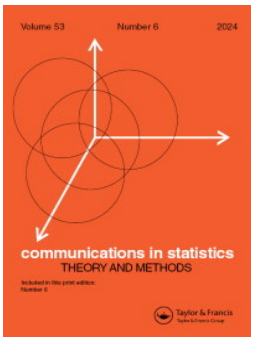

Communications in Statistics - Theory and Methods

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/lsta20

Bayesian inference in a sample selection model with multiple selection rules

Alireza Rezaee, Mojtaba Ganjali & Ehsan Bahrami Samani

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
2023, VOL. 00, NO. 0, 1–21
https://doi.org/10.1080/03610926.2023.2178260

Bayesian inference in a sample selection model with multiple selection rules

Alireza Rezaee, Mojtaba Ganjali, and Ehsan Bahrami Samani

Department of Statistics, Shahid Beheshti University, Tehran, Iran

ABSTRACT
Sample selection model is a solution to eliminate the nonresponse bias. In some applications nonresponse is a multilevel variable with respect to its reasons of occurring. In these cases, the sample selection model can be extended such that a model to be considered for each of the nonresponse reasons. Also, in many cases, the reasons for nonresponse have priority over each other. In other words, it is not possible to observe all of the nonresponse reasons simultaneously. For example, in a survey with two noncontact and refusal reasons, noncontact has priority over refusal and refusal can be observed if the contact to the respondent can be established. For analyzing such extended model, a Bayesian inference approach with multiple selection rules using multivariate normal, inverse gamma and LKJ distributions as prior distributions for parameters and possibility of priority for nonresponse reasons is presented. Simulation studies are performed and an establishment survey data set is analyzed to demonstrate the performance of the proposed method. For sensitivity analysis of nonresponse on the parameters of interest, posterior displacement is applied.

ARTICLE HISTORY
22 September 2021
4 February 2023

KEYWORDS
Data augmentation; establishment survey; Heckman correction; nonresponse mechanism; truncated multivariate normal distribution

SUBJECT CLASSIFICATION CODES:
62D10; 62J12; 62H12

1. Introduction

Data analyzers often face the problem of missing data. The results of their analyses will be distorted and erroneous if the appropriate methods for the treatment of nonresponse are not considered, especially in the case of nonignorable nonresponse mechanism (Little and Rubin 1987). In analysis of a model to investigate the effects of explanatory variables on the variable of interest, a suitable method, to eliminate the bias that may cause by nonresponse, is the use of sample selection model. A sample selection model typically consists of two components (Hasselt 2011). The first is to determine the relationship between dependent variable of interest and some independent variables, and the second is to describe the selection mechanism which determines whether the outcome is missed randomly or not randomly.

Heckman (1976) proposed the univariate sample selection model to correct selection bias in estimating of model parameters. Afterward, many articles were devoted to the use of the Heckman method, both in the development of the model and in facing applied issues. Some articles are also devoted to increase the number of outcome variables, i.e., considering multivariate responses, such as Hanoch (1976), Li and Rahman (2011) and Jolani (2014). Moreover, Bayesian inference for sample selection model was worked by some of researchers. Chib et al. (2008) proposed the Bayesian method for analysis of a regression model with three

CONTACT Mojtaba Ganjali m-ganjali@sbu.ac.ir Department of Statistics, Shahid Beheshti University, Tehran, Iran.
This article has been corrected with minor changes. These changes do not impact the academic content of the article.

response variables and a selection model. Maksym and Obrizan (2011) provided estimates for the parameters of the outcome and the selection models using a Bayesian method where the outcome is discrete. Hasselt (2011) proposed an algorithm for Bayesian analysis of univariate selection model by assuming a bivariate normal distribution for the errors of the selection and outcome models and then extended it in the case of non-normally distributed errors. Ding (2014) assumed t distributions for the errors of the selection model and the outcome model and proposed a method for Bayesian inference of parameters.

In many of the selection models, the nonresponse variable is considered as a binary variable. However, nonresponse may occur for several reasons and therefore we should consider different nonresponse mechanisms. For more than two reasons (levels) of nonresponse, Rezaee, Ganjali, and Bahrami Samani (2022) showed that merging the levels of nonresponse into one level will result in biased estimates when the mechanisms of nonresponse are different at the different levels of nonresponse or when the correlations of different mechanisms of nonresponse with the variable of interest are different.

There are few articles examining the sample selection model with multiple selection rules. Catsiapis and Robinson (1978, 1982) developed Heckman model by two and then multi sample selection models with independent assumptions between sample selection models. Kim and Kim (2016) presented a method to model multivariate responses with multiple selection rules. They assumed elliptically contoured (EC) distribution for the errors in the models to obtain robustness against departures from normality. Rezaee, Ganjali, and Bahrami Samani (2022) presented an approach, for survey data analysis with multilevel nonresponse, by extending the number of selection models to be the same as that of the reasons of the nonresponse. They assumed a multivariate normal distribution for the error components of the models.

Bayesian inference generally has many attractive features of which some of them can be used in the case of sample selection model with multiple selection rules. For example, if there are some information about parameters, they can be incorporated into the analysis through prior distributions for parameters. Example of extracting information can be the use information that is available based on analysis of previous survey. The reason for using Bayesian inference is that, when the number of selection rules is increased, the number of samples will be decreased in each of them and this will increase the standard errors of the estimates.

In this paper, we will present a Bayesian inference in the case of sample selection model with multiple selection rules by developing the Hasselt (2011) method. We ordered the nonresponse reasons by using the natural priority on them. For example, in a survey with two reasons for nonresponse, e.g., noncontact and refusal, noncontact reason has priority over refusal reason. This means that, refusal will be recorded when contact to the respondent can be done. Sequential or nested modeling approach can potentially be a solution for such cases, but by this approach, the correlation between the reasons of the nonresponse can't be taken into account. See Steele and Durrant (2011). We will examine the performance of our model using the presented Bayesian approach by a simulation study and its implementation on an establishment survey data. Also, posterior displacement will be used for sensitivity analysis of the effect of selection models on parameters of interest. The novelties of this article are the development of the number of selection models according to the number of nonresponse reasons, its Bayesian approach considering the order on the nonresponse reasons the use of posterior displacement for the sensitivity analysis. In addition, an establishment survey data has been used for real data analysis. There are a few articles in this field with respect to the analysis of household survey.

This paper is organized as following. In Section 2, sample selection model with multiple selection rules is presented and discussed. The use of the proposed method is illustrated in Section 3 with some simulated data. Section 4, contains an application of the method on analysis of an establishment survey with some sensitivity analysis. In Section 5, conclusion and discussion are given.

2. Sample selection model with multiple selection rules

To illustrate the sample selection model with multiple selection rules, suppose that the data set to be analyzed is $\{y_1, y_2, \ldots, y_n\}$, where some of y_is are not observed. The reasons for the nonresponse are known and divided into K categories. For each corresponding category of nonresponse, we consider a selection model and use the following joint model to show their relationship with the explanatory variables and the variable of interest.

$$y_i = x_i\beta + e_i, i = 1, 2, \ldots, n \tag{1}$$

$$y^*_{ij} = w_{ij}\alpha_j + u_{ij}, i = 1, 2, \ldots, n; j = 1, 2, \ldots, K \tag{2}$$

where $x_i = (x_{i1}, \ldots, x_{ip}), \beta = (\beta_1, \ldots, \beta_p)', w_{ij} = \left(w_{i1}, \ldots, w_{iq_j}\right)$ and

$\alpha_j = \left(\alpha_{j1}, \ldots, \alpha_{jq_j}\right)'$. M_{ij} is a nonresponse indicator defined as $M_{ij} = 0 \Leftrightarrow y^*_{ij} \geq 0$ and $M_{ij} = 1 \Leftrightarrow y^*_{ij} < 0$ for $j = 1, 2, \ldots, K$. Therefore, y_i is observed if $y^*_{i1} \geq 0, y^*_{i2} \geq 0, \ldots, y^*_{iK} \geq 0$ and is non response if $y^*_{ij} < 0$ for at least one $j = 1, 2, \ldots, K$.

In order to simplify the calculations, we rewrite Equations (1) and (2) as follows:

$$y_i = x_i\beta + e_i, i = 1, 2, \ldots, n \tag{3}$$

$$y^*_i = \mathbf{W}_i\alpha + u_i, \tag{4}$$

where

$$y^*_i = \begin{bmatrix} y^*_{i1} \\ y^*_{i2} \\ \vdots \\ y^*_{iK} \end{bmatrix}, u_i = \begin{bmatrix} \mathrm{u}_{i1} \\ \mathrm{u}_{i2} \\ \vdots \\ \mathrm{u}_{iK} \end{bmatrix}, \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_K \end{bmatrix}, \mathbf{W_i} = \begin{bmatrix} w_{i1} & O_2 & O_3 & \ldots & O_K \\ O_1 & w_{i2} & O_3 & \ldots & O_K \\ O_1 & O_2 & w_{i3} & \ldots & O_K \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ O_1 & O_2 & O_3 & \ldots & w_{iK} \end{bmatrix}. \tag{5}$$

In the equations (5), α is a $\sum_{j=1}^{K} q_j$ dimension column vector, O_j is a q_j dimension row vector with 0 values and $\mathbf{W_i}$ is a $K \times \left(\sum_{j=1}^{K} q_j\right)$ dimension matrix. We assume $\left(y_i, y^*_i\right)$ has a $(K + 1)$ variate normal distribution as follows:

$$\left(y_i, y^*_i\right) \mid (\beta, \alpha, \mathbf{\Sigma}) \sim N_{K+1}\left(\begin{bmatrix} x_i\beta \\ \mathbf{W_i}\alpha \end{bmatrix}, \mathbf{\Sigma}\right), i = 1, 2, \ldots, n$$

where $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{00} & \Sigma_{eu} \\ \Sigma_{ue} & \boldsymbol{\Sigma_{uu}} \end{bmatrix}$, $\Sigma_{eu} = [\sigma_{01}, \sigma_{02}, \ldots, \sigma_{0K}] = \Sigma'_{ue}$ and

$$\boldsymbol{\Sigma_{uu}} = \begin{bmatrix} 1 & \rho_{12} & \ldots & \rho_{1K} \\ \rho_{12} & 1 & \ldots & \rho_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1K} & \rho_{2K} & \ldots & 1 \end{bmatrix}.$$

Setting the diagonal elements of $\boldsymbol{\Sigma_{uu}}$ equal to one is the typical identifiability constraint for a sample selection model with multiple selection rules. But, it should be noted that in a Bayesian treatment of this model it is not necessary to impose this constraint (Hasselt 2011).

In order to eliminate the bias caused by the nonresponse samples, Rezaee, Ganjali, and Bahrami Samani (2022) considered Equation (3) as follows

$$y_i = x_i\beta + \sum_{j=1}^{K} \sigma_{0j}\lambda_{ij} + v_i, i = 1, \ldots, n \tag{6}$$

and presented a method for estimating the model parameters. In equation (6), λ_{ij} is the Inverse Mills Ratio, $\sum_{j=1}^{K} \sigma_{0j}\lambda_{ij}$ is the bias component related to nonresponse and v_i is the error. If the correlation of y_i and M_{ij} is zero for all of $i = 1, \ldots, n$, then λ_{ij} will be zero and there is no bias related to the j-th reason of nonresponse. Formulas for Bayesian parameter estimation are given in the next section and the Appendix.

In this paper, we assume that the reasons for the nonresponse are prioritized. Priority of observing nonresponse reasons is common in surveys. For example, if the nonresponse in a survey is due to noncontact or refusal then noncontact has priority over refusal, because, refusal will be observable if contact with respondent can be reached. We set $\mathrm{M_i} = 0$ if we have $\mathrm{M}_{ij} = 0$ for all $\mathrm{j} = 1, \ldots, \mathrm{K}$ and $\mathrm{M_i} = j$ if $\mathrm{M}_{ij} = 1$ and $\mathrm{M}_{it} = 0$ for all of $\mathrm{t} = 1, \ldots, \mathrm{j} - 1$. In this case the values of M_{it}s are unobservable for all of $\mathrm{t} = j + 1, \ldots, \mathrm{K}$. By prioritizing the nonresponse reasons, observing $\mathrm{M_i} = j$ means that nonresponse reasons were not related to 1 to $\mathrm{j} - 1$ first items of reasons and also, we cannot have any judgment for reasons $\mathrm{j} + 1$ to K. However, the model can be also developed for studies where it is possible to simultaneously observe all nonresponse reasons.

