Journal of Choice Modelling 34 (2020) 100200

Contents lists available at ScienceDirect

Journal of Choice Modelling

journal homepage: http://www.elsevier.com/locate/jocm

Endogeneity in adaptive choice contexts: Choice-based recommender systems and adaptive stated preferences surveys

Mazen Danaf[a,*], Angelo Guevara[b,e], Bilge Atasoy[c], Moshe Ben-Akiva[d]

[a] Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, 02139, USA
[b] Civil Engineering Department and Instituto Sistemas Complejos de Ingeniería (ISCI), Universidad de Chile, Blanco Encalada 2002, Santiago, Chile
[c] Department of Maritime and Transport Technology, Delft University of Technology, Mekelweg 2, Delft, 2628 CD, the Netherlands
[d] Edmund K. Turner Professor of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, 02139, USA
[e] Instituto Sistemas Complejos de Ingeniería (ISCI), República 695, Santiago, Chile

ARTICLE INFO

Keywords:
Discrete choice models
Endogeneity
Adaptive stated preferences surveys
Choice-based recommender systems

ABSTRACT

Endogeneity arises in discrete choice models due to several factors and results in inconsistent estimates of the model parameters. In adaptive choice contexts such as choice-based recommender systems and adaptive stated preferences (ASP) surveys, endogeneity is expected because the attributes presented to an individual in a specific menu (or choice situation) depend on the previous choices of the same individual (as well as the alternative attributes in the previous menus). Nevertheless, the literature is indecisive on whether the parameter estimates in such cases are consistent or not. In this paper, we discuss cases where the estimates are consistent and those where they are not. We provide a theoretical explanation for this discrepancy and discuss the implications on the design of these systems and on model estimation. We conclude that endogeneity is not a concern when the likelihood function properly accounts for the data generation process. This can be achieved when the system is initialized exogenously and all the data are used in the estimation. In line with previous literature, Monte Carlo results suggest that, even when exogenous initialization is missing, empirical bias decreases with the number of choices per individual. We conclude by discussing the practical implications and extensions of this research.

1. Introduction

Endogeneity can arise in discrete choice models due to several factors including measurement errors, selection bias, omitted variables, and simultaneity, and results in inconsistent estimates of the model parameters (Guevara, 2015). The textbook definition of endogeneity is a correlation between the independent/observed variables in the model and the unobserved error term. A broader definition of endogeneity has been provided by Louviere et al. (2005), in which they defined “endogenous” as “all effects that are not exogenous”. In the latter work, the authors attribute endogeneity to model misspecification.

In linear regression models, different methods have been proposed to address endogeneity, the most common of which are instrumental variables (IV) and Two-Stage Least Squares (TSLS). These methods use instruments that are correlated with the endogenous variables, but not with the error term. Other corrections include Heckman correction for selection bias (Heckman, 1979),

* Corresponding author.
E-mail addresses: mdanaf@mit.edu (M. Danaf), crguevar@ing.uchile.cl (A. Guevara), b.atasoy@tudelft.nl (B. Atasoy), mba@mit.edu (M. Ben-Akiva).

https://doi.org/10.1016/j.jocm.2019.100200
Received 27 February 2019; Received in revised form 26 September 2019; Accepted 29 December 2019
Available online 2 January 2020

and the estimation of simultaneous equation models.

In discrete choice models, several corrections have been proposed that mainly fall into two categories; the BLP method (Berry et al., 1995), and the control-function method (Heckman, 1977; Hausman, 1978). The BLP method (also known as the "product-market" control approach) suggests aggregating disaggregate consumer preferences obtained from discrete choice models into an aggregate market-level system, allowing for the application of standard instrumental variable methods. On the other hand, control-function methods use extra variables in the utility specification that are obtained using exogenous instruments. Different control-functions have been proposed, the most common of which are by Petrin and Train (2010), Villas-Boas and Winer (1999), Blundell and Powell (2004), Guevara and Ben-Akiva (2006, 2012), and Guevara and Polanco (2016).

These methods are convenient when a few endogenous variables are identified and relevant instruments are available. However, the cases considered in this paper are of a different nature; we consider adaptive choice contexts such as adaptive stated preferences (ASP) surveys and choice-based recommender systems. When individuals are presented with successive menus (or choice situations), the alternatives and attributes in each menu depend on the attributes and choices in the previous menus. Recommender systems recommend the "best" alternatives to a user (e.g. alternatives that are most likely to be chosen), while ASP surveys usually do the opposite; the attributes of the chosen alternative are deteriorated in order to test whether a respondent will switch.

In the presence of taste variation (e.g. logit mixture with random parameters), the distributions of heterogeneity are assumed to be uncorrelated with the covariates (see Wooldridge, 2010). To illustrate the source of the endogeneity that may arise in adaptive choice contexts, we consider the simple logit mixture model shown in equation (1), where individual n's utility of alternative j includes an alternative specific constant (β_j) and one attribute with a random parameter β_n:

$$\begin{aligned} U_{jn} &= \beta_j + \beta_n X_{jn} + \varepsilon_{jn} \\ &= \beta_j + (\overline{\beta} + \eta_n \sigma) X_{jn} + \varepsilon_{jn} \end{aligned} \tag{1}$$

where $\beta_n \sim N(\overline{\beta}, \sigma^2)$, and η_n is a standard normal random variable independently and identically distributed across individuals.

Endogeneity arises as a result of the correlation between the unobserved heterogeneity (η_n) and the independent variable (X_{jn}). In recommender systems, individuals with a (supposedly) higher preference for the attribute X (i.e., positive and high value of η_n) are recommended alternatives having higher values of this attribute. In turn, in ASP surveys, individuals with a (supposedly) higher preference for X are usually presented with a lower value of that attribute in their following choice tasks. If recommendations/choice tasks are generated based on multiple attributes, the above reasoning can be extended to conclude that endogeneity arises in all attributes (as X_{jn} and β_n will be multidimensional).

Under the context described before for recommender systems or ASP surveys, endogeneity can also be explained by model misspecification. The values of X_{jn} depend on the user's previous choices (and thus on this user's preferences β_n). A correctly specified model should account for the joint likelihood of the choices (d_n) and the observed attributes (X_{jn}) as shown in equation (2). On the other hand, models that do not account for this dependency ($P(X_n|\beta_n)$) will be misspecified and might result in biased estimates.

$$P(d_n, X_n|\beta_n) = P(d_n|X_n, \beta_n)P(X_n|\beta_n) \tag{2}$$

In this paper, we extend the theoretical analyses of Liu et al. (2007) on "adaptive metric utility balance" to choice contexts to demonstrate how endogeneity can cause inconsistent estimation results, and how this inconsistency can be avoided. We show that when the system is initialized with exogenous attributes, and when all menus (or choice tasks) are included in the estimation, the estimates are consistent. On the other hand, excluding data from the estimation leads to inconsistent estimates. We note that this paper only addresses endogeneity that arises as a result of the adaptive nature of these contexts, however, there might be other sources of endogeneity that are outside the scope of this paper, such as measurement errors, omitted variables, and self-selection.

The remainder of this paper is organized as follows. Section 2 presents an overview of adaptive contexts such as recommender systems and ASP surveys. Section 3 presents the methodology used in adaptive choice scenarios and a theoretical analysis of endogeneity. A Monte Carlo experiment mimicking a dynamic recommender system is presented in Section 4. Section 5 discusses the practical implications on recommender systems, ASP surveys, and RP/SP estimation. Finally, Section 6 concludes the paper.

2. Background

This paper primarily focuses on adaptive choice contexts whereby endogeneity might arise because the attributes of alternatives presented to a user are determined by the previous choices done by the same user. In such applications, all attributes are assumed to be endogenous. We start by discussing two applications of this nature; choice based-recommender systems and adaptive stated preferences surveys. In addition, we briefly discuss adaptive metric utility experiments.

2.1. Choice-based recommender systems

Discrete choice models are used in recommender systems due to their ability to integrate item specific, user specific, and contextual data in a single model (Chaptini, 2005; Danaf et al., 2019a; Jiang et al., 2014; Polydoropoulou and Lambrou, 2012). These applications are generally based on estimating individual-level parameters, and using them in menu/assortment optimization.

In traditional applications of choice-based recommender systems, individual preferences are estimated beforehand (offline) and used in generating personalized recommendations. For example, Chaptini (2005) presents an online academic advisor for MIT students that recommends academic courses based on observed and latent attributes of the courses (e.g. difficulty, workload, overall impression, etc.). The model is estimated with data collected via an online revealed preferences (RP)/stated preferences (SP) survey, and the estimated individual level parameters are used in generating course recommendations.

A similar approach is used by Ansari et al. (2000), who present a Hierarchical Bayes approach to movie recommendations on the internet. The underlying model accounts for systematic (observed) and random (unobserved) heterogeneity in user preferences. Customer ratings are modeled as a function of product attributes, customer characteristics, and expert evaluations. Rubin and Steyvers (2009) present another application to movie recommendations, modeling the process by which an individual selects and later rates an item. A Latent Dirichlet Allocation (LDA) model is used to model the probability of selecting a movie given a set of movies classified by topic, and an ordered logit model (with an individual-specific error term) is used to model movie ratings.

The above applications are static, and individual-level parameters are estimated beforehand using SP or RP data. More recently, dynamic online applications of choice models in recommender systems have been proposed. Teo et al. (2016) present a choice-based recommender system used by Amazon Stream, whereby a probit model is estimated at the individual level using the users' click data. Individual preferences are learnt from repeated choices, and updated dynamically using Bayesian estimation.

Estimating models at the individual level is feasible when the number of choices per individual is large (e.g. click data). However, long panels are not available in many applications. To solve this problem, Danaf et al. (2019a) propose a methodology for estimating and updating individual-specific preferences in real-time, using population-level parameters as priors. The underlying model accounts for inter- and intra-consumer heterogeneity (corresponding to random taste variations among individuals and among different choices of the same individual respectively). The estimation method extends the three-step Gibbs sampler of the logit mixture model (Train, 2009) to a five-step Gibbs sampler (Becker et al., 2018) to account for intra-consumer heterogeneity. A computationally efficient Bayesian estimation is used to update preferences (i.e., individual specific coefficients and their distributions) after each choice, while population-level parameters are updated offline periodically. Related work by Song et al. (2017, 2018) demonstrates how the estimated individual-specific coefficients are used in generating personalized recommendations through online assortment optimization. This model maximizes hit-rate (i.e., the probability of choosing one of the recommended alternatives[1]) or consumer-surplus (CS) in the form of log-sum, subject to constraints on the maximum number of alternatives to be shown in a menu. While this method allows for estimating individual preferences with a small number of choices per individual, dynamic recommendations coupled with taste heterogeneity can result in endogeneity bias.

