On Ride-Sharing: A Departure Time Choice Analysis with Latent Carpooling Preference

Chenfeng Xiong*, Pratt Hetrakul†, and Lei Zhang‡

December 19, 2014

Abstract

This paper presents a departure time choice analysis, based on the notion of a latent carpooling preference. The study is based on combined revealed preference and stated preference survey data collected on the Maryland side of the Capital Beltway (I-495). A conditional logit model is first estimated to identify drivers’ choice of departure time when tolls and congestion management strategies, including high-occupancy vehicle (HOV) lanes and high-occupancy toll (HOT) lanes, are implemented. Then a latent class model accounting for heterogeneity across categories of drivers is proposed to examine difference in behavioral preferences across classes. The latent class model result reveals significant heterogeneity in drivers’ latent preference toward ride-sharing, which can potentially support ranges of transportation policy and incentive design related to congestion management strategies such as HOV/HOT lane usage.

1 Introduction

Having way too many people traveling during morning and afternoon peak hours and to the same destination, the congestion level on D.C.’s Capital Beltway is far from satisfactory. Multiple congestion management initiatives, including dynamic tolling and managed lanes such as high-occupancy vehicle (HOV) and high-occupancy toll (HOT) lanes, have received increasing attention recently, due to their efficiency in altering travelers’ departure time and route choice [Arnott et al., 1990]. Time varying charges stimulate toll road users to make departure time decision so as to reduce their travel cost. Vehicles on HOV/HOT lanes can move at a more enjoyable speed, even when parallel lanes suffer delays from queueing. This convenience is often traded off for extra charges on HOT lanes and additional time loss and inconvenience for ride-sharing on HOV lanes [Burris et al., 2007; Horowitz and Sheth, 1977; Li et al., 2007]. The consideration of departure time choice

*Graduate Research Assistant, Dept. of Civil Eng., Univ. of Maryland, 1173 Glenn Martin Hall Bldg., College Park, MD 20742. E-mail: cxiong@umd.edu.
†Graduate Research Assistant, E-mail: hetrakul@umd.edu.
‡Associate Professor, E-mail: lei@umd.edu.
behavior is therefore essential for modeling and understanding the impact of these congestion management strategies. The new toll road facility and the HOV/HOT lanes built on freeway corridors in Maryland have further emphasized the need for an appropriate departure time model that realistically captures drivers’ response to time-varying toll charges and their inherent preference toward ride-sharing.

Many publications (Jou et al., 1997; Kroes et al., 1996; Mahmassani and Liu, 1999) emphasize the importance of departure time choice as a potential response to congestion while a number of studies incorporated departure time choice in various micro-simulation models or other assignment models (Hu and Mahmassani, 1997; Rossetti and Liu, 2005; Van Berkum and Van Der Mede, 1993). However, these studies do not contain a number of aspects which we believe are crucial for modeling the response to congestion management policy appropriately.

First of all, user’s responses to travel conditions, including the shift in departure time and the change in travel routes, are generally governed by individual trip makers’ preferences on travel time and the cost they perceive toward the earliness/lateness of their preferred schedule. These behavioral characteristics vary significantly across users (Brownstone and Small, 2005). Capturing the heterogeneity of users in this regard is critical in predicting the impact of dynamic pricing schemes (Lu et al., 2008). However, limited number of studies have considered the heterogeneity of users in the underlying departure time and path choice decision framework (Ran and Boyce, 1996; Ran et al., 1996?). Secondly, most of the departure time choice models only consider travel time and schedule delay variables, without the capability to represent drivers’ responses to pricing, managed lanes, and other congestion management strategies.

In practice, there is imperative need for incorporating departure time choices into both microscopic traffic operations analysis and micro/macro-level travel demand studies. Nowadays, in many real-world applications the capability of analyzing peak-period pricing, HOV/HOT lanes, traveler information systems, and operational improvements becomes readily available. For instance, a recent study (Lee et al., 2009) evaluated HOV/HOT scenarios by integrating travelers’ departure time response to value pricing in a real-world TRANSIMS simulation model. To support the growing number of this type of applications, it is important to have a suitable departure time choice model which is developed based on carefully designed survey and is of decent model sensitivity. And potentially it would be helpful to combine the survey data with real-world data on the use of HOV/HOT facilities. These efforts are particularly necessary for the analysis of congestion management strategies which often induce departure time shifts and peak spreading (i.e. trips diverting from peak to off-peak, and from peak-peak to shoulder-peak period). In total, the heterogeneity issue and the response to pricing and other congestion management strategies, inevitably need to be better addressed in modeling departure time choice.

