Journal of Choice Modelling 46 (2023) 100400

Contents lists available at ScienceDirect

Journal of Choice Modelling

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

A Bayesian instrumental variable model for multinomial choice with correlated alternatives

Hajime Watanabe [a,*], Takuya Maruyama [b]

[a] Department of Civil Engineering, The University of Tokyo Japan, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan
[b] Faculty of Advanced Science and Technology, Kumamoto University Japan, 2-39-1 Kurokami, Kumamoto, 860-8555, Japan

ARTICLE INFO

Keywords:
Multinomial probit
Endogeneity
Instrumental variable
Bayesian Markov chain Monte Carlo
Travel behavior

ABSTRACT

Endogeneity and correlated alternatives are major concerns to be addressed in travel behavior analysis. However, these issues have rarely been dealt with simultaneously in advanced discrete choice models. This study proposes a multinomial probit model that incorporates the instrumental variable method, namely, a fully parametric instrumental variable model for a multinomial choice. The proposed model has the following three characteristics: (1) it allows binary and/or continuous endogenous variables; (2) it allows any number of instrumental variables in each alternative; and (3) it allows positive and/or negative correlations between any choice alternatives. For parameter estimation, we also propose a Bayesian Markov chain Monte Carlo algorithm that can be accommodated in more extended model structures. The simulation study demonstrates that the proposed model addresses endogeneity while allowing correlations between the choice alternatives. Meanwhile, the simulation also implies that the users need to pay attention to the setting of the prior distribution when an endogenous variable of interest is binary, even if the sample size is moderate. The proposed model will be a useful tool in disciplines in which both endogeneity and correlations between choice alternatives are major concerns.

1. Introduction

Many discrete choice models have addressed issues that occur due to unobserved variables. Many studies have addressed the correlation between choice alternatives and endogeneity, which are major concerns in discrete choice modeling. Two alternative choices are correlated if they share unobserved variables, which does not meet the property of independence from irrelevant alternatives (IIA) assumed in the basic logit model. Additionally, if unobserved variables are correlated with an observed explanatory variable in a model specification, the corresponding parameter is inconsistently estimated, which is referred to as the endogeneity problem. Ignoring these correlations and endogeneity leads to a misspecification of the fundamental behavioral assumptions. Therefore, both the correlation and endogeneity caused by unobserved variables must be adequately addressed to make robust quantitative claims using discrete choice models.

However, in practice, the problems of these correlations and endogeneity seem to be rarely addressed simultaneously. Although they can be caused by a common factor (i.e., unobserved variables), their solutions have been developed and implemented separately. For example, non-IIA models, which are used when choice alternatives are likely to be correlated, cannot address the problem of endogeneity by themselves. Additionally, many empirical cases applying typical solutions for endogeneity in discrete choice models,

* Corresponding author.
E-mail addresses: hwatanabe@bin.t.u-tokyo.ac.jp (H. Watanabe), takumaru@kumamoto-u.ac.jp (T. Maruyama).

https://doi.org/10.1016/j.jocm.2023.100400
Received 14 June 2022; Received in revised form 6 December 2022; Accepted 18 January 2023
Available online 19 January 2023

for example, the control-function (CF) method, assume independent alternatives for simplicity. Incorporating such solutions in non-IIA models can be a challenge in terms of computational efficiency, especially when the non-IIA model itself is complicated. Therefore, although these correlations and endogeneity can occur simultaneously in reality, these two issues are rarely addressed simultaneously.

This study aims to connect the solutions for correlations and endogeneity methodologically in a new manner. Specifically, we propose a multinomial probit model incorporating the instrumental variable (IV) method, which may be referred to as an instrumental variable model for a multinomial choice. The error structure of the proposed model explicitly describes the correlations between alternatives and between explanatory variables and unobserved variables. Thereby, the model simultaneously addresses the correlations and endogeneity.

The proposed model can be flexibly extended to other models by accommodating the estimation algorithm to wider error structures. The error structure follows the variance-covariance matrix of a multivariate normal distribution that many existing choice models employ. Additionally, parameter estimation is based on the Bayesian Markov chain Monte Carlo (MCMC) method that has been used to estimate parameters in the multivariate normal distribution. Hence, the proposed model can be extended to an estimable and more specific one while addressing the correlations and endogeneity at the same time.

Although this paper shows a base model structure that can be extended to more advanced ones, the model has the following practical advantages:

- It allows binary and/or continuous endogenous variables.
- It allows any number of instrumental variables in each alternative.
- It allows positive and/or negative correlations between any choice alternatives.

The remainder of this paper is organized as follows. Section 2 reviews IV methods for addressing endogeneity in discrete choice models. Sections 3 and 4 present the proposed model and the MCMC algorithm, respectively. Section 5 demonstrates whether the proposed model can address endogeneity while allowing correlations between the choice alternatives. Finally, Section 6 summarizes this study.

2. Literature review

In this section, we first review the CF method addressing endogeneity in discrete choice models. Subsequently, we review existing instrumental variable models, that is, probit models that incorporate IV methods (Guevara, 2015; Rivers and Vuong, 1988).

2.1. Control-function method addressing endogeneity in discrete choice models

There are various methods for addressing endogeneity in discrete choice models. Many studies have applied these methods to cross-sectional survey data in the academic literature on transportation (Guevara et al., 2020; Guevara and Hess, 2019; Vij and Walker, 2014; Walker et al., 2011; Xu et al., 2017). Guevara (2015) reviews the following popular methods: the use of Proxys, the multiple indicator solution, the integration of latent variable, and IV methods (the two-step CF method and the simultaneous estimation of the CF method via Maximum-Likelihood (ML) estimation). He discussed the pros and cons of these methods and pointed out that ML estimation of the CF method can cost a computational burden with potential difficulties in identification.

Many studies have applied the CF method to address endogeneity in discrete choice models. However, most of these studies assume the property of IIA for the simplicity of the estimation, even though endogeneity and correlations between choice alternatives can often occur simultaneously in real-world data (Lurkin et al., 2017; Pike and Lubell, 2018; Ren et al., 2022; Tan and Xiao, 2021). A typical way to address both simultaneously is by incorporating the CF method in a mixed logit model (Li et al., 2018; Petrin and Train, 2010; Xu et al., 2017). However, extending it to more advanced ones can require more complex simulation procedures for parameter estimation. This can be computationally intensive because of the wider model structure or simply impossible due to identifiability reasons.

2.2. Probit models incorporating instrumental variable methods

Herein, we explain the existing instrumental variable models, namely, probit models that incorporate IV methods (Guevara, 2015; Rivers and Vuong, 1988). Note that the instrumental variable models are equivalent to the CF method via ML estimation, except that they are probit-based models rather than logit-based ones (Train, 2003). These two approaches provide a tradeoff of extendibility versus efficiency. For instance, the CF method via ML estimation provides closed-form choice probability expressions, which leads to easy and efficient parameter estimation. However, as we mentioned above, its extendibility is limited since logit-based modeling is relatively restrictive for the expression of the error structure.

