



MONASH University

Australia

Department of Econometrics and Business Statistics

<http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/>

**A Bayesian Sampling Algorithm for a Discrete-Response Model
With Double Rules of Sample Selection**

Rong Zhang, Brett A Inder and Xibin Zhang

November 2013

Working Paper 24/13 (Revised 05/12)

Bayesian estimation of a discrete response model with double rules of sample selection

Rong Zhang[†], Brett A. Inder[‡], Xibin Zhang^{‡,1}

[†]Department of Marketing, Monash University, Australia

[‡]Department of Econometrics and Business Statistics, Monash University, Australia

November 12, 2013

Summary We present a Bayesian sampling algorithm for parameter estimation in a discrete-response model, where the dependent variables contain two layers of binary choices and one ordered response. Our investigation is motivated by an empirical study using such a double-selection rule for three labour-market outcomes, namely labour-force participation, employment and occupational skill level. It is of particular interest to measure the marginal effects of some mental health factors on these labour-market outcomes. The contribution of our investigation is to present a sampling algorithm, which is a hybrid of Gibbs and Metropolis-Hastings algorithms. In Monte Carlo simulations, numerical maximization of likelihood fails to converge for more than half of the simulated samples. Our Bayesian method represents a substantial improvement: it converges in every sample, and performs with similar or better precision than maximum likelihood. We apply our sampling algorithm to the double-selection model of labour-force participation, employment and occupational skill level, where marginal effects of explanatory variables, in particular the mental health factors, on the three labour-force outcomes are assessed through 95% Bayesian credible intervals. The proposed sampling algorithm can easily be modified for other multivariate nonlinear models that involve selectivity and are difficult to estimate by other means.

Keywords: *Gibbs sampler, Marginal effects, Mental illness, Metropolis-Hastings algorithm, Ordered outcome.*

JEL Classification: C01, C11, C35

¹Address: 900 Dandenong Road, Caulfield East, Victoria 3145, Australia. Telephone: +61 3 99032130. Fax: +61 3 99032007. E-mail: xibin.zhang@monash.edu

1 Introduction

Modeling non-random samples has been an important issue in microeconometrics since the seminal work of [Heckman \(1979\)](#) on sample selection. For example, sample selection models are widely used in models of female labour supply and health expenditures ([Amemiya, 1985](#); [Cameron and Trivedi, 2005](#), among others). Although Heckman's model has had a major impact on empirical studies in economics, some researchers argue that the selection mechanism can be misspecified in some cases (see for example, [Blundell, Ham, and Meghir, 1987](#)). If in reality there are two layers of selection, which for example, are the participation and employment selection decisions in the labour market, the standard Heckman's model actually lumps together the two selection rules into a single selection mechanism. This single selection equation is likely to be misspecified, as the parameters are held fixed across the two groups of non-working individuals — non-participants and unemployed. The bias which results from misspecification would also affect estimates of parameters and marginal effects in the outcome equation. This issue has often been dealt with by restricting the sample to labour market participants, and focusing only on selection into employment. Alternatively, one can set up a model that allows for double selection, where participation determines the first selection, and employment determines the second. For example, [Henneberger and Sousa-Poza \(1998\)](#) and [Mohanty \(2001\)](#) applied such a double-selection rule to the wage equations in their models. In this paper, we aim to extend the double-selection model to allow for a discrete response variable in the final outcome equation under two layers of sample selection.

Our investigation is motivated by an empirical study involving three labour-market outcomes, which are the labour-force participation, employment and occupational skill level. In this situation, an individual's occupational skill level can only be observed when she/he passes two barriers of sample selection in the following manner. Employment status can be observed after an individual chooses to participate in the labour force, and the labour outcomes such as income and occupation can only be observed after the individual is employed. The occupational skill

level is of particular interest and becomes the focus of our model. Therefore, our model has three discrete outcomes modeled by three equations, in which the error terms are correlated with each other. In addition, the selection rule of participation precedes the selection rule of employment in our model. The same model was discussed by [Smith \(2003\)](#) to illustrate the computation of the likelihood under the Archimedean copula, and an application of this model was recently studied by [Cornwell, Forbes, Inder, and Meadows \(2009\)](#).

