Monte Carlo Simulation of Adaptive Stated Preference Survey with a case study: Effects of Aggregate Mode Shares on Individual Mode Choice

Carlos Carrion* Nebiyou Tilahun † David Levinson ‡

August 1, 2012

Abstract

Monte Carlo experiments are used to study the unbiasedness of several common random utility models for a proposed adaptive stated preference survey. This survey is used to study the influence of the knowledge of existing mode shares on travelers mode choice. Furthermore, the survey is applied to a sample of subjects selected from the University of Minnesota. The results indicate that the presence of mode shares in the mode choice model does influence the decision of travelers. Unfortunately, the estimates are found to be biased by the Monte Carlo experiments.

Keywords: *mode choice, mode shares, mixed logit, stated preference, monte carlo.*

Word Count: *5,746*
Word Count incl. figures and tables: *7,496*

*Corresponding Author, University of Minnesota, Department of Civil Engineering, carri149@umn.edu

†Assistant Professor, Department of Urban Planning and Policy, University of Illinois at Chicago, 412 S. Peoria St, Suite 254, Chicago, IL 60607, USA, ntilahun@uic.edu

‡RP Braun-CTS Chair of Transportation Engineering; Director of Network, Economics, and Urban Systems Research Group; University of Minnesota, Department of Civil Engineering, 500 Pillsbury Drive SE, Minneapolis, MN 55455 USA, dlevinson@umn.edu, `http://nexus.umn.edu`

1 Introduction

This study presents the addition of a new possible factor that influences the travelers' preferences towards travel modes. This factor (i.e. social influence variables) represents the influence of the decision of others travelers. Researchers have argued that social networks, and the interactions of social contacts within these networks may influence decisions related to travel. Recent studies have focused on travel choices influenced by distinct layers of proximity of social contacts in travelers' social networks (household members: (1, 2, 3, 4); friends, colleagues: (5, 6, 7)). However, endogeneity has been a significant issue in studies using discrete choice modeling in the random utility framework as social influence variables are likely to be correlated with unobserved factors leading to inconsistency, and biasedness of estimates. This is because it has been more difficult to isolate the correlation of social influence variables with unobserved factors in revealed preference data.

This study uses stated preference survey design to model the subjects' mode choice decisions on distinct level of aggregate mode shares as reported for each choice situation. In essence, the study explores the persuasion of mode shares as a source of information for travelers to base their choices. It should be noted that the interpretation travelers may give to mode shares may be different across them, but this is not the purpose but rather the influence of mode shares. The stated choice experiment design allows the researcher to present the social influence variables as external information to the travelers. At the moment, there are no studies similar to this one with regards to stated preference surveys and social influence variables.

The recoverability of the true parameters from stated preference experiments using econometric models is a property of superlative importance. It is related to the unbiasedness (or consistency depending on sample size) of the estimators of the econometric models using data collected from subjects answering hypothetical scenarios from stated choice experiments. In this study, the proposed adaptive stated preference design (see section 3.2) is investigated using Monte Carlo methods (8, 9, 10).

Four data generating processes are simulated: no heterogeneity (parameters are assumed to be homogeneous across the population); random intercepts (only the intercepts are assumed to be independent and identically normally distributed across the population); random coefficients (only the coefficients of the regressors are assumed to be independent and identically normally distributed across the population); and random coefficients and intercepts (both coefficients of the regressors and intercepts are assumed to be independent and identically normally distributed across the population). In addition, it is assumed that there is zero covariance (i.e. no correlation) between random parameters (i.e. coefficients and/or intercepts) in the data generating processes with heterogeneity. Also, it is assumed that the unobserved heterogeneity is only manifested in the coefficients of the regressors and/or the intercepts; there is no scale heterogeneity (see (11) for scale heterogeneity). Moreover, it is assumed that subjects follow a compensatory behavior (i.e. subjects choose the alternative with the highest utility). It is also assumed that the subjects have linear in parameters systematic utility functions. The results identify additional biasedness of the estimators using data from adaptive stated preference experiments further extending previous results by (12). In addition, only (12), and this study have significantly explored the recoverability of true parameters from adaptive stated preference choice experiments.

The remainder of the study is organized as follows: Section 2 presents a literature review briefly covering the principal areas of research in travelers' mode choice. Section 3 presents the data collection effort, descriptive statistics of the data, econometric models used in the analysis, and the proposed Monte Carlo experiments. That is followed in Sections 5 and 6 with a discussion of the results and concluding remarks respectively.

2 Literature Review

Researchers have looked at several aspects of the trip within mode choice models, including trip purpose (e.g. commute, leisure), trip attributes (e.g. travel time which may be different by mode), measures of level of service associated with a mode, travelers' characteristics (e.g. income), features of the built environment, social influence effects, as well as data type/sources for estimating such models (13).

Trip purpose refers to the travelers' intentions with regards to their prospective destinations and activities. Generally, mode choice models has been developed for commute trips. This may be because of data availability. The general idea is that travelers will evaluate their mode choices differently depending on their trip purpose (14).

Travel time and out of pocket travel costs (e.g. fares, tolls) constitute the main relevant factors in explaining mode choice decisions. Travelers have a fixed amount of time to allocate to different activities as well as a fixed amount of wealth (i.e. income) to allocate to distinct consumption activities. Increased expenditure in either of these therefore translates into disutilities to travelers. Disutilities attached to travel time could further be divided into other components. For example, travelers may incur higher disutility for time spent waiting in comparison to the time spent traveling inside their vehicles (15, 16).

Travelers' characteristics have been incorporated in mode choice models in order to control for (observed) heterogeneity. The evaluation of attributes may also differ across travelers, and thus the inclusion of travelers' characteristics allows for market segmentation. Several studies have shown the importance of income, gender, auto ownership, age, occupation, number of licensed drivers in the household, and others (17).

Researchers have also argued that the formation of social networks, and the interactions between social contacts may influence decisions related to travel (18). Recent studies have focused on travel choices influenced by distinct layers of proximity of social contacts in travelers' social networks (household members: (1, 2, 3, 4) ; friends, colleagues: (5, 6, 7)). In the random utility framework, the social influence effect is abstracted into variables such as the share of decision-makers selecting a specific choice. Thus, the coefficient of social influence variables may represent (as previously discussed) distinct behaviors: imitation, herd behavior, and others. In addition, researchers have distinguished between global (decision-makers are influenced by all decision-makers) versus local (decision-makers are influenced by a subset of decision-makers) social interaction effects (19). The local effects may be grouped by similar socio-economic attributes (e.g. income), and spatial proximity of residential location (20). Furthermore, endogeneity has been considered a potential issue with social effect variables (21). It refers to the correlation of the social effect variables with unobserved factors leading to inconsistency, and biasedness of estimates. In the social networks literature, there are studies dealing explicitly with endogeneity. (22) includes instruments based on excluded trips to study the bicycle cultural effects in german cities. (21) based on (23) discusses methods (e.g. control function) to address the endogeneity issues in discrete choice models.

Several mode types can be considered as part of the choice set of travelers in mode choice analysis. The inclusion of modes in the travelers' choice set when using revealed data depends on the existence of the mode in the market. These choices can be limited to the automobile and transit or may include carpools and non-motorized alternatives. There are also cases where researchers desire to ascertain the possible demand for modes entering the current market (see for example (24)). Situations where the choices of interest are not yet part of the market can be handled by the collection of stated preference (SP) data. Stated preference experiments put decision makers in a simulated (or fictional) market while revealed preference (RP) refers to observed behavior in an actual market (25).

It has been well known that SP experiments may differ in results from RP. One of the main reasons is the difference behind what individuals say and what they actually do. This difference may be due to a myriad of reasons that may be related to how the stated preference experiments resemble reality or emulates the situation the individual will confront in a real market. Unfortunately, it is typically hard to obtain revealed preference data. In some cases, the variables exhibit high levels of multicollinearity as there is not sufficient variation of values of the variables in the real market, and thus stated preference experiments may help. In other cases, real market situations (e.g. a new mode) may yet not exist, and thus revealed preference data cannot be collected. The validity of the preferences collected from SP data may be affected by the lack of realism, and the subject's understanding of the abstract situations. Thus, the subject's mode preferences may not be similar to the ones during their actual trips (25, 26). Also, the stated preference design (e.g. adaptive, fixed; see (12)) may exert an influence on the unbiasedness of the estimators of econometric models. Moreover, new modeling techniques have been developed to combine RP and SP data, and to correct for the scale issues of one over the other (25). The idea behind these techniques is to ground stated choices to real choices, and to use SP data to stabilize RP data allowing more precise estimates.