2 Literature Review

Vickrey (1969) examined a single bottleneck and derived supply-demand equilibrium conditions with departure time considerations. Inspired by Vickrey’s seminal contribution, several researchers, including Arnott et al. (1990), De Palma et al. (1997), and Van Vuren et al. (1999), have developed continuous departure time modeling frameworks, collectively referred to as the “Equilibrium Scheduling Theory” (EST). Later, EST was integrated with route choice analysis (Chen and Chang, 2000; Mahmassani et al., 1997; Mannering, 1989). Bates (1996) employed EST for network-level
transportation models and for dynamic traffic assignment. The HADES (i.e. Heterogeneous Arrival and Departure times based on EST) model represents the latest achievement of the EST-line of research.

Another line of research focuses on discrete choice modeling. Small (1982) adopted the multinomial logit (MNL) approach to model departure time decision-making. Although being the most widely used method, MNL approach has its well-known limitations. Firstly, the underlying assumption of independence from irrelevant alternatives (IIA) may not hold for departure time choice analysis because adjacent departure time options tend to exhibit correlated unobservable factors. In other words, if a decision-maker prefers departure interval $t$, the same decision-maker probably also prefers departure time intervals $t-1$ and $t+1$ to other intervals due to some unobservable factors (e.g. dropping off/picking up a child as part of the trip, unknown scheduling constraints). Nested logit (NL) models have been used to identify the correlated departure time intervals (Bowman et al., 1998; Polak and Jones, 1993; Yun et al., 2000). Small (1987) and Bhat (1998b) both tested the NL and a more general ordered generalized extreme value (OGEV) model, which allows for a correlation parameter, for a pair of alternatives, depending on the distance between these alternatives along some natural ordering such as the clock time in departure time choice. And they concluded that both the NL and the OGEV models performed better than MNL. More general model specifications include paired combinatorial logit (Koppelman and Wen, 2000), cross-nested logit model (Ben-Akiva and Bierlaire, 1999; Papola, 2004; Vovsha, 1997; Wen and Koppelman, 2001), and continuous cross-nested logit models (Lemp et al., 2010).

Another limitation of MNL specifications is its inability to capture taste variation. Mixed logit (ML) models have been widely employed to address this limitation (Bhat, 1998a; De Jong et al., 2003). ML is highly flexible that it generalizes a standard MNL by allowing its parameter(s) to vary with a known population distribution across individual. And they can approximate any random utility models (McFadden and Train, 2000). Börjesson (2008) developed a mixed logit model considering three alternatives (two departure time alternatives for autos and one for public transit) using stated preference (SP) data. She has tested different specifications on parameter distributions, as well as on drivers’ preferred departure/arrival schedules. One weakness of ML, as they are often criticized for, is that it requires the analyst to make specific assumptions about the distributions of parameters across individuals (Greene and Hensher, 2003; Shen et al., 2006).

Compared to ML, a relatively simple and reasonably plausible model specification to consider individual heterogeneity is the latent class (LC) model, which is often employed to model consumers’ latent preference in marketing research (e.g. Swait and Adamowicz, 2001). Unlike the ML which assumes specific distributions of parameters, LC captures unobserved preference heterogeneity by assuming that dividing the population into a discrete number of classes can sufficiently represent the taste variation (Shen et al., 2006). Greene and Hensher (2003) and Shen et al. (2006) compared LC and ML and promoted the merits offered by the LC models. Very recently, Ben-Elia et al. (2010) applied LC based on the notion of a latent preferred arrival time to study the reward effect on peak-hour avoidance. Based on the LC specification, they drew interesting conclusions regarding the latent heterogeneity in preferred arrival schedule.