Meanwhile, the instrumental variable models can benefit from probit-based modeling that brings a more flexible error structure. For example, various probit-based models are proposed to describe the unobserved dependencies between different discrete and continuous outcomes (Bhat, 2011; Bhat et al., 2016; Fang, 2008; Mondal and Bhat, 2022; Watanabe and Maruyama, 2022). This enables us to develop an extended choice model while addressing the endogeneity issue; however, its parameter estimation can be computationally intensive. Thus, the simulation for approximating the integrals or Bayesian estimation can be required.

The instrumental variable models can be roughly classified into (1) fully parametric and (2) semi- or non-parametric models. Specifically, fully parametric instrumental variable models require strong assumptions on the error structure, whereas semi- or non-parametric instrumental variable models do not. Thus, semiparametric instrumental variable models estimate causal effects more robustly than fully parametric models (Chesher, 2010; Chesher et al., 2013; Chesher and Rosen, 2017). However, such semi- or non-parametric models are quite complicated, which impedes empirical application (Guevara, 2015). For the same reason, it can be quite difficult to extend these semi- or non-parametric models to other models. Therefore, the remainder of this paper focuses on fully parametric instrumental variable models.

Fully parametric instrumental variable models are popular in political science. In particular, probit models incorporating IV methods have been frequently used in political science literature (Arceneaux and Nickerson, 2009; Gerber and Green, 2000). However, these probit models handle binary outcomes, not multinomial and other discrete ones. In the transportation literature, Brownstone and Fang (2014) employed a similar model to deal with endogenous residential density in vehicle ownership and utilization. Their proposed model is a multivariate ordered probit and Tobit model that incorporates IV methods. Following these existing studies, we propose a multinomial probit model that incorporates IV methods while allowing for correlations between alternatives. To the best of our knowledge, no such model currently exists.

Again, analysts can use the CF method incorporated with the mixed logit model when dealing with endogeneity and correlated alternatives simultaneously (Train, 2003; Villas-Boas and Winer, 1999), which is much easier than our proposed model. Nevertheless, this study provides useful information for developing a more extended choice model while addressing these two issues. Notably, we also propose a tailored MCMC algorithm for the Bayesian parameter estimation of the proposed and other extended models. In short, towards moving towards a more advanced model, this paper offers a base model and estimation procedure and examines their properties in a simulation study.

3. Model structure

In this section, we first introduce a simple binary probit model that incorporates instrumental variable methods fundamentally related to the model proposed in this study. Then, we explain the proposed model, extended from the binary probit model, while maintaining the incorporated instrumental variable methods.

3.1. An instrumental variable model for a binary outcome

In a discrete choice model, endogeneity is a situation in which an explanatory variable is correlated with the error term. In this case, the explanatory variable is referred to as the endogenous variable. To address the endogeneity, a straightforward approach explicitly describes the correlation between the endogenous variable and error term. As a typical example, Maddala (1983) presented the following a fully parametric instrumental variable model structure describing the correlation between a typical binary probit model and an endogenous variable:

$$\begin{aligned} y_i^* &= x_i^{'}\beta + z_i\delta + \varepsilon_i, y_i = 1 \text{ if } y_i^* > 0, \\ z_i &= w_i^{'}\alpha + \xi_i, \end{aligned} \tag{1}$$

where y_i^* is the latent utility of the binary probit model and x_i and z_i are vectors of the exogenous and endogenous variables for individual i, respectively. Here, we assume that endogenous variable z_i is continuous for simplicity. β and δ are the vectors of the corresponding parameters, w_i is a vector of instrumental variables that must be correlated with z_i but cannot be correlated with the error term ε_i, and α is the corresponding parameter. In this case, z_i and ε_i are correlated because z_i is an endogenous variable, which leads to an inconsistent estimate of δ. To describe this correlation explicitly, it is assumed that the error terms ε_i and ξ_i follow a bivariate normal distribution, with the covariance matrix shown below:

$$\begin{pmatrix} \varepsilon_i \\ \xi_i \end{pmatrix} = N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma \\ \sigma & \nu^2 \end{pmatrix}\right], \tag{2}$$

where σ is the covariance to capture the correlation which causes endogeneity. By freely estimating σ, this instrumental variable model can address endogeneity and estimate δ consistently. Otherwise, endogeneity renders the estimate of δ inconsistent. Therefore, the key point of the instrumental variable model is to describe explicitly how endogeneity occurs through a fully parametric covariance structure of a multivariate normal distribution.

3.2. An instrumental variable model for a multinomial outcome

We then extend the above instrument variable model for a binary choice to one for a multinomial choice that allows correlations between choice alternatives. To address endogeneity in a multinomial discrete choice in the same way, we explicitly describe how endogeneity occurs through the covariance of a multivariate normal distribution herein. First, let $\boldsymbol{y} = (y_1, y_2, \dots, y_n)$ denote multinomial discrete choice results from J alternatives for individuals $i \in \{1, 2, \dots, n\}$, where $y_i \in \{1, 2, \dots, J\}$. The proposed model assumes that the latent utilities determine the choice results $y_i^* = (y_{i1}^*, y_{i2}^*, \dots, y_{iJ}^*)'$, as illustrated below:

$$y_i = j \text{ if } \max\left(y_i^*\right) = y_{ij}^*, \tag{3}$$

where y^*_{ij} is the utility of choosing alternative j for individual i and is further specified as follows:

$$y^*_{ij} = x'_{ij}\beta_j + Z'_{ij}\Delta_j + \varepsilon_{ij}, \tag{4}$$

where x_{ij} and β_j are vectors of exogenous variables; the corresponding parameters, including the constant, $Z_{ij} = (z_{i1}, \ldots, z_{il}, \ldots, z_{iL_j})'$, $\Delta_j = (\delta_1, \ldots, \delta_l, \ldots, \delta_{L_j})'$, z_{il} and δ_l are vectors of l th endogenous variables and the corresponding parameters; and ε_{ij} is the error term. The proposed model allows any number of endogenous variables to be considered for each alternative j, where L_j is the set of endogenous variables in the specification of y^*_{ij}. Note that, L_j and the set of the other alternatives L_{-j} are disjoint sets; that is, $L_j \cap L_{-j} = \varnothing$ for all j. Thus, $L = \bigcup_{j=1}^{J} L_j$ where L is the set of the endogenous variables in the overall model specification, and hence $l \in L$. Then, we assume the endogenous variables to be specified as follows:

$$z_{il} = \begin{cases} z^*_{il} & \text{if } z_{il} \text{ is continuous} \\ I\{z^*_{il} > 0\} & \text{if } z_{il} \text{ is binary} \end{cases}, \tag{5}$$

where $z^*_{il} = w'_{il}\alpha_l + \xi_{il}$, w_{il} and α_l are vectors of instrumental variables and the corresponding parameters, including the constant, and ξ_{il} is the error term with zero mean and variance ν_l^2. Note that the variance ν_l^2 is fixed at 1 when z_{il} is binary.[1] $I\{\bullet\}$ denotes the indicator function defined as 1 if $\{\bullet\}$ is true; otherwise, it is 0.