We define $S_j = \{i|\mathrm{M_i} = j\}, j = 0, 1, 2, \ldots K$ as the index sets of the observed and missing outcomes due to $j - th$ reason. The likelihood function is as follows:

$$\begin{aligned}
&L\left(\mathrm{y}_{obs}, M|\beta, \alpha, \boldsymbol{\Sigma}\right) \\
&= \prod_{i\in S_0} \phi_1\left(\frac{y_i - x_i\beta}{\sigma_0}\right) \int_{-w_{i1}\alpha_1}^{+\infty} \cdots \int_{-w_{iK}\alpha_K}^{+\infty} \phi_K\left(u_{i1}, u_{i2}, \ldots, u_{iK}|y_i, \boldsymbol{\Sigma_{uu,0}}\right) d_{u_{i1}} d_{u_{i2}} \ldots d_{u_{iK}} \\
&\quad \times \prod_{i\in S_1} \Phi_1\left(w_{i1}\alpha_1\right) \times \prod_{i\in S_2} \int_{-w_{i1}\alpha_1}^{+\infty} \int_{-\infty}^{-w_{i2}\alpha_2} \varphi_2\left(u_{i1}, u_{i2}, \boldsymbol{\Sigma}_{\boldsymbol{uu}}^{1:2}\right) d_{u_{i2}} d_{u_{i1}} \times \ldots \\
&\quad \times \prod_{i\in S_{K-1}} \int_{-w_{i1}\alpha_1}^{+\infty} \cdots \int_{-w_{i[K-2]}\alpha_{[K-2]}}^{+\infty} \int_{-\infty}^{-w_{i[K-1]}\alpha_{[K-1]}} \varphi_{K-1}\left(u_{i1}, u_{i2}, \ldots, u_{i[K-1]}, \boldsymbol{\Sigma}_{\boldsymbol{uu}}^{1:[K-1]}\right) \\
&\qquad d_{u_{i[K-1]}} \ldots d_{u_{i1}} \\
&\times \prod_{i\in S_K} \int_{-w_{i1}\alpha_1}^{+\infty} \cdots \int_{-w_{i[K-1]}\alpha_{[K-1]}}^{+\infty} \int_{-\infty}^{-w_{iK}\alpha_K} \varphi_K\left(u_{i1}, u_{i2}, \ldots, u_{iK}, \boldsymbol{\Sigma_{uu}}\right) d_{iK} \ldots d_{u_{i1}}
\end{aligned} \tag{7}$$

where the $i\left[K-j\right]$ is the notation for subscript i and $K-j$, $\varphi_1\left(\frac{y_i-x_i\beta}{\sigma_0}\right)$ is the density function of the univariate standard normal distribution evaluated at $\frac{y_i-x_i\beta}{\sigma_0}$, $\phi_K(u_{i1},u_{i2},\ldots,u_{iK}|y_i,\boldsymbol{\Sigma}_{\boldsymbol{uu},\mathbf{0}})$ is the conditional density function of a K variates normal distribution evaluated at the $(u_{i1},u_{i2},\ldots,u_{iK})$ given y_i, $\Sigma_{uu.0}=\Sigma_{uu}-\frac{1}{\sigma_{00}}\Sigma_{ue}\Sigma_{eu}$, $\Phi_1\left(w_{i1}\alpha_1\right)$ is the cumulative distribution function (cdf) of the standard normal distribution evaluated at $w_{i1}\alpha_1$, $\phi_j\left(u_{i1},\ldots,u_{ij},\boldsymbol{\Sigma}_{uu}^{1:j}\right)$ is the density function of a j variate normal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}_{uu}^{1:j}$ evaluated at $\left(u_{i1},\ldots,u_{ij}\right)$ and $\boldsymbol{\Sigma}_{uu}^{1:j}$ is the covariance matrix of $\left(u_{i1},\ldots,u_{ij}\right)$.

By specifying the prior distribution for the model parameters, the posterior distribution will be obtained. Similar to Hasselt (2011), for simplicity in calculations, we assume that α,β and $\boldsymbol{\Sigma}$ priorly are independent. In other words:

$$\pi\left(\beta,\alpha,\Sigma_{eu},\boldsymbol{\Sigma}_{\boldsymbol{uu}}\right)=\pi\left(\beta\right)\pi\left(\alpha\right)\pi\left(\Sigma_{eu}\right)\pi\left(\boldsymbol{\Sigma}\right),$$

with the likelihood function given in Equation (4) and the prior distributions, the posterior distribution may be obtained as follows.

$$\pi\left(\beta,\alpha,\Sigma_{eu},\boldsymbol{\Sigma}_{\boldsymbol{uu}}|\,\mathrm{y}_{obs},M\right)\propto\pi\left(\beta\right)\pi\left(\alpha\right)\pi\left(\Sigma_{eu}\right)\pi\left(\boldsymbol{\Sigma}\right)L\left(\mathrm{y}_{obs},M|\beta,\alpha,\Sigma_{eu},\boldsymbol{\Sigma}_{\boldsymbol{uu}}\right).$$

The posterior distribution has not explicit form, but it is possible to generate samples from it by Gibbs sampling. To do this, we partition unknown parameters $\boldsymbol{\theta}=(\beta,\alpha,\sigma_{00},\Sigma_{eu},\boldsymbol{\Sigma}_{\boldsymbol{uu}})$ to $\theta_1=\boldsymbol{\Sigma}_{\boldsymbol{uu}},\theta_2=\alpha,\theta_3=(\beta,\Sigma_{eu})$ and $\theta_4=\sigma_{00}$, then with generating samples from full conditional posterior distributions $\pi\left(\theta_1|y,y^*,\theta_2,\theta_3,\theta_4\right),\pi\left(\theta_2|y,y^*,\theta_1,\theta_3,\theta_4\right),\pi(\theta_3|y,y^*,\theta_2,\theta_2,\theta_4)$ and $\pi\left(\theta_4|y,y^*,\theta_1,\theta_2,\theta_3\right)$, we will obtain samples from posterior distribution $\pi\left(\boldsymbol{\theta}|y,y^*\right)$. The problem here is the existence of latent variables. In this case, the values of these variables are unknown. As a result, the values of the parameters of the four conditional posterior distributions will be unknown, and it will not be possible to produce a sample of these full conditional distributions.

Data augmentation is a tool to solve the problem of obtaining full conditional distributions of model parameters condition on the latent variables. Data augmentation was first proposed by Tanner and Wong (1987) and later was used in the analysis of models with latent variables by researchers such as Albert and Chib (1993), McCulloch and Rossi (1994), Munkin and Trivedi (2003) and Hasselt (2011). In data augmentation, the values of latent variables are generated based on the conditional distribution of latent variables given observations and model parameters, and these values are used to determine the conditional distributions of the desired parameters in Gibbs sampling. $\left(y_i^*|y_i\right)$ $i=1,\ldots,\mathrm{n}$ has a truncated normal distribution as follows, and we use this distribution to augment the data.

$$\left(y_i^*|y_i,\boldsymbol{\theta}\right)\sim\begin{cases}TN_{K,(0,+\infty)^K}\left(\mu_{i.0},\boldsymbol{\Sigma}_{\boldsymbol{uu}.0}\right) & i\in S_0\\ TN_{K,I_j}\left(\mathbf{W}_i\alpha,\boldsymbol{\Sigma}_{\boldsymbol{uu}}\right) & I_j=(0,+\infty)^{j-1}\left(-\infty,0\right)\left(-\infty,+\infty\right)^{K-j}\\ & i\in S_j\quad j=1,\ldots,K\end{cases}$$

where

$$\mu_{i.0}=\mathbf{W}_i a+\frac{y_i-x_i\beta}{\sigma_{00}}\Sigma_{ue},\boldsymbol{\Sigma}_{\boldsymbol{uu}.\mathbf{0}}=\boldsymbol{\Sigma}_{\boldsymbol{uu}}-\frac{1}{\sigma_{00}}\Sigma_{ue}\Sigma_{eu}.$$

There are different algorithms for generating the latent variables from the truncated multivariate normal distribution. In the rejection algorithm, samples are generated from the multivariate normal distribution and they are either accepted if they are inside the support region or otherwise rejected. This algorithm may be very inefficient when the support region is small (i.e., in higher dimensions). Geweke (1991) presented a method for generating samples from the truncated multivariate normal distribution using Gibbs sampling. The samples are generated from a series of conditional posteriors which are univariate truncated normal distributions. This method is more efficient than those of rejection sampling for the small support regions but it may suffer from poor mixing and convergence. An alternative algorithm can be in form of generating jointly latent variables and α in one block within the Gibbs sampler to avoid poor mixing properties of the Markov chain. Whereas w_i and x_i are related together, the correlation between (α, β) arises and then between $\left(y_i^*, \alpha\right)$ and this correlation is potential for poor mixing. For more details about the algorithm of the sample generation of the $\left(y_i^*, \alpha\right)$, see Rahman and Vossmeyer (2019). We use the rejection algorithm to generate sample from the truncated multivariate normal distribution and extract the results, however using Rahman and Vossmeyer (2019) algorithm can be applicable.

In order to select the appropriate prior distribution for Σ_{eu} instead of considering σ_{00} directly in the set of parameters, like Koop and Poirier (1997) and Hasselt (2011), we use the conditional variance value i.e.,

$$\xi^2 = Var\left(e_i|u_i\right) = \sigma_{00} - \Sigma_{eu}\boldsymbol{\Sigma}_{\boldsymbol{uu}}^{-1}\Sigma_{ue}.$$

In this way we can consider the prior distribution of Σ_{eu} to be dependent on ξ^2 and thus this enables the selection of prior distributions with different shapes.

2.1. Posterior distribution of Σ_{uu}

The $\boldsymbol{\Sigma}_{\boldsymbol{uu}}$ is a correlation matrix with diagonal elements of one and off-diagonal elements of the correlation coefficients. These restrictions make it impossible to use the inverse Wishart distribution as the prior distribution. A suitable prior distribution for the correlation matrix is the Lewandowski-Kurowicka-Joe (LKJ) introduced by Lewandowski et al. (2009). distribution. This distribution is obtained by converting partial correlations into correlation matrices. Wang, Wu, and Chu (2018) showed that the restricted Wishart distribution is equivalent to the LKJ distribution. The density function of this distribution is as follows:

$$f\left(\boldsymbol{\Sigma}_{\boldsymbol{uu}}\right) = 2^{\sum_{k=1}^{K-1}(2(\upsilon-1)+K-k)(K-k)} \prod_{k=1}^{K-1}\left[B\left(\upsilon + \frac{K-k-1}{2}, \upsilon + \frac{K-k-1}{2}\right)\right]^{K-k} \left|\boldsymbol{\Sigma}_{\boldsymbol{uu}}\right|^{\nu-1}$$

$$f\left(\boldsymbol{\Sigma}_{\boldsymbol{uu}}\right) \propto \left|\boldsymbol{\Sigma}_{\boldsymbol{uu}}\right|^{\nu-1}$$

where K is the rank of the correlation matrix and the parameter ν is the scale parameter. This parameter is used to determine the shape of the probability density function. It can be interpreted as the shape parameter of a symmetric beta distribution. If $\nu = 1$, the density function has an almost uniform shape on all correlation coefficient matrices. If $\nu > 1$, matrices with smaller correlation have larger density and for larger ν there are sharper peaks in the density around the identity matrix. If $0 < \nu < 1$, the density has a trough at the identity matrix. The expectation of this distribution is the identity matrix and its variance is $\frac{4\left(\eta+\frac{K}{2}-1\right)^2}{(2\eta+K-2)^2(2\eta+K-1)}$. This prior is not a conjugate and posterior distribution is as follows:

$$f\left(\boldsymbol{\Sigma}_{\boldsymbol{uu}}|y, y^*, \alpha, \beta, \sigma_{00}, \Sigma_{eu}\right)$$

$$\propto \prod_{i\in S_0} f\left(\mathrm{y}_{\mathrm{i}}^{*}|\left(y_i,\beta,\sigma_{00},\Sigma_{eu},\boldsymbol{\Sigma}_{\boldsymbol{uu}}\right)\right)\prod_{i\notin S_0} f\left(\mathrm{y}_{\mathrm{i}}^{*}|\left(\alpha,\boldsymbol{\Sigma}_{\boldsymbol{uu}}\right)\right)\pi\left(\boldsymbol{\Sigma}_{\boldsymbol{uu}}\right).$$

Once the above equation is obtained, the Metropolis-Hastings method can be used to generate samples from the above posterior. In addition, Bayesian analysis with this prior distribution can be performed with the STAN package in R software, (see Carpenter et al. 2017). In the next subsections we drive conditional posterior distribution of the parameters. Additional details are given in the Appendix.

