive data processing procedure, which includes the map-matching process, is introduced. In Section 6, the predicted probability and the parameter estimation results are documented and analyzed. This dissertation closes with conclusions and a discussion in Section 7.

2. LITERATURE REVIEW

2.1. Decision rules of route choice models

2.1.1. Random utility theory

The underlying assumptions of random utility theory are that the individual decision-making process is rational by maximizing the perceived utilities of alternatives, in which there is a random utility term that reflects the heterogeneity among travelers. The utility theory is a disaggregate theory that builds on individual travelers. Each individual is endowed with a finite choice set of available routes, denoted by J and indexed by j. The practical models in either the form of logit or probit all assume i.i.d. for the random error term. The property of i.i.d. implies that individual decisions are independent of each other. The decision-makers have a perceived utility for each alternative a_j, denoted by U_j, and this utility is assumed to be characterized by a finite number of measurable attributes that may contain alternative-specific and individual-specific factors.

With the definitions above, the fundamental hypothesis of random utility theory can be expressed as

$$P(j|A) = P\left(U_j \geq U_{j'}\right), \forall\, j' \neq j. \tag{2.1}$$

According to Manski (1977), there are four different sources of uncertainty for the decision-making process: unobserved alternative attributes, unobserved individual characteristics or unobserved taste variations, measurement errors, and proxy, or instrumental, variables. Considering those uncertainties, for individual n, the individual

perceived utility U_{nj} is modelled as a random variable, which contains a systematic part V_{nj} and a stochastic part ε_{nj}.

The systematic part, or deterministic part of the utility V_{nj}, is expressed as a function of attributes relative to each alternative. Although the functional form of attributes is flexible, the most commonly used one is linear. Let X_n denote the attributes matrix for each individual, where n index the number of total decision-makers of the study; X_{nj} is the j^{th} row of X_n and denotes the vector of attributes with respect to alternative j. Let β denote the vector taste parameters to be estimated. The linear form of systematic utility can therefore be written as

$$V_{nj} = \beta^T X_{nj}. \tag{2.2}$$

To model the stochastic part of the utility or the random error ε_{nj}, different statistical distributions are applied. Among the variety of potential distributions reported in the existing literature, Gumbel and Normal are the most popular choices. With the Gumbel distribution (generalized extreme value distribution type II) assumption, the probability density function and the cumulative density function of the random error are

$$f(\varepsilon_{nj}) = \frac{1}{\alpha} e^{-(\frac{\varepsilon_{nj}-\mu}{\alpha} + e^{-\frac{\varepsilon_{nj}-\mu}{\alpha}})} \tag{2.3}$$

and

$$F(\varepsilon_{nj}) = e^{-e^{-(\varepsilon_{nj}-\mu)/\alpha}} \tag{2.4}$$

where μ is the location parameter and α is the scale parameter. This assumption brings in the logit family of route choice models, which is further specified in Section 2.1.

2.1.2. Other decision rules

Besides the mainstream random utility decision rule, other decision rules have been proposed for discrete choice models. A famous alternative is random regret minimization (RRM). The convectional linear-additive utility in a random utility model shows full compensation among attributes, while the random regret in RRM is defined with a non- or semi-compensatory scheme. RRM assumes that individuals would choose an alternative with a minimum of anticipated regret. A line of research compares the parameters such as the value of time given by the two decision rules (C. Chorus 2012; Masiero, Yang, and Qiu 2019). Many studies have explored the performance of different route choice models and the corresponding network assignment problem under RRM (Li and Huang 2017). Some shed light on the generalization form, where both RUM and RRM properties are considered (Chorus, 2014).

Bounded rationality is another alternative for route choice. Both RUM and RRM assume that the traveler chooses the optimal option (i.e., perfect rationality). However, the lack of accurate information or the complexity of real-world situations makes people boundedly rational, and their decision may thus be a satisfactory choice. For a detailed review of concepts, formulations, and methods on boundedly rational route choice, see Di and Liu (2016). Moreover, it is shown that the route choice decision is affected by inertia, which is understood as the reluctance to adjust the current choice. In other words, individuals would change their route choice only if the perceived utility of other choices significantly outweighed that of the current one (J. Zhang and Yang 2015).

2.2. Logit family of models

The logit family model documented in this section is based on random utility theory. The stochastic part of utility is assumed to be Gumbel distributed with mean 0. Statistically, this distribution is characterized by a location parameter μ and a scale parameter α, that is, $\varepsilon_{nj} \sim Gumbel(\mu, \alpha)$. To reflect the zero-mean assumption, this distribution is rewritten as $\varepsilon_{nj} \sim Gumbel(-\alpha\gamma, \alpha)$, where γ is the Euler-Mascheroni constant. In this case, each random utility U_{nj} can be viewed as a Gumbel variable with mean V_{nj}.

2.2.1. Binary logit and multinomial logit

In binary logit, there are only two alternatives in the choice set, that is, $J = \{j_1, j_2\}$, and the outcome of discrete choice Y_n is a binary variable, which only takes a value of 0 or 1. The individual n chooses an alternative a_1 if and only if $U_{n1} \geq U_{n2}$. The Gumbel distribution is stable with respect to maximization, which is one of the critical reasons that logit family models are popular. The maximum of independent Gamble variables with the same scale parameter α is also a Gumbel variable with scale parameter α and with the mean being the expected maximum perceived utility. Therefore, the probability of choosing alternative a_1 is

$$P_n(a_1) = \frac{\exp(V_{n1}/\alpha)}{\exp(V_{n1}/\alpha) + \exp(V_{n2}/\alpha)}. \tag{2.5}$$

Using the linear functional form of the systematic utility and combining the scale parameter with the taste parameter, one gets

$$P_n(a_1) = \frac{1}{1 + \exp(\beta^T X_n)}. \tag{2.6}$$

Then, the probability of choosing a_2 is simply 1-P_n.

When the choice set contains multiple choices, the binary logit progresses to the multinomial logit model, which is the simplest and most popular discrete choice model. With the i.i.d. assumption where the stochastic error is independent and identically Gumbel distributed, the probability of individual n choosing alternative j is written as

$$P_{nj} = \frac{\exp\left(\beta^T X_{nj}\right)}{\sum_{j' \neq j} \exp\left(\beta^T X_{nj'}\right)}. \tag{2.7}$$

2.2.2. Path size logit

In real transportation network applications, the independence of irrelevant alternatives (IIA) property limits the performance of the multinomial logit model due to shared links in the roadway network. The stochastic part of the perceived utility is not independent, but is correlated when alternative routes contain shared links. Path size logit is a remedy to correlated error components by implementing an adjustment or correction factor that is proportional to the degree of roadway overlapping in the perceived utility. This adjustment factor is named the path size, with a generalized form proposed by Ramming (2001), specified as follows:

$$PS_j = \sum_{a \in \Gamma_i} \left(\frac{l_a}{L_j} * \frac{1}{\sum_{j \in J} \left(\frac{l_a}{L_j}\right)^{\gamma} \delta_{aj}}\right) \tag{2.8}$$

where l_a denotes the length of link a , L_j denotes the length of path j, and Γ_j denotes the set of links in path j. Further, δ_{aj} is the link-path incidence dummy, and $\delta_{aj} = 1$ if path j otherwise uses link a and 0. When the parameter $\gamma = 0$, the formulation of the path size reduces to the first formulation proposed by Ben-Akiva and Bierlaire (1999).

Recent studies have revealed the pros and cons of the parameter γ, which is designed for penalizing extremely long routes in the choice set. According to Bekhor et al. (2006), a positive value of the parameter γ can improve model fit empirically. However, Frejinger and Bierlaire (2007) have pointed out that the utility adjustments with a positive γ may be questionable and thus the corresponding path probabilities may be unrealistic and illogical.

To avoid theoretical confusion, the basic formulation of path size is adopted in this paper, which takes the form

$$PS_j = \sum_{a \in \Gamma_i} \left(\frac{l_a}{L_j} * \frac{1}{\sum_{j \in J} \delta_{aj}}\right). \tag{2.9}$$

The path size logit model is given by

$$P_{nj} = \frac{\exp\left(\beta^T X_{nj} + \ln\left(PS_j\right)\right)}{\sum_{j' \neq j} \exp\left(\beta^T X_{nj'} + \ln\left(PS_{j'}\right)\right)}. \tag{2.10}$$

As the value of the path size is between 0 and 1, the logarithm makes the utility adjustment factor non-positive. Precisely, the perceived utility of a unique route will not be adjusted while the perceived utility of a partially overlapped route is reduced based on the degree of overlapping.

It is shown by a handful of studies that the path size logit (PSL) performs well in addressing network overlaps and thus is widely used in route choice modeling and large-size traffic assignment problems (Prashker and Bekhor 2004). The current research effort is a direct application of PSL and generalized PSL to different travel modes and trip purposes to reveal factors that are significant to route choice.

Broach et al. (2012) applied PSL in bicycle route choice. The GPS data used was collected from 164 cyclists in Portland, Oregon. It is noted that the parameter of the path size term should not be arbitrarily set to 1. Instead, the coefficient of the path size from the model estimation is 1.8, which is significantly different from 1. Moreover, the model statistical fit is significant with estimated parameters. Lue and Miller (2019) modeled pedestrian route choice using PSL with smartphone GPS data. The generalized formulation of path size was used with gamma equal to 14. Model results showed that the parameter of the logarithm of path size is 1.5. The model also included a stochastic alternative route generation process that accounts for the randomness that a path would be included in the choice set, with the consideration of limited route information.

Due to the closed-form formulation of the choice probability and simple analytical form, the path size term can be flexibly embedded in many logit family models. Yang et al. (2016) applied a nest logit model to modeling the route choice of battery electric vehicles. In route selection, the choice of charging or no charging makes two nests of the model, and alternatives under the second nest are modeled by a path size logit. The empirical results of the study demonstrated the effectiveness of the nested structure.

Dalumpines and Scott (2017) applied path size logit to the network of Halifax, Canada which contains 21,782 nodes and 26,917 links. This study compared route choice under different trip purposes (work and shopping) with 13-month GPS trajectories. Results show that the adjustment factor for utility and other parameters should separate for different trip purposes. Moreover, the path size factor captures

overlaps in route alternatives, but its effect may be reduced by other latent factors that affect the route choice behavior.

However, Duncan et al. (2020) have pointed out the key issue of path size logit. Without consideration of choice probability, all routes in the choice set contributed equally to the path size term. This results in the path size term of routes that are likely to be used being negatively affected by the routes that are not likely to be used. This issue is not solved with the generalized form. Since the parameter lambda is meant to punish long routes, the path size term is heavily affected by link length when lambda is large. This formulation is internally inconsistent in the assessment of feasibility and scaling parameters. The proposed adaptive path size logit solves these problems by setting the link contribution to the path size term proportional to choice probability ratios. It is shown that the proposed model can be reduced to a fixed-point problem; thus, a unique solution is guaranteed.

2.2.3. C-logit

The availability of computation powers and the emphasis on behavioral aspects make the random utility model the mainstream method of route choice models. With the consideration of network topologies, the probit model has been proposed where the stochastic random errors are assumed to be normally distributed and correlated among overlapped alternatives. However, the probit model cannot be solved explicitly like logit family models; instead, the Monte Carlo method is necessary to estimate choice probabilities. As a result, the application of the probit model in the path enumeration and network assignment is limited by the computation burden from the Monte Carlo draws

and straightforward behavioral interpretations. From the point of improving computation and calibration efficiency, a C-logit formulation is proposed by Cascetta, Russo, and Vitetta (1997), with the overlapping alternatives deal with a punishing term to the systematic utility named as a CF. The authors give three analytical formulations; only the one used for this paper is displayed, which is expressed as

$$CF_j = -\theta \ln \left(\sum_{a \in \Gamma_i} \frac{l_a}{L_j} * N_a \right) \tag{2.11}$$

where $N_a = \sum_{j \in J} \delta_{aj}$, δ_{aj} is the link-path incidence dummy, and $\delta_{aj} = 1$ if path j uses link a and 0 otherwise.

The choice probability is given by

$$P_{nj} = \frac{\exp\left(\beta^T X_{nj} + CF_j\right)}{\sum_{j' \neq j} \exp\left(\beta^T X_{nj'} + CF_{j'}\right)}. \tag{2.12}$$

The direct application of C-logit is limited due to a lack of theoretical guidance to the proper functional form of the CF.

Tests of network performance of the C-logit model see the following. Russo and Vitetta (2003) formulate a link-based assignment model with C-logit and propose a solution algorithm called D-C-logit that can provide closed-form path choice probabilities and avoids explicit path enumeration. Zhou, Chen, and Bekhor (2012) provide the route length-based and congestion-based stochastic user equilibrium with two formulations of the CF: a flow-dependent cost form and a flow-independent form. Results show that the CF can effectively overcome the overlapping issue. The equilibrium flow of the multinomial logit model is significantly different from that of

length-based C-logit and congestion-based C-logit, with a slight difference observed between the last two models.

2.2.4. Nested logit and cross nested logit

The nested or hierarchical logit is an extension of multinomial logit designed for nested choice sets. The multinomial logit model is characterized by the IIA with the independent error term. According to Hall (1999), this property is described as "Where any two alternatives have a non-zero probability of being chosen, the ratio of one probability over the other is unaffected by the presence or absence of any additional alternative in the choice set." Since the IIA property guarantees independence among alternatives, it fails in addressing the situation with correlated choice alternatives. One option to account for the overlapping route problem is to add a correction or adjustment factor to the systematic utility, leading to path size logit and C-logit. The second option is to address the problem with correlated errors, which includes the multinomial probit model, the mixed logit model, and part of the family of generalized extreme value models such as the nested logit, cross-nested logit, and the paired combinatorial logit model.

Distinguished from other models in the generalized extreme value model family that have a single-level structure, the nested logit model is characterized by a two-level tree structure. The choice set is first clustered or grouped to several subsets called nests, denoted by $C_1, C_2, \dots, C_m, \dots C_M$, and the MNL model is then applied to each of the nest, which brings in the probability of choosing alternative j_c in nest C_m:

$$P_n(c, j_c) = p(j_c|C_m)p(C_m|J) = \frac{\exp(\mu V_{C_m})}{\sum_{m'} \exp(\mu V_{C_{m'}})} * \frac{\exp(\mu_m \beta^T X_{nj})}{\sum_{j_{c'} \in C_m} \exp(\mu_m \beta^T X_{nj'})} \quad (2.13)$$

where μ and μ_m are scaled parameters and V_{C_m} is the composite utility of nest C_m. The cross nested logit model is a direct extension to the nested model, with the relaxation that one alternative is allowed to be assigned in more than one nest. For applications, see Abbe, Bierlaire, and Toledo (2007); Lai and Bierlaire (2015); and Papola (2004).

2.2.5. Mixed logit or error component logit

Unlike path size logit and C-logit, which is a direct extension of the multinomial logit model with a utility adjustment factor, the mixed logit model is designed to capture individuals' preference heterogeneity by the combined structure of error terms or by random parameters. There are two formulations of random utilities for the mixed logit model: random parameters formulation and error component formulation. The first formulation is written as

$$U_{ntj} = {\beta_n}^T x_{ntj} + \varepsilon_{ntj}. \quad (2.14)$$

Within a mixed logit setting, each individual would face a set of different choice situations, denoted by T and indexed by t. Note that the taste parameter is no longer the same for all individuals. Instead, it is defined as a random variable, whose behavior is reflected by some statistical distributions. The other formulation of the random utility is expressed as

$$U_{ntj} = \alpha^T x_{ntj} + {d_n}^T x_{ntj} + \varepsilon_{ntj} \quad (2.15)$$

where α denotes the vector of fixed variables and d_n denotes the vector of random variables with 0 means. The relationship between the two formulations is then revealed

by decomposing the random parameter β_n to a mean of α and deviation d_n. By substituting the decomposition of β_n to the first formulation, it is shown that these two formulations are mathematically equivalent.

The probability of the mixed logit model is given by the mixed distribution

$$P_{nj} = \int \frac{\exp(\beta_n{}^T X_{nj})}{\sum_{j' \neq j} \exp(\beta_n{}^T X_{nj'})} f(\beta_n) d(\beta_n) \tag{2.16}$$

where $f(\beta_n)$ is the probability density function of β_n. Since the behavior of β_n is unknown, different probability distributions are assumed to represent $f(\beta_n)$. The most commonly used ones are normal distribution, log-normal distribution (when parameters are restricted to a specific sign), uniform distribution, and triangular distribution. Interested readers are referred to Hensher and Greene (2003) for a detailed case study to compare the performance of those distributions.

Growing attention to the prevalence of heterogeneity that influences discrete choice behavior has led to a line of research with the application of the mixed logit model. The preference heterogeneity or taste heterogeneity is observed in that individuals consider different attributes or perceive the same attributes differently when making choices, which is captured by this model's random parameter setting. In the freight industry, drivers' taste heterogeneity in mode choice is represented by data categorization according to industry activities or commode types. Within each category, the mode choice is assumed to be homogeneous and thus modeled with multinomial logit or nested logit. However, Arunotayanun and Polak (2011) have pointed out that, although a strong difference is shown between categories, a substantial amount of inter-

individual heterogeneity remains among carriers. This conclusion is reached with a comparison between a multinomial logit and a mixed logit that are applied respectively in each category. The other point of the study was reached from a latent class analysis, which indicates that commodity type is not necessarily the determining factor in the context of taste heterogeneity. Instead, a combination of different attributes provides better justification.

Due to the complexity of human thought in decision making, an extension of the mixed logit model has been developed as a treatment for intra-individual taste heterogeneity. The random parameters in the model are then specified by each user in each choice situation. A detailed exploration of the performance of mixed logit models in recovering inter- and intra-individual preference is given by Hess and Train (2011). The simulation results have shown that model performance in retrieving the true patterns of heterogeneity would be limited by the number of choice situations. The model that includes intra-personal specification provides excellent recovery of heterogeneity, yet requires high computation time. There are other modeling procedures within a mixed logit framework that aim at deriving the taste heterogeneity. Hong et al. (2017) modeled the random taste parameters with a joint probability density function instead of assuming that they are independent of each other. It is indicated that both mean standard error and mean absolute error were reduced with this modeling procedure.

Another type of heterogeneity is called scale heterogeneity. Since the mathematical form of the utility function only includes observable attributes, the effect of unobservable factors on an individual’s choice behavior is indicated by scale

heterogeneity. Specifically, if an individual's choice behavior is dominated by observable attributes, a massive magnitude of the coefficient of utility attributes is determined. In contrast, if the individual's choice behavior is dominated by unobservable factors, a small magnitude of coefficients should be used instead. Such a phenomenon is called "scale heterogeneity," which is reflected by the magnitude of utility parameters and is inversely and analytically proportional to the error component's standard deviation. Greene and Hensher (2010) compared several forms of multinomial logit, which include the multinomial logit, the standard mixed logit, the mixed logit with extension to include scale heterogeneity, and the generalized mixed logit model (Fiebig et al. 2010), with both taste and scale heterogeneity included. It is found that a model that only includes scale heterogeneity performs worse than both the standard mixed logit model and the generalized mixed logit model. Recently, Hess and Train (2017) have further examined the source of correlations among utility parameters. The latent class and scale-adjust latent class (Magidson, Vermunt, and Madura 2004) are included in comparison, except for the mixed logit and the generalized mixed logit. The mixed logit model with full covariance allows all correlation forms and captures both scale and preference heterogeneity, as does the scale-adjusted latent class model. It is noted that the model specification does not necessarily include all sources of correlation and should be data specified.

The mixed logit model is also characterized by its high flexibility. On the one hand, with the error component formulation, an appropriate choice of variables for the error components makes the mixed model approximation of other logit family models.

For example, with an indicator variable categorizing the nests, the mixed logit is an analog of the nested logit. If the error components are defined for each alternative, the mixed logit is then an analog of heteroskedastic logit (Train 2003). On the other hand, variables that enter the error components have random parameters, whereas variables that sit in the utility function's systematic part have a fixed parameter, which allows for flexible parameter structures. Tsirimpa, Polydoropoulou, and Antoniou (2007) examined traveler's behavior using three different mixed logit models with variables specific to the departure time change, route change, and no change entering the error component. These models realistically predicted the traveler's pattern of switching departure time and the route and were shown to outperform the multinomial models in goodness-of-fit with the same setting. Zimmermann et al. (2018) applied a recursive formulation of the utility function using the Bellman equation to model activity-travel scheduling decisions. The current location and activity are defined as state, and combinations of activity, location, and transport modes are defined as actions. The mixed structure is reflected by the random parameter of different travel modes.

Moreover, various model specifications make the mixed logit model widely applied in modeling discrete choice. The n in the model, for instance, can be either defined as each individual or as groups of individuals from a broad scope. An example is given by Han, Ren, and Bao (2020). They used a mixed distribution of three groups of individuals (low-income, medium-income, and high-income) to model the market share of railway modes ($n = 3$ in this case).

Similar to the multinomial logit, the mixed logit cannot address the overlapping route alternative issue when modeling route choice. Extensions such as mixed path size logit, mixed paired combinatorial logit (Shahhoseini, Haghani, and Sarvi 2015), and the mixed nested logit (Haghani, Sarvi, and Shahhoseini 2015) have been proposed as remedies for this issue. Tang et al. (2020) modeled taxi route choice with mixed path size logit using GPS trajectory data. With the customer generation process incorporated, five periods in the context of different times of day were modeled respectively with five models. It is worth mentioning that the parameter settings (fixed or random) were different for each model, even though the attributes were the same. The mixed path size logit was shown to be superior to path size logit and multinomial logit.

2.3. Parameter estimation procedure of mixed logit models

2.3.1. Maximum simulated likelihood

Simulation is the most commonly used method for solving the mixed logit model. Specifically, the parameter estimation results from maximizing the simulated likelihood, which is derived from the average of numerous sequences of draws. For cross-sectional data, where each individual is faced with one choice situation (i.e., $t = 1$), the simulated likelihood is expressed as

$$SLL = \sum_{n=1}^{N} \ln \left(\frac{1}{R} \sum_{r=1}^{R} P_n(j|\beta_{r,n})\right) \tag{2.17}$$

where $r = 1,2, \dots R$ denotes the rt$^{\text{h}}$ simulation draws, with R being the total number of draws. Additionally, $P_n(j|\beta_{r,n})$ denotes the choice probability logit probability of individual n choosing choice j given β_n on the r^{th} draw.

For penal data, where each individual is faced with a set of choice situations (i.e., $t > 1$), the simulated likelihood is written as

$$SLL = \sum_{n=1}^{N} \ln \left(\frac{1}{R}\sum_{r=1}^{R}\left(\prod_{t=1}^{T} P_{n,t}(j|\beta_{r,n})\right)\right). \quad (2.18)$$

Here, $P_{n,t}(j|\beta_{r,n})$ denotes the choice probability logit probability of individual n choosing choice j in choice situation t given β_n on the r^{th} draw.

More details of the sampling procedure for different settings of mixed logit models with cross-sectional or penal data can be found in Train (2003) and Hess and Train (2011).

2.3.2. Variational Bayes inference

The choice probability from the mixed distribution can be solved with Bayesian analysis. Instead of directly assessing the value of the high-dimensional integration, the Bayesian procedure approximates the true value interested through the establishment of Bayesian inference. The first step is introducing the latent variables or hidden variables, denoted by φ, which bridges the gap between the observed data and the unknown parameters interested. With Bayes' theorem, the conditional probability of latent variables, given the observed data, also known as the posterior distribution, is then derived as

$$\mathrm{p}(\varphi|\mathrm{data},\lambda) = \frac{\mathrm{p}(\mathrm{data}|\varphi,\lambda)\mathrm{p}(\varphi,\lambda)}{\mathrm{p}(\mathrm{data},\lambda)} \quad (2.19)$$

where λ represents unknown parameters to be estimated.

The evaluation of the posterior distribution is called the Bayesian inference problem. Explicit analytical form for the prior joint distribution $p(\varphi, \lambda)$ and the likelihood (i.e., the conditional probability of an unknown parameter given observed data $p(\text{data}|\varphi, \lambda)$) is addressed by construction, leaving the main difficulty of Bayes inference in the computation of marginal density of the data and the evidence. Mathematically, this density is computed by integrating the latent variables:

$$p(\text{data}, \lambda) = \int p(\text{data}|\varphi, \lambda)\ p(\varphi, \lambda) d\varphi. \tag{2.20}$$

However, in many cases of interest, this integral is high dimensional and is thus intractable. As a result, the assessment procedure of this integral falls into two categories: a sampling-based procedure and variational inference procedure. The sampling-based Monte Carlo method, particularly the Markov Chain Monte Carlo (MCMC) simulation, is an established paradigm in estimating the posterior density. The general idea of MCMC is to evaluate the density of interest stochastically with the stationary distribution of an ergodic Markov chain, which enables repeated sampling from the joint posterior distribution. The final empirical estimate of the posterior density is based on the subset of samples collected. By construction, the sampling approach returns an unbiased estimation and converges toward the true posterior with increased dependent samples. Consequently, the MCMC algorithm has been widely investigated and a handful of state-of-art developments has been proposed such as the Metropolis-Hastings algorithm and Gibbs sampler.

Notwithstanding that the exactness of the solutions is guaranteed with the MCMC algorithm, it is not always easily applicable when the model is complicated or

when the dataset is large due to the slow convergence rate, high storage requirement, and long computation time under such circumstances. According to Depraetere and Vandebroek (2017), the assessment of convergence of the Markov chain remains open. However, the variational inference procedure is characterized by computational efficiency and scalability to large datasets compared to the sampling procedure. Instead of directly assessing the joint posterior density, variational inference constructs an approximated posterior density deterministically with an explicit and tractable analytical, usually parametric, form. The critical problem of variational inference is to find an approximate density as close as possible to the true posterior, where the closeness is measured by the Kullback-Leibler (KL) divergence. Since the KL divergence is non-negative, the minimization of the KL divergence is transferred into the maximization of the evidence lower bound (ELBO) that is restricted by the marginal data likelihood. In this case, the variational inference procedure is also called an optimization-based procedure (Zhang et al., 2019). Without a sampling process, the variational inference does not require a large memory for samples, and computation time is reduced accordingly.

The development of variational inference fosters a wide application, including computational biology, neuroscience, computer vision, robotics, and text processing and recognition (Blei, Kucukelbir, and McAuliffe 2017). As it is a machine learning process, recent research interests are to achieve the scalability for large data sets, to relax the limitation on the class of models, to improve the accuracy of posterior approximation, to change the model structure with a neural network embedded, and to extend its

application on Bayesian deep learning (C. Zhang et al. 2019). One of the most commonly used types of variational inferences is the mean-field variational inference (MFVI). The variational density (i.e., density chosen as an approximation of the true posterior) is factorized by the density of each group of latent variables. It is assumed that latent variables are independent of between groups, with each group expressed by a tractable family of densities, which is recognized by the exponentiated partial expectation of the log joint density. Braun and McAuliffe (2010) first applied the MFVI in the mixed logit discrete choice models, with two formulations of the Bayesian structure. The empirical Bayes formulation only assumes a prior distribution of the taste parameter, while the hierarchical Bayes formulation additionally assumes hyperprior distributions for both the mean and variance of the prior distribution of the taste parameter. The results showed that MFVI has significantly reduced the computation times of both models for all scenarios compared to the MCMC algorithm, yet it does not suffer in prediction accuracy.

Following this trend, Tan (2017) extended the MFVI in the mixed logit model that applies to large datasets with a stochastic gradient ascent algorithm. In the context of parameter updates, three methods are compared, namely, Laplace approximations for the variational density, fixed point update via variational message passing, and stochastic linear regression. Moreover, the half-t distribution was used as the hyperprior distribution of the taste parameter's standard deviation and was shown to practically perform better than the inverse Wishart distribution that was used in the study of Braun and McAuliffe (2010) and Huang and Wand (2013). The stochastic linear regression was

the most accurate and stable one among the three methods. The accuracy of variational message passing closely follows the stochastic linear regression, yet failure in convergence was encountered. Using the same hierarchical Bayes formulation as used by Braun and McAuliffe (2010), Depraetere and Vandebroek (2017) explored the approximation and bounding method for evaluating the log-sum exponential term of the ELBO. Results implied that the approximation method was more accurate than the bounding method, and the proposed quasi-Monte Carlo approximation performed equally well as the second-order Taylor series approximation used in Braun and McAuliffe (2010).

The other line of research focuses on the application of MFVI on various types of mixed logit models. Bansal et al. (2020) developed the MFVI process for the mixed logit model with a combination of fixed and stochastic parameters. Krueger et al. (2020) constricted this process for mixed models with inter-and intra-individual heterogeneity.

3. TRUCK ROUTE CHOICE MODELING

3.1. Model formulation

In this section, a mixed C-logit model is introduced, which is solved via the establishment of three different formulations, resulting in three Bayesian approaches: BMCL, HBMCL1, and HBMCL2, respectively. First, we introduce a generalized form of the mixed logit model with the utility adjustment factor, named the adjusted mixed multinomial logit model (AMML). Next, we show that the proposed mixed C-logit is a specific type of AMML. The Bayesian formulations and the numerical tests included in this dissertation are then based on the proposed mixed C-logit model. In what follows, the AMML is specified in detail.

Let N denote the set of individuals and $|N|$ be the total number of individuals. The set of available alternatives for a trucker's O-D travel is denoted by J, which is indexed by j, and the total number of available routes is $|J|$. In this mixed setting of route choice, each individual $n \in N$ will make multiple trips on the same O-D pair with each trip associated with a choice situation, denoted by T_n. In addition, each faces a total of $|T_n|$ choice situations. In this study, a choice situation corresponds to a traffic situation, characterized by observed attributes of the route alternatives, such as the travel time, the toll price, and others.

Let X denote the set of attributes matrixes. Each individual $n \in N$, under choice situation $t \in T_n$, is endowed a matrix of attributes $x_{nt} \in X$, with x_{ntj} representing its j^{th}

row. Analytically, the utility function of alternative j for an individual given some choice situation is written as

$$U_{nt}^{j} = \beta_n{}^{T} x_{ntj} + \theta_{nj} * a_j + \varepsilon_{nt}^{j}. \quad (3.1)$$

The term U_{nt}^{j} in Equation 3.1 is the perceived utility of route j of the n^{th} driver in under the t^{th} choice situation. Further, β_n denotes the $K * 1$ coefficient for choice attributes, where K is the total number of explanatory variables considered in this study. In some studies, this term is named as the taste parameter, because it reflects the individual preferences or tastes of the decision-makers for the observed attributes.

It is worth noting that the most significant difference between the multinomial logit model and the mixed logit model is the specification of the taste parameters. In the multinomial logit model, the taste parameter β is assumed to be constant over individuals, whereas in the mixed logit model, the taste parameter β_n is individual based. It is assumed, in this study, that β_n only contains random coefficients. The preference heterogeneity of individuals is then reflected by the interaction between the random coefficient and its covariate. Mathematically, the preference heterogeneity around the mean of a random parameter is equivalent to its coefficient of variation.

Additionally, x_{nt} denotes the $|J| * K$ attribute matrix where the term x_{ntj} in Equation 3.1 denotes the j^{th} row that contains attributes for calculating part of the systematic component of the utility of the j^{th} alternative.

To account for the roadway overlapping in the network, this study adopted a utility adjustment factor a_j. The detailed mathematical form of this term is discussed in the following section. In Equation 3.1, it is a fixed value for all drivers $n \in N$ in all

choice instances $t \in T_n$ of alternative j. Additionally, θ_{nj} is the coefficient of the utility adjustment factor, and it is assumed to be random instead of a fixed value.

The last term in Equation 3.1, ε_{nt}^{j} represents the random component of the utility function of choice j, which is assumed to be stochastic for each individual in each situation. This term can also be called the stochastic random error and is assumed to be Gumbel (generalized extreme value type II) distributed in this model.

3.1.1. The utility adjustment factor and choice probability

Network topology characteristics are necessary considerations for route choice models. One efficient and straightforward procedure to include the roadway correlation information in modeling is adding a utility adjustment factor to the deterministic part of the utility. In Equation 3.1, we show a generalized form of the utility adjustment factor, a_j. In the literature, there are two alternatives to the mathematical form of this term. One form is the CF. The substitution of the CF with the a_j leads to an extension of the classic logit model, the C-logit model. The other form is the path size, which leads to the extension of the path size logit. Both models have been used for modeling route choices, yet seldom have studies compared the performance of the two models due to the various functional forms proposed for the CF. Only Ben-Akiva et al. (2004) have shown a small discrepancy in model goodness of fit with the two alternatives. Moreover, the path size logit has been extended to a mixed setting, named the mixed path size logit model in existing studies, while there is no such extension for the C-logit model. In this case, this study adopts a properly selected form of the CF. Applied to a mixed setting where each

individual faces a certain number of choice situations, the proposed model is thus named the mixed C-logit model.

In our model, with a necessary reformulation, the term CF is defined with the following functional form:

$$CF_j = -\ln\left(\sum_{a\in\Gamma_j} \frac{l_a}{L_j} * N_a\right) \tag{3.2}$$

where l_a is the length of link a, L_j is the length of path j, $N_a = \sum_{j\in J} \delta_{aj}$ and δ_{aj} is the link-path incidence dummy, and $\delta_{aj} = 1$ if path j uses link a and 0 otherwise. Theoretically, N_a is the number of paths trespassing the link a; it represents a weight to the link, and Γ_j is the set of links in path j.

When the CF is substituted into Equation 3.1, it becomes

$$U_{nt}^j = {\beta_n}^T x_{ntj} + \theta_{nj} * CF_j + \varepsilon_{nt}^j. \tag{3.3}$$

This model is referred to as the mixed C-logit model from now on.

For a thorough understanding of the utility adjustment term, the other alternative to the path size is also provided here, even though it is not examined in the numerical tests. The path size is defined as follows:

$$PS_j = \sum_{a\in\Gamma_j} \left(\frac{l_a}{L_j} * \frac{1}{\sum_{j\in J} \delta_{aj}}\right) \tag{3.4}$$

where the l_a, L_j, Γ_j, and δ_{aj} are the same as C-logit. In both Equation 3.2 and 3.4, a larger number of paths trespassing a link result in a smaller term in the utility function. They each represent a different way to consider the negative effect of the commonality of a shared link.

It is pointed out that the existing mixed path size logit uses a single fixed parameter for its path size term. As a contrast, in the mixed C-logit model proposed in this study, θ_{nj} is assumed to be a random variable that follows a certain probability distribution, which is estimated for other coefficients.

The conditional probability of the mixed C-logit with utility adjustment factor is given by

$$P(y_{nt}^{j} = 1|\beta_n, x_{ntj}, \theta_{nj}, CF_j) = \frac{exp(\ \beta_n^T x_{ntj} + \theta_{nj}\ * CF_j)}{\sum_{j\prime \in J} exp(\ \beta_n^T x_{ntj\prime} + \theta_{nj}\ * CF_{j\prime})} \tag{3.5}$$

where x_{nt} denotes the $K * |J|$ attributes matrix for driver $n \in N$ in his or her choice event $t \in T$; ${x_{ntj}}^T$ is the j^{th} row vector, and y_{nt} is the observed choice vector of individual n at situation t. The input matrix is given by $[x_{nt}\ a]$ with dimension $(K + 1) * |J|$. $a = [a_1 \dots a_j]$. The marginal probability of individual n at situation t may be calculated in the following way:

$$P(y_{nt}^{j} = 1|x_{ntj}, CF_j) = \iint \frac{exp(\ \beta_n^T x_{ntj} + \theta_{nj} * CF_j)}{\sum_{j\prime \in J} exp(\ \beta_n^T x_{ntj\prime} + \theta_{nj} * CF_{j\prime})} d\beta_n d\theta_{nj}. \tag{3.6}$$

3.1.2. Bayesian modeling procedure

To solve the proposed mixed C-logit model stated in the previous section with a utility function of Equation 3.3, we develop a Bayesian procedure that specifically deals with this particular utility function to address the truckers' route choice decision. With the different levels of hierarchy, three Bayesian formulations are introduced in this section: BMCL, HBMCL1, and HBMCL2.

In Bayesian modeling, the information contained in the priors affects the performance of model prediction. Theoretically, a strong informative prior would have better performance if the prior belief is close to the fact. However, the prior guess can be vague and non-informative when there is little prior information available. In this case, we applied different priors for the coefficients to explore how the informativity of the prior affects the posterior prediction of the mixed C-logit model. The informativity of priors is revealed by the correlation among the coefficients. A weak correlation would result in a distribution with a high level of dispersion, while a strong correlation would lead to a concentrated distribution for the coefficients. Since HBMCL1 has additional assumptions for the mean and variance of the coefficients, the correlation among its coefficients would be thus stronger than BMCL would be, and thus the HBMCL1 would be more informative than would BMCL. Similarly, with more assumptions for the hyperparameters, HBMCL2 is more informative than are HBMCL1 and the BMCL. As a strongly informative prior would result in a more complicated model, the examination of different priors would also shed light on the balance between model complexity and prediction precision. The distributions for the priors are chosen empirically.

First, we introduce a basic formulation with two levels of hierarchy. The coefficient β_n is assumed to follow a k-dimensional multivariate Gaussian distribution with mean ξ and variance Ω. The term θ_n is assumed to follow a Laplace distribution with the location parameter c and the scale parameter d. The following summarizes the above assumptions:

$$\theta_n | c, \mathrm{d} \sim Laplace(c, \mathrm{d})$$

$$\beta_n|\xi,\Omega \sim Gaussian_k(\xi,\Omega)$$

$$y_{nt}|x_{nt},\beta_n,a,\theta_n \sim MNL(x_{nt},\beta_n,a,\theta_n)$$

The first model is named the Bayesian mixed C-logit (BMCL) model in this dissertation. Graphically, the structure of this model is shown in Figure 3.1. The top-level parameters $\xi,\Omega,$ and c,d describe the distribution of individual preferences of the population.