As discussed in the previous section, the key point of the proposed model is to explicitly describe the error correlations between the error terms $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iJ})'$ and $\xi_i = (\xi_{i1}, \xi_{i2}, \ldots, \xi_{iL})'$, as follows:

$$\begin{pmatrix} \varepsilon_i \\ \xi_i \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_y & \Sigma_{y,z} \\ \Sigma_{y,z}^{\mathrm{T}} & \Sigma_z \end{pmatrix}\right], \tag{6}$$

where $\Sigma_{y,z} = (\Sigma_{y,z_1}, \ldots, \Sigma_{y,z_L})$, $\Sigma_{y,z_l} = (\sigma_{1l}, \ldots, \sigma_{jl}, \ldots, \sigma_{Jl})'$, σ_{jl} is a covariance parameter between ε_{ij} and ξ_{il}, $\Sigma_z = \mathrm{diag}(\nu_1^2, \ldots, \nu_L^2)$, and

$$\Sigma_y = \Sigma_{y|z} + \Sigma_{y,z}\Sigma_z^{-1}\Sigma_{y,z}^{\mathrm{T}}, \tag{7}$$

where

$$\Sigma_{y|z} = \begin{pmatrix} 1 & \gamma_{1,2} & \cdots & \gamma_{1,J-1} & \gamma_{1,J} \\ \gamma_{1,2} & 1 & \cdots & \gamma_{2,J-1} & \gamma_{2,J} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{1,J-1} & \gamma_{2,J-1} & \cdots & 1 & \gamma_{J-1,J} \\ \gamma_{1,J} & \gamma_{2,J} & \cdots & \gamma_{J-1,J} & 1 \end{pmatrix}, \tag{8}$$

where $\gamma_{p,q}$ is the correlation coefficient between ε_{ip} and ε_{iq}. This means that the proposed model describes any correlations between choice alternatives. Note that the diagonal elements of $\Sigma_{y|z}$ are fixed to 1 for simplicity, following Albert and Chib's (1993) probit model, which limits the flexibility of the error structure defined in Eq. (7). The diagonal elements are estimable in existing probit models (see Train (2003) for further details). However, it requires transforming the covariance matrix of errors to the covariance matrix of error differences for identifiability reasons (Daganzo, 1979). Unfortunately, this step will complicate the structure and estimation procedure of the instrumental variable model. Hence, the proposed model assumes the diagonal elements of $\Sigma_{y|z}$ are fixed to 1 to make the model structure and estimation simple in exchange for flexibility.

Here, we explain the specification of Σ_y in Eq. (7), in detail. An endogenous variable z_{il} is specified in the latent utility of only one alternative but not in the others. Thus, assuming an endogenous variable z_{il} in the latent utility y^*_{ij} of alternative j, the elements in Σ_{y,z_l} except for σ_{jl} are fixed to zero. Therefore, only the diagonal elements of Σ_y increase (i.e., the variance of error term ε_{ij} increases). In other words, the proposed model assumes that every variance of the error term ε_{ij} is fixed at one, excluding the amount of that increase. This assumption makes the estimation procedure quite efficient, as will be discussed in Section 4.

Finally, we summarize the characteristics of the proposed model. In short, the proposed model is a multinomial probit model that incorporates instrumental variable methods. As specified in Eq. (4), any number of endogenous variables can be accommodated in the latent utility. Additionally, both continuous and binary endogenous variables can be accommodated. The proposed model maintains the advantage of a typical multinomial probit model by allowing correlations between choice alternatives. Therefore, the proposed

[1] A typical example of binary endogenous variables in the context of a travel behavior analysis is a dummy variable of residential location that takes 1 if a respondent lives in a certain area and 0 otherwise. In such cases, the endogeneity issue arises owing to the residential self-selection problem (Mokhtarian and Cao, 2008).

model simultaneously addresses these correlations and endogeneity. It should be noted that, as illustrated in Eq. (7), the proposed model assumes that the variance of the error term ε_{ij} increases due to the correlations between the endogenous variables Z_{ij} and the error term ε_{ij}. Thus, analysts need to be careful when comparing the estimates of parameters β_j, Δ_j between choice alternatives. These estimated parameters are not directly comparable between alternatives since the variance of the error term ε_{ij}; that is, the diagonal elements of Σ_y can vary across alternatives.

3.3. The likelihood

For the proposed model, $y_i^* = (y_{i1}^*, y_{i2}^*, \ldots, y_{iJ}^*)'$ and $z_i^* = (z_{i1}^*, \ldots, z_{il}^*, \ldots, z_{iL}^*)'$ follow a multivariate normal distribution with mean μ and covariance matrix Σ, as illustrated below:

$$\begin{pmatrix} y_i^* \\ z_i^* \end{pmatrix} \sim N[\mu, \Sigma] = N\left[\begin{pmatrix} X_i\beta + Z_i\Delta \\ W_i\alpha \end{pmatrix}, \begin{pmatrix} \Sigma_y & \Sigma_{y,z} \\ \Sigma_{y,z}^{\mathrm{T}} & \Sigma_z \end{pmatrix}\right], \tag{9}$$

where $X_i = \mathrm{diag}(x_{i1}', x_{i2}', \ldots, x_{iJ}')$, $\beta = (\beta_1', \beta_2', \ldots, \beta_J')'$, $Z_i = \mathrm{diag}(Z_{i1}', Z_{i2}', \ldots, Z_{iJ}')$, $\Delta = (\Delta_1', \Delta_2', \ldots, \Delta_J')'$, $W_i = \mathrm{diag}(w_{i1}', \ldots, w_{il}', \ldots, w_{iL}')$, and $\alpha = (\alpha_1', \ldots, \alpha_l', \ldots, \alpha_L')'$. Let $\theta' = (\alpha', \beta', \gamma', \delta', \sigma')$ be the vector of the model parameters and the probability $f(y_i = j, z_i|\theta)$ where $z_i = (Z_{i1}', Z_{i2}', \ldots, Z_{iJ}') = (z_{i1}, \ldots, z_{il}, \ldots, z_{iL})$ can be expressed as

$$f(y_i = j, z_i|\theta) = \int_{y_i^*, z_i^* \in \mathscr{U}_i} \varphi(y_i^*, z_i^*|\mu, \Sigma)\, dy_i^* dz_i^*, \tag{10}$$

where $\mathscr{U}_i$ represents the set of values of (y_i^*, z_i^*) for which $(y_i = j, z_i)$ are realized simultaneously. Thus, the choice probability component in this definition is expressed as J-dimensional integrals over the J errors $(\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iJ})$. Note that the choice probabilities of many multinomial probit models are expressed as a function of the difference in utilities, namely, $(J-1)$-dimensional integrals over all possible values of the error differences (Daganzo, 1979; Train, 2003). As mentioned above, we assume that the diagonal elements of $\Sigma_{y|z}$ are fixed to 1 for simplicity, following Albert and Chib's (1993) multinomial probit model. Thus, we do not need to transform the covariance matrix of errors to the covariance matrix of error differences because the proposed model is identifiable without such a transformation step. Then, the likelihood is given by:

$$f(\boldsymbol{y}, z|\theta) = \prod_{j \in J} \prod_{i: y_i = j} f(y_i = j, z_i|\theta), \tag{11}$$

where $\boldsymbol{y}$ and z are the n-dimensional sample vectors of $(y_i = j, z_i)$ values, θ is a vector of the model parameters and $i : y_i = j$ indicates an individual i who chooses an alternative j.