One specific purpose of this study is to investigate heterogeneous ride-sharing preferences among drivers. A latent class model is used to identify the number of classes, the class membership function, and the different attitudes to time-varying toll and HOV variables of different class members. The remainder of the paper is organized as follows: the carefully designed survey data collection for this study is described in Section 3. The survey was designed to investigate trav-
lers’ departure time according to the hypothetical time-of-day congestion pricing scheme on the Capital Beltway (I-495). Based on the review findings in Section 2, the latent class methodology is selected and the departure time choice model estimation is presented in Section 4. The conclusions and discussions of this paper are presented in Section 5.

3 Behavioral Survey

To initiate the development of the departure time choice model, Maryland State Highway Administration (SHA) and University of Maryland conducted a web-based joint stated preference (SP) and revealed preference (RP) departure time, mode/lane choice survey. It consisted of two waves of data collection collecting a total number of 151 effective samples. The first-wave survey was conducted on March 21-25, 2011. And the second one was conducted on May 23-27, 2011.

The survey questionnaire consisted of two parts. The first part was a list of RP questions about the current travel behavior of respondents and their socio-economic status. The second part contained an SP experiment, which used the answers to the RP questions to determine the values of attributes.

3.1 Sampling and Data Collection

The survey used random sampling method. The population from which the respondents were recruited consisted of car drivers traveling on the Capital Beltway in the Washington D.C. metropolitan area during weekday morning and afternoon peak period (6:00 a.m.-10:00 a.m., and 3:00 p.m.-7:00 p.m.). Drivers were first intercepted during their journeys when they were waiting at traffic lights on several off-ramps of the Capital Beltway. Flyers designed for this study were distributed to the drivers. And the link to the web-based survey was provided in the flyers. In total, around 4,000 drivers have been reached during the data collection period. A total number of 304 drivers responded the survey questionnaire, which resulted in the overall response rate of 7.6%. Within the 304 responded surveys, 151 of the respondents completed the survey.

Figure 1 compared the distributions of age and income of the final sample with those of the 2010 U.S. Census for Montgomery County, Maryland. In general, the sample was representative of all drivers in the study area in terms of socio-economic characteristics (including gender, age, and ethnicity) with slight oversampling of higher-income drivers (typical for web-based surveys).

The driving characteristics also resembled the area population well. Figure 2 compared departure time choice distribution of the final sample with the departure time distribution of the commuting trips collected by Transportation Planning Board (TPB)/Baltimore Metropolitan Council (BMC) Household Travel Survey. The sample was representative with slightly higher peak period samples. Other descriptive statistics are presented in Table 1. The average reported vehicle travel time was 30.22 minutes (23.96 minutes in the TPB/BMC survey). And the percentage of current carpoolers in the survey sample was 19% (the percentage of carpoolers got from the TPB/BMC survey was 18.8%).
Figure 1: Comparison of the Distributions of Age and Income

### 3.2 Revealed Preference

The RP questionnaire consisted of two sections: (1) respondents’ socioeconomics, and (2) information about the most recent trip. In the first section, drivers’ gender, age, household income range, education, occupation,
number of workers in the household, number of vehicles in the household, vehicle type, vehicle age, and work place ZIP code information have been collected.

The second section gathered data about the drivers’ most recent trip on the Beltway. The purpose of this section was to use their experienced trip condition as the pivot point when designing the stated preference (SP) questions. This ensured that the stated scenario in the SP part were realistic for each individual in the survey. Each respondent was asked to recall and describe his/her most recent trip information on the Beltway via the following constructs:

- General information (mode, number of passengers, trip purpose, work starting/ending time, and schedule flexibility);
- Trip time (departure time, arrival time);
- Preferred schedule (preferred departure time, and preferred arrival time);
- Trip cost (fuel cost, toll cost, park cost);
- Travel uncertainty information (experienced shortest and longest total travel time, the shortest and longest time spent on the Beltway).