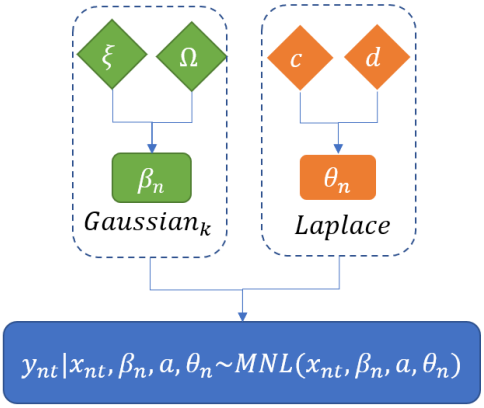


Figure 3.1 Modeling structure for BMCL.

Compared to the first model, the mean and variance of β_n is further assumed to have additional distributions, leading to an additional hierarchy in the second model. Specifically, ξ is assumed to follow another k-dimensional multivariate Gaussian distribution with mean ξ_0 and variance Ω_0, and Ω follows an inverse Wishart distribution

with scale matrix s and degree of freedom ν. This full Bayesian model can be written as follows:

$$\Omega|\nu, \mathrm{s} \sim IW_k(\nu, \mathrm{s}^{-1})$$

$$\xi|\xi_0, \Omega_0 \sim Gaussian_k(\xi_0, \Omega_0)$$

$$\theta_n|c, \mathrm{d} \sim Laplace(c, \mathrm{d})$$

$$\beta_n|\xi, \Omega \sim Gaussian_k(\xi, \Omega)$$

$$y_{nt}|x_{nt}, \beta_n, a, \theta_n \sim MNL(x_{ntj}, \beta_n, a, \theta_n)$$

The second model is named the hierarchical Bayesian mixed C-logit model with inverse Wishart prior for the variance (HBMCL1). The graphic modeling structure is displayed in Figure 3.2.

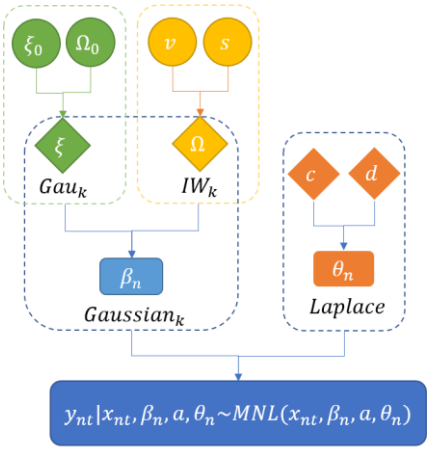


Figure 3.2 Modeling structure for HBMCL1.

Another popular prior distribution for the variance of β_n is Huang's half-t distribution. With this prior, the third model provided for this dissertation is summarized as follows:

$$\mathrm{g}|G_k \sim IG_k\left(0.5, \frac{1}{{G_k}^2}\right), G = [G_1, G_2 \dots G_k]^T$$

$$\Omega|\nu, \mathrm{g} \sim IW_k\big(\nu + k, 2\nu diag(g)\big), g = [g_1, g_2 \dots g_k]^T$$

$$\xi|\xi_0, \Omega_0 \sim Gaussian_k(\xi_0, \Omega_0)$$

$$\theta_n|c, \mathrm{d} \sim Laplace(c, \mathrm{d})$$

$$\beta_n|\xi, \Omega \sim Gaussian_k(\xi, \Omega)$$

$$y_{nt}|x_{nt}, \beta_n, a, \theta_n \sim MNL(x_{nt}, \beta_n, a, \theta_n)$$

In this case, the third model in this dissertation is named hierarchical Bayesian adapted mixed multinomial logit model with Huang's half-t prior for the variance (HBMCL2). The graphical display of the model structure is illustrated in Figure 3.3.

Compared to the HBMCL1, where the Ω is assumed to follow an inverse Wishart distribution, HBMCL2 defines the scale matrix as a diagonal matrix with the diagonals following an inverse gamma distribution with a shape parameter $\frac{1}{2}$ and a scale parameter $\frac{1}{{G_k}^2}$.

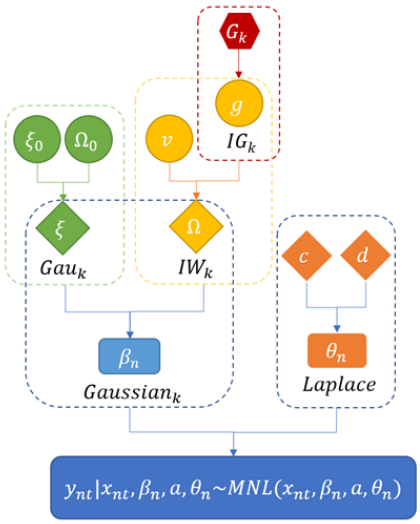


Figure 3.3 Modeling structure for HBMCL2.

3.2. Mean-field variational inference

After reformulating the mixed C-logit model with different prior distributions for the parameters of interest, the next step is to find out their posterior distributions, that is, $p(\beta_n, \theta_n | data, model\ parameters)$ for each model. Since there are no closed forms for those full conditional posterior distributions, the variational Bayesian inference is applied to estimate those posteriors.

The variational inference estimates distributions by finding tractable densities that are close to the true density. The closeness is measured by the KL divergence. The KL divergence is a positive value, with a smaller value indicating a better estimation. Let φ denote the parameters to be estimated, let $q(\varphi)$ define the variational distribution selected for estimating parameters, and let $p(\varphi|data)$ define the true posterior density. The KL divergence is given by

$$KL[q(\varphi)||p(\varphi|data)] = E_{q(\varphi)}\, log\frac{q(\varphi)}{p(\varphi|data)}$$
$$= E_{q(\varphi)}\log\big(q(\varphi)\big) - E_{q(\varphi)}\log\big(p(\varphi, data)\big) \tag{3.7}$$
$$+ E_{q(\varphi)}\log\big(p(data)\big)\,.$$

Interested readers are referred to, for example, Blei, Kucukelbir, and McAuliffe (2017) for a detailed description of the KL divergence.

Then the ELBO is defined as

$$ELBO = E_{q(\varphi)}\log\big(p(\varphi, data)\big) - E_{q(\varphi)}\log\big(q(\varphi)\big). \tag{3.8}$$

By introducing the ELBO, we can simplify the problem by getting rid of the marginal distribution of the data $p(data)$, which is a constant that does not affect the minimization of the KL divergence. Eventually, the minimization of the KL divergence is equivalent to the maximization of the ELBO. Analytically, this problem can be written as

$$\varphi^* = arg\min_{\varphi} KL[q(\varphi)||p(\varphi|data)] = arg\max_{\varphi} ELBO. \tag{3.9}$$

With the MFVI, the joint distribution of parameters $q(\varphi)$ can be represented as a factorized distribution of each parameter, as parameters are assumed to be independent of each other. In this case, the variational probabilities, as well as the ELBOs of the aforementioned Bayesian models, can be defined with closed forms. The maximization problem can then be solved with the block coordinate ascent algorithm. In what follows, we explain the variational inference for the three Bayesian models posed in the previous section in detail. In Section 3.3, we explain the corresponding block coordinate ascent algorithm for those models.

3.2.1. Variational inference of BMCL

As mentioned, the variational inference finds the posterior distribution by approximating it with a factorized distribution. For BMCL, a family of factorized distributions Q is defined as $Q = \{ q(\lambda|\varphi') : \lambda \in \Lambda, \varphi' \in \Phi' \}$ where $q(\lambda|\varphi') = \prod_{n \in N} q(\lambda_n|\varphi'_n)$. Additionally, $\varphi'_n \in \Phi'$ are variational parameters that represent $q(\cdot)$, and Φ' denotes the set of variational parameters. This section specifies the distribution of each $\varphi'_n \in \Phi'$ for BMCL.

The formulation of BMCL has the simplest structure among the three models, with a one-level prior distribution for β_n and θ_n, respectively. Its posterior distribution is written as

$$p(\beta_n, \theta_n|data, \xi, \Omega, \mathrm{c}, \mathrm{d}) = \frac{p(\lambda, y_{nt}|x_{nt}, CF, \varphi)}{p(y_{nt}|x_{nt}, a, \varphi)}$$
$$= \prod_{n=1}^{N} \frac{p(\theta_n|c, \mathrm{d})p(\beta_n|\xi, \Omega) \prod_{t=1}^{T} p(y_{nt}|x_{nt}, \beta_n, CF, \theta_n)}{\int p(\theta_n|c, \mathrm{d})p(\beta_n|\xi, \Omega) \prod_{t=1}^{T} p(y_{nt}|x_{nt}, \beta_n, CF, \theta_n)\, d\,\beta_n d\theta_n dc dd} \quad (3.10)$$

where

$\lambda \in \Lambda$ and Λ is the set of taste parameters to be estimated: $\Lambda = \{\beta_{1:n}, \theta_{1:n}\}$.

$\varphi \in \Phi$ and Φ is the set of modeling parameters: $\Phi = \{\xi, \Omega, \mathrm{c}, \mathrm{d}\}$.

The variation inference requires a proper section of densities to each variational parameter. For each variational factor $q(\lambda_n|\varphi'_n)$, the optimal density is proportional to the exponentiated expectation of the logarithm of the joint density of y and other variational parameters (Krueger et al., 2020). With this procedure, the corresponding

optimal density of each variational parameter in Φ' is listed in Table 3.1; $\Phi' = \{\mu_n, \Sigma_n, c_\theta, d_\vartheta\}$.

Table 3.1 List of variational parameters (BMCL).

Parameters to be estimated λ_n	Modeling parameter	Distribution	Variational parameters φ'_n	Distribution
$\beta_{1:n}$	ξ, Ω	Gaussian(ξ, Ω)	μ_n, Σ_n	Gaussian(μ_n, Σ_n)
$\theta_{1:n}$	c, d	$Laplace(c, d)$	c_θ, d_ϑ	$Laplace(c_\theta, d_\vartheta)$

The variational joint density can then be written as

$$q(\lambda|\varphi') = q(\beta_{1:n}, \theta_{1:n}|\mu_n, \Sigma_n, c_\theta, d_\vartheta) = \prod_{n=1}^{N} q(\beta_n|\mu_n, \Sigma_n) \prod_{n=1}^{N} q(\theta_n|c_\theta, d_\vartheta). \quad (3.11)$$

The ELBO of BMCL is

$$\begin{aligned} ELBO_B = &\sum_{n=1}^{N} E_{q(\varphi_B)} \log\big(q(\varphi_B)\big) + \sum_{n=1}^{N} E_{q(\varphi_B)} \log(p(\beta_n|\xi, \Omega)) \\ &+ \sum_{n=1}^{N} E_{q(\varphi_B)} \log(p(\theta_n|c, \mathrm{d})) \\ &+ \sum_{n=1}^{N} \sum_{t=1}^{T} E_{q(\varphi_B)} \log(p(y_{nt}|x_{nt}, \beta_n, a, \theta_n)). \end{aligned} \quad (3.12)$$

The first term of Equation 3.12 is the sum of the information entropy of $q(\beta_{1:n}|\mu_n, \Sigma_n)$ and $q(\theta_{1:n}|c_\theta, d_\vartheta)$, which is

$$\sum_{n=1}^{N} E_{q(\varphi_B)} \log\big(q(\varphi_B)\big) = H\big(q(\beta_{1:n}|\mu_n, \Sigma_n)\big) + H\big(q(\theta_{1:n}|c_\theta, d_\vartheta)\big)$$
$$= \frac{1}{2}\sum_{n=1}^{N} \log((2\pi e)^K|\Sigma_n|) + \log(2d_\vartheta e). \tag{3.13}$$

The second term of Equation 3.12 is the negative cross-entropy or relative entropy of $p(\beta_n|\xi, \Omega)$ relative to $q(\beta_{1:n}|\mu_n, \Sigma_n)$, which can be written as $-H(q,p) = -H(p) - KL(q||p)$. More specifically,

$$\sum_{n=1}^{N} E_{q(\varphi_B)} \log(p(\beta_n|\xi, \Omega))$$
$$= -\frac{N}{2}\log((2\pi)^K|\Omega|) - \frac{1}{2}tr\{\Omega^{-1}\sum_{n=1}^{N}(\Sigma_n + (\mu_n - \xi)(\mu_n - \xi)^T)\}. \tag{3.14}$$

Similar to the second term, the third term of Equation 3.12 is negative cross-entropy of $p(\theta_n|c, \mathrm{d})$ relative to $q(\theta_{1:n}|c_\theta, d_\vartheta)$, which is

$$\sum_{n=1}^{N} E_{q(\varphi_B)} \log(p(\theta_n|c, \mathrm{d})) = \frac{d}{d_\vartheta} e^{\frac{-|c-c_\theta|}{d}} + \log\left(\frac{d_\vartheta}{d}\right) - 1 - \frac{1}{d_\vartheta}|c - c_\theta|$$
$$- \log(2de) \tag{3.15}$$

where B in the subscript is short for BMCL.

With the functional forms listed above, the estimation of the ELBO has a closed-form except for the last term. Braun and McAuliffe (2010) provided the application of

first-order delta estimation of the expectation of the log-sum-exponential term, and we adopt this estimation in this study. The last term in Equation 3.12 is

$$\sum_{n=1}^{N}\sum_{t=1}^{T} E_{q(\varphi_{BG})} \log\left(p\left(y_{nt}\middle|x_{ntj},\beta_n,a,\theta_n\right)\right)$$

$$\begin{aligned} &= \sum_{n=1}^{N}\sum_{t=1}^{T}[\left[{\mu'}_n{}^T\ \mu_\theta\right][x_{nt}\ a]y_{nt} \\ &- \log\left(\sum_{j=1}^{J} e^{[{x_{ntj}}^T\ a_j][{\mu'}_n{}^T\ \mu_\theta]^T}\right) - \frac{1}{2}tr({x_{nt}}^T(diag(z_{nt}) \\ &- z_{nt}{z_{nt}}^T)x_{nt}\Sigma_n)] \end{aligned} \tag{3.16}$$

where

$$z_{nt} = \frac{e^{[x_{nt}\ a][{\mu'}_n{}^T\ \mu_\theta]^T}}{1^T e^{[x_{nt}\ a][{\mu'}_n{}^T\ \mu_\theta]^T}} \tag{3.17}$$

where $\lambda.$ denotes the set of unknown parameters to find out for each model and $\varphi.$ denotes the set of variational parameters. In addition, $\mu_n = [{\mu'}_n{}^T\ \mu_\theta]^T$; let μ_θ be the mean of θ_n and ${\mu'}_n$ be the mean of β_n.

3.2.2. Variational inference of HBMCL1

Similar to BMCL, the joint distribution of variational parameters is represented by $Q = \{\, q(\lambda|\varphi'): \lambda \in \Lambda, \varphi' \in \Phi'\}$, and the joint distribution can be factorized according to the basic assumption of the MFVI that the variational parameters are mutually independent. The posterior distribution of HBMCL1 is

$$p(\beta_n, \theta_n, \xi, \Omega | data, c, d) = \frac{p(\lambda_{H1}, y_{nt} | x_{nt}, CF, \varphi_{H1})}{p(y_{nt} | x_{nt}, a)}$$
$$= \frac{p(\xi|\xi_0, \Omega_0) p(\Omega|\nu, \mathrm{s}) \prod_{n=1}^{N} p(\theta_n|c, \mathrm{d}) \prod_{n=1}^{N} p(\beta_n) \prod_{t=1}^{T} p(y_{nt})}{\int p(\xi) p(\Omega) \prod_{n=1}^{N} \int p(\theta_n|c, \mathrm{d}) \prod_{n=1}^{N} \int p(\beta_n) \prod_{t=1}^{T} p(y_{nt}) d\beta_n d\theta_n d\xi d\Omega} \quad (3.18)$$

where

$\lambda_{H1} \in \Lambda_{H1}$, and Λ_{H1} is the set of parameters to be estimated: $\Lambda_{H1} = \{\beta_{1:n}, \theta_{1:n}, \xi, \Omega\}$.

$\varphi_{H1} \in \Phi_{H1}$, and Φ_{H1} is the set of modeling parameters: $\Phi_{H1} = \{\mathrm{c}, \mathrm{d}, \xi_0, \Omega_0, \nu, \mathrm{s}\}$.

The subscript in λ_{H1}, φ_{H1}, Λ_{H1}, and Φ_{H1} is short for HBMCL-1.

Additionally, ${\varphi_{H1}}' \in {\Phi_{H1}}'$ are variational parameters, and ${\Phi_{H1}}'$ is the set of variational parameters. Further, ${\Phi_{H1}}' = \{\mu_n, \Sigma_n, \mu_\xi, \Sigma_\xi, \omega, \Upsilon^{-1}, c_\theta, d_\vartheta\}$. The corresponding optimal density of each variational parameter in ${\Phi_{H1}}'$ is listed in Table 3.2.

Table 3.2 List of variational parameters (HBMCL1).

Parameters to be estimated λ_{H1}	Modeling parameter	Distribution	Variational parameters ${\varphi_{H1}}'$	Distribution
$\beta_{1:n}$	ξ, Ω	$\mathrm{Gaussian}(\xi, \Omega)$	μ_n, Σ_n	$\mathrm{Gaussian}(\mu_n, \Sigma_n)$
ξ	ξ_0, Ω_0	$\mathrm{Gaussian}(\xi_0, \Omega_0)$	μ_ξ, Σ_ξ	$\mathrm{Gaussian}(\mu_\xi, \Sigma_\xi)$
Ω	ν, s	$\mathrm{IW}(\nu, \mathrm{s})$	ω, Υ	$\mathrm{IW}(\omega, \Upsilon^{-1})$
$\theta_{1:n}$	c, d	$Laplace(c, d)$	c_θ, d_ϑ	$Laplace(c_\theta, d_\vartheta)$

Then the joint density of the variational parameters listed above is

$$q(\lambda_{H1}|\varphi_{H1}) = q\left(\beta_{1:n}, \theta_n, \xi, \Omega|\mu_\xi, \Sigma_\xi, \Upsilon, \omega, \mu_n, \Sigma_n, c_\theta, d_\vartheta\right)$$
$$= q\left(\xi\middle|\mu_\xi, \Sigma_\xi\right) q(\Omega|\Upsilon^{-1}, \omega) \prod_{n=1}^{N} q(\theta_n|c_\theta, d_\vartheta) \prod_{n=1}^{N} q(\beta_n|\mu_n, \Sigma_n). \tag{3.19}$$

With the joint density of both true posterior distribution $p(\beta_n, \theta_n, \xi, \Omega|data, c, d)$ and its approximation $q(\lambda_{H1}|\varphi_{H1})$ defined above, the ELBO of HBMCL1 can be expressed as the following:

$$\begin{aligned} ELBO_{H1} = & \sum_{n=1}^{N} E_{q(\varphi_{H1})} \log\left(q(\beta_n|\mu_n, \Sigma_n)\right) + \sum_{n=1}^{N} E_{q(\varphi_{H1})} \log\left(q(\theta_n|c_\theta, d_\vartheta)\right) \\ & + \sum_{n=1}^{N} E_{q(\varphi_{H1})} \log\left(q\left(\xi\middle|\mu_\xi, \Sigma_\xi\right)\right) \\ & + \sum_{n=1}^{N} E_{q(\varphi_{H1})} \log\left(q(\Omega|\Upsilon^{-1}, \omega)\right) \\ & + \sum_{n=1}^{N} E_{q(\varphi_{H1})} \log(p(\beta_n|\xi, \Omega)) + \sum_{n=1}^{N} E_{q(\varphi_{H1})} log(p(\theta_n|c, d)) \\ & + \sum_{n=1}^{N} E_{q(\varphi_{H1})} \log(p(\xi|\xi_0, \Omega_0)) + \sum_{n=1}^{N} E_{q(\varphi_{H1})} \log(p(\Omega|\nu, \mathrm{s})) \\ & + \sum_{n=1}^{N} \sum_{t=1}^{T} E_{q(\varphi_{H1})} \log(p(y_{nt}|x_{nt}, \beta_n, a, \theta_n)). \end{aligned} \tag{3.20}$$

Each term of Equation 3.20 is specified as follows.

The first two terms are similar to the first term of $ELBO_B$:

$$\sum_{n=1}^{N} E_{q(\varphi_{H1})} \log\big(q(\varphi_{H1})\big) = H\big(q(\beta_{1:n}|\mu_n, \Sigma_n)\big) + H\big(q(\theta_{1:n}|c_\theta, d_\vartheta)\big)$$
$$= \frac{1}{2}\sum_{n=1}^{N} \log((2\pi e)^K|\Sigma_n|) + \log(2d_\vartheta e). \tag{3.21}$$

The second and third terms are each the sum of the information entropy of $q\big(\xi|\mu_\xi, \Sigma_\xi\big)$ and $q(\Omega|\Upsilon^{-1}, \omega)$, which is

$$\sum_{n=1}^{N} E_{q(\varphi_{H1})} \log\Big(q\big(\xi|\mu_\xi, \Sigma_\xi\big)\Big)$$
$$+ \sum_{n=1}^{N} E_{q(\varphi_{H1})} \log\big(q(\Omega|\Upsilon^{-1}, \omega)\big) = H(q\big(\xi|\mu_\xi, \Sigma_\xi\big))$$
$$+ H(q(\Omega|\Upsilon^{-1}, \omega)) \tag{3.22}$$
$$= \frac{1}{2} log\big((2\pi e)^K\big|\Sigma_\xi\big|\big) - \frac{\omega - K - 1}{2} D(\omega, \Upsilon) + \frac{\omega K}{2} + A_\omega(\Upsilon)$$

where

$$D(\omega, \Upsilon) = Klog(2) + \log(|\Upsilon|) + \sum_{i=0}^{K-1} \Psi(\frac{\omega - i}{2}) \tag{3.23}$$

and

$$A_\omega(\Upsilon) = \frac{\omega K}{2}\log(2) + \log\big(\Gamma_K(\omega)\big) + \frac{\omega}{2}\log(|\Upsilon|). \tag{3.24}$$

The $\Psi(\cdot)$ is the digamma function, and $\Gamma(\cdot)$ is the multivariate gamma function where $\Gamma_K(\omega) = \pi^{K(K-1)/4} \prod_{k=0}^{K-1} \Gamma((\omega - i)/2)$ and $\Gamma(\cdot)$denotes the standard gamma function.

The fourth and fifth terms are similar to the second and third terms of the $ELBO_B$:

$$\begin{aligned}\sum_{n=1}^{N} E_{q(\varphi_{H1})} \log(p(\beta_n|\xi,\Omega)) + \sum_{n=1}^{N} E_{q(\varphi_{H1})} log(p(\theta_n|c,d)) \\ = -\frac{N}{2}\log((2\pi)^K) + \frac{N}{2}D(\omega,\Upsilon) - \frac{\omega}{2}tr\Bigg\{\Upsilon^{-1}N\Sigma_\xi \\ + \Upsilon^{-1}\sum_{n=1}^{N}\left(\Sigma_n + (\mu_\xi - \mu_n)(\mu_\xi - \mu_n)^T\right)\Bigg\} + \frac{d}{d_\vartheta}e^{\frac{-|c-c_\theta|}{d}} \\ + \log\left(\frac{d_\vartheta}{d}\right) - 1 - \frac{1}{d_\vartheta}|c - c_\theta| - \log(2de).\end{aligned} \tag{3.25}$$

The sixth and seventh terms are the sum of cross-entropy of $p(\xi|\xi_0,\Omega_0)$ to $q(\xi|\mu_\xi,\Sigma_\xi)$ and $p(\Omega|\nu,\mathrm{s})$ to $q(\Omega|\Upsilon^{-1},\omega)$, which is

$$\begin{aligned}\sum_{n=1}^{N} E_{q(\varphi_{H1})} \log(p(\xi|\xi_0,\Omega_0)) + \sum_{n=1}^{N} E_{q(\varphi_{H1})} \log(p(\Omega|\nu,\mathrm{s})) \\ = -\frac{1}{2}\log((2\pi)^K|\Omega_0|) \\ -\frac{1}{2}tr\{{\Omega_0}^{-1}\left(\Sigma_\xi + (\mu_\xi - \xi_0)(\mu_\xi - \xi_0)^T\right)\} - A_\nu(s^{-1}) \\ + \frac{\nu - K - 1}{2}D(\omega,\Upsilon) + \frac{\omega}{2}tr(s^{-1}\Upsilon).\end{aligned} \tag{3.26}$$

The last term is the same as the last term in $ELBO_B$.

3.2.3. Variational inference of HBMCL2

Similarly, to get the variational distributions for the HBMCL2, we first specify its posterior distribution, which takes the form of

$$p(\beta_n, \theta_n, \xi, \Omega, l|data) = \frac{p(\lambda_{H2}, y_{nt}|x_{nt}, CF, \varphi_{H2})}{p(y_{nt}|x_{nt}, a)}$$
$$= \frac{\prod_{k=1}^{K} p(l)\, p(\xi|\xi_0, \Omega_0) p(\Omega|\nu, \mathrm{s}) \prod_{n=1}^{N} p(\theta_n) \prod_{n=1}^{N} p(\beta_n) \prod_{t=1}^{T} p(y_{nt})}{\int \prod_{k=1}^{K} p(l) \int p(\xi) p(\Omega) \prod_{n=1}^{N} \int p(\beta) \prod_{n=1}^{N} \int p(\theta) \prod_{t=1}^{T} p(y) d\beta d\theta d\xi d\Omega \mathrm{d}l} \quad (3.27)$$

where

$\lambda_{H2} \in \Lambda_{H2}$, and Λ_{H2} is the set of parameters to be estimated.

$\Lambda_{H2} = \{\beta_{1:n}, \theta_{1:n}, \xi, \Omega, l_{1:K}\}$, $\varphi_{H2} \in \Phi_{H2}$, and Φ_{H2} is the set of modeling parameters.

$\Phi_{H2} = \{\mathrm{c}, \mathrm{d}, \xi_0, \Omega_0, G = [G_1, G_2 \dots G_k]^T, \nu\}$.

The subscript in λ_{H2}, φ_{H2}, Λ_{H2}, and Φ_{H2} is short for HBMCL2.

The joint distribution of variational parameters is represented by $Q = \{ q(\lambda_{H2}|\varphi_{H2}'): \lambda_{H2} \in \Lambda_{H2}, \varphi_{H2}' \in \Phi_{H2}'\}$, where $\varphi_{H2}' \in \Phi_{H2}'$ are variational parameters and Φ_{H2}' is the set of variational parameters. In this model, $\Phi_{H2}' = \{\mu_n, \Sigma_n, \mu_\xi, \Sigma_\xi, h_k, g = [g_1, g_2 \dots g_k]^T, \omega, \Upsilon, c_\theta, d_\vartheta\}$. The correspondence between variational parameters and modeling parameters are displayed in Table 3.3.

Table 3.3 List of variational parameters (HBMCL2).

Parameters to be estimated λ_{H1}	Modeling parameter	Distribution	Variational parameters $\varphi_{H1}{}'$	Distribution
$\beta_{1:n}$	ξ, Ω	$\text{Gaussian}(\xi, \Omega)$	μ_n, Σ_n	$\text{Gaussian}(\mu_n, \Sigma_n)$
ξ	ξ_0, Ω_0	$\text{Gaussian}(\xi_0, \Omega_0)$	μ_ξ, Σ_ξ	$\text{Gaussian}(\mu_\xi, \Sigma_\xi)$
$l_{1:K}$	G	$IG\left(0.5, \frac{1}{G_k{}^2}\right)$	$h_k, g = [g_1, g_2 \dots g_k]^T$	$IG(h_k, g_k)$
Ω	ν, G	$\text{IW}(\nu + K - 1{,}2\nu diag(G))$	ω, Υ	$\text{IW}(\omega, \Upsilon)$
$\theta_{1:n}$	c, d	$Laplace(c, d)$	c_θ, d_ϑ	$Laplace(c_\theta, d_\vartheta)$

With the mutual independence assumption of the MFVI, the variational distribution can be factorized as follows:

$$q(\lambda_{H2}|\varphi_{H2}) = q\left(\beta_{1:n}, \theta_{1:n}, \xi, \Omega, \text{G}|\mu_n, \Sigma_n, c_\theta, d_\vartheta, \mu_\xi, \Sigma_\xi, \omega, h_k, g_k\right)$$
$$= q\left(\xi\middle|\mu_\xi, \Sigma_\xi\right) q(\Omega|\omega, \Upsilon) \prod_{k'=1}^{k+1} q(l|h_k, g_k) \prod_{n=1}^{N} q(\beta_n|\mu_n, \Sigma_n) \prod_{n=1}^{N} q(\theta_n|c_\theta, d_\vartheta). \quad (3.28)$$

The ELBO for HBMCL2 can then be described as

$$
\begin{aligned}
ELBO_{H2} = & \sum_{n=1}^{N} E_{q(\varphi_{H2})} \log\Big(q\big((\beta_n|\mu_n, \Sigma_n)\big)\Big) \\
& + \sum_{n=1}^{N} E_{q(\varphi_{H2})} \log\Big(q\big((\theta_n|c_\theta, d_\vartheta)\big)\Big) \\
& + \sum_{n=1}^{N} E_{q(\varphi_{H2})} \log\Big(q\big(\xi|\mu_\xi, \Sigma_\xi\big)\Big) \\
& + \sum_{n=1}^{N} E_{q(\varphi_{H2})} \log\big(q(\Omega|\omega, \Upsilon)\big) + \sum_{k=1}^{k+1} E_{q(\varphi_{H2})} \log\big(q(l|h_k, g_k)\big) \\
& + \sum_{n=1}^{N} E_{q(\varphi_{H2})} \log(p(\beta_n|\xi_0, \Omega_0\ )) \\
& + \sum_{n=1}^{N} E_{q(\varphi_{H2})} \log(p(\theta_n|c, \mathrm{d}\ )) \\
& + \sum_{n=1}^{N} E_{q(\varphi_{H2})} \log(p(\xi|\xi, \Omega)) + \sum_{k\prime=1}^{k+1} E_{q(\varphi_{H2})} \log\left(p\left(l\middle|\frac{1}{2}, G_k\right)\right) \\
& + \sum_{n=1}^{N} E_{q(\varphi_{H2})} \log(p(\Omega|\nu, G\ )) \\
& + \sum_{n=1}^{N} \sum_{t=1}^{T} E_{q(\varphi_{H2})} \log(p(y_{nt}|x_{nt}, \beta_n, a, \theta_n)).
\end{aligned}
\tag{3.29}
$$

Each term in Equation 3.29 is specified as follows. The first two terms are the same as the first two terms of $ELBO_{H1}$. The third term is the entropy of the normal distribution with mean μ_ξ and variance Σ_ξ, which is

$$\sum_{n=1}^{N} E_{q(\varphi_{H2})} \log\left(q\left(\xi|\mu_\xi, \Sigma_\xi\right)\right) = \frac{1}{2} log\left((2\pi e)^K|\Sigma_\xi|\right). \tag{3.30}$$

The fourth and fifth terms are

$$\sum_{n=1}^{N} E_{q(\varphi_{H2})} \log\big(q(\Omega|\omega, \mathrm{g})\big) + \sum_{k=1}^{k+1} E_{q(\varphi_{H2})} \log\big(q(l|h_k, g_k)\big)$$

$$= -\frac{\omega - K - 1}{2} D(\omega, \Upsilon) + \frac{\omega K}{2} + A_\omega(\Upsilon) \tag{3.31}$$

$$+ \sum_{k=1}^{K} \log\left(g_k\right) + \mathrm{K}(h_k + \log\big(\Gamma(h_k)\big) - (h_k + 1)\Psi(h_k)).$$

The sixth, seventh, and eighth terms are the same as the fifth, sixth, and seventh terms of $ELBO_{H1}$. The ninth and tenth terms are the negative relative entropies of $p\left(l\middle|\frac{1}{2}, G_k\right)$ to $q(l|h_k, g_k)$ and $p(\Omega|\nu, G\ )$ to $q(\Omega|\omega, \Upsilon)$. Explicitly,

$$\sum_{k=1}^{k+1} E_{q(\varphi_{H2})} \log\left(p\left(l\middle|\frac{1}{2}, G_k\right)\right) + \sum_{n=1}^{N} E_{q(\varphi_{H2})} \log(p(\Omega|\nu, G\ )) =$$

$$= \frac{\nu - K - 1}{2} D(\omega, \Upsilon) - \log\big(\Gamma_K(\nu + K + 1)\big)$$

$$- \sum_{k=1}^{K} \left(\nu\omega {\Upsilon_{kk}}^{-1} + \frac{1}{{G_k}^2}\right)\frac{h_k}{g_k} \tag{3.32}$$

$$- \frac{\nu + K + 2}{2} \sum_{k=1}^{K} (\log(g_k) - \Psi(h_k)) + \frac{(\nu + K - 1)K}{2} \log\left(\nu\right)$$

$$- \sum_{k=1}^{K} \log(G_k) - \mathrm{K}\log\left(\Gamma\left(\frac{1}{2}\right)\right)$$

where Υ_{kk} is the k^{th} diagonal term in Υ.

3.3. Block coordinate ascent algorithm

With the MFVI formulation of the ELBO as in Equation 3.8 of Section 3.2, the variational parameters in BMCL, HBMCL1, and HBMCL2 can be solved by the block coordinate ascent algorithm, which is described in detail below. It goes through iterations. In each iteration of the algorithm, parameter estimations are updated through the maximization of the ELBO with respect to each parameter while keeping other variables fixed. For example, one iteration of the BMCL is the following: We first update the $\mu_{1:n}$, using $\Sigma_{1:n}$ ξ, c_θ, d_θ and Ω from the last iteration, and then we update the $\Sigma_{1:n}$, using the updated $\mu_{1:n}$ and ξ, c_θ, d_θ, Ω from the last iteration. We continue the procedure for each variable until all of them get updated. Then the algorithm moves on to the next iteration.

According to Braun and McAuliffe (2010), the concavity of the ELBO simplifies the maximization process of each parameter. For BMCL, the ELBO shows concavity on ξ, c_θ, d_θ and Ω^{-1}; the updates of ξ, c_θ, d_θ and Ω are thus derived using the first-order condition of this unconstrained convex optimization problem with closed forms. Similarly, for HBMCL1 and HBMCL2, the updates for $\mu_\xi, \Sigma_\xi, \omega, \Upsilon, c_\theta, d_\theta, g$ are also derived from the first-order conditions in which the nonnegativity of Σ_ξ is guaranteed (Braun and McAuliffe, 2010). There is no closed-form solution for μ_n and Σ_n; thus, their updates are derived directly from $argmax\ ELBO$ by the use of the Broyden-Fletcher-Goldfarb-Shanno optimization solver from Python SciPy Library (Jones et al., 2001).

Algorithm 1: Block coordinate algorithm

Input:
Data:X,y
Hyper-parameters:$\mu_0, \Sigma_0, c_0, d_0, \nu, \xi_0, \Omega_0, G_k, \omega = \nu + N, g_0$
Starting values:$\Phi = \{\mu_{1:n}, \Sigma_{1:n}, c, d, \mu_\xi, \Sigma_\xi, \xi, \Omega, \Upsilon, g\}$
Output: Φ

if *BMCL* **then**
  while *Convergence not reached* **do**
    $\mu_{1:n} \leftarrow \text{argmax}_\mu \, ELBO_{BMCL}$
    $\Sigma_{1:n} \leftarrow \text{argmax}_\Sigma \, ELBO_{BMCL}$
    $c_0 \leftarrow \text{argmax}_c \, ELBO_{BMCL}$
    $d_0 \leftarrow \text{argmax}_d \, ELBO_{BMCL}$
    $\xi \leftarrow \frac{1}{N} \sum_{n=1}^{N} \mu_n$
    $\Omega \leftarrow \frac{1}{N} \sum_{n=1}^{N} \Sigma_n + \text{Cov}(\mu_{1:n})$
else if *HBMCL1* **then**
  while *Convergence not reached* **do**
    $\mu_{1:n} \leftarrow \text{argmax}_\mu \, ELBO_{HBMCL1}$
    $\Sigma_{1:n} \leftarrow \text{argmax}_\Sigma \, ELBO_{HBMCL1}$
    $c_0 \leftarrow \text{argmax}_c \, ELBO_{HBMCL1}$
    $d_0 \leftarrow \text{argmax}_d \, ELBO_{HBMCL1}$
    $\mu_\xi \leftarrow (\Omega_0^{-1} + N\omega\Upsilon)^{-1}(\Omega_0^{-1}\xi_0 + \omega\Upsilon \sum_{n=1}^{N} \mu_n)$
    $\Sigma_\xi \leftarrow (\Omega_0^{-1} + N\omega\Upsilon)^{-1}$
    $\Upsilon \leftarrow (N\Sigma_\xi + S^{-1} + \sum_{n=1}^{N}(\Sigma_n + (\mu_\xi - \mu_n)(\mu_\xi - \mu_n)^T))^{-1}$
else
  while *Convergence not reached* **do**
    $\mu_{1:n} \leftarrow \text{argmax}_\mu \, ELBO_{HBMCL2}$
    $\Sigma_{1:n} \leftarrow \text{argmax}_\Sigma \, ELBO_{HBMCL2}$
    $c_0 \leftarrow \text{argmax}_c \, ELBO_{HBMCL2}$
    $d_0 \leftarrow \text{argmax}_d \, ELBO_{HBMCL2}$
    $\mu_\xi \leftarrow (\Omega_0^{-1} + N\omega\Upsilon)^{-1}(\Omega_0^{-1}\xi_0 + \omega\Upsilon \sum_{n=1}^{N} \mu_n)$
    $\Sigma_\xi \leftarrow (\Omega_0^{-1} + N\omega\Upsilon)^{-1}$
    $\Upsilon \leftarrow (N\Sigma_\xi + 2\nu \, \text{diag}(\mathbf{1}/g_0) + \sum_{n=1}^{N}(\Sigma_n + (\mu_\xi - \mu_n)(\mu_\xi - \mu_n)^T))^{-1}$
    $g \leftarrow \nu\omega\Upsilon_{kk} + \frac{1}{G_k^2} \forall k = 1, ..., k$

Figure 3.4 Block coordinate algorithm.

