Computational Statistics and Data Analysis 55 (2011) 1099–1108

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

# Estimation of sample selection models with two selection mechanisms

Phillip Li

*Department of Economics, University of California, Irvine, 3151 Social Science Plaza, Irvine CA 92697-5100, United States*

ARTICLE INFO

*Article history:*
Received 25 February 2010
Received in revised form 28 August 2010
Accepted 7 September 2010
Available online 21 September 2010

*Keywords:*
Sample selection
Markov chain Monte Carlo
Data augmentation

ABSTRACT

This paper focuses on estimating sample selection models with two incidentally truncated outcomes and two corresponding selection mechanisms. The method of estimation is an extension of the Markov chain Monte Carlo (MCMC) sampling algorithm from Chib (2007) and Chib et al. (2009). Contrary to conventional data augmentation strategies when dealing with missing data, the proposed algorithm augments the posterior with only a small subset of the total missing data caused by sample selection. This results in improved convergence of the MCMC chain and decreased storage costs, while maintaining tractability in the sampling densities. The methods are applied to estimate the effects of residential density on vehicle miles traveled and vehicle holdings in California.



## 1. Introduction

The seminal sample selection model of Heckman (1976, 1979) has generated a vast amount of theoretical and empirical research across a variety of disciplines. Sample selection, also referred to as incidental truncation, occurs when a dependent variable of interest is non-randomly missing for a subset of the sample as a result of a separate selection variable. A well-known application involves market wages (the outcome of interest) and labor force participation (the selection variable), in which wages are missing for individuals who are not participating in the labor force. Consequently, the remaining observations available to the researcher are non-random and do not represent the population of interest. As a result, estimation based only on this selected sample may lead to specification errors. This problem is ubiquitous in economics and disciplines that use observational data, therefore estimation techniques that address this issue are of substantial interest.

The conventional sample selection model with a single selection mechanism and its variants have been extensively estimated. Common classical estimation methods are developed and discussed in Amemiya (1984), Gronau (1973), Heckman (1976, 1979) and Wooldridge (1998, 2002), while semiparametric estimation and a variety of extensions are discussed in Heckman (1990), Manski (1989) and Newey et al. (1990). Extensions in the direction of multiple selection mechanisms are discussed in Shonkwiler and Yen (1999), Yen (2005) and Poirier (1980), where the two former articles discuss a model similar to that presented here, and the latter discusses observability of a single binary outcome as a result of two binary selection variables. The preceding procedures generally involve two classes of estimators: (1) two-step estimators that are consistent, asymptotically normal, but inefficient, and (2) maximum likelihood estimators that depend on evaluations of integrals. Puhani (2000) studies the practical performance of both classes of estimators using a Monte Carlo framework and criticizes their small sample properties. Alternatively, Bayesian estimation results in finite sample inference and avoids direct evaluations of integrals. Recent developments with one selection mechanism include Chib et al. (2009), Greenberg (2007) and van Hasselt (2009); extensions such as semiparametric estimation, endogeneity, and multiple outcome types are also discussed.

The model being analyzed contains a correlated system of equations with two continuous dependent variables of interest, each with an incidental truncation problem, and two corresponding selection variables. A major difference between this

*E-mail address:* phillil@uci.edu.

0167-9473/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.csda.2010.09.006

**Table 1**
Variable observability. The symbols ◯ and ✓ denote whether the variable is missing or observed in the sample partition, respectively.

| Variables | $N_1$ | $N_2$ | $N_3$ | $N_4$ |
|---|---|---|---|---|
| $y_{i,1}$ | ✓ | ◯ | ✓ | ◯ |
| $y_{i,2}$ | ✓ | ✓ | ◯ | ◯ |
| $y_{i,3}$ | ✓ | ✓ | ✓ | ✓ |
| $y_{i,4}$ | ✓ | ✓ | ✓ | ✓ |

model and much previous work is that there are two incidentally truncated outcomes being considered instead of one, resulting in four possible combinations of missing data for any observational unit and a complex pattern of missing data across the entire sample. The main contribution of this article is in the extension of the Markov chain Monte Carlo (MCMC) algorithm with "minimal data augmentation" to accommodate the nature of missing data for the model being analyzed. The minimal data augmentation technique first appeared in Chib (2007) to estimate the Bayesian version of the Roy model and was later extended in Chib et al. (2009) to estimate a semiparametric model with endogeneity and sample selection. This paper proposes an algorithm that only involves a minimal subset of the total missing data in the sampling scheme, resulting in improved convergence of the Markov chain and decreased storage costs, while maintaining tractability of the sampling densities without the complete data. The sampling densities are easy to draw from and result in samples that are close to iid for many parameters. A simulation study is included to study the performance of the algorithm.

The methods are applied to study the effects of residential density on vehicle miles traveled and vehicle holdings in California. A careful analysis is needed since data for vehicle miles traveled is only observable for households that own vehicles. The resulting estimation results will supplement the current literature and be informative for policy decisions.

## 2. Sample selection model

The model is given by

$$y_{i,1} = x'_{i,1}\beta_1 + \epsilon_{i,1}, \tag{1}$$

$$y_{i,2} = x'_{i,2}\beta_2 + \epsilon_{i,2}, \tag{2}$$

$$y^*_{i,3} = x'_{i,3}\beta_3 + \epsilon_{i,3}, \tag{3}$$

$$y^*_{i,4} = x'_{i,4}\beta_4 + \epsilon_{i,4}, \tag{4}$$

$$y_{i,j} = t_j \quad \text{if } \alpha_{t_j-1,j} < y^*_{i,j} \leq \alpha_{t_j,j}, \tag{5}$$

$$\delta_{t_j,j} = \ln\left\{\frac{\alpha_{t_j,j} - \alpha_{t_j-1,j}}{1 - \alpha_{t_j,j}}\right\}, \tag{6}$$

for observational units $i = 1, \ldots, N$, equations $j = 3, 4$, ordered categories $t_j = 1, \ldots, T_j$, ordered cutpoints $\alpha_{0,j} = -\infty < \alpha_{1,j} = 0 < \alpha_{2,j} = 1 < \alpha_{3,j} < \cdots < \alpha_{T_j-1,j} < \alpha_{T_j,j} = +\infty$, and transformed cutpoints $\delta_{t_j,j}$ for $t_j = 3, \ldots, T_j - 1$. The cutpoint restrictions are discussed in Section 3.3. The continuous dependent variables of interest are $y_{i,1}$ and $y_{i,2}$. Due to sample selection, their observability depends on the values of two ordered selection variables, $y_{i,3}$ and $y_{i,4}$ from (5), respectively. Following Albert and Chib (1993), the ordered variables are modeled in a threshold-crossing framework with the latent variables $y^*_{i,3}$ and $y^*_{i,4}$ according to Eqs. (3) through (5). In addition, a re-parameterization of the ordered cutpoints according to Eq. (6) is performed to remove the ordering constraints (Chen and Dey, 2000). The row vector $x'_{i,j}$ and conformable column vector $\beta_j$ are the exogenous covariates and corresponding regression coefficients, respectively. The vector of error terms $(\epsilon_{i,1}, \epsilon_{i,2}, \epsilon_{i,3}, \epsilon_{i,4})'$ is distributed independent multivariate normal, $\mathcal{N}(0, \Omega)$, where $\Omega$ is an unrestricted covariance matrix. This normality assumption for the error terms results in ordered probit models for Eqs. (3) through (5).

