

ANALYZING COMMUTER TRAIN USER BEHAVIOR: A DECISION FRAMEWORK FOR ACCESS MODE AND STATION CHOICE

Vincent Chakour
Junior Engineer
Transports Québec
Montréal, Québec, H2Z 1W7
Canada
Ph: 514 864 1750, Fax: 514 864 1765
Email: vincent.chakour@mtq.gouv.qc.ca

Naveen Eluru*
Assistant Professor
Department of Civil Engineering and Applied Mechanics
McGill University
Montréal, Québec, H3A 2K6
Canada
Ph: 514 398 6823, Fax: 514 398 7361
Email: naveen.eluru@mcgill.ca

*corresponding author

November, 2013

Abstract

The purpose of the current research effort is to develop a framework for a better understanding of commuter train users' access mode and station choice behavior. Typically, access mode and station choice for commuter train users is modeled as a hierarchical choice with access mode being considered as the first choice in the sequence. The current study proposes a latent segmentation based approach to relax the hierarchy. In particular, this innovative approach simultaneously considers two segments of station and access mode choice behavior: Segment 1 - station first and access mode second and Segment 2 – access mode first and station second. The allocation to the two segments is achieved through a latent segmentation approach that determines the probability of assigning the individual to either of these segments as a function of socio-demographic variables, level of service (LOS) parameters, trip characteristics, land-use and built environment factors, and station characteristics. The proposed latent segment model is estimated using data from an on-board survey conducted by the *Agence Métropolitaine de Transport (AMT)* for commuter train users in Montreal region. The model is employed to investigate the role of socio-demographic variables, level of service (LOS) parameters, trip characteristics, land-use and built environment factors, and station characteristics on commuter train user behavior. The results indicate that as the distance from the station by active forms of transportation increases, individuals are more likely to select a station first. Young persons, females, car owners, and individuals leaving before 7:30 am have an increased propensity to drive to the commuter train station. The station model indicates that travel time has a significant negative impact on station choice, whereas, presence of parking and increased train frequency encourages use of the stations.

Key words: Access mode choice, station choice, commuter train user behavior, latent segmentation model

1. Introduction

Transportation professionals in developed countries such as Canada and USA are focused on improving the sustainability of the transportation system. In this regard, the high share of personal automobile travel is of particular concern. The negative externalities of excessive dependence on personal vehicles are well documented. An often suggested alternative to reducing personal automobile travel is the transit mode (Hodges, 2009). A well-designed transit system can provide equitable access to employment and recreational opportunities for the entire urban population, while simultaneously offering significant environmental benefits by offsetting emissions from personal vehicles (FHWA, 2002). Naturally, the recent decade has seen substantial interest within the travel behavior community on examining the key determinants of transit mode usage. The emphasis of this stream of research is on identifying the impact of individual and household socio-demographics, household residential neighborhood characteristics, transportation network attributes, transit service characteristics, and spatial and temporal transit accessibility on transit usage.

Montréal, with its unique multimodal transit system consisting of bus, metro and commuter train, offers a rich array of public transit alternatives to individuals travelling to and from different parts of the city. The commuter train provides access to the urban population from the suburbs to the central business district of Montréal. In this research, we examine the behavior of the commuter train riders in terms of their commuter train station and travel mode to commuter train choices (access mode). The focus of the analysis is on developing a behaviorally representative framework for understanding the decision processes involved in the station and access mode choice.

We propose an innovative latent segmentation approach that simultaneously considers two segments of station and access mode choice behavior: Segment 1 - station first and access mode second and Segment 2 – access mode first and station second. The allocation to the two segments is achieved through a latent segmentation approach that determines the probability of

assigning the individual to either of these segments as a function of socio-demographic variables, level of service (LOS) parameters, trip characteristics, land-use and built environment factors, and station characteristics. Within each segment, the sequence structure imposed is followed to examine the choice processes. To elaborate, in the first segment, access mode choice is modeled first and the station decision is modeled using the access mode choice decision. In the second segment, the choices are reversed. The latent segmentation based framework will allow us to identify important factors that affect the choice sequence decision while simultaneously modeling the access mode and station choices. In fact, through this approach, we allow for two distinct choice hierarchies (access mode first and station second (MS) and station first and access mode second (SM)) to be simultaneously considered in the analysis as two segments for individuals.

The remainder of the paper is organized as follows; Section 2 will provide a brief review of earlier research while positioning the current research effort in context. Section 3 discusses the econometric methodology employed in our research. In section 4, details about the survey and data assembly procedures are outlined. The next section presents the results of the model estimation. Further, we also undertake a policy exercise to illustrate the applicability of the proposed model. Section 6 concludes the paper.

2. Earlier Research and Current Study in Context

The travel behavior community has examined travel mode choice decision in substantial detail. A complete review of literature on travel mode choice is beyond the scope of our study. Briefly, earlier research on travel mode choice (not just access mode choice) has shown that individual and household socio-demographic characteristics such as age, gender, income, and vehicle ownership influence mode choice decisions (Bhat, 1997; Cervero & Gorham, 1995). The local built environment, population density and urban form affect travel mode choice; denser areas

increase the likelihood of choosing the transit mode (Pinjari et al., 2007; Rajamani et al., 2003; van Wee & van Baren, 2002).

In terms of transit behavior, the decision framework for boarding station choice and access mode choice has received extensive attention in the transportation research community (for example see Liou and Talvitie (1974) for a research effort from the 70s). A large proportion of these studies focused on access mode choice. The findings from studies investigating mode choice to train stations are analogous to those obtained from studies on general mode choice. Givoni and Rietveld (2007) show that the availability of a car does not have a strong effect on the choice of access mode to the station. Further, the authors find that improving accessibility to stations by adding newer stations will only result in a mode shift from transit to active transportation (walking and cycling), leaving the car mode share unchanged. Keijer and Rietveld (2000) found that the mode choice behavior depends strongly on distance to station. Specifically, active modes of transportation are preferred for shorter distances, whereas driving and transit are favored for longer distances. Krygsman et al. (2004) found that if the distance to the station exceeds a certain threshold, users will not consider transit alternatives. Bergman et al., 2011 examined access mode choice behavior using data from the Portland region. In this study, the authors explored the impact of historical mode choice behavior, subjective assessment of transit attributes on access mode choice. Finally, researchers have also examined the access to rail stations by active modes of transportation (Park and Kang, 2008, and Appleyard, 2012).