To evaluate the convergence of this algorithm, the following criterion is used. First, we define a vector that contains the parameters of interest for each model. Then we evaluate its difference between two consecutive iterations. When this difference is negligible, the convergence is considered to have been reached, and the algorithm will stop.

Specifically, for BMCL, HBMCL1, and HBMCL2, we define $\vartheta = [\xi^T, diag(\Omega), c_\theta, d_\vartheta]^T$, $\vartheta = [{\mu_\xi}^T, diag(\Sigma_\xi), diag(\Upsilon), c_\theta, d_\vartheta]^T$ and $\vartheta = [{\mu_\xi}^T, diag(\Sigma_\xi), diag(\Upsilon), \mathrm{g}^T, c_\theta, d_\vartheta]^T$, respectively, with each element being the average over the last five iterations. In addition, ϑ_i^τ denotes the i^{th} element at iteration τ. The algorithm stops when $argmax_i \frac{|\vartheta_i^{\tau+1} - \vartheta_i^\tau|}{\vartheta_i^\tau} < 0.001$.

3.4. Performance evaluation

In the simulation study documented in Section 4, we evaluated the Bayesian models from the perspective of parameter estimation and prediction precision.

The model performance in the parameter estimation is assessed by the root mean squared error (RMSE) for each interested parameter. The RMSE is defined as follows:

$$RMSE(\vartheta) = \sqrt{\frac{1}{M}(\hat{\vartheta} - \vartheta)^T(\hat{\vartheta} - \vartheta)} \tag{3.33}$$

where $\hat{\vartheta}$ is the estimates of ϑ and M is the total number of scalar parameters in ϑ. In HBMCL1 and HBMCL2, the estimation accuracy of $\mu_{1:n}$ and $\Sigma_{1:n}$ is evaluated through their mean $\bar{\mu} = \frac{1}{N}\sum_{n=1}^{N}\mu_n$ and $cov(\mu_{1:n}) = \frac{1}{N}\sum_{n=1}^{N}(\mu_n - \bar{\mu})(\mu_n - \bar{\mu})^T$.

The prediction error is measured by mean and maximum error, which are written as

$$mean\ error = \frac{1}{|J|}\sum_{j\in J} |P_{true}(y_{nt}^{test} = j|X_{nt}^{test}) - P_{predicted}(\widehat{y_{nt}} = j|X_{nt}^{test}, y_{nt}^{test})| \tag{3.34}$$

and

$$max\ error = max_{j\in J}\ (|P_{true}(y_{nt}^{test} = j|X_{nt}^{test}) - P_{predicted}(\widehat{y_{nt}} = j|X_{nt}^{test}, y_{nt}^{test})|) \tag{3.35}$$

where $P_{true}(\cdot)$ denotes the true probability of individual n selecting the j^{th} choice in situation t, while $P_{predicted}(\cdot)$ denotes the corresponding predicted probability. The calculation for each posterior predicted probability is through the MC integration with 1,000,000 draws. According to Braun and McAuliffe (2010), when the number of draws is sufficiently large, the predicted probability from the sample would converge to the true value. Thus, with the MC integration, it is claimed that the difference between the predicted probability and the true probability would be only from the model formulation and the mean field variational inference procedure; no error results from the calculation of the predicted probability.

In Section 6, the proposed models are applied to real data. For model selection and evaluation, a 10-fold cross-validation procedure was applied to each part of the data to compare the proposed Bayesian models and other machine learning models such as the support vector machine and random forests. The main idea behind cross-validation is to split the dataset into different subsets in which the training sample and the testing

sample are drawn independently. With multiple training and testing processes, the model that outperforms others in regards to the selected assessment method is then selected as the best model. The creation of cross-validation dates back to the 1930s, to noticing the problem of training and testing models using the same dataset, which would cause over-optimistic results. A detailed introduction of the cross-validation procedure is given by Arlot and Celisse (2010). The accuracy assessment method used in the cross-validation is $1 - max\ error$.

4. TWO CASE STUDIES ON A THREE-PATH NETWORK

4.1. Simulation setting of Test 1

Two simulation tests are included in this section and Section 4.2 for a thorough examination of the performance of the proposed models using a three-path network, as shown in Figure 4.1. In this network, Node 0 is the origin, and Node 6 is the destination. The numbers next to the links are link distances, which are also displayed in Table 4.1. The CF and the path size are calculated using Equation 3.2 and 3.4 from Section 3. To simulate data for this test, three attributes are considered in this case study, which are the travel time, the toll indicator, and the highway indicator.

The travel time of the three path alternatives, which are continuous variables, are generated from $Uniform(120,180)$, $Uniform(130,200)$, and $Uniform(100,160)$, respectively. The toll indicator and the highway indicator are binary variables and are randomly generated. We set $c = -1.5$ and $d = -0.5$ in each formulation. For BMCL, the β_ns are generated from multivariate Gaussian distribution $Gaussian(\xi, \Omega)$, where $\xi = diag(Identity_3)$ and $\Omega = \frac{1}{2} Identity_3$. For HBMCL1 and HBMCL2, $\xi = diag(Identity_3)$ is set the same as BMCL, while $\Omega \sim IW_3(\nu, \mathrm{s}^{-1})$, $\nu = 5$ and $\mathrm{s}^{-1} = inv[(\nu + |N|) \cdot Identity_3]$ for HBMCL1 and $\Omega \sim IW_3\big(\nu + K, 2\nu diag(g)\big)$, $g = \frac{3}{2} diag(Identity_3)$ for HBMCL2.

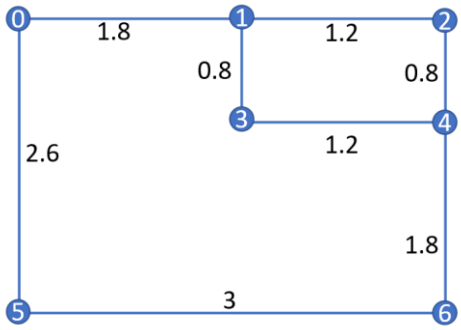


Figure 4.1 Example network.

Table 4.1 Example Summary.

	Link distance			
Path 1	l_{01}=1.8	l_{12}=1.2	l_{24}=0.8	l_{46}=1.8
0-1-2-4-6	Commonality factor: -0.49643688632			
	Path size: 0.67857142857		$ln(PS) = -0.38776553101$	
	Link distance			
Path 2	l_{01}=1.8	l_{13}=0.8	l_{34}=1.2	l_{46}=1.8
0-1-3-4-6	Commonality factor: -0.49643688632			
	Path size: 0.67857142857		$ln(PS) = -0.38776553101$	
	Link distance			
Path 3	l_{05}=2.6	l_{56}=3		
0-5-6	Commonality factor: 0			
	Path size: 1		$ln(PS) = 0$	

The error is generated from $Gumbel(0,1)$; $|N| = 500$ and $|T| = 50$ are used. When simulating the true utility adjustment factor, a fixed value of -0.435 and a group of stochastic values that are drawn from $Laplace(-0.435,0.5)$ are used.

4.2. Results of Test 1

This test aims at exploring the performance of the proposed mixed C-logit model with the utility function displayed in Equation 3.3, with different Bayesian formulations. Three models are compared in this test, namely, the proposed mixed C-logit model (MCL), which has a random coefficient for the CF term; the mixed path size logit model (MPSL), which only uses a fixed coefficient for the path size term; and the original mixed logit model (MXL), which contains no utility adjustment factor. All mixed models are respectively formulated with the three proposed Bayesian structures, which leads to a total of nine mixed models.

By solving the MFVI solution provided in Section 3, 20 replications are conducted for each model. The RMSE of the parameters of interest ($\xi, \Omega, c, d, \beta_{1:n}$) and the mean error and max error in predicting probabilities of each route are displayed in Table 4.2, with mean and variances (in parentheses) over the 20 replications. The corresponding prediction error of the classic multinomial logit is also included for comparison.

The RMSE represents the accuracy of parameter estimation. A smaller RMSE indicates a slight deviation from the true parameter, and thus the corresponding model provides a more precise parameter prediction. In general, HBMCL2 outperforms HBMCL1 and BMCL in parameter estimation. The RMSE of ξ and Ω reduces

significantly with the increase of formulation complexity. For example, the RMSE of ξ with HBMCL2 is approximately 0.08, which is approximately the same as the HBMCL1 and 0.04 smaller than that of BMCL. The RMSE of Ω indicates a similar performance of HBMCL2 and HBMCL1; both are superior to the BMCL. The RMSEs of c, d, and $\beta_{1:n}$ among HBMCL2, HBMCL1, and BMCL are close to each other.

Besides, a smaller variance of the RMSE reflects a more concentrated estimation value of a parameter. It is observed that HBMCL1 provides the most concentrated estimations, followed closely by HBMCL2 and then by BMCL. Take ξ as an example. The variance of ξ's RMSE for most of the models with HBMCL2 and HBMCL1 is smaller than that for the models with the BMCL.

Another trend that is clear in the results comes from the comparison between the MXL, MPSL, and the proposed MCL. It is noted that in most cases, the MCL provides more reliable parameter estimation than the MPSL, and both outperform the MXL.

The prediction error of each model is one of the key concerns for model selection. Obviously, the mixed models with all Bayesian formulations outperform the classic multinomial logit model, with evidence in both mean and maximum prediction errors. The smallest mean error is given by HBMCL1, closely followed by HBMCL2. Since the BMCL is inferior to these two formulations in parameter estimation, a higher mean and maximum predicting error are expected. It is shown, by comparing each mixed model within the same formulation, that the proposed mixed C-logit model is superior to other mixed models. For example, in HBMCL1, the mean prediction errors of the MCL are approximately 0.003 smaller than that of the MPSL and 0.013 smaller than that of

MXL; and the MCL mean errors are 0.004 and 0.012 smaller than that of the MPSL and that of the MXL in HBMCL2. The maximum prediction error displays a similar trend as the mean prediction errors, with MCL in the HBMCL1 and HBMCL2 outperforming MCL in BMCL and the MCL outperforming other models with the same formulation.

To summarize, using a random parameter for the utility adjustment factor achieves better performance in parameter estimation and prediction accuracy, as shown by the outperformance of the MCL over the MPSL, MXL, and the MNL. Additional structures with hyperparameters strengthen model performance, which is indicated by the outperformance of HBMCL2 and HBMCL1 over BMCL.

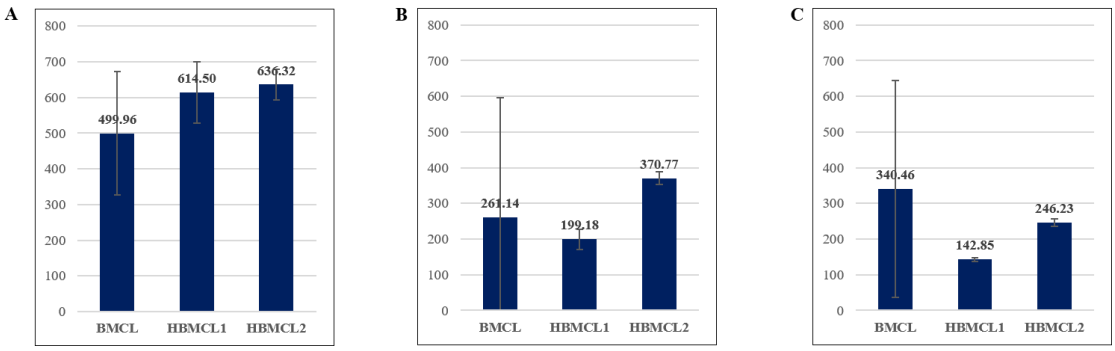


Figure 4.2 Computation time of each model.
(A) MCL using data simulated. (B) MPSL using the same data as with (A). (C) MXL using the same data as with (A).

The other concern in model comparison is the computational efficiency, usually evaluated by the total computation time, which is shown in Figure 4.2 by the number displayed on the top end of each shaded bar.

It is shown in Figure 4.2 (A) and (B) that the computation time of the three formulations all notably decreased when the coefficient of the utility adjustment factor changes from a random value with a Laplace distribution to a fixed value. Specifically, when an additional Bayesian structure associated with this random coefficient is applied to each formulation, the computation time of each model ends with the longest time, 636.32, of HBMCL-2 and the shortest time, 499.96, of BMCL. The computation time of HBMCL1 falls in the middle, with a value of 614.50.

Through comparison of Figure 4.2 (B) and (C), it is found that an additional attribute (the path size) with a fixed coefficient in the utility function does not necessarily result in a longer computation time. For example, in BMCL, without the utility adjustment factor, the computation time is 340.46 and reduces to 261.14 when the path size attribute was added to the model. In contrast, increases in the computation time have been observed in both HBMCL1 and HBMCL2 when the path size term is added to the model.

The error bars in each plot show the standard deviation of the computation time in 20 replications, in which shorter error bars reflect higher model stability. It can be seen, from Figure 4.2 (A) and (B), that the HBMCL2 has the smallest error bar, followed closely by the HBMCL1, and that the BMCL has the largest error bar. This indicates that even though the random coefficient of the utility adjustment factor costs increased the computation time, it helps in the stability of model performance. Similarly, in Figure 4.2 (B) and (C), the HBMCL2 is the most stable Bayesian formulation.

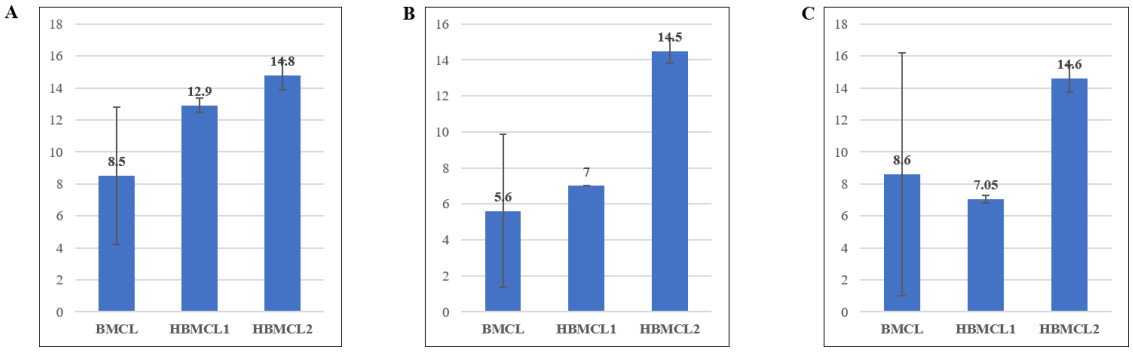


Figure 4.3 Number of iterations of each model.
(A) MCL using data simulated. (B) MPSL using the same data as with (A). (C) MXL using the same data as with (A).

The number of iterations of each model is displayed in Figure 4.3. Overall, the HBMCL2 requires the greatest number of iterations to converge, which conforms to its long computation time shown in Figure 4.2. Generally speaking, adding a utility adjustment factor with either a fixed coefficient or a random coefficient does not affect the number of iterations, which is observed in the similarity between Figure 4.3 (A) and (B). From the perspective of stability, which is shown by the error bars in Figure 4.3, both HBMCL1 and HBMCL2 are more stable than is BMCL. Specifically, HBMCL-2 requires approximately 14 iterations, and BMCL requires five to nine iterations to converge with each formulation in each setting.

Table 4.2 RMSE of parameters and prediction errors, Test 1.

Formulation	Model	RMSE (ξ)	RMSE (Ω)	RMSE (c)	RMSE (d)	RMSE (β)	Max predicted error	Mean predicted error
BMCL	MCL	0.1065 (0.0279)	0.1224 (0.0009)	0.003 (0.0293)	0.1066 (0.005)	0.3364 (0.0101)	0.0126 (0.0015)	0.0077 (0.001)
	MPSL	0.1097 (0.0275)	0.1242 (0.0005)			0.3402 (0.0095)	0.0166 (0.0015)	0.0111 (0.001)
	MXL	0.1454 (0.0256)	0.1259 (0.0003)			0.356 (0.0088)	0.0262 (0.0024)	0.0174 (0.0016)
HBMCL1	MCL	0.079 (0.0148)	0.0823 (0.0191)	0.0044 (0.0223)	0.1113 (0.0046)	0.3353 (0.0043)	0.0122 (0.0011)	0.0084 (0.0007)
	MPSL	0.0843 (0.0156)	0.0755 (0.0217)			0.3387 (0.0045)	0.016 (0.0014)	0.0107 (0.0009)
	MXL	0.1276 (0.0286)	0.0948 (0.9231)			0.3766 (0.0683)	0.0243 (0.0019)	0.0162 (0.0013)

*Variance in parenthesis.

Table 4.2 Continued.

HBMCL2	MCL	0.0766 (0.0222)	0.0828 (0.0269)	0.0012 (0.0093)	0.1237 (0.0069)	0.3788 (0.0169)	0.0123 (0.003)	0.0082 (0.002)
	MPSL	0.0828 (0.0211)	0.0794 (0.0274)			0.386 (0.0169)	0.0157 (0.0038)	0.0105 (0.0025)
	MXL	0.1173 (0.0214)	0.1513 (0.0171)			0.402 (0.0168)	0.0265 (0.0065)	0.0176 (0.0043)
MNL							0.0649 (0.013)	0.0974 (0.0196)

*Variance in parenthesis.

Table 4.3 RMSE of parameters and prediction errors, Test 2.

Model	RMSE (ξ)		RMSE (Ω)		RMSE (c)		RMSE (d)		Max error	
BMCL	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
N=10	0.3078	0.0100	0.4765	0.0776	0.3374	0.0022	0.1213	0.0024	0.0392	0.0003
N=50	0.1816	0.0041	0.1690	0.0066	0.0623	0.0003	0.1179	0.0026	0.0206	0.0000
N=100	0.1318	0.0026	0.1537	0.0032	0.0043	0.0000	0.1180	0.0001	0.0185	0.0001
N=200	0.1219	0.0004	0.1364	0.0050	0.0026	0.0000	0.1175	0.0001	0.0174	0.0000
N=500	0.1029	0.0001	0.1089	0.0018	0.0015	0.0000	0.1187	0.0001	0.0171	0.0000
HBMCL1										
N=10	0.2989	0.0089	0.4029	0.0532	0.0874	0.0003	0.1071	0.0012	0.0364	0.0001
N=50	0.1711	0.0054	0.1630	0.0037	0.0244	0.0000	0.1237	0.0004	0.0189	0.0001
N=100	0.1310	0.0028	0.1235	0.0040	0.0194	0.0000	0.1226	0.0001	0.0169	0.0000
N=200	0.1239	0.0005	0.1156	0.0010	0.0078	0.0000	0.1246	0.0001	0.0159	0.0000
N=500	0.0914	0.0002	0.0739	0.0003	0.0082	0.0000	0.1236	0.0000	0.0157	0.0000

Table 4.3 Continued.

HBMCL2										
N = 10	0.3088	0.0097	0.4933	0.0501	0.1843	0.0009	0.1308	0.0038	0.0364	0.0004
N = 50	0.1698	0.0050	0.1722	0.0064	0.0340	0.0000	0.1039	0.0005	0.0225	0.0001
N = 100	0.1130	0.0009	0.1220	0.0024	0.0335	0.0000	0.1024	0.0000	0.0164	0.0000
N = 200	0.1170	0.0003	0.1033	0.0022	0.0113	0.0000	0.1127	0.0001	0.0159	0.0000
N = 500	0.1001	0.0001	0.0601	0.0006	0.0039	0.0000	0.1381	0.0001	0.0156	0.0000

Table 4.4 Max errors of the C-logit.

N	10	50	100	200	500
Mean	0.0968	0.0960	0.1086	0.0912	0.0959
Variance	0.0009	0.0007	0.0001	0.0001	0.0000

4.3. Simulation setting of Test 2

The second simulation study is designed to test the performance of the proposed model with different numbers of individuals. In this test, $n = 10, 50, 100, 200$, and 500 are used. The network used for this test is the same as in Test 1. Three attributes are considered in this test. The first variable is the travel time, which is continuous and is generated from Lognormal (4.72,0.12), Lognormal (4.75,0.08), and Lognormal (4.68,0.08), respectively for each alternative. The second variable is the toll indicator, which is a random binary variable. The third variable is the highway indicator, and it is a categorical variable with values [0,1,0] for each alternative. Other settings are the same as in Test 1. The simulation results displayed in the following subsection are summarized from 10 replications for each case.

4.4. Results of Test 2

The RMSEs for variables to the interest are displayed in Table 4.2. A clear trend that can be observed is that both the precision of the parameter estimation and the prediction accuracy increased with the growth of the number of individuals. The RMSE of ξ and Ω drastically reduces when N increases from 10 to 100. Additionally, it becomes relatively stable when N is over 100. The RMSE of c and d is small, compared to the RMSE of ξ and Ω. Similarly, the maximum error of each model becomes stable when N is over 200.

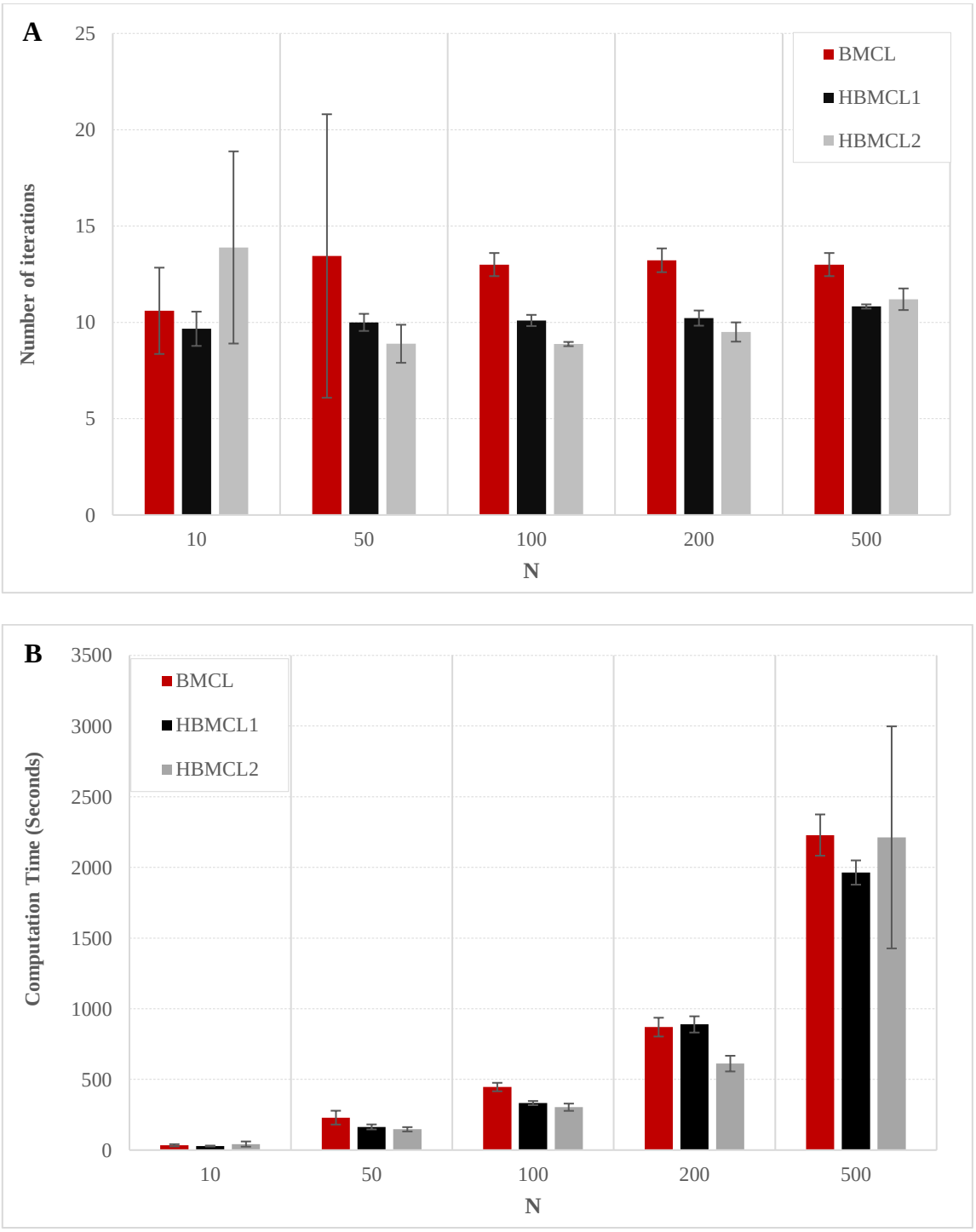


Figure 4.4 Number of iterations and computation time.
(A) Number of iterations of each model with various numbers of *N*. (B) Computation time of each model with various numbers of *N*.

It is shown by the maximum prediction error that the HBMCL1 and the HBMCL2 perform equally well and that both outperform the BMCL. As a comparison, the prediction errors of the classic C-logit model with a fixed parameter for the CF term are displayed in Table 4.3 with various values of *N*. The precision of the proposed models is notably superior to that of the C-logit.

The number of iterations until convergence as well as the computation time of each model is displayed in Figure 4.4 (A) and (B). Model computation time is heavily affected by the number of iterations. BMCL generally needs the greatest number of iterations to converge and thus has the longest computation time. The number of iterations until convergence is not heavily affected by the number of individuals, but when *N* is large, the algorithm is more stable. The computation time is proportional to the number of individuals for all three models.

It is concluded from the second simulation test that additional structures on the variance of the taste parameters contribute to the prediction accuracy and computation time, as shown by the outperformance of HBMCL1 and HBMCL2. With more hyperparameters included, the initial guesses of HBMCL1 and HBMCL2 are closer to the global optimal solution than is the BMCL, which notably reduced the number of iterations for the block coordinate algorithm. Moreover, the increase in the number of individuals enhances model reliability and stability. When *N* is less than 200, diverting cases are observed for each model, while most of the cases end with a converged solution when *N* is over 200.

To summarize, two simulation tests were conducted via a three-path network, which is documented in Section 4.1 and 4.2. In the first simulation test, Test 1, it is found that the proposed mixed C-logit model had better performance in prediction precision and coefficient estimation compared to the mixed path size logit model with a fixed coefficient for the path size term and the classical mixed logit model without a utility adjustment factor. The second finding in Test 1 is that in the proposed Bayesian formulations for the mixed C-logit model, HBMCL2 slightly outperforms HBMCL1, and both outperform BMCL, yet they have a longer computation time. In Test 2, it is found that when the number of individuals in the model exceeds 200, the convergence of all three Bayesian models becomes stable.

5. DATA PROCESSING

5.1. Overview of data processing

The GPS data utilized in this study is from INRIX, which was collected from January to April, 2016, in the Dallas metropolitan area in Texas. Figure 5.1 displays an overview of this dataset.

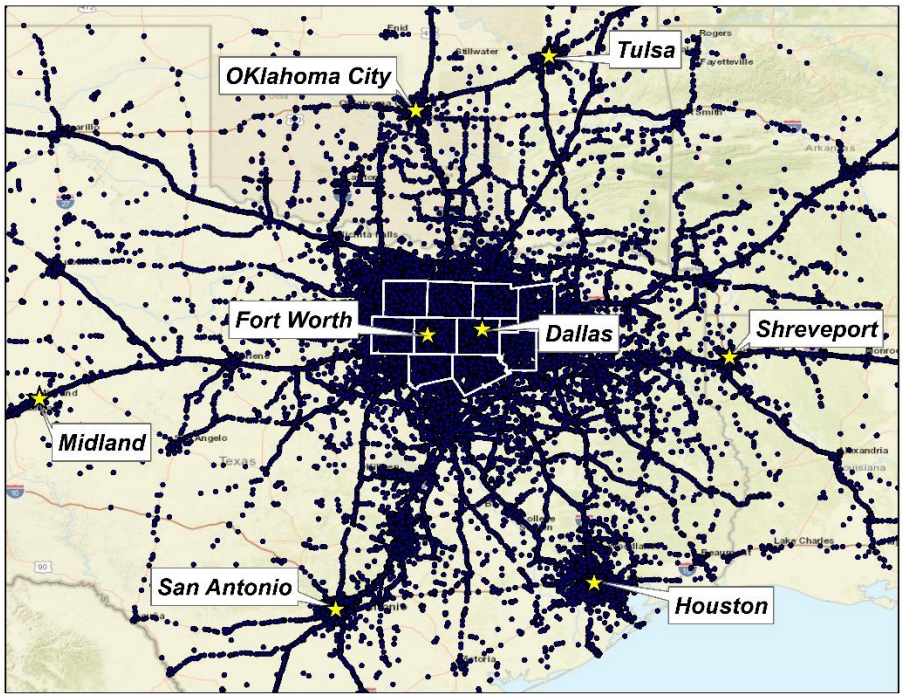


Figure 5.1 GPS dataset overview.

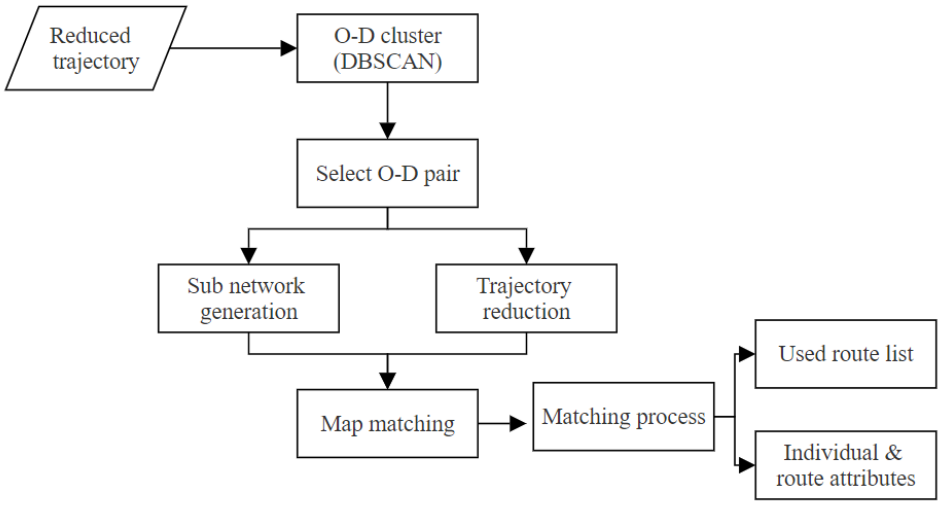


Figure 5.2 Flow chart of the data processing procedure.

The data processing procedure is shown in Figure 5.2 above. The first step of data processing is the identification of origins and destinations using the density-based spatial clustering of application with noise (DBSCAN) method based on the longitudes and the latitudes. This spatial data clustering algorithm provides the total number of clusters and how many data points fall in each cluster and thus makes it possible to capture the most frequently traveled O-D pairs. The second reason to use this clustering procedure is that the recorded start points and end points of trips that start from the same location are usually close to but do not overlap with each other. In this case, direct extraction of O-D pairs using recorded longitudes and latitudes would end with incomplete trip selection.

After applying the DBSCAN method for origins, the origin where most trips started is selected. The associated destinations of the selected origins are shown in Figure 5.3.

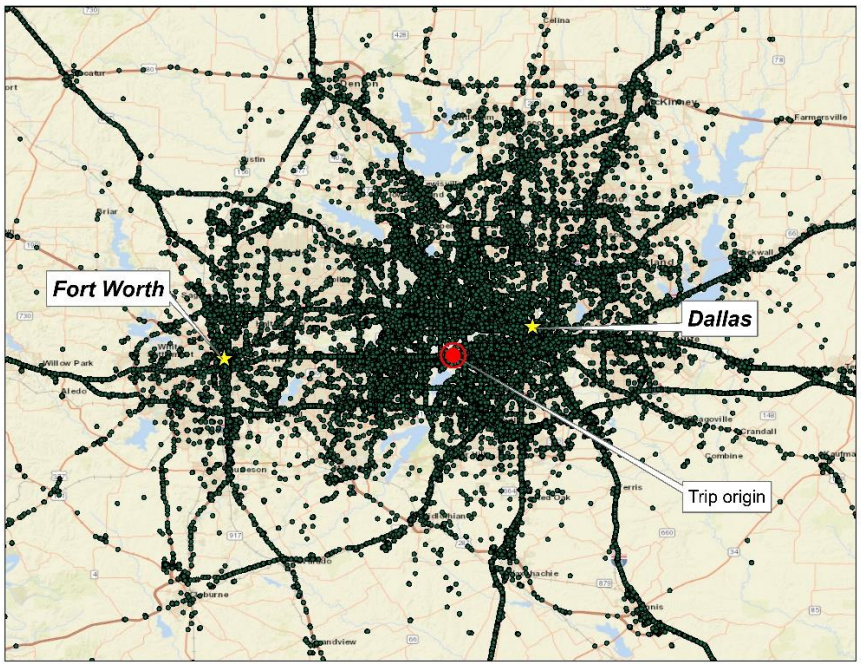


Figure 5.3 Illustration of destinations.

Next, this clustering procedure is applied to the selected destinations. The final results yield two O-D pairs for this study, which are shown in Figure 5.4. The selected O-D pair represent the two scenarios of this study. In the first scenario, the decision-makers are faced with alternatives that contain a toll road. In addition to other factors, their decisions reflect their willingness to use the toll road to save their travel time and

their willingness to pay for the time savings. In the second scenario, decision-makers only need to select among freeways, informed by the route characteristics such as the travel time.

5.2. Difficulty in the data processing

The main difficulty in data processing comes from trajectory cleaning and map-matching. With the O-D pair selected, the corresponding subnetwork that only contains relevant and realistic alternatives with respect to this O-D pair can thus be constructed, using common choice set generation algorithms. In this study, we applied the *k* shortest travel time algorithm. Specifically, for each interested O-D pair, the subnetwork is determined by applying the *k* shortest travel time algorithm on the whole network, with the parameter k being sufficiently large to cover all possible alternatives that drivers may use. However, when k is very large, the subnetworks will resemble the ones displayed in Figure 5.5. The over-detailed network would significantly reduce the computational efficiency of the map-matching process and would result in imprecise capture of reasonable trajectories.

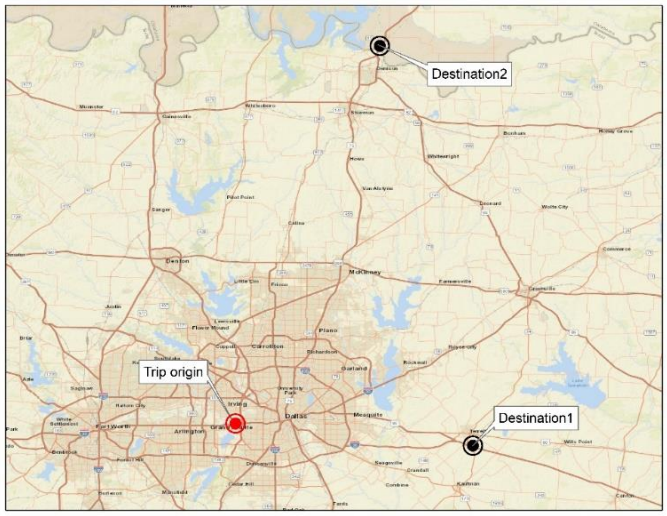


Figure 5.4 Selected O-D pairs.

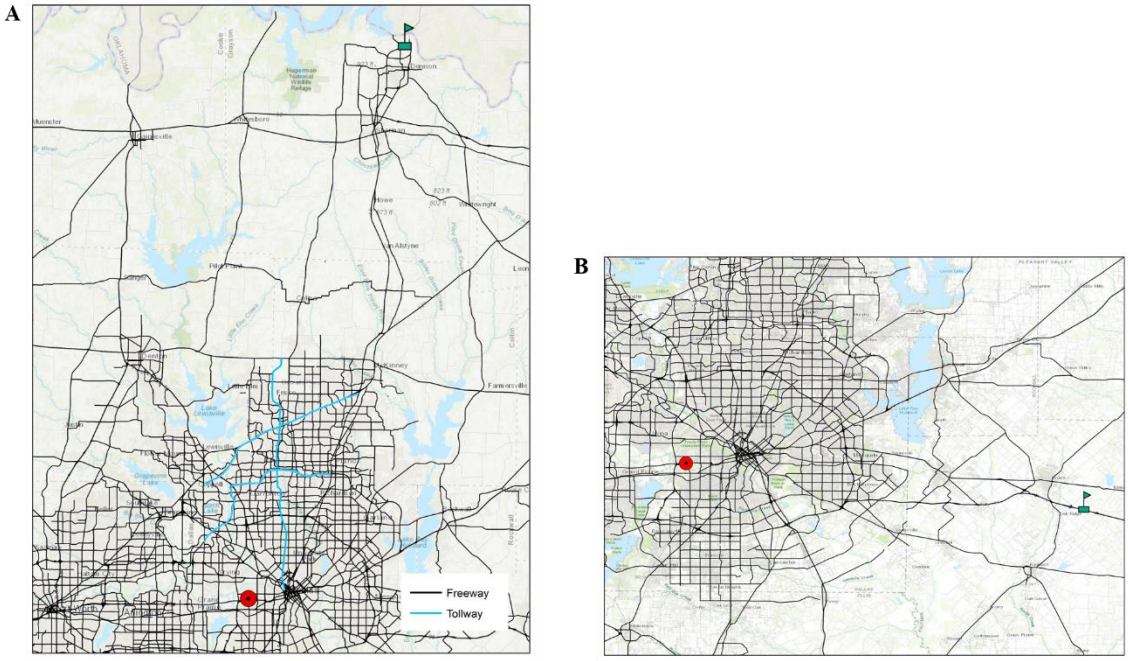


Figure 5.5 Original subnetworks of selected O-D pairs.
(A) Subnetwork of Scenario 1, before reduction. (B) Subnetwork of Scenario 2, before reduction.