### 3.3 Stated Preference

The second part of the survey was based on stated preference design, wherein the respondents were asked to make choices in seven hypothetical choice situations. The experiment was designed to explore the trade-off that drivers make between shifts in their actual departure time, travel time, fuel/toll costs, and lane usage.
Table 1: Descriptive Statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>gender</td>
<td>Dummy= 1 if the respondent is male</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>age</td>
<td>The respondent’s age</td>
<td>43.85</td>
<td>13.26</td>
</tr>
<tr>
<td>workers</td>
<td>Number of workers in the household</td>
<td>1.76</td>
<td>0.64</td>
</tr>
<tr>
<td>cars</td>
<td>number of vehicles in the household</td>
<td>1.98</td>
<td>0.87</td>
</tr>
<tr>
<td>income</td>
<td>Household income</td>
<td>$50K - $100K</td>
<td>-</td>
</tr>
</tbody>
</table>

Most Recent Trip Statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>travel time</td>
<td>Travel time (minute)</td>
<td>30.22</td>
<td>22.28</td>
</tr>
<tr>
<td>fuel cost</td>
<td>Fuel cost ($)</td>
<td>6.24</td>
<td>10.63</td>
</tr>
<tr>
<td>distance</td>
<td>Travel distance (mile)</td>
<td>28.56</td>
<td>57.38</td>
</tr>
<tr>
<td>carpool</td>
<td>Dummy= 1 if the respondent is carpooling</td>
<td>0.19</td>
<td>0.40</td>
</tr>
</tbody>
</table>

Respondents’ actual travel times were very close to their preferred departure time (statistical hypothesis has been tested), thus it was assumed that travelers were already traveling at a sub-optimal time-of-day (only sub-optimal given the congestion level in D.C. area). In this sense, the SP experiment did capture respondents’ decision taste on the preferred schedule. The questionnaire first listed the information about the most recent commuting trip observed from RP to remind respondents about the circumstances of the trip. Then unique SP alternatives were designed for each respondent based on her/his reported commuting trip attributes. Each choice situation included three departure time alternatives with different travel delay, schedule delay, and monetary costs: (1) solo driver on normal lane, (2) high occupancy toll (HOT) lane, and (3) high occupancy vehicle (HOV) lane. Figure 3 shows the interface of the departure time choice on the website.

As illustrated in Figure 3, each alternative has been specified with five variables, each of which had up to five levels of variation per alternative. The variables included in the departure time choice experiment includes: (1) Departure time, (2) Travel time range, (3) Arrival time range, (4) Fuel cost, and (5) Toll. These variables have been designed to account for traffic conditions by time of day by taking into account the observed respondents’ departure time (Crunkleton 2008). The description of the variables used in the SP scenarios is as follows:

- Departure time: Departure time is pivoted from respondent’s reported departure time in the RP.
- Total travel time range: This variable is designed to account for both time-of-day conditions based on the respondent’s reported departure time and travel condition on toll lane. It is aimed at capturing travel time uncertainty. This information is not used in the current study. We leave it in future work following this paper.
- Arrival time range: This variable is calculated corresponded to the departure time and travel time range of the scenario provided to the respondent.
- Fuel cost: The fuel cost is designed to reflect higher expenses in the peak period and on the normal lane. The fuel cost is pivoted from the reported fuel cost in the RP part.
• Toll cost: The toll cost is designed as a mileage based using the Inter-County Connector toll rates as a reference. The toll rate for the HOT lane accounts varies depending on whether the respondents’ reported departure time is in the peak or non-peak period.

The survey was designed with orthogonal design approach where numerical evaluations in a wide range of parameter values were undertaken to guarantee sufficient efficiency of the design. Similar design can be found in other studies (e.g. Bhat and Sardesai [2006] Börjesson [2008] among others). A pilot study, in combination with expert judgments, was used to arrive at the final levels of attribute in the SP experiment.

4 Model Formulation

This section develops discrete choice models investigating the impact of various congestion management policies such as tolling and managed lanes. We first estimate a conditional logit model to identify travelers’ choice. And then a new modeling framework assuming latent class (LC) in carpooling preference has been proposed and estimated.

4.1 Conditional Logit Model

One of the most widely used models when the number of possible choices is large is the conditional logit form: \( V_{ij} = \beta'X_{ij} + \varepsilon_{ij} \) which arises when each \( \varepsilon_{ij} \) is an independent and identically distributed (iid) draw from the type I extreme value distribution with scale parameter \( \mu \). The probability (Equation 1) that individual \( i \) prefers alternative \( j \) takes the well known form (McFadden [1974]):

\[
P_{ij} = \frac{\exp(\beta'x_{ij})}{\sum_k \exp(\beta'x_{ik})}
\]  

(1)