On the other hand, research on boarding station choice has found that frequency of trains at the station, parking availability, station facilities, and travel time to station (always considered along with mode choice) plays a major role in the decision process (Debrezion et al., 2007, 2009; Fan et al., 1993; Wardman & Whelan, 1999). The most common approach employed when modeling mode and station choice simultaneously is the nested logit model with mode as the choice in the upper level. It is important to note that only Fan et al. (1993) and

Wardman and Whelan (1999) employ disaggregate individual level models. The other studies (Debrezion, et al., 2007, 2009) develop aggregate models at the postal code level (not individual level). The aggregate studies employ socio-demographic information at a postal code level and individual level information is not considered. Moreover, most of the access mode and station choice research has been undertaken in the European context where car mode share to train station (drive alone or shared ride) is lower than 15% (Givoni & Rietveld, 2007). The behavioral processes under consideration might be different in the North American context, especially given that the car mode share to station is greater than 60% (much higher for most urban regions).

2.1. Current Study

All the above studies examining station choice consider a very small sample of stations (2 or 3) in the choice set. We observed from the Montréal commuter train data that people exhibit a great variability in terms of the station choice in the database. Residents from the same neighborhood are observed to have boarded the commuter trains at varying locations, indicating that the station choice is not merely a decision to arrive at the nearest commuter train station (not even the nearest 3 stations). For a variety of reasons such as seat availability, parking, fare, or better transit coverage, some respondents travel to stations farther from home (probably in the direction of the destination) to board the commuter train. The current research, in addition to examining the access mode choice (drive alone, shared ride, transit, and active transportation), will also investigate the heterogeneity among individuals in choosing the commuter train stations (50 stations in the Montréal metropolitan region).

The decision framework of determining the station at which to board the commuter train and the corresponding travel access mode are interconnected. There is reason to believe that these are potentially simultaneous decisions. There are two approaches that have been employed to study these choices. The first approach employs a discrete choice model that has composite alternatives of station and travel mode combination (i.e. every combination of travel

mode and station is considered as an alternative). In this approach, it is important to recognize the potential correlations between sets of alternatives. Towards accommodating such correlations, some studies have considered nested logit version of the composite alternative models where one of the decision is placed in the upper level and the other in the lower level (Debrezion, et al., 2007, 2009). The approach, though plausible, imposes a hierarchy that is very hard to validate in the dataset. Further, the number of alternatives explodes very quickly in this approach. For instance, the number of possible combination alternatives might go as high as 200 (4 modes and 50 stations). The second approach employed in literature to account for the simultaneity involved in the decision process is to develop a simultaneous equation model that explicitly accounts for common unobserved heterogeneity across the two decisions (see Eluru et al., 2009, Pinjari et al., 2011 for examples of such approaches). These approaches are simulation intensive and focus predominantly on the unobserved correlation across the choice processes.

In our paper, we propose an alternate approach to study such simultaneous choices. Specifically, we employ a new latent segmentation based approach that allows us to incorporate simultaneously the two possible sequences (MS and SM). To elaborate, we hypothesize that individuals are likely to consider joint choices or interconnected decisions in a sequence, even if the time difference between these decisions is infinitesimally small. Now, if there was a way to determine the hierarchy (i.e. whether individuals decide first on station or access mode), we can develop an appropriate sequential approach to modeling the decision process (see Liou and Talvitie (1974) for a study with two sequences handled separately). Unfortunately, the true sequence is latent to the analyst. Hence, we propose a latent segmentation approach where the first segment follows the station first and mode second sequence and the second segment follows the mode first and station second sequence. The individuals are then allocated to these two segments based on a host of exogenous variables, including socio-demographic variables, LOS parameters, trip characteristics, land-use and built environment factors, and station

characteristics. For instance, workers have primary access to automobile in the household and are probably more likely to decide on their mode (automobile) while subsequently depending on the perception of parking availability to decide on the station. Similarly, individuals residing close to the station might decide on the station first and then either walk or take transit (in inclement weather) to arrive at the station.

3. Methodology

The modeling approach proposed consists of three components: (1) latent segmentation component, (2) Mode choice component for each segment and (3) Station choice component for each segment. The first component represents a binary logit model while the latter two components are two multinomial logit models (see Waddell et al., 2007 for a similar approach).

Let q be the index for commuters ($q = 1, 2, \dots, Q$) and i be the index for segment ($i = 1$ or 2), m be the index for mode choice alternative ($m = 1, 2, \dots, M$), and s be the index for station alternative ($s = 1, 2, \dots, S$). With this notation, the random utility formulation takes the following form:

$$u_{qi}^* = \alpha'x_{qi} + \varepsilon_{qi} \quad (1)$$

$$u_{qim}^* = \beta_i'x_{qim} + \varepsilon_{qim} \quad (2)$$

$$u_{qis}^* = \gamma_i'x_{qis} + \varepsilon_{qis} \quad (3)$$

where u_{qi}^* represents the utility obtained by the q^{th} commuter in selecting the i^{th} segment, u_{qim}^* represents the utility obtained by choosing mode alternative m in the i^{th} segment, and u_{qis}^* represents the utility obtained by choosing station alternative s in the i^{th} segment. x_{qi} , x_{qim} , x_{qis} are column vector of attributes influencing the choice framework. ε_{qi} , ε_{qim} and ε_{qis} are assumed to follow Type 1 Gumbel distribution. The commuter q will choose the alternative that offers the highest utility. $\alpha, \beta_i, \gamma_i$ are corresponding coefficient column vector of parameters to be estimated. The second model in each segment is conditional on the first model in the segment. x_{qij} , x_{qik} incorporate the information available to the commuter at that instant in the choice

process. For example, if the mode choice is the first alternative, level of service attributes to the chosen station by the chosen mode are unavailable in the model.