Moreover, the extracted GPS trajectories of selected O-D pairs contain sub-trips. Three types of sub-trips are identified in the dataset, which is displayed in Figure 5.6. The first type of sub-trip contains very short detours that are negligible compared to the whole trip, which is named “no detour.” The second type of sub-trips has an extended detour, but the driver has returned to the same link before the detour. These two types of sub-trips affect the travel time only by the detours, such that the trajectory cleaning process is straightforward, with GPS points out of the main trip removed. The cleaning process of the third type of sub-trips requires additional procedures, where the sub-trip starts and ends at different links of the main trip. As a remedy, we first remove the sub-trips’ trajectory points and then fill in the gap on the main trip left in the first step by GPS points on the smallest cost path in the gap.

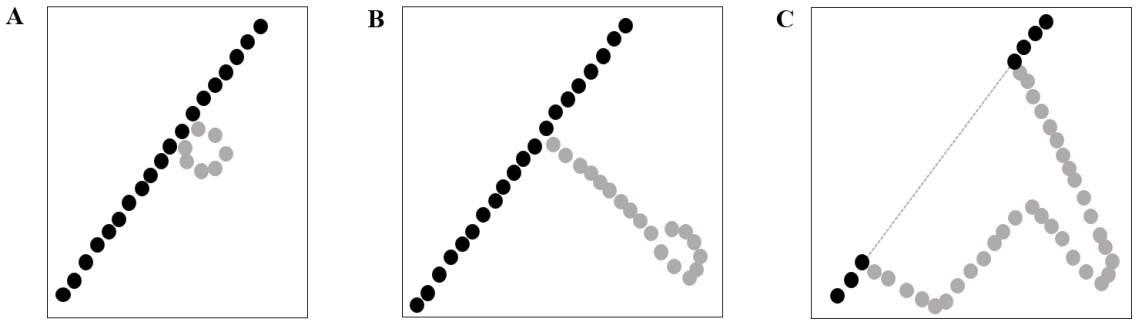


Figure 5.6 Illustration of sub-trip type.
(A) Type 1 sub-trip: no detour. (B) Type 2 sub-trip: detour and return to the main trip. (C) Type 3 sub-trip: detour and use of a different route.

To this end, the first goal of this study is to find a map-matching method to identify the used trajectory and address the sub-trip issue of this dataset. The following section introduces the map-matching process used in this dissertation.

5.3. The map-matching process

In this section, we introduce the map-matching process that can address the sub-trip issue.

There are two inputs for the map-matching process: the extracted trajectory GPS points and the reduced network. Since the generated *k* shortest travel time path should cover all possible routes that have been used, the number *k* would be huge if applied to the original network shown in Figure 5.5. The reduced subnetworks are shown in Figure 5.7 (A) and (B). Node 1 is the origin of both subnetworks, and Node 34 is the destination of Subnetwork 1, while Node 17 is the destination of Subnetwork 2. Each link in the reduced subnetworks contains one or multiple segments in the original network, and therefore, the link attributes are the averages of the included segments weighted by their length.

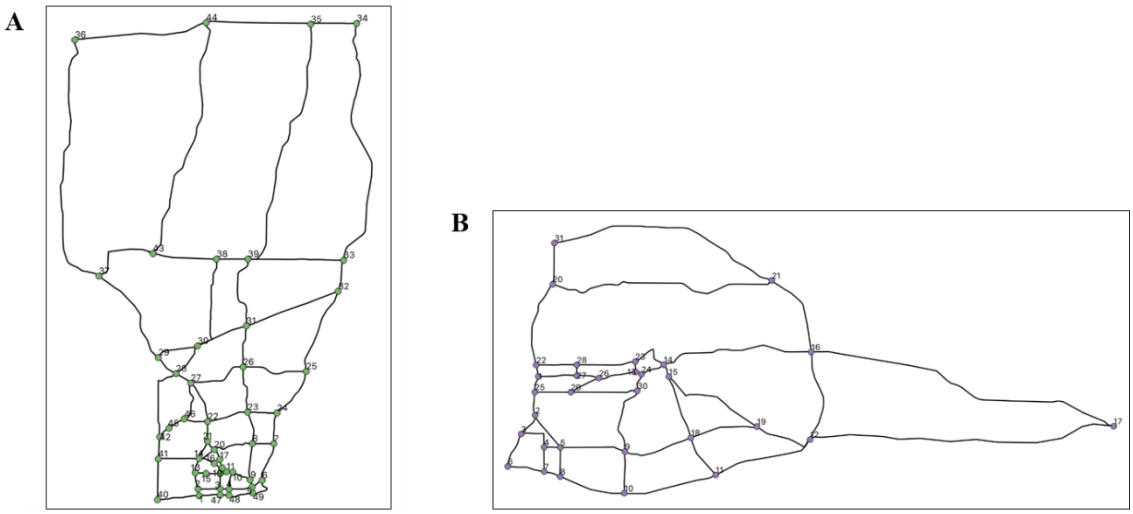


Figure 5.7 Reduced subnetworks.
(A) Subnetwork 1 after reduction. (B) Subnetwork 2 after reduction.

Applying the NetworkX Python library, the *k* shortest time paths are derived using the travel time as the weight. It is worth pointing out that the number *k* should be large enough to cover all GPS trajectory points to avoid wrong matches. In this study, for Scenario 1, *k* is set to 300, with consideration of the combination of interstate and state highways, local highways, and toll roads. For Scenario 2, *k* is set to 75. With the proper least time path generated, the GPS trajectory points are then matched to those paths. The paths identified for each trajectory is determined using the steps as follows:

1. Each GPS point of a trip is assigned a distance to all alternative routes.
2. The route with the shortest cumulative distance for a trip is identified as the path used for this trip.

3. After the used path has been determined for the trip, it is checked whether there is any GPS point that has a distance of over 60 feet. Such points would then be labeled as sub-trips.

The threshold is determined to be 60 feet based on the maximum number of lanes. The maximum number of lanes among all segments is five in this study, and we consider 12 feet for each lane, which gives the threshold of 60. The flow chart of this process is displayed in Figure 5.8.

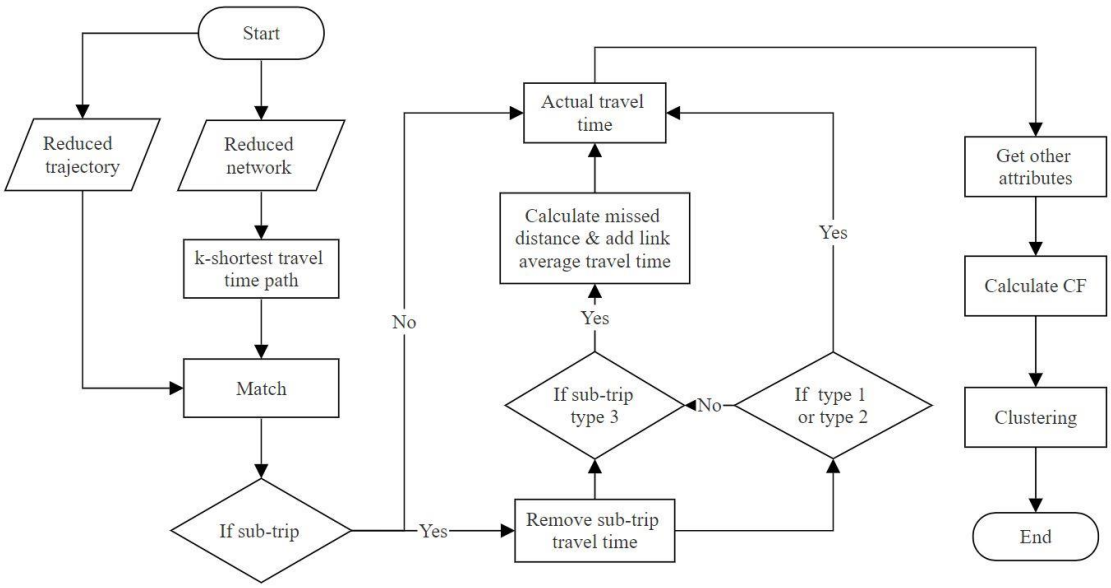


Figure 5.8 Flow chart of the map-matching process.

This procedure labeled the trajectories that belong to sub-trips. The travel time of each trip that contains sub-trips will be adjusted by removing the travel time during the

sub-trip. However, simply removing the time on the sub-trip results in imprecise travel times for trips that contain a Type 3 sub-trip. As the driver used a different link to return to the main trip, there are missed travel distances and time intervals between the start point and the end point of the sub-trip.

In this case, after removing all the labeled points, the distance between the points that are not consecutive is checked. If this distance is longer than the link average speed/60, in other words, the missed travel time between the two points is longer than 1 min, this trip is identified as containing a Type 3 sub-trip. Then we add the travel time between the nonconsecutive points using the average travel time of the missed link length. With the missed travel time for all trips that contain the Type 3 sub-trip calculated, we adjust the travel time of those trips by adding their missed travel time.

As shown in the data, that the majority of the trips contain a sub-trip, the route choice pattern of truckers in this study would be heavily affected by the delivery purpose in addition to factors considered for this study. To address this issue, one efficient method is to cluster the route according to its network geometry using the CF. Taking Scenario 1 as an example, it is shown in Figure 5.10 that clustering using the CF clearly grouped the routes according to their geometric locations. Since the intention of fulfilling the sub-trip leads to the fact that longer travel time routes or a lower level of service routes have been used, route clustering will narrow down the whole choice set into subsets, in which only alternatives that can effectively reach the intermediate destinations of sub-trips are considered. Consequently, routes that fall in the same cluster

should be viewed as the effective potential alternatives that the trucker considers when making route decisions.

However, suppose we cluster using the CF and jointly with the travel time or other route level of service attributes, such as the number of lanes, average truck speed, annual average daily traffic (AADT), and annual average daily truck traffic. In that case, the full choice set cannot be effectively divided into subsets, as shown in Figure 5.9. In contrast to the clustering results with the CF only, many routes fall into multiple subsets that do not meet the goal of classifying sub-trips. It can be seen from Figure 5.11 and 5.12 that when using travel time as well as other route attributes in clustering, the algorithm tends to emphasize travel time over the CF. Thus, the cluster pattern is clearly observed in travel time instead of in the path.

Figures 5.13 to 5.16 display the clustering results for Scenario 2, which shows the same conclusion as for Scenario 1, that using the CF only as clustering input can effectively filter out trips that have the same sub-trips and that using a combination of multiple attributes cannot reach this goal. The CF cannot directly reflect the location of the intermediate pickup/delivery stops. However, using the CF as the input for the clustering algorithm can split the entire network into different parts. The pick/delivery stops will then be included in each part.

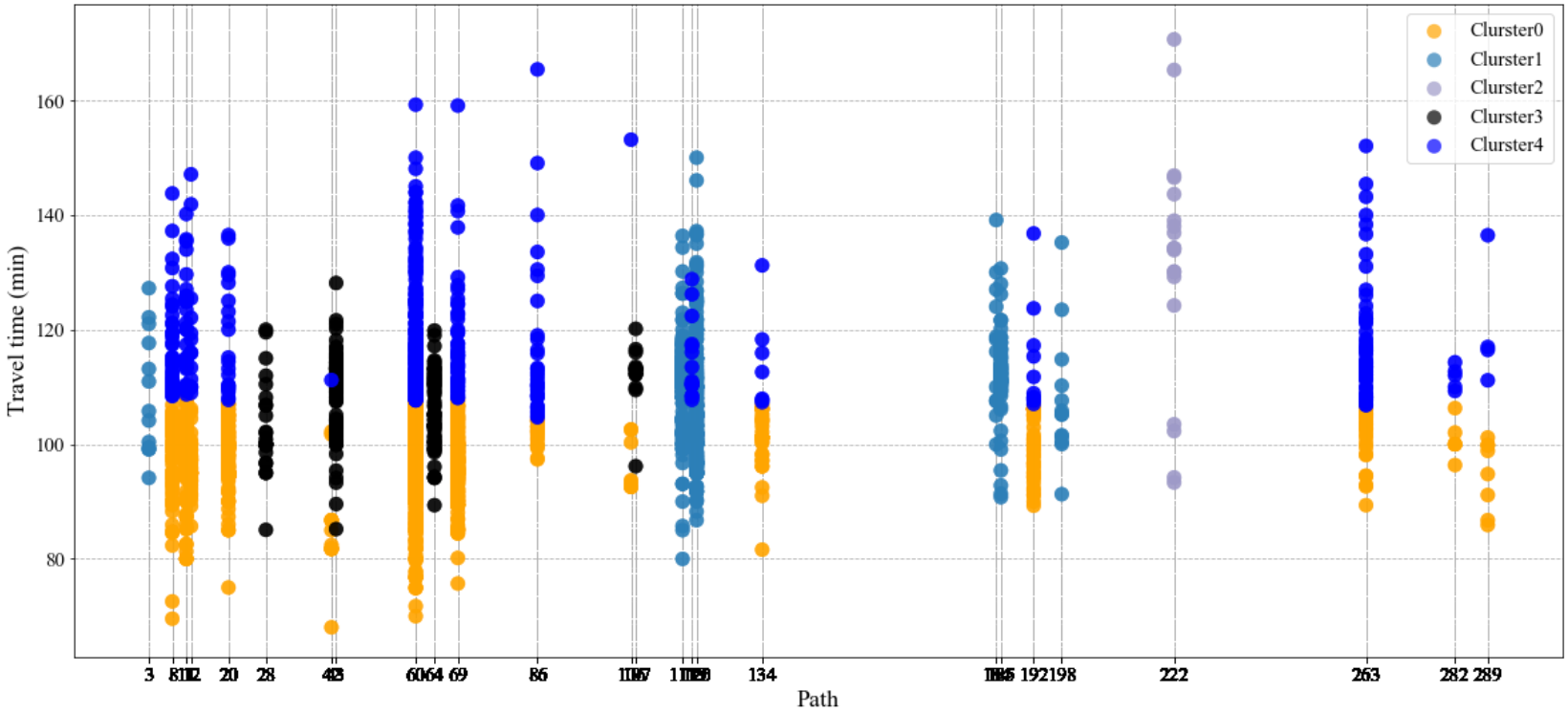


Figure 5.9 Clustering results using travel time, AADT, truck AADT, truck speed, and CF, by path (Scenario 1).

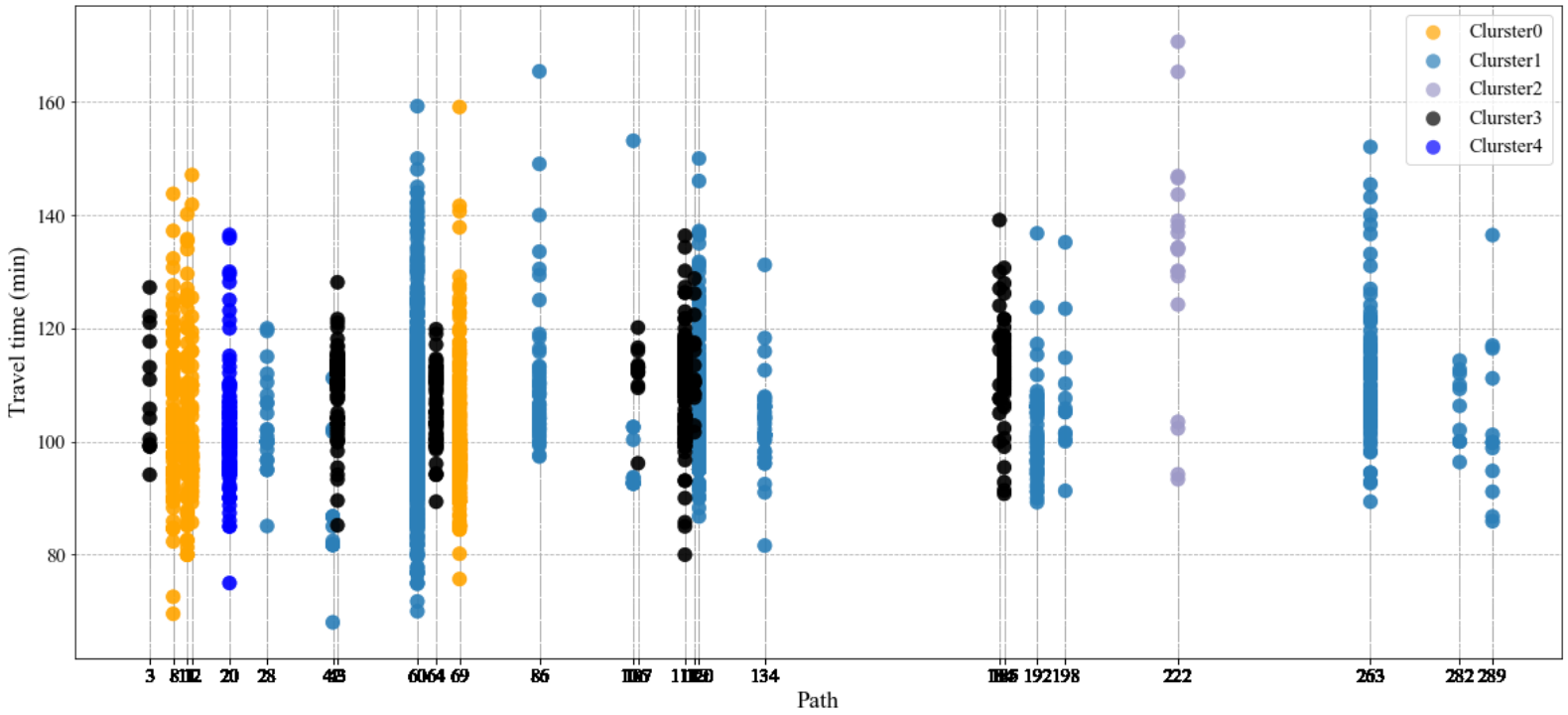


Figure 5.10 Clustering results using CF only, by path (Scenario 1).

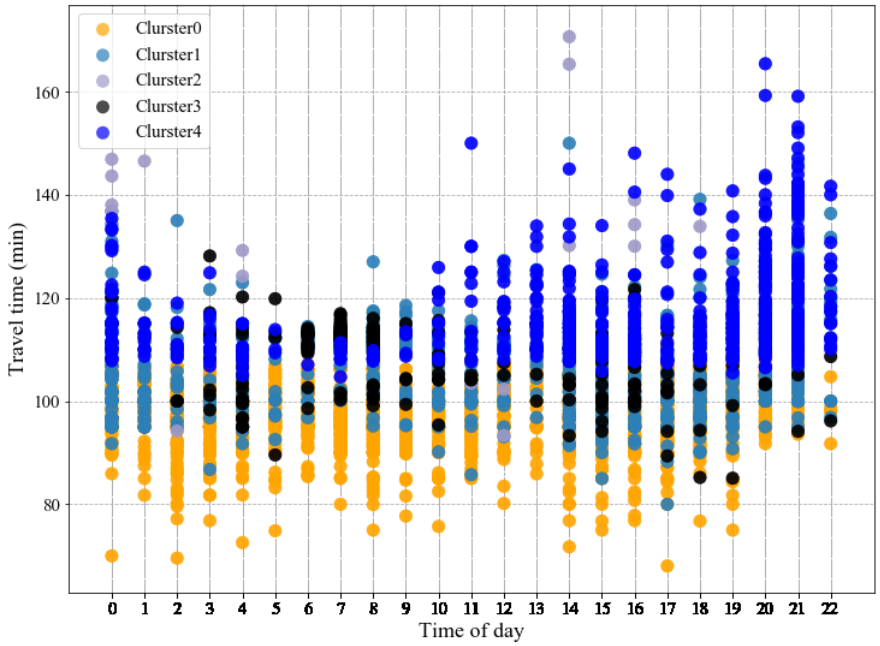


Figure 5.11 Clustering results using travel time, AADT, truck AADT, truck speed, and CF, by the time of day (Scenario 1).

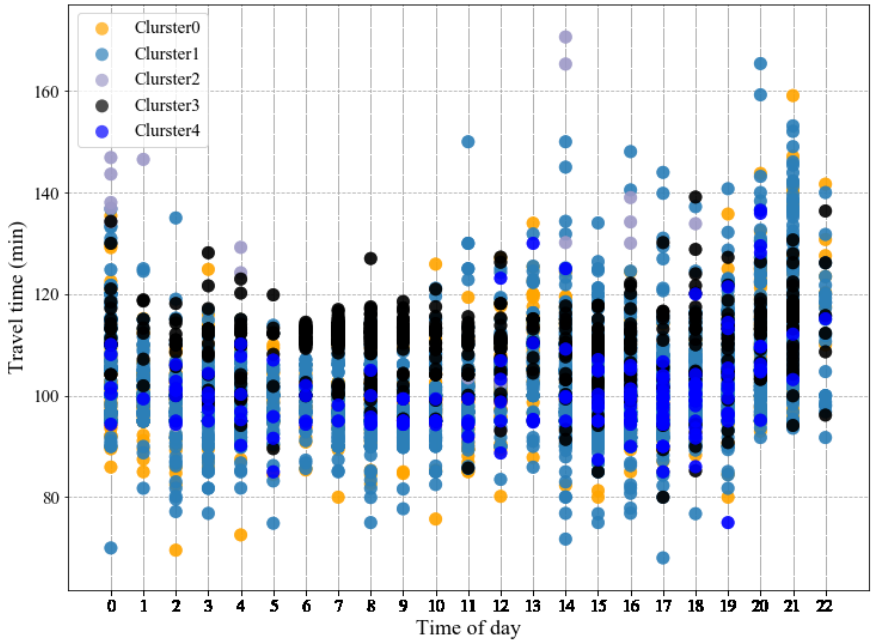


Figure 5.12 Clustering results using CF only, by the time of day (Scenario 1).

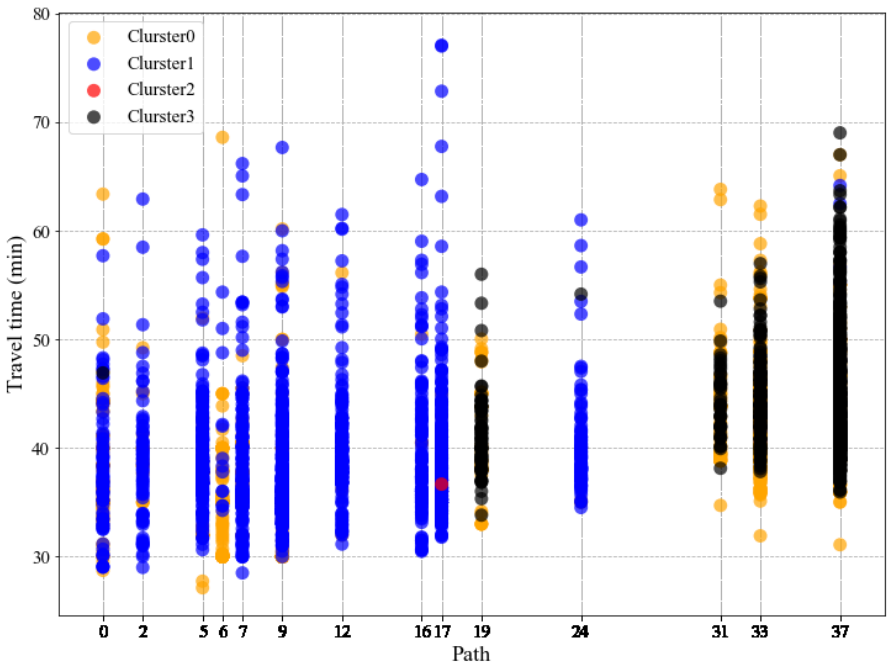


Figure 5.13 Clustering results using travel time, AADT, truck AADT, truck speed, and CF, by path (Scenario 2).

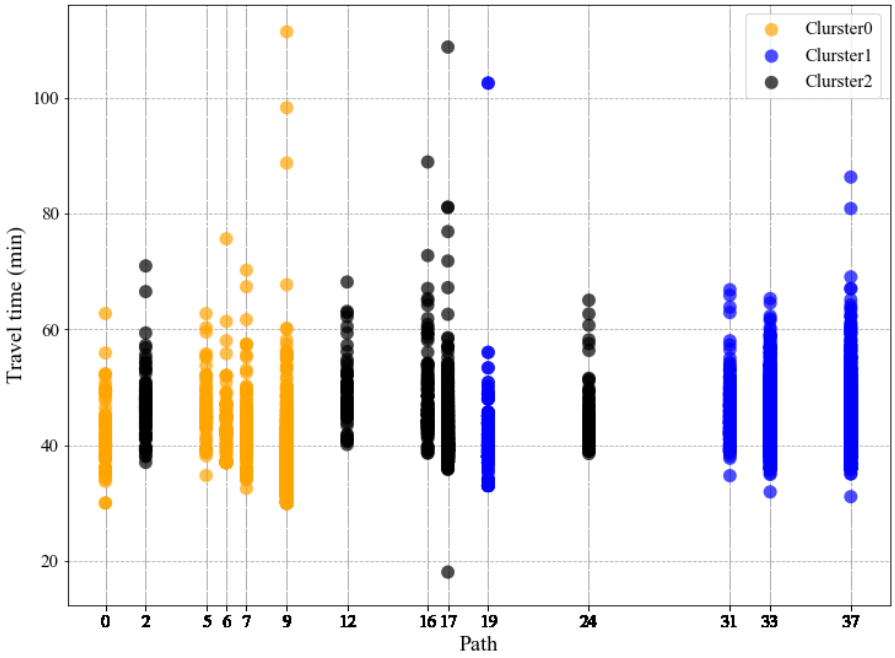


Figure 5.14 Clustering results using CF only, by path (Scenario 2).

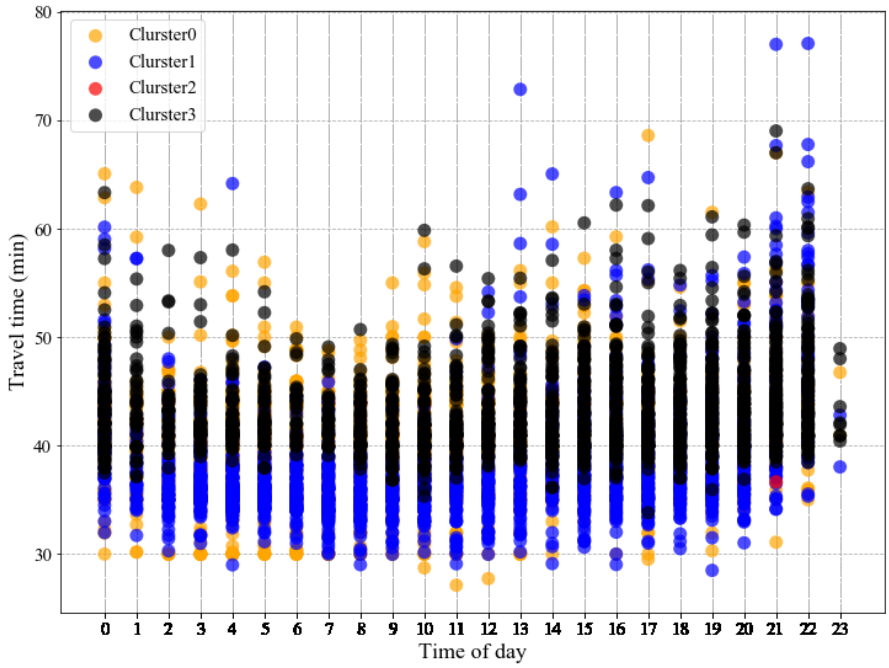


Figure 5.15 Clustering results using travel time, AADT, truck AADT, truck speed, and CF, by the time of day (Scenario 2).

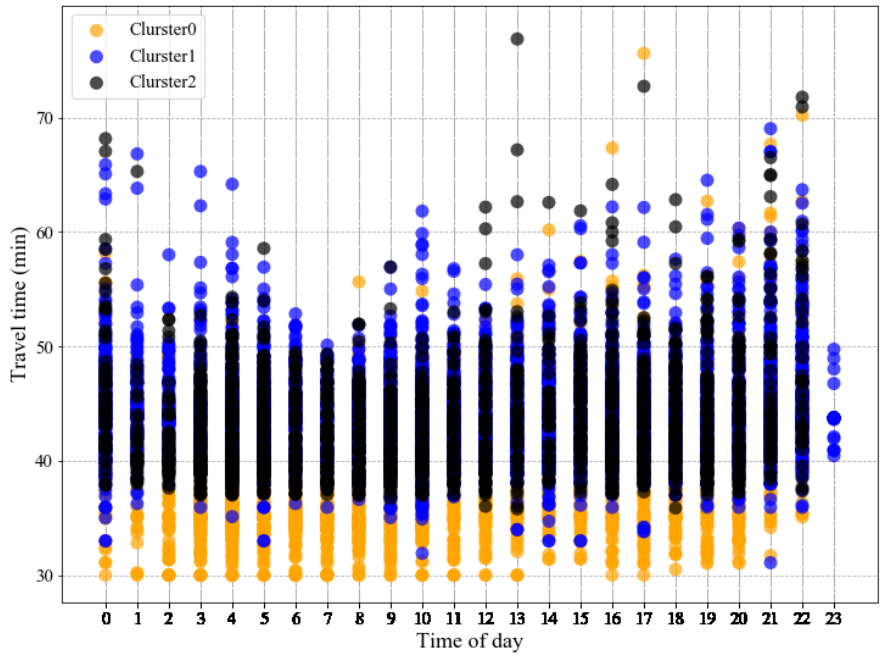


Figure 5.16 Clustering results using CF only, by the time of day (Scenario 2).

The classic *k*-mean cluster is used for both scenarios. The value of k is determined using the "elbow" method, which determines the optimal *k* value using the within-cluster sum of squared distance (SSD). Analytically, the sum of the squared distance is

$$SSD = \sum_{k=1}^{K} \sum_{i \in c_k} \sum_{j=1}^{P} (x_{ij} - \overline{x_{kj}})^2 \quad (5.1)$$

where x_{ij} is observation i at the j^{th} variable, c_k is the set of observations in the k^{th} cluster, and $\overline{x_{kj}}$ is the k^{th} cluster center at the j^{th} variable. Figure 5.17 (A) displays the sum of squared distance for Scenario 1. It can be seen from this figure that when k = 3, the slope of the SSD curve begins to reduce, and when k = 5, the slope remains stable. If k = 3 is used, one of the clusters would have 20 routes. Considering that there are only several dominant routes, if the number of routes is too high in one cluster, the probability of many routes may close to 0, which affects the model's accuracy.

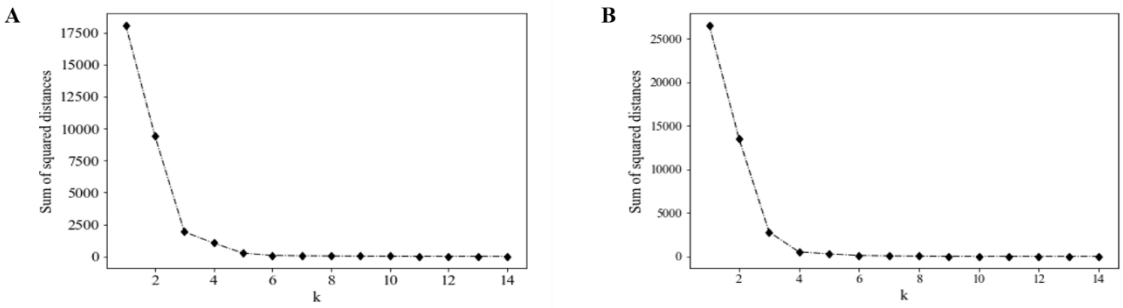


Figure 5.17 Elbow plots of both scenarios.
(A) Elbow plot for Scenario 1. (B) Elbow plot for Scenario 2.

We use k = 5 for the first scenario, where Cluster 0 contains 12 routes, Cluster 1 contains 8 routes, Cluster 2 contains 3 routes, and Cluster 4 and Cluster 5 contain 1 route.

For Scenario 2, as shown in Figure 5.9 (B), k = 3 is the optimal number of clusters. The number of routes that are in Cluster 0, Cluster 1, and Cluster 2 is 5, 5, and 4, respectively. The adjusted travel time for later analysis is shown in Figure 5.18 (A) and (B).

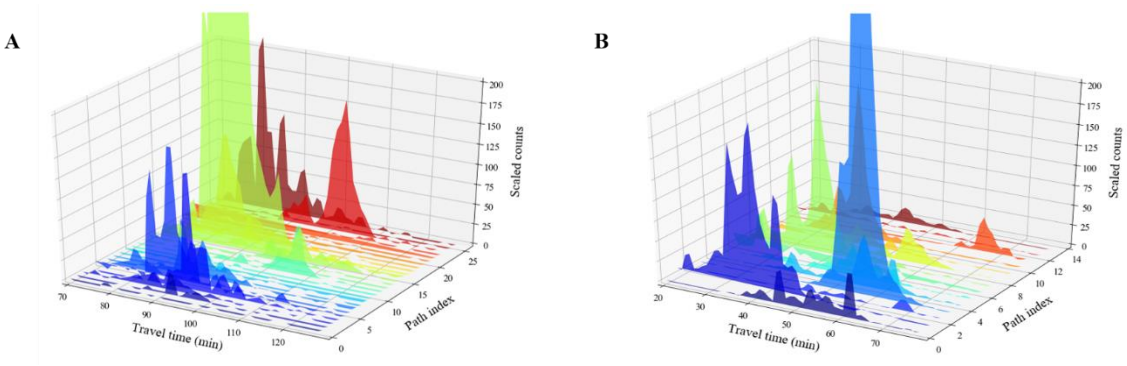


Figure 5.18 Histogram of travel time by path.
(A) Travel time for Scenario 1, by path. (B) Travel time for Scenario 2, by path.

6. RESULTS

This section documents the results of modeling with the GPS dataset cleaned in Section 5. Applied to the GPS dataset, the results of model precision compared to other machine learning methods, the estimated parameters, the VTTS as well as the predicted probabilities are documented in this section. In Section 6.1, the data used for this dissertation is summarized by clusters. The cross-validation results of the proposed models as well as other machine learning techniques are compared. In Section 6.2, the detailed parameter estimation results are displayed, and the preference heterogeneity reflected by the proposed models is analyzed. Section 6.3 describes the estimated VTTS. The last section, Section 6.4, shows the predicted percentages of road use.

6.1. Cross-validation results

After the map-matching process, routes that are used in subnetwork 1 are clustered into five clusters. The number of trip counts in each cluster and the corresponding probability that each cluster takes is shown in Table 6.1.

The total number of trips collected for Scenario 1 and 2 is 9,040 and 13,304, respectively. With a mixed experimental setting, the total number of drivers in each cluster is displayed in the third column of Table 6.1. The number of drivers in each scenario is approximately 160 and 240, respectively; each driver in both scenarios have a fixed choice set size of 50. Drivers who have less than 50 choices observed are removed in the data cleaning process.

Table 6.1 Data summary.

Scenario	Cluster	*N*	Probability	Travel time			
				mean	variance	min	max
Scenario 1	Cluster 0	110	0.6115	101.6770	97.1350	68.0530	165.4000
	Cluster 1	24	0.1398	110.5362	46.5266	80.0000	139.1170
	Cluster 2	37	0.2100	101.0561	88.5803	69.5650	159.0930
	Cluster 3		0.0042	131.3765			
	Cluster 4		0.0345	100.5309			
Scenario 2	Cluster 0	26	0.1757	43.1784	26.3379	18.0667	108.7000
	Cluster 1	76	0.2865	38.4301	33.9807	30.0000	111.3330
	Cluster 2	143	0.5377	44.3662	29.7483	31.0991	102.4790

Table 6.1 Continued.

Scenario	Cluster	Delay per lane		Toll	S/IS	CF	
		min	max	count	count	min	max
Scenario 1	Cluster 0	1,020.4003	1,532.7228	6	1	-4.9664	-4.8946
	Cluster 1	1,086.2630	1,539.5200	8	0	-4.8921	-4.8342
	Cluster 2	1,086.1660	1,087.5340	0	2	-4.8927	-4.8848
	Cluster 3	310.7956		0	0	-3.2266	
	Cluster 4	1,084.1845		0	0	-4.8631	
Scenario 2	Cluster 0	292.9980	450.4120	0	1	-3.0585	-2.8741
	Cluster 1	971.8600	1,035.1300	0	2	-2.2361	-2.0609
	Cluster 2	75.1582	90.7710	0	1	-2.8274	-2.8762

Table 6.2 Toll price of each link.

Nodes		Toll rates (2-axle)	Toll rates (3-axle)	Station begin	Station end
From	To	Sam Rayburn Tollway			
28	30	0.47	0.94	Lake Vista	Old Denton
30	31	1.62	3.24	Old Denton	Parkwood
31	32	2.24	4.48	Parkwood	US 75
From	To	President George Bush Turnpike			
42	45	0.6	1.2	Belt Line W**	SH 114/Royal
45	46	0.31	0.62	SH 114/Royal	IH 635/Valley View/Las Colinas
46	47	1.04	2.08	IH 635/Valley View/Las Colinas	IH 35E/Sandy Lake/Old Denton
27	26	1.31	2.62	IH 35E/Sandy Lake/Old Denton	DNT
26	25	1.41	2.82	DNT	US 75/Alma/Plano
From	To	Dallas North Tollway			
9	8	1.65	3.3	IH 35E/Oaklawn/Wycliff	Walnut Hill/Royal
8	23	0.43	0.86	Walnut Hill/Royal	IH 635
23	26	1.18	2.36	IH 635	PGBT
26	31	1.05	2.1	PGBT	SRT

With the aim of finding out the VTTS on commercial truckers' route choices, two O-D pairs corresponding to two scenarios are included in this study, where five out of 25 routes in Scenario 1 are associated with different toll prices according to the distance traveled on the tollway, while no tollway is available in Scenario 2.

Table 6.3 Toll price of each route.