2.2. Adaptive SP surveys

In stated choice experiments, a respondent is presented with hypothetical alternatives and asked to choose one among those. Efficient designs, which have been proposed to reduce the standard errors of the estimated parameters, rely heavily on priors (Walker et al., 2018). To overcome this problem, Kanninen (2002) and Johnson et al. (2006, 2013) suggested updating the design during the data collection phase as knowledge of the true parameters increases. With increasing computational capabilities, adaptive stated preferences surveys (ASP) have been proposed in which the attributes presented to the users are determined based on the choices they have already made. Different studies have investigated endogeneity in ASP surveys, however they do not have a consensus on whether the estimation results are consistent in such applications or not.

Several studies have found that ASP surveys might result in inconsistent or inefficient estimates. Bradley and Daly (1993) analyzed several variations of ASP designs and concluded that endogenous SP designs might result in bias in the presence of taste variation in the sample. To counter this phenomenon, the authors suggested a few remedies such as (1) using market segmentation in order to ensure that each segment is as homogenous as possible, (2) using a Fixed/Adaptive approach (which uses a fixed basic design but avoids presenting certain choice options if they are deemed redundant based on prior choices), or (3) using an exogenous adaptive design in which an exogenous variable is used to adapt the design levels prior to the experiment. Similarly, Toubia et al. (2003) and Abernethy et al. (2008), who developed polyhedral methods for survey designs that "reduce the feasible set of parameters as rapidly as possible", indicated that these methods are susceptible to endogeneity bias. However, Abernethy et al. (2008) observed that the magnitude of bias decreases with more questions per individual.

Fowkes (2007) analyzed endogeneity bias in the Leeds Adaptive Stated Preferences (LASP) survey (Fowkes and Shinghal, 2002; Shinghal, 1999), and found significant bias when the models are calibrated over several respondents, but not at the level of the respondent. In line with the findings of Abernethy et al. (2008), Fowkes (2007) found that the accuracy of an estimate improves and the apparent bias is reduced as more questions are added (due to alterations to the design), and concluded that LASP is, what he termed, "asymptotically unbiased".

In other applications, the estimates obtained from ASP surveys were found to be consistent. Richardson (2002) developed a simulation study to estimate individual-specific values of time (VOT) using an ASP Survey, and concluded that this method produces unbiased estimates of VOT. Similarly, Tilahun et al. (2007) used an ASP survey in order to evaluate individual preferences for different cycling environments in Minnesota, and concluded that ASP surveys allow for measuring the exact values individuals attach to attributes of interest.

Related work has been done on SP surveys constructed based on RP surveys (SP-off-RP and pivoted SP designs) by Train and Wilson

[1] In the context of recommender systems, hit-rate refers to the probability of choosing one of the recommended alternatives (e.g. Danaf et al., 2019a,b; Song et al., 2018). Another common use of the term in the context of discrete choice models refers to the probability that the chosen alternative has the highest predicted probability (e.g. Donkers and Melenberg, 2002; Natter and Feurstein, 2002, etc.). In this paper, we refer to the former definition.

(2008, 2009). They found that the estimates based on properly combined RP and SP data are consistent, while those based on SP data alone are not. The authors attributed this inconsistency to the conditional distribution of coefficients, which differs over respondents and cannot be calculated without the RP data.

2.3. Adaptive metric utility balance

Hauser and Toubia (2005) and Liu et al. (2007) analyzed metric utility balance in adaptive conjoint analysis, where the dependent variable is continuous rather than discrete. In these experiments, respondents are presented with paired-comparisons of two different items. The preference scale is metric, and questions are chosen based on previous responses to result in utilities that are nearly equal. Software packages such as Sawtooth Software (2003) use an adaptive conjoint analysis (ACA) method, which focuses on the attributes that are most relevant to the respondent, thus avoiding information overload. This method generates pairwise comparisons based on the users' responses to "self-explicated" data (where users indicate how important an attribute is to them), and is referred to as "adaptive metric utility balance".

Hauser and Toubia (2005) argue that adaptive metric utility balance in conjoint analysis results in biases, inefficiencies, and higher response errors. Their Monte Carlo simulations show that estimates are biased downwards compared to those obtained from an orthogonal design. They attribute this bias to endogeneity since new questions depend upon the errors made by respondents in their previous answers. However, unlike Fowkes (2007) and Abernethy et al. (2008), their results show that bias increases in the beginning as more questions are asked, and decreases afterwards. In order to avoid endogeneity bias and lowered efficiency, the authors suggest using polyhedral methods instead of adaptive metric utility balance.

In response to the work of Hauser and Toubia (2005), Liu et al. (2007) show that in adaptive metric utility balance experiments (such as the one implemented in Sawtooth software (Sawtooth Software, 2003)), endogeneity becomes ignorable for estimation once the data has been collected because of weak exogeneity and the "likelihood principle". Weak exogeneity holds when the error term is independent of the current values of the regressors (but not of all past and future values). Under weak exogeneity, estimates might be biased, but are consistent. The likelihood principle states that "the likelihood function contains all the information in the data about the model parameters" (Fisher, 1922). This implies that the likelihood is meant to represent the true data generating mechanism. This has several practical implications, such as sequential sampling and pre-survey questions. According to Liu et al. (2007), in such cases one should simply condition on these questions instead of accounting for all the possible answers.

2.4. Summary of background literature

Endogeneity bias could arise in adaptive choice contexts especially in the presence of heterogeneity (Bradley and Daly, 1993), which is at the core of recommender systems and ASP surveys. In choice-based recommender systems, the estimated individual-specific parameters are usually used in menu optimization in order to recommend the "best" alternatives to users. This introduces correlations between the independent variables and the error term. However, applications of discrete choice models in recommender systems are very limited, and endogeneity has not been studied in these few applications.

On the other hand, ASP surveys are used to obtain more efficient estimates by updating the design as the respondents make more choices. Based on the limited literature on ASP surveys, it is undecided whether the estimation results are consistent in such applications or not. Most studies were based on Monte Carlo experiments, some of which have concluded that estimation under ASP is consistent, while others have concluded that it is not. In the applications where significant bias was observed, the magnitude of this bias was found to decrease as more observations per individual are included (Abernethy et al., 2008; Fowkes, 2007).

A theoretical investigation of endogeneity bias and a formal demonstration were presented by Liu et al. (2007) in the context of adaptive regression-based metric utility balance, where SP questions are generated beforehand based on self-explicated data. The authors concluded that consistent estimates can be obtained by conditioning on the self-explicated data. This is in accordance with the findings of Train and Wilson (2008, 2009), who found that SP estimates are only consistent if the RP questions are included in the estimation.

In the following sections, we extend the analysis by Liu et al. (2007) to dynamic choice contexts (where menus or SP questions are generated dynamically based on the responders' previous responses rather than self-explicated data), and demonstrate that the estimates are consistent in certain cases and inconsistent in others, depending on the data used and the likelihood function. We then present a Monte Carlo simulation that replicates a recommender system context, however, the findings can be generalized to ASP surveys as well. Our results also show that the biased estimates of ASP surveys become closer to their true values as more questions are included in the estimation, which is consistent with the literature on ASP surveys.

3. Methodology

In this section, we analyze adaptive choice contexts using discrete choice models with random parameters and provide theoretical explanations for consistent and inconsistent estimation strategies.

In this paper, we are interested in *consistency* rather than *unbiasedness*. Bias refers to the difference between the true value of a parameter (β) and the expectation of its estimate ($E(\widehat{\beta})$), while consistency indicates that the estimate converges to the true value as the sample size increases ($\lim_{n\to\infty} \text{pr}(|\widehat{\beta} - \beta| < \varepsilon) = 1$) (see, e.g. Ben-Akiva and Lerman, 1985). Despite its importance in small samples, unbiasedness cannot be established with standard frequentist estimators (e.g., Maximum Likelihood), and Bayesian estimators are almost always biased (Koop et al., 2007). Therefore, researchers in discrete choice models are primarily concerned with consistency. Maximum Simulated Likelihood (MSL) estimators of discrete choice models are consistent with a sufficient number of draws (Train, 2009), and Bayesian estimators are consistent and independent of the prior distribution in most cases as the sample size increases according to the Bernstein–von Mises theorem (Ghosal, 1997; Train, 2009; Van der Vaart, 2000). Later in our Monte Carlo experiments (Section 4), we refer to "bias" as the difference between the true values of the model parameters and their corresponding estimates (i.e., finite sample empirical bias).

3.1. Model specification

We consider a system (e.g., recommender system or ASP survey) in which an individual n ($n = 1, 2, \ldots, N$) is presented with a menu m ($m = 1, 2, \ldots M_n$), with J_{mn} ($j = 1, 2, \ldots J_{mn}$) alternatives (each *menu* corresponds to a *choice task*). The underlying behavioral model is a logit mixture with the linear utility specification shown in equation (3).

$$U_{jmn} = -P_{jmn} + X_{jmn}\beta_n + \exp(\alpha_n)\varepsilon_{jmn} \tag{3}$$

U_{jmn} in equation (3) is individual n's unobserved utility of alternative *j* in menu m, P_{jmn} is the price of alternative *j* (with its coefficient fixed to -1), X_{jmn} represents a vector of individual characteristics and alternative attributes, β_n represents a vector of parameters/preferences, α_n is a scale parameter, and ε_{jmn} is an error term following the extreme value distribution EV (0, 1).

The model uses the money-metric utility specification as in Ben-Akiva et al. (2019), whereby the price coefficient is fixed to -1. This specification is equivalent to assuming that the scale is fixed and the price coefficient is distributed, but it is advantageous because in it all other coefficients represent the willingness-to-pay for the corresponding attributes (Train and Weeks, 2005). Since the price coefficient is fixed, the scale parameter α_n can be estimated. We can divide all the elements of equation (3) by $\exp(\alpha_{mn})$ in order to obtain equation (4), where the error term ε_{jmn} is distributed as EV (0,1). This setting is used to be able to distinguish between the impact of endogeneity on the scale and on the ratio of the model coefficients (Guevara and Ben-Akiva, 2012).

$$U_{jmn} = \frac{1}{\exp(\alpha_n)}\left(-\ P_{jmn}\ +\ X_{jmn}\beta_n\right)\ +\ \varepsilon_{jmn} \tag{4}$$

We define ζ_n as a vector of individual-specific coefficients which includes both β_n and α_n. We assume that ζ_n is normally distributed in the population with mean μ and covariance matrix Ω.

$$\zeta_n \sim N(\mu,\ \ \Omega) \tag{5}$$

We also define θ as the set of all model parameters (in this case, it includes μ, Ω, and ζ_n).