The conditional logit model embodies the IIA property (which means that the odds ratio for any two alternatives is unaffected by the inclusion of any third alternative). And individual’s taste variance is not very
well represented since in the random utility function, each explanatory variable has the same marginal utility for each individual. Albeit this method has these two well-known limitations, it has been estimated as a benchmark to explore the dataset and reveal the econometric nature. The following variables are considered in the utility function:

- Travel time: mean travel time, as provided by the SP survey;
- Fuel cost: in US dollars;
- Toll cost: in US dollars;
- Schedule delay early;
- Schedule delay late;
- HOV usage: a dummy variable which equals one if the alternative is associated with carpooling;
- Income*travel time: interaction variable for high income level (> $150K) and travel time;
- Age*fuel cost: interaction variable for age (> 45) and fuel cost.

The schedule delay early (SDE) and schedule delay late (SDL) variables measure the deviance of individuals’ actual departure time (DT) from their reported preferred departure time (PDT) schedule. These two variables are defined by the following equations:

\[
SDE = \max (0, PDT - DT) \tag{2}
\]

\[
SDL = \max (0, DT - PDT) \tag{3}
\]

Other specifications of SDE and SDL based on preferred arrival time (PAT) and actual arrival time (AT) have also been tested by the authors. The estimation results of these specifications turned out to be insignificant. It suggests that in this study, temporal constraints at the origin seem to primarily restrict early and late departures rather than early or late arrivals. The estimation results are listed in Table 2. The coefficients of the attributes of the choice model are all significant and with the correct sign.

The first five variables are impedances, thus having negative signs as expected. This model shows the different levels of impact resulted from travel time, fuel cost, toll cost, and schedule delays. HOV usage has a negative coefficient, which agrees with the intuition that ride-sharing is generally considered as a discomfort. The rest two variables show different effects of level-of-service on different socioeconomic groups. The coefficient of the interaction of high income and travel time is -0.017, which indicates that travel time has an extra penalty on higher-income group of people. The coefficient of the interaction of age and fuel cost shows that when choosing departure time, fuel cost is more influential to younger drivers. These socio-demographical indications are considered when we construct the latent class model and specify the class membership model in the following subsection.

### 4.2 Latent Class Model

This subsection aims at improving the conditional logit model using latent class specifications. This method identifies heterogeneity without assuming a feasible distribution for certain parameters which may require considerable testing work for different types of density functions. Instead, latent
Table 2: Estimation Results for the Conditional Logit Model

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficients</th>
<th>Std. Err.</th>
<th>t-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Travel time (min.)</td>
<td>-0.022</td>
<td>0.009</td>
<td>-3.389</td>
</tr>
<tr>
<td>Fuel cost ($)</td>
<td>-0.339</td>
<td>0.147</td>
<td>-4.371</td>
</tr>
<tr>
<td>Toll cost ($)</td>
<td>-0.403</td>
<td>0.066</td>
<td>-11.204</td>
</tr>
<tr>
<td>Schedule delay early (min.)</td>
<td>-0.007</td>
<td>0.005</td>
<td>-2.241</td>
</tr>
<tr>
<td>Schedule delay late (min)</td>
<td>-0.006</td>
<td>0.003</td>
<td>-2.562</td>
</tr>
<tr>
<td>HOV usage</td>
<td>-2.200</td>
<td>0.191</td>
<td>-16.319</td>
</tr>
<tr>
<td>Income*travel time</td>
<td>-0.017</td>
<td>0.019</td>
<td>-2.421</td>
</tr>
<tr>
<td>Age*fuel cost</td>
<td>0.395</td>
<td>0.174</td>
<td>4.240</td>
</tr>
</tbody>
</table>

# of Obs. 1,457
Log Likelihood -707.01
Pseudo $R^2$ 0.300

class model assumes that there are certain number of latent classes among individuals and uses a discrete number of segments to describe the density function of the parameters.

In this study, it is assumed that drivers have different latent attitudes toward ride-sharing. Intuitively, those who are already sharing their ride with others are more likely to use HOV lanes rather than paying the toll or using normally more congested general purpose lanes. Many respondents in the survey (about 19%) reported that they have already been carpooling with passengers. Therefore, these drivers were likely to belong to a certain group who relatively favored ride-sharing more than others. This preference is arguably latent since we observed from the SP survey data that a number of respondents who originally drove alone have chosen alternative departure times with HOV lane usage. These drivers may also belong to the aforementioned latent class. Although they did not indicate explicitly that they were carpoolers, these drivers may be inherently more flexible to ride-sharing. And in practice, certain thresholds (e.g. in travel time, cost, etc.) may exist in encouraging these potential managed-lane users to share ride and/or use HOV/HOT lanes.