There exist other examples of the double-selection mechanism in the literature. [Caudill and Oswald \(1993\)](#) described an arbitrator's decision through a sequential selectivity model, which contains double rules of selection. An arbitrator decides whether the termination should be sustained or overturned. If the termination is overturned, the arbitrator must decide whether to partially or fully reinstate the grievant. To simplify the estimation, they assumed no correlation exists between the 'overturn' decision and the 'partially reinstate' decision. This allows Heckman's two-step procedure to be used to estimate parameters. [Maumbe and Swinton \(2003\)](#) discussed a similar sequential selection process. In their empirical study, farmers first decide whether to adopt soil testing and then whether to adopt a variable rate technology after soil testing. The outcome is the farm's nitrogen productivity, which can be categorised into three groups after the aforementioned double selection.

Estimation of of single selectivity models has been discussed by [Maddala \(1983\)](#) and [Vella \(1998\)](#). The most common approach is [Heckman's \(1979\)](#) two-step method, which corrects selection bias by including an inverse Mills ratio as an additional regressor. The inclusion of the inverse Mills ratio is based on a bivariate normality assumption, or normality in the selection equation and the regression of the main residual on the selection residual being linear. [Heckman's \(1979\)](#) method was proposed for a single-selection model with a continuous outcome in the main equation. In double-selection models, one might sacrifice a certain degree of estimation accuracy for the convenience of using the two-step method. For example, [Cornwell et al. \(2009\)](#) used a two-step estimation procedure twice in order to take the error-term correlations into the estimation without specifying such correlations explicitly. However, this estimation

method ignores such correlations when estimating parameters in the first equation, and it cannot reveal the strength of correlation between any pair of error terms. Moreover, the nonlinear feature of the main equation in our model also makes the two-step method inappropriate. On the surface it might seem plausible to construct an equivalent to the inverse Mills ratio for nonlinear models, but appropriate evaluation of the conditional expectation of the outcome variable in the presence of selection does not yield a correction term that is analytically tractable. [Greene \(2006\)](#) provided extensive discussion of this issue.

An alternative estimation method is the full information maximum likelihood (FIML) estimation that is often used in empirical studies. However, numerical optimization of FIML often encounters convergence problems even for models with one barrier of sample selection. In the double-selection model under our investigation, such numerical optimization is likely to result in serious convergence problems due to the complicated nature of the model, even though a full likelihood can be obtained under the normality assumption for the error terms.

This paper aims to provide a Bayesian sampling algorithm for parameter estimation in the three-equation model with double rules of sample selection. In the literature on sample selection models, [van Hasselt \(2011\)](#) presented a Bayesian sampling algorithm to estimate parameters in a single selection model, where the outcome of the second equation is missing if the selection indicator in the first equation is zero. [Chib, Greenberg, and Jeliazkov \(2009\)](#) presented a Bayesian sampling approach to parameter estimation for semiparametric models in the presence of endogeneity and sample selection. The model under our investigation is different from the above-mentioned two types of models, and therefore, a new sampling algorithm has to be developed. A remarkable benefit of our proposed Bayesian sampling algorithm is that it allows for explicit specification of correlation parameters among the three error terms, and these parameters can be sampled at the same time as when the coefficients in the mean equations are sampled. Therefore, all parameters in the three-equation model can be sampled within a hybrid of the Gibbs sampler and Metropolis-Hastings algorithm. The proposed Bayesian estimation also facilitates the computation of the 95% Bayesian credible interval for the marginal effect of any

regressor on its corresponding response variable. Moreover, the proposed Bayesian approach can always produce reasonable results even when the numerical optimization of FIML fails to converge.

We present a new reparameterization method, whose purpose is to derive conditional posteriors of some parameters, and therefore, these parameters can be sampled conditional on the other parameters using either the Gibbs sampler or the Metropolis-Hastings algorithm. The reparameterization is necessary because the derivation of such conditional posteriors can speed up the convergence of the resulting sampling procedure. In the literature on Bayesian sampling for discrete-response models, [Cowles \(1996\)](#) found that slow mixing was sometimes caused by high correlation between the estimated threshold and latent variables in ordered probit models. [Li and Tobias \(2006\)](#) applied [Nandram and Chen's \(1996\)](#) reparameterization method to a bivariate ordered probit model to solve this problem. [Li and Tobias \(2006\)](#) reparameterised the parameters in the variance-covariance matrix and derive their conditional posteriors that are the inverse Wishart densities. However, this reparameterization method cannot be directly used in our model, which has only one equation of ordered response. [McCulloch, Polson, and Rossi \(2000\)](#) presented a reparameterization of the parameters of the error variance-covariance matrix in a subset of their multinomial model. In our three-equation model, we propose a reparameterization strategy that combines the techniques from [Li and Tobias \(2006\)](#) and [McCulloch, Polson, and Rossi \(2000\)](#). With this strategy, the conditional posteriors of some parameters can be derived.