3.2. Menu generation

In both ASP surveys and recommender systems, menus are generated using the previous choices and attributes. Without loss of generality, we assume that:

1. The system is initialized with one or more menus having exogenous attributes X_0 (which can also be pre-survey questions, or any exogenous screening questions). The choices/responses in these exogenous menus are denoted as d_0.
2. A known function (Q) is used to generate menu m with alternative attributes X_m based on the previous choices ($d_0, \ldots,\ d_{m-1}$) and attributes ($X_0, \ldots,\ X_{m-1}$) (where $Q(X_m|X_0, \ldots,\ X_{m-1}, d_0, \ldots,\ d_{m-1})$ represents the probability of menu X_m conditional on these choices and attributes).

In the case of recommender systems, this function is an assortment optimization method that generates a personalized menu of alternatives. A detailed example demonstrating this process is presented in Section 4. At each menu m, we have a universal set of alternatives to recommend from $\mathscr{S}_{mn}$. This universal set is assumed to be exogenous; it is independent of any of the model parameters, the error terms, and the individual-specific coefficients of individual n. The goal is to generate a recommended subset $\mathscr{S}^*_{mn} \in \mathscr{S}_{mn}$ with J_{mn} alternatives by maximizing a specific objective function given our estimates of the individual-specific coefficients of individual n. Song et al. (2017, 2018) provide a methodology for generating recommendations with different objective functions (e.g., maximizing consumer surplus, hit-rate, etc.). In each menu m, individuals make the choice between any of the J_{mn} recommended alternatives and opting-out (i.e., not choosing any of the recommendations).

The recommended alternatives in menu m are generated as follows:

1. Individual-specific coefficients are estimated using all menus up to $(m - 1)$
2. These coefficients are fed into an assortment optimization to determine the J_{mn} alternatives to be included in the optimized menu ($\mathscr{S}^*_{mn}$).
3. The choice among these J_{mn} alternatives and opting-out in menu m is observed.
4. Steps 1–3 are repeated after the new menu m is included in the next estimation.

Choice-based recommender systems typically use estimates of the individual-specific parameters ζ_n, which are obtained either using the Hierarchical Bayes (HB) estimator of logit mixture (Train, 2009), or using MSL estimation. These parameters are usually not estimated precisely due to shrinkage and simulation variance, especially when the number of observations per individual is small. However, Danaf et al. (2019a,b) show that using these parameters in personalized recommendations improves the hit-rates substantially (even with a small number of observations per individual).

In the case of ASP surveys, the menu generation function can be a deterministic (or even probabilistic) function which determines the attributes of menu m based on the previous choices. For example, a researcher might increase the price of an alternative if it was chosen before in order to see whether the subject will switch to another alternative or not.

3.3. Consistent estimation

When a typical model estimation is done using data from ASP surveys or recommender systems, the likelihood function that is being considered is presented in equation (6).

$$P(d_0, d_1, \dots, d_m \,|X, \theta) = P\left(d_0, d_1, \dots, d_m | X_0, X_1, \dots, X_m, \theta\right) \tag{6}$$

Inconsistency arises due to the misspecification of this likelihood function. Ideally, the likelihood function should include two components:

1. The probability of the observed choices conditional on the attributes $P(d_m|X_m,\theta)$: This is conditional on the true model parameters θ because the choices are generated based on these parameters.
2. The probability of a menu with attributes X_m conditional on the previous choices and attributes, $Q(X_m|d_1, \dots, d_{m-1}, X_1, \dots, X_{m-1}, \mathscr{S}_{mn})$: This is dependent on the true model parameters only through the choices.

In the case of recommender systems, the attributes are determined by the analysts' estimates of the model parameters $\widehat{\theta}$, and not their true values. However, since $\widehat{\theta}$ are estimated using the previous attributes and choices, the two expressions $Q(X_m|\widehat{\theta}, \mathscr{S}_{mn})$ and $Q(X_m|d_1, \dots, d_{m-1}, X_1, \dots, X_{m-1}, \mathscr{S}_{mn})$ are equivalent.

The joint likelihood of the choices and attributes of the presented alternatives up to menu m is presented in equation (7).

$$\begin{aligned}&P\left(d_0, \dots, \ d_m, \ X_1, \dots, \ X_m | X_0, \ \theta\right) = P\left(d_0 \ |X_0, \theta\right) Q\left(X_1 | d_0, X_0, \mathscr{S}_1\right) P\left(d_1 | X_1, \theta\right) \dots \\ &Q\left(X_m | d_0, \dots, d_{m-1}, X_0, \dots, X_{m-1}, \mathscr{S}_m\right) P\left(d_m | X_m, \theta\right)\end{aligned} \tag{7}$$

Q can either be a deterministic or a non-deterministic function. If it is deterministic, then the outcome is either 0 or 1; it is 1 if the combination of $\{d_0, \dots, d_{m-1}, X_0, \dots, X_{m-1}, \mathscr{S}_m\}$ will result in presenting attributes X_m in the next menu and 0 otherwise. Danaf et al. (2019a) and Song et al. (2018) present deterministic recommendation functions (where recommendations are generated by integrating over the posterior distributions of the individual-specific parameters, or by using the means of these distributions). If it is non-deterministic, then the probabilities of presenting different attributes can be specified by the researcher. Teo et al. (2016) and Song (2018) present applications with non-deterministic recommendation functions that use multi-armed bandit methods such as Thompson sampling (which use random draws from the posterior distributions).

Therefore, conditional on the previous choices, their attributes, and the universal set to recommend from ($\mathscr{S}_{mn}$), the probability of the recommended alternatives in menu m is a constant. For example, in a recommender system with a deterministic function Q, given the previous choices and menus presented to a specific individual, and given the universal set of alternatives, the recommendation becomes deterministic:

$$\begin{aligned}&Q\left(X_1 | d_0, X_0, \mathscr{S}_1\right) = 1 \\ &Q\left(X_m | d_0, \dots, d_{m-1}, X_0, \dots, X_{m-1}, \mathscr{S}_m\right) = 1\end{aligned} \tag{8}$$

Furthermore, if the universal sets of alternatives $\mathscr{S}_m$ are not known, but exogenous, even without conditioning on these universal sets, the menu generation probability is a constant and only depends on the distribution of attributes in the universal sets:

$$Q\left(X_m | d_0, \dots, d_{m-1}, X_0, \dots, X_{m-1}\right) = \int_{\mathscr{S}_m} Q\left(X_m | d_0, \dots, d_{m-1}, X_0, \dots, X_{m-1}, \mathscr{S}_m\right) f(\mathscr{S}_m) d\mathscr{S}_m \tag{9}$$

where $f(\mathscr{S}_m)$ is the distribution of attributes in the universal sets. $Q(X_m|d_0, \dots, d_{m-1}, X_0, \dots, X_{m-1}, \mathscr{S}_m)$ is equal to 1 for all universal sets that result in recommending X_m and 0 otherwise. Therefore, this probability is a constant and it is independent of the model parameters.

As a result, these constant terms can be dropped from the likelihood without having any effect on the estimation results (in classical estimation, these terms drop out of the log-likelihood function, and in Bayesian estimation, the numerator and the denominator of the posterior distribution are scaled by the same constant). Using the conditional independence of choices, we can conclude that the two likelihood functions presented in equations (6) and (7) have the same maximum. In this case, the likelihood function fully explains the underlying data generation process, and therefore, endogeneity is not a concern.

3.4. Inconsistent estimation

Inconsistent estimation results are obtained when the likelihood function does not fully reflect the data generation process. We consider the example in which the exogenous menus $\{X_0,\ d_0\}$ are used to generate menu X_1. However, estimation is done using menus $1, 2, \ldots m$ (i.e., excluding the exogenous menus, X_0). The (misspecified) likelihood function that is being considered is:

$$P(d_1, d_2, \ldots, d_m\ |X, \theta) = P(d_1, d_2, \ldots, d_m | X_1, X_2, \ldots,\ \ X_m,\ \theta), \tag{10}$$

while the correct likelihood function is presented in equation (11):

$$P(d_1,\ \ldots d_m, X_1, \ldots, X_m|\theta) = Q(X_1|\theta)P(d_1|X_1, \theta)\ldots Q(X_m|d_1, \ldots, d_{m-1}, X_1, \ldots, X_{m-1})P\left(d_m|X_m, \theta\right), \tag{11}$$

where $Q(X_1|\theta)$ is the marginal probability of observing menu X_1 (assuming that X_0 and d_0 are unknown). $Q(X_1|\theta)$ can be obtained by integrating over the joint distribution of X_0, d_0, and X_1 given by equation (12), which is a function of θ because it depends on the choice probabilities $P(d_0|X_0, \theta)$ (which are functions of θ).

$$f(X_0, d_0, X_1|\theta) = P\left(d_0|X_0, \theta\right)f(X_0)Q\left(X_1|d_0, X_0\right) \tag{12}$$

where $f(X_0)$ is the joint distribution of all the attributes in the exogenous menus, $P(d_0|X_0, \theta)$ is the probability of the choices given the exogenous attributes and model parameters, and $P(X_1|d_0, X_0)$ is the menu generation probability of the first menu.

$$Q(X_1|\theta) = \int_{X_0} \sum_{d_0} P(d_0|X_0, \theta)f(X_0)Q\left(X_1|d_0, X_0\right)dX_0 \tag{13}$$

In equation (13), $Q(X_1|d_0, X_0)$ cannot be treated as a constant and dropped from the equation as before because the outcome is different for each combination of $\{d_0, X_0\}$. For example, if a deterministic function is used, this probability is either equal to zero or one, depending on the values of d_0 and X_0 (it is equal to one only for combinations of d_0 and X_0 that will result in recommending the first observed menu X_1).

Therefore, the probability of observing the first endogenous menu is a function of the model parameters θ, and excluding this expression from the likelihood function results in inconsistent estimates. Since the two likelihood functions are not equivalent, the one presented in equation (10) does not account for the data generation process, and therefore, the estimates are inconsistent. In addition, the expression in equation (13) is difficult to evaluate because it requires multi-dimensional integration over all the possible attribute levels in the exogenous menus. The case presented above demonstrates an example of when inconsistency can arise, however, it is not the only one. Inconsistent estimates will be obtained if any of the following menus is excluded (if this menu was used in generating the subsequent menus). For example, Hauser and Toubia (2005) obtained biased estimates in a Monte Carlo simulation of ACA after excluding the "self-explicated" data from their estimation.