The probability expression for each model component takes the usual multinomial logit form given by:

$$P_{qi} = \frac{\exp(\alpha_i'x_{qi})}{\sum_{j=1}^I \exp(\alpha_i'x_{qj})} \quad (4)$$

$$P_{qim} = \frac{\exp(\beta_i'x_{qim})}{\sum_{m=1}^M \exp(\beta_i'x_{qim})} \quad (5)$$

$$P_{qis} = \frac{\exp(\gamma_i'x_{qis})}{\sum_{s=1}^S \exp(\gamma_i'x_{qis})} \quad (6)$$

With these preliminaries, the latent segmentation based probability for joint choice of mode m and station s with two segments can be formulated as follows:

$$P_{qms} = P_{q1}P_{q1m}P_{q1s} + P_{q2}P_{q2s}P_{q2m} \quad (7)$$

The first term in Equation (7) reflects the first sequence - mode first and station second while the second term reflects the second sequence - station first and mode second. The exogenous variables in the second choice are generated while recognizing the chosen alternative attributes from the first choice process in the segment.

The log-likelihood at the individual q is defined as:

$$L_q = \delta_{ms} \ln(P_{qms}) \quad (8)$$

where $\delta_{ms} = 1$ if the mode and station combination is the chosen alternative and 0 otherwise.

$$L = \sum_q L_q \quad (9)$$

The log-likelihood function is constructed based on the above probability expression, and maximum likelihood estimation is employed to estimate the $\alpha_i, \beta_i, \gamma_i$ parameters. The model is programmed in GAUSS matrix programming language.

4. Data

The primary source of data for the research was based on an onboard survey conducted by the AMT for commuter train users in the month of September 2010. The information compiled includes individual and household socio-demographics such as age, gender, vehicle ownership, and occupational status. Also included are residential location, boarding and alighting commuter stations, final destination location, travel mode to the boarding commuter train station and from the alighting commuter train station, and travel departure times. The exhaustive database on the commuter train travel is compiled for analysis by eliminating missing records and inconsistent information.

4.1. Level of Service Variable Generation

To undertake travel access mode choice analysis, assembly of LOS attributes for all available alternative modes under consideration is required. In our study, we are faced with the challenge of generating these measures for all the alternatives as well as for all stations possible. A Google Maps based algorithm was used to generate the walk, cycle, drive, and transit time for all viable stations (more details on the process of compiling viable stations is described below). Further, transit alternatives available to the chosen station based on the departure times provided in the survey were also generated using a Google Maps based algorithm. The information on a transit trip was compiled only for those individuals for whom a transit alternative was available. Transit can be unavailable if the station is very close to the individual's residence or if there are no transit services within 37 minutes of walking for the individual (a threshold implicitly established in Google Maps). For our model analysis, we randomly sample 3,902 individuals from the 24,000 survey responses. The reason for sampling was to reduce the computational burden of generating level of service attributes. The survey database is appropriately augmented with the LOS attribute database generated. Also, parking data and train frequency for each commuter train station was obtained from the AMT.

4.2. Station Choice Set Formation

To generate a behaviorally representative station choice set, we focussed on individual level choice set preparation. Considering all the station alternatives in the region as part of the choice set would not truly represent individual behavior.

The objective was to identify the maximum distance commuters are willing to travel vis-a-vis the closest station, where this distance is not necessarily in the direction of travel. Towards this end, we compute a distance ratio measure for every respondent. The measure is defined as:

$$D = \frac{\text{Distance to Chosen Station} - \text{Distance to Closest Station}}{\text{Distance to Closest Station}} \quad (10)$$

Once we compute D for all the respondents in the sample, the 95th percentile was used to determine the threshold value. Based on the threshold value, viable alternatives can be generated. However, generating a single ratio measure would be inadequate. For example, a person living 500 m from the closest station might be willing to travel further – 3 km (high D, say 5) whereas the person living 15 km from the closest station might not consider travelling 90 km (assuming same D = 5). So we adopted a distance based D measure that reduces as the distance to closest station increases. We considered D values for the following intervals – 0-0.5 km, 0.5 – 1 km, 1 – 2 km, 2 – 4 km and > 4 km. The maximum viable distance thus computed was used to generate the set of alternatives that are feasible for each respondent. The number of alternatives varies from 1 to 18, with 91% of respondents having between 1 and 5 alternatives.

However, creating only one set of viable stations means that irrespective of chosen mode, the station set is generated in the same fashion for all individuals. This may lead to potential inaccuracy in the analysis, as the station set might also depend on mode (for the Mode first and Station second segment). For instance, an individual who walks to the train station should have a station set for which all stations are at a walkable distance. If this same individual would have driven to train station, it is possible that his or her station set would expand, as more

stations are accessible by driving than walking. To address this, a new set of viable stations based on the chosen mode (driver/shared ride transit, active mode (walk/bike)) is generated. The same methodology previously stated is employed for each of the sub-samples to create mode specific station sets. This station set is exclusively employed when modeling station as second choice. In station as first choice model, mode to station is unknown to the analyst and is therefore not incorporated when generating station choice set.

4.3. Sample Statistics

Table 1 contains summary statistics for the sample compiled. The following observations can be made based on the summary statistics. The sample consists of 3902 commuters of whom more than half reported driving to the commuter rail station. Females represent a slightly higher proportion of the sample. The majority of the respondents are between 26 and 54 years of age. Vehicle ownership (recorded as a binary variable) is constant in the different age groups at about 90%, with the exception of individuals 25 and under. The sample predominantly consists of workers, and almost half of the respondents claim to have departed between 6:30am and 7:30am. Further, we can see that individuals do not always board the nearest station – average distance to the nearest station is lower than the average distance to chosen station. Figure 1 illustrates the distribution of the geographic locations of the sample throughout the region. Evidently, there is a strong concentration of respondents around the train lines, notably in the western part of the Island of Montréal and Laval, as well as on the North Shore and the South Shore.