Therefore, the authors investigate a latent class model in order to model and understand the latent carpooling preference. A class membership function is specified and estimated to examine the significance of different socio-demographic attributes in determining drivers’ class membership. Latent class model recognizes latent heterogeneity by classifying individuals into several groups. And thus the taste variation is realized among travelers belonging to different latent classes. Thus, the choice probability that individual traveler $i$ of class $c$ prefers alternative $j$ from a particular set $k$ is defined by the Equation 4:

$$P_{ij|c} = \frac{\exp(\beta_j^c x_{ij})}{\sum_k \exp(\beta_k^c x_{ik})}$$

(4)

The class membership model is constructed as a multinomial logit choice model. LC model simultaneously estimates Equation 4 for the classes and predicts the membership probability $Q_{ic}$ as individual $i$ in class $c$. Then Equation 5 gives the unconditional probability of individual $i$ choosing
After testing variables including age, gender, income, and education, the following variables have been selected to build the class membership model, which is based on the previous conditional logit work and after corrections of trial and error estimation and clearing out of non-significant coefficients.

- **Age**: dummy variable for the age (= 1 if age > 45 which is the mean value of the sample respondents); and

- **Carpool**: dummy variable for carpooling (equals 1 if the driver in the RP is recorded having at least one passenger).

Then this study needs to decide the number of classes, \( C \). The authors have tested a number of methods to decide \( C \) based on the Akaike Information Criterion (AIC) and its variant: Consistent AIC (CAIC) \[ \text{(Louviere et al., 2000)} \]. They are given by Equations 6 and 7:

\[
\text{AIC} = -2[LL(\hat{\beta}) - C \cdot K_c - (C - 1)K_s]
\]

\[
\text{CAIC} = -2 \cdot LL(\hat{\beta}) - [C \cdot K_c + (C - 1)K_s - 1] \cdot [\log(2N) + 1]
\]

\( LL(\hat{\beta}) \) is the log likelihood at the estimated parameters \( \hat{\beta} \). \( K_c \) denotes the number of parameters in the utility function of the class-specific choice models. \( K_s \) denotes the total number of parameters in the class membership model. \( N \) is the total number of observations used by the model. In this study, LC model with two latent classes has the best AIC and CAIC measures.

After deciding the number of classes, three latent class models have been estimated and the results are reported and compared in Table 3. As suggested by the class membership model, whether or not one particular respondent is a current carpooler partially determines his/her class membership. Therefore, latent classes reflect different latent attitudes toward ride-sharing. Presumably, **Latent Class 1** represents a group that relatively favors ride-sharing, while, on the other hand, **Latent Class 2** represents a group that dislikes ride-sharing. The class-dependent choice has been modeled with various explanatory variables. Model 1 is a full model with 15 coefficients. Model 2 reduces the degree of freedom from 15 to 13 by fixing the coefficient of schedule delay early in the **Latent Class 1** and that of fuel cost in the **Latent Class 2**. Model 3 further reduces the degree of freedom by regulating the coefficients of HOV usage in the **Latent Class 1** to be equal to zero.

The coefficients of the variables in Model 3 are with the expected signs and are all significant at 90% confidence level and above. The coefficients of the class membership probability model are also significant and with the correct signs. This significance supports the claim that the latent carpooling preference is related to individuals’ choice. Model 1 seems to have the best statistical goodness of fit in terms of log likelihood and pseudo Rho-squared statistic. Model 2 and 3 put restrictions on the parameters in the likelihood expression that effectively reduce the total number of unknown parameters. Likelihood ratio tests for Model 2 and Model 3 have been conducted. The null hypothesis that the more restricted models 2 and 3 are statistically as good as the full model 1 has been accepted at the 95% significance level and thereby has all the assumptions about the
Table 3: Estimation Results for the Latent Class Logit Model