2.2. Posterior distribution of α

We have $\mathrm{y}_{\mathrm{i}}^{*}|\left(y_i,\alpha,\beta,\sigma_{00},\Sigma_{eu},\boldsymbol{\Sigma}_{\boldsymbol{uu}}\right)\sim TN_{K,(0,+\infty)^K}\left(\mu_{i.0},\boldsymbol{\Sigma}_{uu.0}\right), i\in S_0, \mathrm{y}_{\mathrm{i}}^{*}|\left(\alpha,\Sigma_{\boldsymbol{uu}}\right)\sim N_K$ $(\mathbf{W}_i\alpha,\boldsymbol{\Sigma}_{\boldsymbol{uu}}) i\notin S_0$ and assuming $\alpha\sim N_{\sum_{k=1}^{K} q_k}(\mathrm{a},\mathbf{A})$ as prior, the posterior distribution of α given observations and other parameters is in the form of:

$$f\left(\alpha|y,y^{*},\beta,\sigma_{00},\Sigma_{eu},\boldsymbol{\Sigma}_{\boldsymbol{uu}}\right)$$
$$\propto \prod_{i\in S_0} f\left(\mathrm{y}_{\mathrm{i}}^{*}|\left(y_i,\alpha,\beta,\sigma_{00},\Sigma_{eu},\boldsymbol{\Sigma}_{\boldsymbol{uu}}\right)\right)\prod_{i\notin S_0} f\left(\mathrm{y}_{\mathrm{i}}^{*}|\left(\alpha,\boldsymbol{\Sigma}_{\boldsymbol{uu}}\right)\right)\pi\left(\alpha\right).$$

It can be showed that $\alpha|\left(y,y^{*},\beta,\sigma_{00},\Sigma_{eu},\boldsymbol{\Sigma}_{\boldsymbol{uu}}\right)\sim N_{\sum_{k=1}^{K} q_k}\left(\bar{\mathrm{a}}\bar{\mathbf{A}}\right)$, where

$$\bar{\mathbf{A}}=\mathbf{W}^{-1},$$

$\mathbf{W}$ with the following form has of dimension $\sum_{k=1}^{K} q_k\times\sum_{k=1}^{K} q_k$

$$\mathbf{W}=\sum_{i\in S_0}\mathbf{W}_i'\boldsymbol{\Sigma}_{\boldsymbol{uu}.0}^{-1}\mathbf{W}_{\mathbf{i}}+\sum_{i\notin S_0}\mathbf{W}_i'\boldsymbol{\Sigma}_{\boldsymbol{uu}}^{-1}\mathbf{W}_{\mathbf{i}}+\mathbf{A}^{-1}$$

and
$\bar{\mathrm{a}}=\bar{\mathbf{A}}\mathrm{z}$ where z is a vector of dimension $\sum_{k=1}^{K} q_k$ and

$$\mathrm{z}=\sum_{i\in S_0}\mathbf{W}_i'\boldsymbol{\Sigma}_{\boldsymbol{uu}.0}^{-1}\left[\mathrm{y}_{\mathrm{i}}^{*}-\frac{y_i-x_i\beta}{\sigma_{00}}\Sigma_{ue}\right]+\sum_{i\notin S_0}\mathbf{W}_i'\boldsymbol{\Sigma}_{\boldsymbol{uu}}^{-1}\mathrm{y}_{\mathrm{i}}^{*}+\mathbf{A}^{-1}\mathrm{a}.$$

2.3. Posterior distribution of β,Σ_{eu}

We set $\mathrm{g}=\begin{bmatrix}\beta\\ \Sigma_{\mathrm{ue}}\end{bmatrix}$ and assume that $\beta\sim N_p(\mathrm{b},\mathbf{B})$ and $\Sigma_{ue}|\xi^2\sim N_p(0,\mathbf{R})$ where $\mathbf{R}=\begin{bmatrix} r_1^2 & & 0\\ & \ddots & \vdots\\ 0 & & r_K\xi^2\end{bmatrix}$, then we have $\mathrm{g}|\xi^2\sim N_{p+K}\left(\mathrm{g}_0,\mathbf{G}_0\right)$ where $\mathrm{g}_0=\begin{bmatrix} b\\ 0\end{bmatrix}$ and $\mathbf{G}_0=\begin{bmatrix}\mathbf{B}_{p\times p} & 0_{p\times K}\\ 0_{K\times p} & \mathrm{R}_{K\times K}\end{bmatrix}$.
The posterior distribution is in the form of:

$$f\left(\mathrm{g}|\,\sigma_{00},\boldsymbol{\Sigma}_{\boldsymbol{uu}},\alpha,y,y^{*}\right)\propto\prod_{i\in S_0} f\left(y_i|y_i^{*},\alpha,\beta,\sigma_{00},\boldsymbol{\Sigma}_{\boldsymbol{uu}}\right)\times\pi\left(\mathrm{g}|\,\sigma_{00}\right),$$

it can be shown that $\mathrm{g} \big| \left(\sigma_{00}, \boldsymbol{\Sigma}_{\boldsymbol{uu}}, \alpha, y, y^{*}\right) \sim N_{p+K}\left(\bar{\mathrm{g}}, \bar{\mathbf{G}}\right)$ where $\bar{\mathbf{G}} = (\mathbf{G}_{\mathbf{0}}^{-\mathbf{1}} + \xi^{-2}\mathbf{D}'\mathbf{D})^{-\mathbf{1}}$,

$$\bar{\mathrm{g}} = \bar{\mathbf{G}}(\mathbf{G}_{\mathbf{0}}^{-\mathbf{1}}\mathrm{g}_0 + \xi^{-2}\mathbf{D}'y_0), \mathbf{D} = \begin{bmatrix} x_1 & \left(y_1^{*} - \mathbf{W}_1' a\right)' \Sigma_{\boldsymbol{uu}}^{-1} \\ \vdots & \vdots \\ x_{n_0} & \left(y_{n_0}^{*} - \mathbf{W}_{n_0}' a\right)' \Sigma_{\boldsymbol{uu}}^{-1} \end{bmatrix}, y_0 = \begin{bmatrix} y_1 \\ \vdots \\ y_{n_0} \end{bmatrix},$$

$$\mathrm{y}_{\mathrm{i}}^{*} = \begin{bmatrix} y_{i1}^{*} \\ y_{i2}^{*} \\ \vdots \\ y_{iK}^{*} \end{bmatrix},$$

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_K \end{bmatrix} \text{ and } \mathbf{W_i} = \begin{bmatrix} w_{i1} & O_2 & O_3 & \dots & O_K \\ O_1 & w_{i2} & O_3 & \dots & O_K \\ O_1 & O_2 & w_{i3} & \dots & O_K \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ O_1 & O_2 & O_3 & \dots & w_{iK} \end{bmatrix}$$

2.4. Posterior distribution of ξ^2

We assume that $\xi^2 \sim IG\left(c_0, d_0\right)$ and we have

$$f\left(\xi^2 \big| y, y^{*}, \alpha, \beta, \Sigma_{eu}, \boldsymbol{\Sigma}_{\boldsymbol{uu}}\right) \propto \prod_{i \in S_0} f\left(y_i | \xi^2, y_i^{*}, \alpha, \beta, \Sigma_{eu}, \boldsymbol{\Sigma}_{\boldsymbol{uu}}\right) f\left(\Sigma_{eu} | \, \xi^2\right) \pi\left(\xi^2\right).$$

Then it can be shown that $\left(\xi^2 \big| y, y^{*}, \alpha, \beta, \Sigma_{eu}, \boldsymbol{\Sigma}_{\boldsymbol{uu}}\right) \sim IG(\bar{c}, \bar{d})$ where $\bar{c} = \frac{n_0+K}{2} + C_0$, $\bar{d} = d_0 + \frac{1}{2}\Sigma_{i\epsilon S_0}(y_i - \gamma_i)^2 + \frac{1}{2}\Sigma_{k=1}^{K}\frac{\sigma_{0k}^2}{r_k}$ and $\gamma_i = x_i\beta + \Sigma_{eu}\boldsymbol{\Sigma}_{\boldsymbol{uu}}^{-\mathbf{1}}\left(y_i^{*} - \mathbf{W_i}a\right)$.

2.5. An algorithm for sample generation

For given starting value of $\boldsymbol{\theta}$:

1. For $i \in S_0$ sample y_i^{*} from $TN_{K,(0,+\infty)^K}\left(\mu_{i.0}, \boldsymbol{\Sigma}_{\boldsymbol{uu}.\mathbf{0}}\right)$ and for $i \in S_k$ sample y_i^{*} from $TN_{K,I_j}\left(\mathbf{W_i}a, \boldsymbol{\Sigma}_{\mathbf{uu}.}\right), I_j = (0, +\infty)^{j-1}(-\infty, 0)(-\infty, +\infty)^{K-j}, j = 1, \dots, K$
2. Sample α from $\alpha | \left(y, y^{*}, \beta, \xi^2, \Sigma_{eu}, \boldsymbol{\Sigma}_{\boldsymbol{uu}}\right) \sim N_{\sum_{k=1}^{K} q_k}\left(\bar{\mathrm{a}}, \bar{\mathbf{A}}\right)$
3. Sample β and Σ_{eu} from $(\beta, \Sigma_{eu}) | \left(y, y^{*}, \alpha, \xi^2\right) \sim N_{p+K}\left(\bar{\mathrm{g}}, \bar{\mathbf{G}}\right)$
4. Sample Σ_{eu} from posterior distribution (2) by Metropolis-Hastings algorithm,
5. Sample ξ^2 from $\xi^2 \big| \left(y, y^{*}, \beta, \alpha, \Sigma_{eu}, \boldsymbol{\Sigma}_{\mathbf{uu}}\right) \sim IG\left(\bar{c}, \bar{d}\right)$,
6. Return step 1 and repeat.