A key feature of the model is the inclusion of two incidentally truncated outcomes, which results in four cases of observability. For any observational unit $i$, only one of the following vectors is observed

$$(y_{i,1}, y_{i,2}, y_{i,3}, y_{i,4})', \qquad (y_{i,2}, y_{i,3}, y_{i,4})', \qquad (y_{i,1}, y_{i,3}, y_{i,4})', \quad (y_{i,3}, y_{i,4})', \tag{7}$$

where $y_{i,1}$ and $y_{i,2}$ are missing if and only if $y_{i,3}$ and $y_{i,4}$ are in known, application-specific categories $\gamma$ and $\lambda$, respectively. In the context of the empirical application, the mileage driven with trucks and cars ($y_{i,1}$ and $y_{i,2}$) are missing when the number of trucks and cars owned by the household ($y_{i,3}$ and $y_{i,4}$) equal zero, expressed as $y_{i,3} = \gamma = 0$ and $y_{i,4} = \lambda = 0$. The rules involving $y_{i,3}$ and $y_{i,4}$ that affect the observability are known as the selection mechanisms. To be specific about where incidental truncation occurs, let $N_r$ $(r = 1, \ldots, 4)$ denote partitions of the sample set that correspond to the four aforementioned cases of observability in (7). In addition, let $n_r$ denote their sizes such that $\sum_{r=1}^{4} n_r = N$. Using this notation, the variable $y_{i,1}$ is only observed for units in $N_1 \cup N_3$, and $y_{i,2}$ is only observed for units in $N_1 \cup N_2$, as illustrated in Table 1. Other quantities such as the ordered variables and explanatory variables are always observed.

**Table 2**
Minimal data augmentation scheme. The symbols ✓, ×, ⊗, and ◯ denote whether the variable is observed, latent but augmented, missing but augmented, or missing but not augmented in the posterior, respectively.

| Variables | $N_1$ | $N_2$ | $N_3$ | $N_4$ |
|---|---|---|---|---|
| $y_{i,1}$ | ✓ | ◯ | ✓ | ◯ |
| $y_{i,2}$ | ✓ | ✓ | ⊗ | ◯ |
| $y^*_{i,3}$ | × | × | × | × |
| $y^*_{i,4}$ | × | × | × | × |

## 3. Estimation

The proposed estimation algorithm uses MCMC methods with minimal data augmentation (MDA) based on Chib (2007) and Chib et al. (2009). The idea, motivation, and implementation of MDA are described in Section 3.1. Section 3.2 provides the data-augmented likelihood, priors, and data-augmented posterior. Section 3.3 presents the sampling algorithm in detail.

### 3.1. Minimal data augmentation (MDA)

The aim of MDA is to augment the posterior with the least amount of missing outcomes possible while keeping the densities of interest tractable for sampling. By introducing all the latent and missing data along the lines of Tanner and Wong (1987), many complex econometric models can be estimated as linear regression models with Gibbs or Metropolis–Hastings sampling (see Chapter 14 of Koop et al. (2007) for many examples). This approach is often desirable since given the "complete" data, the full conditional densities for $\tilde{\beta}$, $\Omega$, and other quantities are in standard forms (Chib and Greenberg, 1995). However, as noted in Chib et al. (2009), such a "naive" approach would degrade the mixing of the Markov chains and increase computation time. This problem is especially intensified when the quantity of missing outcomes due to the selection mechanism is large or when the model contains a sizable number of unknown parameters. Even if these impediments are disregarded, sample selection makes simulating the missing outcomes difficult as influential covariates may also be missing as a result of sample selection. For these reasons, it is generally desirable to minimize the amount of missing outcomes involved in the algorithm.

The proposed algorithm only augments the posterior with the missing variable $y_{i,2}$ in $N_3$ and the latent variables $\{y^*_{i,3}\, y^*_{i,4}\}$ for all observations, while leaving $y_{i,1}$ in $N_2 \cup N_4$ and $y_{i,2}$ in $N_4$ out of the sampler, as illustrated in Table 2. While the choices of variables and observations for augmentation appear arbitrary, they are specifically chosen to facilitate the sampling of $\Omega$ (see Section 3.3 for more details). By assuming that $y_{i,1}$ is missing more than $y_{i,2}$, this algorithm includes less than 50% of all missing data, which results in lower storage costs. In the vehicle choice application with 2297 observations, only 18% of the total missing data is used.

### 3.2. Posterior analysis

The data-augmented posterior density is proportional to the product of the data-augmented likelihood and the prior density for the unknown parameters:

$$\pi(\theta, y_{miss}, y^*|y_{obs}) \propto f(y_{obs}, y_{miss}, y^*|\theta)\pi(\theta). \tag{8}$$

Define the vector $\theta = (\tilde{\beta}, \delta, \Omega)$, where $\tilde{\beta} = (\beta_1', \beta_2', \beta_3', \beta_4')'$ and $\delta = \{\delta_{t_j,j}\}$, to contain all the unknown parameters. Also, define $y_{miss}$ and $y^*$ to respectively contain the augmented missing outcomes and latent variables from Table 2, and define $y_{obs}$ to contain all the observed data from Table 1.

Due to the intricate pattern of missing outcomes, specific quantities for each case of observability need to be defined. Let

$$\tilde{y}_{i,1:4} = (y_{i,1}, y_{i,2}, y^*_{i,3}, y^*_{i,4})', \qquad \tilde{y}_{i,2:4} = (y_{i,2}, y^*_{i,3}, y^*_{i,4})', \qquad \tilde{y}_{i,134} = (y_{i,1}, y^*_{i,3}, y^*_{i,4})', \qquad \tilde{y}_{i,3:4} = (y^*_{i,3}, y^*_{i,4})',$$

and using similar notation, let $\tilde{X}_{i,1:4}$, $\tilde{X}_{i,2:4}$, $\tilde{X}_{i,134}$, and $\tilde{X}_{i,3:4}$ be block-diagonal matrices with the corresponding vectors of covariates on the block diagonals and zeros elsewhere. Similarly, define $S'_{2:4}$, $S'_{134}$, and $S'_{3:4}$ to be conformable matrices that "select out" the appropriate regression coefficients when pre-multiplied to $\tilde{\beta}$. For example,

$$\tilde{X}_{i,3:4} = \begin{pmatrix} x'_{i,3} & 0 \\ 0 & x'_{i,4} \end{pmatrix}, \qquad S_{3:4} = \begin{pmatrix} 0 \\ I \end{pmatrix}, \quad \text{and} \quad S'_{3:4}\tilde{\beta} = \begin{pmatrix} \beta_3 \\ \beta_4 \end{pmatrix}.$$

Now, define and partition $\Omega$ and $\Omega_{22}$ as

$$\Omega = \begin{pmatrix} \underset{(1\times 1)}{\Omega_{11}} & \Omega_{12} \\ \Omega_{21} & \underset{(3\times 3)}{\Omega_{22}} \end{pmatrix}, \qquad \Omega_{22} = \begin{pmatrix} \underset{(1\times 1)}{\overline{\Omega}_{11}} & \overline{\Omega}_{12} \\ \overline{\Omega}_{21} & \underset{(2\times 2)}{\overline{\Omega}_{22}} \end{pmatrix},$$

and denote the covariance matrix for $\tilde{y}_{i,134}$ as $\Omega_{134}$.