These results are in line with the "*initial conditions*" problem (Heckman, 1981), who analyzed time-series probit models with lagged dependent variables. If the initial conditions are assumed to be exogenous, estimation with the entire history results in consistent estimates. If the analyst does not have access to the entire history (e.g., because the process has been in operation prior to the time it is sampled), or if initialization is not truly exogenous, the estimation results are inconsistent. In analyzing this problem, Akay (2012) found similar results to those obtained by Abernethy et al. (2008) and Fowkes (2007), where Monte Carlo experiments showed that the magnitude of bias decreases with longer panels.

4. Monte Carlo experiment

In this section, we simulate a dynamic recommender system following the methodology of Danaf et al. (2019a) and Song et al. (2017, 2018), mimicking the choice of Mobility-as-a-Service (MaaS) plans. Based on the analysis presented in Section 3, we present examples of consistent and inconsistent estimation results.

4.1. Data set description

The Monte Carlo data assumes that 10,000 individuals are presented with 16 successive menus. A sufficiently large sample size is used because we are interested in analyzing the consistency of the estimates (lower samples result in larger standard errors or standard

Table 1
Monte Carlo attributes and levels.

Attribute	Symbol	Levels
Monthly Price	P	\$0 to \$480*
Transit	T	Available (1) or unavailable (0)
Bike Sharing	B	Available (1) or unavailable (0)
On Demand Trips	OD	2 to 12 trips/month

*Price is positively correlated with all other attributes deviations of the posterior distribution, which makes it difficult to distinguish bias from statistical discrepancies and simulation errors). While this sample is large compared to traditional applications, it is typical in web or app-based contexts such as recommender systems (where the number of users is large).

Each menu includes three alternatives (different MaaS plans) with varying attributes and an opt-out alternative. Each plan has a different monthly price and three attributes: access to transit, access to bike sharing, and the number of on-demand trips (e.g., taxi, Uber. Lyft etc.) per month. Table 1 shows the distributions of the attributes in the universal sets ($\mathscr{S}_{\mathrm{m}}$). The dependent variable is the choice between the three different plans or opting-out (indicating that the individual does not purchase any of the MaaS plans, or chooses an outside alternative).

The utility equations (normalized to the opt-out alternative) are presented in equation (14).

$$\begin{aligned} &\mathrm{U_{jmn}} \equiv \frac{1}{\exp(\alpha_{\mathrm{n}})}\left(-\mathrm{P_{jmn}} + \beta_{\mathrm{T_n}}\mathrm{T_{jmn}} + \beta_{\mathrm{B_n}}\mathrm{B_{jmn}} + \exp(\beta_{\mathrm{OD_n}})\mathrm{OD_{jmn}} + \beta_{\mathrm{q_n}}\right) + \varepsilon_{\mathrm{jmn}} \\ &\mathrm{U_{opt-out,mn}} \equiv 0 + \varepsilon_{\mathrm{opt-out,mn}} \end{aligned} \tag{14}$$

where:

- n is an index for users ($\mathrm{n} = 1, 2, \ldots, \mathrm{N}$), m is an index for menus ($\mathrm{m} = 1, 2, \ldots, \mathrm{M_n}$), and j is an index for alternatives in the menu ($\mathrm{j} = 1, 2, \ldots, \mathrm{J_{mn}}$).
- $\mathrm{U_{jmn}}$ represents the utility of alternative j in menu m faced by individual n, and $\mathrm{U_{opt-out,mn}}$ is the opt-out utility.
- $\mathrm{P_{jmn}}$ is the monthly price (in $100's) of alternative j in menu m faced by individual n, with its coefficient normalized to -1.
- $\mathrm{T_{jmn}}$, $\mathrm{B_{jmn}}$ and $\mathrm{OD_{jmn}}$ represent access to transit, access to bike sharing, and the number of on-demand trips per month of alternative j with coefficients $\beta_{\mathrm{T_n}}$, $\beta_{\mathrm{B_n}}$, and $\exp(\beta_{\mathrm{OD_n}})$ respectively (exponentiation is used to obtain a lognormally distributed coefficient).
- β_{qn} is a constant term for choosing any plan (rather than opting out).
- α_{n} is a scale parameter (the lognormal distribution is also used to guarantee that the scale is positive for everyone).
- $\varepsilon_{\mathrm{jmn}}$ is an error component following the extreme value distribution ($\mathrm{EV}(0, 1)$).

All coefficients are normally distributed in the sample. The true values of the population means and inter-consumer covariance matrix are shown in Table 2. The true values of the individual specific coefficients are generated only once from their corresponding distributions, and then used in generating the choices in each menu. However, the attributes of the alternatives in the universal sets are different (both across different individuals and different menus).

The choices are simulated by calculating the systematic utilities (using the true individual-specific coefficients and the attributes) and adding EV (0,1) error terms to these systematic utilities to obtain the total utilities. The alternative with the highest total utility is chosen.

4.2. Dynamic recommendations

We assume that in each menu, three alternatives are recommended (and included in this menu) from a universal set ($\mathscr{S}_{\mathrm{mn}}$) of 10 alternatives. This universal set is different for each individual and for each menu, and the 10 alternatives in this set are generated by drawing the attributes from their corresponding distributions shown in Table 1. We also assume that the individual can choose any of the recommended alternatives or opt-out, and that the remaining (non-recommended) alternatives are not available in that choice situation.

In the first two menus, the recommendations are random; i.e., the three recommended alternatives are randomly chosen from the 10 alternatives in the universal set. Therefore, these two menus have exogenous attributes.

The following menus (3–16) are generated in a dynamic manner based on each user's previous choices. Recommendations in menu m are generated using the same procedure as in Section 3.2:

1. Estimating individual-specific coefficients ($\widehat{\zeta}_{\mathrm{n}}$) using menus $1, 2, \ldots, \mathrm{m} - 1$ (e.g. using the HB procedure presented in following section, Section 4.3).

Table 2
True values of the parameters.

Parameter	True mean	Covariances				
		α_n	β_{T_n}	β_{B_n}	β_{OD_n}	β_{q_n}
α_n	−1.0	0.25	0	0	0	0
β_{T_n}	1.0	0	2.0	0.3	0.1	−0.3
β_{B_n}	0.5	0	0.3	1.0	0.2	−0.2
β_{OD_n}	−2.0	0	0.1	0.2	0.5	−0.1
β_{q_n}	−0.5	0	−0.3	−0.2	−0.1	1.0

2. Calculating the systematic utilities of each alternative in the universal set using the estimated individual-specific coefficients:

$$\widehat{V}_{jmn} \equiv \frac{1}{\exp(\widehat{\alpha}_n)}\left(-\ P_{jmn} + \widehat{\beta}_{S_n}S_{jmn} + \widehat{\beta}_{C_n}C_{jmn} + \exp(\widehat{\beta}_{L_n})L_{jmn} + \widehat{\beta}_{q_n}\right) \qquad j = 1,\ 2, \ldots,\ 100 \tag{15}$$

3. Selecting the three alternatives with the highest calculated utilities and including them in the recommended menus.

In each menu, the choice among the three MaaS plans and opting-out is simulated using the true individual-specific coefficients (ζ_n) (and not their estimates). The procedure is then repeated for the next menu ($m+1$).

4.3. Model estimation

To estimate this model, we use the Hierarchical Bayes procedure described below, which is based on a three-step Gibbs sampler with an embedded Metropolis-Hastings algorithm (Train, 2009):

1. Drawing $\mu|\Omega,\ \zeta_n$ using a Normal Bayesian update with unknown mean and known variance.
2. Drawing $\Omega|\mu,\ \zeta_n$ using a Normal Bayesian update with known mean and unknown variance.
3. Drawing $\zeta_n|\mu,\ \Omega$ using the Metropolis-Hastings algorithm.

In this procedure, the individual-specific coefficients ζ_n are treated as additional model parameters. Sequential draws are obtained from the three conditional distributions shown above in order to eventually obtain draws from the unconditional posterior distribution $K(\mu,\Omega,\ \zeta_n|X,d)$. A diffuse prior is used on the population means (i.e., with large variances), and a weakly informative Half-t prior is used on the covariance matrix as recommended by Akinc and Vandebroek (2018), and Huang and Wand (2013).

Bayesian estimation is used to replicate MSL estimates (and standard errors) at a lower computational cost. The point estimates of the population means and variances are obtained by averaging the posterior draws of these parameters, which pertains to minimizing the quadratic loss function (Koop et al., 2007). Because of the uninformative priors and the large sample size, Bayesian and classical estimates of μ, Ω, and ζ_n are virtually identical, despite their algorithmic differences (Ben-Akiva et al., 2019; Huber and Train, 2001; Train, 2009). Train (2009) shows that the mean of the posterior is an equivalent to the maximum likelihood estimator, and the standard deviations of the posterior provide classical standard errors for the estimator. Huber and Train (2001) extend this result to show that the individual-specific parameters obtained from both estimation methods are identical. Therefore, estimation using MSL and extracting individual-specific coefficients does not affect the estimation results in this case (and hence the findings of this paper).

Table 3
Estimation results with varying panel sizes and exogenous initialization.