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff. (Std. Err.)</td>
<td>Coeff. (Std. Err.)</td>
<td>Coeff. (Std. Err.)</td>
</tr>
<tr>
<td>Latent Class 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Travel time (min.)</td>
<td>-0.129 (0.023)***</td>
<td>-0.133 (0.071)*</td>
<td>-0.151 (0.091)*</td>
</tr>
<tr>
<td>Fuel cost ($)</td>
<td>-0.828 (0.212)***</td>
<td>-0.836 (0.409)**</td>
<td>-0.818 (0.458)*</td>
</tr>
<tr>
<td>Toll cost ($)</td>
<td>-1.331 (0.210)***</td>
<td>-1.346 (0.519)***</td>
<td>-1.387 (0.564)**</td>
</tr>
<tr>
<td>Schedule delay early (min.)</td>
<td>0.002 (0.009)</td>
<td>0 (fixed)</td>
<td>0 (fixed)</td>
</tr>
<tr>
<td>Schedule delay late (min.)</td>
<td>-0.029 (0.008)***</td>
<td>-0.029 (0.016)*</td>
<td>-0.031 (0.017)*</td>
</tr>
<tr>
<td>HOV usage</td>
<td>-0.453 (0.267)*</td>
<td>-0.446 (0.529)</td>
<td>0 (fixed)</td>
</tr>
<tr>
<td>Latent Class 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Travel time (min.)</td>
<td>-0.023 (0.007)***</td>
<td>-0.022 (0.005)***</td>
<td>-0.022 (0.005)***</td>
</tr>
<tr>
<td>Fuel cost ($)</td>
<td>-0.001 (0.054)</td>
<td>0 (fixed)</td>
<td>0 (fixed)</td>
</tr>
<tr>
<td>Toll cost ($)</td>
<td>-0.339 (0.039)***</td>
<td>-0.340 (0.025)***</td>
<td>-0.343 (0.025)***</td>
</tr>
<tr>
<td>Schedule delay early (min.)</td>
<td>-0.010 (0.004)***</td>
<td>-0.010 (0.003)***</td>
<td>-0.010 (0.002)***</td>
</tr>
<tr>
<td>Schedule delay late (min.)</td>
<td>-0.005 (0.003)*</td>
<td>-0.005 (0.002)**</td>
<td>-0.005 (0.001)**</td>
</tr>
<tr>
<td>HOV usage</td>
<td>-3.098 (0.771)***</td>
<td>-3.081 (0.178)***</td>
<td>-3.121 (0.182)***</td>
</tr>
<tr>
<td>Class Membership Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC</td>
<td>-2.374 (0.350)***</td>
<td>-2.406 (0.429)***</td>
<td>-3.253 (1.369)***</td>
</tr>
<tr>
<td>Age</td>
<td>-1.360 (0.563)**</td>
<td>-1.371 (0.496)***</td>
<td>-1.334 (0.466)**</td>
</tr>
<tr>
<td>Carpool</td>
<td>2.826 (0.771)***</td>
<td>2.844 (0.573)***</td>
<td>2.754 (0.536)***</td>
</tr>
<tr>
<td># of Obs.</td>
<td>1,457</td>
<td>1,457</td>
<td>1,457</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-681.58</td>
<td>-681.59</td>
<td>-681.91</td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>0.319</td>
<td>0.319</td>
<td>0.319</td>
</tr>
</tbody>
</table>

*significance level: * - 90% ** - 95% *** - 99%
coefficients accepted. The calculation procedure of the Model 2 is shown by Equation 8, where $L(\beta_i)$ denotes the likelihood of Model $i$.

$$\chi^2 = -2 \ln \frac{L(\beta_2)}{L(\beta_1)} = 0.020 < \chi^2(0.95, 2) = 5.99$$

(8)

The class membership model suggests that younger drivers and current carpoolers are more likely to belong to the **Latent Class 1**. For this class, drivers are more sensitive to congestion, fuel cost, pricing, and schedule delay late, as shown in the model. **Latent Class 1** has a very low coefficient of the variable: HOV usage, which indicates this class is less penalized by ride-sharing. As shown, schedule delay early is only significant for **Latent Class 2**, which means **Latent Class 1** may be less sensitive to early departures. In addition, the taste variation of toll cost is also shown explicitly in the model as the toll cost has a higher negative impact to **Latent Class 1**.