5. Analysis and Discussion

5.1. Variables Considered

The universal set of variables considered in our analysis for the latent segmentation sequence choice, mode choice, and station choice components of the proposed model includes individual

and household socio-demographics, trip specific attributes, level of service attributes, station attributes, residential and station level land-use and built environment variables.

The individual and household variables considered include age, gender, and vehicle ownership (characterized as Yes/No). The LOS attributes considered include travel time by different modes, average travel times to viable stations, and travel time to closest and chosen stations. The trip specific attributes considered include egress mode and departure time. Station level attributes considered include parking and fare information for the station.

To generate residential and station level land-use and built environment information, a large database of land-use, socio-demographic, transportation network, and vehicle ownership information was generated at the Traffic Analysis Zone (TAZ) level for the Montréal metropolitan region. Naturally, using all these variables simultaneously in model estimation was not possible due to potential collinearity effects. Hence, a principal component analysis (PCA) based factor analysis was conducted (see Pinjari et al., 2008 for example of PCA based analysis). For factor loading and other technical information of the PCA analysis please review Sider et al. (2013). For this study, the results from the factor analysis are directly employed. The variables from the large dataset were classified into two categories: (1) demand-side variables and (2) supply-side variables. For the demand-based category, three orthogonal factors were derived: (1) zones with high median income and high proportion of newer vehicles, (2) zones with high vehicle ownership and high proportion of larger vehicles, and (3) zones with large proportion of older vehicles. The supply-based variables provided three orthogonal factors: (1) zones with high density, high walkability, and transit oriented developments (TOD), (2) zones with commercial land-use, and (3) zones with government & institutional land-use. All of these six factors were considered in our model estimation.

5.2. Model Estimation Results

The universal set of variables was considered in the estimation of the three components – latent segmentation sequence choice, mode choice and station choice. The variables were carefully

chosen so as to correspond to the sequence under consideration for each segment. The specification process was based on a systematic process of removing statistically insignificant variables. The process was guided by intuition and findings from earlier literature.

5.2.1. Model Fit Measures

In order to adequately assess the explanatory power of our latent model, its log-likelihood was compared to the log-likelihood obtained from the sequential models run separately. The log-likelihood for separate MS, separate SM, and our latent segmentation model are -6,525.33, -6,198.65, and -5,603.00 respectively. The models are not nested within one another and hence we use the Bayesian Information Criterion (BIC) measure to examine statistical significance. The BIC is defined as $-2\ln(L) + K \ln(Q)$, where $\ln(L)$ is the log-likelihood value at convergence, K is the number of parameters, and Q is the number of observations. The model with the lower BIC value is the preferred model. The corresponding values for the separate MS, separate SM, and latent segmentation models are 13,094.65, 12,437.51, and 11,288.90 respectively. The BIC values illustrate the statistical advantage offered by the latent segmentation model. The model which imposes SM sequence offers better fit compared to the model with MS sequence. Moreover, the improvement in the combined model indicates that the population share for MS sequence is far from negligible and requires a careful consideration for policy analysis. The results provide evidence to the presence of two hierarchies in the choice process, thus providing support to our hypothesis in Section 2.1.

5.2.2. Latent Segmentation Shares

Prior to examining the individual parameters, we generated the overall aggregate share of the two segments in the population. We find that about 36% of the respondents' share is allocated to the MS segment, while the remainder is allocated to the SM segment. The segment shares indicate that the population share for MS and SM sequences is substantial and requires a careful consideration for policy analysis. We find that for the MS segment, the access mode shares are 42% (drive alone), 12% (shared ride), 12% (transit), and 34% (active) while for the

SM segment, the access mode shares are 62% (drive alone), 15% (shared ride), 12% (transit), and 11% (active). These results indicate that at the aggregate level there are significant differences in mode share across the two segment choices. The MS segment has a large share of respondents choosing auto mode and active transportation while in the SM segment large share of respondents rely on automobile (drive or shared ride). Quite interestingly, the share of transit remains constant across the two segments.

5.3. Model Parameters

The model is divided into three distinct components: the sequence choice component (Table 2a), the mode choice first and station choice second component, and the station choice first and mode choice second component. Please note that there are two versions of mode choice and station choice components. To facilitate the ease of comparison across the two models for each choice we present the mode choice models in Table 2b and station choice models in Table 2c. Specifically, in Table 2n and 2c, model estimates presented in columns two and three correspond to the MS sequence while model estimates presented in columns four and five correspond to SM sequence.

5.3.1. Sequence Component

The latent segmentation component examines the sequence of the decision process (i.e. whether the individual considers mode first station second sequence (MS) or the station first and the mode second sequence (SM)). The parameter results offer quite interesting insights on the decision process.

The constant term indicates that given everything else remains the same the MS segment is more preferred relative to the SM segment (of course, the nature of the relationship might be skewed by the values of independent variables as is the case in our context where SM segment has 64% share). The inclination for workers to opt for MS segment indicates that workers are likely to decide on their mode decision first. This is probably a manifestation of the impact of a rigid schedule.

Individuals with longer walk times to the nearest station prefer to make station decisions first. This attribute, used as a surrogate for proximity to station, suggests that the further you live to a station, the more likely you are to choose the station first. The result appears counter intuitive; however, the individuals that live far from any station are likely to be willing to consider more options for station. Hence, these respondents decide their station first and then choose their mode to arrive there. Individuals leaving before 7.30 am are likely to be assigned to the SM segment.

5.3.2. MS Segment

The MS segment considers mode choice as the first decision, followed by station choice. In this section, the results for Mode as the first choice and station as the second choice are presented – the columns two and three from Table 2b and 2c. To incorporate for similarities across the two alternatives (drive and shared ride) and to estimate a parsimonious specification, we retained the same parameters for drive and shared ride alternatives.