3 Data and Methodology

This study is based on a stated choice experiment. The following subsections describe the collection of the data, the administered surveys to the subjects, the econometric modeling effort, and the Monte Carlo experiments.

3.1 Recruitment

Subjects for the survey were randomly selected from a University of Minnesota staff list excluding students and faculty. Subject recruitment was done through announcements sent by email in the Summer of 2004. Each email addressed the individual by name and offered them a gift of USD$15 for participating in the survey. A total of 91 subjects participated in the study, of which 77 subjects were left, after dropping subjects that did not answer most of the survey questions, and the travel diary. Furthermore, subjects had to fulfill the following requirements for their participation:

1. Legal driver,
2. Full-time job and follow a "regular" work schedule
3. The main mode of travel is in the study's choice set (automobile, bike/walking, and transit).

3.2 Survey Design

This study uses data collected from a computer based adaptive stated preference survey on individual mode preference based on regional aggregate mode shares. In addition to the SP questions, individuals are asked about their sociodemographic background and mode preferences (e.g. auto/bike ownership, biking frequency, mode to work). Subjects are also asked to provide a one day travel diary.

The SP experiment gives the subjects a hypothetical mode share for the Twin Cities area and asks the respondent which mode they would use. All respondents face the same first alternative where the mode shares are 85% auto, 10% transit and 5% bike/walk. Mode shares on subsequent presentations are each informed by the alternatives in the previous question and the choice made by the respondent. Based on their choice, respondents face one of three potential mode share distributions for presentation 2, one of nine possible mode shares for presentation 3, and one of twenty seven alternatives for the final presentation.

The underlying assumption in this survey is that if a person exhibits a preference for mode m when it constitutes a share $x\%$ in the population, then any larger percentage would also be preferred by the respondent. Lowering the preferred mode's share suggests to the respondent that the mode has become less attractive to the regional population, which implies that either its quality of service has declined or that of the alternatives has improved. The survey takes these shifts to be implicit and doesn't explicitly communicate the implications to respondents. When a respondent switches modes, in some contexts it is a response to perceived change in quality, and in others may be driven by mimicking of others.

The way in which the mode shares are generated for each presentation is by lowering the proportion of people that are using the mode chosen in the immediate prior presentation. If on presentation i, the respondent chose mode m, on the next presentation, mode m's share goes down to 75% of what it was previously. The reduction from mode m is equally divided between the remaining two modes. With $S_{i,m}$ representing share for the chosen mode m on presentation i, the shares S for modes m, $m1$, and $m2$ on the next presentation $(i+1)$ are:

$$\begin{aligned} S_{i+1,m} &= 0.75 * S_{i,m} \\ S_{i+1,m1} &= \frac{0.25 * S_{i,m}}{2} + S_{i,m1} \\ S_{i+1,m2} &= 100 - S_{i+1,m} - S_{i+1,m1} \end{aligned}$$

To make the presentation questions easier for respondents, only the integer portion of $S_{i+1,m}$ and $S_{i+1,m1}$ are taken. There are a possible 3^4 (eighty one) different choice patterns for any given individual over these presentations. In addition, though the enumeration of presentations leads to 40 presentations (1+3+9+27 as described above), because some choice paths lead to identical mode shares, the number of unique combinations in the design is 34 presentation. Each individual therefore faces four of thirty four different possible presentations. Since each presentation depends on the choices prior to it, some of these may not be presented to any given individual while others may appear more frequently in the final dataset. In the survey, 21 of these 34 unique mode share distributions are presented. An example survey presentation is shown in Figure 1.

Following the stated preference experiment, the survey asks the subjects to report other demographic variables as well as their current mode which may be important indicators of choice behavior. These include questions about the subjects' age, income, auto/bike ownership as well as questions about frequency of biking/walking, and preferred mode for distinct situations such as mode used to get to work today, during the summer period, and others.

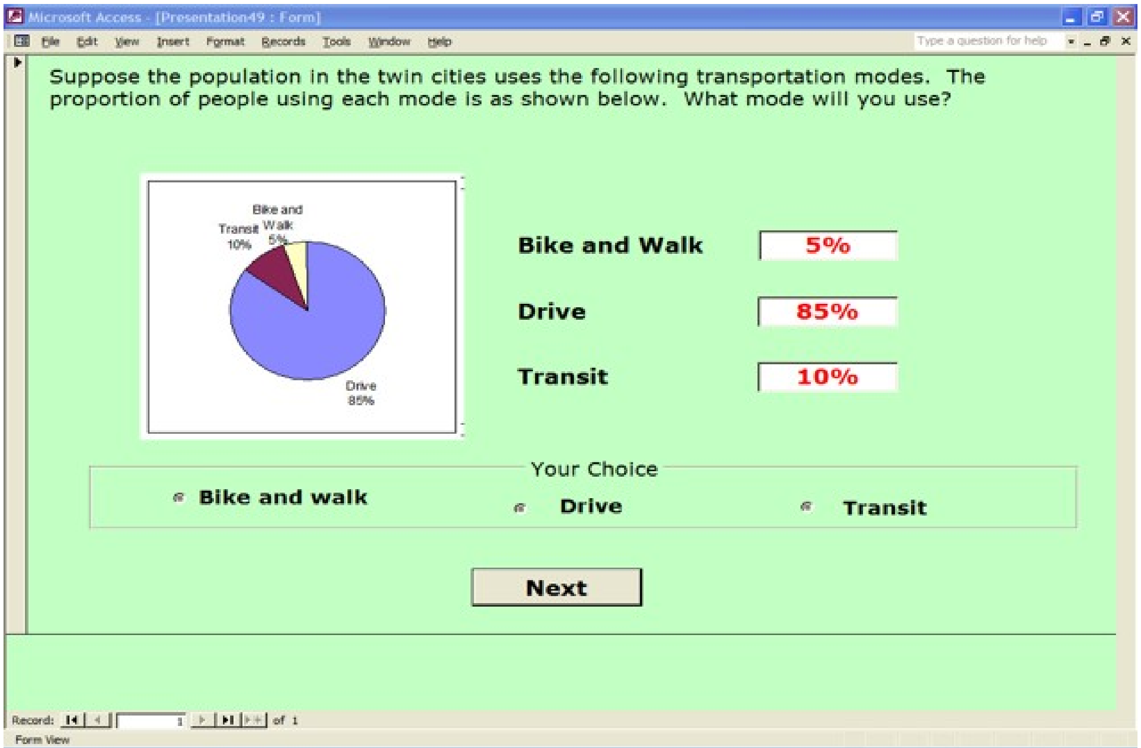

Figure 1: Sample screenshot of survey questions.

3.3 Descriptive Statistics

Table 1, summarizes socio-demographic information of the subjects. The main differences between the sample and the population of the Twin Cities are a higher proportion of females, subjects that are on average older, more educated, and have higher incomes.

Table 1: Socio-Demographics attributes of the sample

Number of Subjects		77	
		Sample	Twin Cities
Sex	Male	31.17%	49.40%
	Female	68.83%	50.60%
Age (Mean, Std. Deviation)		(43.37, 10.46)	(34.47, 20.9)
Education	11th grade or less	0.00%	9.40%
	High School	2.60%	49.60%
	Associate	12.99%	7.70%
	Bachelors	49.35%	23.20%
	Graduate or Professional	35.06%	10.10%
Household Income	$49,999 or less	33.77%	45.20%
	$50,000 to $74,999	28.57%	23.30%
	$75,000 to $99,999	16.88%	14.60%
	$100,000 to $149,999	16.88%	11.00%
	$150,000 or more	3.90%	5.90%

The Twin Cities population statistics are obtained from the 2006-2008 American Community Survey (27)

3.4 Econometric Models: Specification and Estimation

The administered survey is analyzed through a random utility model (28). Three systematic utility functions are specified for each alternative in the choice set. The alternatives considered are obtained directly from the survey design, and these are: Bike and Walk, Drive (or auto), and Transit. Furthermore, a linear in parameters functional form is used for the systematic utility functions. It is unknown at the moment to the authors what type of nonlinearities may be present, and the main purpose is to study whether aggregate mode shares have any influence on the mode choice of the travelers.