Cluster 0		Cluster 1	
Path	Toll price	Path	Toll price
28	4.46	3	8.66
42	0.86	43	6.54
60	0	64	9.34
86	0	107	2.1
106	3.3	117	10.48
120	8.66	119	3.9
134	0	184	7.72
192	3.3	185	10.48
198	8.66		
263	0		
282	0		
289	3.3		

Among all the alternatives in Scenario 1, Cluster 2 has the least mean travel time, while Cluster 1 has the longest travel time with the smallest variance. Routes in Cluster 0 dominate the trips from Node 1 (the origin) to Node 34 (the destination), in which both maximum and minimal travel times have fallen. In Scenario 2, routes in Cluster 2 cover more than half of the trips while having the longest mean travel time. Cluster 1 has the shortest mean travel time with the largest variance. Instantaneous travel time in Cluster 0 is the most stable, with the smallest variance. The delay per lane tells a different story than does the travel time. In Scenario 1, the per-lane truck delay varies widely from approximately 1,000 to 1,500 minutes for Cluster 0 and 1. However, routes in Cluster 2 all have similar delays. In Scenario 2, a similar truck delay was observed in routes for each cluster.

The probability column in Table 6.1 shows the percentage of each cluster that takes in the total number of trips of each scenario. As each link may be used repeatedly in a multi-O-D network setting, the predicted probability of each link is thus the aggregated route probabilities in which this link is included.

Consequently, for each scenario, the final probability of route usage will then be calculated with a nested structure

$$P\left(j_{c_i}\right) = P\left(j_{c_i}\middle|c_i\right) * P(c_i) \;\; \forall j_{c_i} \in c_i, i \in \{0,1,2 \dots\} \tag{6.1}$$

where the probability that a route has been used $P\left(j_{c_i}\right)$ equals the product of the probability of the route being used in its cluster $P\left(j_{c_i}\middle|c_i\right)$ and the probability of the cluster $P(c_i)$. For instance, the probability that Route 60, which is in Scenario 1, Cluster 0, has been used is 0.8960 * 0.6115 = 54.79%. The true and estimated probabilities of

each route are listed in Appendix 1. Meanwhile, in Scenario 1, since there is only one route in Cluster 3 and 4, the probability that the only alternative in the two clusters is 0.42% of Route 222 and 3.45% of Route 20, respectively.

The cross-validation process provides a large picture of the model prediction's accuracy and stability among all clusters in our two study scenarios. In addition to our proposed mixed C-logit model with three Bayesian formulations, namely BMCL, HBMCL1, and HBMCL2, nine additional models are included in the 10-fold cross-validation model evaluation process. The error displayed in Figure 6.1 is the maximum error, which is the maximum discrepancy between the predicted probability and the real probability among potential alternatives in each cluster. The constructions of the additional models are derived from the Python library scikit-learn 0.23.2, all with the default setting.

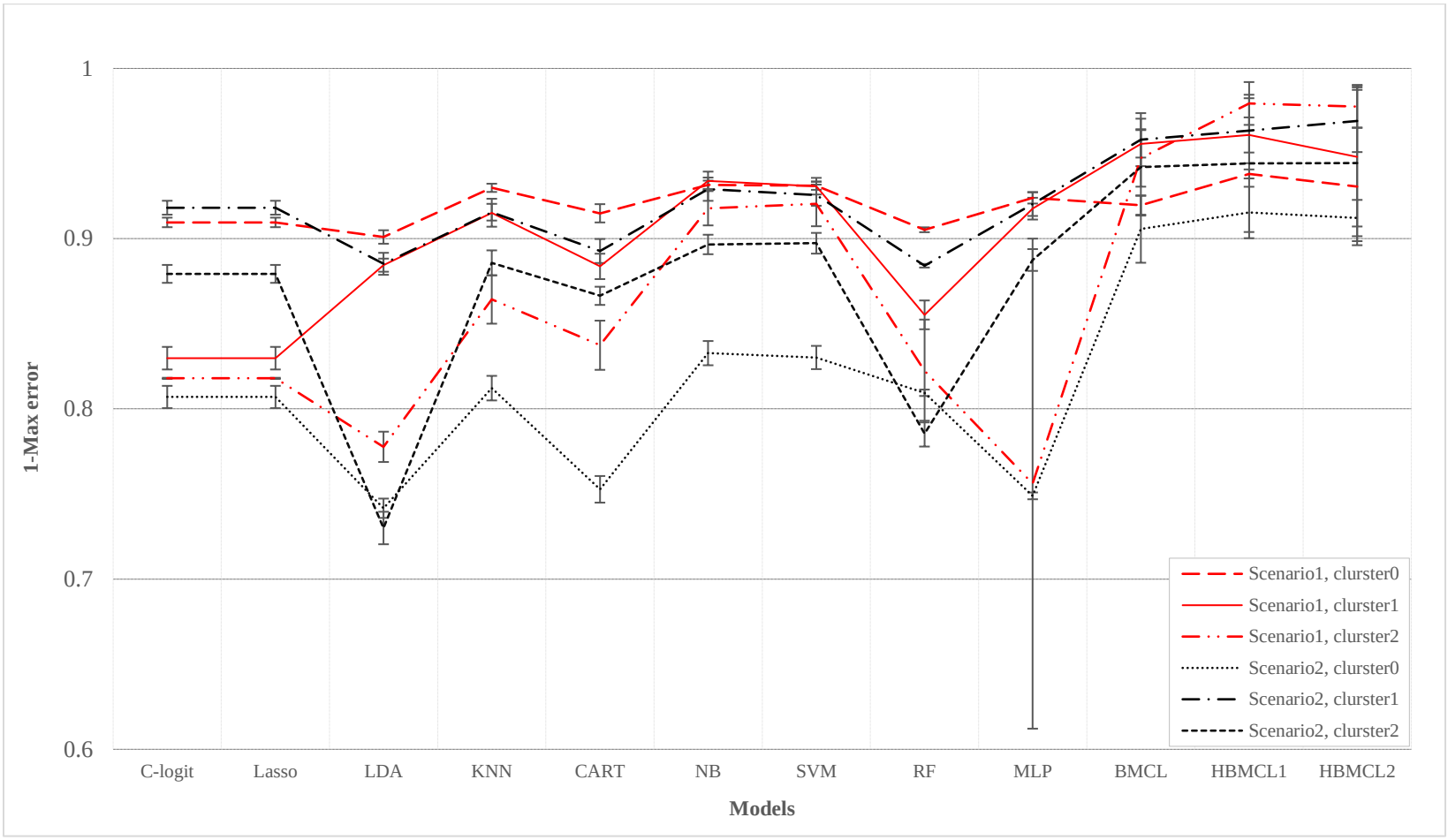


Figure 6.1 Cross-validation error.

Overall, the outperformance of BMCL, HBMCL1, and HBMCL2 over machine learning procedures is obvious, judging from the high prediction accuracy of these three models. All other models perform badly in Scenario 2, Cluster 0, with accuracy varying between 0.68 and 0.78. The linear discriminant analysis (LDA) reported the lowest accuracy of 0.69, with little variance indicated by the error bar in the figure. This was followed by the multi-layer perceptron classifier (MLP), which had an average accuracy of 0.70, yet a huge variance was observed for this classifier. In all clusters of Scenario 2 and Cluster 2 of Scenario 1, the accuracy of the LDA and of the random forest classifier (RF) are notably inferior to other classifiers and the logit models. Except for the LDA, RF, and MLP that differ significantly in the performance of different clusters, other models are relatively stable in model performance over six clusters. It is observed that the C-logit model performs equivalently to the Lasso classifier in all clusters. Both have a higher accuracy than the classification and regression tree (CART) classifier, and both perform slightly worse than the *k*-nearest neighbors (KNN) classifier. Among all the machine learning methods listed in the figure above, the support vector machine (SVM) and the naïve Bayes (NB) classifier are the best classifiers for truck route choice in both performance stability and prediction accuracy.

Comparing the three Bayesian formulations, HBMCL1 outperforms BMCL and HBMCL2, which have similar performance in four out of six cases. The only situation where BMCL notably outperform HBMCL1 and HBMCL2 is Scenario 2, Cluster 2, yet the latter is still remarkably superior to other classifiers and the C-logit model.

Moreover, the proposed model shows a more significant variance in performance compared to the path size logit and machine learning classifiers. In Scenario 2, Cluster 2, all three Bayesian models have large variances in the accuracy reported from the 10 cross-validation tests. In other clusters, however, HBMCL2 is distinct with its high variance.

6.2. Estimated parameters and the preference heterogeneity

After the evaluation of the model performance using the maximum prediction error with a 10-fold cross-validation process, this section analyzes the predicted parameters from the proposed BMCL, HBMCL1, and HBMCL2. We run the proposed model on the whole dataset with an 80%-20% split for each cluster's training set and test set. The parameter estimation and probability prediction are all based on this final test (Section 6.2, 6.3, and 6.4). Results for each cluster are shown in Table 6.5 to 6.10 in this section.

6.2.1. Estimated parameters for Scenario 1

In Scenario 1, it can be seen, from the magnitude of parameters, that the travel time is the first consideration in a trucker's route choice no matter the existence of a toll alternative. Among all clusters in Scenario 1, the parameter of travel time has a mean over 2.5 in the absolute value, while the parameters of the S/IS indicator are smaller than 2, except in Cluster 0, where the parameter of the S/IS road indicator has a value of 2.7 from BMCL. Compared to the S/IS road indicator, the delay per lane attribute is slightly more significant and generally has the largest variance. The range between differences in the absolute value of parameters from each model in Cluster 0 and 2 is about 1.4 and

0.64, respectively. Among all clusters in Scenario 1, the CF is the least significant one, whose value varies from -0.11 to -0.63, with the smallest variance.

Table 6.4 Estimated parameter, Scenario 1, Cluster 0.

Model	BMCL		HBMCL1		HBMCL2	
Variable	Mean	Variance	Mean	Variance	Mean	Variance
Travel time (min)	-5.4734	5.4131	-3.9002	1.6961	-4.1575	1.3298
Delay per lane	-1.5160	4.5838	-1.0132	1.6768	-1.5727	8.4855
Toll price	-2.8555	4.7323	-2.1415	1.8007	-2.0669	1.6971
S/IS indicator	-0.2856	2.1916	-0.6403	1.1676	-0.3961	1.1044
CF	0.1208	1.5279	0.1206	1.8301	0.1161	1.8259
Hyper-parameter	HBMCL1	μ_ξ	[-3.89, -1.01, -2.14, -6.39e-01]			
		Σ_ξ	[2.08e-03, 2.21e-03, 2.11e-03, 2.13e-03]			
		$diag(\Upsilon)$	[9.43e-02, 8.85e-02, 9.28e-02, 9.18e-02]			
	HBMCL2	μ_ξ	[-4.34, -1.81, -2.35, 0.435]			
		Σ_ξ	[8.36e-02, -2.78e-01, 2.62e-01, 1.27e-01]			
		$diag(\Upsilon)$	[1.24e+00, 1.035e+00, 8.99e-01, 5.12e-01]			
		g	[1.59e+03, 1.32e+03, 1.14e+03, 6.54e+02]			

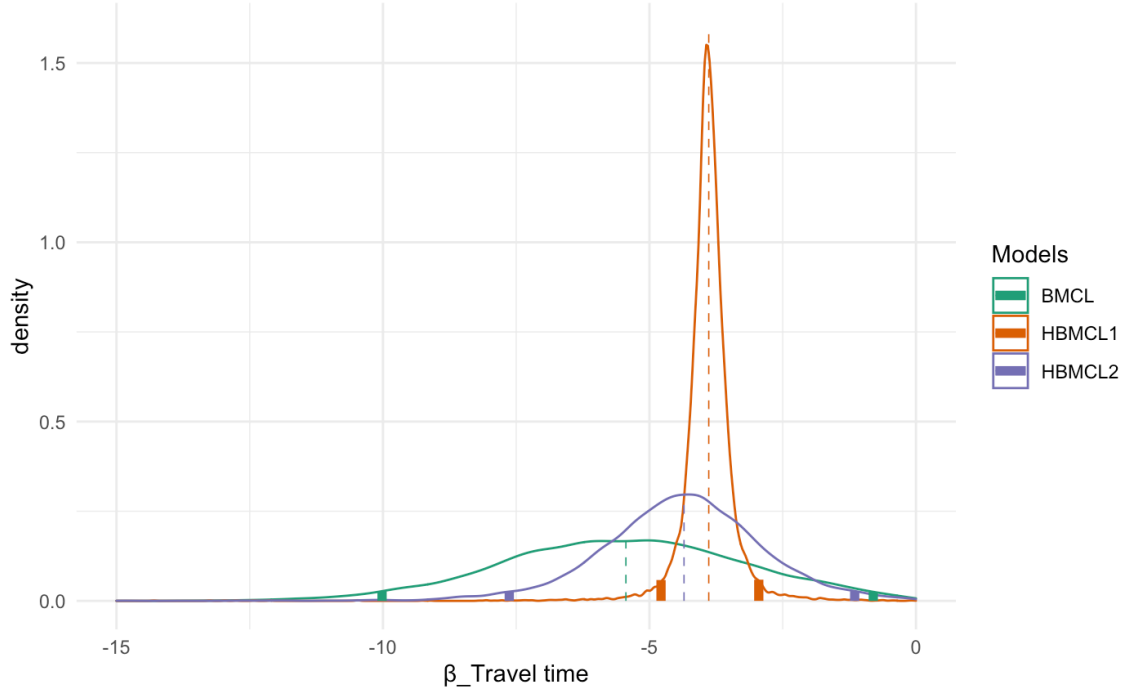


Figure 6.2 Kernel density of $\beta_{Travel\ Time}$ for Scenario 1, Cluster 0.

Figure 6.2 shows the kernel density of the coefficient of travel time, $\beta_{Travel\ Time}$ for Scenario 1, Cluster 0. Overall, the BMCL and the HBMCL2 provide more disperse density of estimated taste parameters than does HBMCL1. For the coefficient of travel time, the 95% confidence interval (CI) from BMCL, HBMCL1, and HBMCL2 is (-0.80, 10.00), (-2.95, -4.78), and (-1.15, -7.63), respectively. In some studies, the $\beta_{Travel\ Time}$ is modeled with a log-normal prior distribution for restricting a negative sign of the estimated value. However, this asymmetric prior suffers from low precision in

estimating the mean and variance parameters of interest (Hensher and Greene 2003). The normal prior that has been used for this study well addressed this sign consideration. It can be seen that the bulk of the density of $\beta_{Travel\ Time}$ from three models is negative, with all bounds of confidence intervals being negative.

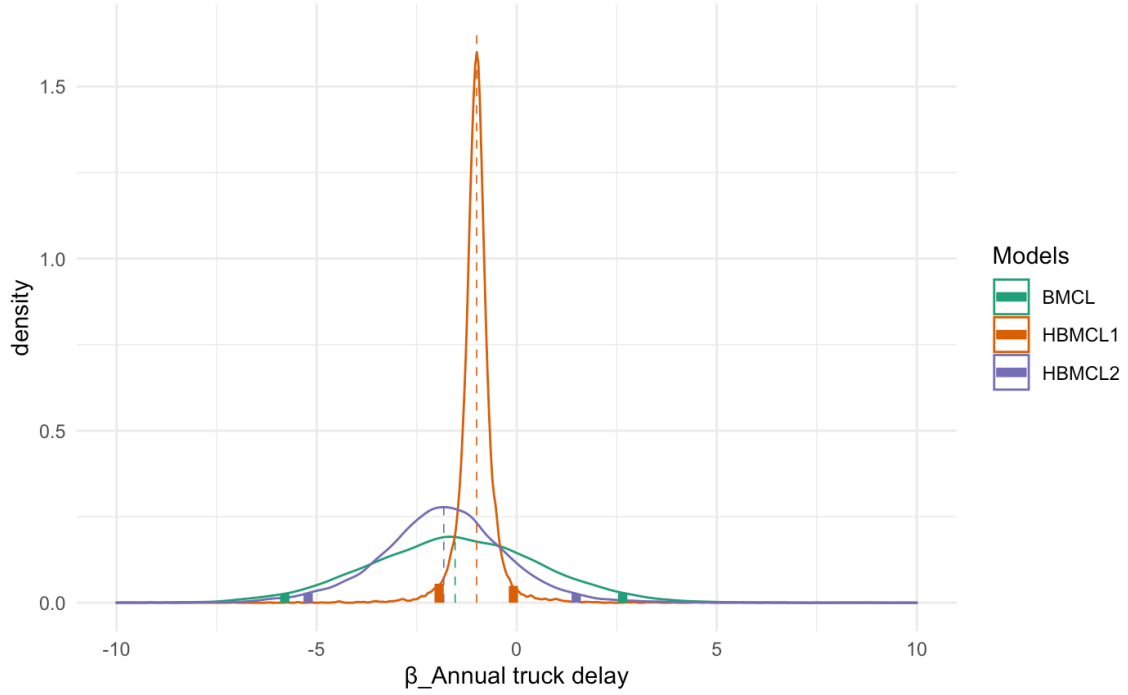


Figure 6.3 Kernel density of $\beta_{Anual\ truck\ delay}$ for Scenario 1, Cluster 0.

The kernel density of $\beta_{Anual\ truck\ delay}$ for Scenario 1, Cluster 0 is displayed in Figure 6.3. HBMCL1 provides the most concentrated density of $\beta_{Anual\ truck\ delay}$ compared to the HBMCL2 and the BMCL, with density from the former relatively more

concentrated than that from the latter. The confidence interval of the three models is (-2.65, 5.80) in BMCL, (-1.94, -0.082) in HBMCL1, and (-5.21, 1.49) in HBMCL2. Both the confidence intervals from HBMCL2 and the BMCL1 included 0, whereas the confidence interval from HBMCL1 has both bounds less than 0.

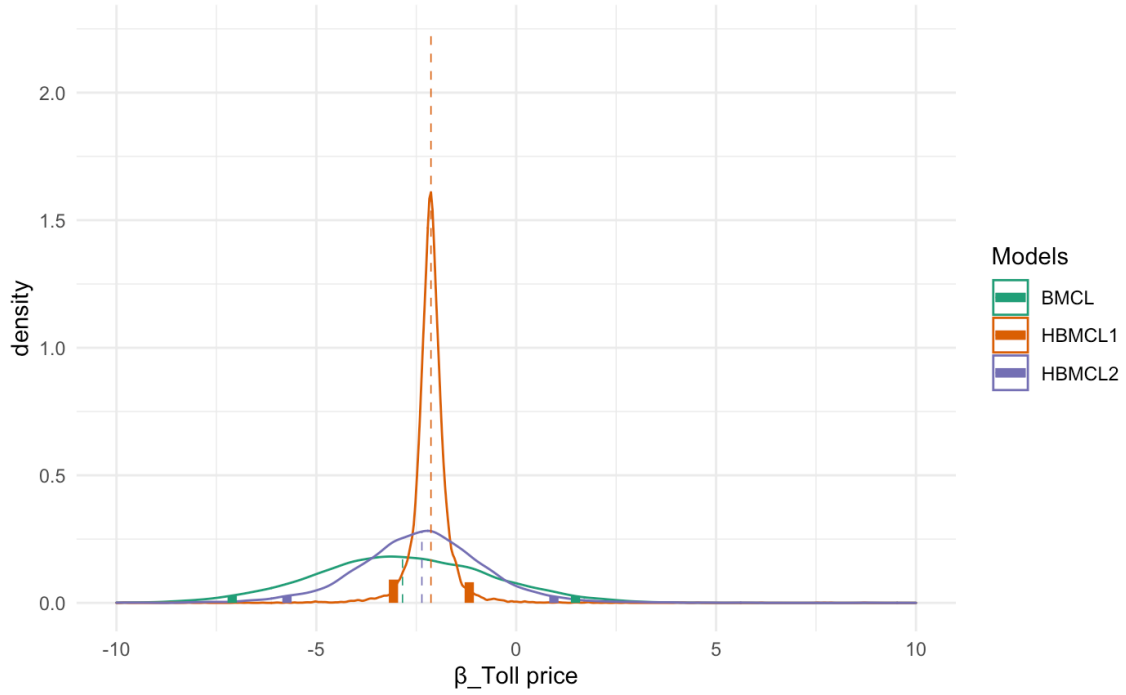


Figure 6.4 Kernel density of $\beta_{Toll\ price}$ for Scenario 1, Cluster 0.

Similar to the $\beta_{Travel\ Time}$ and $\beta_{Annual\ truck\ delay}$, Figure 6.4 shows that the HBMCL1 provides the least dispersed distribution for the coefficient of toll price compared to the other two models. The 95% confidence interval of HBMCL1 is (-1.17, -

3.07), which is negative. Both the 95% confidence interval of BMCL and of HBMCL1 contain both a negative value and positive value, (-7.11, 1.48) and (-5.73, 0.943).

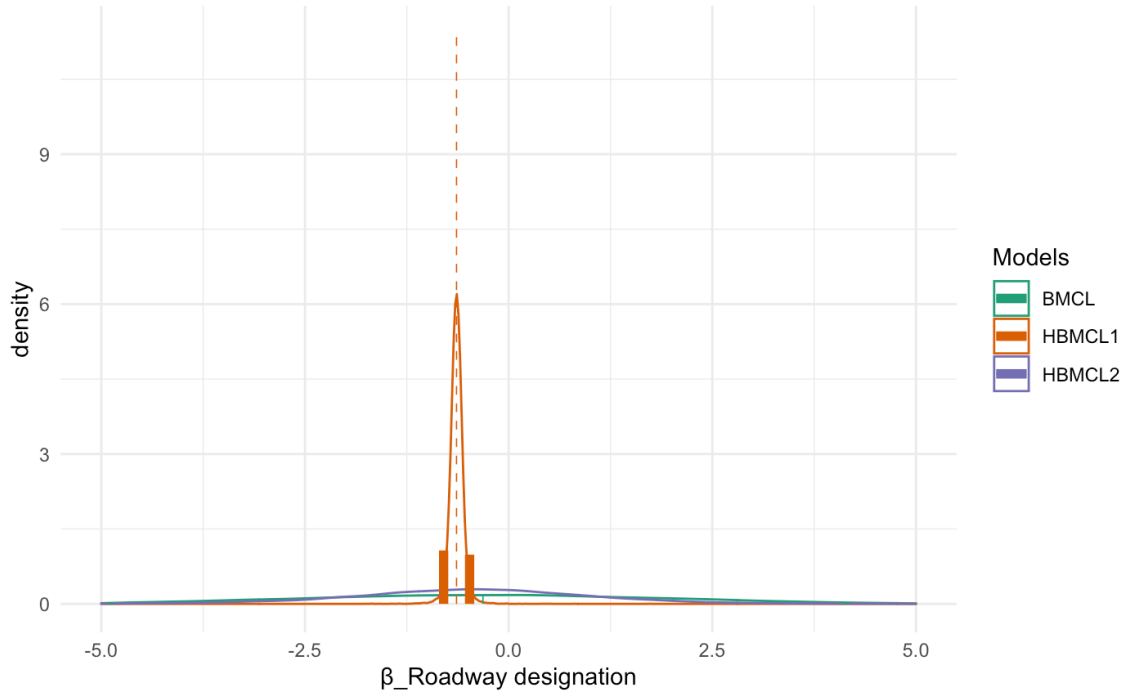


Figure 6.5 Kernel density of $\beta_{Roadway\ designation}$ for Scenario 1, Cluster 0.

The kernel density of the coefficient of roadway designation shows the greatest difference among models, more than other coefficients. According to Figure 6.5, the 95% confidence interval of each model is (-4.88, 4.14), (-0.798, -0.480), and (-3.75, 2.73), respectively. The confidence interval of HBMCL1 is significantly smaller than that of BMCL and HBMCL2.

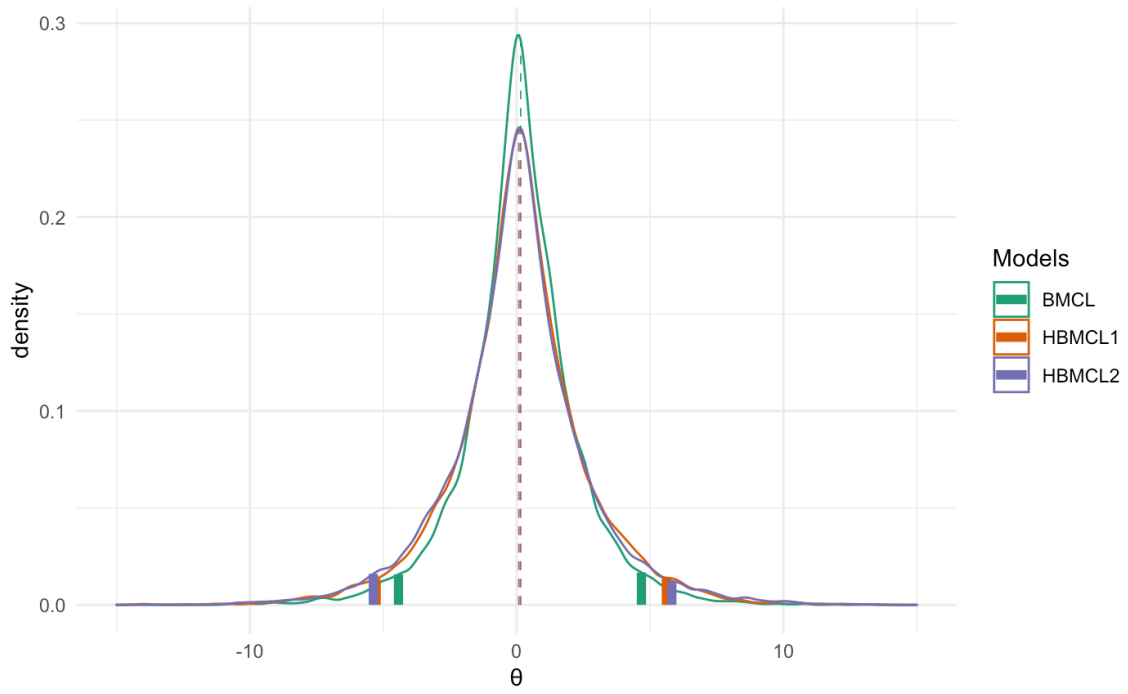


Figure 6.6 Kernel density of θ for Scenario 1, Cluster 0.

Since each model's CF is estimated using the Laplace distributions for the prior, the resulting kernel density for the coefficient of CF is different from other coefficients. Generally, all the three models have provided a similar density for θ, with BMCL giving a slightly more concentrated one than do the other two. The means of the densities are all close to 0, and variances are similar. In particular, the 95% confidence interval of BMCL, HBMCL1, and HBMCL2 is (-4.43, 4.68), (-5.27,5.61), and (-5.37, 5.82), respectively.

To reveal the relationship between the predicted parameters, the pairwise scatter plot of the coefficient of each attribute is presented, based on a Gibbs sampler with 10,000 draws.

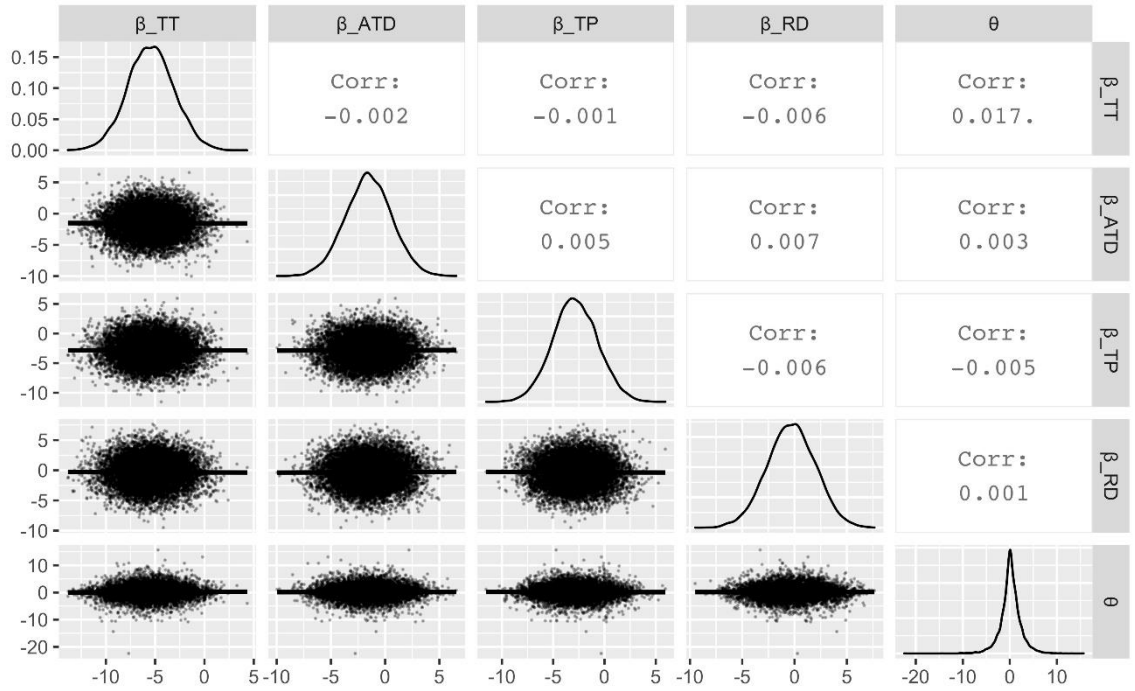


Figure 6.7 Pairwise scatter plot of parameters for BMCL.

Figure 6.7 displays the scatter plot of parameters for BMCL. As indicated by the model structure, each predicted parameter is not correlated with each other, with all the correlations less than 0.02.

In HBMCL1, the means of the taste parameters are assumed to follow a *k*-dimensional distribution, with *k* being the number of attributes considered for the model.

The variances, however, are assumed to follow an inverse Wishart distribution, which resulted in the correlation between the coefficients.

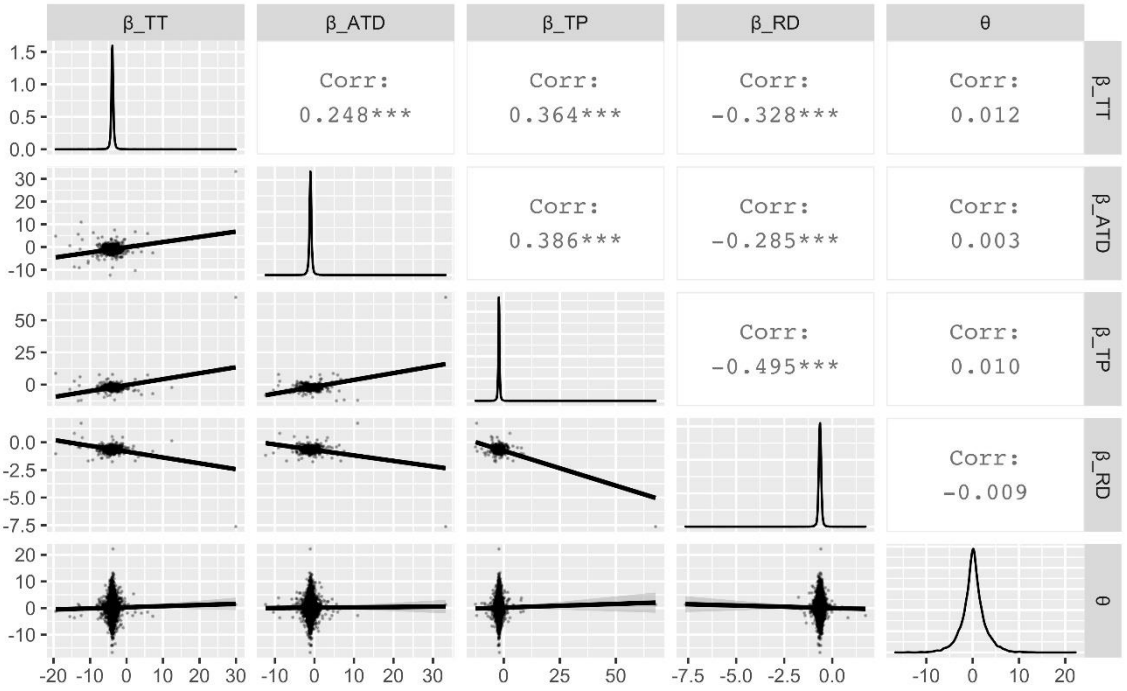


Figure 6.8 Pairwise scatter plot of parameters for HBMCL1.

It can be seen from Figure 6.8 that mild correlations are observed in the first four parameters of the HBMCL1. Since θ is assumed to follow a distinct Laplace distribution, the correlation between θ and each β is negligible.

In addition to assuming that the variance of the taste parameters follows an inverse Wishart distribution in HBMCL1, the HBMCL2 future assumes an inverse gamma prior to this inverse Wishart prior, which makes Huang's half-t prior. In this

case, the parameters in HBMCL2 have a much stronger correlation between each other, which can be as high as 0.961, as indicated in Figure 6.9.

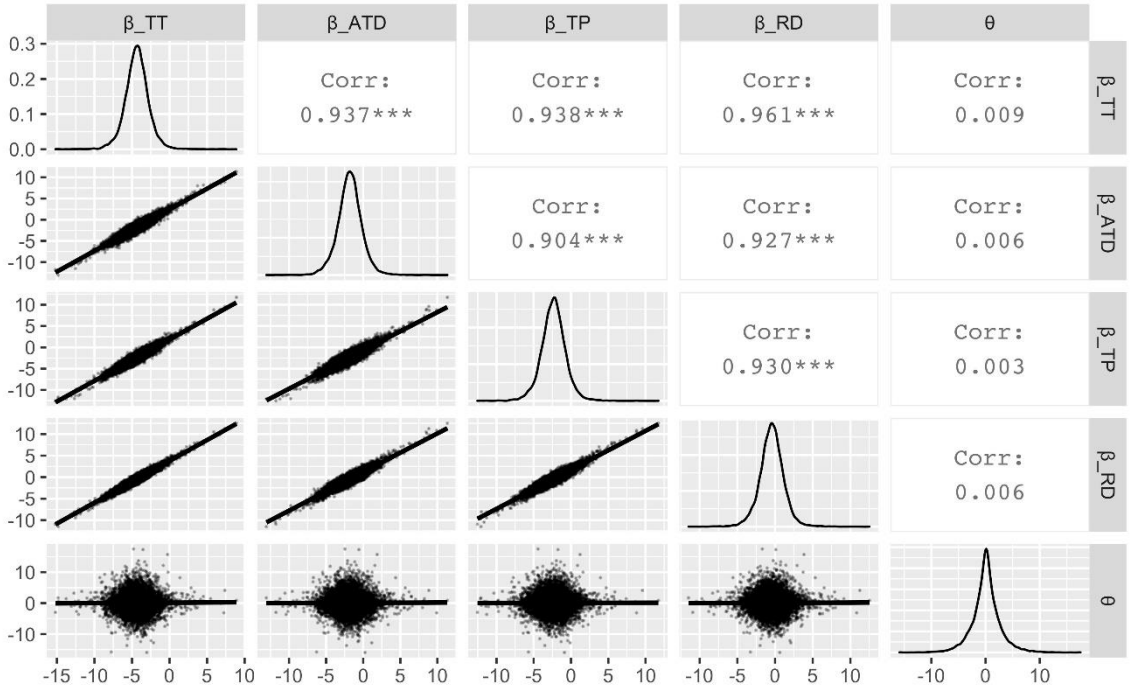


Figure 6.9 Pairwise scatter plot of parameters for HBMCL2.

The estimated parameters of Scenario 1, Cluster 1 are listed in Table 6.5. Since all the routes in this cluster do not contain any local segments, the roadway designation S/IS indicator is not included in the model.

Table 6.5 Estimated parameter, Scenario 1, Cluster 1.

Model	BMCL		HBMCL1		HBMCL2	
Variable	Mean	Variance	Mean	Variance	Mean	Variance
Travel time (min)	-2.8676	8.2319	-2.1529	4.0826	-2.3158	2.8952
Delay per lane	1.5644	4.5985	1.2989	3.0713	1.4520	5.8408
S/IS indicator	-1.3400	6.2264	-0.9850	3.3914	-1.0311	1.4721
CF	0.3700	1.2826	0.3128	0.7840	0.3289	1.0582
Hyper-parameter	HBMCL1	μ_ξ	[2.13, 0.564, -1.09]			
		Σ_ξ	[2.62e+00, 5.71e+00, 3.87e+00]			
		$diag(\Upsilon)$	[7.13e-02, -3.72e-01,8.04e-01]			
	HBMCL2	μ_ξ	[-2.30, 1.45, -1.03]			
		Σ_ξ	[1.57e-01, 1.55e-02, 7.53e-02]			
		$diag(\Upsilon)$	[2.46e-02, 1.07e-02, 6.61e-02]			
		g	[3.16e+01, 1.38e+01, 8.45e+01]			

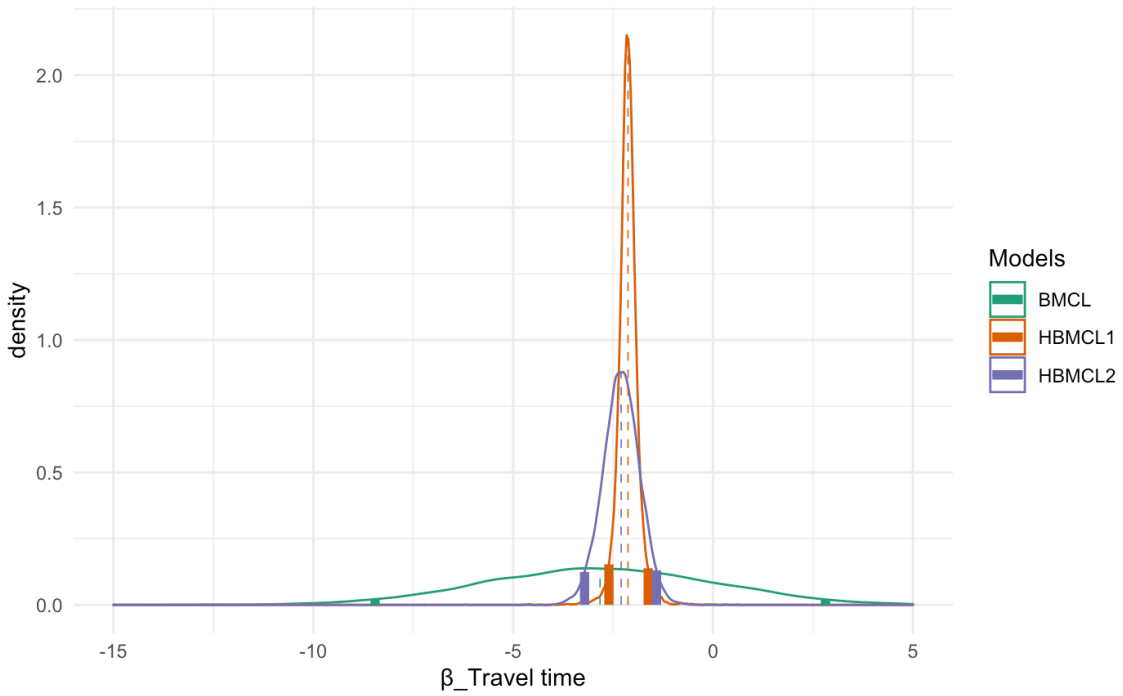


Figure 6.10 Kernel density of $\beta_{Travel\,Time}$ for Scenario 1, Cluster 1.