The constants of the mode choice model do not have any behavior interpretation because the alternative utility value is determined by the value of the independent variables in the model. Men are less likely to use the car alternatives for traveling to the station. Household car ownership coefficient indicates that individuals from households with cars are unlikely to choose the transit alternative. The finding is contrary to earlier research (Debrezion, et al., 2009; Givoni & Rietveld, 2007). However, we believe it is more intuitive that individuals with car exhibit preference to drive (as indicated by the overall modal share).

The travel time variables offer interesting results. Two variables are significant in our analysis – average travel time to viable stations and travel time to the closest station. The former coefficient has a negative value while the latter coefficient has a positive value. The net effect is clearly negative because the numerical value of the travel time to closest station is always lower than the average travel time to viable stations. Hence, as expected in any mode choice decision, increasing travel times exerts an overall negative effect on the utilities of the

modes (see Debrezion, et al., 2009; Givoni & Rietveld, 2007; Keijer & Rietveld, 2000 for similar results). The direction of the train trip significantly impacts mode choice. Specifically, individuals whose trips are in the direction of the central business district (CBD) are more likely to choose a car mode to get to the station. Also, we see that individuals residing in areas with high factor values for government and institution area are likely to opt for transit or active transportation alternatives. Finally, individuals residing in zones with high score for the factor high car ownership and high percentage of larger vehicles are less likely to use automobile alternatives (drive alone or shared ride).

In terms of the station choice parameters for the MS segment, the variable effects are intuitive. Travel time by the chosen mode has a significant negative impact on station choice. Further, car owners find travel time even more burdensome. As is expected, we observe that train frequency impacts station choice positively. The parking inventory variable has the expected result; higher parking inventory encourages use of the stations. The higher parking inventory serves as a surrogate for station attractiveness. The results are very similar to findings from earlier research (Debrezion, et al., 2007, 2009; Fan, et al., 1993; Wardman & Whelan, 1999). Of course, there is a clear case of endogeneity that we are not considering for this variable. It is entirely possible that these stations have high parking because of the high demand at these stations. The examination of the endogeneity is a direction for further investigation in the future.

5.3.3. SM Segment

In this section, the results for Station as the first choice and Mode as the second choice are presented – columns four and five from Table 2c and 2b.

In this segment, the surrogate for travel time – considered as drive time to station – impacts the station choice negatively. It is observed that individuals prefer to choose stations that have higher factor values for commercial zones. These zones usually reflect central business districts, and the result is not surprising given the higher accessibility by different

modes to stations in these areas. The distance from station to CBD variable presents a negative value, highlighting a preference for stations that are in the direction of the final destination. The frequency and parking inventory variables offer very similar results compared to the station as second choice. The negative coefficient for parking between 1 and 200 can be explained by a better accessibility to some stations with 0 parking, as they can be located in denser and more accessible areas.

The mode second parameter estimates offer similar results as mode first, specifically for males, travel time, and train direction. Younger users (25 and younger) have a higher likelihood of choosing the drive alone and shared ride modes in the decision processes.

Individuals who have indicated their egress mode as transit are more favorably inclined to use the latter in the access trip, indicating that they are not averse to using transit. In this segment, individuals residing in zones with high score for the factor high car ownership and high percentage of larger vehicles are more likely to use automobile alternatives. Finally, if the chosen station is located in an area with high factor values for government and institution area, the transit and active modes are less likely to be chosen.

5.4. Policy Analysis

The model estimates provide useful insight on the impact of various parameters on mode and station choice. However, to highlight the utility of the proposed framework in evaluating the impact of various policy alterations, we undertake a series of sensitivity analysis. Specifically, we examine how mode and station choice decisions alter with changes to a set of the exogenous variables such as individual characteristics, LOS measures, land use attributes, and train frequency. The analysis is undertaken as elasticity values for mode and station choices (see Eluru et al., 2012 for a similar analysis in the context of public transportation). The exact procedure employed for elasticity analysis depends on the nature of the variable (categorical or continuous). For more details on the procedure employed for elasticity analysis please Eluru

and Bhat, 2007. For continuous variables such as travel time changes in probability for mode choice and station choice for an increase/decrease of 5 minutes was considered.

The results of the computation, in the form of percentage change, are presented in Table 3. We observe that being male strongly increases the likelihood of transit and active transportation use while being employed has a positive influence on active transportation usage. Car ownership has a very strong negative effect on choosing the transit mode. Reducing the drive time by 5 minutes has a significant negative impact on transit and active transportation. Reducing drive time by 5 minutes to a station (while maintaining the same times to other stations) significantly increases the odds of choosing that station. It is quite interesting to observe that improving transit service to reduce transit times by 5 minutes has a substantially large positive increase in transit usage. Reducing transit time by 5 minutes is relatively inelastic for station choice. The effect of a 5 minute reduction in walk time impacts the automobile and transit modes negatively and active transportation very positively. Station choice remains inelastic for walking time. All results from land use changes are inelastic. Particularly, 15% increases in government and institutional area, high vehicle ownership/ high LTD share area, and commercial area almost all present changes of less than 1% in mode share and station choice. Increasing train frequency, as expected, provides increased station level usage with about an 8.9% increase per unit increase.

The policy implications of these findings are quite clear and provide straightforward interpretations. Car ownership plays a major role in transit use, but there creating policies to reduce car ownership are far from straight forward. Based on our results, increasing public transportation service and accessibility is the most beneficial method to increase transit mode share. To this end, reducing transit travel time to stations by expanding the feeder bus network or increasing bus frequency can be effective policies. On the other hand, mode and station choice do not seem to react to land use changes. Therefore, the priority for transit agencies

should be to provide stronger transit accessibility to train stations all the while increasing train frequency to attract more users.