The explanatory variables considered in the study relate to those discussed previously in the literature review, and that are available in the collected data. In addition, the mode shares distributions presented to each traveler for the last choice situation are included.

The final selection of the explanatory variables and their specification as either generic or alternative-specific variables was done based on the goodness of fit of the discrete choice model with and without the variables. The variables selected will be discussed in the subsequent sections along with explanations about why other variables were not selected. Moreover, the analysis is performed on panel from the choice situations of the stated choice experiment.

The estimated models are based according to the specific characteristics of the data. For the panel data (4 choice situations per subject), a random effects model is specified, and estimated within the mixed logit framework, and also multinomial logit models with corrected standard errors. Also, it should be remembered that the systematic utilities across all models are the same. Only the unsystematic parts follow distinct covariance structures, and statistical distributions.

The analysis of panel data such as this one (repeated observations per subject) requires a model that handles explicitly the individual-specific variation (or heterogeneity). Both (29) and (30) discuss and recommend several parametric approaches to model the heterogeneity. In this study, a parametric method of random effects is adopted. The assumption is that the observations for each subject represent a cluster with its own variation (within subject variation), but also variation across clusters may be present (between subject variation).

The random effects specification can be formulated in a mixed multinomial logit model (31). Assume that the utility function a decision-maker k in the set of decision-makers $\mathcal{N}$ associates with alternative j in the set of choices $\mathcal{C}$ for a given choice situation t in the set of choice situations $\mathcal{T}$ is given by:

$$U_{jt}^k = V_{jt}^k + \xi_{jt}^k \tag{1}$$

$$U_{jt}^k = V_{jt}^k + [\eta^k + \epsilon_{jt}^k] \tag{2}$$

For this case of mixed logit model, the functional form is given by equation (2). The random term is partitioned into two additive parts: The first (η^k) is an individual-specific random vector distributed as a bivariate normal density function (with zero mean vector) as is typically done for random intercept logits (29), and the second (ϵ_{jt}^k) is a random vector identically and independently distributed (i.i.d.) over alternatives and decision-makers following a extreme value type 1 (or Gumbel) distribution.

The likelihood for this mixed logit model is given by:

$$L(\beta, \Sigma) = \Pi_{\forall k \in \mathcal{N}} \int_{-\infty}^{\infty} \Pi_{\forall t \in \mathcal{T}} \Pi_{\forall j \in \mathcal{J}} \left(\frac{e^{V_j^k(\beta)}}{\sum_{j=1}^{J} e^{V_j^k(\beta)}} \right)^{\gamma_{kjt}} f(\eta^k|0, \Sigma) d\eta^k \tag{3}$$

Where the γ_{kjt} variable is one for the chosen j alternative of the k decision-maker for choice situation t, and zero otherwise. The function $f(\eta^k|0, \Sigma)$ represents the bivariate normal density with zero mean vector (the mean is estimated by the alternative specific constants of the alternatives), and a zero off diagonal for the covariance matrix (the covariance is assumed to be zero between alternatives). Furthermore, the parameters (for a linear in parameters specification, $V_j^k = \beta^T x_j^k$), where β is the coefficient vector, and x_j^k are the vectors of explanatory variables in the regressors matrix) in this model are estimated using STATA (32) with Maximum Simulated Likelihood using 300 Halton draws.

In addition, multinomial logits were estimated using the panel data for comparison purposes. It should be noted that these models consider each observation as an individual (or pseudo individuals), and thus are inappropriate without at least correction for the standard errors. A correction of the standard errors is done through nonparametric bootstraps (33, 34) clustered (i.e. resampling with replacement over subjects instead of individual observations). 250 resamples were used for the nonparametric bootstraps.

The likelihood for these multinomial logit models is given by:

$$L(\beta) = \Pi_{\forall k \in \mathcal{N}} \Pi_{\forall t \in \mathcal{T}} \Pi_{\forall j \in \mathcal{J}} \left(\frac{e^{V_j^k(\beta)}}{\sum_{j=1}^{J} e^{V_j^k(\beta)}} \right)^{\gamma_{kjt}} \tag{4}$$

3.4.1 Systematic Utility for the models

The additive linear in parameters systematic utility for the alternatives for all models is:

$$V_j^k = f(S, C, A; \beta) \tag{5}$$

where

- S: SP Mode Shares variables
- C: Characteristics of the Travelers
- A: Alternative specific constants (ASC)

3.4.2 SP Mode Shares

Two variables are considered to capture the effects of the SP mode shares: ratio of Bike/Walking share to Auto share; and ratio of Transit share to Auto share. The value of these variables will vary from values close to 0 to values close to 1 as the redistribution of mode shares never reduces the auto share below the other two shares. Higher values of the ratios means that the Bike/Walking and Transit shares are closer to the auto share (see section 3.2). These variables are alternative specific to the Bike/Walking and Transit alternatives.

3.4.3 Characteristics of the Travelers

Three characteristics are considered: travelers preference with regards to biking (a dummy variable indicates whether travelers have biked or not to work before; Biking Preference); traveler's age (a dummy variable indicating whether a traveler's age is between 40 and 50); traveler's income (a dummy variable indicating whether a traveler's income is between USD$60,000 and USD$100,000); and the number of vehicles per adults in the household.

3.4.4 Alternative specific constants

For the multinomial logit, the alternative specific constant of the auto is set to 0. For the random effects multinomial logit (mixed logit), the variance of the auto must be set to zero as only two variances can be estimated (see (35)). Furthermore, the random effect can be understood as a random intercept (or alternative specific constants) model. Thus, alternative specific constants represent mean values, and the variances are the random effects deviations.

4 Monte Carlo experiments of the ASP survey

The recoverability of the true parameters from stated preference surveys using econometric models is a property of superlative importance. It is related to the unbiasedness (or consistency depending on sample size) of the estimators of the econometric models using data collected from subjects answering hypothetical scenarios from stated choice experiments. For this purpose, the proposed adaptive stated preference design (see section 3.2) is studied using Monte Carlo Simulations (i.e. simulated data) focusing only on the survey generated mode shares. The mode shares are generated at each choice situation based on the previous choices by the subjects according to rules described in section 3.2. In addition, the effects of "random taste variation" or "unobserved heterogeneity" in the estimates of econometric models is also explored using simulated data (31, 36).

Four data generating processes are simulated using Monte Carlo methods (8, 9, 10): no heterogeneity (parameters are assumed to be homogeneous across the population); random intercepts (only the intercepts are assumed to be independent and identically normally distributed across the population); random coefficients (only the coefficients of the regressors are assumed to be independent and identically normally distributed across the population); and random coefficients and intercepts (both coefficients of the regressors and intercepts are assumed to be independent and identically normally distributed across the population). In addition, it is assumed that there is zero covariance (i.e. no correlation) between random parameters (i.e. coefficients and/or intercepts) in the data generating processes with heterogeneity. Also, it is assumed that the unobserved heterogeneity is only manifested in the coefficients of the regressors and/or the intercepts; there is no scale heterogeneity (see (11) for scale heterogeneity). Moreover, it is assumed that subjects follow a compensatory behavior (i.e. subjects choose the alternative with the highest utility). It is also assumed that the subjects have linear in parameters systematic utility functions. Each of the four data generating processes is represented mathematically as follows:

Nomenclature

- A: It refers to the Auto alternative.
- T: It refers to the Transit alternative.
- BW: It refers to the Bike/Walking alternative.
- γ^{kt}: It is one for the chosen alternative of subject k in choice situation t; zero otherwise.
- U^{k_t}: It is the utility function of subject k in choice situation t for a specific alternative.
- $T_{share}, BW_{share}, A_{share}$: These terms represent the mode shares generated by the adaptive stated preference survey for transit, bike/walking, and auto, respectively.