Figure 6.10 shows the density of $\beta_{Travel\,Time}$ for this cluster. Similar to Cluster 0, the HBMCL1 provides the most concentrated distribution. Both HBMCL1 and HBMCL2 have a negative 95% confidence interval, which is (-2.61, -1.62) and (-3.22, -1.41). However, the confidence interval of BMCL is (-8.46, 2.81), which suggests that the estimation of $\beta_{Travel\,Time}$ is unrealistic since this interval includes part of the positive values.

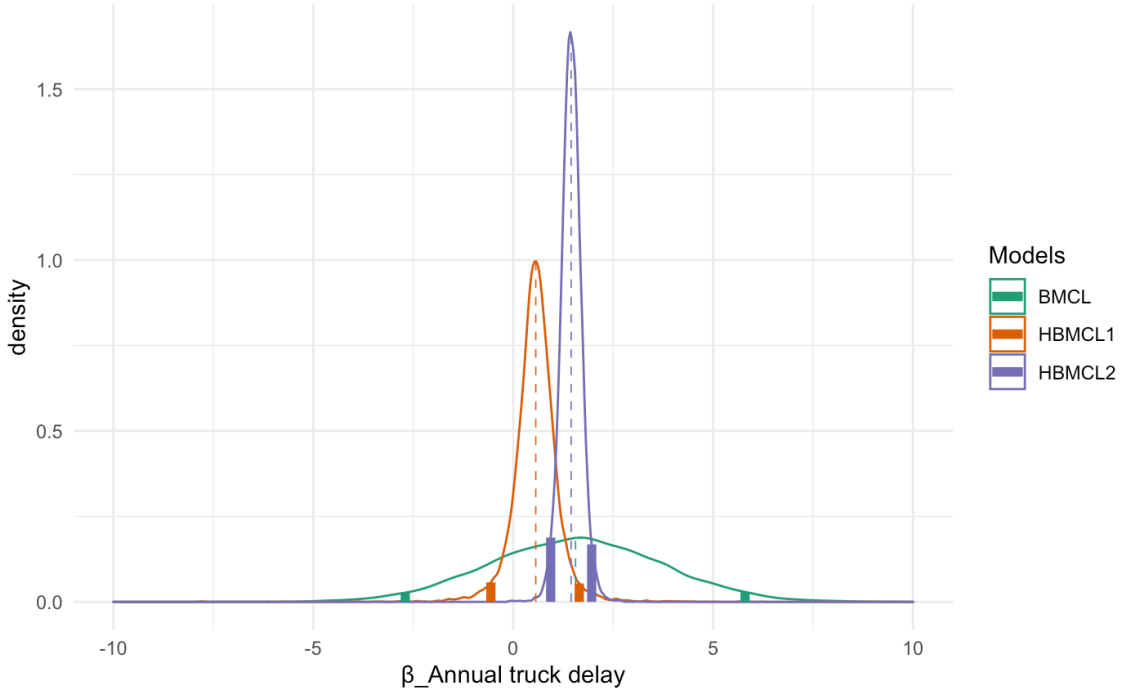


Figure 6.11 Kernel density of $\beta_{Anual\ truck\ delay}$ for Scenario 1, Cluster 1.

As shown by Figure 6.11, the mean of $\beta_{Annual\ truck\ delay}$ estimated by the two hierarchical Bayesian models is different, yet the mean of this coefficient estimated by the BMCL and the HBMCL2 is close. The 95% confidence interval for BMCL, HBMCL1, and HBMCL2 is (-2.70, 5.80), (-0.559,1.65), and (0.938, 1.96). All three CIs have both positive and negative parts, indicating that drivers have fluid opinions about the annual truck delay when making route choices.

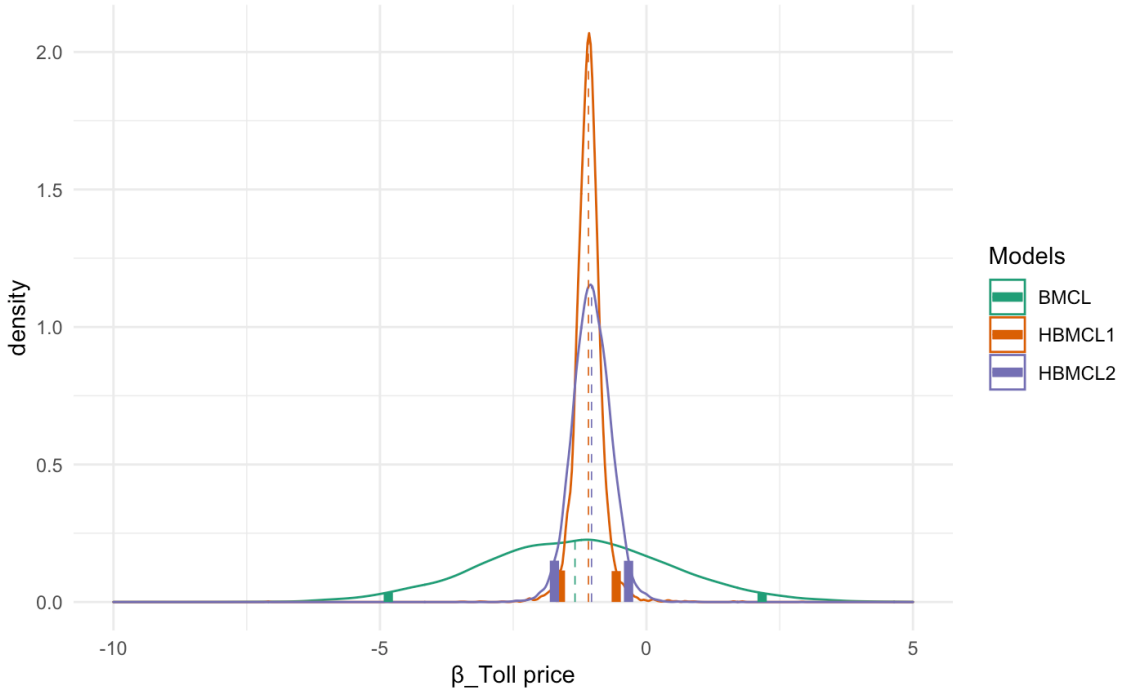


Figure 6.12 Kernel density of $\beta_{Toll\ price}$ for Scenario 1, Cluster 1.

The coefficient of the toll price of Cluster 1 in Scenario 1 displays a similar density of the coefficient of travel time, judging from the similarity of Figure 6.10 and 6.12. The 95% CIs of $\beta_{Toll\ price}$ BMCL, HBMCL1, and HBMCL2 are (-4.82, 2.19), (-1.61, -0.559), and (-1.72 and -0.326). The BMCL is not recommended to use for this scenario since it reported a positive upper bound of the CIs for both $\beta_{Travel\ time}$ and $\beta_{Toll\ price}$.

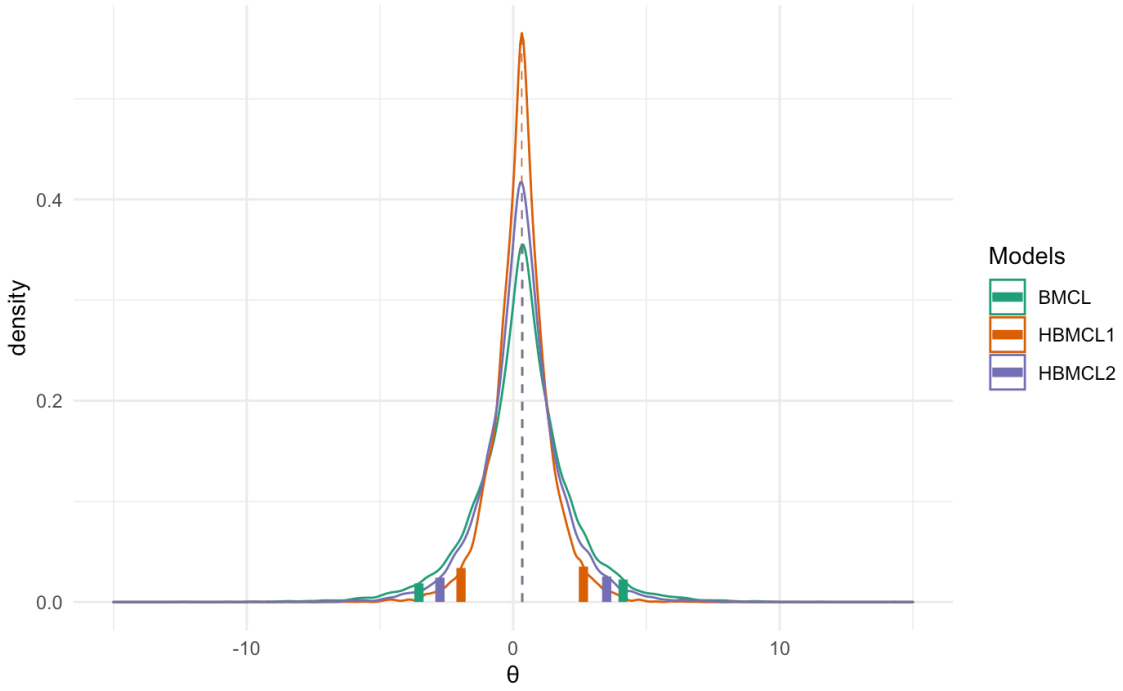


Figure 6.13 Kernel density of θ for Scenario 1, Cluster 1.

The density of the coefficient of the CF in Cluster 1, Scenario 1 is illustrated in Figure 6.13, which shows similarity to the density of θ in Cluster 0 of this scenario. The estimated means of θ from the three models are the same, with the 95% CIs being (-3.51, 4.12), (-2.06, 2.63), and (-2.84, 3.63). The mean of θ from the three models is close to 0, indicating that the network density is not a priority consideration in the truckers' route choice.

The estimated parameters for the third cluster of this scenario are shown in Table 6.6. Overall, compared to Cluster 1, the importance of travel time in affecting truck drivers' route choice is notable.

Table 6.6 Estimated parameter, Scenario 1, Cluster 2.

Model	BMCL		HBMCL1		HBMCL2	
Variable	Mean	Variance	Mean	Variance	Mean	Variance
Travel time (min)	-2.5511	5.6776	-1.6376	1.6786	-1.6840	1.8090
Delay per lane	-0.0638	0.1374	0.0445	0.0317	-0.0588	0.0718
S/IS indicator	-0.0597	0.1723	-0.0681	0.0214	-0.0528	0.0342
CF	0.6383	1.8776	0.5268	2.5678	0.5383	2.5511
Hyper-parameter	HBMCL1	μ_ξ	[-1.64 4.52e-02 6.76e-01]			
		Σ_ξ	[1.11e-03 7.10e-03 7.73e-04]			
		$diag(\Upsilon)$	[2.00e-01 3.25e-02 2.68e-01]			
	HBMCL2	μ_ξ	[-1.74 -4.61e-02 6.61e-02]			
		Σ_ξ	[3.97e-03 9.86e-03 6.24e-03]			
		$diag(\Upsilon)$	[3.27e-01 1.20e-01 8.64e-01]			
		g	[4.17e+02 1.53e+02 1.10e+03]			

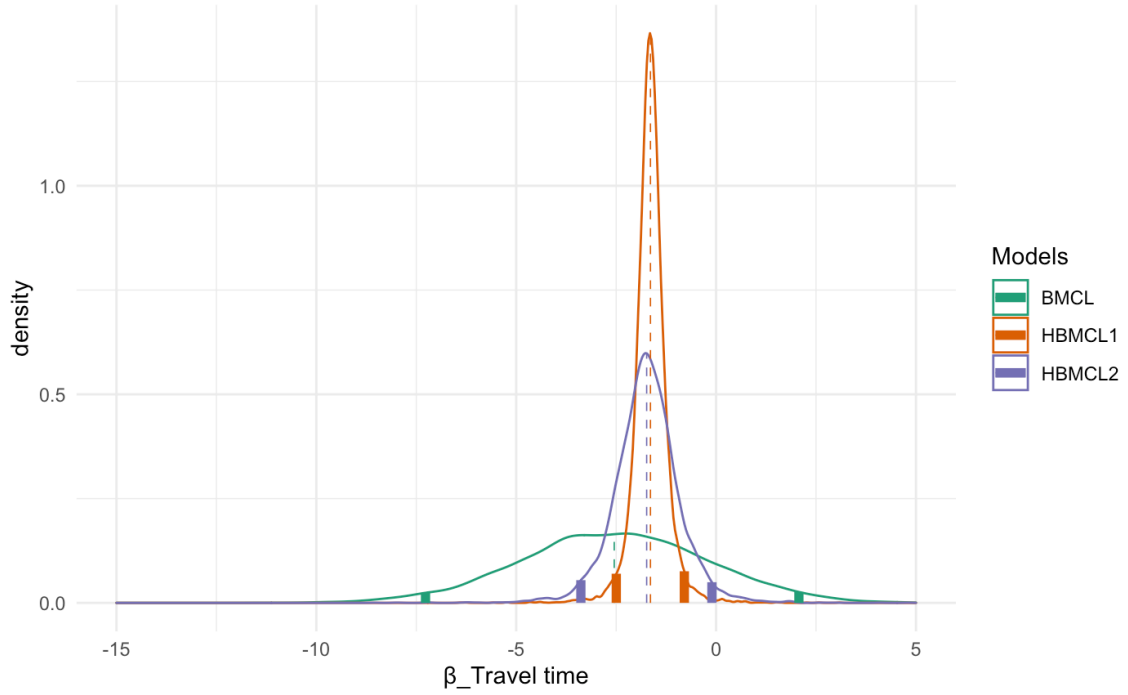


Figure 6.14 Kernel density of $\beta_{Travel\,Time}$ for Scenario 1, Cluster 2.

The distribution of this cluster's coefficient of travel time displays a similar trend as for Cluster 0 and 1. Additionally, the 95% CIs for the three models are (-7.18, 2.15), (-0.795, -2.47), and (-0.0863, -3.44). The 95% CI of BMCL has a positive upper bound, which indicates that this model is not a good choice for modeling the truck route choice of this cluster.

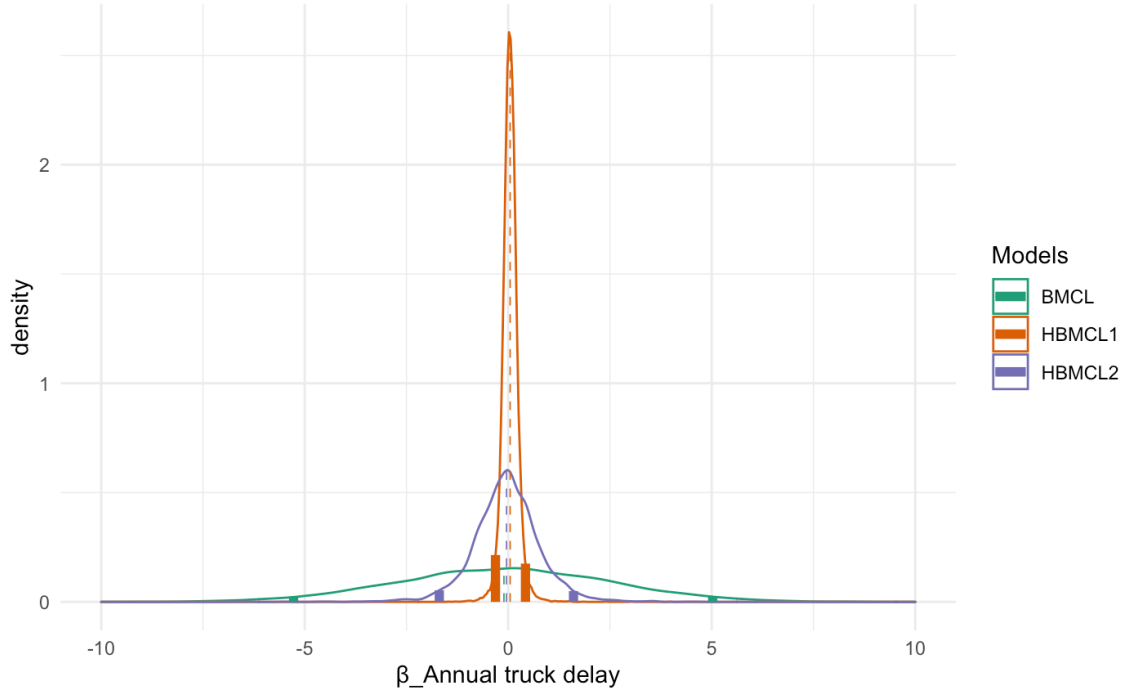


Figure 6.15 Kernel density of $\beta_{Anual\ truck\ delay}$ for Scenario 1, Cluster 2.

The 95% CIs for the coefficient of the annual truck delay from the three models are (4.82, -5.11), (0.426, -0.320), and (-1.75, 1.62). Figure 6.15 shows a strong contrast of the predicted densities of $\beta_{Annual\ truck\ delay}$ among the three models. The variance of $\beta_{Annual\ truck\ delay}$ with HBMCL1 is much smaller than with HBMCL2 and BMCL, with the former smaller than the latter. The annual truck delay does not affect truck drivers' opinions much, as can be seen from the mean from three models being all close to 0.

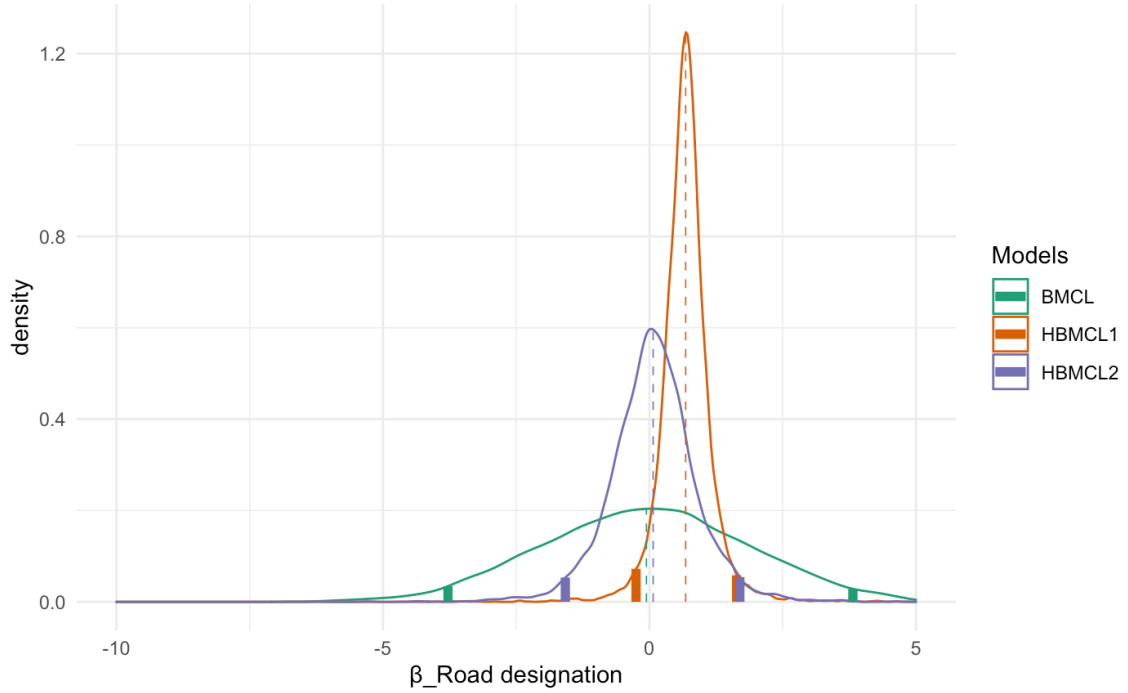


Figure 6.16 Kernel density of $\beta_{Roadway\ designation}$ for Scenario 1, Cluster 2.

The impact of road designation on affecting drivers' route choice in Cluster 2 is mild, compared to the travel time, while being slightly greater than the annual truck delay, according to the estimated mean of $\beta_{Roadway\ designation}$. Specifically, both BMCL and HBMCL reported a mean value close to 0, yet HBMCL1 shows that the mean is approximately 0.5. In addition, the 95% CIs for the coefficient of the roadway designation from the three models are (-3.93, 3.84), (-0.295, 1.64), and (-1.63, 1.74), respectively.

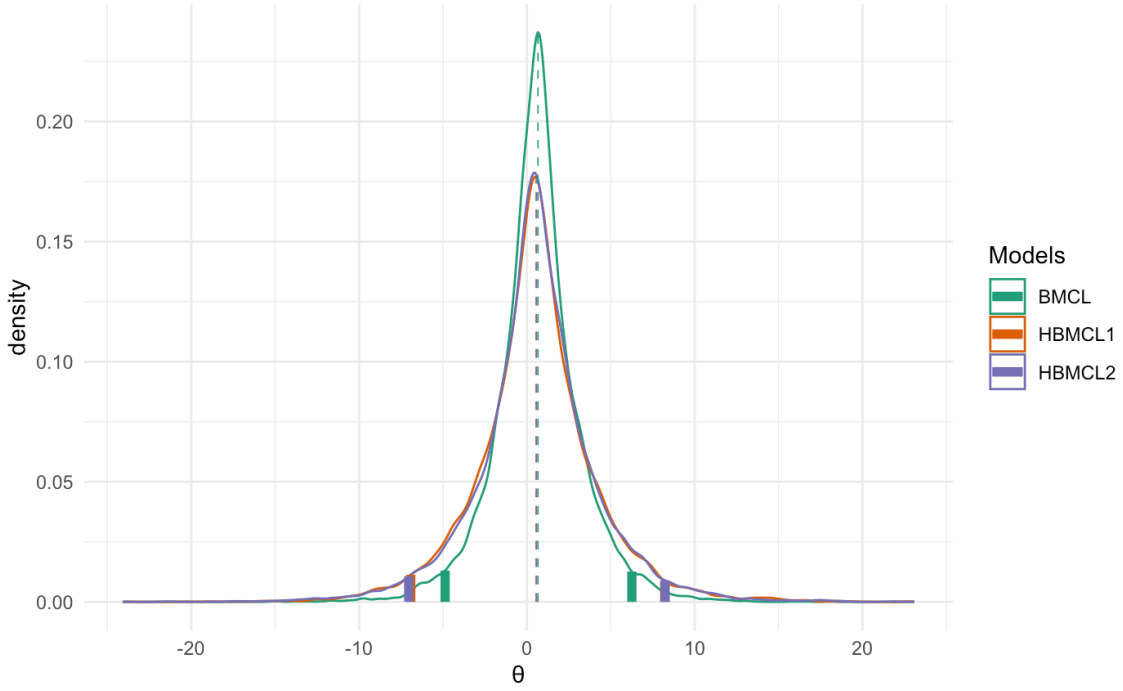


Figure 6.17 Kernel density of θ for Scenario 1, Cluster 2.

The 95% CIs for the coefficient of the annual truck delay from the three models are (-5.10, 5.99), (-6.88, 8.23), and (-6.97, 8.03), as shown in Figure 6.17. The estimated distribution of θ is similar across clusters of this scenario, with BMCL1 having a smaller variance than the other two models.

6.2.2. Estimated parameters for Scenario 2

In Scenario 2, where all alternatives are freeways, the travel time attribute is no longer the most significant factor in truckers' perceived utilities. For example, in Cluster 1, the parameter of the S/IS road indicator has means of 2.99, 1.95, and 2.12 from each model, yet the mean of the travel time parameter has absolute values of 2.49, 1.71, and 1.97. The absolute value of delay per lane is between 0.65 and 1.32.

Table 6.7 Estimated parameter, Scenario 2, Cluster 0.

Model	BMCL		HBMCL1		HBMCL2	
Variable	Mean	Variance	Mean	Variance	Mean	Variance
Travel time (min)	-1.3843	7.0109	-0.8165	1.4568	-1.0094	2.7116
Delay per lane	-1.5094	6.8204	-0.9497	1.7365	-1.1204	2.5347
S/IS indicator	1.1585	3.2178	1.1983	1.2935	1.1748	1.3231
CF	-0.0700	1.1585	-0.0900	1.3843	-0.0725	1.2011
Hyper-parameter	HBMCL1	μ_ξ	[-0.9594 -1.1822 0.9902]			
		Σ_ξ	[0.0024 0.0022 0.0012]			
		$diag(\Upsilon)$	[0.0919 0.1025 0.1721]			
	HBMCL2	μ_ξ	[-1.1348 -1.1804 1.0263]			
		Σ_ξ	[7.14e-03 5.79e-03 6.04e-01]			
		$diag(\Upsilon)$	[4.13e-02 2.12e-01 2.28e-01]			
		g	[5.29e+01 2.71e+02 2.92e+02]			

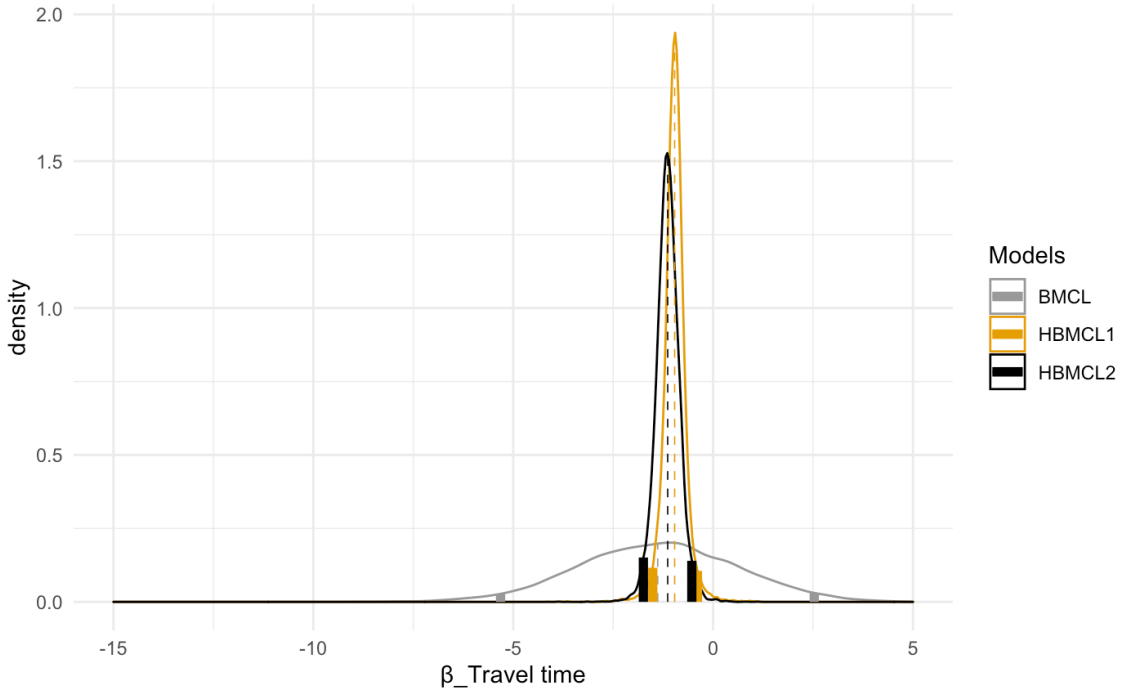


Figure 6.18 Kernel density of $\beta_{Travel\,Time}$ for Scenario 2, Cluster 0.

The estimated density of $\beta_{Travel\,time}$ in Scenario 2 is different from that of clusters in Scenario 1, according to Figure 6.18. Generally, HBMCL1 and HBMCL2 have similar estimations of the coefficient of the travel time; both are significantly more concentrated than BMCL. The corresponding 95% confidence interval of the three models are (-1.52, -0.402), (-1.74, -0.518), and (-5.23, 2.50). Since BMCL has a positive upper bound of the confidence interval, it is not suggested to use it for modeling this cluster.

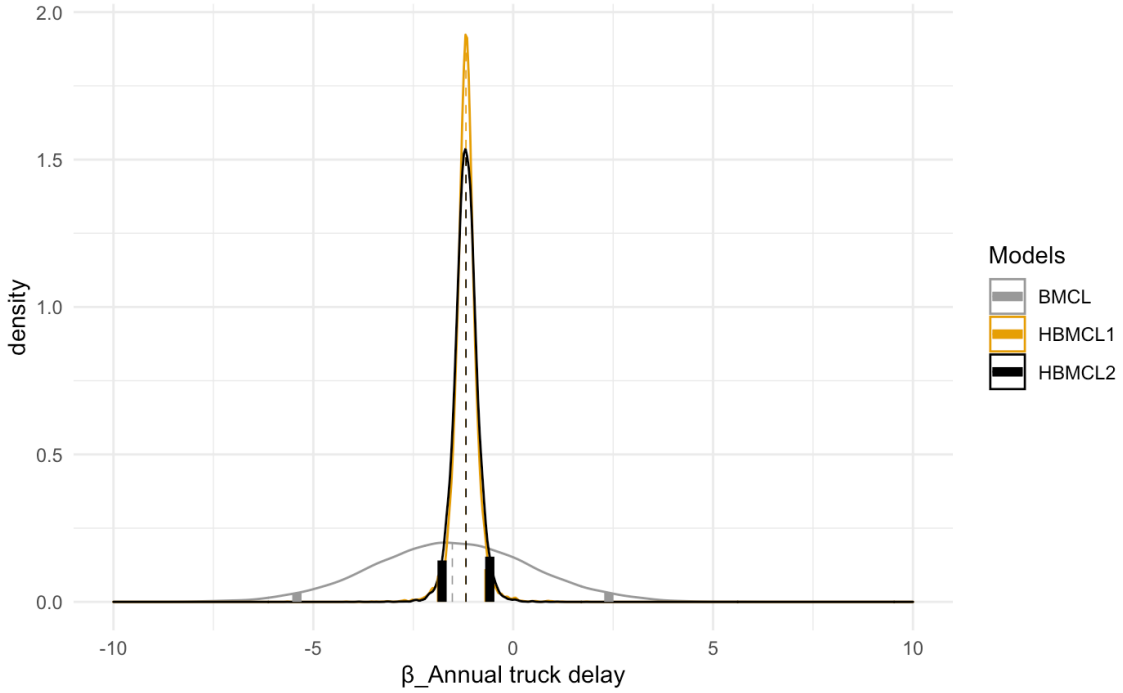


Figure 6.19 Kernel density of $\beta_{Anual\ truck\ delay}$ for Scenario 2, Cluster 0.

Similar to the $\beta_{Travel\ time}$, the $\beta_{Annual\ truck\ delay}$ also indicates that both hierarchical models perform better than BMCL does. HBMCL1 provides a 95% confidence interval of (-1.79, 0.606). The 95% CI of HBMCL2 is (-1.77, -0.568), and the 95% CI of BMCL is (-5.33, 2.36).

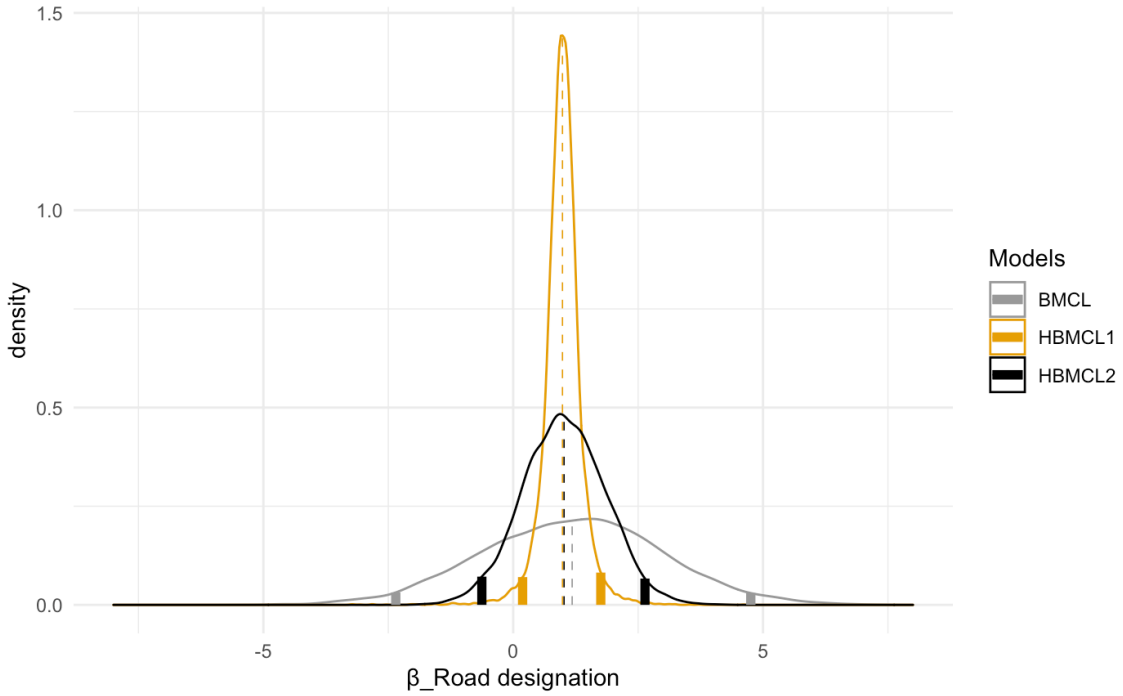


Figure 6.20 Kernel density of $\beta_{Roadway\ designation}$ for Scenario 2, Cluster 0.

Different from $\beta_{Travel\ time}$ and $\beta_{Annual\ truck\ delay}$, the $\beta_{Road\ designation}$ shows that HBMCL1 provides a density with the smallest degree of dispersion, compared to BMCL and HBCML2. The corresponding 95% CIs of the three models are (0.248,1.75), (-2.36, 4.69), and (-2.36,4.69), which is shown in Figure 6.20. Figure 6.21 displays the estimated density of θ. The 95% CIs of BMCL, HBMCL1, and HBMCL2 are (-3.55,3.60), (-4.27, 4.11), and (-4.35,4.07), respectively.

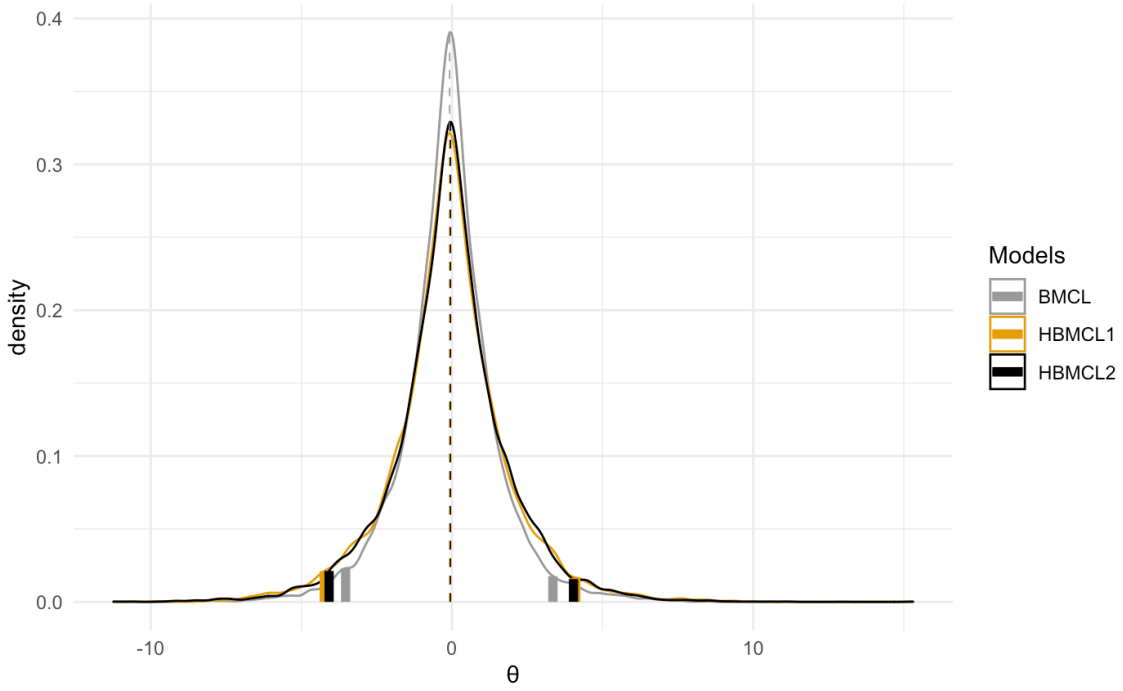


Figure 6.21 Kernel density of θ for Scenario 2, Cluster 0.