2.6. Sensitivity analysis

The difference in the Bayesian parameter estimation method with that of the ordinary least square method is the existence of the correlation of each of the non-response reasons with the variable of interest. In the multivariate sample selection model, if the value of one of the correlation coefficients is zero, the nonresponse mechanism will be random at that nonresponse level and the corresponding Inverse Mills ratio factor will be removed from the outcome model. The likelihood displacement is a useful method to investigate the influence of correlation coefficients on the estimation of the parameters of the original model. The method of local influence was introduced by Cook (1986) and developed by others as a general tool for assessing the influence of local departures from the assumptions underlying the models. These assumptions for our research, since we desire to study the departure of random nonresponse to nonrandom nonresponse, are about the elements of Σ, i.e., deviation from $\rho_{01} = 0$, or $\rho_{02} = 0$ or $\rho_{01} = \cdots = \rho_{0K} = 0$ is of our interest. Kass, Tierne, and Kadane (1989) investigated sensitivity analysis in Bayesian approach. They introduced a perturbation ρ, so that $\tilde{l}_{NEW}(\theta)9\rho(\theta)$ where $\tilde{l}_{NEW}(\theta)$ is proportional to posterior density with perturbation and $\tilde{l}(\theta) = \log(\pi(\theta)f(y|\theta))$. Then they considered $E_{NEW}\left[g\left(\theta\right)\right] - E\left[g\left(\theta\right)\right]$ as a criterion for assessment of perturbation function $\rho\left(\theta\right)$ where $E_{NEW}\left[g\left(\theta\right)\right]$ is the posterior expectation of $g\left(\theta\right)$ with perturbation, $E\left[g\left(\theta\right)\right]$ is the posterior expectation with no perturbation and $g\left(\theta\right)$ is a real valued function of θ. They showed that the change in posterior mean is equal to

$$E_{NEW}\left[g\left(\theta\right)\right] - E\left[g\left(\theta\right)\right] = \left(D_{\rho}\right)' \tilde{\Lambda}(D_g) + o(n^{-2}),$$

where $\tilde{\Lambda} = [-D^2\tilde{l}(\tilde{\theta})]^{-1}$ and $\tilde{\theta}$ is the joint posterior mode and is equal to change in posterior mode to order $o\left(n^{-2}\right)$. Moreover, they considered maximum standardized change in posterior mean over function g as:

$$Max_g \frac{E_{NEW}\left[g\left(\theta\right)\right] - E\left[g\left(\theta\right)\right]}{\left(Var\left[g\left(\theta\right)\right]\right)^{1/2}} = MSC \approx \left[\left(D_{\rho}\right)' \tilde{\Lambda}(D_{\rho})\right]^{1/2}.$$

They defined posterior displacement in form of: $PD\left(\rho\right) = 2[\tilde{l}(\tilde{\theta}|y) - \tilde{l}(\tilde{\theta}_{NEW}|Y)$ and showed that:

$$MSC \approx PD\left(\rho\right)^{1/2}. \tag{8}$$

We want to assess sensitivity to j-th nonresponse reason or a set of nonresponse reasons (J), so, we fix the value of ρ_{0j} or $\left\{\rho_{0j}, j \in J\right\}$ at 0 or in an interval near 0, we consider $\rho\left(\theta\right) = log\frac{f(y|\theta^{-S})\pi(\theta^{-S})}{f(y|\theta)\pi(\theta)}$, where θ^{-S} is the θ which is fixed for the elements of $\rho_{0j} \in \left\{\rho_{0j}, j \in J\right\}$ and then apply the equation (8). Here the parameters which should be estimated are the elements of $\boldsymbol{\theta}$ except the $\left\{\rho_{0j}, j \in J\right\}$, so the functions g is clear. If the graph of these values versus ρ_{0j} is strongly curved at zero, it means that sample selection is nonrandom at j-th nonresponse reason or a set of nonresponse reasons in J and the parameters are estimated with high precision using selection model. For more details, see Kass, Tierne, and Kadane (1989), Cook (1986), Ganjali and Rezaei (2005) and Razie et al. (2013).

3. Simulation studies

This section evaluates the performance of the algorithm in a simulation study, where the data are generated from a sample selection model with the outcome model and two selection rules as follows:

$$\begin{aligned} y_i &= \beta_1 + \beta_2 x_i + e_i, i = 1, 2, \ldots, n \\ y_{ij}^* &= \alpha_{j1} + \alpha_{j2} w_{ij} + u_{ij}, i = 1, 2, \ldots, n; j = 1, 2. \end{aligned}$$

In order to cover different nonresponse mechanism at the levels of nonresponse, we will consider three cases as follows:

- Case 1: The mechanism of nonresponse at one level is random and at another level is nonrandom with parameters:

$$\alpha_{11} = 4.5, \alpha_{12} = -0.6, \alpha_{21} = 1, \alpha_{22} = 0, \sigma_{01} = -0.4, \sigma_{02} = 0, \rho_{12} = 0.$$

- Case2: Missing not at random (MNAR) in the same direction with the variable of interest at both levels of nonresponse with parameters:

$$\alpha_{11} = 2, \alpha_{12} = -0.2, \alpha_{21} = 5, \alpha_{22} = -0.7, \sigma_{01} = -0.4, \sigma_{02} = -0.6, \rho_{12} = 0.7.$$

- Case3: MNAR at both levels of nonresponse and non-directional with the variable of interest with parameters:

$$\alpha_{11} = 4.5, \alpha_{12} = -0.5, \alpha_{21} = -3, \alpha_{22} = 1, \sigma_{01} = -0.5, \sigma_{02} = 0.6, \rho_{12} = -0.4.$$

We set $\beta_1 = -1, \beta_2 = 1.5, \sigma_{00} = 1, (\alpha_1, \alpha_2) \sim N_4\left(\mathbf{0}, 10\mathbf{I}_4\right), \beta \sim N_2\left(\mathbf{0}, 10\mathbf{I}_2\right)$ where $\mathbf{I}_k$ is the identity matrix of size k, $\Sigma_{eu} \sim N_2\left(\mathbf{0}, \xi^2\mathbf{I}_2\right)$, $\nu = 1, \xi^2 \sim IG(3, 6)$ and a sample size of $\mathrm{n} = 500$ is considered to generate data. The algorithm is run for 10,000 iterations and the first 2000 draws are discarded as a burn in period. The initial value for the σ_{00} is set to be 1 and for the other parameters is set to be 0. The same explanatory variable, extracted from a uniform distribution between 1 and 10, was chosen for the selection models and the outcome model. Moreover, the sample size 50 is used to investigate the efficiency of the method, when there are small sample sizes with suitable prior information. In this case, the mean of priors of $\alpha, \beta, \sigma_{01}$and σ_{02} are the considered values in the cases 1 to 3 and for covariance matrix of those priors is identity matrix. The average response rate is about 60% in case 1 and about 63% in cases 2 and 3. The average nonresponse rates at level 1 in cases 1 to 3 are about 29, 21 and 15 percents respectively, and at level 2 are around 11, 16 and 23 percents respectively. It is possible to run simulations for any number of selection models, but we considered two models since due to the number of different combinations of the nonresponse mechanism at the nonresponse levels, there will be many cases and reporting them is out of the aim of this paper. We consider the Gelman–Rubin statistic[1] with two chains of 10,000 iterations, a lag 1 to 3 auto correlation function and inefficiency factor[2] as convergence diagnostics for samples generation. All analyses were done in R. We applied some packages such as “mvtnorm” to do calculations. The corresponding source code is available on request.

Results for the sample size of 500 and the sample size of 50 are given in Tables 1 and 2 respectively. It can be seen from Table 1 that the values of the parameters used to generate the data all lie well within the 95% highest posterior density (HPD) intervals except that for the σ_{00} in the case 1. The Gelman–Rubin statistic suggests that the Markov chain is mixing well. In particular, the autocorrelation function decreases rapidly, leading to a low value of the inefficiency factor. The simulated draws of the correlation coefficient ρ_{12} display more autocorrelation. All the findings of Table 1 except that for the σ_{00} in the case 2 are also valid for Table 2. Table 2 shows that the proposed method can work well when there are small sample size and proper information in the form of priors.

[1]The Gelman-Rubin statistic is $\frac{\frac{L-1}{L}W + \frac{1}{L}B}{W}$, where W is the within chain variance, B is the between chain variance and L is the number of chains

[2]Inefficiency factor is $1 + 2\sum_{k=1}^{\infty} AC(k)$ where $AC(k)$ is the lag k autocorrelation function

Table 1. Posterior summary for simulated data with sample size $n = 500$.

Case	Parameter	True	Mean	Std. dev.	95% HPD lower bound	95% HPD upper bound	GR	AC(1)	AC(2)	AC(3)	IF
1	α_{11}	4.5	4.53	0.13	4.31	4.80	1.02	0.44	0.19	0.08	3.67
	α_{12}	−0.6	−0.61	0.02	−0.65	−0.56	1.00	0.45	0.21	0.11	4.15
	α_{21}	1	0.83	0.13	0.60	1.13	1.01	0.33	0.23	0.15	2.76
	α_{22}	0	0.03	0.03	−0.03	0.08	1.02	0.59	0.44	0.32	4.69
	β_1	−1	−1.22	0.14	−1.48	−0.92	1.02	0.18	0.04	0.06	1.67
	β_2	1.5	1.54	0.03	1.48	1.59	1.00	0.12	0.04	0.04	1.47
	ρ_{01}	−0.5	−0.46	0.05	−0.55	−0.36	1.00	0.01	−0.06	−0.03	0.87
	ρ_{02}	0	0.10	0.07	−0.02	0.24	1.00	0.16	0.13	0.06	1.95
	ρ_{12}	0	−0.05	0.07	−0.22	0.06	1.00	0.58	0.39	0.26	4.25
	σ_{00}	1	1.12	0.07	1.00	1.26	1.00	0.07	−0.06	−0.03	0.77
2	α_{11}	2	1.98	0.11	1.76	2.18	1.00	0.21	0.04	0.01	0.83
	α_{12}	−0.2	−0.20	0.02	−0.23	−0.16	1.00	0.25	0.07	0.03	1.05
	α_{21}	5	5.15	0.12	4.91	5.38	1.01	0.42	0.21	0.13	3.14
	α_{22}	−0.7	−0.74	0.02	−0.78	−0.69	1.02	0.56	0.38	0.26	5.89
	β_1	−1	−0.99	0.13	−1.26	−0.73	1.00	0.23	0.09	0.01	0.98
	β_2	1.5	1.48	0.03	1.42	1.54	1.00	0.17	0.11	0.04	1.44
	ρ_{01}	−0.5	−0.46	0.05	−0.56	−0.36	1.00	0.04	0.03	0.01	0.87
	ρ_{02}	−0.5	−0.50	0.04	−0.58	−0.42	1.00	0.01	0.02	0.03	0.97
	ρ_{12}	0.5	0.55	0.04	0.49	0.63	1.00	0.72	0.52	0.36	6.03
	σ_{00}	1	1.15	0.08	1.00	1.31	1.00	-0.01	0.00	0.01	0.65
3	α_{11}	4.5	4.70	0.15	4.37	4.97	1.00	0.64	0.43	0.29	4.29
	α_{12}	−0.5	−0.52	0.02	−0.56	−0.47	1.00	0.57	0.38	0.26	3.91
	α_{21}	−3	−2.96	0.14	−3.23	−2.68	1.00	0.54	0.30	0.15	2.37
	α_{22}	1	0.99	0.02	0.94	1.03	1.00	0.55	0.32	0.17	2.82
	β_1	−1	−1.12	0.19	−1.49	−0.73	1.00	0.17	0.12	0.09	2.09
	β_2	1.5	1.52	0.03	1.46	1.58	1.00	0.16	0.10	0.07	2.22
	ρ_{01}	−0.5	−0.47	0.05	−0.56	−0.37	1.00	0.04	0.05	0.00	1.16
	ρ_{02}	0.5	0.46	0.04	0.38	0.55	1.01	0.05	0.06	0.08	1.98
	ρ_{12}	−0.5	−0.57	0.03	−0.64	−0.50	1.00	0.71	0.52	0.39	5.12
	σ_{00}	1	1.14	0.08	0.99	1.30	1.00	-0.01	0.04	0.03	1.38

Note: Estimates are based on the sample selection model with two selection rules.
Case 1: Random nonresponse at one level and MNAR at another level of nonresponse,
Case 2: MNAR in the same direction with the variable of interest at both levels of nonresponse,
Case 3: MNAR at both levels of nonresponse and non-directional with the variable of interest.

4. Empirical application

The survey of manufacturing with 10 employees or more is one of the most important surveys in the statistical center of Iran. Its results are used for calculation of value added in the manufacturing sector in whole country and provinces. We use this survey to investigate the performance of the proposed method in an establishment survey because there are many researches about nonresponse in household survey, but studies on establishment surveys are a few. This survey covers all activities in manufacturing. We select industry of manufacture of rubber and plastics products. However, other industries may also be considered.

Noncontact and refusal are two reasons of nonresponse in this survey and so we consider two selection models. The value of all sales of goods and services for each of establishment, i.e., output, is our variable of interest and we want to model it with some of explanatory variables.

Table 3 shows the explanatory variables used in main model and selection models. A number of establishments are registered in "Securities and Exchange Organization (Bourse)". These establishments generally are large establishments and have high turnover. Many of the survey questions are filled based on their financial statement which is released in public and the remaining questions, in form of supplementary questionnaire with a small number of

Table 2. Posterior summary for simulated data with sample size $n = 50$.