We carry out a Monte Carlo simulation study to compare the performance of the proposed Bayesian sampling method with that of FIML estimation. The numerical optimization of FIML fails for more than half of the simulated samples due to the problem that the Hessian matrix cannot be inverted. The Bayesian approach is comparable with FIML in terms of the mean and variation measures of parameter estimates for the simulated samples where FIML works. For the simulated samples where FIML fails, the Bayesian method continues to perform as well as it does for the simulated samples where FIML works. The proposed Bayesian estimation method is

applied to the three-equation model that models an individual’s participation, employment and occupational skill level in the labour force. We present the point and interval estimates of the marginal effect of each explanatory variable on its associated response.

The rest of this paper is organised as follows. The next section describes the formulation of the model. In Section 3, we derive the joint posterior, as well as the conditional posteriors of some parameters through a new reparameterization technique. A Bayesian sampling procedure is also presented. Section 4 presents a Monte Carlo simulation study to compare the performance of the proposed Bayesian method with that of FIML. We use the proposed model and its estimation method to investigate the effect of mental illness on labour-force outcomes in Section 5. The last section concludes the paper.

2 A discrete-response model with double selection

The model of interest has three equations, where in the first equation, the response variable denoted as y_1 , is binary and decides the first hurdle of selection. In the second equation, the response variable denoted as y_2 , is binary and is only observable when $y_1 = 1$. In other words, the second selection rule is censored based on the outcome of the first selection rule. In the third equation, the main equation, the dependent variable denoted as y_3 , has ordered categorical outcomes, which can only be observed after individuals pass the two selection rules. The response variable in our main equation, however, is different from that in Heckman’s (1979) main equation, where the latter is continuous. To fully describe the features of each response variable, we assume that y_{i1} , y_{i2} and y_{i3} are generated respectively, from the following reduced latent variable models:

$$\begin{cases} z_{i1} = x'_{i1}\beta_1 + \varepsilon_{i1}, \\ z_{i2} = x'_{i2}\beta_2 + \varepsilon_{i2}, \\ z_{i3} = x'_{i3}\beta_3 + \varepsilon_{i3}, \end{cases} \quad (1)$$

for $i = 1, 2, \dots, n$, with n being the sample size, where x_{i1} , x_{i2} and x_{i3} are respectively, vectors of explanatory variables, and β_1 , β_2 and β_3 are parameter vectors. It is assumed that in each

equation, the errors are independent and identically distributed (iid).

The binary choice response y_{i1} is defined according to the value of the latent variable z_{i1} . If an individual does not pass the first selection rule, it means that this individual does not participate in the labour force, and thus, y_{i1} is labeled with a zero value. If an individual passes the first selection rule, y_{i1} is labeled with a value of one, and consequently, the response variable of the second equation is observable with $y_{i2} = 1$, for $z_{i2} > 0$; and $y_{i2} = 0$, for $z_{i2} < 0$. In the third equation, the ordered outcomes of y_{3i} can only be observed when $y_{i2} = 1$; otherwise, this response variable is missing and labeled with a zero value. The ordered outcomes are characterised by threshold values $\{\gamma_0, \gamma_1, \gamma_2, \dots, \gamma_J\}$, which divide the values of the latent variable z_{i3} into J categories, where we assume that $\gamma_0 = -\infty$, $\gamma_1 = 0$ and $\gamma_J = \infty$ to avoid any possible identification problem. Therefore, the three observed dependent variables are defined as

$$\begin{cases} y_{i1} = I(z_{i1} > 0), \\ y_{i2} = I(z_{i2} > 0) \times y_{i1}, \\ y_{i3} = j \times y_{i2}, & \text{if } \gamma_{j-1} \leq z_{i3} \leq \gamma_j, \quad \text{for } 1 \leq j \leq J, \end{cases} \quad (2)$$

for $i = 1, 2, \dots, n$, where $I(\cdot)$ is the indicator function with a value one if its argument is true. The largest threshold value γ_{J-1} will be reparameterised, and therefore, the vector of threshold parameters is $\gamma = (\gamma_2, \dots, \gamma_{J-2})'$.