Parameters estimated in Cluster 2 have a similar pattern as from Cluster 1, for which the existence of an S/IS road takes a larger share in the perceived utility, as shown in Table 6.8. Even though the parameters of travel time and delay per lane in Cluster 2 are in a smaller magnitude compared to Cluster 1, the travel time is indeed more significant than delay per lane, as judged from the relative ratio between the two factors' parameters. Travel time, delay per lane, as well as the S/IS road are of similar importance in affecting trucks' route choice. Variances of all parameters except the CF range from 1.45 to 9.29. In addition, consistent with results in Scenario 1, the road network geometry factor in Scenario 2 is also the lowest priority when selecting routes.

Table 6.8 Estimated parameter, Scenario 2, Cluster 1.

Model	BMCL		HBMCL1		HBMCL2	
Variable	Mean	Variance	Mean	Variance	Mean	Variance
Travel time (min)	-2.4934	6.4543	-1.7091	3.0109	-1.9749	2.4169
Delay per lane	-1.3186	7.2611	-0.6540	1.5204	-0.8105	1.5873
S/IS indicator	2.9975	5.6455	1.9549	3.2178	2.1239	2.7396
CF	0.1050	1.4934	0.0890	1.3186	0.1054	1.0166
Hyper-parameter	HBMCL1	μ_ξ	[-1.82 -7.21e-01 2.22]			
		Σ_ξ	[1.32e-01 2.80e-01 1.11e-01]			
		$diag(\Upsilon)$	[1.54e-01 8.65e-02 1.81e-01]			
	HBMCL2	μ_ξ	[-2.02e+00 -8.17e-01 4.03e+01]			
		Σ_ξ	[-7.03e-02 -2.11e-02 2.33e-02]			
		$diag(\Upsilon)$	[5.74e-01 2.65e-01 8.46e-01]			
		g	[7.32e+01 3.39e+01 1.07e+02]			

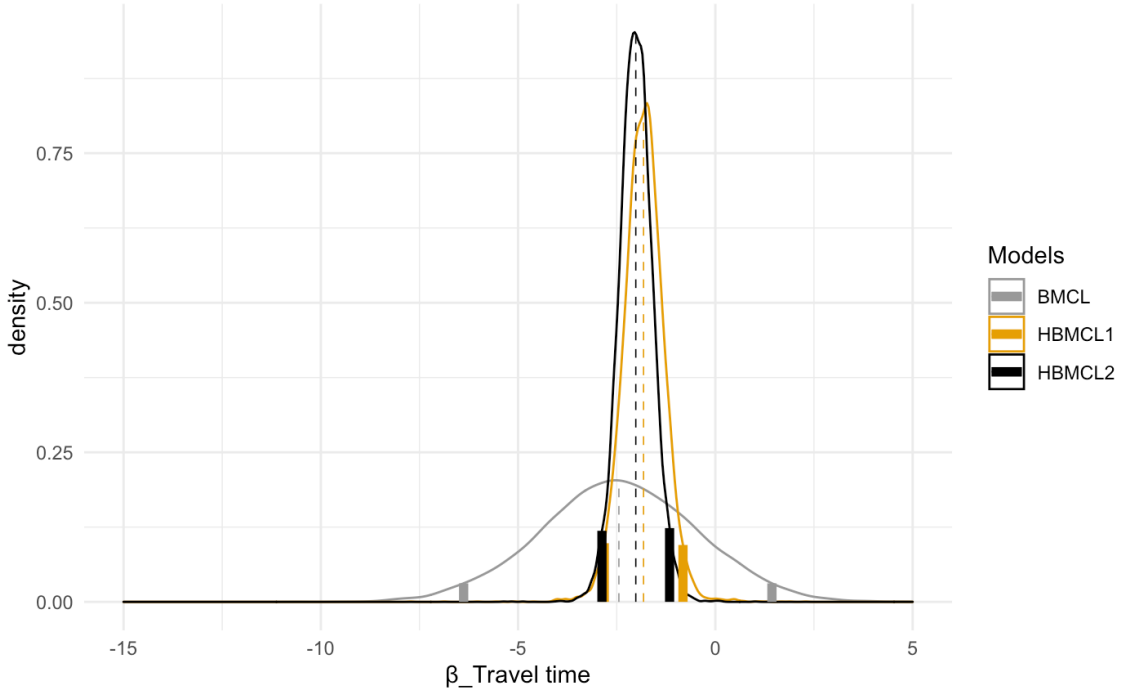


Figure 6.22 Kernel density of $\beta_{Travel\,Time}$ for Scenario 2, Cluster 1.

The corresponding 95% confidence intervals of the three model are (-2.82, -0.821), (-2.88, -1.16), and (-6.38, 1.43). Different from Cluster 1 and all clusters in Scenario 1, $\beta_{Travel\,time}$ estimated with HBMCL2 have the lowest degree of dispersion, compared to other models. The $\beta_{Annual\,truck\,delay}$ indicate similar performance of the models. HBMCL1 provides a 95% confidence interval of (-1.89, 0.439). The 95% CI of HBMCL2 is (-1.55, -0.0815), and the 95% CI of BMCL is (-5.22, 2.52), as displayed in Figure 6.23.

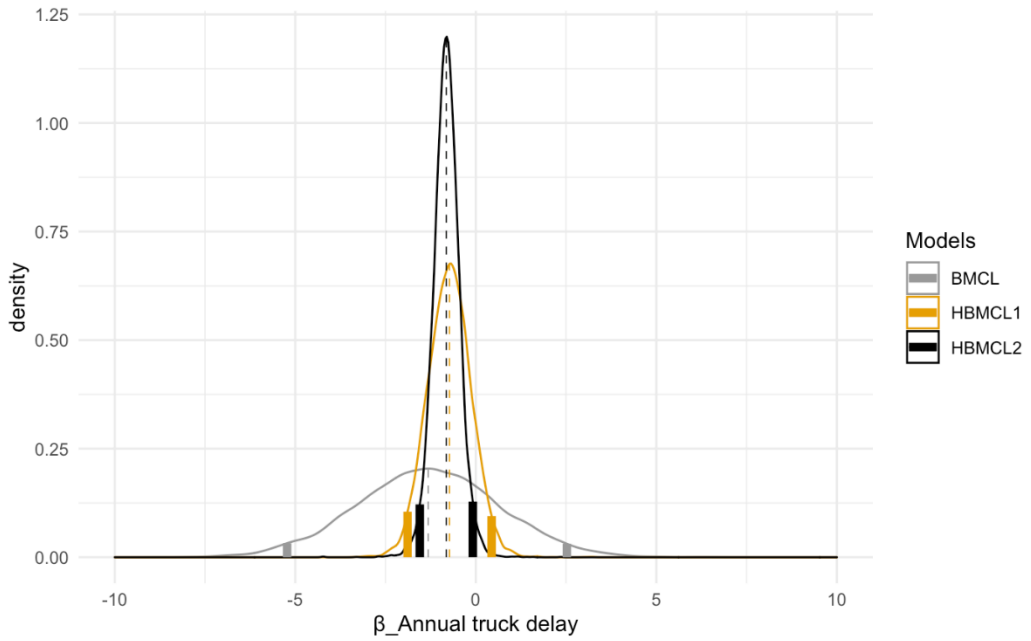


Figure 6.23 Kernel density of $\beta_{Anual\ truck\ delay}$ for Scenario 2, Cluster 1.

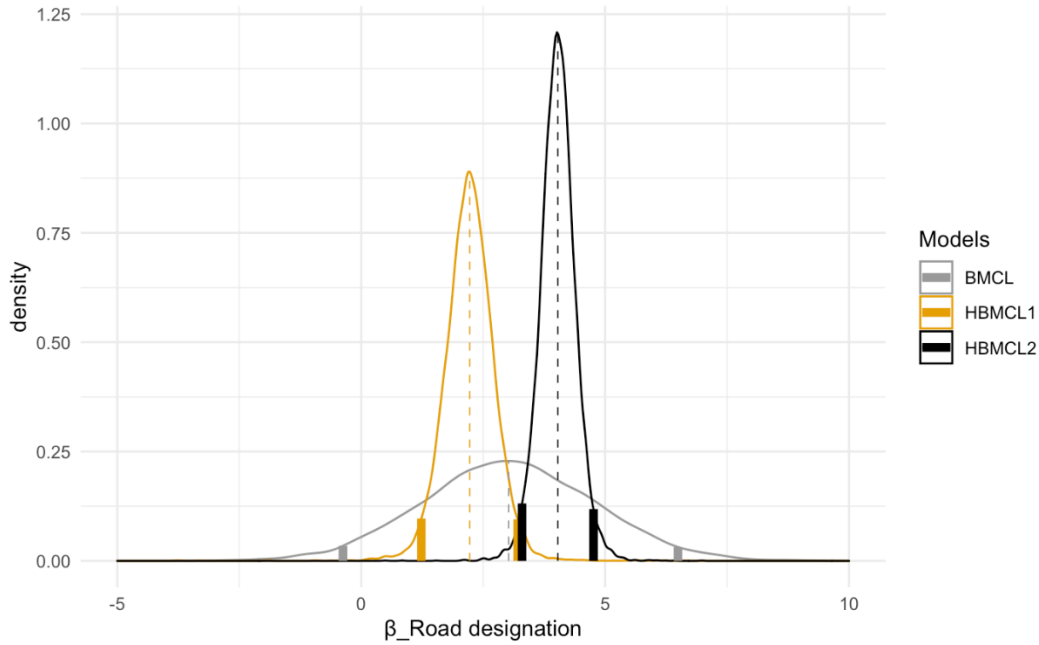


Figure 6.24 Kernel density of $\beta_{Roadway\ designation}$ for Scenario 2, Cluster 1.

Figure 6.24 shows the density of $\beta_{Road\ designation}$. It is noted that the estimated means of the $\beta_{Road\ designation}$ from each of the three models are different from each other, with the largest value from HBMCL1 and the smallest from HBMCL2 and BMCL in between, from which the 95% CIs are (1.23, 3.21), (3.30, 4.76), and (-0.374, 6.49). Figure 6.25 below shows the density of θ.

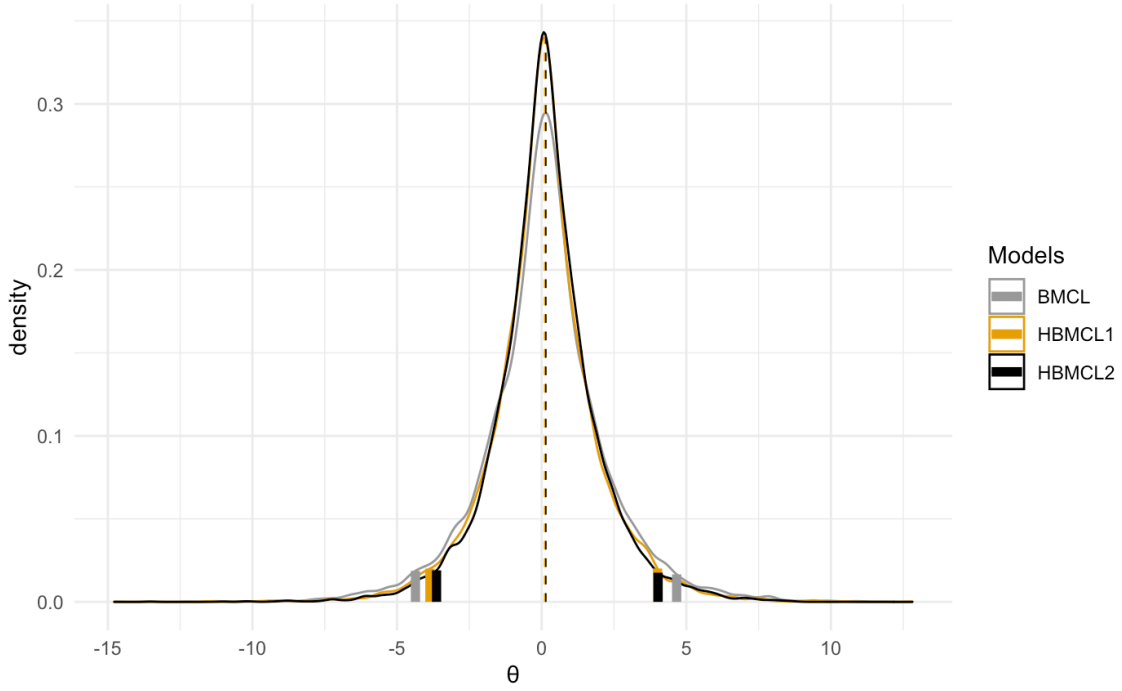


Figure 6.25 Kernel density of θ for Scenario 2, Cluster 1.

Table 6.9 Estimated parameter, Scenario 2, Cluster 2.

Model	BMCL		HBMCL1		HBMCL2	
Variable	Mean	Variance	Mean	Variance	Mean	Variance
Travel time (min)	-0.9335	5.9556	-0.7034	2.9456	-0.7234	1.2727
Delay per lane	-0.1419	0.7211	-0.0782	0.1917	-0.0761	0.1211
S/IS indicator	3.2197	5.7544	2.7979	4.1738	3.1538	3.1637
CF	-0.1753	1.2099	-0.0924	0.6248	-0.1634	1.2689
Hyper-parameter	HBMCL1	μ_ξ	[-7.14e-01 -8.40e-02 2.92]			
		Σ_ξ	[2.48e-02 2.45e-02 7.41e-02]			
		$diag(\Upsilon)$	[1.17e-01 1.16e-01 2.64e-01]			
	HBMCL2	μ_ξ	[-9.01e-01 -2.88e-01 3.48e+00]			
		Σ_ξ	[1.84e-02 1.26e-02 2.93e-02]			
		$diag(\Upsilon)$	[2.53e-01 1.18e-01 1.73e-01]			
		g	[3.23e+01 1.51e+01 2.21e+01]			

The estimated parameters of Cluster 2 in Scenario 2 are documented in Table 6.9. Generally speaking, both Cluster 2 and 1 show that HBMCL2 provides a slightly more

aggregated density than HBMCL1, with both significantly more aggregated than BMCL. In this scenario, the roadway designation is the dominant attribute that heavily affects drivers' route choice, followed by the travel time, which is the second more important factor. The roadway congestion situation represented by the S/IS indicator and the network density are less important factors to consider. Figure 6.26 shows the density of $\beta_{travel\ time}$. The 95% CIs from each model are (-5.73, 3.84), (-1.39, -0.0153), and (-1.43, -0.376). The estimated densities of $\beta_{Annual\ truck\ delay}$, $\beta_{roadway\ designation}$, and θ are shown in Figure 6.27, 6.28, and 6.29.

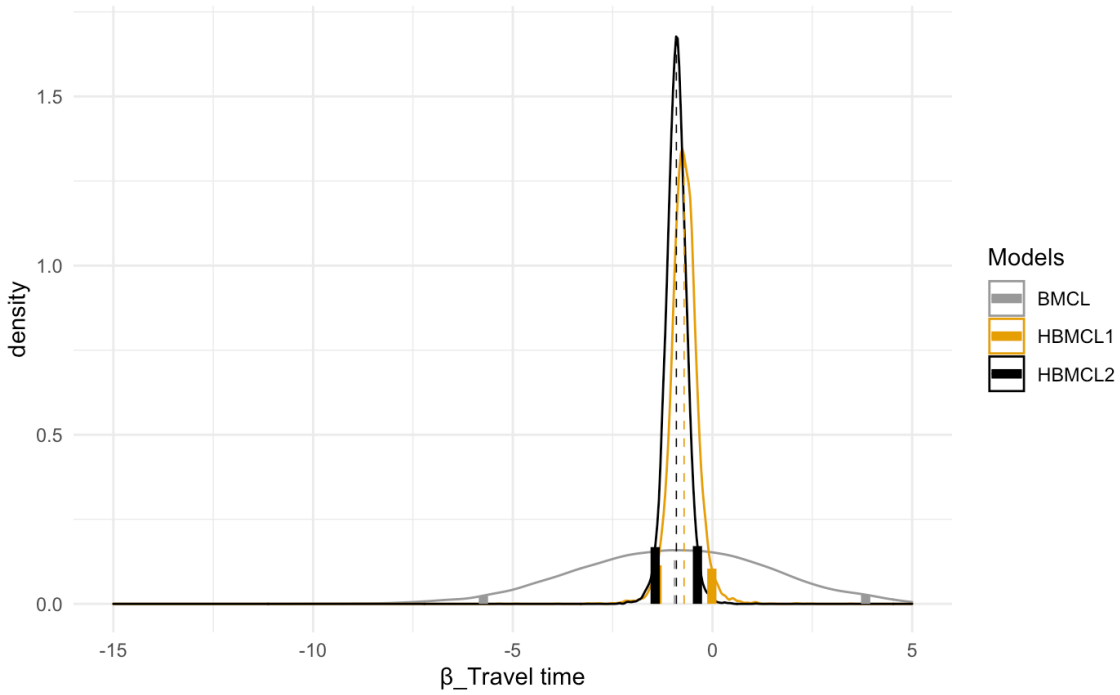


Figure 6.26 Kernel density of $\beta_{Travel\ Time}$ for Scenario 2, Cluster 2.

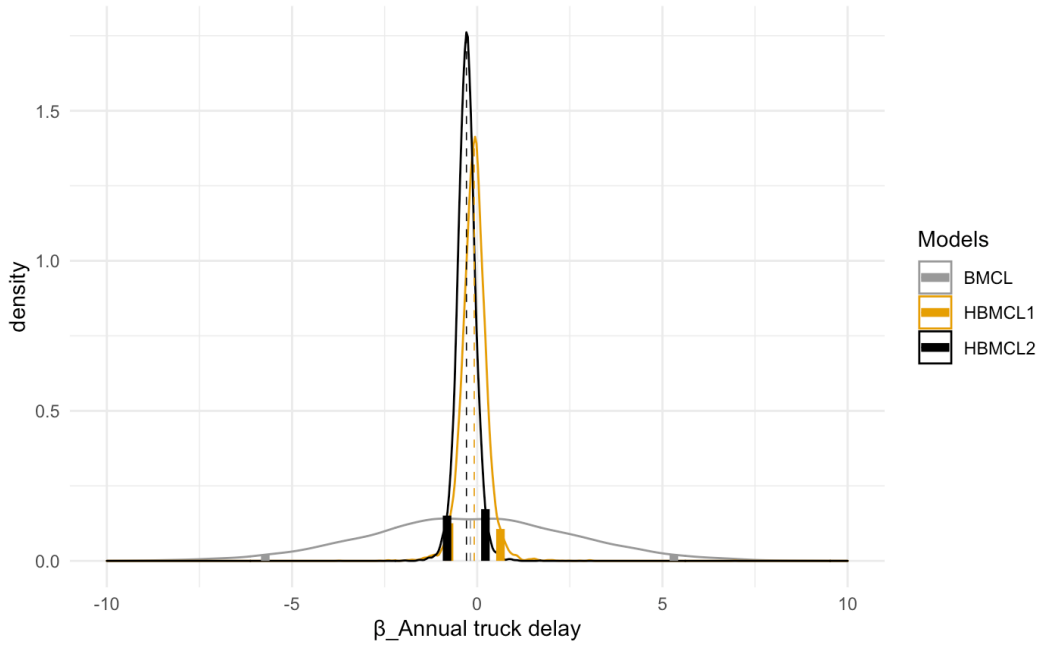


Figure 6.27 Kernel density of $\beta_{Anual\ truck\ delay}$ for Scenario 2, Cluster 2.

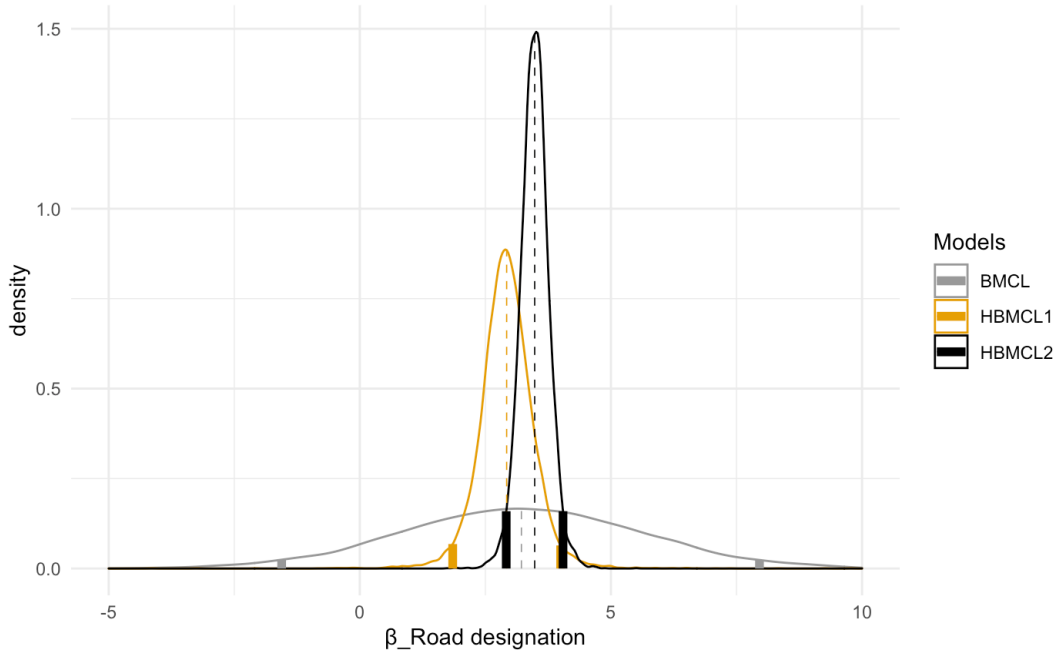


Figure 6.28 Kernel density of $\beta_{Roadway\ designation}$ for Scenario 2, Cluster 2.

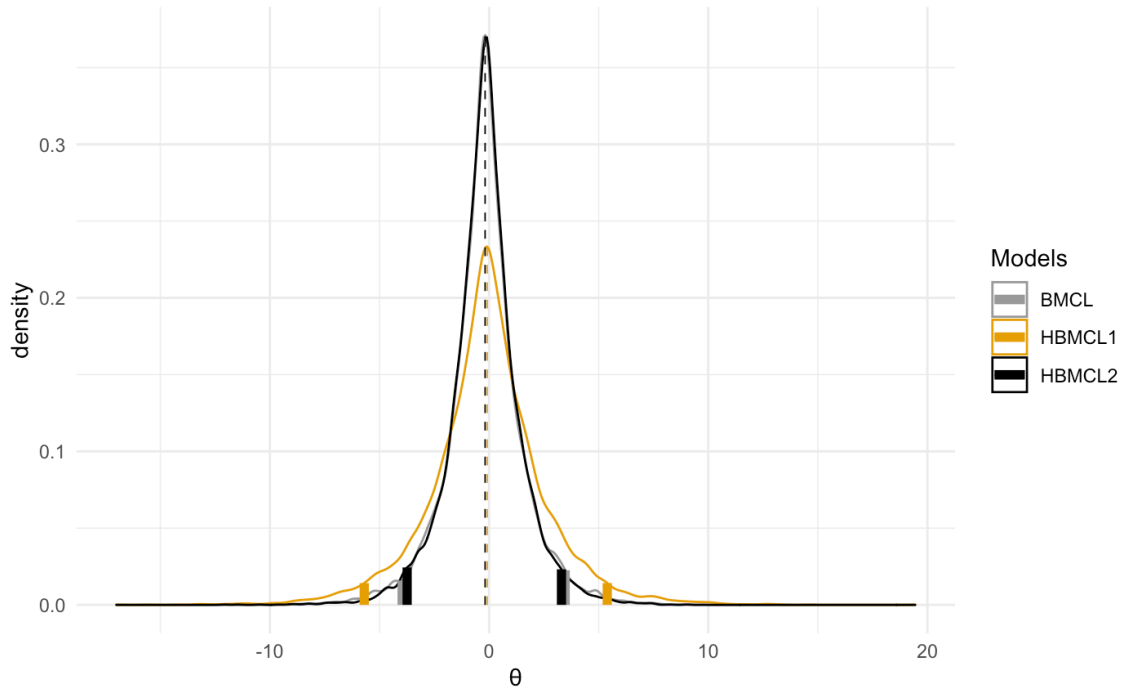


Figure 6.29 Kernel density of θ for Scenario 2, Cluster 2.

To summarize, the 95% confidence interval of the estimated density for the coefficients can be used for model validation. If the predicted 95% CI for the coefficient of travel time contains a large proportion of positive values, then the model prediction is questionable since the coefficient of travel time should be negative in most cases. There are situations in which the coefficient of travel time is a positive value where the driver has used a path with a long travel time. This can be addressed by the right proportion out of the 95% CI in the estimated density. From this point of view, HBMCL1 and HBMCL2 are preferred.

6.2.3. The preference heterogeneity

As mentioned earlier, preference heterogeneity refers to the variation of an individual's taste in evaluating information when making decisions. In our model, the truckers' preference heterogeneity is shown by the relative variation of the mean of a random coefficient to its standard deviation, which is the coefficient of variation. The mathematical form for the coefficient of variation *cv* is $cv = \frac{\sigma}{|\mu|}$, where σ is the standard deviation of a random variable and $|\mu|$ is the absolute mean. For a better illustration of the preference heterogeneity, Table 6.10 lists the value of it.

The magnitude of the *cv* for each coefficient indicates the consistency among drivers' preferences of the corresponding attribute. A large value of *cv* shows that drivers' opinions vary largely on this attribute, whereas a small value of *cv* represents consistent ideas among drivers. Combining the results from the previous section, if an attribute is significant with a small *cv* of its coefficient, it is concluded that this attribute is considered critical for most drivers. In contrast, if an attribute is not significant and has a large *cv*, it is concluded that it is only considered important by a minority.

Table 6.10 The coefficient of variation for each parameter.

		Scenario 1			Scenario 2		
Attributes	Model	Cluster 0	Cluster 1	Cluster 2	Cluster 0	Cluster 1	Cluster 2
Travel time	BMCL	0.4251	1.0005	0.9340	1.9128	1.0189	2.6142
	HBMCL1	0.3339	0.9385	0.7912	1.4783	1.0153	2.4400
	HBMCL2	0.2774	0.7347	0.7987	1.6313	0.7872	1.5594
Truck delay per lane	BMCL	1.4123	1.3707	5.8075	1.7302	2.0436	5.9860
	HBMCL1	1.2780	1.3492	3.9966	1.3876	1.8854	5.5986
	HBMCL2	1.8522	1.6644	4.5597	1.4210	1.5545	4.5750
Toll price	BMCL	0.7618	1.8621				
	HBMCL1	0.6266	1.8696				
	HBMCL2	0.6303	1.1767				
S/IS indicator	BMCL	5.1837		6.9578	1.5484	0.7927	0.7450
	HBMCL1	1.6875		2.1470	0.9491	0.9176	0.7302
	HBMCL2	2.6531		3.5005	0.9791	0.7793	0.5640
CF	BMCL	10.2355	3.0609	2.1468	10.9355	11.6378	6.2738
	HBMCL1	11.2201	2.8310	3.0416	11.3289	12.9023	8.5500
	HBMCL2	11.6432	3.1277	2.9673	11.5974	9.5685	6.8941

Generally, in all clusters of Scenario 1, the travel time has the lowest coefficient of variation. In other words, truckers' preference for travel time is the least heterogeneous compared to attributes considered in this study. With a combination of the results from Tables 6.4 to 6.9, this implies that travel time is an essential factor for most truck drivers when making route choice decisions. In Cluster 0, Scenario 1, the truckers' preference for toll alternatives only shows slight heterogeneity. In contrast, truckers' preference in Cluster 1 of this scenario shows notable heterogeneity on the use of tollways. Additionally, the truck drivers' preference for using a road that does not contains any local road segments displays tremendous heterogeneity in Scenario 1.

Compared to Scenario 1, most of the truckers in Scenario 2 think that using a route that only contains interstate highways or state highways is important, shown by the last three columns in Table 6.10. The preference of travel time displays slightly higher heterogeneity than the highway indicator.

Additionally, the preference of truck drivers varies drastically on the CF, which indicates that drivers' preference diverges largely on whether to use a route with many shared segments. In summary, it is found that the most significant attribute has the least preference heterogeneity while the least significant parameter has the most remarkable preference heterogeneity.

6.3. The estimated value of travel time savings

This part analyzes the value of travel time savings (VTTS) from the proposed models. The kernel densities of the VTTS reflected by the random coefficient in each model are listed in Figure 6.30. Since the kernel density has a very long right tail, the

density shown in the graph is truncated, with records that fall out of the 0.5th and 99.5th percentile removed for a better representation of the majority. In Scenario 1, Cluster 0, three models reported a similar mean estimation of the VTTS while the estimated variance displays a large discrepancy.

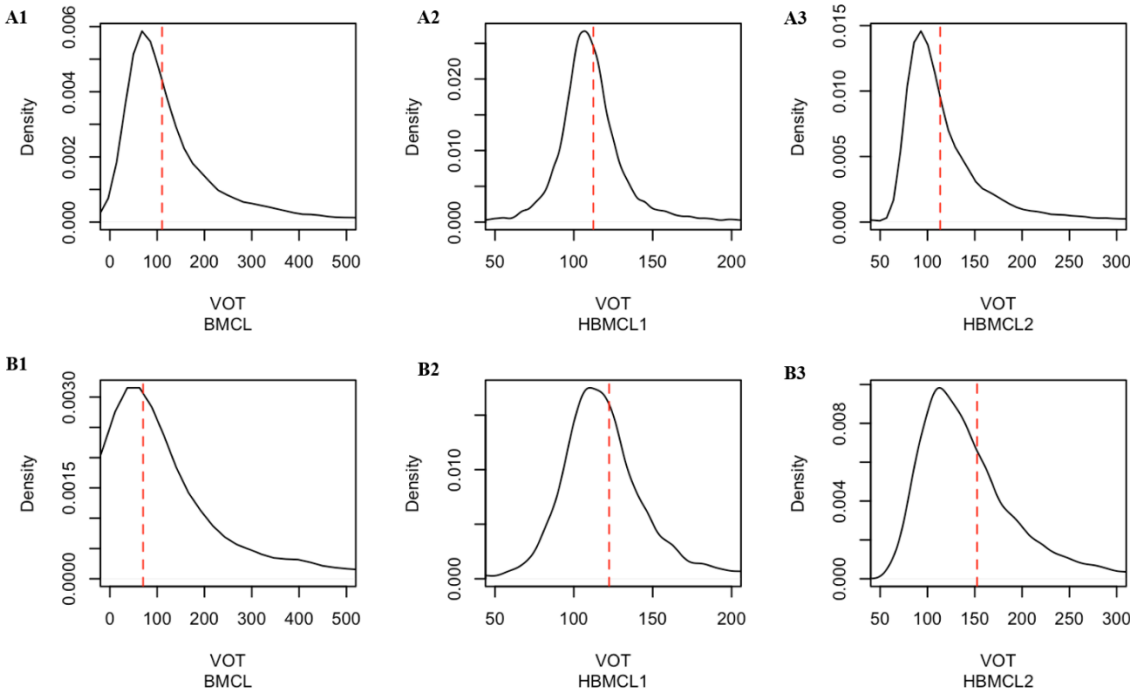


Figure 6.30 Kernel density of value of travel time savings.
(A1-A3) Kernel density of VTTS for Scenario 1, Cluster 0. (B1-B3) Kernel density of VTTS for Scenario 1, Cluster 1. The red dashed line in each subplot displays the mean of each density.

The mean VTTS from BMCL is $110.3056 per hour, and the variance is 205,771.3. The mean VTTSs from HBMCL1 and HBMCL2 are $112.3983 per hour and $113.4255 per hour, with a variance of 794.5488 and 30,444.79, respectively.

In contrast to Cluster 0, the discrepancy in the mean VTTS reported by each model in Cluster 1 varies notably. The mean and variance in BMCL are $70.20283 per hour and 457,290.2. The mean VTTS reported by HBMCL1 is $122.3652 per hour and $152.5073 per hour. Similar to Scenario 0, the variances in VTTS from each model in Scenario 1 differ from each other, which are 457,290.2, 1,275.678, and 5,524.264, respectively.

Moreover, with a comparison of Figure 6.2 (A1), (A2), and (A3), it is pointed out that HBMCL2 and HBMCL1 have a much more concentrated distribution of the VTTS than does BMCL, which indicates that adding additional hierarchy to the variance of the Bayesian model contributes to smaller variances of the estimated parameters.

6.4. Predicted percentages of road use

This section displays the predicted road use percentages of BMCL, HBMCL1, and HBMCL2, which are exhibited in Figure 6.31 and 6.32. As shown in the cross-validation process that the proposed model outperforms the C-logit model, each route's predicted probabilities among all clusters of the proposed models are closer to the true probability, compared to the C-logit model with fixed parameters. In each figure, the black bar stands for the true probability, and bars with different colors stand for models used. In the first cluster of Scenario 1, even though 12 routes are included, approximately 90% of the trips are on route 60. The HBMCL1 provides the closest prediction of the route probability, followed by the HBMCL2.

All the proposed models give the highest predicted percentage on route 60. The main difference among those three models is the probability of route 42, 86, and 198.

Since there are 12 routes in this cluster, even though the C-logit did not perform well in probability prediction, the maximum prediction error is still close to the proposed models.

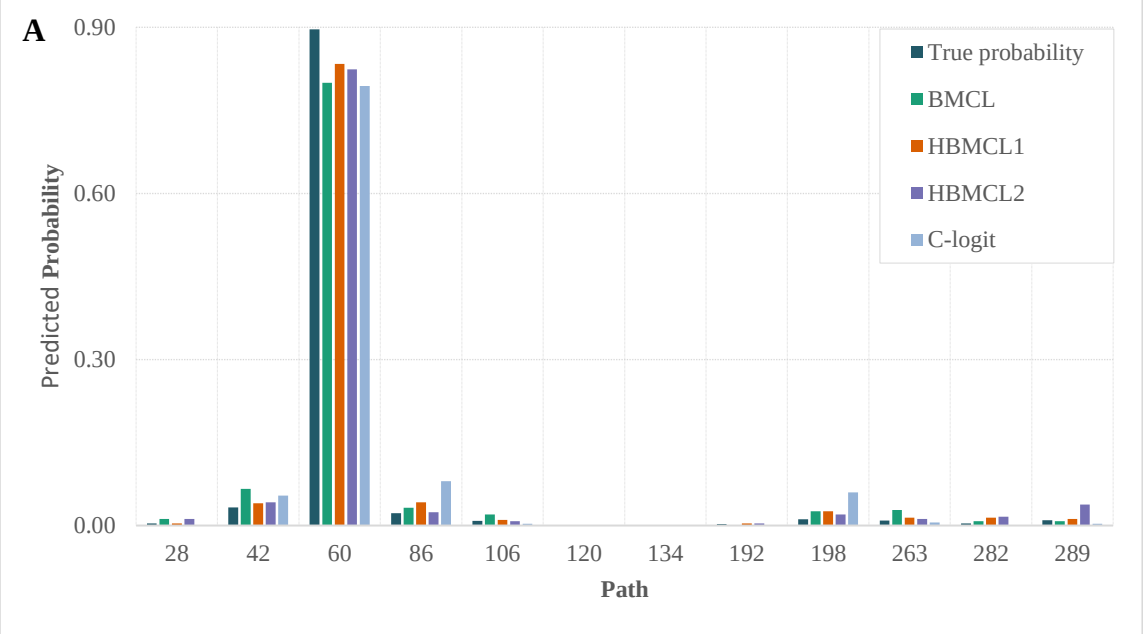
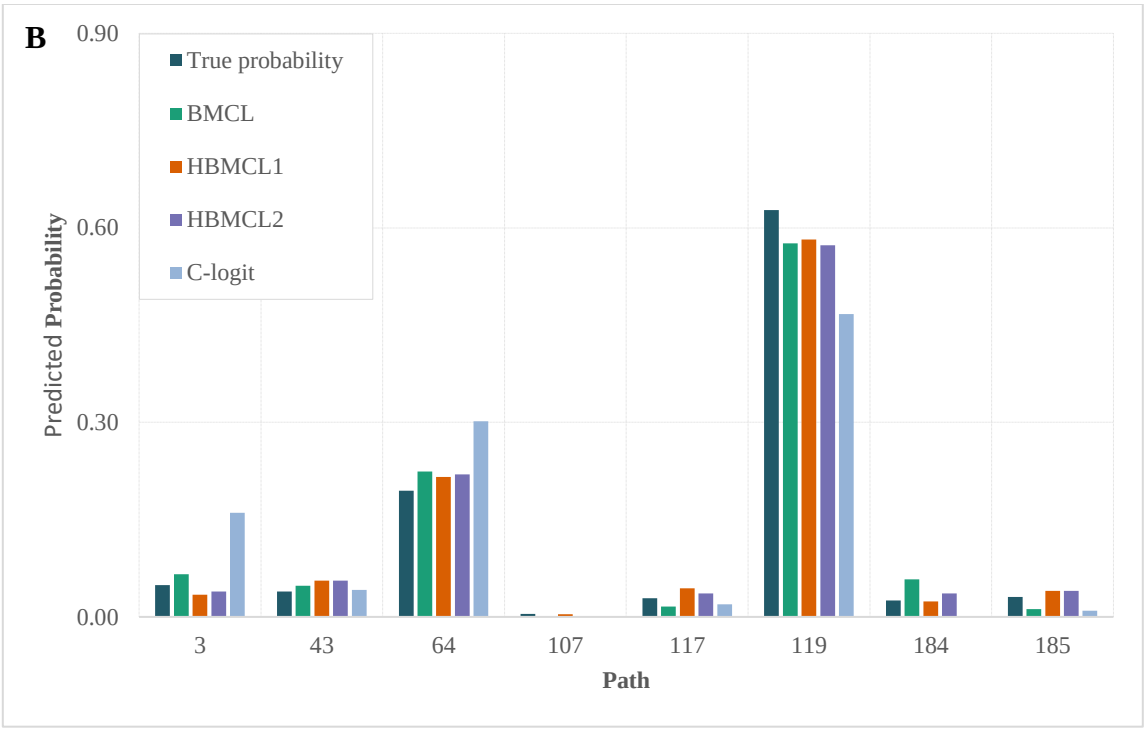


Figure 6.31 Probability by path (Scenario 1).
(A) Predicted probability, Cluster 0.