Case	Parameter	True	Mean	Std. dev.	95% HPD lower bound	95% HPD upper bound	GR	AC(1)	AC(2)	AC(3)	IF
1	α_{11}	4.5	4.01	0.37	3.29	4.74	1.01	0.40	0.23	0.07	1.95
	α_{12}	−0.6	−0.56	0.06	−0.69	−0.45	1.02	0.45	0.25	0.08	2.85
	α_{21}	1	0.93	0.35	0.23	1.60	1.03	0.30	0.20	0.13	4.07
	α_{22}	0	−0.06	0.08	−0.21	0.09	1.01	0.66	0.53	0.45	9.71
	β_1	−1	−1.03	0.48	−1.94	−0.04	1.02	0.09	0.05	−0.01	2.35
	β_2	1.5	1.54	0.10	1.36	1.73	1.01	0.08	0.04	0.02	2.14
	ρ_{01}	−0.5	−0.39	0.15	−0.65	−0.09	1.00	−0.01	0.03	−0.05	0.81
	ρ_{02}	0	−0.29	0.15	−0.59	0.01	1.00	0.08	0.00	0.01	1.25
	ρ_{12}	0	0.28	0.20	−0.07	0.66	1.01	0.69	0.53	0.43	7.00
	σ_{00}	1	1.56	0.33	1.04	2.22	1.00	0.05	0.00	0.01	0.90
2	α_{11}	2.00	1.95	0.33	1.34	2.64	1.00	0.26	0.04	0.02	1.78
	α_{12}	−0.20	−0.21	0.06	−0.32	−0.10	1.00	0.26	0.03	0.01	1.63
	α_{21}	5.00	5.01	0.34	4.34	5.68	1.00	0.31	0.08	0.01	1.60
	α_{22}	−0.70	−0.69	0.06	−0.81	−0.56	1.01	0.37	0.13	0.05	2.54
	β_1	−1.00	−0.97	0.48	−1.89	−0.03	1.00	0.25	0.03	0.00	1.98
	β_2	1.50	1.46	0.10	1.26	1.63	1.00	0.17	0.01	−0.01	1.85
	ρ_{01}	−0.50	−0.62	0.09	−0.77	−0.45	1.00	0.05	0.03	−0.06	1.48
	ρ_{02}	−0.50	−0.64	0.09	−0.79	−0.45	1.00	0.03	0.04	0.00	2.02
	ρ_{12}	0.50	0.78	0.05	0.67	0.86	1.00	0.61	0.37	0.21	4.68
	σ_{00}	1.00	1.86	0.36	1.19	2.56	1.00	0.11	0.07	−0.01	2.40
3	α_{11}	4.50	4.10	0.52	3.12	5.08	1.03	0.64	0.42	0.25	5.62
	α_{12}	−0.50	−0.46	0.08	−0.61	−0.30	1.02	0.59	0.37	0.21	4.99
	α_{21}	−3.00	−3.27	0.42	−4.10	−2.46	1.00	0.48	0.23	0.11	2.74
	α_{22}	1.00	1.07	0.07	0.94	1.22	1.00	0.52	0.26	0.11	2.97
	β_1	−1.00	−1.45	0.63	−2.56	−0.19	1.00	0.09	0.02	0.01	0.93
	β_2	1.50	1.55	0.11	1.33	1.75	1.00	0.10	0.05	0.01	0.85
	ρ_{01}	−0.50	−0.44	0.15	−0.69	−0.13	1.00	0.11	0.07	0.04	1.16
	ρ_{02}	0.50	0.53	0.09	0.35	0.69	1.00	0.06	0.05	0.02	1.27
	ρ_{12}	−0.50	−0.62	0.09	−0.78	−0.44	1.01	0.50	0.28	0.19	3.63
	σ_{00}	1.00	1.39	0.25	0.96	1.88	1.00	0.07	0.04	0.05	0.93

Note: The same as Table 1.

questions, are filled by establishments. ISIC (International Standard Industrial Classification of All Economic Activities) is the international reference classification of productive activities and published by the United Nations Statistics Division. Organization of conducting survey is a local organization at province level which is responsible for field conducting survey. There are 31 Organization of conducting survey. We specify each of them by their number. The values of the explanatory variables are known for all samples before conducting the survey. We used separate Probit models for refusal and noncontact to determine effective and noneffective explanatory variables and exclude variables which were not significant. The number of samples is 1236 establishments of which 865 establishments are respondent, 55 establishments are noncontact and 316 establishments are refusal.

4.1. Estimation

Rezaee, Ganjali, and Bahrami Samani (2022) estimated the effects of explanatory variables on nonresponse and output variable using maximum likelihood method and obtained that the mechanism of nonresponse due to refusal is nonrandom, due to noncontact is random and refusal and noncontact are independent for this survey. They showed that the estimations will be biased using univariate selection model when the nonresponse mechanism is different

Table 3. Explanatory variables used in models.

Dependent variable	Explanatory variables	Values
Logarithm of output	Logarithm of the number of employees	Real number
	Status of registration in Bourse	2 categories: 1 for registered and 2 for not registered
	Economic activity code (ISIC)	2 categories: 1 for Manufacture of rubber tyres and tubes plastics products and 2 for Manufacture of other rubber products (Isic4_2219)
Noncontact	Organization of conducting survey	3 categories: Org. 1, Org. 7 and Org. 11
	Status of participation in previous survey	2 categories: 1 for refusal and 2 for noncontact
	Having ancillary unit	2 categories: 1 for having and 2 for not
	Location of establishment	2 categories: 1 for in industrial zone and 2 for out of industrial zone
Refusal	Organization of conducting survey	6 categories: Org. 3, Org. 10, Org. 12, Org. 20, Org. 23 and Org. 25
	Status of participation in previous survey	3 categories: 0 for respondent, 1 for refusal and 2 for noncontact
	Status of being in sample in previous survey	2 categories: 1 for in sample and 2 for not in sample
	Having ancillary unit	2 categories: 1 for having and 2 for not having

Table 4. Posterior summaries of parameters for empirical application (Parameters that are not significant at the 5% level are highlighted).

			95% HPD						
Parameter	**Mean**	Std. dev.	lower bound	upper bound	**Gelman. Rubin**	**AC(1)**	**AC(2)**	**AC(3)**	**IF**
Intercept1(α_{11})	1.58	0.12	1.34	1.81	1.00	0.79	0.62	0.49	8.73
Ostan1(α_{12})	−0.94	0.29	−1.48	−0.35	1.00	0.71	0.51	0.39	9.00
Ostan7(α_{13})	−0.92	0.20	−1.31	−0.54	1.00	0.62	0.41	0.29	7.32
Ostan11(α_{14})	−1.36	0.46	−2.27	−0.47	1.00	0.48	0.25	0.16	6.66
M962(α_{15})	−0.34	0.18	−0.70	0.01	1.01	0.73	0.55	0.42	8.06
CO2(α_{16})	0.40	0.17	0.08	0.73	1.02	0.86	0.75	0.65	15.72
IZ1(α_{17})	0.35	0.13	0.10	0.62	1.00	0.82	0.68	0.56	8.96
Intercept2(α_{21})	1.10	0.09	0.92	1.29	1.00	0.75	0.59	0.48	23.51
Ostan3(α_{22})	0.45	0.20	0.06	0.86	1.00	0.72	0.52	0.37	6.63
Ostan10(α_{23})	0.23	0.16	−0.09	0.54	1.01	0.69	0.48	0.34	6.97
Ostan12(α_{24})	0.70	0.27	0.19	1.23	1.00	0.75	0.57	0.43	6.60
Ostan20(α_{25})	−1.02	0.18	−1.37	−0.66	1.00	0.52	0.30	0.16	4.36
Ostan23(α_{26})	−0.65	0.11	−0.87	−0.43	1.00	0.54	0.31	0.19	4.51
Ostan25(α_{27})	−1.01	0.19	−1.41	−0.65	1.00	0.52	0.27	0.16	4.00
M962(α_{28})	−1.07	0.14	−1.34	−0.82	1.00	0.52	0.28	0.15	4.46
M966(α_{29})	−0.57	0.09	−0.75	−0.40	1.00	0.63	0.39	0.26	5.16
CO2(α_{210})	0.21	0.10	0.01	0.40	1.00	0.60	0.36	0.24	8.00
Intercept(β_1)	21.09	0.12	20.86	21.33	1.00	0.23	0.13	0.11	3.52
log.Worker(β_2)	1.14	0.03	1.08	1.20	1.00	0.10	0.02	0.02	1.30
Bourse1(β_3)	0.58	0.26	0.06	1.07	1.00	0.08	0.00	0.02	1.03
Isic4.2219(β_4)	−0.67	0.11	−0.87	−0.43	1.00	0.09	0.00	0.01	1.24
ρ_{01}	−0.03	0.20	−0.38	0.35	1.02	0.97	0.95	0.92	52.79
ρ_{02}	−0.31	0.08	−0.46	−0.14	1.01	0.81	0.67	0.56	10.99
ρ_{12}	0.09	0.38	−0.54	0.70	1.02	1.00	0.99	0.99	68.64
σ_{00}	1.11	0.05	1.01	1.22	1.00	0.51	0.40	0.33	5.14

among the levels of nonresponse. We derive the estimation of sample selection model with two selection rules for these data using our Bayesian approach. We use the $(\alpha_1, \alpha_2) \sim N_{17}(0, \mathbf{I}_{17})$, $\beta \sim N_4(\mathbf{b}, \mathbf{I}_3)$ where $\mathbf{I}_k$ is the identity matrix of size k, $\mathrm{b} = (20, 0, 0, 0)$, $\Sigma_{eu} \sim N_2(\mathrm{p}, \xi^2\mathbf{I}_2)$ where $\mathrm{p} = (0, 0)$ and $\xi^2 \sim IG(3, 6)$ to obtain Bayes estimates of the parameters. The algorithm is run for 10000 iterations. The first 2000 draws are discarded as a burn-in. We use the same convergence diagnostics used in the simulation section. The results consist of mean, standard deviation, 95% HPD lower and upper bounds credible sets, Gelman Rubin Index value, lag 1 to lag 3 autocorrelation and Inefficiency factor which are given in Table 4.

It can be seen from Table 4 that the Gelman-Rubin statistics for all parameters are approximately equal to 1 and suggests that the Markov chain is mixing well. The autocorrelation function decreases rapidly except for the parameters $\alpha_{11}, \alpha_{12}, \alpha_{24}$ and ρ_{02}. Parameters that are not significant at the 5% level are highlighted.

Figure 1 displays the posterior density of the outcome model parameters and the correlations.