The value of time (VOT) found for **Latent Class 1** is $11.1/h, which is reasonably consistent with VOT found by other researchers (e.g. $12/h in Cirillo and Axhausen [2006] and $8/h in Algers et al. [1998]). For **Latent Class 2**, the coefficient of fuel cost is very close to zero. The VOT calculated based on the Toll cost coefficient is relatively low ($4.1/h). This result is consistent to other studies on HOT tolling (e.g. Carrion and Levinson [2011] their multinomial logit models yielded around $4/h VOT) Although a higher VOT seems more plausible, one reasonable explanation could be that the marginal disutility of cost variables for **Latent Class 2** drivers is too high, given that the cost is actually generated by HOT tolling. And drivers probably do not have tight time constraints (61.6% of samples stated that they have flexible work schedule) and thus have less incentive to trade-off toll cost for potential travel time savings.

The average probability that individuals belong to **Latent Class 1** is 11.5%. Note that 19% of the survey sample is already carpooling. Thus a proportion of the current carpoolers (most likely to be senior drivers, as suggested by the class membership model) actually latently dislikes ride-sharing and belongs to **Latent Class 2**. Table 4 reports the probability to belong to the class of individuals with ride-sharing penalty for each segment of the population.

<table>
<thead>
<tr>
<th>Age</th>
<th>Currently Carpooling?</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 45</td>
<td>Yes</td>
<td>71.3%</td>
</tr>
<tr>
<td>&gt; 45</td>
<td>No</td>
<td>97.7%</td>
</tr>
<tr>
<td>≤ 45</td>
<td>Yes</td>
<td>38.9%</td>
</tr>
<tr>
<td>≤ 45</td>
<td>No</td>
<td>91.5%</td>
</tr>
</tbody>
</table>

The cross tabulation of the conditional logit model, and the latent class model has been examined by the authors. The latent class model has a considerably better success rate in predicting the outcomes within the sample. The prediction accuracy of the latent class model is 79.00% (1,151 successful predicts out of 1,457 observations), while the accuracy of the conditional logit model is only 71.60%. In terms of log likelihood and pseudo Rho squared statistics, the latent class model also does a better job.
5 Closing Remarks and Further Work

This empirical research presents a model of departure time choice based on the Washington D.C. Capital Beltway travel survey data. This dataset captures individuals’ behavior responses to road pricing, managed lane policy, and other travel condition dynamics. A conditional logit model and a latent class model have been developed and estimated. Heterogeneity in behavior is apparent in the latent class model. The latent class model, with specified latent carpooling preference, has superior explanatory power compared to other models. The class probability model suggests that younger drivers and current carpoolers are more likely to belong to Latent Class 1, whose departure time choice is more sensitive to travel time, toll cost, and fuel cost, as these coefficients are negative and statistically significant. This finding may justify further exploration of age-specific considerations in incentive-based programs. For instance, there may be possibilities in designing psychologically based interventions targeted at younger generation to increase their awareness of the personal and social benefits of ride-sharing. Similar policies may be initiated and justified to maintain the existing carpoolers. This is an important policy implication especially when many of the existing programs promoting ride-sharing are found to make little impact on daily travel pattern (Wang and Chen 2012). Compared to Class 2, departure time decision made by individuals in Class 1 are more dependent on schedule delay late and are more biased toward ride-sharing. Aside from avoiding toll charge, using HOV/HOT lanes is also often associated with narrower travel time bandwidth, which means less travel time uncertainty. This finding suggests that there exists a group of drivers who have more fixed travel schedule and prefer carpooling rather than paying toll to avoid possible extreme delays. While this paper contributes to the travel behavior research by revealing individuals’ relative preference towards tolling and HOV usage, future research can take full advantage of the dataset by combining the SP data with RP data and jointly model departure time and HOV/HOT/general-purpose lane choices.

6 Acknowledgment

This research is partially funded by the U.S. Federal Highway Administration Exploratory Advanced Research Program and the Maryland State Highway Administration. Views and opinions herein do not necessarily represent the views of research sponsors. The authors are responsible for all statements in the paper. The authors would like to thank graduate students in the Transportation Systems Research Lab at University of Maryland for implementing the behavioral survey. The authors would also like to thank Dr. Shanjiang Zhu and Nathan Hopper for their valuable suggestions and comments.

References