4. Bayesian estimation

4.1. Prior

First, we specify the prior distributions of the parameters as $\beta_j \sim N(b_{0j}, B_{0j})$, $\gamma_{p,q} \sim N(0, \gamma_{0p,q})$, $\Delta_j \sim N(\delta_{0j}, \Delta_{0j})$, $\sigma_{jl} \sim N(0, \sigma_{0jl})$, $\alpha_l \sim N(a_{0l}, A_{0l})$, and $\nu_l^2 \sim IG(c_{0l}, d_{0l})$, where $N(\bullet)$ and $IG(\bullet)$ denote the normal and inverse-gamma distributions, respectively. In this paper, we set the prior distributions on the parameters to be diffuse following the related literature (Brownstone and Fang, 2014; Fang, 2008; van Hasselt, 2011). Specifically, we set b_{0j}, δ_{0j}, and a_{0l} to be a vector of zeros, and B_{0j}, Δ_{0j}, and A_{0l} to be a diagonal matrix with 100 on the diagonal. Additionally, we set c_{0l} and d_{0l} to be 3 and 6, respectively, and $\gamma_{0p,q}$ to be 0.5. Note that we test three different prior variances $\sigma_{0jl} = 0.1$, 0.3, 0.5 on σ_{jl} in a simulation study in Section 5. These prior settings on σ are based on the settings of existing studies, which employed relatively narrow prior distributions on the parameters in the error component of the multivariate normal distribution (Chib and Greenberg, 1998; Rajbhandari, 2014; van Hasselt, 2011; Watanabe and Maruyama, 2022). For example, Chib and Greenberg (1998) employed prior distributions with prior means of 0 and variances of 0.5 on the correlation parameters of the error component in the multivariate probit model. Additionally, Rajbhandari (2014) specified prior distribution with a prior mean of 0 and a variance of 0.063 on the error correlation parameter of a partially observable probit model. By doing so, we can secure the identification of the parameters even if the sample size is inadequate and the asymptotic properties are not fully exhibited (see Gelman and Shalizi, 2013; Rossi and Allenby, 2003 for further discussions).

4.2. Estimation framework

An application of Bayes' theorem leads to the joint posterior distribution as illustrated below:

$$f(\theta|\boldsymbol{y}, z) \propto f(\boldsymbol{y}, z|\theta) f(\theta), \tag{12}$$

where $f(\boldsymbol{y}, z|\theta)$ is the likelihood function in Eq. (9). However, the computational effort required to directly evaluate the likelihood increases exponentially with the dimensionality of the integral illustrated in Eq. (10). Thus, for computational efficiency, we augmented $\boldsymbol{y}^* = (y_1^*, y_2^*, \ldots, y_n^*)$. Therefore, we obtain the approximated joint posterior $f(\theta, \boldsymbol{y}^* | \boldsymbol{y}, z)$, which makes parameter estimation straightforward. This technique is known as data augmentation (Albert and Chib, 1993; Tanner and Wong, 1987) and enables us to efficiently estimate parameters of more advanced choice models which are capable of handling other outcomes at the same time (Brownstone and Fang, 2014; Fang, 2008; van Hasselt, 2011). In addition to $\boldsymbol{y}^*$, we augment $z_l^* = (z_{1l}^*, z_{2l}^*, \ldots, z_{nl}^*)$ if $z_l = (z_{1l}, z_{2l}, \ldots, z_{nl})$ is binary. The overall estimation algorithm is as follows.

Step 1. Set the initial values $\beta^{(0)}$, $\delta^{(0)}$, $\sigma^{(0)}$, $\alpha^{(0)}$, $\nu^{2\ (0)}$, $\gamma^{(0)}$ and go to Step 2 with $k = 1$.

Step 2. Sample $\boldsymbol{y}^{*(k)}$ using data augmentation.

Step 3. If z_l is binary, sample $z_l^{*(k)}$ using data augmentation and if z_l is continuous, $z_l^{*(k)} = z_l$, for all $l \in L$.

Step 4. Sample $\beta^{(k)}$, $\delta^{(k)}$, $\sigma^{(k)}$ from $N(h, H)$ using Gibbs sampling.

Step 5. Sample $\alpha_l^{(k)}$ from $N(g_l, G_l)$ using Gibbs sampling for all $l \in L$.

Step 6. If z_l is continuous, the sample $\nu_l^{2(k)}$ from $IG(c_l, d_l)$ using Gibbs sampling and if z_l is binary, $\nu_l^{2(k)} = 1$, for all $l \in L$.

Step 7. Sample $\gamma^{(k)}$ using Metropolis-Hastings sampling.

Step 8. Return to Step 2 until $k = K$.

We assume K is the total number of MCMC iterations $(k = 1, \ldots, K)$ and S is the number of retained iterations $(s = 1, \ldots, S)$ after a burn-in period. Each step of the algorithm is explained in detail in the next subsection.

The proposed estimation procedure can be applied to some advanced models with minor changes according to their extended model structures. For instance, when applied to instrumental variable models for ordered discrete and ranked discrete choices (Mondal and Bhat, 2022; Train, 2003), only the data augmentation of the latent utility in Step 2 will need to be changed to correspond to their choice rules. Similarly, only the same change in Step 2 will be required when we estimate the multivariate probit-based instrumental variable model (Chib and Greenberg, 1998; Choo and Mokhtarian, 2008). Any change will not be necessary except for the error structure when applied to the instrumental variable model with a structured covariance (Yai et al., 1997). Even when we estimate a more advanced model beyond these existing model structures, its estimation algorithm would have some common steps with the proposed one.

4.3. Posterior

We begin by sampling latent data y_i^* and z_{il}^*. In step 2 of the algorithm, we augment the latent utility y_i^* conditional on z_i. Specifically, the sampling of latent utility y_i^* is as follows:

$$y_i^* \Big| [z_i, \theta] \sim N\left(\mu_{y_i|z_i}, \Sigma_{y|z}\right), \text{s.t. } \max\left(y_i^*\right) = y_{ij}^* \text{ if } y_i = j, \tag{13}$$

where $\Sigma_{y|z}$ is illustrated in Eq. (8) and

$$\mu_{y_i|z_i} = X_i\beta + Z_i\Delta + \Sigma_{y,z}\Sigma_z^{-1}\left(z_i^* - W_i\alpha\right). \tag{14}$$

An efficient way to sample y_i^* is the method of Geweke (1991) to compose a cycle of J times Gibbs sampling steps for $(y_{i1}^*, y_{i2}^*, \ldots y_{ij}^*, \ldots, y_{iJ}^*)$. Specifically, in the j th step of this cycle, y_{ij}^* is sampled from $y_{ij}^* \Big| [y_{i-j}^*, z_i, \theta]$, which is a truncated univariate normal distribution whose lower bound is the maximum of the other latent utilities y_{i-j}^*. Thereby, y_i^* is directly sampled without an accept-reject sampling that repeats sampling y_i^* until the choice rule $\max(y_i^*) = y_{ij}^*$ is met.