Population Means					
	Transit	Bike Sharing	On Demand	Constant	Scale
True Value	1.000	0.500	−2.000	−0.500	−1.000
Value in Sample	0.996	0.500	−1.998	−0.507	−0.999
Estimation with m = 1, 2	0.942 (0.025)	0.497 (0.022)	−1.997 (0.021)	−0.498 (0.035)	−1.022 (0.044)
Estimation with m = 1, 2, 3	0.960 (0.021)	0.486 (0.017)	−2.007 (0.018)	−0.501 (0.027)	−0.988 (0.024)
Estimation with m = 1, 2, 3,4	0.964 (0.019)	0.490 (0.015)	−2.004 (0.016)	−0.491 (0.025)	−0.989 (0.018)
Estimation with m = 1, 2, …,5	0.973 (0.018)	0.489 (0.014)	−2.008 (0.013)	−0.497 (0.022)	−0.990 (0.014)
Estimation with m = 1, 2, …,6	0.980 (0.017)	0.494 (0.013)	−1.997 (0.012)	−0.514 (0.022)	−0.984 (0.013)
Estimation with m = 1, 2, …,7	0.979 (0.017)	0.494 (0.013)	−2.004 (0.012)	−0.512 (0.021)	−0.987 (0.011)
Estimation with m = 1, 2, …,8	0.983 (0.017)	0.496 (0.013)	−2.002 (0.011)	−0.516 (0.020)	−0.993 (0.011)
Estimation with m = 1, 2, …,12	0.982 (0.017)	0.496 (0.012)	−2.007 (0.010)	−0.510 (0.018)	−0.992 (0.009)
Estimation with m = 1, 2, …,16	0.985 (0.016)	0.496 (0.011)	−2.005 (0.010)	−0.507 (0.017)	−0.996 (0.008)
Variances					
	Transit	Bike Sharing	On Demand	Constant	Scale
True Value	2.000	1.000	0.500	1.000	0.25
Value in Sample	2.009	1.013	0.507	1.011	0.248
Estimation with m = 1, 2	2.127 (0.112)	1.105 (0.068)	0.523 (0.041)	1.142 (0.109)	0.308 (0.089)
Estimation with m = 1, 2, 3	2.059 (0.088)	1.037 (0.050)	0.532 (0.029)	1.122 (0.086)	0.233 (0.049)
Estimation with m = 1, 2, 3,4	2.056 (0.073)	1.032 (0.040)	0.519 (0.026)	1.055 (0.065)	0.243 (0.028)
Estimation with m = 1, 2, …,5	2.052 (0.069)	1.013 (0.034)	0.529 (0.021)	1.052 (0.050)	0.230 (0.024)
Estimation with m = 1, 2, …,6	2.045 (0.062)	1.034 (0.032)	0.534 (0.018)	1.101 (0.052)	0.235 (0.023)
Estimation with m = 1, 2, …,7	2.050 (0.060)	1.036 (0.029)	0.537 (0.017)	1.093 (0.047)	0.218 (0.017)
Estimation with m = 1, 2, …,8	2.063 (0.056)	1.042 (0.027)	0.536 (0.017)	1.073 (0.043)	0.218 (0.015)
Estimation with m = 1, 2, …,12	2.053 (0.047)	1.033 (0.023)	0.527 (0.013)	1.045 (0.037)	0.232 (0.008)
Estimation with m = 1, 2, …,16	2.053 (0.047)	1.033 (0.021)	0.527 (0.012)	1.027 (0.031)	0.243 (0.007)

4.4. Results

The results presented in Table 3 show the estimates (posterior means) and standard errors (posterior standard deviations) of the population means and inter-consumer variances. The estimates of the off-diagonal elements of the covariance matrix are presented in Appendix A. These results are based on a single estimation, however, in order to avoid incidental simulation results, a large sample size is used, and variations of the experimental setting are presented in Appendices B and C. The results indicate that when all menus (up to menu m) are included in the estimation, there is no significant bias in any of the parameters (bias is defined as the difference between the estimates and their corresponding true values in the population).

On the other hand, Table 4 presents similar estimation results (obtained with the same datasets), but the first two menus are excluded from the estimation. In this case, significant differences are observed in most of the parameters (at the 95% level of confidence). For example, in the first few menus, all the population means (including the means of the constant and the scale parameter) are over-estimated. In addition, the variances of the constant and scale parameters are substantially overestimated in the first two menus. These results indicate that any model estimation with menus $1, \ldots, m$ results in consistent estimates. However, if some menus prior to menu m are excluded from the estimation, the estimates deviate substantially from their true values.

Fig. 1 shows the bias in the population means (difference between the estimates and the true values) as a function of the number of menus included in estimation. When the first two menus are excluded, the observed bias decreases as more choices are included in the estimation. This is because the effect of the omitted term $(Q(X_3|\theta))$ on the likelihood decreases. The attributes in any menu m after the third one are determined partially by the attributes and choices of menus $3, \ldots, m$. Therefore, the bias decreases since the misspecified likelihood function accounts partially for the data generation process. The results are also consistent with the findings of Fowkes (2007), who concluded that LASP is asymptotically unbiased, and Abernethy et al. (2008), who concluded that bias decreases as more questions are observed.

5. Practical implications

This section presents the practical implications on recommender systems, ASP surveys, and RP/SP estimation. In some of these applications, the analyst might have access to the entire choice history, and thus consistent estimation is possible. However, this might not apply to all cases.

5.1. Recommender systems

The results presented in Section 4.4 indicate that the entire choice history of each individual is needed to estimate the behavioral models in recommender systems. However, in such systems, the number of individuals and choices per individual can be excessively large, resulting in computational constraints.

There are a few ways for the analyst to deal with this problem. The results presented in Table 4 and Fig. 1 suggest including as much observations per individual as possible, to reduce the bias. However, the analyst cannot know how many choices are needed, or

Table 4
Estimation results with varying panel sizes, excluding the first two menus (endogenous initialization).

Population Means					
	Transit	Bike Sharing	On Demand	Constant	Scale
True Value	1.000	0.500	−2.000	−0.500	−1.000
Value in Sample	0.996	0.500	−1.998	−0.507	−0.999
Estimation with m = 3	1.911 (0.090)	0.909 (0.045)	−1.892 (0.051)	0.994 (0.239)	−0.792 (0.066)
Estimation with m = 3, 4	1.331 (0.038)	0.686 (0.029)	−1.866 (0.029)	−0.185 (0.086)	−0.604 (0.037)
Estimation with m = 3, 4, 5	1.182 (0.028)	0.588 (0.020)	−1.945 (0.021)	−0.154 (0.064)	−0.709 (0.026)
Estimation with m = 3, 4, …, 6	1.102 (0.022)	0.550 (0.017)	−1.960 (0.019)	−0.196 (0.047)	−0.765 (0.02)
Estimation with m = 3, 4, …, 7	1.059 (0.020)	0.531 (0.015)	−1.987 (0.016)	−0.236 (0.039)	−0.819 (0.016)
Estimation with m = 3, 4, …, 8	1.041 (0.018)	0.521 (0.014)	−1.986 (0.014)	−0.264 (0.031)	−0.855 (0.013)
Estimation with m = 3, 4, …, 12	1.004 (0.016)	0.509 (0.012)	−2.005 (0.010)	−0.370 (0.027)	−0.921 (0.01)
Estimation with m = 3, 4, …, 16	1.001 (0.015)	0.503 (0.012)	−1.999 (0.009)	−0.417 (0.021)	−0.948 (0.008)
Variances					
	Transit	Bike Sharing	On Demand	Constant	Scale
True Value	2.000	1.000	0.500	1.000	0.25
Value in Sample	2.009	1.013	0.507	1.011	0.248
Estimation with m = 3	8.539 (1.416)	2.053 (0.342)	0.751 (0.117)	50.964 (6.924)	0.941 (0.160)
Estimation with m = 3, 4	2.171 (0.177)	1.011 (0.088)	0.504 (0.050)	7.602 (0.865)	1.634 (0.174)
Estimation with m = 3, 4, 5	1.765 (0.095)	0.998 (0.059)	0.551 (0.036)	3.823 (0.329)	0.826 (0.114)
Estimation with m = 3, 4, …, 6	1.655 (0.074)	1.026 (0.046)	0.563 (0.031)	2.729 (0.204)	0.549 (0.054)
Estimation with m = 3, 4, …, 7	1.698 (0.064)	1.021 (0.039)	0.563 (0.022)	2.120 (0.134)	0.387 (0.039)
Estimation with m = 3, 4, …, 8	1.711 (0.056)	1.023 (0.032)	0.537 (0.021)	1.780 (0.102)	0.331 (0.028)
Estimation with m = 3, 4, …, 12	1.786 (0.051)	0.997 (0.024)	0.530 (0.015)	1.313 (0.050)	0.271 (0.012)
Estimation with m = 3, 4, …, 16	1.847 (0.043)	1.007 (0.022)	0.520 (0.012)	1.202 (0.043)	0.269 (0.010)

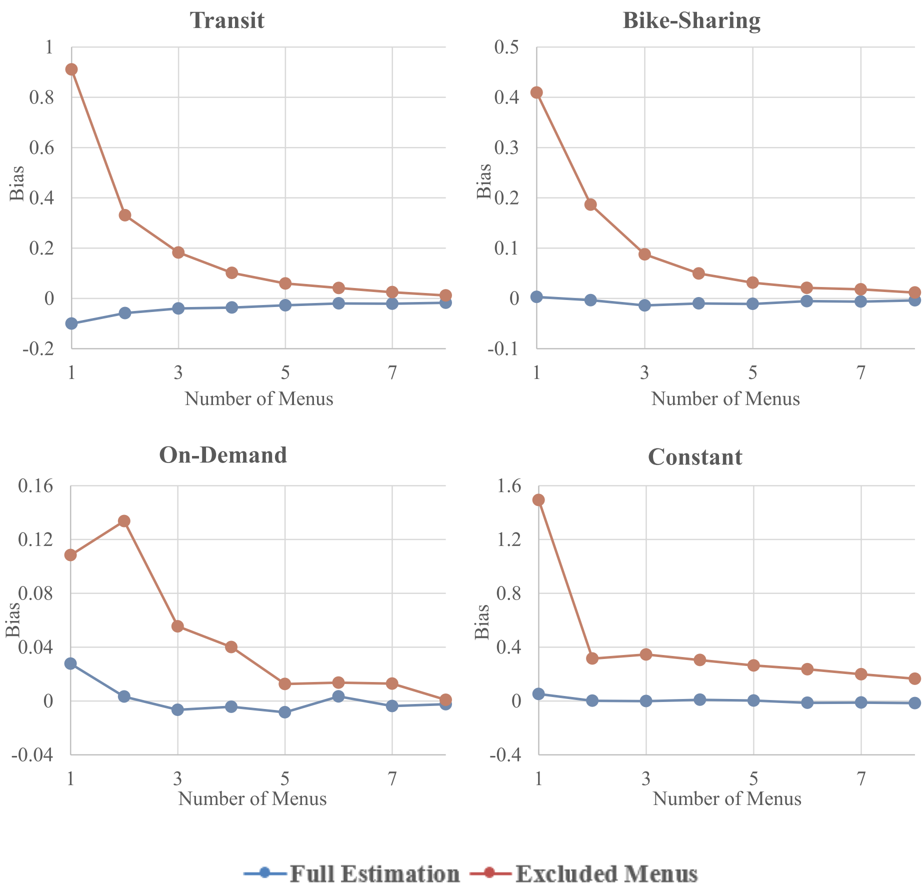

Fig. 1. Bias as a function of the number of menus under "Full Estimation" (with the entire choice history) and "Excluded Menus" (where the first two menus are excluded). Fig. 1 shows the number of menus used in estimation, and not the menu number. (e.g., in the case where the first two menus are excluded, estimation with one menu refers to estimation with menu 3).

whether the bias is eliminated or not.