Choice behavior

$\gamma_A^{kt} = I(U_A^{k_t} > U_T^{k_t} \wedge U_A^{k_t} > U_{BW}^{k_t})$

$\gamma_T^{kt} = I(U_T^{k_t} > U_A^{k_t} \wedge U_T^{k_t} > U_{BW}^{k_t})$

$\gamma_{BW}^{kt} = I(U_{BW}^{k_t} > U_A^{k_t} \wedge U_{BW}^{k_t} > U_T^{k_t})$

No heterogeneity

$$U_A^{k_t} = \epsilon_A^{k_t}$$

$$U_T^{k_t} = \beta_{ASC}^{T} + \beta_{Ratio-T}^{Transit} \frac{T_{share}}{A_{share}} + \epsilon_T^{k_t}$$

$$U_{BW}^{k_t} = \beta_{ASC}^{BW} + \beta_{Ratio-BW}^{BW} \frac{BW_{share}}{A_{share}} + \epsilon_{BW}^{k_t}$$

$$\begin{aligned} \epsilon_A^{k_t} &\sim Gumbel(0,1) \\ \epsilon_T^{k_t} &\sim Gumbel(0,1) \\ \epsilon_{BW}^{k_t} &\sim Gumbel(0,1) \end{aligned}$$

Random intercepts

$$U_A^{k_t} = \epsilon_A^{k_t}$$

$$U_T^{k_t} = \beta_{ASC}^{T} + \nu_T^{k}\sigma_{ASC}^{T} + \beta_{Ratio-T}^{T}\frac{T_{share}}{A_{share}} + \epsilon_T^{k_t}$$

$$U_{BW}^{k_t} = \beta_{ASC}^{BW} + \nu_{BW}^{k}\sigma_{ASC}^{BW} + \beta_{Ratio-BW}^{BW}\frac{BW_{share}}{A_{share}} + \epsilon_{BW}^{k_t}$$

$$\begin{aligned}
&\epsilon_A^{k_t} \sim Gumbel(0,1)\\
&\epsilon_T^{k_t} \sim Gumbel(0,1)\\
&\epsilon_{BW}^{k_t} \sim Gumbel(0,1)\\
&\nu_T^{k} \sim N(0,1)\\
&\nu_{BW}^{k} \sim N(0,1)
\end{aligned}$$

Random coefficients

$$U_A^{k_t} = \epsilon_A^{k_t}$$

$$U_T^{k_t} = \beta_{ASC}^{T} + (\beta_{Ratio-T}^{T} + \nu_T^{k}\sigma_{Ratio-T}^{T})\frac{T_{share}}{A_{share}} + \epsilon_T^{k_t}$$

$$U_{BW}^{k_t} = \beta_{ASC}^{BW} + (\beta_{Ratio-BW}^{BW} + \nu_{BW}^{k}\sigma_{Ratio-BW}^{BW})\frac{BW_{share}}{A_{share}} + \epsilon_{BW}^{k_t}$$

$$\begin{aligned}
&\epsilon_A^{k_t} \sim Gumbel(0,1)\\
&\epsilon_T^{k_t} \sim Gumbel(0,1)\\
&\epsilon_{BW}^{k_t} \sim Gumbel(0,1)\\
&\nu_T^{k} \sim N(0,1)\\
&\nu_{BW}^{k} \sim N(0,1)
\end{aligned}$$

Random coefficients and intercepts

$$U_A^{k_t} = \epsilon_A^{k_t}$$

$$U_T^{k_t} = \beta_{ASC}^T + \nu_{Transit}^{k^1} \sigma_{ASC}^T + (\beta_{Ratio-T}^{Transit} + \nu_T^{k^2} \sigma_{Ratio-T}^T)\frac{T_{share}}{A_{share}} + \epsilon_T^{k_t}$$

$$U_{BW}^{k_t} = \beta_{ASC}^{BW} + \nu_{BW}^{k} \sigma_{ASC}^{BW} + (\beta_{Ratio-BW}^{BW} + \nu_{BW}^{k} \sigma_{Ratio-BW}^{BW})\frac{BW_{share}}{A_{share}} + \epsilon_{BW}^{k_t}$$

$$\begin{aligned}
&\epsilon_A^{k_t} \sim Gumbel(0,1)\\
&\epsilon_T^{k_t} \sim Gumbel(0,1)\\
&\epsilon_{BW}^{k_t} \sim Gumbel(0,1)\\
&\nu_{Transit}^{k^1} \sim N(0,1)\\
&\nu_{Transit}^{k^2} \sim N(0,1)\\
&\nu_{Bike/Walking}^{k^1} \sim N(0,1)\\
&\nu_{Bike/Walking}^{k^2} \sim N(0,1)
\end{aligned}$$

Simulated (panel) data sets for any of the previously described data generating processes are obtained by drawing from uniform distributions, and evaluating inverse cumulative distribution functions (ICDF) to obtain realizations for the normal variates and/or Gumbel variates (8, 9, 10). The realizations are substituted in the utility functions for each subject, and/or iteration depending on the variate, and these values are added to the systematic utility portion of ratio of the mode shares (e.g. transit mode share divided by auto mode share). The simulated choice behavior is compensatory, and subjects for each iteration choose the mode with the highest utility as indicated previously mathematically. The true parameters for the data generating processes are presented in Table 2. The values of the true parameters are chosen based on the results of econometric models with the actual data in Table 3. The Monte Carlo simulations are coded in STATA (37).

Table 2: True Parameters

β_{ASC}^T	β_{ASC}^{BW}	$\beta_{Ratio-T}^{Transit}$	$\beta_{Ratio-BW}^{BW}$	σ_{ASC}^T	σ_{ASC}^{BW}	$\sigma_{Ratio-T}^{Transit}$	$\sigma_{Ratio-BW}^{BW}$
2.00	1.00	6.00	4.00	2.00	0.50	4.00	0.50

Two sample sizes are considered: 70 subjects or 280 observations (4 choice situations per subject); and 500 subjects or 20000 observations (4 choice situations per subject). The econometric models studied are estimated using Maximum Likelihood methods, and these estimators may be biased, but consistent under several conditions (see (38) for details). Thus, the sample of 500 subjects keeps the conclusions objective for this study looking at consistency of the estimators. The sample of 70 subjects is close to the sample of 77 subjects (308 observations) of the actual data set (not simulated) for comparison purposes. The econometric models described in section 3.4.1 are fitted to the simulated panel data sets as follows: multinomial logit, and mixed logit (random intercepts, random coefficients, and both random coefficients and intercepts) for the simulated panel data sets. The probability densities for the random parameters (i.e. coefficients and intercepts) are always independent and identically distributed as normal densities. Furthermore, 1000 replications are conducted of the Monte Carlo experiments for each of the four previously described data generating processes. In other words, 10000 estimates of the parameters in Table 2 are obtained for the fitted econometric models for both the 70 subject simulated data sets, and the 500 simulated data sets. The 1000 replications allow to construct the sampling distributions of the estimators (these are roughly normal as expected; see (38)). It also allows to calculate confidence intervals to control for simulation error of the Monte Carlo estimates of the expected value of the each of the estimators for the parameters in Table 2.

5 Results and Discussion

Table 3 present the estimates of the panel models (stated choice experiment). The SP transit share variable (i.e. Ratio - Transit to Auto share) is statistically significant at 5% level, but the SP bike mode share variable is not found statistically significant. This confirms in part the original hypothesis of mode shares influencing the mode choice of travelers. Thus, subjects were more susceptible to changes in the transit shares than the bike/walking shares. In addition, the signs are positive for the random effects multinomial logit model, but negative for the multinomial logit with bootstrap standard errors. This is a problem as it shows contradicting results. Positive signifies that subjects are likely to consider Bike/Walking or Transit alternative as the mode share for the auto reduces, and the mode share for these alternatives increases. This indicates an underlying behavior that higher value of mode shares means a pull (or attraction) of these shares over the travelers. A possible reason for transit (the statistically significant SP share) is that more passengers may be related to higher frequency, or better service. In contrast, a negative sign indicates that subjects are likely to not consider Bike/Walking or Transit alternatives as the mode share for the auto reduces, and the mode share for the other these alternatives increases. This demonstrates a underlying behavior of the subjects that higher value of mode shares means an increase in the push (or repulsion) of this share over travelers. There are several possible reasons behind the repulsion: contrarian behavior (subjects may not favor the alternative with higher shares because others have preferred it) or higher mode share may be correlated with crowding.