Subsequently, in Step 3, we augment z_{il}^* conditional on y_i^* if z_{il} is binary. Here, we rewrite the overall model structure as Eq. (9) as

$$\begin{pmatrix} y_i^* \\ z_{i-l}^* \\ z_{il}^* \end{pmatrix} \sim N\left[\begin{pmatrix} X_i\beta + Z_i\Delta \\ W_{i-l}\alpha_{-l} \\ w_{il}'\alpha_l \end{pmatrix}, \begin{pmatrix} \Sigma_y & \Sigma_{y,z_{-l}} & \Sigma_{y,z_l} \\ \Sigma_{y,z_{-l}}^{\mathrm{T}} & \Sigma_{z_{-l}} & \mathbf{0} \\ \Sigma_{y,z_l}^{\mathrm{T}} & \mathbf{0}^{\mathrm{T}} & \nu_l^2 \end{pmatrix}\right] = N\left[\begin{pmatrix} \mu_{i-l} \\ w_{il}'\alpha_l \end{pmatrix}, \begin{pmatrix} \Sigma_{-l} & \Sigma_{-l,l} \\ \Sigma_{-l,l}^{\mathrm{T}} & \nu_l^2 \end{pmatrix}\right], \tag{15}$$

where $\mathbf{0}$ is the zero vector. Accordingly, we sample z_{il}^* from

$$\begin{aligned} z_{il}^* \Big| [y_i^*, z_{i-l}^*, \theta] &\sim TN_{(0,\infty)}\left(w_{il}'\alpha_l + \Sigma_{-l,l}^{\mathrm{T}}\Sigma_{-l}^{-1}\left(U_{i-l}^* - \mu_{i-l}\right), \nu_l^2 - \Sigma_{-l,l}^{\mathrm{T}}\Sigma_{-l}^{-1}\Sigma_{-l,l}\right), \text{if } z_{il} = 1, z_{il}^* \Big| [y_i^*, z_{i-l}^*, \theta] \\ &\sim TN_{(-\infty,0)}\left(w_{il}'\alpha_l + \Sigma_{-l,l}^{\mathrm{T}}\Sigma_{-l}^{-1}\left(U_{i-l}^* - \mu_{i-l}\right), \nu_l^2 - \Sigma_{-l,l}^{\mathrm{T}}\Sigma_{-l}^{-1}\Sigma_{-l,l}\right), \text{if } z_{il} = 0, \end{aligned} \tag{16}$$

where $U_{i-l}^* = (y_i^{*'}, z_{i-l}^{*'})'$. If z_{il} is continuous, $z_{il}^* = z_{il}$.

Subsequently, the posterior of (β, δ, σ) satisfies

$$f\left(\beta,\delta,\sigma\middle|\mathbf{y}^*,z,\theta_{-\beta,\delta,\sigma}\right)\propto f(\mathbf{y}^*|z,\theta)f(\beta,\delta,\sigma)=\left[\prod_{i\in n}f\left(y_i^*\middle|\mu_{y_i|z_i},\Sigma_{y|z}\right)\right]f(\beta,\delta,\sigma), \tag{17}$$

where $\theta_{-\beta,\delta,\sigma}$ are the collection of parameters, excluding (β,δ,σ). Thus, in Step 4, we sample (β,δ,σ) from

$$\beta,\delta,\sigma\left|\left[\mathbf{y}^*,z,\theta_{-\beta,\delta,\sigma}\right]\right.\sim N(h,H), \tag{18}$$

where

$$H=\left(H_0^{-1}+\sum_{i\in n}V_i'\Sigma_{y|z}^{-1}V_i\right)^{-1},h=H\left(H_0^{-1}h_0+\sum_{i\in n}V_i'\Sigma_{y|z}^{-1}Y_i^*\right), \tag{19}$$

$$H_0=\text{diag}\left(H_{01},\dots,H_{0j},\dots,H_{0J}\right),h_0=\left(h_{01}',\dots,h_{0j}',\dots,h_{0J}'\right)',H_{0j}=\begin{bmatrix}B_{0j}&\mathbf{0}&\mathbf{0}\\\mathbf{0}&\Delta_{0j}&\mathbf{0}\\\mathbf{0}&\mathbf{0}&\sigma_{0j}\end{bmatrix},h_{0j}=\begin{pmatrix}b_{0j}\\\delta_{0j}\\\mathbf{0}\end{pmatrix}, \tag{20}$$

$$\sigma_{0j}=\left(\sigma_{0j1}',\dots,\sigma_{jl}',\dots,\sigma_{jL_j}'\right)',V_i=\text{diag}\left(V_{i1},\dots,V_{ij},\dots,V_{iJ}\right),V_{ij}=\left\{x_{ij}',Z_{ij}',Z_{ij}^*-W_{ij}\alpha_j\right\},W_{ij}=\text{diag}\left(w_{i1}',\dots,w_{il}',\dots,w_{iL_j}'\right),$$
$$\text{and }\alpha_j=\left(\alpha_1',\dots,\alpha_l',\dots,\alpha_{L_j}'\right)'.$$

In Step 5, we sample α_l from

$$\alpha_l\left|\left[\mathbf{y}^*,z,\theta_{-\alpha_l}\right]\right.\sim N(a_l,A_l), \tag{21}$$

where

$$A_l=\left(A_{0l}^{-1}+\sum_{i\in n}w_{il}'\nu_l^{-2}w_{il}\right)^{-1},a_l=A_l\left(A_{0l}^{-1}a_{0l}+\sum_{i\in n}w_{il}'\nu_l^{-2}z_{il}^*\right). \tag{22}$$

Then, if z_l is continuous, we sample ν_l^2 from $IG(c_l,d_l)$

$$\text{where }c_l=c_{0l}+\frac{n}{2},d_l=d_{0l}+\frac{1}{2}\sum_{i\in n}\left(z_{il}^*-w_{il}'\alpha_l\right)^2. \tag{23}$$

If z_l is binary, $\nu_l^2=1$. This corresponds to Step 6. Finally, the posterior of γ is given by:

$$f\left(\gamma\middle|\mathbf{y}^*,z,\theta_{-\gamma}\right)\propto I(\gamma\in C)f(\gamma)f(\mathbf{y}^*,z^*|\theta), \tag{24}$$

where C is the region satisfying the positive-definite condition of Σ. This is not a standard distribution; hence, we cannot sample γ directly. Accordingly, in Step 7, sampling γ requires a Metropolis-Hastings (MH) step (Chib and Greenberg, 1995, 1998).

5. Simulation study

In this section, we estimate the proposed model using simulated data and examine whether it can address endogeneity while allowing correlations between choice alternatives. In addition, we conduct a prior sensitivity analysis that indicates how different prior settings affect the estimates of σ in the proposed model.