Table 5 shows the estimation of parameters with different size of iterations. It is observed that the estimation values of the parameters, have slight changes except large changes the

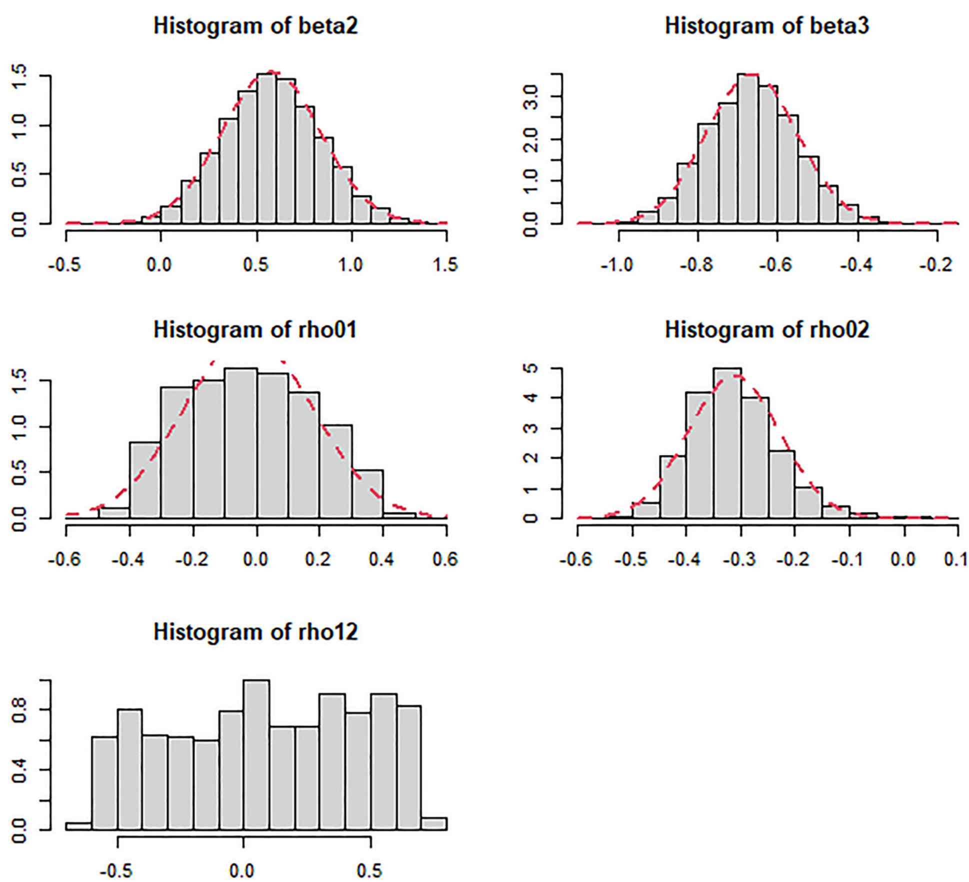

Figure 1. Posterior density of the outcome model parameters and the correlations.

Table 5. The mean of posterior distributions with different size of iterations for empirical application.

Parameter	n = 3000	n = 4000	n = 5000	n = 6000	n = 7000	n = 8000	n = 9000	n = 10000
Intercept1(α_{11})	1.5792	1.5775	1.5814	1.5831	1.5807	1.5793	1.5802	1.5788
Ostan1(α_{12})	−0.9018	−0.9134	−0.9161	−0.9355	−0.9302	−0.9292	−0.9322	−0.9380
Ostan7(α_{13})	−0.8891	−0.9156	−0.9176	−0.9120	−0.9161	−0.9149	−0.9183	−0.9166
Ostan11(α_{14})	−1.3610	−1.3855	−1.3784	−1.3541	−1.3564	−1.3545	−1.3664	−1.3588
M962(α_{15})	−0.3146	−0.3283	−0.3326	−0.3438	−0.3406	−0.3433	−0.3414	−0.3431
CO2(α_{16})	0.4310	0.4162	0.4037	0.3982	0.3950	0.3952	0.3973	0.3998
IZ1(α_{17})	0.3294	0.3414	0.3379	0.3461	0.3489	0.3513	0.3520	0.3535
Intercept2(α_{21})	1.1118	1.1132	1.1154	1.1017	1.1044	1.1013	1.1052	1.1038
Ostan3(α_{22})	0.4558	0.4570	0.4568	0.4579	0.4553	0.4539	0.4506	0.4499
Ostan10(α_{23})	0.2256	0.2335	0.2306	0.2284	0.2298	0.2257	0.2262	0.2265
Ostan12(α_{24})	0.7191	0.7051	0.7014	0.7123	0.7145	0.7148	0.7063	0.7037
Ostan20(α_{25})	−1.0166	−1.0151	−1.0245	−1.0194	−1.0184	−1.0191	−1.0196	−1.0177
Ostan23(α_{26})	−0.6412	−0.6442	−0.6467	−0.6467	−0.6448	−0.6450	−0.6456	−0.6475
Ostan25(α_{27})	−1.0068	−1.0055	−1.0069	−1.0022	−1.0048	−1.0060	−1.0089	−1.0117
M962(α_{28})	−1.0711	−1.0706	−1.0720	−1.0732	−1.0766	−1.0745	−1.0732	−1.0745
M966(α_{29})	−0.5726	−0.5719	−0.5737	−0.5710	−0.5732	−0.5716	−0.5718	−0.5715
CO2(α_{210})	0.2033	0.1996	0.2005	0.2079	0.2054	0.2049	0.2048	0.2055
Intercept(β_1)	21.0895	21.0877	21.0876	21.0936	21.0947	21.0957	21.0927	21.0939
log.Worker(β_2)	1.1405	1.1409	1.1406	1.1400	1.1398	1.1399	1.1401	1.1401
Bourse1(β_3)	0.5873	0.5859	0.5828	0.5799	0.5788	0.5778	0.5783	0.5769
Isic4.2219(β_4)	−0.6678	−0.6727	−0.6696	−0.6683	−0.6668	−0.6662	−0.6661	−0.6655
ρ_{01}	0.0359	0.0265	0.0228	−0.0214	−0.0234	−0.0273	−0.0168	−0.0257
ρ_{02}	−0.3182	−0.3119	−0.3104	−0.3129	−0.3132	−0.3154	−0.3122	−0.3121
ρ_{12}	−0.0112	−0.0193	−0.0180	0.0747	0.0798	0.0951	0.0642	0.0862
σ_{00}	1.1109	1.1102	1.1096	1.1098	1.1092	1.1096	1.1084	1.1079

parameters $\rho_{01}$and ρ_{12}. Their changes can be acceptable considering the HPD credible sets for these parameters which contain zero.

Table 6 shows the sensitivity of Bayesian estimates to the values of the variance of the prior distribution. For this purpose, the range of diagonal element of matrices A, B and the values of r_1 and r_2 are considered to be changed from 0.1 to 100. It is observed that for the parameter of variance of prior larger than 1, the Bayesian estimates tend to a constant. Moreover, non-significance of the ρ_{01} with respect to 95% HPD credible set, among the different values of variance of the prior distribution.

Table 7 shows the results of Bayesian estimation of parameters along with maximum likelihood estimates. It can be seen that for those parameters highlighted in the table, the signs of estimates are different in Bayesian and frequentist methods. This is due to small sample size for noncontact. In these cases, the coefficients have a bit large standard error in frequentist method and the Bayes estimator tends to prior mean which is zero and they are not significant. In the other cases, it can be seen that the estimators in two methods are nearly the same. Also, it can be seen that the correlation coefficient between output value and nonresponse due to refusal (ρ_{02}) is negative and is significant at the 5% level in both methods, while the correlation coefficient between output value and noncontact is not significant by both methods. Therefore, the mechanism of nonresponse is different among the levels of nonresponse and if sample selection model with one selection is used for analysis, the estimation will be biased. The standard errors of the Bayesian estimators are in most cases less than those of the Maximum likelihood method.

Table 6. The mean of posterior distributions with different values of variance parameters of priors for $\alpha, \beta, \sigma_{01}$ and σ_{02}.

Prior parameter* Parameter	0.1	1	10	100
Intercept1(α_{11})	1.4047	1.5788	1.5927	1.6139
Ostan1(α_{12})	−0.4812	−0.9380	−1.0811	−1.0599
Ostan7(α_{13})	−0.6093	−0.9166	−0.9552	−0.9526
Ostan11(α_{14})	−0.4617	−1.3588	−1.6149	−1.6601
M962(α_{15})	−0.1957	−0.3431	−0.3629	−0.3752
CO2(α_{16})	0.4053	0.3998	0.4116	0.3831
IZ1(α_{17})	0.3822	0.3535	0.3763	0.3542
Intercept2(α_{21})	0.9950	1.1038	1.1039	1.1201
Ostan3(α_{22})	0.3614	0.4499	0.4484	0.4519
Ostan10(α_{23})	0.2300	0.2265	0.2438	0.2310
Ostan12(α_{24})	0.4542	0.7037	0.7224	0.7543
Ostan20(α_{25})	−0.7491	−1.0177	−1.0509	−1.0527
Ostan23(α_{26})	−0.5602	−0.6475	−0.6527	−0.6682
Ostan25(α_{27})	−0.7410	−1.0117	−1.0575	−1.0651
M962(α_{28})	−0.8650	−1.0745	−1.1063	−1.1078
M966(α_{29})	−0.4476	−0.5715	−0.5727	−0.5933
CO2(α_{210})	0.1616	0.2055	0.2098	0.2182
Intercept (β_1)	20.9724	21.0939	21.0981	21.1112
log.Worker(β_2)	1.1642	1.1401	1.1383	1.1363
Bourse1(β_3)	0.3445	0.5769	0.6225	0.6206
Isic4.2219(β_4)	−0.5921	−0.6655	−0.6726	−0.6767
ρ_{01}	0.0378	−0.0257	−0.0552	−0.0582
ρ_{02}	−0.2520	−0.3121	−0.3020	−0.3133
ρ_{12}	−0.0355	0.0862	0.1624	0.1711
σ_{00}	1.0857	1.1079	1.1053	1.1071

Prior parameter*: diagonal element of matrices A, B and the values of r_1 and r_2.

Table 7. Estimation of the parameters and standard errors in frequentist (MLE) and Bayesian methods.

		Maximum likelihood estimation		Bayesian estimation	
Model	parameter	Estimate	SE.	Estimate	SE
Selection Model 1	Intercept1(α_{11})	1.604	0.127	1.579	0.122
	Ostan1(α_{12})	−0.970	0.297	−0.938	0.291
	Ostan7(α_{13})	−1.012	0.199	−0.917	0.198
	Ostan11(α_{14})	−1.943	0.509	−1.359	0.461
	M962(α_{15})	−0.377	0.185	−0.343	0.179
	CO2(α_{16})	0.378	0.179	0.400	0.166
	IZ1(α_{17})	0.358	0.141	0.353	0.135
Selection Model 2	Intercept2(α_{21})	1.183	0.080	1.104	0.093
	Ostan3(α_{22})	0.471	0.208	0.450	0.203
	Ostan10(α_{23})	0.153	0.166	0.227	0.161
	Ostan12(α_{24})	0.675	0.284	0.704	0.265
	Ostan20(α_{25})	−1.076	0.181	−1.018	0.182
	Ostan23(α_{26})	−0.681	0.110	−0.647	0.112
	Ostan25(α_{27})	−1.059	0.195	−1.012	0.194
	M962(α_{28})	−1.024	0.162	−1.074	0.135
	M966(α_{29})	−0.567	0.093	−0.572	0.088
	CO2(α_{210})	0.146	0.102	0.205	0.098
Outcome Model	Intercept (β_1)	21.156	0.119	21.094	0.119
	log.Worker(β_2)	1.132	0.031	1.140	0.031
	Bourse1(β_3)	0.611	0.270	0.577	0.261
	Isic4.2219(β_4)	−0.678	0.112	−0.666	0.114
	σ_{00}	1.080	0.065	1.108	0.055
Correlations	ρ_{01}	0.052	0.279	−0.026	0.203
	ρ_{02}	−0.411	0.099	−0.312	0.084
	ρ_{12}	−0.657	0.785	0.086	0.380

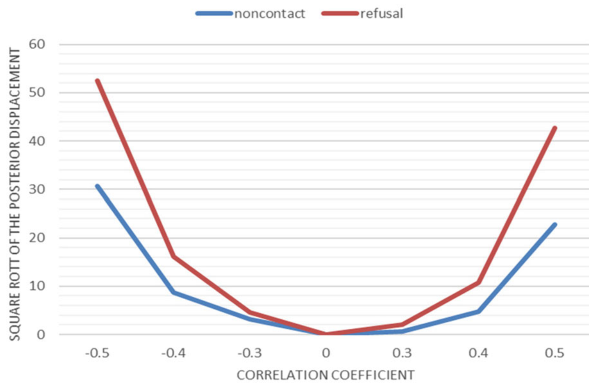

Figure 2. Square root of the posterior displacement for assessing the influence of the sample selection on estimating the parameters from random nonresponse to nonrandom nonresponse for noncontact and refusal reasons.

4.2. Sensitivity analysis

In order to assess the influence of the sample selection model, we consider two cases $\rho_{01} = 0$ and $\rho_{02} = 0$. Figure 2 shows the graph of the posterior displacement for the two cases. It can be seen in the Figure 2, the graph of the PD for ρ_{01} is not curved around zero and it can be concluded that the parameter estimation will not be affected by noncontact nonresponse. This is consistent with the HPD credible set for ρ_{01}, cause this HPD credible set covers 0. But for refusal nonresponse, it can be seen that the value of the PD for ρ_{02} is more curved than that of noncontact and it can be said that the refusal nonresponse has large effect on the estimation of the parameters.