The question becomes whether the sign of these variables is positive according to the random effects multinomial logit or negative according to the multinomial logit with bootstrap standard errors. (34) demonstrated that presence of random taste heterogeneity leads to erroneous estimates in random utility models that ignore it. Thus, the multinomial logit with bootstrap standard errors correct the bias of the standard error for the panel data, but likely do not lead to consistent estimates. The taste heterogeneity is present as the random effects multinomial logit uncovered with statistically significant random effects at 1%. This question is explored using Monte Carlo experiments (see section 4). In tables 4, 5, 6, and 7, it shows how the estimates for the ratio of mode shares for the multinomial logit become negative in the presence of random taste variation or unobserved heterogeneity in the alternative specific constants. However, the absolute value of the estimates is close to the true value of the parameters, but exhibits significant bias likely to be attributed to the adaptive stated preference survey.

In terms of travelers' characteristics, subjects with ages between 40 and 50 were found to favor Bike/Walking, and Transit relative to the auto. This is puzzling as other variables such auto ownership, bike ownership... were found statistically not significant. In addition, it is clear from table 1 that although a significant number of the subjects (25 in the age bracket) fall into this category, there are still subjects who do not gain the additional utility. Other sociodemographic variables (and interactions) were tested, but not found statistically significant.

Furthermore, subjects that indicated that they have never biked to work were founded to be less likely to favor the Bike/Walking and Transit alternative, and subjects' households with high proportion of vehicles per adults were found to favor the Auto over Bike/Walking or Transit alternatives. Both variables are found statistically significant at a 5% in the stated and revealed choice dimensions. In the panel, only the transit alternative is affected at statistical significant levels (at a 5%) with the proportion of vehicles per adults. Thus, certain preferences of subjects translated across stated and revealed choices.

Lastly, the Monte Carlo experiments (see section 4) conducted on the adaptive stated preference survey (see section 3.2) of this study highlighted an important concern: difficulty of the estimators in recovering the true parameters even with no random taste variation or unobserved heterogeneity. This is agreeable with (12); see their discussion of the endogenous adaptive stated preference survey, although the authors further explore the issues in this study. For the panel data (tables 4, 5, 6, and 7), all the studied data generating processes (discussed in section 4) indicate that the multinomial logit, and mixed logit estimators for the coefficients of the regressors (ratios of the mode shares) exhibit small bias to significant bias regardless of the presence of unobserved heterogeneity (random coefficients and/or intercepts). The multinomial logit estimates of the alternative specific constants are unbiased or exhibit very small bias when there is no presence of unobserved heterogeneity in the intercepts (or alternative specific constants). The mixed logit models exhibit significant bias when misspecified (i.e. the sigma parameters are not statistically significant). Thus, mixed logit models with lack of statistical significance in this adaptive stated survey must not be trusted for inferences. For the actual data set (not simulated), the random effects model (or mixed logit with random intercepts) was the only one with statistically significant parameters. In addition, the presence of random taste variation in the intercepts leads to negative signs in models that do not account for it. In contrast, the presence of random taste variation in the coefficient of the regressors (the ratio of mode shares) does not lead to negative signs in the model that do not account for it, but there is still presence of bias. In fact, it does not seem to matter if the models are correctly specified for the data generating process, there will still be bias in the coefficients of the regressors (ratio of mode shares), and none to small bias in the alternative specific constants. The mixed logit models exhibit greater bias even when correctly specific compared to the multinomial logit models correctly specified for the data generating process.

The Monte Carlo experiments help explain the reason behind the previous question whether the sign of SP transit share variable (i.e. Ratio - Transit to Auto share) and SP bike/walking share variable (i.e. Ratio - Bike/Walking to Auto share) is positive or negative. In addition, they show that the adaptive stated preference survey does induces bias to the estimators of the econometric models. However, the actual data generating process behind the actual data is likely to be more complex (e.g. presence of nonlinearities; distinct probability density function for random parameters), and thus the results of the Monte Carlo experiments may not fully apply.

6 Conclusion

The use of disaggregate mode choice modeling has become standard among practitioners and researchers in the travel demand field. In this framework decisions are modeled as individual choices made within the confines of a time and income budget, trip characteristics, mode availability, and household constraints. Each decision maker is considered to be independent. Despite these assumptions, that the choice of others is likely to influence our decisions is intuitive - either directly through copying behavior, or indirectly, through the improvements in service that are likely to accompany the well used alternative. However, these influences are difficult to test using revealed data, and more so for mode choice, which does not change significantly over a short period of time. In this study we use Stated Preference data instead to test the influence of changing mode share on individual decisions.

While one additional traveler's mode choice is not likely to change the magnitude of the mode shares dramatically, larger shifts can have a self propagating quality further pushing their own share illustrating the feedback process of the subjects' choices.

In addition, persistent preferences of the subjects are shown to exist in both the stated and revealed choice dimension as variables representing subjects that have never biked to work, and vehicles per adults in subjects' households are statistically significant at a 5% in both dimensions.

Furthermore, it is shown and discussed that care must be taken in modeling panel data, and especially if the panel data comes from adaptive stated preference surveys. The presence of unobserved taste heterogeneity may lead to inconsistent estimates, and erroneous conclusions, if it is ignored in the models. In addition, the adaptive stated preference survey may also add bias to the estimates of the econometric explored as the Monte Carlo experiments suggest.

References

[1] J. Gliebe and F. Koppelman. A model of joint activity participation between household members. *Transportation*, 29:49–72, 2002.

[2] J. Gliebe and F. Koppelman. Modeling household activity-travel interactions as parallel constrainted choices. *Transportation*, 32:449–471, 2005.

[3] D. Scott and P. Kanaroglou. An activity-episode generation model that captures interactions between household heads: development and empirical analysis. *Transportation Research Part B*, 36:875–896, 2002.

[4] S. Srinivasan and C. Bhat. Modeling household interactions in daily in-home and out-of-home maintenance activity participation. *Transportation*, 32:523–544, 2005.

[5] K. Axhausen. Social networks, mobility biographies, and travel: survey challenges. *Enviroment and Planning B*, 35(6):981–996, 2008.

[6] J. Carrasco, B. Hogan, B. Wellman, and E. Miller. Collecting social network data to study social activity-travel behavior: an egocentric approach. *Enviroment and Planning B*, 35(6):961–980, 2008.

[7] N. Tilahun and D. Levinson. Work and home location: Possible role of social networks. *Transportation Research Part A*, 45(4):323–331, 2011.

[8] J. Hammersley and D. Handscomb. *Monte Carlo Methods*. Wiley, 1964.

[9] J. Johnston and J. DiNardo. *Econometric methods*. McGraw-Hill, 1997.

[10] W. Greene. *Econometric Analysis*. Prentice-Hall, 7th edition, 2012.

[11] D. Fiebig, M. Keane, J. Louviere, and N. Wasi. The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. *Marketing Science*, 29(3):393–421, 2010.

[12] M. Bradley and A. Daly. New analysis issues in stated preference research. In J. Ortuzar, editor, *Stated preference modeling techniques*. PTRC Education and Research Services, 2000.

[13] R. Barff, D. Mackay, and R. Olshavsky. A selective review of travel-mode choice models. *Journal of Consumer Research*, 8(4):370–380, 1982.

[14] F. De Donnea. *Determinants of Transportation Mode Choice in Dutch Cities*. University Press, 1971.

[15] D. Quarmby. Choice of travel mode for the journey to work: Some findings. *Journal of Transport Economics and Policy*, 1(3):273–314, 1967.

[16] M. Ben-Akiva. The structure of passenger demand models. *Highway Research Record*, 526:26–42, 1974.

[17] S. Lerman and M. Ben-Akiva. Disaggregate behavioral model of auto-ownership. *Transportation Research Record*, 569:34–51, 1976.

[18] E. Dugundji, A. Paez, and T. Arentze. Social networks, choices, mobility, and travel. *Enviroment and Planning B*, 35(6):956–960, 2008.

[19] E. Dugundji and L. Gulyas. An exploration of the role of global versus local and social versus spatial networks in transportation mode choice behavior in the netherlands. In *Proceedings of AGENT 2003: Challenges in Social Simulation, Argonne National Laboratory, Chicago, USA.*, 2003.