Under the normality assumption for ε_{i1} , ε_{i2} and ε_{i3} , the first two equations in (2) are probit models, while the last is an ordered probit model. It is usually required that the variance of the error term in each equation be one for model identification reasons. Although Bayesian methods can be used for estimation purposes in unidentified models (McCulloch and Rossi, 1994), we focus on standardised equations for simplicity. Importantly, the errors of the three equations are correlated with each other. For example, in the labour market, the decision to participate and the possibility of finding jobs are likely to be driven by common unobservable factors; and people who target certain types of jobs may be less likely to be employed than those who target other jobs. Therefore, it is necessary to allow for correlated error terms. Thus, we assume that

$(\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3}) \sim N(\mathbf{0}_3, \Omega)$ with

$$\Omega = \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{pmatrix}, \quad (3)$$

and $\mathbf{0}_3$ being a column vector of three zeros.

3 A Bayesian sampling algorithm

In this paper, we derive the posterior of all parameters in (2) and develop a sampling algorithm. Due to the complicated feature of this model, we are interested in deriving conditional posteriors of some parameters, and thus, certain types of reparameterization are necessary.

3.1 Reparameterization

In a multivariate ordered probit model, [Li and Tobias \(2006\)](#) proposed to scale the equation with ordered outcome by the largest threshold value. As our third equation has an ordered outcome, we propose to divide the latent equation, which is the third equation of (1), by the largest threshold value. In this way, the differences between two successive reparameterised threshold parameters are sampled through the Metropolis-Hastings algorithm with a Dirichlet proposal density. The original threshold parameters can be solved back from the sampled values of the reparameterised threshold parameters.

Let $\beta_3^* = \beta_3/\gamma_{J-1}$, $z_{i3}^* = z_{i3}/\gamma_{J-1}$ and $\varepsilon_{i3}^* = \varepsilon_{i3}/\gamma_{J-1}$. The latent variables that determine the corresponding responses are modeled as

$$\begin{cases} z_{i1} = x'_{i1}\beta_1 + \varepsilon_{i1}, \\ z_{i2} = x'_{i2}\beta_2 + \varepsilon_{i2}, \\ z_{i3}^* = x'_{i3}\beta_3^* + \varepsilon_{i3}^*, \end{cases} \quad (4)$$

for $i = 1, 2, \dots, n$, where we assume that $(\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3}^*)' \sim N(0, \Sigma^*)$ with

$$\Sigma^* = \begin{pmatrix} 1 & \rho_1 & \rho_2/\gamma_{J-1} \\ \rho_1 & 1 & \rho_3/\gamma_{J-1} \\ \rho_2/\gamma_{J-1} & \rho_3/\gamma_{J-1} & 1/\gamma_{J-1}^2 \end{pmatrix}.$$

Therefore, the double-selection model given by (2) becomes

$$\begin{cases} y_{i1} = I(z_{i1} > 0), \\ y_{i2} = I(z_{i2} > 0) \times y_{i1}, \\ y_{i3} = j \times y_{i2}, \quad \text{if } \gamma_{j-1}^* \leq z_{i3}^* \leq \gamma_j^*, \text{ for } 1 \leq j \leq J, \end{cases} \quad (5)$$

for $i = 1, 2, \dots, n$, where $\gamma_1^* = 0$, $\gamma_j^* = \gamma_j / \gamma_{j-1}$, for $j = 2, \dots, J-2$, $\gamma_{J-1}^* = 1$ and $\gamma_J^* = \infty$. Let the vector of transformed threshold parameters be denoted as $\gamma^* = (\gamma_2^*, \dots, \gamma_{J-2}^*)'$.