The data-augmented likelihood needed in Eq. (8) is given by

$$f(y_{obs}, y_{miss}, y^*|\theta) \propto \prod_{N_1\cup N_3} \phi(\tilde{y}_{i,1:4}|\tilde{X}_{i,1:4}\tilde{\beta}, \Omega) \prod_{N_2} \phi(\tilde{y}_{i,2:4}|\tilde{X}_{i,2:4}S'_{2:4}\tilde{\beta}, \Omega_{22}) \times \prod_{N_4} \phi(\tilde{y}_{i,3:4}|\tilde{X}_{i,3:4}S'_{3:4}\tilde{\beta}, \overline{\Omega}_{22}) \prod_{i=1}^{N}\prod_{j=3}^{4} \mathbb{I}(\alpha_{y_{i,j}-1,j} < y^*_{i,j} \le \alpha_{y_{i,j},j}), \tag{9}$$

where $\phi(x|\mu, \Sigma)$ denotes the density of a multivariate normal distribution with mean vector $\mu$ and covariance matrix $\Sigma$, and $\mathbb{I}(\cdot)$ denotes an indicator function. The last product in (9) is the joint probability function of the ordered selection variables, which is known with certainty conditional on the latent variables. For some calculations, the data-augmented likelihood marginally of the missing outcomes is needed; it is obtained by integrating $\{y_{i,2}\}_{i\in N_3}$ out of Eq. (9) and is given by

$$f(y_{obs}, y^*|\theta) \propto \prod_{N_1} \phi(\tilde{y}_{i,1:4}|\tilde{X}_{i,1:4}\tilde{\beta}, \Omega) \prod_{N_2} \phi(\tilde{y}_{i,2:4}|\tilde{X}_{i,2:4}S'_{2:4}\tilde{\beta}, \Omega_{22}) \times \prod_{N_3} \phi(\tilde{y}_{i,134}|\tilde{X}_{i,134}S'_{134}\tilde{\beta}, \Omega_{134}) \prod_{N_4} \phi(\tilde{y}_{i,3:4}|\tilde{X}_{i,3:4}S'_{3:4}\tilde{\beta}, \overline{\Omega}_{22}) \prod_{i=1}^{N}\prod_{j=3}^{4} \mathbb{I}(\alpha_{y_{i,j}-1,j} < y^*_{i,j} \le \alpha_{y_{i,j},j}). \tag{10}$$

Prior independence is assumed for simplicity. Let

$$\tilde{\beta} \sim \mathcal{N}(\beta_0, B_0), \qquad \Omega \sim \mathcal{IW}(\nu_1, Q), \qquad \delta \sim \mathcal{N}(\delta_0, D_0), \tag{11}$$

where the priors for $\tilde{\beta}$ and $\delta$ are multivariate normal, and the prior for $\Omega$ is inverse-Wishart. The hyperparameters are set to reflect prior information. To be non-informative, set the mean vectors $\beta_0$ and $\delta_0$ to zeros, the covariance matrices $B_0$ and $D_0$ to diagonal matrices with 100 on the diagonals, $\nu_1$ to 4, and $Q$ to an identity matrix.

### 3.3. Sampling algorithm

For the computations that will follow, define $\delta_j$ and $\delta_{(-j)}$ to contain all the transformed cutpoints for equations $j$ and other than $j$, respectively. Similarly, define $y^*_j$ and $y^*_{(-j)}$ to contain the latent variables from $y^*$ for equations $j$ and other than $j$.

The posterior distribution is simulated by MCMC methods. The algorithm, which omits extraneous quantities from the conditioning set, is summarized as follows:

1. Sample $\tilde{\beta}$ from the distribution $\tilde{\beta}|y_{obs}, \Omega, y^*$.
2. Sample $(\delta_j, y^*_j)$ for $j = 3, 4$ from the distribution $\delta_j, y^*_j|y_{obs}, \tilde{\beta}, \Omega, \delta_{(-j)}, y^*_{(-j)}$.
3. Sample $\Omega$ from the distribution $\Omega|y_{obs}, \tilde{\beta}, y_{miss}, y^*$.
4. Sample $y_{i,2}$ for $i \in N_3$ from the distribution $y_{i,2}|y_{obs}, \tilde{\beta}, \Omega, y^*$.

Note that the quantities $\tilde{\beta}$, $\delta_j$, and $y^*_j$ are sampled without conditioning on the missing outcomes as this improves the mixing of the Markov chain. As the number of iterations approaches infinity, the draws can be shown to come from the posterior distribution of interest by collapsed MCMC theory (Liu, 1994) and Metropolis–Hastings convergence results (Chib and Greenberg, 1995; Tierney, 1994).

Identification in the ordered probit equations is achieved by imposing multiple cutpoint restrictions, following Fang (2008) and Jeliazkov et al. (2008). The cutpoints $\alpha_{1,j}$ and $\alpha_{2,j}$ are fixed at zero and one, respectively, along with $\alpha_{0,j} = -\infty$ and $\alpha_{T_j,j} = +\infty$. The proposed restrictions offer two advantages. First, the elements of $\Omega$ corresponding to the ordered variables are not restricted to be in correlation form, which allows for straightforward interpretation. Second, the transformed cutpoints do not need to be sampled when the selection variables only have three categories.

*Sampling $\tilde{\beta}$*

The conditional distribution for $\tilde{\beta}$ can be easily derived by combining (10) and the normal prior for $\tilde{\beta}$. By completing the square in the exponential functions, the distribution of interest can be recognized as $\mathcal{N}(\overline{\beta}, \overline{B})$, where

$$\overline{\beta} = \overline{B}\left(\begin{array}{c}\sum_{N_1} \tilde{X}'_{i,1:4}\Omega^{-1}\tilde{y}_{i,1:4} + \sum_{N_2} S_{2:4}\tilde{X}'_{i,2:4}\Omega_{22}^{-1}\tilde{y}_{i,2:4} + \\ \sum_{N_3} S_{134}\tilde{X}'_{i,134}\Omega_{134}^{-1}\tilde{y}_{i,134} + \sum_{N_4} S_{3:4}\tilde{X}'_{i,3:4}\overline{\Omega}_{22}^{-1}\tilde{y}_{i,3:4} + B_0^{-1}\beta_0\end{array}\right),$$

$$\overline{B} = \left(\begin{array}{c}\sum_{N_1} \tilde{X}'_{i,1:4}\Omega^{-1}\tilde{X}_{i,1:4} + \sum_{N_2} S_{2:4}\tilde{X}'_{i,2:4}\Omega_{22}^{-1}\tilde{X}_{i,2:4}S'_{2:4} + \\ \sum_{N_3} S_{134}\tilde{X}'_{i,134}\Omega_{134}^{-1}\tilde{X}_{i,134}S'_{134} + \sum_{N_4} S_{3:4}\tilde{X}'_{i,3:4}\overline{\Omega}_{22}^{-1}\tilde{X}_{i,3:4}S'_{3:4} + B_0^{-1}\end{array}\right)^{-1}.$$