[20] E. Dugundji and L. Gulyas. Socio-dynamic discrete choice on networks in space: impacts of agent heterogeneity on emergent outcomes. *Enviroment and Planning B*, 35(6):1028–1054, 2008.

[21] J. Walker, E. Ehlers, I. Banerjee, and E. Dugundji. Correcting for endogeneity in behavioral choice models with social influence variables. *Transportation Research Part A*, 45:362–374, 2011.

[22] F. Goetzke and T. Rave. Bicycle use in germany: explaining differences between municipalities with social network effects. *Urban Studies*, 48(2):427–437, 2011.

[23] E. Dugundji and J. Walker. Discrete choice with social and spatial network interdependencies: an empirical example using mixed generalized extreme value models with field and panel effects. *Transportation Research Record*, 1921:70–78, 2005.

[24] M. Bierlaire, K. Axhausen, and G. Abay. The acceptance of modal innovation: The case of swissmetro. In *1st Swiss Transport Research Conference*, 2001.

[25] J.J. Louviere, D.A. Hensher, and J.D. Swait. *Stated choice methods: analysis and applications*. Cambridge University Press, 2000.

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

[27] 2006-2008 american community survey 3-year estimates, minneapolis-st. paul-bloomington, mn-wi metropolitan statistical area, retrieved november 25, 2009. URL http://factfinder.census.gov/.

[28] M. Ben-Akiva and S. Lerman. *Discrete choice analysis: theory and application to travel demand*. MIT Press, 1985.

[29] A. Agresti. *Categorical Data Analysis*. Wiley Series in Probability and Statistics, 2nd edition, 2002.

[30] C. Hsiao. *Analysis of Panel Data*. Cambridge Univ. Press, 2nd edition, 2003.

[31] K. Train. *Discrete choice methods with simulation*. Cambridge University Press, 2nd edition, 2009.

[32] A. Hole. Fitting mixed logit models using maximum simulated likelihood. *The Stata Journal*, 7: 388–401, 2007.

[33] P. K. Trivedi and A. C. Cameron. *Microeconometrics: Methods and Applications*. Cambridge Univ. Press, 2005.

[34] A. Daly and S. Hess. Simple approaches for random utility modelling with panel data. In *Presented at the European Transport Conference, Glasgow, UK*, 2010.

[35] J. Walker. *Extended discrete choice models: integrated framework, flexible error structures, and latent variables*. PhD thesis, Massachusetts Institute of Technology (USA), 2001.

[36] J. Ortuzar and L. Willumsen. *Modelling Transport*. Wiley, 4th edition, 2011.

[37] C. Baum. *An introduction to Stata programming*. Stata Press, 2009.

[38] J. Cramer. *Econometric applications of Maximum Likelihood methods*. Cambridge Univ. Press, 1986.

Table 3: Random Utility Models

Variables	Mixed logit (Random effects)	Multinomial logit (Bootstrap Std. Errors)
	Estimates (T-Stats)	Estimates (T-Stats)
Ratio - Bike to Auto - [Bike/Walking] Alternative Specific Variable (ASV) It is the ratio of the SP bike/walking share to the auto share	1.10 (0.51)	-6.76 (-1.53)
Ratio - Transit to Auto - [Transit] Alternative Specific Variable (ASV) It is the ratio of the SP transit share to the auto share	4.48 (1.74) **	-3.83 (-2.78) **
Biking Preference - [Bike/Walking] Alternative Specific Variable (ASV) Dummy variable indicating whether subjects have never biked to work	-7.05(-2.93) **	-2.00 (-2.62) **
Biking Preference - [Transit] Alternative Specific Variable (ASV) Dummy variable indicating whether subjects have never biked to work	-4.10 (-1.88) **	-1.23 (-1.88) **
Vehicles per adults - [Bike/Walking] Alternative Specific Variable (ASV) Number of vehicles per adults in the household	-4.38 (-1.90) **	-1.14 (-1.34)
Vehicles per adults - [Transit] Alternative Specific Variable (ASV) Number of vehicles per adults in the household	-1.82 (-0.83)	-0.51 (-0.88)
Age $[40, 50)$ - [Bike/Walking] Alternative Specific Variable (ASV) Dummy variable indicating whether subjects have ages of $[40, 50)$	4.83 (2.51) **	1.28 (2.04) **
Age $[40, 50)$ - [Transit] Alternative Specific Variable (ASV) Dummy variable indicating whether subjects have ages of $[40, 50)$	2.81 (1.44)	0.90 (1.54)
Alternative Specific Constant for Bike/Walking Alternative Specific Variable (ASV): Intercept	7.48 (2.48) **	3.36 (3.19) **
Standard Deviation for Bike/Walking Alternative Specific Variable (ASV): Random Effect for Bike/Walking	4.15 (4.08) ***	
Alternative Specific Constant for Transit Alternative Specific Variable (ASV) Intercept	2.14 (0.70)	2.39 (2.60) **
Standard Deviation for Transit Alternative Specific Variable (ASV): Random Effect for Transit	5.98 (3.00) ***	
Intercept Log-Likelihood $ll_{A\hat{S}C}$	-336.14938	-336.14938
Final Log-Likelihood $ll_{\hat{\beta}}$	-208.19424	-252.98452
Likelihood ratio index ρ^2	0.38064964	0.24740447
Number of observations	308	308
Number of subjects	77	77

* is 10% significance level, ** is 5% significance level, *** is 1% significance level

See the section 3.4.1 for details on the econometric models.

Table 4: Monte Carlo Simulations - No heterogeneity (1000 Replications - Panel data)

No heterogeneity **70 Subjects** **280 Observations**	**True parameters**	**No heterogeneity** **(Multinomial logit)**	**Heterogeneity (Mixed logit)** **Random intercepts**	 **Random coefficients**	 **Random intercepts and coefficients**
$\beta_{\mathbf{ASC}}^{\mathbf{Bike/Walking}}$	2.00	1.98 CI-95% [1.962, 2.001]	1.82 CI-95% [1.794, 1.84]	1.85 CI-95% [1.831, 1.874]	1.69 CI-95% [1.663, 1.714]
$\beta_{\mathbf{ASC}}^{\mathbf{Transit}}$	1.00	0.97 CI-95% [0.940, 0.997]	0.59 CI-95% [0.549, 0.635]	0.72 CI-95% [0.682, 0.759]	0.37 CI-95% [0.319, 0.418]
$\beta_{\mathbf{Ratio-Bike/Walking-to-Auto}}^{\mathbf{Bike/Walking}}$	6.00	6.77 CI-95% [6.573, 6.966]	10.34 CI-95% [9.969, 10.702]	9.58 CI-95% [9.260, 9.904]	13.32 CI-95% [12.846, 13.786]
$\beta_{\mathbf{Ratio-Transit-to-Auto}}^{\mathbf{Transit}}$	4.00	4.30 CI-95% [4.127, 4.48]	7.04 CI-95% [6.744, 7.343]	6.14 CI-95% [5.876, 6.399]	8.77 CI-95% [8.420, 9.129]
$\sigma_{\mathbf{ASC}}^{\mathbf{Bike/Walking}}$	0.00		0.25 CI-95% [0.221, 0.270]		0.28 CI-95% [0.249, 0.304]
$\sigma_{\mathbf{ASC}}^{\mathbf{Transit}}$	0.00		0.23 CI-95% [0.206, 0.257]		0.21 CI-95% [0.180, 0.232]
$\sigma_{\mathbf{Ratio-Bike/Walking-to-Auto}}^{\mathbf{Bike/Walking}}$	0.00			1.30 CI-95% [1.098, 1.502]	0.62 CI-95% [0.428, 0.820]
$\sigma_{\mathbf{Ratio-Transit-to-Auto}}^{\mathbf{Transit}}$	0.00			1.32 CI-95% [1.112, 1.530]	0.84 CI-95% [0.589, 1.090]
No heterogeneity **500 Subjects** **2000 Observations**	**True parameters**	**No heterogeneity** **(Multinomial logit)**	**Heterogeneity (Mixed logit)** **Random intercepts**	 **Random coefficients**	 **Random intercepts and coefficients**
$\beta_{\mathbf{ASC}}^{\mathbf{Bike/Walking}}$	2.00	1.97 CI-95% [1.946, 1.985]	1.80 CI-95% [1.777, 1.822]	1.84 CI-95% [1.819, 1.862]	1.69 CI-95% [1.660, 1.710]
$\beta_{\mathbf{ASC}}^{\mathbf{Transit}}$	1.00	0.97 CI-95% [0.940, 0.997]	0.57 CI-95% [0.527, 0.611]	0.69 CI-95% [0.661, 0.736]	0.35 CI-95% [0.304, 0.401]
$\beta_{\mathbf{Ratio-Bike/Walking-to-Auto}}^{\mathbf{Bike/Walking}}$	6.00	7.02 CI-95% [6.817, 7.231]	10.60 CI-95% [10.244, 10.967]	9.71 CI-95% [9.38, 10.04]	13.22 CI-95% [12.765, 13.667]
$\beta_{\mathbf{Ratio-Transit-to-Auto}}^{\mathbf{Transit}}$	4.00	4.42 CI-95% [4.261, 4.588]	7.23 CI-95% [6.943, 7.524]	6.34 CI-95% [6.080, 4.594]	8.91 CI-95% [8.562, 9.266]
$\sigma_{\mathbf{ASC}}^{\mathbf{Bike/Walking}}$	0.00		0.25 CI-95% [0.222, 0.269]		0.26 CI-95% [0.236, 0.289]
$\sigma_{\mathbf{ASC}}^{\mathbf{Transit}}$	0.00		0.25 CI-95% [0.222, 0.274]		0.22 CI-95% [0.193, 0.245]
$\sigma_{\mathbf{Ratio-Bike/Walking-to-Auto}}^{\mathbf{Bike/Walking}}$	0.00			1.34 CI-95% [1.132, 1.553]	0.60 CI-95% [0.396, 0.809]
$\sigma_{\mathbf{Ratio-Transit-to-Auto}}^{\mathbf{Transit}}$	0.00			1.07 CI-95% [0.881, 1.26]	0.95 CI-95% [0.734, 1.176]