5.1. Data generation

Here, we assume three choice alternatives and one endogenous variable z_l for each alternative. A sample size of $n=3,000$ is generated using Eq. (9) with

$$\begin{aligned}y_{i1}^*&=1.0-0.5x_{i1}-0.5z_{i1}^*+\varepsilon_{i1},\\y_{i2}^*&=0.0-0.5x_{i2}+0.5z_{i2}^*+\varepsilon_{i2},\\y_{i3}^*&=-1.0-0.5x_{i3}+1.5z_{i3}^*+\varepsilon_{i3},\\z_{i1}^*&=1.0+1.0w_{i1}+\xi_{i1},\\z_{i2}^*&=1.0+1.0w_{i2}+\xi_{i2},\\z_{i3}^*&=1.0+1.0w_{i3}+\xi_{i3},\end{aligned} \tag{25}$$

where $x_{i1}\sim N(1,1)$, $x_{i2}\sim N(1,1)$, $x_{i3}\sim N(1,1)$, $w_{i1}\sim U(-1,1)$, $w_{i2}\sim U(-1,1)$, and $w_{i3}\sim U(-1,1)$. The error terms are assumed to follow a six-variate normal distribution:

$$\begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \xi_{i1} \\ \xi_{i2} \\ \xi_{i3} \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \Sigma_y & \Sigma_{y,z} \\ \Sigma_{y,z}^{\mathrm{T}} & \Sigma_z \end{pmatrix}\right] = N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1.09 & 0 & -0.3 & 0.3 & 0 & 0 \\ 0 & 1.09 & 0.3 & 0 & 0.3 & 0 \\ -0.3 & 0.3 & 1.09 & 0 & 0 & 0.3 \\ 0.3 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0.3 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0.3 & 0 & 0 & 1 \end{pmatrix}\right], \tag{26}$$

where $\Sigma_{y|z} = \begin{pmatrix} 1 & 0 & -0.3 \\ 0 & 1 & 0.3 \\ -0.3 & 0.3 & 1 \end{pmatrix}$, $\Sigma_{y,z} = \mathrm{diag}(0.3, 0.3, 0.3)$, $\Sigma_z = \mathrm{diag}(1,1,1)$,

and Σ_y is specified in Eq. (7). Thus, in this setting, there is a negative correlation between y_{i1}^* and y_{i3}^* and a positive correlation between y_{i2}^* and y_{i3}^*. Specifically, $\gamma_{1,3} = -0.3$ and $\gamma_{2,3} = 0.3$. Additionally, the endogenous variables z_{i1}^*, z_{i2}^*, and z_{i3}^* are correlated to the error terms ε_{i1}, ε_{i1}, and ε_{i3} via $\sigma = (\sigma_{11}, \sigma_{21}, \sigma_{31}) = (0.3, 0.3, 0.3)$, respectively. This means that endogeneity exists among the three latent utilities y_{i1}^*, y_{i2}^*, and y_{i3}^*.

Here, we examine the properties of the proposed model in the following two cases using single simulated data: (1) when all z_{il} is continuous ($z_{il} = z_{il}^*$) and (2) when all z_{il} is binary ($z_{il} = I\{z_{il}^* > 0\}$). Specifically, we check whether the proposed model can address endogeneity under the given assumption while allowing correlations between choice alternatives. In addition, we conduct a prior sensitivity analysis that indicates how prior settings affect the estimates of σ in the proposed model.

Table 1
Estimated results when z_l is continuous.

Variables [true value]	σ is freely estimated			σ is fixed to zero		
	Parameter	t-value		Parameter	t-value	
Multinomial discrete choice						
β_{10} [0.5]	0.43	5.36	**	0.20	2.95	**
β_{11} [-1.0]	−1.02	−20.89	**	−1.01	−19.40	**
δ_1 [-0.5]	−0.44	−6.14	**	−0.26	−7.20	**
β_{20} [0.0]	Fixed to 0			Fixed to 0		
β_{21} [-0.5]	−0.53	−13.19	**	−0.54	−13.55	**
δ_2 [0.5]	0.52	8.13	**	0.78	−18.35	**
β_{30} [-0.5]	−0.62	−5.63	**	−0.47	−5.32	**
β_{31} [0.5]	0.56	11.43	**	0.59	9.91	**
δ_3 [1.0]	1.03	11.33	**	1.37	14.00	**
$\gamma_{1,2}$ [0.0]	Fixed to 0			Fixed to 0		
$\gamma_{1,3}$ [-0.3]	−0.44	−3.73	**	−0.47	−2.37	*
$\gamma_{2,3}$ [0.3]	0.19	2.21	*	0.07	0.56	
Endogenous variables						
α_{10} [0.5]	0.49	−27.04	**	0.49	27.21	**
α_{11} [-1.0]	−1.00	−31.90	**	−1.00	−32.17	**
α_{20} [-0.5]	−0.51	−27.72	**	−0.51	−27.44	**
α_{21} [-1.0]	−1.00	−31.60	**	−1.00	−31.54	**
α_{30} [-1.0]	−1.01	−54.92	**	−1.01	−55.12	**
α_{31} [-1.0]	−0.96	−30.21	**	−0.96	−30.33	**
ι_1^2 [1.0]	0.99	38.90	**	0.99	38.76	**
ι_2^2 [1.0]	1.03	38.71	**	1.03	39.14	**
ι_3^2 [1.0]	1.00	38.73	**	1.00	38.74	**
Error covariances						
σ_{11} [0.3]	0.23	2.91	**	Fixed to 0		
σ_{21} [0.3]	0.36	4.84	**	Fixed to 0		
σ_{31} [0.3]	0.41	4.71	**	Fixed to 0		
Sample size	3,000			3,000		
Prior variance	0.50			–		

Notes: * represents p < 0.05, while ** represents p < 0.01.

6. Results and discussion

Table 1 shows the estimated results with estimated and fixed σ when z_l is continuous, and the prior variance is 0.5. When σ is fixed to zero, the estimates of $\delta = (\delta_1, \delta_2, \delta_3)$ are relatively far from the true values compared to the results when σ is freely estimated. This is because the endogeneity issue occurs when σ is fixed at zero, whereas it is adequately addressed by freely estimating σ. It is noteworthy that the estimates of the correlation coefficients are relatively close to the true values when σ is freely estimated. Meanwhile, the estimates of the correlation coefficients when σ is fixed to zero are relatively far from the true values. Thus, it may be said that the parameters of correlations between choice alternatives are affected by the endogeneity issue as well as the parameters of endogenous variables.

Likewise, Table 2 shows the estimated results when z_l is binary, and the prior variance is 0.5. Compared to the results when z_l is continuous, as illustrated in Table 1, the estimated parameters are relatively far from the true values, despite using the same samples. This is because the amount of information obtained from the data when z_l is binary is less than when z_l is continuous (i.e., information is abandoned when z_l^* is converted to z_l following $z_{il} = I\{z_{il}^* > 0\}$).

For the same reason, the dependence of the estimates on prior settings when z_l is binary is stronger than when z_l is continuous, as indicated in Tables 3 and 4. Tables 3 and 4 demonstrate how prior settings affect the estimates of σ when z_l is continuous and binary, respectively. Notably, Table 4 indicates that the estimates largely depend on prior settings when endogenous variables are binary. This is due to a potential parameter identification issue of the instrumental variable and other related models pointed out by Freedman and Sekhon (2010). This parameter identification issue arises even if the model itself is identifiable when the information obtained from data is limited; that is, the sample size is not so adequate that the asymptotic properties are fully exhibited (Hollenbach et al., 2019). When the asymptotic properties are not fully exhibited, priors influence the posteriors (Train, 2003). Therefore, the estimates can be affected by the prior information in this simulation. Hence, this simulation demonstrates that, in this setting with a sample size of 3, 000, analysts should be careful about the prior variance when z_l is binary. If this parameter identification issue arises, analysts should

Table 2
Estimated results when z_l is binary.