*Sampling* $(\delta_j, y_j^*)$

The pair $(\delta_j, y_j^*)$ is sampled in one block from the joint distribution $\delta_j, y_j^*|y_{obs}, \tilde{\beta}, \Omega, \delta_{(-j)}, y^*_{(-j)}$ for $j = 3, 4$, as proposed in Albert and Chib (2001) and Chen and Dey (2000). The vector of transformed cutpoints $\delta_j$ is first sampled marginally of $y_j^*$ from $\delta_j|y_{obs}, \tilde{\beta}, \Omega, \delta_{(-j)}, y^*_{(-j)}$, and then $y_j^*$ is sampled conditionally on $\delta_j$ from $y_j^*|y_{obs}, \tilde{\beta}, \Omega, \delta, y^*_{(-j)}$. Sampling is performed jointly, because drawing $\delta_j$ and $y_j^*$ each from their full conditional distributions may induce high autocorrelation in the MCMC chains (Nandram and Chen, 1996).

The marginal distribution of $\delta_j$, recovered by integrating $y_j^*$ out of the joint distribution, is difficult to sample from directly. Instead, an independence chain Metropolis–Hastings step is used. A new draw, $\delta_j'$, is proposed from a multivariate $t$ distribution with $\nu_2 = 5$ degrees of freedom, $f_T(\delta_j|\hat{\delta}_j, \hat{D}_j, \nu_2)$, where $\hat{\delta}_j$ and $\hat{D}_j$ are the maximizer and negative Hessian of $f(y_j|y_{obs(-j)}, \tilde{\beta}, \Omega, \delta_j, y^*_{(-j)})\pi(\delta_j|\delta_{(-j)})$ evaluated at the maximum, respectively. The vectors $y_j$ and $y_{obs(-j)}$ respectively contain all elements in $y_{obs}$ associated with equations $j$ and other than $j$. The acceptance probability for $\delta_j'$ is

$$\alpha_{MH}(\delta_j, \delta_j') = \min\left\{1, \frac{f(y_j|y_{obs(-j)}, \tilde{\beta}, \Omega, \delta_j', y^*_{(-j)})\pi(\delta_j'|\delta_{(-j)})f_T(\delta_j|\hat{\delta}_j, \hat{D}_j, \nu_2)}{f(y_j|y_{obs(-j)}, \tilde{\beta}, \Omega, \delta_j, y^*_{(-j)})\pi(\delta_j|\delta_{(-j)})f_T(\delta_j'|\hat{\delta}_j, \hat{D}_j, \nu_2)}\right\}, \tag{12}$$

where the conditional probabilities of $y_j$ can be calculated as products of univariate normal distribution functions (Chib et al., 2009, Section 2.1).

By independence across observational units, the vector $y_j^*$ can be recovered by sampling $y_{i,j}^*$ from $y_{i,j}^*|y_{obs}, \tilde{\beta}, \Omega, \delta, y^*_{(-j)}$ for $i = 1, \ldots, N$. From Eq. (10), this distribution is truncated normal. Let $\mathcal{TN}(\mu, \sigma^2, a, b)$ denote a univariate normal distribution truncated to the region $(a, b)$ with mean $\mu$ and variance $\sigma^2$. The distribution of interest is given by

$$y_{i,j}^*|y_{obs}, \tilde{\beta}, \Omega, \delta, y^*_{(-j)} \sim \mathcal{TN}(\mu_{i,j}, \sigma^2_{i,j}, \alpha_{y_{i,j-1},j}, \alpha_{y_{i,j},j}), \tag{13}$$

where $\mu_{i,j}$ and $\sigma^2_{i,j}$ are the conditional mean and variance for a normal distribution.

*Sampling* $\Omega$

Due to the non-standard form of the posterior density in Eq. (8), the covariance matrix $\Omega$ cannot be sampled in one block from the usual inverse-Wishart distribution. Instead, one-to-one transformations of $\Omega$ and $\Omega_{22}$ will be sampled and used to construct a draw for $\Omega$. The presented derivations are an extension of Chib et al. (2009) by applying two sets of transformations instead of one due to the additional incidentally truncated outcome.

Define the transformations

$$\Omega_{11\cdot 2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}, \qquad B_{21} = \Omega_{22}^{-1}\Omega_{21}, \qquad \overline{\Omega}_{11\cdot 2} = \overline{\Omega}_{11} - \overline{\Omega}_{12}\overline{\Omega}_{22}^{-1}\overline{\Omega}_{21}, \qquad \overline{B}_{21} = \overline{\Omega}_{22}^{-1}\overline{\Omega}_{21},$$

and partition $Q$ and $Q_{22}$ as

$$Q = \begin{pmatrix} \underset{(1\times 1)}{Q_{11}} & Q_{12} \\ Q_{21} & \underset{(3\times 3)}{Q_{22}} \end{pmatrix}, \qquad Q_{22} = \begin{pmatrix} \underset{(1\times 1)}{\overline{Q}_{11}} & \overline{Q}_{12} \\ \overline{Q}_{21} & \underset{(2\times 2)}{\overline{Q}_{22}} \end{pmatrix}.$$

To sample $\Omega_{22}$, a change of variables from $\Omega_{22}$ to $(\overline{\Omega}_{22}, \overline{\Omega}_{11\cdot 2}, \overline{B}_{21})$ is applied to the density $\Omega_{22}|y_{obs}, \tilde{\beta}, y^*$ with Jacobian $|\overline{\Omega}_{22}|$. The resulting density is proportional to a product of three recognizable distribution kernels, namely two inverse-Wisharts and one matrix normal. They are

$$\overline{\Omega}_{22}|y_{obs}, \tilde{\beta}, y^* \sim \mathcal{IW}\left(\nu_1 + N - 1, \overline{Q}_{22} + \sum_{i=1}^{N} \tilde{\epsilon}_{i,3:4}\tilde{\epsilon}'_{i,3:4}\right), \tag{14}$$

$$\overline{\Omega}_{11\cdot 2}|y_{obs}, \tilde{\beta}, y^* \sim \mathcal{IW}(\nu_1 + n_1 + n_2, \overline{R}_{11\cdot 2}), \tag{15}$$

$$\overline{B}_{21}|\overline{\Omega}_{11\cdot 2}, y_{obs}, \tilde{\beta}, y^* \sim \mathcal{MN}_{(2\times 1)}(\overline{R}_{22}^{-1}\overline{R}_{21}, \overline{\Omega}_{11\cdot 2} \otimes \overline{R}_{22}^{-1}), \tag{16}$$

where $\tilde{\epsilon}_{i,3:4} = (\tilde{y}_{i,3:4} - \tilde{X}_{i,3:4}S'_{3:4}\tilde{\beta})$, $\tilde{\epsilon}_{i,2:4} = (\tilde{y}_{i,2:4} - \tilde{X}_{i,2:4}S'_{2:4}\tilde{\beta})$, $R_{22} = (Q_{22} + \sum_{N_1 \cup N_2} \tilde{\epsilon}_{i,2:4}\tilde{\epsilon}'_{i,2:4})$ is partitioned to be conformable with $Q_{22}$ using similar notation, and $\overline{R}_{11\cdot 2} = \overline{R}_{11} - \overline{R}_{12}\overline{R}_{22}^{-1}\overline{R}_{21}$. By drawing from (14) to (16) and manipulating the inverted quantities, a draw of $\Omega_{22}$ marginally of the missing data can be recovered.