See the section 4 for details on the Monte Carlo Simulations.

Table 5: Monte Carlo Simulations - Random intercepts (1000 Replications - Panel data)

Random intercepts **70 Subjects** **280 Observations**	**True parameters**	**No heterogeneity** **(Multinomial logit)**	**Heterogeneity (Mixed logit)** **Random intercepts**	**Random coefficients**	**Random intercepts and coefficients**
$\beta_{\mathrm{ASC}}^{\mathrm{Bike/Walking}}$	2.00	2.66 CI-95% [2.644, 2.685]	1.96 CI-95% [1.936, 1.992]	2.41 CI-95% [2.388, 2.436]	1.85 CI-95% [1.814, 1.877]
$\beta_{\mathrm{ASC}}^{\mathrm{Transit}}$	1.00	2.12 CI-95% [2.092, 2.152]	0.64 CI-95% [0.592, 0.686]	1.50 CI-95% [1.453, 1.549]	0.16 CI-95% [0.0986, 0.229]
$\beta_{\mathrm{Ratio-Bike/Walking-to-Auto}}^{\mathrm{Bike/Walking}}$	6.00	-6.32 CI-95% [-6.451, -6.191]	7.68 CI-95% [7.332, 8.032]	-2.03 CI-95% [-2.358, -1.699]	12.35 CI-95% [11.824, 12.876]
$\beta_{\mathrm{Ratio-Transit-to-Auto}}^{\mathrm{Transit}}$	4.00	-4.12 CI-95% [-4.263, -3.968]	6.69 CI-95% [6.350, 7.035]	-0.23 CI-95% [-0.544, 0.084]	10.94 CI-95% [10.424, 11.458]
$\sigma_{\mathrm{ASC}}^{\mathrm{Bike/Walking}}$	2.00		2.07 CI-95% [2.024, 2.114]		2.30 CI-95% [2.248, 2.355]
$\sigma_{\mathrm{ASC}}^{\mathrm{Transit}}$	0.50		0.34 CI-95% [0.284, 0.400]		0.39 CI-95% [0.319, 0.460]
$\sigma_{\mathrm{Ratio-Bike/Walking-to-Auto}}^{\mathrm{Bike/Walking}}$	0.00			3.94 CI-95% [3.478, 4.394]	0.21 CI-95% [-0.153, 0.580]
$\sigma_{\mathrm{Ratio-Transit-to-Auto}}^{\mathrm{Transit}}$	0.00			3.72 CI-95% [3.121, 4.322]	0.67 CI-95% [0.0749, 1.263]

Random intercepts **500 Subjects** **2000 Observations**	**True parameters**	**No heterogeneity** **(Multinomial logit)**	**Heterogeneity (Mixed logit)** **Random intercepts**	**Random coefficients**	**Random intercepts and coefficients**
$\beta_{\mathrm{ASC}}^{\mathrm{Bike/Walking}}$	2.00	2.68 CI-95% [2.665, 2.708]	1.99 CI-95% [1.956, 2.015]	2.45 CI-95% [2.424, 2.472]	1.86 CI-95% [1.829, 1.899]
$\beta_{\mathrm{ASC}}^{\mathrm{Transit}}$	1.00	2.13 CI-95% [2.110, 2.167]	0.67 CI-95% [0.632, 0.715]	1.54 CI-95% [1.502, 1.586]	0.22 CI-95% [0.168, 0.278]
$\beta_{\mathrm{Ratio-Bike/Walking-to-Auto}}^{\mathrm{Bike/Walking}}$	6.00	-6.46 CI-95% [-6.603, -6.326]	7.57 CI-95% [10.244, 10.967]	-2.39 CI-95% [-2.721, -2.056]	12.41 CI-95% [11.846, 12.971]
$\beta_{\mathrm{Ratio-Transit-to-Auto}}^{\mathrm{Transit}}$	4.00	-4.16 CI-95% [-4.299, -4.016]	6.50 CI-95% [6.199, 6.804]	-0.47 CI-95% [-0.744, -0.191]	10.53 CI-95% [10.094, 10.963]
$\sigma_{\mathrm{ASC}}^{\mathrm{Bike/Walking}}$	2.00		2.09 CI-95% [0.222, 0.269]		2.31 CI-95% [2.258, 2.364]
$\sigma_{\mathrm{ASC}}^{\mathrm{Transit}}$	0.50		0.30 CI-95% [0.243, 0.357]		0.36 CI-95% [0.293, 0.432]
$\sigma_{\mathrm{Ratio-Bike/Walking-to-Auto}}^{\mathrm{Bike/Walking}}$	0.00			3.31 CI-95% [2.853, 3.767]	0.18 CI-95% [-0.158, 0.512]
$\sigma_{\mathrm{Ratio-Transit-to-Auto}}^{\mathrm{Transit}}$	0.00			3.65 CI-95% [3.0495, 4.250]	0.78 CI-95% [0.128, 1.424]

See the section 4 for details on the Monte Carlo Simulations.

Table 6: Monte Carlo Simulations - Random coefficients (1000 Replications - Panel data)