Variables [true value]	σ is freely estimated			σ is fixed to zero		
	Parameter	t-value		Parameter	t-value	
Multinomial discrete choice						
β_{10} [0.5]	0.38	2.03	*	0.36	3.38	**
β_{11} [-1.0]	−1.05	−16.42	**	−1.05	−17.50	**
δ_1 [-0.5]	−0.21	−0.87		0.01	0.08	
β_{20} [0.0]	Fixed to 0			Fixed to 0		
β_{21} [-0.5]	−0.58	−12.14	**	−0.56	−11.20	**
δ_2 [0.5]	0.58	3.51	**	0.98	10.61	**
β_{30} [-0.5]	−0.44	−4.14	**	−0.50	−5.69	**
β_{31} [0.5]	0.56	10.15	**	0.56	9.83	**
δ_3 [1.0]	0.59	2.41	*	1.57	10.23	**
$\gamma_{1,2}$ [0.0]	Fixed to 0			Fixed to 0		
$\gamma_{1,3}$ [-0.3]	−0.18	−1.08		−0.34	−2.04	*
$\gamma_{2,3}$ [0.3]	0.15	1.09		0.02	0.11	
Endogenous variables						
α_{10} [0.5]	0.48	18.82	**	0.48	18.92	**
α_{11} [-1.0]	−0.98	−21.27	**	−0.98	−21.53	**
α_{20} [-0.5]	−0.49	−19.06	**	−0.49	−19.15	**
α_{21} [-1.0]	−0.99	−21.14	**	−0.99	−21.44	**
α_{30} [-1.0]	−0.97	−31.14	**	−0.97	−33.11	**
α_{31} [-1.0]	−0.89	−16.33	**	−0.90	−17.37	**
ν_1^2 [1.0]	Fixed to 1			Fixed to 1		
ν_2^2 [1.0]	Fixed to 1			Fixed to 1		
ν_3^2 [1.0]	Fixed to 1			Fixed to 1		
Error covariances						
σ_{11} [0.3]	0.15	0.96		Fixed to 0		
σ_{21} [0.3]	0.30	2.63	**	Fixed to 0		
σ_{31} [0.3]	0.62	4.02	**	Fixed to 0		
Sample size	3,000			3,000		
Prior variance	0.50			–		

Notes: * represents $p < 0.05$, while ** represents $p < 0.01$.

Table 3
Prior sensitivity analysis when z_l is continuous.

Estimates [true value]	Prior variance $\sigma_{011} = \sigma_{021} = \sigma_{031}$		
	0.10	0.30	0.50
σ_{11} [0.3]	0.22	0.23	0.23
σ_{21} [0.3]	0.34	0.35	0.36
σ_{31} [0.3]	0.38	0.41	0.41
δ_1 [-0.5]	−0.42	−0.43	−0.44
δ_2 [0.5]	0.52	0.52	0.52
δ_3 [1.0]	1.02	1.03	1.03

Table 4
Prior sensitivity analysis when z_l is binary.

Estimates [true value]	Prior variance $\sigma_{011} = \sigma_{021} = \sigma_{031}$		
	0.10	0.30	0.50
σ_{11} [0.3]	0.13	0.16	0.15
σ_{21} [0.3]	0.26	0.28	0.30
σ_{31} [0.3]	0.52	0.60	0.62
δ_1 [-0.5]	−0.18	−0.22	−0.21
δ_2 [0.5]	0.61	0.60	0.58
δ_3 [1.0]	0.69	0.61	0.59

conduct a prior sensitivity analysis and explain the reason for their prior settings.

In summary, in this setting, we confirm that the proposed model addresses endogeneity while allowing correlations between choice alternatives. In addition, special attention to the prior settings on σ will be needed when z_l is binary and the sample size is insufficient.

7. Conclusion

In this study, we proposed a multinomial probit model that incorporates the instrumental variable method, that is, a fully parametric instrumental variable model for a multinomial choice. Specifically, the proposed instrumental variable model has the following three characteristics: (1) it allows binary and/or continuous endogenous variables; (2) it allows any number of instrumental variables in each alternative; and (3) it allows positive and/or negative correlations between any choice alternatives. We conducted the simulation study and confirmed that the proposed model addresses endogeneity properly while allowing correlations between alternatives. Additionally, we revealed that the users need to pay attention to the setting of the prior distribution when an endogenous variable of interest is binary, even if the sample size is moderate. While maintaining those three characteristics and estimability, the model structure can be extended to more advanced ones capable of handling other outcomes simultaneously because the Bayesian estimation framework employs a data augmentation technique.

The three characteristics will be helpful in real-world applications. For instance, some existing methods only assume continuous endogenous variables, although the endogenous variables of interest are not necessarily continuous in practice. Moreover, several (or more) explanatory variables can be endogenous in many empirical cases. Thus, it is generally recommended to incorporate more instrumental variables in a model specification. To this end, the proposed model can be useful because it can incorporate any number of instrumental variables correlated with binary and/or continuous endogenous variables. Additionally, in recent empirical travel behavior analyses, considering correlations between choice alternatives seems to have become more frequent. Thus, it is desirable not to assume independent choice alternatives, even if the research aims to address endogeneity. The proposed model will play an important role in disciplines where both endogeneity and correlations between choice alternatives are major concerns, such as transportation research.

Author contributions

Hajime Watanabe: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Software; Validation; Visualization; Writing - original draft; Takuya Maruyama: Funding acquisition; Project administration; Supervision; Writing - review & editing.

Declaration of competing interest

The authors declare no conflicts of interest.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work is supported by the Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for JSPS Fellows (Grant Numbers JP20J15157, JP22J01164) and JSPS KAKENHI (Grant Numbers JP18H01561, JP21H01458).

References

Albert, J.H., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. J. Am. Stat. Assoc. 88, 669–679. https://doi.org/10.1080/01621459.1993.10476321.

Arceneaux, K., Nickerson, D.W., 2009. Who is mobilized to vote? A re-analysis of 11 field experiments. Am. J. Pol. Sci. 53, 1–16. https://doi.org/10.1111/j.1540-5907.2008.00354.x.

Bhat, C.R., 2011. The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. Transp. Res. Part B Methodol. 45, 923–939. https://doi.org/10.1016/j.trb.2011.04.005.

Bhat, C.R., Astroza, S., Bhat, A.C., 2016. On allowing a general form for unobserved heterogeneity in the multiple discrete-continuous probit model: formulation and application to tourism travel. Transp. Res. Part B Methodol. 86, 223–249. https://doi.org/10.1016/j.trb.2016.01.012.

Brownstone, D., Fang, H., 2014. A vehicle ownership and utilization choice model with endogenous residential density. J. Transp. Land Use 7, 135. https://doi.org/10.5198/jtlu.v7i2.468.

Chesher, A., 2010. Instrumental variable models for discrete outcomes. Econometrica 78, 575–601. https://doi.org/10.3982/ecta7315.

Chesher, A., Rosen, A.M., 2017. Generalized instrumental variable models. Econometrica 85, 959–989. https://doi.org/10.3982/ecta12223.

Chesher, A., Rosen, A.M., Smolinski, K., 2013. An instrumental variable model of multiple discrete choice. Quant Econom 4, 157–196. https://doi.org/10.3982/qe240.