To sample $\Omega$, a similar change of variables from $\Omega$ to $(\Omega_{22}, \Omega_{11\cdot 2}, B_{21})$ is applied to $\Omega|y_{obs}, \tilde{\beta}, y_{miss}, y^*$ with a Jacobian of $|\Omega_{22}|$. The resulting distributions of interest are

$$\Omega_{11\cdot 2}|y_{obs}, \tilde{\beta}, y_{miss}, y^* \sim \mathcal{IW}(\nu_1 + n_1 + n_3, R_{11\cdot 2}), \tag{17}$$

$$B_{21}|\Omega_{11\cdot 2}, y_{obs}, \tilde{\beta}, y_{miss}, y^* \sim \mathcal{MN}_{(3\times 1)}(R_{22}^{-1}R_{21}, \Omega_{11\cdot 2} \otimes R_{22}^{-1}), \tag{18}$$

where $\tilde{\epsilon}_{i,1:4} = (\tilde{y}_{i,1:4} - \tilde{X}_{i,1:4}\tilde{\beta})$, $R = (Q + \sum_{N_1 \cup N_3} \tilde{\epsilon}_{i,1:4}\tilde{\epsilon}'_{i,1:4})$ is partitioned to be conformable with $Q$, and $R_{11 \cdot 2} = R_{11} - R_{12}R_{22}^{-1}R_{21}$. The covariance matrix $\Omega$ can now be recovered using draws from (14) to (18).

*Sampling* $\tilde{y}_{i,2}$

From (8), the conditional distributions of $\tilde{y}_{i,2}$ are easily recognized as

$$y_{i,2}|y_{obs}, \tilde{\beta}, \Omega, y^* \sim \mathcal{N}(\eta_i, \omega_i^2) \quad \text{for } i \in N_3, \tag{19}$$

where $\eta_i$ and $\omega_i^2$ are the conditional mean and variance of $y_{i,2}$.

## 4. Simulation Study

This section evaluates the performance of the MDA algorithm from Section 3.3 using simulated data. For reference, the algorithm is compared to a similar one that augments all the missing data from Table 1, denoted as $y^*_{miss}$, and conditions on them at every step. Specifically, this benchmark algorithm with full data augmentation (FDA) recursively samples from $[\tilde{\beta}|y_{obs}, y^*_{miss}, \Omega, y^*]$, $[\delta_j, y^*_j|y_{obs}, y^*_{miss}, \tilde{\beta}, \Omega, \delta_{(-j)}, y^*_{(-j)}]$, $[\Omega|y_{obs}, y^*_{miss}, \tilde{\beta}, y^*]$, and $[y^*_{miss}|y_{obs}, \tilde{\beta}, \Omega, \delta, y^*]$, which are densities similar to those found in Section 3.3, but without the missing data integrated out.

The model being considered is from Eqs. (1) to (6). It contains five explanatory variables for each equation (including the constant), three ordered categories for each selection variable, and $N = 500$ observations. The unknown parameters are chosen so that the pattern of missing data is similar to the application in Section 5, where roughly 50% and 25% of the outcomes are missing from Eqs. (1) and (2), respectively. As a result, the MDA algorithm only augments about 20% of the total missing data. The covariance matrix $\Omega$ is set to $I + 0.7ii'$, where $i$ is a column of ones, implying a correlation of about 0.4 between the equations. Let $\omega_{ij}$ denote the element of $\Omega$ corresponding to the $i$-th row and $j$-th column. The prior hyperparameters are chosen to be non-informative, as discussed in Section 3.2. Note that various combinations of observations, explanatory variables, and proportions of missing data were considered, but the results are not presented as they did not vary much from the performance pattern in this base case.

The algorithms are iterated 12,500 times with a burn-in of 2500 draws. The MDA algorithm takes 17 s to perform 1000 iterations on average, and the FDA algorithm takes 19 s. Following Chib et al. (2009), the performance is studied with inefficiency factors over 30 Monte Carlo replications. The inefficiency factor for the $k$-th parameter is defined as $1 + 2\sum_{l=1}^{L} \rho_k(l)(1 - \frac{l}{L})$, where $\rho_k(l)$ is the sample autocorrelation at the $l$-th lag for the $k$-th parameter, and $L$ is the lag in which the autocorrelations taper off (Chib, 2001). This quantity measures the efficiency loss when using correlated MCMC samples instead of independent samples; values close to 1 generally indicate an efficient sampler.

Before the results are discussed, note that the sampling density for $\overline{\Omega}_{22}$ does not depend on $y_{miss}$, while the densities involving the other elements of $\Omega$ do depend on the missing data through $\tilde{y}_{i,1:4}$ and $\tilde{y}_{i,2:4}$ in $R$ and $R_{22}$, respectively. This suggests that the benefits of MDA relative to FDA will be more evident for the elements of $\Omega$ involving Eqs. (1) and (2) since they depend on the missing data. Using a similar argument, the elements of $\beta_1$ and $\beta_2$ from $\tilde{\beta}$ are also expected to have larger gains in performance relative to $\beta_3$ and $\beta_4$.

Boxplots of the inefficiency factors from both algorithms are displayed in Fig. 1. The plots for $\tilde{\beta}$ suggest that these parameters are generally sampled efficiently, except for the constant in $\beta_1$ due to the large fraction of missing outcomes in Eq. (1). The median inefficiency factors for $\beta_3$ and $\beta_4$ (excluding the constant) are between 1.1 and 1.4 under both algorithms. As previously discussed, the two algorithms perform similarly here, although the inefficiency factors are slightly lower for MDA. For $\beta_1$ and $\beta_2$, the median inefficiency factors are around 1 under MDA, but fall between 1.5 and 3 under FDA, indicating that MDA provides very efficient draws. Even with the inclusion of more explanatory variables (e.g. 21 variables for each equation, as in the empirical application), the resulting inefficiency factors under MDA are close to 1 (the results are not included due to graphical limitations). For $\Omega$, the inefficiency factors are predictably lower when using MDA. For example, the median factors for $\omega_{11}$ and $\omega_{31}$ decrease from 9 to 3 and 17 to 4, respectively.