Random coefficients **70 Subjects** **280 Observations**	**True parameters**	**No heterogeneity (Multinomial logit)**	**Heterogeneity (Mixed logit)** **Random intercepts**	**Random coefficients**	**Random intercepts and coefficients**
$\beta^{\mathbf{Bike/Walking}}_{\mathbf{ASC}}$	2.00	2.097 CI-95% [2.078, 2.115]	1.96 CI-95% [1.940, 1.983]	1.87 CI-95% [1.847, 1.894]	1.72 CI-95% [1.692, 1.747]
$\beta^{\mathbf{Transit}}_{\mathbf{ASC}}$	1.00	1.07 CI-95% [1.044, 1.099]	0.74 CI-95% [0.700, 0.778]	0.69 CI-95% [0.645, 0.727]	0.36 CI-95% [0.312, 0.412]
$\beta^{\mathbf{Bike/Walking}}_{\mathbf{Ratio-Bike/Walking-to-Auto}}$	6.00	4.40 CI-95% [4.196, 4.613]	7.28 CI-95% [6.919, 7.638]	9.39 CI-95% [8.973, 9.809]	12.84 CI-95% [12.29, 13.39]
$\beta^{\mathbf{Transit}}_{\mathbf{Ratio-Transit-to-Auto}}$	4.00	3.61 CI-95% [3.446, 3.781]	6.00 CI-95% [5.729, 6.270]	6.50 CI-95% [6.204, 6.787]	8.92 CI-95% [8.555, 9.290]
$\sigma^{\mathbf{Bike/Walking}}_{\mathbf{ASC}}$	0.00		0.20 CI-95% [0.181, 0.227]		0.22 CI-95% [0.199, 0.257]
$\sigma^{\mathbf{Transit}}_{\mathbf{ASC}}$	0.00		0.22 CI-95% [0.197, 0.246]		0.21 CI-95% [0.181, 0.234]
$\sigma^{\mathbf{Bike/Walking}}_{\mathbf{Ratio-Bike/Walking-to-Auto}}$	4.00			3.38 CI-95% [3.0734, 3.688]	2.69 CI-95% [2.314, 3.069]
$\sigma^{\mathbf{Transit}}_{\mathbf{Ratio-Transit-to-Auto}}$	0.50			1.81 CI-95% [1.578, 2.035]	0.77 CI-95% [0.538, 1.000]
Random coefficients **500 Subjects** **2000 Observations**	**True parameters**	**No heterogeneity (Multinomial logit)**	**Heterogeneity (Mixed logit)** **Random intercepts**	**Random coefficients**	**Random intercepts and coefficients**
$\beta^{\mathbf{Bike/Walking}}_{\mathbf{ASC}}$	2.00	2.11 CI-95% [2.0896, 2.128]	1.98 CI-95% [1.961, 2.00]	1.90 CI-95% [1.882, 1.928]	1.75 CI-95% [1.725, 1.779]
$\beta^{\mathbf{Transit}}_{\mathbf{ASC}}$	1.00	1.07 CI-95% [1.046, 1.101]	0.76 CI-95% [0.727, 0.800]	0.72 CI-95% [0.673, 0.751]	0.38 CI-95% [0.338, 0.433]
$\beta^{\mathbf{Bike/Walking}}_{\mathbf{Ratio-Bike/Walking-to-Auto}}$	6.00	7.04 CI-95% [4.187, 4.586]	7.28 CI-95% [6.713, 7.372]	8.81 CI-95% [8.44, 9.175]	12.29 CI-95% [11.792, 12.794]
$\beta^{\mathbf{Transit}}_{\mathbf{Ratio-Transit-to-Auto}}$	4.00	3.69 CI-95% [3.524, 3.850]	5.90 CI-95% [5.652, 6.156]	6.38 CI-95% [6.117, 6.651]	8.82 CI-95% [8.479, 9.168]
$\sigma^{\mathbf{Bike/Walking}}_{\mathbf{ASC}}$	0.00		0.19 CI-95% [0.181, 0.227]		0.22 CI-95% [0.197, 0.252]
$\sigma^{\mathbf{Transit}}_{\mathbf{ASC}}$	0.00		0.20 CI-95% [0.197, 0.246]		0.21 CI-95% [0.178, 0.230]
$\sigma^{\mathbf{Bike/Walking}}_{\mathbf{Ratio-Bike/Walking-to-Auto}}$	4.00			2.93 CI-95% [2.642, 3.214]	2.77 CI-95% [2.423, 3.12]
$\sigma^{\mathbf{Transit}}_{\mathbf{Ratio-Transit-to-Auto}}$	0.50			1.64 CI-95% [1.428, 1.847]	0.77 CI-95% [0.551, 0.977]

See the section 4 for details on the Monte Carlo Simulations.

Table 7: Monte Carlo Simulations - Random intercepts and coefficients (1000 Replications - Panel data)

Random intercepts and coefficients 70 Subjects - 280 Observations	**True parameters**	**No heterogeneity (Multinomial logit)**	**Heterogeneity (Mixed logit)** Random intercepts	Random coefficients	Random intercepts and coefficients
$\beta_{\mathbf{ASC}}^{\mathbf{Bike/Walking}}$	2.00	2.72 CI-95% [2.078, 2.115]	2.05 CI-95% [2.018, 2.074]	2.45 CI-95% [2.426, 2.475]	1.90 CI-95% [1.867, 1.934]
$\beta_{\mathbf{ASC}}^{\mathbf{Transit}}$	1.00	2.19 CI-95% [2.162, 2.217]	0.82 CI-95% [0.776, 0.856]	1.50 CI-95% [1.452, 1.539]	0.33 CI-95% [0.276, 0.412]
$\beta_{\mathbf{Ratio-Bike/Walking-to-Auto}}^{\mathbf{Bike/Walking}}$	6.00	-6.98 CI-95% [-7.116, -6.843]	6.07 CI-95% [5.746, 6.391]	-2.12 CI-95% [-2.489, -1.754]	11.47 CI-95% [10.915, 12.029]
$\beta_{\mathbf{Ratio-Transit-to-Auto}}^{\mathbf{Transit}}$	4.00	-4.37 CI-95% [-4.502, -4.225]	5.55 CI-95% [5.255, 5.841]	0.063 CI-95% [-0.236, 0.363]	9.91 CI-95% [9.460, 10.360]
$\sigma_{\mathbf{ASC}}^{\mathbf{Bike/Walking}}$	2.00		2.01 CI-95% [1.967, 2.050]		2.24 CI-95% [2.181, 2.297]
$\sigma_{\mathbf{ASC}}^{\mathbf{Transit}}$	0.50		0.31 CI-95% [0.260, 0.366]		0.38 CI-95% [0.309, 0.444]
$\sigma_{\mathbf{Ratio-Bike/Walking-to-Auto}}^{\mathbf{Bike/Walking}}$	4.00			5,04 CI-95% [4.352, 5.739]	1.27 CI-95% [0.566, 1.979]
$\sigma_{\mathbf{Ratio-Transit-to-Auto}}^{\mathbf{Transit}}$	0.50			3.30 CI-95% [2.827, 3.777]	0.50 CI-95% [0.166, 0.837]
Random intercepts and coefficients 500 Subjects - 2000 Observations	**True parameters**	**No heterogeneity (Multinomial logit)**	**Heterogeneity (Mixed logit)** Random intercepts	Random coefficients	Random intercepts and coefficients
$\beta_{\mathbf{ASC}}^{\mathbf{Bike/Walking}}$	2.00	2.71 CI-95% [2.688, 2.730]	2.03 CI-95% [2.018, 2.074]	2.44 CI-95% [2.416, 2.464]	1.88 CI-95% [1.848, 1.918]
$\beta_{\mathbf{ASC}}^{\mathbf{Transit}}$	1.00	2.16 CI-95% [2.127, 2.184]	0.73 CI-95% [0.684, 0.771]	1.48 CI-95% [1.439, 1.527]	0.22 CI-95% [0.161, 0.276]
$\beta_{\mathbf{Ratio-Bike/Walking-to-Auto}}^{\mathbf{Bike/Walking}}$	6.00	-6.87 CI-95% [-6.998, -6.734]	6.36 CI-95% [6.020, 6.694]	-2.19 CI-95% [-2.554, -1.831]	11.98 CI-95% [11.389, 12.575]
$\beta_{\mathbf{Ratio-Transit-to-Auto}}^{\mathbf{Transit}}$	4.00	-4.25 CI-95% [-4.391, -4.107]	5.98 CI-95% [5.671, 6.284]	-0.039 CI-95% [-0.335, 0.256]	10.53 CI-95% [10.064, 10.992]
$\sigma_{\mathbf{ASC}}^{\mathbf{Bike/Walking}}$	2.00		2.00 CI-95% [1.960, 2.045]		2.27 CI-95% [2.210, 2.320]
$\sigma_{\mathbf{ASC}}^{\mathbf{Transit}}$	0.50		0.38 CI-95% [0.321, 0.432]		0.42 CI-95% [0.352, 0.494]
$\sigma_{\mathbf{Ratio-Bike/Walking-to-Auto}}^{\mathbf{Bike/Walking}}$	4.00			4.61 CI-95% [3.942, 5.274]	0.73 CI-95% [0.015, 1.443]
$\sigma_{\mathbf{Ratio-Transit-to-Auto}}^{\mathbf{Transit}}$	0.50			3.48 CI-95% [3.002, 3.951]	-0.06 CI-95% [-0.404, 0.294]

See the section 4 for details on the Monte Carlo Simulations.