Chib, S., Greenberg, E., 1998. Analysis of multivariate probit models. Biometrika 85, 347–361. https://doi.org/10.1093/biomet/85.2.347.

Chib, S., Greenberg, E., 1995. Understanding the metropolis-hastings algorithm. Am. Statistician 49, 327–335. https://doi.org/10.1080/00031305.1995.10476177.

Choo, S., Mokhtarian, P.L., 2008. How do people respond to congestion mitigation policies? A multivariate probit model of the individual consideration of three travel-related strategy bundles. Trans. (Amst) 35, 145–163. https://doi.org/10.1007/s11116-007-9142-8.

Daganzo, C., 1979. Multinomial Probit. Academic Press, New York.

Fang, H.A., 2008. A discrete-continuous model of households' vehicle choice and usage, with an application to the effects of residential density. Transp. Res. Part B Methodol. 42, 736–758. https://doi.org/10.1016/j.trb.2008.01.004.

Freedman, D.A., Sekhon, J.S., 2010. Endogeneity in probit response models. Pol. Anal. 18, 138–150. https://doi.org/10.1093/pan/mpp037.

Gelman, A., Shalizi, C.R., 2013. Philosophy and the practice of Bayesian statistics. Br. J. Math. Stat. Psychol. 66, 8–38. https://doi.org/10.1111/j.2044-8317.2011.02037.x.

Gerber, A.S., Green, D.P., 2000. The effects of canvassing, telephone calls, and direct mail on voter turnout: a field experiment. Am. Polit. Sci. Rev. 94, 653–663.

Geweke, J., 1991. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities. In: Keramidas, E., Kaufman, S. (Eds.), Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface. Interface Foundation of North America, Fairfax Station, VA, pp. 571–578.

Guevara, C.A., 2015. Critical assessment of five methods to correct for endogeneity in discrete-choice models. Trans. Res. Part A Policy Pract 82, 240–254. https://doi.org/10.1016/j.tra.2015.10.005.

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. https://doi.org/10.1016/j.trb.2019.03.022.

Guevara, C.A., Tirachini, A., Hurtubia, R., Dekker, T., 2020. Correcting for endogeneity due to omitted crowding in public transport choice using the Multiple Indicator Solution (MIS) method. Trans. Res. Part A Policy Pract 137, 472–484. https://doi.org/10.1016/j.tra.2018.10.030.

Hollenbach, F.M., Montgomery, J.M., Crespo-Tenorio, A., 2019. Bayesian versus maximum likelihood estimation of treatment effects in bivariate probit instrumental variable models. Political Sci. Res. Methods 7, 651–659. https://doi.org/10.1017/psrm.2018.15.

Li, H., Li, X., Xu, X., Liu, J., Ran, B., 2018. Modeling departure time choice of metro passengers with a smart corrected mixed logit model - a case study in Beijing. Transport Pol. 69, 106–121. https://doi.org/10.1016/j.tranpol.2018.06.005.

Lurkin, V., Garrow, L.A., Higgins, M.J., Newman, J.P., Schyns, M., 2017. Accounting for price endogeneity in airline itinerary choice models: an application to Continental U.S. markets. Trans. Res. Part A Policy Pract 100, 228–246. https://doi.org/10.1016/j.tra.2017.04.007.

Maddala, G.S., 1983. Limited-dependent and Qualitative Variables in Econometrics. Cambridge University Press, Cambridge, UK.

Mokhtarian, P.L., Cao, X., 2008. Examining the impacts of residential self-selection on travel behavior: a focus on methodologies. Transp. Res. Part B Methodol. 42, 204–228. https://doi.org/10.1016/j.trb.2007.07.006.

Mondal, A., Bhat, C.R., 2022. A spatial rank-ordered probit model with an application to travel mode choice. Transp. Res. Part B Methodol. 155, 374–393. https://doi.org/10.1016/j.trb.2021.12.008.

Petrin, A., Train, K., 2010. A control function approach to endogeneity in consumer choice models. J. Market. Res. 47, 3–13. https://doi.org/10.1509/jmkr.47.1.3.

Pike, S., Lubell, M., 2018. The conditional effects of social influence in transportation mode choice. Res. Transport. Econ. 68, 2–10. https://doi.org/10.1016/j.retrec.2018.05.010.

Rajbhandari, A., 2014. Identification and MCMC estimation of bivariate probit models with partial observability. In: Jeliazkov, I., Yang, X.-S. (Eds.), Bayesian Inference in the Social Sciences. Wiley, Hoboken, New Jersey, pp. 299–316.

Ren, Y., Zhao, J., Glauben, T., Castro Campos, B., 2022. Supermarket environment and nutrition outcomes: evidence from rural China. J. Rural Stud. 92, 79–92. https://doi.org/10.1016/j.jrurstud.2022.03.019.

Rivers, D., Vuong, Q., 1988. Limited information estimators and exogeneity tests for simultaneous probit models. J. Econom. 39, 347–366. https://doi.org/10.1016/0304-4076(88)90063-2.

Rossi, P.E., Allenby, G.M., 2003. Bayesian statistics and marketing. Market. Sci. 22, 304–328. https://doi.org/10.1287/mksc.22.3.304.17739.

Tan, X., Xiao, Y., 2021. The effect of online reviews on addressing endogeneity in discrete choice models. Data Sci. Manag. 2, 1–11. https://doi.org/10.1016/j.dsm.2021.02.006.

Tanner, M.A., Wong, W.H., 1987. The calculation of posterior distributions by data augmentation. J. Am. Stat. Assoc. 82, 528–540. https://doi.org/10.1080/01621459.1987.10478458.

Train, K.E., 2003. Discrete Choice Methods with Simulation, Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511753930.

van Hasselt, M., 2011. Bayesian inference in a sample selection model. J. Econom. 165, 221–232. https://doi.org/10.1016/j.jeconom.2011.08.003.

Vij, A., Walker, J.L., 2014. Preference endogeneity in discrete choice models. Transp. Res. Part B Methodol. 64, 90–105. https://doi.org/10.1016/j.trb.2014.02.008.

Villas-Boas, J.M., Winer, R.S., 1999. Endogeneity in brand choice models. Manag. Sci. 45, 1324–1338. https://doi.org/10.1287/mnsc.45.10.1324.

Walker, J.L., Ehlers, E., Banerjee, I., Dugundji, E.R., 2011. Correcting for endogeneity in behavioral choice models with social influence variables. Trans. Res. Part A Policy Pract 45, 362–374. https://doi.org/10.1016/j.tra.2011.01.003.
Watanabe, H., Maruyama, T., 2022. Bayesian sample selection model with binary travel behavior for addressing residential self-selection. SSRN Electron. J. https://doi.org/10.2139/ssrn.4086629.
Xu, M., Meng, Q., Liu, K., Yamamoto, T., 2017. Joint charging mode and location choice model for battery electric vehicle users. Transp. Res. Part B Methodol. 103, 68–86. https://doi.org/10.1016/j.trb.2017.03.004.
Yai, T., Iwakura, S., Morichi, S., 1997. Multinomial probit with structured covariance for route choice behavior. Transp. Res. Part B Methodol. 31, 195–207. https://doi.org/10.1016/S0191-2615(96)00025-2.