6. Conclusion

The current study proposes a latent segmentation based approach to jointly consider the access mode and station choice decisions without imposing any hierarchy. In our approach, we simultaneously consider the two possible sequences while also determining the individual allocation for each segment as a function of socio-demographic variables, LOS parameters, trip characteristics, land-use and built environment factors, and station characteristics. The proposed model is estimated using an onboard survey conducted by the AMT for commuter train users. The information compiled includes individual and household socio-demographics such as age, gender, vehicle ownership, and occupational status. Also included are residential location, boarding and alighting commuter stations, final destination location, travel mode to the boarding commuter train station and from the alighting commuter train station, and travel departure times. The survey database was appropriately augmented with the LOS attribute database generated as well as parking and train frequency data for each commuter train station.

In the sequence model, the results indicate that as the distance from the station by active forms of transportation increases, individuals are more likely to select a station first. For the mode choice model, an inclination exists for workers and individuals leaving before 7:30 am to opt for mode first. Young persons, females, car owners, and individuals leaving before 7:30 am have an increased propensity to drive to the commuter train station. Travel time by mode has a negative impact on the mode choice. The station model indicates that travel time by the chosen mode has a significant negative impact on station choice. Finally, the presence of parking encourages use of the stations; not surprising because parking spots at a station is a function of parking demand. Frequency also has a positive impact on station choice. The model developed was used to undertake a validation exercise on a hold-out sample. The latent

segment model's performance was compared with the two sequential models. The results highlight an improvement on the latent segment model from two sequential models. The model estimates are employed to undertake policy analysis highlighting the role of gender, car ownership, train frequency and level of service in mode and station choice decisions.

This paper is not without limitations. The study does not account for various station specific attributes such as percentage of seat availability and parking availability (different from inventory). However, data collection for such attributes is rarely considered. Transit agencies need to enhance their database inventory to better understand transit choice behavior. Further, a joint choice of mode and station choice with a large alternative set within the latent segmentation context is a possible direction for future research.

Acknowledgements

The corresponding author would like to acknowledge financial support from Natural Sciences and Engineering Research Council (NSERC) of Canada under the Discovery Grants program for undertaking the research. The authors would like to extend their appreciation to Daniel Bergeron from *Agence Métropolitaine de Transport (AMT)* and Prof. Ahmed El-Geneidy from McGill University for sharing the data used in our study. The authors would also like to acknowledge useful feedback from three anonymous reviewers and the editor Prof. David Hartgen.

References

Appleyard, B., 2012. Sustainable and healthy travel choices and the built environment.

Transportation Research Record: Journal of the Transportation Research Board 2303 (1), 38-45.

- Bergman, A., Gliebe, J., Strathman, J., (2011). Modeling access mode choice for inter-suburban commuter rail. *Journal of Public Transportation* 14 (4), 23-42.
- Bhat, C. (1997). Work travel mode choice and number of non-work commute stops. *Transportation Research Part B: Methodological*, 31(1), 41-54.
- Cervero, R., & Gorham, R. (1995). Commuting in Transit Versus Automobile Neighborhoods. *Journal of the American Planning Association*, 61(2), 210-225.
- Debrezion, G., Pels, E., & Rietveld, P. (2007). Choice of departure station by railway users. *European Transport*, XIII(37), 78-92.
- Debrezion, G., Pels, E., & Rietveld, P. (2009). Modelling the joint access mode and railway station choice. *Transportation Research Part E: Logistics and Transportation Review*, 45(1), 270-283.
- Eluru, N., and C.R. Bhat (2007), "A Joint Econometric Analysis of Seat Belt Use and Crash-Related Injury Severity", *Accident Analysis & Prevention*, Vol. 39, No. 5, pp. 1037-1049
- Eluru, N., I.N. Sener, C.R. Bhat, R.M. Pendyala, and K.W. Axhausen (2009), "Understanding Residential Mobility: A Joint Model of the Reason for Residential Relocation and Stay Duration", *Transportation Research Record*, Vol. 2133, pp. 64-74
- Eluru, N., V. Chakour, and A. El-Geneidy (2012), "Travel Mode Choice and Transit Route Choice Behavior in Montreal: Insights from McGill University Members Commute Patterns," *Public Transport: Planning and Operations* Vol. 4, No. 2, pp. 129-149
- Fan, K., Miller, E., & Badoe, D. (1993). Modeling Rail Access Mode and Station Choice. *Transportation Research Record*, 1413, 49-59.
- FHWA. (2002). *Status of the Nation's Highways, Bridges, and Transit*.
- Givoni, M., & Rietveld, P. (2007). The access journey to the railway station and its role in passengers' satisfaction with rail travel. *Transport Policy*, 14(5), 357-365.
- Hodges, T. (2009). *Public Transportation's Role in Responding to Climate Change*.

- Keijer, M. J. N., & Rietveld, P. (2000). How do people get to the railway station? The dutch experience. *Transportation Planning and Technology*, 23(3), 215-235.
- Krygsman, S., Dijst, M., & Arentze, T. (2004). Multimodal public transport: an analysis of travel time elements and the interconnectivity ratio. *Transport Policy*, 11(3), 265-275.
- Liou, P. S., A. P Talvitie (1974). Disaggregate access mode and station choice models for rail trips, *Transportation Research Record* 526,42-65
- Martens, K. (2004). The bicycle as a feeding mode: experiences from three European countries. *Transportation Research Part D: Transport and Environment*, 9(4), 281-294.
- Park, S., Kang, J., (2008). Factors that influence walking and biking to the station: Modeling commuter rail user's access mode choice. *In: Proceedings of the Transportation Research Board 87th Annual Meeting.*
- Pinjari, A., Eluru, N., Bhat, C., Pendyala, R., & Spissu, E. (2008). Joint Model of Choice of Residential Neighborhood and Bicycle Ownership: Accounting for Self-Selection and Unobserved Heterogeneity. *Transportation Research Record: Journal of the Transportation Research Board*, 2082(-1), 17-26.
- Pinjari, A., Pendyala, R., Bhat, C., & Waddell, P. (2007). Modeling residential sorting effects to understand the impact of the built environment on commute mode choice. *Transportation*, 34(5), 557-573.
- Pinjari, A., Pendyala, R., Bhat, C., & Waddell, P. (2011). Modeling the Choice Continuum: an Integrated Model of Residential Location, Auto Ownership, Bicycle Ownership, and Commute Tour Mode Choice Decisions. *Transportation*, 38(6), 933-958.
- Rajamani, J., Bhat, C., Handy, S., Knaap, G., & Song, Y. (2003). Assessing Impact of Urban Form Measures on Nonwork Trip Mode Choice After Controlling for Demographic and Level-of-Service Effects. *Transportation Research Record: Journal of the Transportation Research Board*, 1831(-1), 158-165.