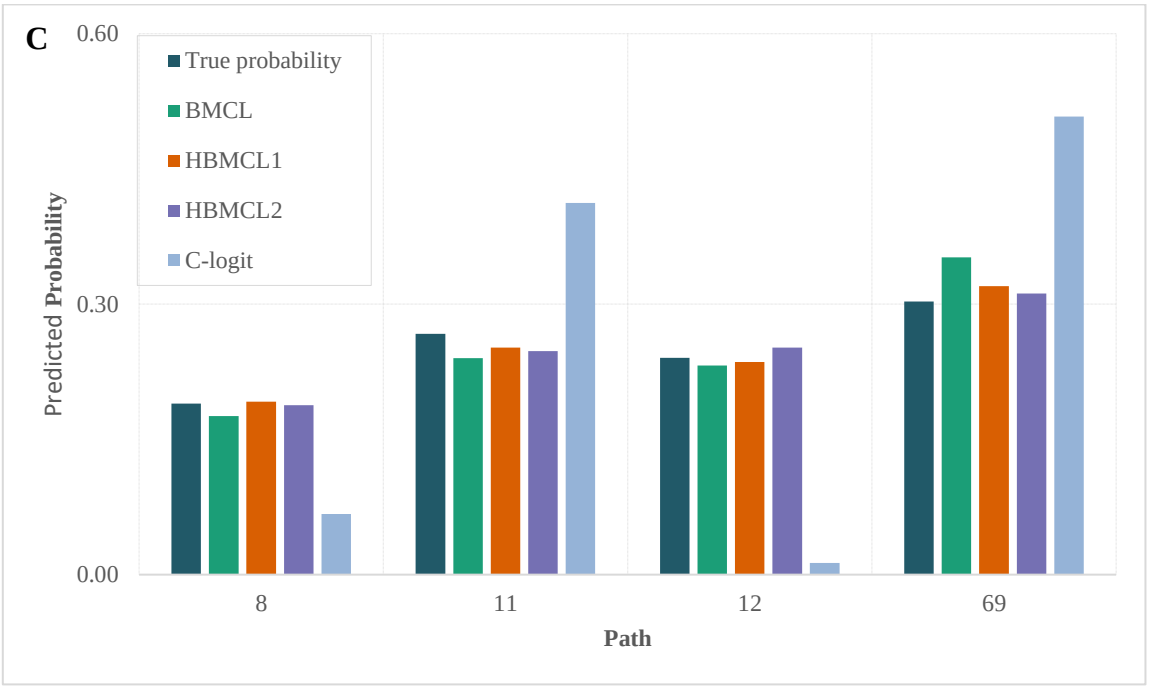


Figure 6.31 Continued.
(B) Predicted probability, Cluster 1. (C) Predicted probability, Cluster 2.

Cluster 1 shows a similar trend as shown by Cluster 0. The most notable difference between the true probability and the predicted probability is observed in Cluster 2. The four routes in Cluster 2 have a similar probability of being used, yet the C-logit provides a distinct prediction on their probabilities. Route 12, for example, is significantly underestimated with a predicted probability close to 0, whereas the true probability is about 25%. In Scenario 2, the HBMCL1 provides more accurate predictions than other models do in Cluster 0 and 1. The BMCL, however, has the most precisely predicted probability in Cluster 2.

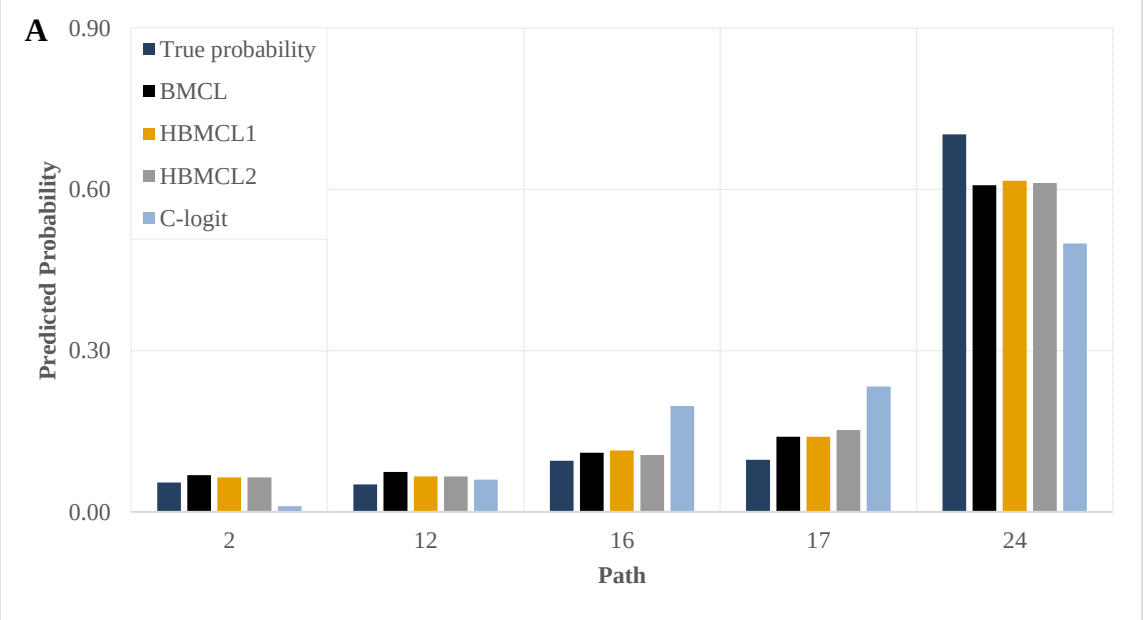
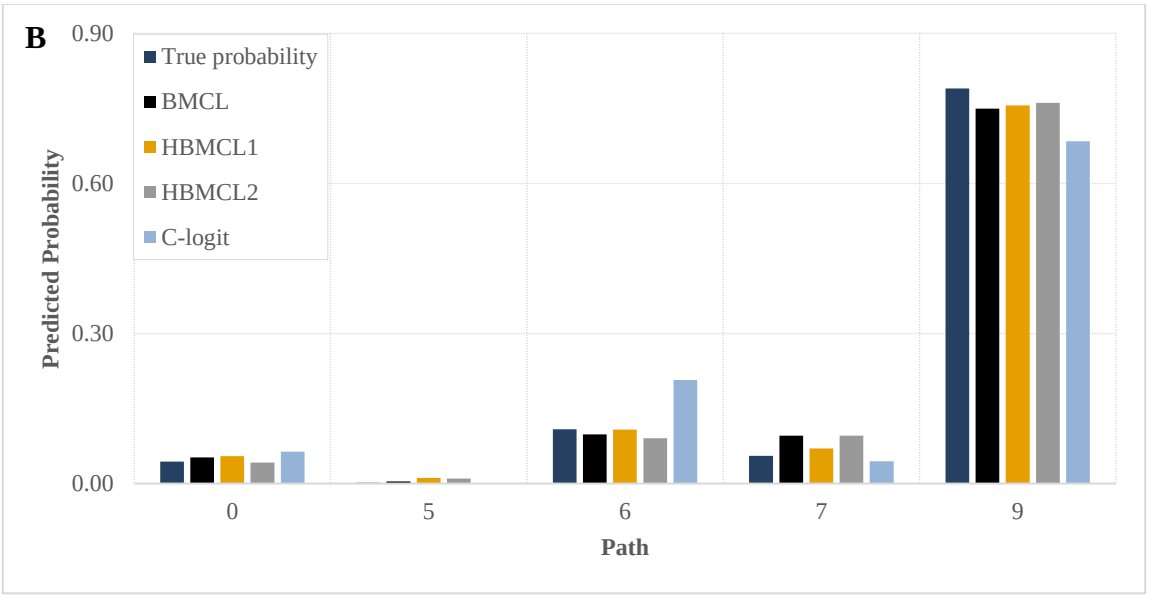


Figure 6.32 Probability by path (Scenario 2).
(A) Predicted probability, Cluster 0.



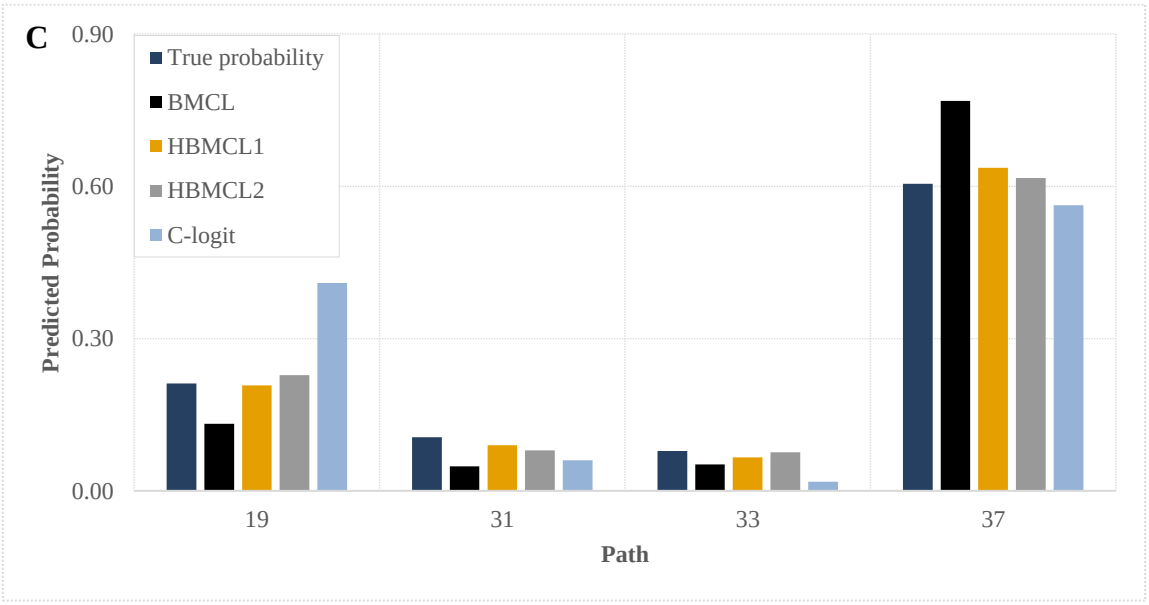


Figure 6.32 Continued.
(B) Predicted probability, Cluster 1. (C) Predicted probability, Cluster 2.

Among the proposed models, the HBMCL1 performs the best with five over six clusters, followed by the HBMCL2. One possible reason that HBMCL2 did not perform as well than did HBMCL1 is that the starting value of hyperparameters is far from the true value, making the algorithm converged to a local optimum that is not close to the global optimum. Consequently, considering the implementation difficulty and the fact that the HABMXL2 does not converge in two cases (Scenario 2, Cluster 2 and Scenario 1, Cluster 0), it is highly recommended that the proposed HBMCL1 model be used by practitioners.

7. CONCLUSION AND DISCUSSION

This study has adopted a mixed C-logit model to study the truck route choice problem and introduced Bayesian models for the first time to calibrate the parameters in the utility function. Three Bayesian models are proposed with different hyperparameters, namely, BMCL, HBMCL1, and HBMCL2, and they are applied to the GPS dataset collected in Dallas, Texas. By using the calibrated utility function, truck route choice probabilities may be calculated. A unique feature of this study is that the coefficient of each attribute is a random variable and is assumed to follow a predetermined probability distribution. The dispersion of the random coefficients can thus reflect drivers' preference heterogeneity, which is the primary contribution of this study. The secondary contribution of this study is the application of a clustering method to the map-matching process in preprocessing the trip itinerary data. The major findings and limitations of this study are specified below.

First, the 10-fold cross-validation process is used when comparing the proposed model to other machine learning methods. Even though the block coordinate ascend algorithm of the mean-field variational Bayesian inference procedure cannot guarantee a global maximization of the ELBO, it still provides superior performance in prediction accuracy. The result shows that even though the third model has the most informative prior, it does not always perform best with the GPS dataset used in this study. The reason is that for real datasets, one usually cannot derive a proper guess for the prior. Using a strongly informative prior is thus risky in such a situation since a strongly

informative prior that is not close to the truth would impede the model prediction. Consequently, a complex model does not always result in better performance; the tradeoff between the model complexity or informativity of priors and model prediction needs to be considered for practitioners.

Additionally, the all-random parameter setting of the mixed C-logit model introduced in this paper strengthens the model stability. The preference heterogeneity is a major consideration for this study. According to Hess and Train (2017), the mixed logit models with fully correlated utility coefficients account for scale heterogeneity and other correlation sources. The scale preference refers to the individual preference variation on the relative weight of attributes that are not included in the utility function to attributes that are included. With a random parameter setting, the proposed model can address both scale and preference heterogeneity. We have found that drivers have a consistent preference for travel time, toll price, and the S/IS indicator. In contrast, the taste for truck delay and network density varies greatly among truckers.

Second, among all the attributes that affect truck drivers' perceived route utilities, travel time and road classification are equally crucial in route selection. Generally, in this study, truckers tend to use a path with a short travel time and without local segments. Three of five clusters indicate that travel time is more significant than the existence of local segments. This finding is in line with most of the existing studies, such as Sharma et al. (2019), Feng et al. (2013), and Tsirimpa et al. (2019). The truck delay per lane is correlated to truck AADT. Both positive and negative effects are observed in the results, which shows that the drivers did not display a specific persistent

preference for the number of trucks on the route. The CF, which shows the truckers' preference for network density, has the smallest effect on the perceived utility, compared to other attributes. By comparing the results of Scenario 1 and of Scenario 2, it is concluded that when toll price is included, the importance of the route level of service in perceiving route utility is reduced.

Third, after the proposed models are applied independently of each cluster, the mean VTTS of the drivers varies from $70 per hour to $150 per hour. The freight value of travel time savings according to the Federal Highway Administration range from $25 per hour to $200 per hour, based on the commodity type and other factors (FHWA, 2001). The mean values of travel time savings from this study are generally above the national freight carriers' average level. It is observed from the kernel density of the VTTS that the HBMCL1 provides the most concentrated density, followed closely by the HBMCL2. The kernel density of the VTTS from BMCL displays a large dispersion with heavy tails on both sides of the density.

Last but not least, the implement of the *k*-mean clustering procedure in the map-matching process shows that using the CF as clustering input can effectively part the whole choice set into subsets in which only sub-trip relevant alternatives are included, and therefore it notably reduces the effect of sub-trips of the route choice model. A sub-trip is a section of a trip with intermediate stops for fueling, breaks, or drop-offs/pick-ups. The clustering method can avoid splitting single trips at intermediate destinations and thus reduces the number of O-D pairs, compared to the way that each sub-trip is

treated an independent trip. Eventually, each O-D pair's probabilities can be calculated with a nested structure.

This study has several limitations. First, the drivers' number of choice situations is pre-set to be 50, resulting in the incomplete use of the data. In future studies, the algorithm may be modified to allow for a flexible number of individuals. Second, the starting values of the variational parameters as well as the guess for the priors are arbitrarily determined. As the block coordinate algorithm converges to local maximums, results need to be derived with multiple trials.

Third, the maximum prediction error reported in the model application to the GPS dataset is larger than that of the simulation test, which implies that attributes that are assumed to affect the perceived utilities in this study may not be complete. The availability of data limits the scope of this study. Generally speaking, truck operations are complicated processes, and the route choices of truckers vary with context. State and local regulations, shipping options, commodity types as well as requirements from different stakeholders all contribute to the routing of truckers. Future studies are encouraged for further exploration of truck route choice with comprehensive consideration of the operating context.

REFERENCES

Abbe, E., M. Bierlaire, and T. Toledo. 2007. "Normalization and Correlation of Cross-Nested Logit Models." *Transportation Research Part B: Methodological* 41(7): 795–808.

Apronti, Dick T., Promothes Saha, Milhan Moomen, and Khaled Ksaibati. 2019. "Truck Safety Evaluation on Wyoming Mountain Passes." *Accident Analysis & Prevention* 122: 342–49.

Arentze, Theo, Tao Feng, Harry Timmermans, and Jops Robroeks. 2012. "Context-Dependent Influence of Road Attributes and Pricing Policies on Route Choice Behavior of Truck Drivers: Results of a Conjoint Choice Experiment." *Transportation* 39(6): 1173–88.

Arunotayanun, Kriangkrai, and John W. Polak. 2011. "Taste Heterogeneity and Market Segmentation in Freight Shippers' Mode Choice Behaviour." *Transportation Research Part E: Logistics and Transportation Review* 47(2): 138–48.

Bansal, Prateek et al. 2020. "Bayesian Estimation of Mixed Multinomial Logit Models: Advances and Simulation-Based Evaluations." *Transportation Research Part B: Methodological* 131: 124–42.

Bekhor, Shlomo, Moshe E. Ben-Akiva, and M. Scott Ramming. 2006. "Evaluation of Choice Set Generation Algorithms for Route Choice Models." *Annals of Operations Research* 144(1): 235–47.

Ben-Akiva, Moshe, and Michel Bierlaire. 1999. “Discrete Choice Methods and Their Applications to Short Term Travel Decisions.” In *Handbook of Transportation Science*, International Series in Operations Research & Management Science, ed. Randolph W. Hall. Boston, MA: Springer US, 5–33. http://link.springer.com/10.1007/978-1-4615-5203-1_2 (August 26, 2020).

Blei, David M., Alp Kucukelbir, and Jon D. McAuliffe. 2017. “Variational Inference: A Review for Statisticians.” *Journal of the American Statistical Association* 112(518): 859–77.

Braun, Michael, and Jon McAuliffe. 2010. “Variational Inference for Large-Scale Models of Discrete Choice.” *Journal of the American Statistical Association* 105(489): 324–35.

Broach, Joseph, Jennifer Dill, and John Gliebe. 2012. “Where Do Cyclists Ride? A Route Choice Model Developed with Revealed Preference GPS Data.” *Transportation Research Part A: Policy and Practice* 46(10): 1730–40.

Cascetta, Ennio, Francesco Russo, and Antonino Vitetta. 1997. “Stochastic User Equilibrium Assignment with Explicit Path Enumeration: Comparison of Models and Algorithms.” *IFAC Proceedings Volumes* 30(8): 1031–37.

Chorus, Caspar. 2012. “Random Regret Minimization: An Overview of Model Properties and Empirical Evidence.” *Transport Reviews* 32(1): 75–92.

Chorus, Caspar G. 2014. “A Generalized Random Regret Minimization Model.” *Transportation Research Part B: Methodological* 68: 224–38.

Dalumpines, Ron, and Darren M. Scott. 2017. “Determinants of Route Choice Behavior: A Comparison of Shop versus Work Trips Using the Potential Path Area - Gateway (PPAG) Algorithm and Path-Size Logit.” *Journal of Transport Geography* 59: 59–68.

Depraetere, Nicolas, and Martina Vandebroek. 2017. “A Comparison of Variational Approximations for Fast Inference in Mixed Logit Models.” *Computational Statistics* 32(1): 93–125.

Di, Xuan, and Henry X. Liu. 2016. “Boundedly Rational Route Choice Behavior: A Review of Models and Methodologies.” *Transportation Research Part B: Methodological* 85: 142–79.

Dowling, Richard et al. 2014. *Incorporating Truck Analysis into the Highway Capacity Manual*. Washington, D.C.: Transportation Research Board. https://www.nap.edu/catalog/22311 (November 15, 2020).

Duncan, Lawrence Christopher et al. 2020. “Path Size Logit Route Choice Models: Issues with Current Models, a New Internally Consistent Approach, and Parameter Estimation on a Large-Scale Network with GPS Data.” *Transportation Research Part B: Methodological* 135: 1–40.

EESI. “Fact Sheet - Vehicle Efficiency and Emissions Standards | White Papers | EESI.” 2005. https://www.eesi.org/papers/view/fact-sheet-vehicle-efficiency-and-emissions-standards (November 16, 2020).

Federal Highway Administration (FHWA). 2020. “Highway Statistics Series - Policy” https://www.fhwa.dot.gov/policyinformation/statistics.cfm (November 15, 2020).

Feng, Tao, Theo Arentze, and Harry Timmermans. 2013. “Capturing Preference Heterogeneity of Truck Drivers’ Route Choice Behavior with Context Effects Using a Latent Class Model.”: 15.

Fiebig, Denzil G., Michael P. Keane, Jordan Louviere, and Nada Wasi. 2010. “The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity.” *Marketing Science* 29(3): 393–421.

Frejinger, E., and M. Bierlaire. 2007. “Capturing Correlation with Subnetworks in Route Choice Models.” *Transportation Research Part B: Methodological* 41(3): 363–78.

Greene, William H., and David A. Hensher. 2010. “Does Scale Heterogeneity across Individuals Matter? An Empirical Assessment of Alternative Logit Models.” *Transportation* 37(3): 413–28.

Haghani, Milad, Majid Sarvi, and Zahra Shahhoseini. 2015. “Accommodating Taste Heterogeneity and Desired Substitution Pattern in Exit Choices of Pedestrian Crowd Evacuees Using a Mixed Nested Logit Model.” *Journal of Choice Modelling* 16: 58–68.

Hall, Randolph W., ed. 1999. 23 *Handbook of Transportation Science*. Boston, MA: Springer US. http://link.springer.com/10.1007/978-1-4615-5203-1 (August 26, 2020).

Han, Bing, Shuang Ren, and Jingjing Bao. 2020. “Mixed Logit Model Based on Improved Nonlinear Utility Functions: A Market Shares Solution Method of Different Railway Traffic Modes.” *Sustainability* 12(4): 1406.

Hensher, David A, and William H Greene. 2003. “The Mixed Logit Model: The State of Practice.” *Transportation* 30(2): 133–76.

Hess, Stephane, and Kenneth Train. 2017. “Correlation and Scale in Mixed Logit Models.” *Journal of Choice Modelling* 23: 1–8.

Hess, Stephane, and Kenneth E. Train. 2011. “Recovery of Inter- and Intra-Personal Heterogeneity Using Mixed Logit Models.” *Transportation Research Part B: Methodological* 45(7): 973–90.

Hong, Sung-Pil, Kyung min Kim, Geunyeong Byeon, and Yun-Hong Min. 2017. “A Method to Directly Derive Taste Heterogeneity of Travellers’ Route Choice in Public Transport from Observed Routes.” *Transportation Research Part B: Methodological* 95: 41–52.

Huang, Alan, and M. P. Wand. 2013. “Simple Marginally Noninformative Prior Distributions for Covariance Matrices.” *Bayesian Analysis* 8(2): 439–52.

Jacob, Bernard, and Véronique Feypell-de La Beaumelle. 2010. “Improving Truck Safety: Potential of Weigh-in-Motion Technology.” *IATSS Research* 34(1): 9–15.

Jiang, Ying, and Junyi Zhang. 2019. “Interaction between Company Manager’s and Driver’s Decisions on Expressway Routes for Truck Transport.” *Transport Policy* 76: 1–12.

Knorring, John H, Rong He, and Alain L Kornhauser. 1923. “Analysis of Route Choice Decisions by Long-Haul Truck Drivers.” *Transportation Research Record*: 15.

Krueger, Rico et al. 2020. “Variational Bayesian Inference for Mixed Logit Models with Unobserved Inter- and Intra-Individual Heterogeneity.” *arXiv:1905.00419 [econ, stat]*. http://arxiv.org/abs/1905.00419 (August 26, 2020).

Lai, Xinjun, and Michel Bierlaire. 2015. “Specification of the Cross-Nested Logit Model with Sampling of Alternatives for Route Choice Models.” *Transportation Research Part B: Methodological* 80: 220–34.

Li, Meng, and Hai-Jun Huang. 2017. “A Regret Theory-Based Route Choice Model.” *Transportmetrica A: Transport Science* 13(3): 250–72.

Lue, Gregory, and Eric J. Miller. 2019. “Estimating a Toronto Pedestrian Route Choice Model Using Smartphone GPS Data.” *Travel Behaviour and Society* 14: 34–42.

Luong, Trang D., Divyakant Tahlyan, and Abdul R. Pinjari. 2018. “Comprehensive Exploratory Analysis of Truck Route Choice Diversity in Florida.” *Transportation Research Record: Journal of the Transportation Research Board* 2672(9): 152–63.

Madadi, Bahman, Rob van Nes, Maaike Snelder, and Bart van Arem. 2019. “Assessing the Travel Impacts of Subnetworks for Automated Driving: An Exploratory Study.” *Case Studies on Transport Policy* 7(1): 48–56.

Magidson, Jay, Jeroen K. Vermunt, and John P. Madura. 2004. *Latent Class Analysis*. 1 Oliver’s Yard, 55 City Road, London EC1Y 1SP United Kingdom: The Sage handbook of quantitative methodology for the social sciences. https://methods.sagepub.com/foundations (August 26, 2020).

Manski, Charles F. 1977. “The Structure of Random Utility Models.” *Theory and Decision* 8(3): 229–54.

Masiero, Lorenzo, Yang Yang, and Richard T.R. Qiu. 2019. “Understanding Hotel Location Preference of Customers: Comparing Random Utility and Random Regret Decision Rules.” *Tourism Management* 73: 83–93.

NYC DOT "Trucks and Commercial Vehicles.” https://www1.nyc.gov/html/dot/html/motorist/trucks.shtml (November 15, 2020).

Oka, Hideki et al. 2019. “Predicting Travel Pattern Changes of Freight Trucks in the Tokyo Metropolitan Area Based on the Latest Large-Scale Urban Freight Survey and Route Choice Modeling.” *Transportation Research Part E: Logistics and Transportation Review* 129: 305–24.

Papola, Andrea. 2004. “Some Developments on the Cross-Nested Logit Model.” *Transportation Research Part B: Methodological* 38(9): 833–51.

Poulopoulou, Maria, Ioanna Spyropoulou, and Constantinos Antoniou. 2015. “Parameters Affecting Professional Driver Response to VMS Messages.” In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, Gran Canaria, Spain: IEEE, 1563–68. http://ieeexplore.ieee.org/document/7313347/ (August 26, 2020).

Prashker, Joseph N., and Shlomo Bekhor. 2004. “Route Choice Models Used in the Stochastic User Equilibrium Problem: A Review.” *Transport Reviews* 24(4): 437–63.

Ramming, MS. 2001. “Network Knowledge and Route Choice.” *Unpublished Ph. D. Thesis, Massachusetts Institute of Technology*.

Rowell, Maura, Andrea Gagliano, and Anne Goodchild. 2014. “Identifying Truck Route Choice Priorities: The Implications for Travel Models.” *Transportation Letters* 6(2): 98–106.

Russo, Francesco, and Antonino Vitetta. 2003. “An Assignment Model with Modified Logit, Which Obviates Enumeration and Overlapping Problems.” *Transportation* 30(2): 177–201.

San Francisco Municipal Transportation Agency (San Francisco MTA, Transit, Streets, Taxi) | SFMTA. https://www.sfmta.com/ (November 15, 2020).

Sathaye, Nakul, Arpad Horvath, and Samer Madanat. 2010. “Unintended Impacts of Increased Truck Loads on Pavement Supply-Chain Emissions.” *Transportation Research Part A: Policy and Practice* 44(1): 1–15.

Schrank, David, Bill Eisele, and Tim Lomax. 2019. “2019 Urban Mobility Report.” : 50.

Schreckenberg, Michael, and Reinhard Selten, eds. 2004. *Human Behaviour and Traffic Networks*. Berlin, Heidelberg: Springer Berlin Heidelberg. http://link.springer.com/10.1007/978-3-662-07809-9 (November 9, 2020).

Shahhoseini, Zahra, Milad Haghani, and Majid Sarvi. 2015. “Estimation and Application of a Multi-Class Multi-Criteria Mixed Paired Combinatorial Logit Model for Transport Networks Analysis.” *Transportmetrica B: Transport Dynamics* 3(1): 59–78.

Sharma, Salil, Maaike Snelder, and Hans van Lint. 2019. “Deriving On-Trip Route Choices of Truck Drivers by Utilizing Bluetooth Data, Loop Detector Data and Variable Message Sign Data.” In *2019 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, Cracow, Poland: IEEE, 1–8. https://ieeexplore.ieee.org/document/8883311/ (August 26, 2020).

Tan, Linda S. L. 2017. “Stochastic Variational Inference for Large-Scale Discrete Choice Models Using Adaptive Batch Sizes.” *Statistics and Computing* 27(1): 237–57.

Tang, Jinjun et al. 2020. “A Mixed Path Size Logit-Based Taxi Customer-Search Model Considering Spatio-Temporal Factors in Route Choice.” *IEEE Transactions on Intelligent Transportation Systems* 21(4): 1347–58.

Toledo, Tomer et al. 2013. “Decision-Making Process and Factors Affecting Truck Routing.” In *Freight Transport Modelling*, eds. Moshe Ben-Akiva, Hilde Meersman, and Eddy Van de Voorde. Emerald Group Publishing Limited, 233–49. https://www.emerald.com/insight/content/doi/10.1108/9781781902868-012/full/html (August 26, 2020).

Train, Kenneth E. 2003. *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press. http://ebooks.cambridge.org/ref/id/CBO9780511753930 (August 26, 2020).

Tsirimpa, Athena, Amalia Polydoropoulou, and Constantinos Antoniou. 2007. “Development of a Mixed Multi-Nomial Logit Model to Capture the Impact of

Information Systems on Travelers' Switching Behavior." *Journal of Intelligent Transportation Systems* 11(2): 79–89.

U.S. Department of Transportation. "Freight Analysis Framework - FHWA Freight Management and Operations." (USDOT FAF4). 2019. https://ops.fhwa.dot.gov/freight/freight_analysis/faf/ (November 15, 2020).

U.S. Department of Transportation. (USDOT). 2007a. "Freight and Congestion - FHWA Freight Management and Operations." https://ops.fhwa.dot.gov/freight/freight_analysis/freight_story/congestion.htm (November 16, 2020).

U.S. Department of Transportation (USDOT). 2007b. "Large and Growing Demand for Freight Transportation - FHWA Freight Management and Operations." https://ops.fhwa.dot.gov/freight/freight_analysis/freight_story/large.htm (November 15, 2020).

U.S. Department of Transportation, Bureau of Transportation Statistics (USDOT BTS). 2019a. "Moving Goods in the United States." https://data.bts.gov/stories/s/Moving-Goods-in-the-United-States/bcyt-rqmu (November 16, 2020).

U.S. Department of Transportation, Bureau of Transportation Statistics (USDOT BTS). 2019b. "Freight Transportation & the Economy." https://data.bts.gov/stories/s/Freight-Transportation-the-Economy/6ix2-c8dn (November 15, 2020).

Wang, Xiaokun (Cara), and Dapeng Zhang. 2017. “Truck Freight Demand Elasticity with Respect to Tolls in New York State.” *Transportation Research Part A: Policy and Practice* 101: 51–60.

Wang, Zun, Anne Goodchild, and Edward McCormack. 2016. “Measuring Truck Travel Time Reliability Using Truck Probe GPS Data.” *Journal of Intelligent Transportation Systems* 20(2): 103–12.

Yang, Yang, Enjian Yao, Zhiqiang Yang, and Rui Zhang. 2016. “Modeling the Charging and Route Choice Behavior of BEV Drivers.” *Transportation Research Part C: Emerging Technologies* 65: 190–204.

Yang, Zhao, Yuanyuan Zhang, and Offer Grembek. 2016. “Combining Traffic Efficiency and Traffic Safety in Countermeasure Selection to Improve Pedestrian Safety at Two-Way Stop Controlled Intersections.” *Transportation Research Part A: Policy and Practice* 91: 286–301.

Zhang, Cheng, Judith Butepage, Hedvig Kjellstrom, and Stephan Mandt. 2019. “Advances in Variational Inference.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41(8): 2008–26.

Zhang, Junlin, and Hai Yang. 2015. “Modeling Route Choice Inertia in Network Equilibrium with Heterogeneous Prevailing Choice Sets.” *Transportation Research Part C: Emerging Technologies* 57: 42–54.

Zhou, Zhong, Anthony Chen, and Shlomo Bekhor. 2012. “C-Logit Stochastic User Equilibrium Model: Formulations and Solution Algorithm.” *Transportmetrica* 8(1): 17–41.

Zimmermann, Maëlle, Oskar Blom Västberg, Emma Frejinger, and Anders Karlström. 2018. “Capturing Correlation with a Mixed Recursive Logit Model for Activity-Travel Scheduling.” *Transportation Research Part C: Emerging Technologies* 93: 273–91.

APPENDIX A

CROSS VALIDATION ERROR

Table A-1 Error of cross-validation of Scenario 1.

Model	Scenario1					
	Cluster0		Cluster1		Cluster2	
	mean	variance	mean	variance	mean	variance
C-logit	0.9095	0.0028	0.8298	0.0066	0.8179	0.0003
Lasso	0.9095	0.0028	0.8298	0.0066	0.8179	0.0003
LDA	0.9010	0.0040	0.8844	0.0039	0.7777	0.0289
KNN	0.9298	0.0024	0.9152	0.0082	0.8642	0.0142
CART	0.9149	0.0054	0.8837	0.0075	0.8374	0.0144
NB	0.9316	0.0025	0.9339	0.0055	0.9178	0.0101
SVM	0.9310	0.0024	0.9308	0.0049	0.9204	0.0131
RF	0.9052	0.0014	0.8552	0.0085	0.8222	0.0302
MLP	0.9239	0.0033	0.9176	0.0064	0.7561	0.1039
BMCL	0.9196	0.0055	0.9557	0.0081	0.9474	0.0169
HBMCL1	0.9380	0.0026	0.9609	0.0103	0.9794	0.0126
HBMCL2	0.9306	0.0345	0.9480	0.0409	0.9776	0.0124

Table A-2 Error of cross-validation of Scenario 2.

Model	Scenario2					
	Cluster0		Cluster1		Cluster2	
	mean	variance	mean	variance	mean	variance
C-logit	0.8070	0.0065	0.9182	0.0041	0.8793	0.0053
Lasso	0.8070	0.0065	0.9182	0.0041	0.8793	0.0053
LDA	0.7416	0.0257	0.8852	0.0065	0.7300	0.0195
KNN	0.8122	0.0072	0.9155	0.0049	0.8858	0.0073
CART	0.7527	0.0078	0.8927	0.0070	0.8664	0.0053
NB	0.8327	0.0071	0.9290	0.0068	0.8966	0.0057
SVM	0.8302	0.0068	0.9255	0.0062	0.8973	0.0061
RF	0.8094	0.0019	0.8841	0.0011	0.7854	0.0096
MLP	0.7489	0.0020	0.9203	0.0070	0.8874	0.0064
BMCL	0.9056	0.0198	0.9582	0.0156	0.9420	0.0285
HBMCL1	0.9154	0.0151	0.9634	0.0191	0.9442	0.0403
HBMCL2	0.9121	0.0107	0.9691	0.0183	0.9444	0.0459

APPENDIX B

PREDICTED PROBABILITIES

Table B-1 Predicted probabilities for Scenario 1.

Path	True probability	Predicted probability			
Cluster0		BMCL	HBMCL1	HBMCL2	C-logit
28	0.0024	0.0073	0.0024	0.0073	0.0000
42	0.0201	0.0404	0.0245	0.0257	0.0332
60	0.5479	0.4892	0.5100	0.5039	0.4856
86	0.0137	0.0196	0.0257	0.0147	0.0488
106	0.0051	0.0122	0.0061	0.0049	0.0019
120	0.0001	0.0000	0.0000	0.0000	0.0000
134	0.0001	0.0000	0.0000	0.0000	0.0000
192	0.0016	0.0000	0.0024	0.0024	0.0000
198	0.0068	0.0159	0.0159	0.0122	0.0365
263	0.0056	0.0171	0.0086	0.0073	0.0033
282	0.0022	0.0049	0.0086	0.0098	0.0000
289	0.0059	0.0049	0.0073	0.0232	0.0021

Table B-1 Continued.

Cluster1					
3	0.0069	0.0092	0.0048	0.0055	0.0225
43	0.0055	0.0067	0.0078	0.0078	0.0058
64	0.0272	0.0313	0.0302	0.0308	0.0422
107	0.0007	0.0000	0.0006	0.0000	0.0000
117	0.0040	0.0022	0.0062	0.0050	0.0027
119	0.0878	0.0805	0.0814	0.0801	0.0653
184	0.0035	0.0081	0.0034	0.0050	0.0000
185	0.0043	0.0017	0.0056	0.0056	0.0013
Cluster2					
8	0.0398	0.0370	0.0403	0.0395	0.0141
11	0.0561	0.0504	0.0529	0.0521	0.0865
12	0.0505	0.0487	0.0495	0.0529	0.0027
69	0.0636	0.0739	0.0672	0.0655	0.1067
Cluster3					
222	0.0042	0.0042			
Clurste4					
20	0.0345	0.0345			

Table B-2 Predicted probabilities for Scenario 2.

Path	True probability	Predicted probability			
Cluster0		BMCL	HBMCL1	HBMCL2	C-logit
2	0.0096	0.0120	0.0112	0.0112	0.0018
12	0.0089	0.0130	0.0116	0.0116	0.0105
16	0.0167	0.0193	0.0200	0.0186	0.0346
17	0.0170	0.0246	0.0246	0.0267	0.0410
24	0.1235	0.1068	0.1083	0.1076	0.0878
Cluster1					
0	0.0126	0.0149	0.0158	0.0120	0.0183
5	0.0006	0.0011	0.0032	0.0029	0.0000
6	0.0312	0.0282	0.0309	0.0260	0.0593
7	0.0158	0.0275	0.0201	0.0275	0.0128
9	0.2264	0.2148	0.2166	0.2181	0.1961
Cluster2					
19	0.1136	0.0710	0.1118	0.1226	0.2201
31	0.0568	0.0258	0.0484	0.0430	0.0324
33	0.0422	0.0280	0.0355	0.0409	0.0096
37	0.3252	0.4130	0.3420	0.3312	0.3025