We reparameterise the parameters in the variance-covariance matrix, and a new parameter defined by $\psi = |\Omega^*| = (1 - \rho_1^2 - \rho_2^2 - \rho_3^2 + 2\rho_1\rho_2\rho_3) / \gamma_{J-1}^2$, is introduced. Let $\lambda_1 = \rho_1$, $\lambda_2 = \rho_2 / \gamma_{J-1}$ and $\lambda_3 = \rho_3 / \gamma_{J-1}$. Thus, the reparameterised variance-covariance matrix is

$$\Omega^* = \begin{pmatrix} 1 & \lambda_1 & \lambda_2 \\ \lambda_1 & 1 & \lambda_3 \\ \lambda_2 & \lambda_3 & (\psi + \lambda_2^2 + \lambda_3^2 - 2\lambda_1\lambda_2\lambda_3) / (1 - \lambda_1^2) \end{pmatrix}.$$

3.2 Joint posterior of parameters and latent variables

Let \mathbf{y} denote the vector of observed y_1 , y_2 and y_3 , $\beta^* = (\beta'_1, \beta'_2, \beta'_3)'$, $p(\theta^*)$ be the joint prior of $\theta^* = (\beta^{*'}, \psi, \lambda_1, \lambda_2, \lambda_3, \gamma^{*'})'$, and $L(\mathbf{y}|\theta^*, Z^*)$ be the likelihood of \mathbf{y} for given θ^* and Z^* , where Z^* is the vector of latent variables corresponding to observed outcomes.

The joint prior of θ^* is the product of independent marginal priors of β^* , ψ , λ_1 , λ_2 , λ_3 and γ^* . The prior of β^* is assumed to be $p(\beta^*) = \phi_k(\beta^*; \beta_0, B_0^{-1})$, where $\phi_k(\cdot; \beta_0, B_0^{-1})$ is the density function of k -dimensional normal distribution with mean vector β_0 and variance-covariance matrix B_0 . The prior of λ_1 denoted as $p(\lambda_1)$, is the uniform density on $(-1, 1)$, and the prior of (λ_2, λ_3) are the density of $N((\lambda_{02}, \lambda_{03})', C_2^{-1})$. The prior of ψ denoted as $p(\psi)$, is assumed to be the inverse Gamma density denoted as $\text{IG}(a_0/2, b_0/2)$ and given as

$$p(\psi) = \frac{(b_0)^{a_0/2}}{\Gamma(a_0/2)} \left(\frac{1}{\psi}\right)^{b_0/2+1} \exp\left\{-\frac{b_0}{2\psi}\right\}.$$

Instead of directly sampling γ^* , we plan to sample the distance between each pair of successive components of γ^* . The priors of such distances are assumed to be the uniform density on $(0, 1)$.

Note that y_{i2} is observed when $y_{i1} = 1$, while y_{i3} is observed when $y_{i1} = 1$ and $y_{i2} = 1$. Let

$$N_1 = \{i : y_{i1} = 0, \text{ for } i = 1, 2, \dots, n\},$$

$$N_2 = \{i : y_{i1} = 1 \text{ and } y_{i2} = 0, \text{ for } i = 1, 2, \dots, n\},$$

$$N_3 = \{i : y_{i1} = 1 \text{ and } y_{i2} = 1, \text{ for } i = 1, 2, \dots, n\},$$

where the numbers of individuals in N_1 , N_2 and N_3 are denoted as n_1 , n_2 and n_3 , respectively.

As unobserved outcomes do not contribute to the likelihood, the posterior of the parameters and reparameterised latent variables is (up to a normalizing constant)