- Sider, T., Alam, A., Zukari, M., Dugum, H., Goldstein, N., Eluru, N., et al. (2013). Quantifying The Effects Of Land-Use And Socio-Economics On The Generation Of Traffic Emissions And Individual Exposure To Air Pollution At A Metropolitan Scale. , Conditionally accepted for publication in the Journal of Transport Geography.
- van Wee, B., & van Baren, R. (2002). Preferences for modes, residential location and travel behaviour: the relevance of land-use impacts on mobility. *European Journal of Transport and Infrastructure Research*, 2(2), 305-315.
- Waddell, P., Bhat, C., Eluru, N., Wang, L., & Pendyala, R. (2007). Modeling Interdependence in Household Residence and Workplace Choices. *Transportation Research Record: Journal of the Transportation Research Board*(2003), 84-92.
- Wardman, M., & Whelan, G. A. (1999). *Using geographical information systems to improve rail demand models*.

Author Biographies

Vincent Chakour is currently employed as a junior engineer in Transports Québec. Vincent recently completed his Masters in Engineering from the Department of Civil Engineering and Applied Mechanics at McGill University. His expertise includes data compilation, and discrete choice analysis with a specific emphasis on public transit.

Naveen Eluru is an Assistant Professor in Transportation at McGill University in Montreal. His primary research areas of interest include transportation planning, socio-demographic and land-use modeling, sustainable urban design, integrated demand supply models, activity time-use analysis and transportation safety. He is currently a member of Transportation Research Board (TRB) committee on Statistical Methods (ABJ80).