To show that the sampling of $\Omega$ is correct under MDA, the posterior means for $vech(\Omega)$ from one of the Monte Carlo replication are $(1.67, 0.76, 1.74, 0.74, 0.67, 1.69, 0.71, 0.71, 0.65, 1.64)'$. The corresponding posterior standard deviations are $(0.09, 0.06, 0.06, 0.08, 0.05, 0.08, 0.06, 0.05, 0.04, 0.06)'$. As the number of observations increase, the posterior means approach the true $\Omega$ from the data generating process.

Overall, the median and average inefficiency factors for all the parameters estimated using the MDA algorithm are less than or equal to their FDA counterparts. This result is consistent with the notion that, in this context, data augmentation is only used to increase the tractability of the sampling densities, so integrating them out of the densities should not reduce the performance of the algorithm. For the majority of the parameters, the MDA algorithm results in lower inefficiency factors, indicative of both lower autocorrelations between the MCMC draws and of improved sampler performance when only a minimal subset of missing data is included. This result is especially evident for the parameters that are highly dependent on the missing data. As for the remaining parameters, they are efficiently estimated in both algorithms.

## 5. Application

Studies suggest that higher urban spatial structure, including residential density, is related to lower vehicle usage (Brownstone et al., 2009; Brownstone and Golob, 2009; Cervero and Kockelman, 1997; Dunphy and Fisher, 1996; Fang, 2008). As a

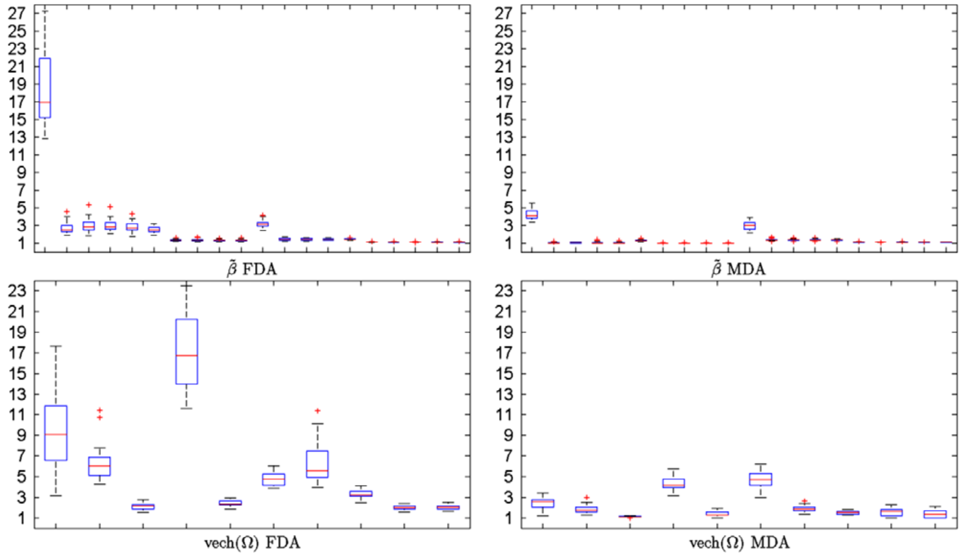


**Fig. 1.** Boxplots of inefficiency factors using the FDA and MDA algorithms for $\tilde{\beta} = (\beta_1', \beta_2', \beta_3', \beta_4')'$ and $vech(\Omega) = (\omega_{11}, \omega_{21}, \omega_{22}, \omega_{31}, \omega_{32}, \omega_{33}, \omega_{41}, \omega_{42}, \omega_{43}, \omega_{44})'$.

result, residential density is one parameter in reducing fuel consumption of automobiles or influencing household travel behavior. Policies targeting residential density can complement traditional ones such as limiting vehicle usage by total mileage driven or enforcing fuel efficiency on vehicles. Improved understanding of this relationship can also influence city development, zoning decisions, congestion growth, and project evaluations. However, vehicle usage data commonly contains a large proportion of missing values due to the lack of vehicle ownership. If these missing values are not modeled correctly or simply omitted from the sample, estimates of interest will suffer from misspecification errors.

The sample selection model is used to jointly study the effects of residential density on vehicle usage and holdings in California. One possible causal relationship suggests that denser areas increase the cost of operating vehicles. Residential areas with more houses per square mile commonly have narrow streets, congested roads, and limited parking spaces, which contribute to higher vehicle fuel consumption and operating costs when traveling around these neighborhoods, especially for less fuel-efficient vehicles. As a result, households will tend to drive less on average and switch to more fuel-efficient vehicles. The data is obtained from the 2001 National Household Travel Survey, from which a subsample 2297 households from California is used. Table 3 provides detailed summary statistics. Outcomes of interest are the annual mileage driven with trucks and cars (measures of vehicle usage) and the number of trucks and cars owned by a household (measures of vehicle holdings). They are modeled jointly with exogenous covariates such as residential density, household size, income, home ownership status, and education levels.

The model is given by

$$\begin{aligned} y_{i,1} &= \beta_{0,1} + \log(DENSITY_i)\beta_{1,1} + x_i'\beta_1 + \epsilon_{i,1}, \\ y_{i,2} &= \beta_{0,2} + \log(DENSITY_i)\beta_{1,2} + x_i'\beta_2 + \epsilon_{i,2}, \\ y_{i,3}^* &= \beta_{0,3} + \log(DENSITY_i)\beta_{1,3} + x_i'\beta_3 + \epsilon_{i,3}, \\ y_{i,4}^* &= \beta_{0,4} + \log(DENSITY_i)\beta_{1,4} + x_i'\beta_4 + \epsilon_{i,4}, \end{aligned} \tag{20}$$

for $i = 1, \ldots, 2297$ households, where $y_{i,1}$ and $y_{i,2}$ are annual mileage driven with trucks and cars, $y_{i,3}^*$ and $y_{i,4}^*$ are the latent variable representations for the number of trucks and cars owned ($y_{i,3}$ and $y_{i,4}$), and $x_i'$ is a row vector of exogenous covariates from Table 3. The equation subscript $j$ is omitted from $x_i'$ since the same covariates are used in every equation, and the covariate $\log(DENSITY_i)$ is separated to emphasize that it is a variable of interest. The error structure is $(\epsilon_{i,1}, \epsilon_{i,2}, \epsilon_{i,3}, \epsilon_{i,4})' \sim \mathcal{N}(0, \Omega)$. The selection variables are the number of trucks and cars a household owns, which have categories of zero, one, or more than two. Sample selection is modeled as follows: $y_{i,1}$ is observed if and only if $y_{i,3} > 0$, and $y_{i,2}$ is observed if and only if $y_{i,4} > 0$. Grouping households that own more than two trucks and cars (2.26% and 4.48% of the sample, respectively) with households that own two trucks and cars is for estimation convenience, because the transformed cutpoints do not need to be sampled. The two combined groups are assumed to be similar, so this grouping should not affect the analysis.

The model estimates are in Table 4, and the marginal effects with respect to residential density are in Table 5. The quantities of interest are obtained by iterating the algorithm in Section 3.3 10,000 times, discarding the first 10,000 iterations for burn-in, and taking the ergodic averages over the associated draws. Prior hyperparameters are set to reflect non-informativeness, since the effects of residential density and other covariates are not known a priori.