If relevant instruments are available, standard correction methods may be used such as the control functions. Ongoing research on recommender systems is investigating potentially relevant instruments, such as the attributes of non-personalized menus (that are recommended based on population parameters rather than individual-specific parameters).

Alternatively, the population parameters can be estimated using a subset of individuals (where the entire choice history of this subset is included), and the online estimation method proposed by Danaf et al. (2019a) can be used to estimate individual-specific parameters. Models calibrated at the individual level do not require the heterogeneity to be uncorrelated with the covariates (as in the case of *fixed effects*). Computational constraints can also be mitigated by parallelization (e.g. Neiswanger et al., 2013), where data is partitioned into different batches that are processed independently on multiple machines (using any classical MCMC method such as Gibbs sampling), and combined afterwards to generate samples from the full posterior.

The analyses presented in Sections 3.3 and 3.4 apply to various recommendation functions as long as the assumptions presented in Section 3.2 are satisfied. In Section 4, we used the estimated individual-specific parameters in order to generate personalized recommendations. The function Q is an assortment optimization that recommends alternatives that are most likely of being chosen. This reasoning can be extended to other applications that do not use individual-specific parameters. For example, Appendix B presents a recommender system that recommends the three nearest neighbors of the alternative chosen in the previous menu (with the weights being the population level means of the corresponding attributes). The results of this experiment are similar to those presented in Tables 3 and 4

5.2. ASP surveys

ASP surveys usually have smaller sample sizes than recommender systems, and therefore, computational constraints are not an issue. In most cases, estimation with the entire data set is feasible. However, we present some cases where specifying the correct likelihood function might not be obvious.

In some cases, the analyst might be interested in estimating a model with a subset of the questions. For example, in transportation mode choice, if the respondent is presented with an alternating design of urban and suburban trips, the analyst cannot estimate two separate models (for urban and suburban trips), as the estimates will be inconsistent.

Some SP questions might be generated using a pre-survey without including the responses to the pre-survey questions in the likelihood function. For example, in a transportation mode choice SP survey, the respondent might only be presented with car alternatives if he/she indicated that they are willing to buy a car. Similarly, the cost of transit can be set to zero in SP questions if the respondent indicated in the pre-survey that he/she is willing to buy a transit pass. In such cases, estimation with the ASP questions alone might result in inconsistent estimates.

The same reasoning applies to adaptive regression-based conjoint analysis (where the density of the continuous response variable $L(Y_m|X_m\theta)$ is used in equation (7) instead of the choice probability $\mathrm{P}(\mathrm{d_m}|\mathrm{X_m},\theta)$). For more details on regression-based ACA, we refer the reader to Liu et al. (2007), who were particularly interested in endogeneity bias that arises as a result of using "*self-explicated*" data in adaptive conjoint analysis. They concluded that endogeneity bias is "ignorable" if SP data and self-explicated data are properly modeled jointly, which can be handled by some modules of the Sawtooth software (Sawtooth Software, 2003).

5.3. Adaptive valuation methods

The findings of this paper can also be relevant to contingent valuation (CV) experiments, where users are presented with a series of double bounded dichotomous choice questions designed to infer their marginal rate of substitution (MRS) or willingness-to-pay (WTP). When the design is not adaptive (e.g. when questions do not depend on previous responses), endogeneity bias is not a concern. However, an adaptive design might be used in order to narrow down the MRS or WTP estimates more efficiently (e.g. Kanninen, 1993a, 1993b; Nyquist, 1992). For example, the researcher might infer a lower bound on the respondents' willingness-to-pay from his/her previous choices, and present the respondent with a price that exceeds the lower bound.

In adaptive CV experiments, considerable care is needed when estimating statistical models over a sample of respondents, especially when heterogeneity is modeled using continuous or discrete distribution (e.g. latent class). In such cases, excluding data from estimation might result in inconsistent estimation results. However, this is not an issue when WTP and MRS values are calculated on the individual level (which is similar to the findings of Fowkes (2007)).

Another relevant application is the "half-space" method proposed by Rouwendal et al. (2010) in order to narrow down the location of an individual's marginal rate of substitution (MRS) or willingness-to-pay (WTP) using successive choice situations that are adjusted dynamically. Each choice situation divides the space of relevant marginal valuations into two "half-spaces", and the choice reveals the half-space to which the respondent belongs. The authors start by assuming that the choices are error free (i.e. individual choices are consistent), and then extend their method to allow for errors using a statistical model.

In the first part of the latter study, the upper and lower bounds of WTP are estimated for each individual separately, without assuming any distribution. Therefore, endogeneity bias is not a concern despite the adaptive nature of the questionnaire. In this case, excluding previous choices might result in wider intervals of WTP, but the true WTP values will certainly lie between the estimated upper and lower bounds (since the choice is error free). In the second part of the study, the data are pooled over all respondents and a statistical model is estimated to account for the probability of committing errors. This is advantageous because it allows for including all the observations in estimation (and not only those whose choices are consistent). However, it also requires the upper and lower bounds of WTP to be estimated using a discrete mixture model (rather than estimating them on the individual level). If an adaptive design is used, excluding previous choices from estimation might result in inconsistent estimates of the population parameters (i.e. the probabilities of belonging to different sub-spaces, and the probabilities of committing errors).

This should rarely be a concern in such applications, because the analyst usually has access to the entire choice history of each respondent, and the sample size is not large enough to impose computational constraints. However, there might be other sources of bias affecting the estimation results in these contexts. For example, Blamey et al. (1999) argue that these experiments generally result in WTP values exceeding those observed in real-life markets, potentially because of "yea-saying" (which refers to the higher probability of responding "yes" in hypothetical dichotomous choice situations where payment is not required). In addition, several studies show that learning and fatigue effects might cause additional biases in the context of contingent valuation (e.g. Adamowicz, and Boxall, 2001; Bateman et al., 2008; Holmes and Boyle, 2005; etc).

5.4. SP/RP estimation

The results obtained in Section 4.4 are in accordance with the findings of Train and Wilson (2008) who concluded that estimation with pivoted SP designs is consistent when RP choices are properly included in the estimation. This can be interpreted as a special case of the framework presented in Section 3. To demonstrate this, we denote by d_{RP} and X_{RP} the choices and attributes in the RP data, and d_{SP} and X_{SP} the choices and attributes in the SP data respectively. Assuming the RP attributes are exogenous and the function Q is used to generate SP attributes, the joint Likelihood of observing the SP attributes and the RP and SP choices is expressed as:

$$P(d_{RP}, X_{SP}, d_{SP}|X_{RP}, \theta) = \mathrm{P}(d_{RP}|X_{RP}, \theta)\mathrm{Q}(X_{SP}|d_{RP}, X_{RP})\mathrm{P}(d_{SP}|X_{SP}, \theta) \tag{16}$$

In this case, $\mathrm{Q}(X_{SP}|d_{RP}, X_{RP})$ is constant (independent of θ), and thus can be dropped. Therefore, joint estimation with SP and RP data ensures that the likelihood function is correctly specified.

However, if we do not include the RP data (d_{RP} and X_{RP}) in the estimation, we have to integrate over their distributions in order to specify the correct likelihood function:

$$\begin{aligned}\mathrm{P}(X_{SP}, d_{SP}|\theta) &= \mathrm{P}(d_{SP}|X_{SP}, \theta)\mathrm{P}(X_{SP}|\theta)\\ &= \mathrm{P}(d_{SP}|X_{SP}, \theta)\int_{X_{RP}}\sum_{d_{RP}}\mathrm{P}(d_{RP}|X_{RP}, \theta)\mathrm{f}(X_{RP})\mathrm{Q}(X_{SP}|d_{RP}, X_{RP})\mathrm{d}X_{RP}\end{aligned} \tag{17}$$

where $\mathrm{f}(X_{RP})$ denotes the distribution of the RP attributes.

Another approach is to use the full information maximum likelihood FIML solution proposed by Train and Wilson (2008), which overcomes this problem by using specially designed maximum simulated likelihood method. Guevara and Hess (2019) also proposed a limited information maximum likelihood (LIML) approach based on a control-function that uses RP attributes as instruments and proved to be more robust to model misspecifications and easier to apply, albeit potentially losing efficiency.

We finally note that the above argument accounts only for endogeneity that is caused by the dependency of SP attributes on RP choices. However, RP data might not be fully exogenous, especially in the presence of measurement errors, self-selection (where the existing market reflects current consumer preferences), and other sources of endogeneity that are outside the scope of this paper.

The same reasoning also applies to efficient SP designs that are generated based on individual-specific priors. Traditional efficient designs aim to improve the efficiency of SP estimates by presenting respondents with more meaningful tradeoffs (e.g. avoiding dominant or inferior alternatives). These designs typically require priors on the model parameters. Endogeneity bias might arise if individual-specific priors are estimated (e.g. using the individuals' previous choices), and used in generating the designs (rather than using the same priors across all individuals). In the latter case, estimation is inconsistent unless the likelihood function accounts for the process by which these priors were generated.

6. Conclusion

This paper investigated endogeneity in adaptive choice contexts, such as choice-based recommender systems and adaptive stated preferences surveys. In these cases, all of the attributes are endogenous, and finding relevant instruments to correct for endogeneity may not be feasible.

A Monte Carlo experiment was used to demonstrate cases where endogeneity results in consistent estimates and cases where it does not, in the context of a choice-based recommender system in which alternatives are recommended based on the estimates of the individual-specific preferences. The results indicate that including all the data in the model estimation (and properly accounting for heterogeneity) results in consistent estimates if the initialization is exogenous. In the latter case, the estimates were not significantly different from their true values. On the other hand, estimates are inconsistent if data that were used in generating subsequent menus are excluded in the estimation. Monte Carlo results also suggest that, even when exogenous initialization is missing, empirical bias decreases with the number of menus, which is consistent with the findings of Fowkes (2007) and Abernethy et al. (2008). Nevertheless, the latter result should be taken with caution because we cannot provide a formal proof for its validity on circumstances beyond the experimental setting considered.

Our findings have implications on the design and analysis of ASP surveys and choice-based recommender systems. Researchers and practitioners should make sure that the likelihood function accounts for the data generation process, which is achieved by conditioning on the previous choices and attributes. It is also important that the system is initialized exogenously, and that this initialization is accounted for in the estimation.