$$\begin{aligned} p(\theta^*, Z^* | \mathbf{y}) &\propto p(\theta^*) \times p(Z^* | \theta^*) \times p(\mathbf{y} | \theta^*, Z^*) \\ &= p(\theta^*) \times \prod_{i \in N_1} p(z_{i1} | \beta_1) L(y_{i1} = 0 | \beta_1, z_{i1}) \times \prod_{i \in N_2} p(z_{i1}, z_{i2} | \beta_1, \beta_2) L(y_{i1} = 1, y_{i2} = 0 | \beta_1, \beta_2, z_{i1}, z_{i2}) \\ &\quad \times \prod_{i \in N_3} p(z_{i1}, z_{i2}, z_{i3}^* | \beta_1, \beta_2, \beta_3^*) L(y_{i1} = 1, y_{i2} = 1, y_{i3} = j | \beta_1, \beta_2, \beta_3^*, z_{i1}, z_{i2}, z_{i3}^*) \\ &\propto p(\theta^*) \times \prod_{i \in N_1} \exp\left\{-\frac{1}{2}(z_{i1} - x'_{i1}\beta_1)^2\right\} I(z_{i1} \leq 0) \times \\ &\quad \prod_{i \in N_2} (1 - \lambda_1^2)^{-1/2} \exp\left\{-\frac{1}{2}(Z_{N_2} - \mu_{N_2})' \Omega_{N_2}^{-1} (Z_{N_2} - \mu_{N_2})\right\} I(z_{i1} > 0) I(z_{i2} \leq 0) \times \\ &\quad \prod_{i \in N_3} \psi^{-1/2} \exp\left\{-\frac{1}{2}(Z_{N_3} - \mu_{N_3})' \Omega_{N_3}^{-1} (Z_{N_3} - \mu_{N_3})\right\} I(z_{i1} > 0) I(z_{i2} > 0) I(\gamma_{j-1}^* < z_{i3}^* < \gamma_j^*), \end{aligned} \quad (6)$$

where $Z_{N_2} = (z_{i1}, z_{i2})'$, $\mu_{N_2} = (x'_{i1}\beta_1, x'_{i2}\beta_2)'$, $\Omega_{N_2} = \begin{pmatrix} 1 & \lambda_1 \\ \lambda_1 & 1 \end{pmatrix}$, $Z_{N_3} = (z_{i1}, z_{i2}, z_{i3}^*)'$, $\mu_{N_3} = (x'_{i1}\beta_1, x'_{i2}\beta_2, x'_{i3}\beta_3^*)'$ and $\Omega_{N_3} = \Omega^*$.

3.3 Conditional posteriors of latent variables

We begin with the conditional posteriors of reparameterised latent variables given as

$$p(z_{i1} | \theta^*, \mathbf{y}) \propto \exp\left\{-\frac{1}{2}(z_{i1} - x'_{i1}\beta_1)^2\right\} I(z_{i1} \leq 0) \quad \text{for } i \in N_1,$$

$$p(z_{i1}, z_{i2} | \theta^*, \mathbf{y}) \propto \exp\left\{-\frac{1}{2}(Z_{N_2} - \mu_{N_2})' \Omega_{N_2}^{-1} (Z_{N_2} - \mu_{N_2})\right\} I(z_{i1} > 0) I(z_{i2} \leq 0) \quad \text{for } i \in N_2,$$

and

$$p(z_{i1}, z_{i2}, z_{i3}^* | \theta^*, \mathbf{y}) \propto \exp\left\{-\frac{1}{2}(Z_{N_3} - \mu_{N_3})' \Omega_{N_3}^{-1} (Z_{N_3} - \mu_{N_3})\right\} I(z_{i1} > 0) I(z_{i2} > 0) I(\gamma_{j-1}^* < z_{i3}^* < \gamma_j^*),$$

for $i \in N_3$. These probability density functions are truncated normal (TN) densities, and we use the Gibbs sampler discussed by Robert (1995) to sample each latent variable from its conditional posterior. In fact, these latent variables are sampled sequentially. The conditional posterior of z_{i1} is given as

$$z_{i1} | z_{i2}, z_{i3}^* \sim \begin{cases} \text{TN}(x'_{i1}\beta_1, 1) |_{(-\infty, 0]} & \text{if } i \in N_1, \\ \text{TN}(x'_{i1}\beta_1 + \lambda_1(z_{i2} - x'_{i2}\beta_2), 1 - \lambda_1^2) |_{(0, \infty)} & \text{if } i \in N_2, \\ \text{TN}(\mu_{z_{i1}}, \sigma_{z_{i1}}^2) |_{(0, \infty)} & \text{if } i \in N_3, \end{cases} \quad (7)$$

where the mean and variance in the last univariate TN distribution are

$$\begin{aligned} \mu_{z_{i1}} &= x'_{i1}\beta_1 + \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}' \begin{pmatrix} 1 & \lambda_3 \\ \lambda_3 & (\psi + \lambda_2^2 + \lambda_3^2 - 2\lambda_1\lambda_2\lambda_3)/(1 - \lambda_1^2) \end{pmatrix}^{-1} \begin{pmatrix} z_{i2} - x'_{i2}\beta_2 \\ z_{i3}^* - x'_{i3}\beta_3^* \end{pmatrix}, \\ \sigma_{z_{i1}}^2 &= 1 - \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}' \begin{pmatrix} 1 & \lambda_3 \\ \lambda_3 & (\psi + \lambda_2^2 + \lambda_3^2 - 2\lambda_1\lambda_2\lambda_3)/(1 - \lambda_1^2) \end{pmatrix}^{-1} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}. \end{aligned}$$