**Table 3**
Descriptive statistics based on 2297 observations.

| Variable | Description | Mean | SD |
|---|---|---|---|
| Dependent variables | | | |
| *TMILE* | Mileage per year driven with trucks (1000 miles) | 7.14 | 10.97 |
| *CMILE* | Mileage per year driven with cars (1000 miles) | 8.90 | 10.00 |
| *TNUM* | Number of trucks owned by the household | 0.72 | 0.79 |
| *CNUM* | Number of cars owned by the household | 1.10 | 0.82 |
| Exogenous covariates | | | |
| *DENSITY* | Houses per square mile | 2564.99 | 1886.09 |
| *BIKES* | Number of bicycles | 0.97 | 1.23 |
| *HHSIZE* | Number of individuals in a household | 2.69 | 1.44 |
| *ADLTS* | Number of adults in a household | 1.99 | 0.79 |
| *URB* | Household is in an urban area | 0.93 | 0.25 |
| *INC1* | Household income is between 20K and 30K | 0.11 | 0.31 |
| *INC2* | Household income is between 30K and 50K | 0.21 | 0.41 |
| *INC3* | Household income is between 50K and 75K | 0.19 | 0.39 |
| *INC4* | Household income is between 75K and 100K | 0.13 | 0.33 |
| *INC5* | Household income is greater than 100K | 0.22 | 0.41 |
| *HOME* | Household owns the home | 0.69 | 0.46 |
| *HS* | Highest household education is a high school degree | 0.31 | 0.46 |
| *BS* | Highest household education is at least a bachelor's degree | 0.46 | 0.50 |
| *CHILD1* | Youngest child is under 6 years old | 0.17 | 0.37 |
| *CHILD2* | Youngest child is between 6 and 15 years old | 0.18 | 0.38 |
| *CHILD3* | Youngest child is between 15 and 21 years old | 0.06 | 0.23 |
| *LA* | Household lives in Los Angeles MSA | 0.42 | 0.49 |
| *SAC* | Household lives in Sacramento MSA | 0.08 | 0.27 |
| *SD* | Household lives in San Diego MSA | 0.09 | 0.28 |
| *SF* | Household lives in San Francisco MSA | 0.23 | 0.42 |

**Table 4**
Model estimates. Posterior means and standard deviations of the coefficients are reported.

| Variable | *TMILE* | | *CMILE* | | *TNUM* | | *CNUM* | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| log(*DENSITY*) | −0.41 | (0.32) | −0.25 | (0.23) | −0.07 | (0.02) | −0.02 | (0.02) |
| *BIKES* | −0.16 | (0.28) | 0.03 | (0.20) | 0.08 | (0.02) | −0.01 | (0.01) |
| *HHSIZE* | 0.45 | (0.52) | 0.73 | (0.42) | 0.05 | (0.03) | −0.06 | (0.03) |
| *ADLTS* | −0.63 | (0.68) | 0.28 | (0.53) | 0.09 | (0.04) | 0.17 | (0.03) |
| *URB* | 0.43 | (1.48) | −0.69 | (1.22) | −0.14 | (0.08) | 0.19 | (0.08) |
| *INC1* | 2.53 | (1.67) | −1.35 | (1.02) | 0.18 | (0.08) | 0.09 | (0.06) |
| *INC2* | 1.28 | (1.46) | 1.15 | (0.88) | 0.41 | (0.07) | 0.11 | (0.05) |
| *INC3* | 2.56 | (1.49) | 1.65 | (0.91) | 0.49 | (0.07) | 0.26 | (0.06) |
| *INC4* | 2.60 | (1.60) | 0.74 | (1.01) | 0.59 | (0.08) | 0.24 | (0.07) |
| *INC5* | 3.63 | (1.58) | 1.86 | (0.97) | 0.61 | (0.08) | 0.31 | (0.06) |
| *HOME* | −0.61 | (0.90) | −1.26 | (0.56) | 0.21 | (0.04) | 0.10 | (0.04) |
| *HS* | −0.41 | (0.98) | 1.28 | (0.70) | 0.02 | (0.05) | 0.11 | (0.04) |
| *BS* | −2.04 | (1.03) | 0.85 | (0.71) | −0.20 | (0.05) | 0.17 | (0.05) |
| *CHILD1* | 1.71 | (1.45) | 0.56 | (1.07) | 0.12 | (0.08) | 0.12 | (0.07) |
| *CHILD2* | 1.24 | (1.31) | 0.61 | (0.98) | 0.08 | (0.07) | 0.06 | (0.06) |
| *CHILD3* | 1.32 | (1.51) | 0.01 | (1.07) | 0.04 | (0.08) | −0.02 | (0.07) |
| *LA* | 2.71 | (0.99) | 1.51 | (0.74) | −0.14 | (0.05) | 0.03 | (0.05) |
| *SAC* | 2.09 | (1.40) | 1.74 | (1.03) | −0.15 | (0.08) | 0.07 | (0.07) |
| *SD* | 1.26 | (1.42) | 0.07 | (1.02) | −0.18 | (0.08) | 0.10 | (0.07) |
| *SF* | 1.58 | (1.17) | −0.06 | (0.81) | −0.27 | (0.06) | 0.15 | (0.05) |

For the truck and car mileage equations, the posterior means for the coefficients of log(*DENSITY*) are −0.41 and −0.25 with posterior standard deviations of 0.32 and 0.23, respectively. The signs suggest that households located in denser neighborhoods, all else equal, are associated with lower truck and car usage on average. For example, the marginal effects from Table 5 show that a 50% increase in residential density is associated with a 168.18 and 98 decrease in annual mileage driven with trucks and cars, respectively. These estimates are small despite increasing residential density by as much as 50%. The results suggest that residential density has a small economic impact on vehicle usage. Also, the differences in magnitudes suggest that less fuel-efficient vehicles are more sensitive to residential density changes than fuel-efficient vehicles on average. The results are consistent with the intuition that households would want to drive less as overall vehicle operating costs increased, which is particularly true for less efficient vehicles. However, the posterior standard deviations are close in magnitude to the coefficient estimates, which suggest some uncertainty in the relationship between residential density and vehicle usage for trucks and cars. This finding is somewhat contrary to the conclusions in Brownstone et al. (2009) and

**Table 5**
Marginal effects of increasing *DENSITY* by 50%. The changes in probabilities are in $10^{-3}$ units, and the changes in truck and car mileage are in annual miles.

| $\Delta Pr(TNUM = 0)$ | $\Delta Pr(TNUM = 1)$ | $\Delta Pr(TNUM \geq 2)$ |
|---|---|---|
| 13.18<br>(3.35) | −6.37<br>(1.67) | −6.81<br>(1.72) |
| $\Delta Pr(CNUM = 0)$<br>2.88<br>(2.89) | $\Delta Pr(CNUM = 1)$<br>−0.23<br>(0.26) | $\Delta Pr(CNUM \geq 2)$<br>−2.64<br>(2.65) |
| $\Delta TMILE$<br>−168.14<br>(130.70) | $\Delta CMILE$<br>−98.00<br>(93.85) | |

Fang (2008), where the vehicle usage variables are modeled as censored (Tobit-type) outcomes instead of potential outcomes. The authors find that residential density does affect truck utilization in a significant way but not for car utilization. This difference arises due to the different modeling strategies.