This paper addressed endogeneity that is caused by the dependencies in adaptive choice contexts. However, there might be other sources of endogeneity in such contexts that are not accounted for (such as measurement errors, omitted variables, etc.). By adjusting the design based on the respondents' previous choices, ASP surveys can result in harder questions that increase the respondents' mental burden. On the other hand, recommender systems are designed to reduce the mental burden caused by information overload, by reducing the size of the choice or consideration set. Both of these effects can result in cognitive biases that are beyond the scope of this paper, and which we leave for future research.

In addition, our analysis assumes that the behavioral model is correctly specified. Model misspecification generally results in inconsistent estimates. For example, Hess and Train (2017) show that ignoring correlations between parameters can result in over- or under-estimating the distribution of ratios of coefficients, representing the values of willingness to pay (WTP) and marginal rates of substitution. Model misspecification can also be attributed to preference formulation, learning, and fatigue, which are not usually accounted for in the behavioral model. Our analysis in Section 3 assumes that preferences are stable over time, however, it can be extended to time-varying preferences, only if the model accounts explicitly for this temporal variability. For example, Danaf et al. (2019b) suggest modeling the scale parameter as a function of the menu/choice situation number in order to capture learning and fatigue effects.

Another key assumption in our analysis is that the universal set of alternatives to recommend from $\mathscr{S}_{\mathrm{mn}}$ is assumed to be exogenous. This might not apply in many cases. For example, a recommender system or an ASP survey cannot present the respondent with car alternatives if the respondent does not have access to a vehicle or a valid driving license. In the latter case, car and driving license ownership are endogenous decisions, which are a function of the respondent's preferences towards car.

While our empirical results show that the magnitude of bias decreases with the number of menus or choice situations, several other factors can come into play. Since the bias is caused by the correlation between the individual-specific effects and the independent variables (as shown in Equation (1)), the magnitude of bias is determined by the degree of correlation. For example, the magnitude of bias is expected to be larger with higher levels of heterogeneity (i.e. higher population variances). On the other hand, with a lower scale (or higher variance of the error term), the learning process from previous choices becomes more difficult (due to noise in the data), resulting in a lower bias (however, lower scale usually results in lower hit-rates in recommender systems). Future research should further investigate the effects of heterogeneity, scale, and experimental design on the magnitude of bias in different settings including recommender systems, ASP surveys, and CV experiments.

Future work should focus on real data applications in order to identify relevant instruments to deal with cases where either the analyst does not have access to the entire choice history, or computational constraints preclude using all the data (e.g. recommender systems). In addition, future research should investigate endogeneity corrections in adaptive choice contexts. For example, the standard corrections to the "initial conditions" problem (Heckman, 1981) (e.g. such as the reduced-form approximations and variations of the Wooldridge method (Wooldridge, 2005)) can be extended. In the context of RP/SP estimation, Train and Wilson (2008) and subsequent work by Guevara and Hess (2019) propose corrections (which are variations of the control-function method) in order to obtain consistent estimates with SP data alone.

Acknowledgements

This work is partially aligned with the TRIPOD: Sustainable Travel Incentives with Prediction, Optimization and Personalization research project sponsored by the U.S. Department of Energy Advanced Research Projects Agency-Energy (ARPA-E). It was awarded through the ARPA-E Traveler Response Architecture using Novel Signaling for Network Efficiency in Transportation (TRANSNET) program. Angelo Guevara was funded in part by CONICYT, FONDECYT 1191104 and by CONICYT PIA/BASAL AFB180003.

Appendix A. Covariance Matrix Estimation

Because of the large number of simulations, we did not include the estimation results of all the off-diagonal elements. However, the results below present the estimation results for 16 menus indicating that the true values are recovered accurately.

Table 5
True and estimated covariances with 16 menus.

Parameter	True Covariances				
	α_n	β_{T_n}	β_{B_n}	β_{OD_n}	β_{q_n}
α_n	0.25	0	0	0	0
β_{T_n}	0	2.0	0.3	0.1	−0.3
β_{B_n}	0	0.3	1.0	0.2	−0.2
β_{OD_n}	0	0.1	0.2	0.5	−0.1
β_{q_n}	0	−0.3	−0.2	−0.1	1.0
	Estimated Covariances				
	α_n	β_{T_n}	β_{B_n}	β_{OD_n}	β_{q_n}
α_n	0.243 (0.007)	0	0	0	0
β_{T_n}	0	2.053 (0.047)	0.329 (0.019)	0.100 (0.014)	−0.320 (0.026)
β_{B_n}	0	0.329 (0.019)	1.033 (0.021)	0.222 (0.010)	−0.227 (0.019)
β_{OD_n}	0	0.100 (0.014)	0.222 (0.010)	0.527 (0.012)	−0.117 (0.016)
β_{q_n}	0	−0.320 (0.026)	−0.227 (0.019)	−0.117 (0.016)	1.027 (0.031)

Appendix B. 3-Nearest Neighbors Recommendations

This appendix presents similar results to Section 4.4, but with a different menu generation procedure. The first two menus are generated exogenously as before. Afterwards, the three nearest neighbors to the alternative that was chosen in the previous menu are recommended. Therefore, only the previous choice is taken into consideration.

The results shown in Tables 6 and 7 are similar to those observed in Section 4.4:

1. When the entire choice history of each individual is included in estimation, no significant bias is observed.
2. Excluding the first two menus results in significant biases.
3. The magnitude of bias decreases as more choices per individual are included in the estimation. The estimates with 16 choices are very close to their true values even when the first two menus are excluded.
4. The coefficients of transit, bike sharing, and on-demand, are biased upwards (because of the positive correlations between these variables and the error terms).

Table 6
Endogeneity results using 3-NN recommendations.

Population Means					
	Transit	**Bike Sharing**	**On-Demand**	**Constant**	**Scale**
True Values	1.00	0.50	−2.00	−0.50	−1.00
Menus 1-2	0.942 (0.025)	0.497 (0.022)	−1.997 (0.021)	−0.498 (0.035)	−1.022 (0.044)
Menus 1, 2, 3	0.965 (0.021)	0.496 (0.018)	−1.998 (0.019)	−0.49 (0.027)	−0.999 (0.021)
Menus 1, …, 4	0.961 (0.021)	0.497 (0.016)	−1.998 (0.018)	−0.476 (0.026)	−0.996 (0.02)
Menus 1, …, 5	0.971 (0.019)	0.494 (0.014)	−1.994 (0.015)	−0.494 (0.023)	−0.994 (0.014)
Menus 1, …, 16	0.986 (0.017)	0.496 (0.012)	−2.001 (0.011)	−0.507 (0.017)	−1.004 (0.008)
Variances					
	Transit	**Bike Sharing**	**On-Demand**	**Constant**	**Scale**
True Values	2.00	1.00	0.50	1.00	0.25
Menus 1-2	2.127 (0.112)	1.105 (0.068)	0.523 (0.041)	1.142 (0.109)	0.308 (0.089)
Menus 1, 2, 3	2.055 (0.09)	1.041 (0.052)	0.494 (0.029)	1.146 (0.084)	0.25 (0.037)
Menus 1, …, 4	2.064 (0.078)	1.008 (0.038)	0.494 (0.026)	1.063 (0.068)	0.242 (0.031)
Menus 1, …, 5	2.065 (0.072)	1.005 (0.036)	0.502 (0.024)	1.072 (0.055)	0.249 (0.022)
Menus 1, …, 16	2.066 (0.051)	1.028 (0.022)	0.514 (0.015)	1.032 (0.03)	0.254 (0.01)

Unlike the case with individual-specific parameters, the constant is biased downwards with 3-NN recommendations (while it is biased upwards with in Table 4).

Table 7
Endogeneity results using 3-NN recommendations, excluding the first 2 menus.

Population Means					
	Transit	**Bike Sharing**	**On-Demand**	**Constant**	**Scale**
True Values	1.00	0.50	−2.00	−0.50	−1.00
Menu 3	1.970 (0.067)	0.951 (0.046)	−1.568 (0.040)	−1.374 (0.149)	−1.019 (0.063)
Menus 3, 4	1.620 (0.038)	0.780 (0.027)	−1.619 (0.028)	−1.370 (0.063)	−0.859 (0.04)
Menus 3, 4, 5	1.436 (0.036)	0.680 (0.023)	−1.717 (0.025)	−1.096 (0.058)	−0.855 (0.026)
Menus 3, …, 16	1.025 (0.017)	0.509 (0.013)	−1.973 (0.012)	−0.550 (0.019)	−0.986 (0.009)
Variances					
	Transit	**Bike Sharing**	**On-Demand**	**Constant**	**Scale**
True Values	2.00	1.00	0.50	1.00	0.25
Menu 3	4.527 (0.616)	1.583 (0.202)	0.875 (0.113)	23.752 (3.57)	0.526 (0.148)
Menus 3, 4	2.86 (0.194)	1.333 (0.102)	0.498 (0.065)	5.959 (0.529)	0.581 (0.127)
Menus 3, 4, 5	2.536 (0.152)	1.285 (0.073)	0.512 (0.046)	3.311 (0.283)	0.319 (0.057)
Menus 3, …, 16	2.113 (0.054)	1.055 (0.027)	0.504 (0.014)	1.130 (0.041)	0.252 (0.009)

References

Abernethy, J., Evgeniou, T., Toubia, O., Vert, J.P., 2008. Eliciting consumer preferences using robust adaptive choice questionnaires. IEEE Trans. Knowl. Data Eng. 20 (2), 145–155.

Adamowicz, V., Boxall, P., 2001. Future Directions of Stated Choice Methods for Environment Valuation. Choice Experiments. A new approach to environmental valuation, London, pp. 1–6.

Akay, A., 2012. Finite-sample comparison of alternative methods for estimating dynamic panel data models. J. Appl. Econom. 27 (7), 1189–1204.

Akinc, D., Vandebroek, M., 2018. Bayesian estimation of mixed logit models: selecting an appropriate prior for the covariance matrix. J. Choice Model. 29, 133–151.

Ansari, A., Essegaier, S., Kohli, R., 2000. Internet recommendation systems. J. Mark. Res. 37 (3), 363–375.

Bateman, I.J., Burgess, D., Hutchinson, W.G., Matthews, D.I., 2008. Learning design contingent valuation (LDCV): NOAA guidelines, preference learning and coherent arbitrariness. J. Environ. Econ. Manag. 55 (2), 127–141.

Becker, F., Danaf, M., Song, X., Atasoy, B., Ben-Akiva, M., 2018. Bayesian estimator for Logit Mixtures with inter-and intra-consumer heterogeneity. Transp. Res. Part B Methodol. 117, 1–17.