5. Conclusion

In this paper, we presented the Bayesian approach for analysis of the sample selection model with multiple selection rules assuming a multivariate normal distribution for the error components of the models. We considered the priorities for the nonresponse reasons to obtain the parameters estimates.

We examined the performance of the approach using a simulation study and analyzed the data of an establishment survey. Bayesian analysis for such problems can be easily performed using Gibbs sampling and the Metropolis-Hastings method using our proposed algorithm. It turns out from the simulation study that when there is small sample size with useful prior information, the proposed method works well. Moreover, this method can be useful for situations where there is information about the parameters of the outcome model and the selection model. It is useful especially for the analysis of periodic statistics, where information is available based on previous survey in form of priors and it can be used to update the prior using the survey data. Although the methodology is directly applicable to survey data, the algorithm of generation the latent variables, when the number of selection model is high, is time-consuming. Sensitivity to the normality assumption and the use of alternative distribution for variables are topics for future researches.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Albert, J. H., and S. Chib. 1993. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 88:669–79. doi:10.2307/2290350

Carpenter, B., A. Gelman, M. D. Hoffman, L. Daniel, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. 2017. Stan: A probabilistic programming language. *Journal of Statistical Software* 76 (1):1–32. doi:10.18637/jss.v076.i01

Catsiapis, G., and C. Robinson. 1978. Sample selection bias with two selection rules: An application to student aid grants. UWO Department of Economics Working Papers 7833. University of Western Ontario, Department of Economics.

Catsiapis, G., and C. Robinson. 1982. Sample selection bias with multiple selection rules: An application to student aid grants. *Journal of Econometrics* 18:351–68. doi:10.1016/0304-4076(82)90088-4

Chib, S., E. Greenberg, and I. Jeliazkov. 2008. Estimation of semiparametric models in the presence of endogeneity and sample selection. *Journal of Computational and Graphical Statistics* 18:321–48. doi:10.1198/jcgs.2009.07070

Cook, R. D. 1986. Assessment of local influence (with discussion). *Journal of the Royal Statistical Society: Series B (Methodological)* 48 (2):133–69. doi:10.1111/j.2517-6161.1986.tb01398.x

Ding, P. 2014. Bayesian robust inference of sample selection using selection-t models. *Journal of Multivariate Analysis* 124:451–64. doi:10.1016/j.jmva.2013.11.014

Ganjali, M., and M. Rezaei. 2005. An influence approach for sensitivity analysis of non-random dropout based on the covariance structure. *Iranian Journal of Science & Technology, Transaction A* 29 (A2):287–94.

Geweke, J. 1991. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. In *Computing science and statistics: Proceedings of the twenty-third Symposium on the interface*, ed. E. Keramidas, 571–8. Fairfax: Interface Foundation of North America, Inc.

Hanoch, G. 1976. *A multivariate model of labor supply: Methodology for estimation*. Santa Monica, CA: RAND Corporation.

Hasselt, V. M. 2011. Bayesian inference in a sample selection model. *Journal of Econometrics* 165 (2):221–32. doi:10.1016/j.jeconom.2011.08.003

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. *Annals of Economic and Social Measurement* 5 (4):475–92.

Jolani, S. 2014. An analysis of longitudinal data with nonignorable dropout using the truncated multivariate normal distribution. *Journal of Multivariate Analysis* 131:163–73. doi:10.1016/j.jmva.2014.06.016

Kass, R. E., L. Tierney, and J. B. Kadane. 1989. Approximate method for assessing influence and sensitivity in Bayesian analysis. *Biometrika* 76 (4):663–74. doi:10.2307/2336626

Kim, H. J., and H. M. Kim. 2016. Elliptical regression models for multivariate sample selection bias correction. *Journal of the Korean Statistical Society* 45 (3):422–38. doi:10.1016/j.jkss.2016.01.003

Koop, G., and D. J. Poirier. 1997. Learning about the across-regime correlation in switching regression models. *Journal of Econometrics* 78:217–27. doi:10.1016/S0304-4076(96)00009-7

Lewandowski, D., D. Kurowicka, and H. Joe. 2009. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis* 100 (9):1989. doi:10.1016/j.jmva.2009.04.008

Li, P., and M. A. Rahman. 2011. Bayesian analysis of multivariate sample selection models using Gaussian copulas. *Advances in Econometrics* 27:269–88. doi:10.1108/S0731-9053(2011)000027A013.

Little, R. J. A., and D. B. Rubin. 1987. *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.

Maksym and Obrizan. 2011. A Bayesian model of sample selection with a discrete outcome variable: Detecting depression in older adults. No 41, Discussion Papers, Kyiv School of Economics. https://EconPapers.repec.org/RePEc:kse:dpaper:41.

McCulloch, R. E., and P. E. Rossi. 1994. An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics* 64:207–40. doi:10.1016/0304-4076(94)90064-7
Munkin, M. K., and P. K. Trivedi. 2003. Bayesian analysis of a self-selection model with multiple outcomes using simulation-based estimation: an application to the demand for healthcare. *Journal of Econometrics* 114:197–220. doi:10.1016/S0304-4076(02)00223-3
Rahman, M. A., and A. Vossmeyer. 2019. Estimation and applications of quantile regression for binary longitudinal data. *Advances in Econometrics* 40B:157–91.
Razie, F., E. Bahrami, and M. Ganjali. 2013. Analysis of mixed correlated bivariate negative binomial and continuous responses. *Application and Applied Mathematics* 8 (2):404–15.
Rezaee, A., M. Ganjali, and E. Bahrami Samani. 2022. Sample selection bias with multiple dependent selection rules: An application to survey data analysis with multilevel nonresponse. *Swiss Journal of Economics and Statistics* 158:8. 10.1186/s41937-022-00089-1
Steele, F., and G. B. Durrant. 2011. Alternative approaches to multilevel modelling of survey noncontact and refusal. *International Statistical Review* 79 (1):70–91. 10.1111/j.1751-5823.2011.00133.x
Tanner, M. A., and W. H. Wong. 1987. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* 82 (398):528–40. 10.1080/01621459.1987.10478458
Wang, Z., Y. Wu, and H. Chu. 2018. On equivalence of the LKJ distribution and the restricted Wishart distribution. Cornell University. arXiv. Stat. Computation (stat. CO). arXiv:1809.04746 (stat. Co).

Appendix

- **Inverse Mills Ratio formula**

In the equation (6),

$$\lambda_{ij} = \frac{\phi_1\left(w_{ij}\alpha_j\right)\Phi^*_{K-1}\left(w_{ij}\alpha_j\right)}{\Phi_K\left(\mathbf{W}_i\alpha\right)},$$

where $\phi_1\left(w_{ij}\alpha_j\right)$ is the density function of the univariate standard normal distribution evaluated at $w_{ij}\alpha_j$, $\Phi_K\left(\mathbf{W}_i\alpha\right)$ is the cumulative distribution function (cdf) of a K-variates normal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma_{uu}}$ evaluated at $\mathbf{W}_i\boldsymbol{\alpha}$, $\Phi^*_{K-1}\left(w_{ij}\alpha_j\right) \equiv 1$ for $\mathrm{K} = 1$ and for $\mathrm{K} > 1$,

$$\Phi^*_{K-1}\left(w_{ij}\alpha_j\right) = \int_{-\infty}^{w_{i(j)}\alpha_{(j)}} \phi_{K-1}\left(\mathrm{u}_{i(j)}\middle|\, \mathrm{u}_{ij} = w_{ij}\alpha_j\right) du_{i(j)},$$

where $\phi_{K-1}\left(\mathrm{u}_{i(j)}\middle|\,\mathrm{u}_{ij} = w_{ij}\alpha_j\right)$ is the conditional density function of a $\mathrm{K}-1$ variates normal distribution evaluated at the $\mathrm{u}_{i(j)}$ (all u_is without j-th variable) given u_{ij} and $\int_{-\infty}^{w_{i(j)}\alpha_{(j)}} \phi_{K-1}\left(\mathrm{u}_{i(j)}\middle|\,\mathrm{u}_{ij} = w_{ij}\alpha_j\right) du_{i(j)}$ is the $\mathrm{K}-1$ integral of $\phi_{K-1}\left(\mathrm{u}_{i(j)}\middle|\,\mathrm{u}_{ij} = w_{ij}\alpha_j\right)$ on all of u_{it}, $\mathrm{t} = 1, \ldots, \mathrm{K}, \mathrm{t} \neq \mathrm{j}$, v_i is the error component in equation (6) with mean 0 and variance of $\sigma_{00} + \Sigma_{eu}H\Sigma_{ue}$, where H is a $K \times K$ matrix with diagonal elements of $h_{jj} = -w_{ij}\alpha_j\lambda_{ij} - \lambda_{ij}^2$ and off-diagonals of $h_{kl} = \lambda^*_{i,kl} - \lambda_{ik}\lambda_{il}$,

$$\lambda^*_{i,kl} = \frac{\phi_2\left(w_{ik}\alpha_k, w_{il}\alpha_l; \boldsymbol{\Sigma^{kl}_{uu}}\right)\Phi^*_{K-2}\left(w_{ik}\alpha_k, w_{il}\alpha_l\right)}{\Phi_K\left(\mathbf{W_i}\boldsymbol{\alpha}\right)}.$$

Here, $\phi_2\left(w_{ik}\alpha_k, w_{il}\alpha_l; \boldsymbol{\Sigma}^{\mathbf{kl}}_{\mathbf{uu}}\right)$ is the density function of the standard bivariate normal distribution evaluated at $\mathrm{w}'_{ik}\alpha_k$ and $w_{il}\alpha_l$, $\boldsymbol{\Sigma}^{\mathbf{kl}}_{\mathbf{uu}}$ is the covariance matrix of u_k and u_l, $\Phi^*_{K-2}\left(w_{ik}\alpha_k, w_{il}\alpha_l\right) \equiv 1$, $\boldsymbol{\Sigma}^{\mathbf{kl}}_{\mathbf{uu}} = \boldsymbol{\Sigma}_{\mathbf{uu}}$ for $\mathrm{K} = 2$ and for $\mathrm{K} > 2$

$$\Phi^*_{K-2}\left(w_{ik}\alpha_k, w_{il}\alpha_l\right) = \int_{-\infty}^{w_{i(j)}\alpha_{(j)}} \phi_{K-2}\left(u_{i(kl)}\middle|\, u_{ik} = w_{ik}\alpha_k, u_{il} = w_{il}\alpha_{kl}\right) du_{i(kl)}$$

where $\phi_{K-2}\left(\mathrm{u}_{i(kl)}\middle|\,\mathrm{u}_{ik} = w_{ik}\alpha_k, \mathrm{u}_{il} = w_{il}\alpha_l\right)$ is the conditional density function of a $\mathrm{K}-2$ variates normal distribution evaluated at the $\mathrm{u}_{i(kl)}$ (all u_i s without k-th and l-th variable) given u_{ik} and u_{il} and $\int_{-\infty}^{w_{i(j)}\alpha_{(j)}} \phi_{K-2}\left(\mathrm{u}_{i(kl)}\middle|\,\mathrm{u}_{ik} = w_{ik}\alpha_k, \mathrm{u}_{il} = w_{il}\alpha_l\right) du_{i(kl)}$ is the $\mathrm{K}-2$ integral of $\phi_{K-2}(\mathrm{u}_{i(kl)}|\,\mathrm{u}_{ik} = w_{ik}\alpha_k$, $\mathrm{u}_{il} = w_{il}\alpha_l)$ on all of u_{it}, $\mathrm{t} = 1, \ldots, \mathrm{K}, \mathrm{t} \neq \mathrm{k}, \mathrm{l}$, and $\sigma_{00} + \Sigma_{eu}H\Sigma_{ue}$ should be positive.