Marginal effects are presented in Table 5 since the coefficients in the ordered equations are difficult to interpret. The estimates suggest that when residential density increases by 50%, the probability of not holding any trucks increases by 1.318%, while the probabilities of holding one and two or more trucks decrease by 0.637% and 0.681%, respectively. The effects on car holdings is practically on the same order of magnitude, but there is sizable uncertainty in the estimates as the posterior standard deviations are large. These estimates are similar to the findings in Fang (2008) and approximately half the size of the estimates in Brownstone et al. (2009).

## 6. Concluding remarks

This paper develops an efficient algorithm to estimate sample selection models with two incidentally truncated outcomes and two separate selection variables. While such models are easily described mathematically, estimation is often difficult due to the intricate pattern of missing data that arises with two incidentally truncated outcomes and the discrete nature of the selection variables. These features result in evaluations of intractable likelihoods, identification issues, and computationally-inefficient algorithms. This paper extends the Markov chain Monte Carlo algorithm with minimal data augmentation, first proposed in Chib (2007) and Chib et al. (2009), to efficiently simulate the joint posterior distribution of interest. A central aspect of the proposed algorithm is that it includes only a small subset of the total missing data in the MCMC sampler, resulting in improved sampling efficiency and decreased computational load, as demonstrated in the simulation study. Also, despite not having the "complete" data, the resulting sampling distributions are well-known and easy to sample from.

The model is applied to estimate the effects of residential density on vehicle usage and holdings in the state of California. Results suggest that large increases in residential density are not strongly associated with changes in either vehicle utilization or the probability of holding cars, but are strongly related to changes in truck holdings. This finding associated with vehicle utilization, especially for truck usage, is contrary to the literature and demonstrates that the sample selection framework can reveal new conclusions in the data.

## Acknowledgements

The author thanks the Editor, Associate Editor, and two anonymous referees for helpful comments. He is especially indebted to Ivan Jeliazkov and David Brownstone for their guidance and insightful suggestions, Audrey Fang for providing the data, and numerous colleagues for proofreading earlier drafts. This research is generously supported by a research grant through David Brownstone and the University of California Transportation Center.

## References

Albert, J.H., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association 88 (422), 669–679.
Albert, J.H., Chib, S., 2001. Sequential ordinal modeling with applications to survival data. Biometrics 57, 829–836.
Amemiya, T., 1984. Tobit models: a survey. Journal of Econometrics 24 (1–2), 3–61.
Brownstone, D., Fang, A., May 2009. A vehicle ownership and utilization choice model with endogenous residential density. Working Papers. University of California, Irvine.
Brownstone, D., Golob, T.F., 2009. The impact of residential density on vehicle usage and energy consumption. Journal of Urban Economics 65 (1), 91–98.
Cervero, R., Kockelman, K., 1997. Travel demand and the 3ds: density, diversity, and design. Transportation Research Part D: Transport and Environment 2 (3), 199–219.
Chen, M., Dey, D., 2000. Bayesian analysis for correlated ordinal data models. In: Generalized Linear Models: A Bayesian Perspective. CRC Press.
Chib, S., 2001. Markov chain Monte Carlo methods: computation and inference. In: Heckman, J., Leamer, E. (Eds.), Handbook of Econometrics. In: Handbook of Econometrics, vol. 5. Elsevier, pp. 3569–3649 (Ch. 57).
Chib, S., 2007. Analysis of treatment response data without the joint distribution of potential outcomes. Journal of Econometrics 140 (2), 401–412.
Chib, S., Greenberg, E., 1995. Understanding the Metropolis–Hastings algorithm. The American Statistician 49 (4), 327–335.

Chib, S., Greenberg, E., Jeliazkov, I., 2009. Estimation of semiparametric models in the presence of endogeneity and sample selection. Journal of Computational and Graphical Statistics 18 (2), 321–348.
Dunphy, R., Fisher, K., 1996. Transportation, congestion, and density: new insights. Transportation Research Record: Journal of the Transportation Research Board 1552, 89–96.
Fang, H.A., 2008. A discrete-continuous model of households' vehicle choice and usage, with an application to the effects of residential density. Transportation Research Part B: Methodological 42 (9), 736–758.
Greenberg, E., 2007. Introduction to Bayesian Econometrics. Cambridge University Press.
Gronau, R., 1973. The effect of children on the housewife's value of time. The Journal of Political Economy 81 (2), S168–S199.
Heckman, J.J., 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In: Annals of Economic and Social Measurement, Volume 5, Issue 4, NBER Chapters. National Bureau of Economic Research, Inc., pp. 120–137.
Heckman, J.J., 1979. Sample selection bias as a specification error. Econometrica 47 (1), 153–161.
Heckman, J.J., 1990. Varieties of selection bias. The American Economic Review 80 (2), 313–318.
Jeliazkov, I., Graves, J., Kutzbach, M., 2008. Fitting and comparison of models for multivariate ordinal outcomes. In: Advances in Econometrics: Bayesian Econometrics.
Koop, G., Poirier, D., Tobias, J., 2007. Bayesian Econometric Methods. Cambridge University Press.
Liu, J.S., 1994. The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. Journal of the American Statistical Association 89 (427), 958–966.
Manski, C.F., 1989. Anatomy of the selection problem. The Journal of Human Resources 24 (3), 343–360.
Nandram, B., Chen, M.-H., 1996. Reparameterizing the generalized linear model to accelerate Gibbs sampler convergence. Journal of Statistical Computation and Simulation 54 (1–3), 129–144.
Newey, W.K., Powell, J.L., Walker, J.R., 1990. Semiparametric estimation of selection models: some empirical results. The American Economic Review 80 (2), 324–328.
Poirier, D.J., 1980. Partial observability in bivariate probit models. Journal of Econometrics 12 (2), 209–217.
Puhani, P.A., 2000. The Heckman correction for sample selection and its critique. Journal of Economic Surveys 14 (1), 53–68.
Shonkwiler, J.S., Yen, S.T., 1999. Two-step estimation of a censored system of equations. American Journal of Agricultural Economics 81 (4), 972–982.
Tanner, M.A., Wong, W.H., 1987. The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82 (398), 528–540.
Tierney, L., 1994. Markov chains for exploring posterior distributions. Annals of Statistics 22, 1701–1762.
van Hasselt, M., 2009. Bayesian inference in a sample selection model. Working papers. The University of Western Ontario.
Wooldridge, J., 2002. Econometric Analysis of Cross Section and Panel Data. MIT press.
Wooldridge, J.M., 1998. Selection corrections with a censored selection variable. Mimeo. Michigan State University Department of Economics.
Yen, S.T., 2005. A multivariate sample-selection model: estimating cigarette and alcohol demands with zero observations. American Journal of Agricultural Economics 87 (2), 453–466.