Ben-Akiva, M., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand, vol. 9. MIT press.

Ben-Akiva, M., McFadden, D., Train, K., 2019. Foundations of stated preference elicitation: consumer behavior and choice-based conjoint analysis. Found. Trends® Econ. 10 (1–2), 1–144.

Berry, S., Levinsohn, J., Pakes, A., 1995. Automobile prices in market equilibrium. Econometrica: J. Econ. Soc. 841–890.

Blamey, R.K., Bennett, J.W., Morrison, M.D., 1999. Yea-saying in contingent valuation surveys. Land Econ. 126–141.

Blundell, R.W., Powell, J.L., 2004. Endogeneity in semiparametric binary response models. Rev. Econ. Stud. 71 (3), 655–679.

Bradley, M.A., Daly, A.J., 1993. New Analysis Issues in Stated Preference Research. PTRC-PUBLICATIONS-P, 75-75.

Chaptini, B.H., 2005. Use of Discrete Choice Models with Recommender Systems. Doctoral dissertation, Massachusetts Institute of Technology.

Danaf, M., Becker, F., Song, X., Atasoy, B., Ben-Akiva, M., 2019a. Online discrete choice models: applications in personalized recommendations. Decis. Support Syst. 119, 35–45.

Danaf, M., Atasoy, B., Ben-Akiva, M., 2019b. Observed and unobserved inter- and intra-consumer heterogeneity. In: Accepted for Presentation in the 6th International Choice Modelling Conference (ICMC) (Kobe, Japan).

Donkers, B.A., Melenberg, B.B., 2002. Testing Predictive Performance of Binary Choice Models (No. EI 2002-01).

Fisher, R.A., 1922. On the mathematical foundations of theoretical statistics. Philos. Trans. R. Soc. Lond. - Ser. A Contain. Pap. a Math. or Phys. Character 222 (594–604), 309–368.
Fowkes, T., 2007. The design and interpretation of freight stated preference experiments seeking to elicit behavioural valuations of journey attributes. Transp. Res. Part B Methodol. 41 (9), 966–980.
Fowkes, A.S., Shinghal, N., 2002. The Leeds adaptive stated preference methodology. Working Pap. Inst. Transp. Stud., Univ. Leeds 2001.
Ghosal, S., 1997. A review of consistency and convergence of posterior distribution. In: Varanashi Symposium in Bayesian Inference. Banaras Hindu University.
Guevara, C.A., 2015. Critical assessment of five methods to correct for endogeneity in discrete-choice models. Transp. Res. A Policy Pract. 82, 240–254.
Guevara, C.A., Ben-Akiva, M.E., 2006. Endogeneity in residential location choice models. Transp. Res. Rec.: J. Transp. Res. Board 60–66, 1977.
Guevara, C.A., Ben-Akiva, M.E., 2012. Change of scale and forecasting with the control-function method in logit models. Transp. Sci. 46 (3), 425–437.
Guevara, C.A., Hess, S., 2019. A control-function approach to correct for endogeneity in discrete choice models estimated on SP-off-RP data and contrasts with an earlier FIML approach by Train & Wilson. Transp. Res. Part B Methodol. 123, 224–239.
Guevara, C.A., Polanco, D., 2016. Correcting for endogeneity due to omitted attributes in discrete-choice models: the multiple indicator solution. Transportmetrica: Transp. Sci. 12 (5), 458–478.
Hauser, J.R., Toubia, O., 2005. The impact of utility balance and endogeneity in conjoint analysis. Mark. Sci. 24 (3), 498–507.
Hausman, J.A., 1978. Specification tests in econometrics. Econometrica: J. Econ. Soc. 1251–1271.
Heckman, J.J., 1977. Dummy Endogenous Variables in a Simultaneous Equation System. NBER Working Paper Series.
Heckman, J.J., 1979. Sample selection bias as a specification error. Econometrica: J. Econ. Soc. 153–161.
Heckman, J., 1981. The incidental parameters problem and the problem of initial conditions in estimating a discrete time - discrete data stochastic process. In: Manski, C., McFadden, D. (Eds.), Structural Analysis of Discrete Data with Econometric Applications. MIT Press, pp. 179–195.
Hess, S., Train, K., 2017. Correlation and scale in mixed logit models. J. Choice Model. 23, 1–8.
Holmes, T.P., Boyle, K.J., 2005. Dynamic learning and context-dependence in sequential, attribute-based, stated-preference valuation questions. Land Econ. 81 (1), 114–126.
Huang, A., Wand, M.P., 2013. Simple marginally noninformative prior distributions for covariance matrices. Bayesian Anal. 8 (2), 439–452.
Huber, J., Train, K., 2001. On the similarity of classical and Bayesian estimates of individual mean partworths. Mark. Lett. 12 (3), 259–269.
Jiang, H., Qi, X., Sun, H., 2014. Choice-based recommender systems: a unified approach to achieving relevancy and diversity. Oper. Res. 62 (5), 973–993.
Johnson, F.R., Kanninen, B., Bingham, M., Özdemir, S., 2006. Experimental design for stated-choice studies. In: Valuing Environmental Amenities Using Stated Choice Studies. Springer, Dordrecht, pp. 159–202.
Johnson, F.R., Lancsar, E., Marshall, D., Kilambi, V., Muehlbacher, A., Regier, D.A., et al., 2013. Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. Value Health 1, 3–13.
Kanninen, B.J., 1993a. Design of sequential experiments for contingent valuation studies. J. Environ. Econ. Manag. 25 (1), S1–S11.
Kanninen, B.J., 1993b. Optimal experimental design for double-bounded dichotomous choice contingent valuation. Land Econ. 69 (2).
Kanninen, B.J., 2002. Optimal design for multinomial choice experiments. J. Mark. Res. 39 (2), 214–227.
Koop, G., Poirier, D.J., Tobias, J.L., 2007. Bayesian Econometric Methods. Cambridge University Press.
Liu, Q., Otter, T., Allenby, G.M., 2007. Investigating endogeneity bias in marketing. Mark. Sci. 26 (5), 642–650.
Louviere, J., Train, K., Ben-Akiva, M.E., Bhat, C., Brownstone, D., Cameron, T.A., Carson, R.T., Deshazo, J.R., Fiebig, D., Greene, W., Hensher, D., 2005. Recent progress on endogeneity in choice modeling. Mark. Lett. 16 (3–4), 255–265.
Natter, M., Feurstein, M., 2002. Real world performance of choice-based conjoint models. Eur. J. Oper. Res. 137 (2), 448–458.
Neiswanger, W., Wang, C., Xing, E., 2013. Asymptotically Exact, Embarrassingly Parallel MCMC arXiv preprint arXiv:1311.4780.
Nyquist, H., 1992. Optimal designs of discrete response experiments in contingent valuation studies. Rev. Econ. Stat. 74 (3), 559–563.
Petrin, A., Train, K., 2010. A control function approach to endogeneity in consumer choice models. J. Mark. Res. 47 (1), 3–13.
Polydoropoulou, A., Lambrou, M.A., 2012. Development of an e-learning recommender system using discrete choice models and bayesian theory: a pilot case in the shipping industry. In: Security Enhanced Applications for Information Systems. IntechOpen.
Richardson, A.J., 2002. Simulation study of estimation of individual specific values of time by using adaptive stated-preference survey. Transp. Res. Rec. 1804 (1), 117–125.
Rouwendal, J., de Blaeij, A., Rietveld, P., Verhoef, E., 2010. The information content of a stated choice experiment: a new method and its application to the value of a statistical life. Transp. Res. Part B Methodol. 44 (1), 136–151.
Rubin, T., Steyvers, M., 2009. A topic model for movie choices and ratings. In: Proceedings of the 9th International Conference on Cognitive Modeling (Manchester).
Sawtooth Software, 2003. ACA/Hierarchical Bayes v2.0. Technical paper. http://www.sawtoothsoftware.com/technicaldownloads.
Shinghal, N., 1999. An Application of Stated Preference Methods to the Study of Intermodal Freight Transport Services in India. Doctoral dissertation, University of Leeds.
Song, X., 2018. Personalization of Future Urban Mobility (Doctoral dissertation). Massachusetts Institute of Technology.
Song, X., Atasoy, B., Ben-Akiva, M., 2017. Smart Mobility through Personalized Menu Optimization. Transportation Research Board Annual Meeting, Washington, DC, USA.
Song, X., Danaf, M., Atasoy, B., Ben-Akiva, M.E., 2018. Personalized menu optimization with preference updater: a Boston case study. Transp. Res. Rec. 2672 (8), 599–607.
Teo, C.H., Nassif, H., Hill, D., Srinivasan, S., Goodman, M., Mohan, V., Vishwanathan, S.V.N., 2016. Adaptive, personalized diversity for visual discovery. In: Proceedings of the 10th ACM Conference on Recommender Systems. ACM, pp. 35–38.
Tilahun, N.Y., Levinson, D.M., Krizek, K.J., 2007. Trails, lanes, or traffic: valuing bicycle facilities with an adaptive stated preference survey. Transp. Res. A Policy Pract. 41 (4), 287–301.
Toubia, O., Simester, D.I., Hauser, J.R., Dahan, E., 2003. Fast polyhedral adaptive conjoint estimation. Mark. Sci. 22 (3), 273–303.
Train, K.E., 2009. Discrete Choice Methods with Simulation. Cambridge university press (Chapter 12).
Train, K.E., Weeks, M., 2005. Discrete choice models in preference space and willingness-to-pay space. In: Applications of Simulation Methods in Environmental and Resource Economics. Springer, Dordrecht, pp. 1–16.
Train, K., Wilson, W., 2008. Estimation on stated-preference experiments constructed from revealed-preference choices. Transp. Res. Part B Methodol. 42 (3), 191–203.
Train, K., Wilson, W., 2009. Monte Carlo analysis of SP-off-RP data. J. Choice Model. 2 (1), 101–117.
Van der Vaart, A.W., 2000. Asymptotic Statistics, vol. 3. Cambridge university press.
Villas-Boas, J.M., Winer, R.S., 1999. Endogeneity in brand choice models. Manag. Sci. 45 (10), 1324–1338.
Walker, J.L., Wang, Y., Thorhauge, M., Ben-Akiva, M.E., 2018. D-efficient or deficient? A robustness analysis of stated choice experimental designs. Theory Decis. 84 (2), 215–238.
Wooldridge, J.M., 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. J. Appl. Econom. 20 (1), 39–54.
Wooldridge, J.M., 2010. Correlated Random Effects Models with Unbalanced Panels. Michigan State University, Department of Economics.