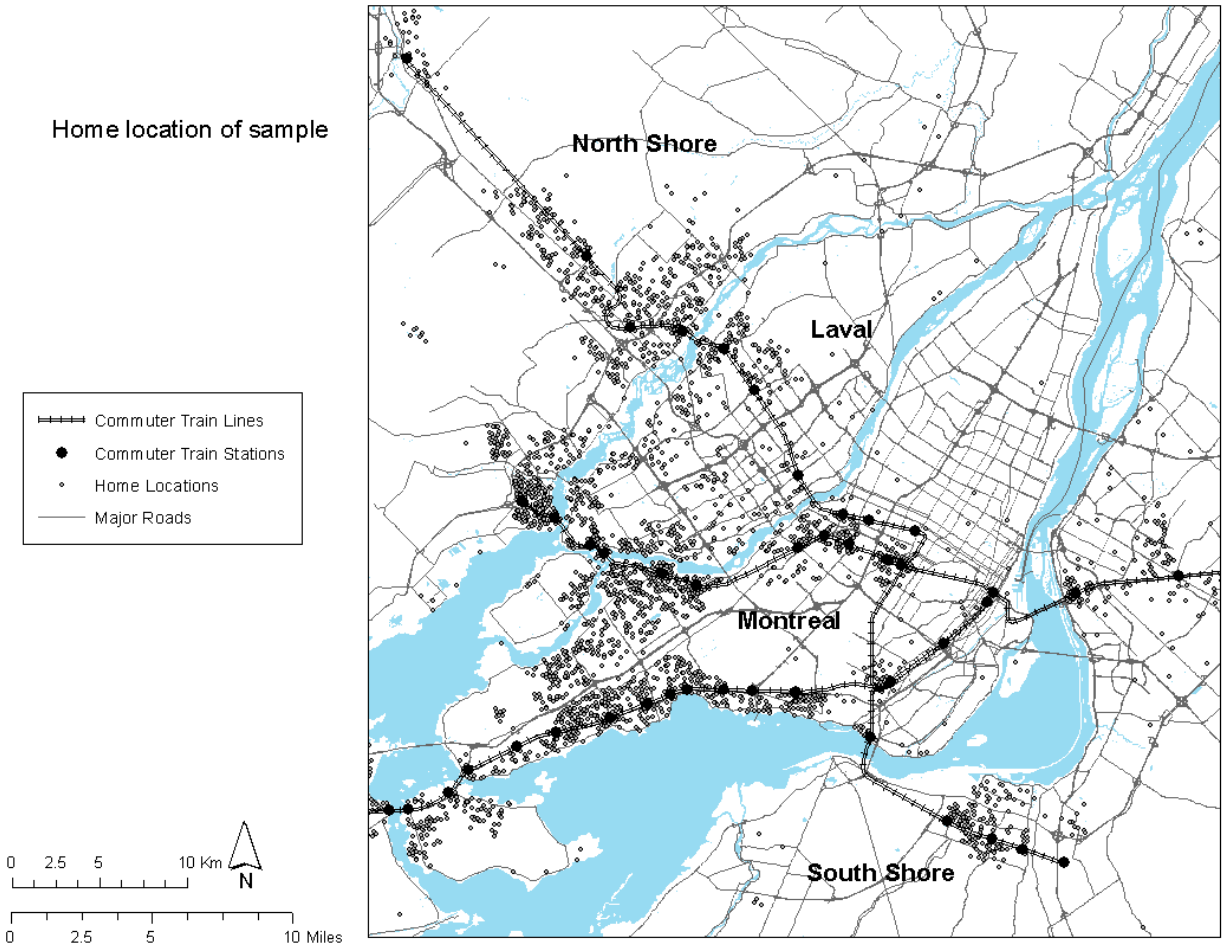


Figure 1: Home Location of Survey Sample

Table 1: Summary Statistics

<u>Summary Statistics (N=3902)</u>	(%)
Drive	51.2
Passenger	12.4
Transit	11.3
Active (Walk/Bicycle)	25.1
<i>Gender</i>	
Males	43.8
Females	56.2
<i>Age</i>	
<25	19.6
25-39	34.1
40-54	37.2
>55	9.1
<i>Car Ownership within age groups</i>	
<25	56.9
25-39	93.1
40-54	92.1
>55	88.3
<i>Status</i>	
Worker	81.9
Student	17.6
Other	0.5
<i>Time left home</i>	
Before 6:30 am	18.4
Between 6:30 am and 7:30 am	47.7
After 7:30 am	33.9
<u>Station Characteristics</u>	
<i>Distance to stations</i>	(km)
Average distance to closest station	3.6
Average distance to chosen station	4.9
<i>Number of stations with:</i>	
No parking at station	14
Between 0 and 200 spots	12
Between 201 and 500 spots	12
Between 501 and 700 spots	7
Between 701 and 1000 spots	4
More than 1000 spots	3
<i>Ratio of stations being boarded</i>	(%)
No parking at station	7.3
Between 0 and 200 spots	11.7
Between 201 and 500 spots	29.9
Between 501 and 700 spots	18.0
Between 701 and 1000 spots	13.8
More than 1000 spots	19.3

Table 2 (a): Latent Segmentation based Mode and Station Choice Model Results

Variables	Sequence Choice			
	<u>Mode First Station Second</u>		<u>Station First Mode Second</u>	
	Coefficient	t-stat	Coefficient	t-stat
Constant	0.788	4.30	-	-
<u>Individual Demographics</u>				
Worker	0.407	2.80	-	-
<u>Travel Time (unit is hours)</u>				
Walk time to closest station	-2.311	-10.11	-	-
<u>Trip Characteristics</u>				
Time left home before 7:30 am	-0.415	-3.47	-	-

Table 3 (b): Access Mode Choice Model Results

	<u>First (MS Sequence)</u>		<u>Second (SM Sequence)</u>	
	Coefficient	t-stat	Coefficient	t-stat
Drive alone constant	-7.421	-4.24	-1.329	-1.53
Shared ride constant	-9.039	-5.15	-3.431	-3.94
Transit constant	-1.437	-4.44	0.060	0.37
<u>Individual Demographics</u>				
25 years old and younger				
-Drive alone	-	-	0.809	5.43
-Shared ride	-	-	0.809	5.43
Male				
-Drive alone	-0.618	-1.89	-0.719	-5.50
-Shared ride	-0.618	-1.89	-0.719	-5.50
Car ownership				
-Transit	-2.268	-5.27	-	-
<u>Travel Time (unit is hours)</u>				
Average travel times to viable stations	-20.927	-10.46	-	-
Travel time to closest station	16.158	8.08		
Total travel time to chosen station	-	-	-0.342	-2.81
<u>Trip Characteristics</u>				
Egress mode is transit				
-Transit	-	-	0.633	4.11
Trip is in the direction of CBD				
-Drive alone	3.692	2.18	3.149	3.68
-Shared ride	3.692	2.18	3.149	3.68
<u>Land-use and Built Environment Factors</u>				
Government and institutional areas (at origin)				

-Transit	0.889	2.00	-	-
-Active transportation	0.889	2.00	-	-
Government and institutional areas (at station)				
-Transit	-	-	-0.103	-3.54
-Active transportation	-	-	-0.103	-3.54
High vehicle ownership/High LTD share				
-Drive alone	-0.919	-2.75	0.655	5.43
-Shared ride	-0.919	-2.75	0.655	5.43

Table 4 (c): Station Choice Model Results

Station Choice				
	<u>Second (MS Sequence)</u>		<u>First (SM Sequence)</u>	
	Coefficient	t-stat	Coefficient	t-stat
<u>Travel Time (unit is hours)</u>				
Drive time to station	-	-	-7.171	-14.58
Travel time with chosen mode to station	-6.448	-3.75	-	-
Travel time for car owners	-2.339	-6.14	-	-
<u>Station Service</u>				
Train frequency	1.573	6.74	0.163	6.40
Distance from station to CBD	-	-	-0.037	-4.20
<u>Land-use and Built Environment Factors</u>				
Commercial area	-	-	0.314	2.66
<u>Parking Inventory (0 parking is base)</u>				
Between 1 and 200 parking spots	3.697	3.79	-1.300	-5.14
Between 201 and 500 parking spots	1.361	1.54	1.082	5.26
Between 501 and 700 parking spots	3.261	3.34	0.935	4.36
Between 701 and 1000 parking spots	3.815	3.88	1.204	5.40
More than 1000 parking spots	4.051	4.16	2.207	9.71

Table 3: Elasticity Analysis

	Mode Choice				Station Choice
	<u>Drive Alone</u>	<u>Shared Ride</u>	<u>Transit</u>	<u>Active</u>	
Individual Characteristics					-
Male	-10.37	-19.28	35.57	15.94	-
Car Ownership	4.49	7.12	-50.71	10.45	-
Worker	-3.46	-3.22	-0.43	9.41	-0.26
Travel Time					
Drive time (-5min)	1.74	2.73	-4.51	-3.09	15.35
Transit time (-5min)	-0.74	-1.36	8.37	-1.63	0.61
Walk time (-5min)	-2.71	-2.84	-3.50	9.02	0.21
Land Use					
Gov Area (+15%)	0.10	0.18	-0.26	-0.18	-
High LTD/Veh (+15%)	0.38	1.02	-1.76	-0.53	-
Commercial Area (+15%)	-	-	-	-	-0.28
Frequency					
+1	-	-	-	-	8.87
+2	-	-	-	-	16.14