- **Posterior for $\boldsymbol{\alpha}$**

$$
\begin{aligned}
&\log f\left(\alpha \mid y, y^{*}, \beta, \sigma_{00}, \boldsymbol{\Sigma}_{eu}, \boldsymbol{\Sigma}_{\boldsymbol{uu}}\right) \\
&\propto \sum_{i \in S_0} -\frac{1}{2}\left(\mathrm{y}_{\mathrm{i}}^{*}-\mathbf{W}_{i} \alpha-\frac{y_i-x_i \beta}{\sigma_{00}} \Sigma_{ue}\right)^{\prime} \boldsymbol{\Sigma}_{\boldsymbol{uu}.0}^{-1}\left(\mathrm{y}_{\mathrm{i}}^{*}-\mathbf{W}_{i} \alpha-\frac{y_i-x_i \beta}{\sigma_{00}} \Sigma_{ue}\right) \\
&\quad+\sum_{i \notin S_0}-\frac{1}{2}\left(\mathrm{y}_{\mathrm{i}}^{*}-\mathbf{W}_{i}^{\prime} \alpha\right)^{\prime} \boldsymbol{\Sigma}_{\mathbf{uu}}^{-1}\left(\mathrm{y}_{\mathrm{i}}^{*}-\mathbf{W}_{i}^{\prime} \alpha\right)-\frac{1}{2}(\alpha-\mathrm{a})^{\prime} \mathbf{A}^{\mathbf{-1}}(\alpha-\mathrm{a})+\text{const} \\
&=-\frac{1}{2} \alpha^{\prime}\left(\sum_{i \in S_0} \mathbf{W}_{i}^{\prime} \boldsymbol{\Sigma}_{\boldsymbol{uu}.0}^{-1} \mathbf{W}_{\mathbf{i}}+\sum_{i \notin S_0} \mathbf{W}_{i}^{\prime} \boldsymbol{\Sigma}_{\boldsymbol{uu}}^{-1} \mathbf{W}_{\mathbf{i}}+\mathbf{A}^{\mathbf{-1}}\right) \alpha \\
&\quad+\alpha^{\prime}\left(\sum_{i \in S_0} \mathbf{W}_{i}^{\prime} \boldsymbol{\Sigma}_{\boldsymbol{uu}.0}^{-1}\left(\mathrm{y}_{\mathrm{i}}^{*}-\frac{y_i-x_i \beta}{\sigma_{00}} \Sigma_{ue}\right)+\sum_{i \notin S_0} \mathbf{W}_{i}^{\prime} \boldsymbol{\Sigma}_{\boldsymbol{uu}}^{-1} \mathrm{y}_{\mathrm{i}}^{*}+\mathbf{A}^{\mathbf{-1}} \mathrm{a}\right)+\text{const} \\
&=-\frac{1}{2} \alpha^{\prime} \mathbf{W} \alpha+\alpha^{\prime} \mathbf{z}+\text{const} \\
&=-\frac{1}{2}\left(\alpha-\mathbf{W}^{-1} \mathbf{z}+\mathbf{W}^{-1} \mathbf{z}\right)^{\prime} \mathbf{W}\left(\alpha-\mathbf{W}^{-1} \mathbf{z}+\mathbf{W}^{-1} \mathbf{z}\right)+\alpha^{\prime} \mathbf{z}+\text{const} \\
&=-\frac{1}{2}\left(\alpha-\mathbf{W}^{-1} \mathbf{z}\right)^{\prime} \mathbf{W}\left(\alpha-\mathbf{W}^{-1} \mathbf{z}\right)+\text{const}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{W} &= \sum_{i \in S_0} \mathbf{W}_{i}^{\prime} \boldsymbol{\Sigma}_{uu.0}^{-1} \mathbf{W}_{\mathbf{i}}+\sum_{i \notin S_0} \mathbf{W}_{i}^{\prime} \boldsymbol{\Sigma}_{\boldsymbol{uu}}^{-1} \mathbf{W}_{\mathbf{i}}+\mathbf{A}^{-1} \\
\mathrm{z} &= \sum_{i \in S_0} \mathbf{W}_{i}^{\prime} \boldsymbol{\Sigma}_{uu.0}^{-1}\left(\mathrm{y}_{\mathrm{i}}^{*}-\frac{y_i-x_i \beta}{\sigma_{00}} \Sigma_{ue}\right)+\sum_{i \notin S_0} \mathbf{W}_{i}^{\prime} \boldsymbol{\Sigma}_{\boldsymbol{uu}}^{-1} \mathrm{y}_{\mathrm{i}}^{*}+\mathbf{A}^{-1} \mathrm{a} \\
\overline{\mathbf{A}} &= \mathbf{W}^{\mathbf{-1}} \\
\bar{a} &= \overline{\mathbf{A}} Z
\end{aligned}
$$

- **Posterior for $\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{eu}}$**

$$
\begin{gathered}
\mathrm{g}=\left(\beta, \Sigma_{eu}\right) \\
\mathrm{g} \mid \xi^2 \sim N_{p+K}\left(g_0, \mathbf{G}_{\mathbf{0}}\right)
\end{gathered}
$$

We set: $\mu_{i.u}=E\left(y_i \mid y_i^{*}\right)=x_i \beta+\Sigma_{eu} \boldsymbol{\Sigma}_{\boldsymbol{uu}}^{-1}\left(y_i^{*}-\mathbf{W}_{\mathbf{i}} \alpha\right)$, then we have:

$$
\begin{aligned}
\log \pi\left(\mathrm{g} \mid \xi^2, \boldsymbol{\Sigma}_{\boldsymbol{uu}}, \alpha, y, y^{*}\right) &\propto \sum_{i \in S_0} \log f\left(y_i \mid y_i^{*}, \alpha, \beta, \xi^2, \Sigma_{eu}, \boldsymbol{\Sigma}_{\boldsymbol{uu}}\right)+\log \pi\left(\mathrm{g} \mid \xi^2\right) \\
&=-\frac{1}{2 \xi^2} \sum_{i \in S_0}\left(y_i-\mu_{i.u}\right)^2-\frac{n_0}{2} \log (2 \pi)-\frac{n_0}{2} \log \left(\xi^2\right) \\
&\quad-\frac{1}{2}\left(\mathrm{g}-\mathrm{g}_0\right)^{\prime} \mathbf{G}_0^{\mathbf{-1}}\left(\mathrm{g}-\mathrm{g}_0\right) \\
&\quad-\left(\frac{p+K}{2}\right) \log (2 \pi)-\frac{1}{2} \log \left(|\mathbf{B}| r_1 \ldots r_K \xi^{2K}\right) \\
&=-\frac{1}{2 \xi^2}\left(\mathrm{y}_0-\mathbf{D} \mathrm{g}\right)^{\prime}\left(\mathrm{y}_0-\mathbf{D} \mathrm{g}\right)-\frac{1}{2}\left(\mathrm{g}-\mathrm{g}_0\right)^{\prime} \mathbf{G}_0^{\mathbf{-1}}\left(\mathrm{g}-\mathrm{g}_0\right)+\text{const} \\
&=-\frac{1}{2} \mathrm{g}^{\prime}\left(\xi^{-2} \mathbf{D}^{\prime} \mathbf{D}+\mathbf{G}_0^{\mathbf{-1}}\right)^{\prime} \mathrm{g}+\mathrm{g}^{\prime}\left(\xi^{-2} \mathbf{D}^{\prime} \mathrm{y}_0+\mathbf{G}_0^{\mathbf{-1}} \mathrm{g}_0\right)+\text{const}
\end{aligned}
$$

$$
\begin{aligned}
&= -\frac{1}{2}\mathrm{g}'\bar{\mathbf{G}}^{-\mathbf{1}} + \mathrm{g}'\bar{\mathbf{G}}^{-\mathbf{1}}\bar{\mathrm{g}} + const \\
&= -\frac{1}{2}\left(\mathrm{g} - \bar{\mathrm{g}} + \bar{\mathrm{g}}\right)'\bar{\mathbf{G}}^{-\mathbf{1}}\left(\mathrm{g} - \bar{\mathrm{g}} + \bar{\mathrm{g}}\right) + \ \mathrm{g}'\bar{\mathbf{G}}^{-\mathbf{1}}\bar{\mathrm{g}} + const \\
&= -\frac{1}{2}\left(\mathrm{g} - \bar{\mathrm{g}}\right)'\bar{\mathbf{G}}^{-\mathbf{1}}\left(\mathrm{g} - \bar{\mathrm{g}}\right) + const
\end{aligned}
$$

where $\bar{\mathbf{G}} = (\bar{\mathbf{G}}_{\mathbf{0}}^{-\mathbf{1}} + \xi^{-2}\mathbf{D}'\mathbf{D})^{-\mathbf{1}}$ and $\bar{\mathrm{g}} = \bar{\mathbf{G}}(\bar{\mathbf{G}}_{\mathbf{0}}^{-\mathbf{1}}\mathrm{g}_0 + \xi^{-2}\mathbf{D}'y)$

- **Posterior for $\boldsymbol{\xi^2}$**

$$
\begin{aligned}
\xi^2 &= Var\left(e_i|u_i\right) = \sigma_{00} - \Sigma_{eu}\mathbf{\Sigma}_{\mathbf{uu}}^{-1}\Sigma_{ue} \\
y_i|y_i^* &\sim N\left(\mu_i, \xi^2\right), i \in S_0 \\
\mu_{i.u} &= E\left(y_i\middle|\, y_i^*\right) = x_i\beta + \Sigma_{eu}\Sigma_{\boldsymbol{uu}}^{-1}\left(y_i^* - \mathbf{W}_{\mathbf{i}}'\alpha\right) \\
\pi\left(\xi^2\middle|\, y, y^*, \alpha, \beta, \Sigma_{eu}, \mathbf{\Sigma}_{\boldsymbol{uu}}\right) &\propto \prod_{i\in S_0} f\left(y_i|\xi^2, y_i^*, \alpha, \beta, \Sigma_{eu}, \mathbf{\Sigma}_{\boldsymbol{uu}}\right)\pi\left(\Sigma_{eu}|\,\xi^2\right)\pi\left(\xi^2\right) \\
&\propto \prod_{i\in S_0}\frac{exp\left\{-\frac{1}{2\xi^2}\left(y_i - \mu_{i.u}\right)^2\right\}}{\sqrt{2\pi\xi^2}}\cdot\frac{exp\left\{-\frac{1}{2\xi^2}\sum_{j=1}^{K}\frac{\sigma_{0j}^2}{r_j}\right\}}{\sqrt{(2\pi)^K r_1\ldots r_K\xi^{2K}}} \\
&\cdot\frac{d_0^{c_0}}{\Gamma\left(c_0\right)\left(\xi^2\right)^{c_0+1}}exp\left(\frac{-d_0}{\xi^2}\right) \\
&\propto \exp\left\{-\frac{1}{\xi^2}\left(\frac{1}{2}\sum_{i\epsilon S_0}\left(y_i - \mu_{i.0}\right)^2 + \frac{1}{2}\sum_{j=1}^{K}\frac{\sigma_{0j}^2}{r_j} + d_0\right)\right\} \\
&\frac{1}{\xi^{2\left(\frac{n_0}{2}+\frac{K}{2}+c_0+1\right)}}\frac{\left(\frac{1}{2}\sum_{i\epsilon S_0}\left(y_i - \mu_i\right)^2 + \frac{1}{2}\sum_{j=1}^{K}\frac{\sigma_{0j}^2}{r_j} + d_0\right)^{\frac{n_0}{2}+\frac{K}{2}+c_0}}{\Gamma\left(\frac{n_0}{2} + \frac{K}{2} + c_0\right)} \\
&\equiv IG(\bar{c}, \bar{d}) \\
\bar{c} &= \frac{n_0}{2} + \frac{K}{2} + c_0 \\
\bar{d} &= d_0 + \frac{1}{2}\sum_{i\epsilon S_0}(y_i - \mu_i)^2 + \frac{1}{2}\sum_{j=1}^{K}\frac{\sigma_{0j}^2}{r_j} \\
\mu_i &= x_i\beta + \Sigma_{eu}\Sigma_{\boldsymbol{uu}}^{-1}\left(y_i^* - \mathbf{W}_i\alpha\right).
\end{aligned}
$$