The conditional posterior of z_{i2} is

$$z_{i2} | z_{i1}, z_{i3}^* \sim \begin{cases} \text{TN}(x'_{i2}\beta_2 + \lambda_1(z_{i1} - x'_{i1}\beta_1), 1 - \lambda_1^2) |_{(-\infty, 0]} & \text{if } i \in N_2, \\ \text{TN}(\mu_{z_{i2}}, \sigma_{z_{i2}}^2) |_{(0, +\infty)} & \text{if } i \in N_3, \end{cases} \quad (8)$$

where the mean and variance in the second TN distributions are

$$\begin{aligned} \mu_{z_{i2}} &= x'_{i2}\beta_2 + \begin{pmatrix} \lambda_1 \\ \lambda_3 \end{pmatrix}' \begin{pmatrix} 1 & \lambda_2 \\ \lambda_2 & (\psi + \lambda_2^2 + \lambda_3^2 - 2\lambda_1\lambda_2\lambda_3)/(1 - \lambda_1^2) \end{pmatrix}^{-1} \begin{pmatrix} z_{i1} - x'_{i1}\beta_1 \\ z_{i3}^* - x'_{i3}\beta_3^* \end{pmatrix}, \\ \sigma_{z_{i2}}^2 &= 1 - \begin{pmatrix} \lambda_1 \\ \lambda_3 \end{pmatrix}' \begin{pmatrix} 1 & \lambda_2 \\ \lambda_2 & (\psi + \lambda_2^2 + \lambda_3^2 - 2\lambda_1\lambda_2\lambda_3)/(1 - \lambda_1^2) \end{pmatrix}^{-1} \begin{pmatrix} \lambda_1 \\ \lambda_3 \end{pmatrix}. \end{aligned}$$

The conditional posterior of z_{i3} is

$$z_{i3}^* | z_{i1}, z_{i2} \sim \text{TN}(\mu_{z_{i3}}, \sigma_{z_{i3}}^2) |_{(\gamma_{j-1}^*, \gamma_j^*)}, \text{ for } y_{i3} = j \text{ and } i \in N_3, \quad (9)$$

which is a univariate TN distribution with mean and variance given by

$$\mu_{z_{i3}} = x'_{i3}\beta_3^* + \begin{pmatrix} \lambda_2 \\ \lambda_3 \end{pmatrix}' \begin{pmatrix} 1 & \lambda_1 \\ \lambda_1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} z_{i1} - x'_{i1}\beta_1 \\ z_{i2} - x'_{i2}\beta_2 \end{pmatrix}, \quad (10)$$

$$\sigma_{z_{i3}}^2 = \frac{\psi + \lambda_2^2 + \lambda_3^2 - 2\lambda_1\lambda_2\lambda_3}{1 - \lambda_1^2} - \begin{pmatrix} \lambda_2 \\ \lambda_3 \end{pmatrix}' \begin{pmatrix} 1 & \lambda_1 \\ \lambda_1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \lambda_2 \\ \lambda_3 \end{pmatrix}. \quad (11)$$

3.4 Conditional posteriors of parameters

3.4.1 Conditional posterior of β^*

Let $Z_{N_2} = \{Z_{iN_2} = (z_{i1}, z_{i2})' : i \in N_2\}$, $Z_{N_3} = \{Z_{iN_3} = (z_{i1}, z_{i2}, z_{i3}^*)' : i \in N_3\}$, $X_{iN_2} = \begin{pmatrix} x'_{i1} & 0 \\ 0 & x'_{i2} \end{pmatrix}$, and $X_{iN_3} = \begin{pmatrix} x'_{i1} & 0 & 0 \\ 0 & x'_{i2} & 0 \\ 0 & 0 & x'_{i3} \end{pmatrix}$. The conditional posterior of β^* is

$$\beta^* | Z^*, \Omega^* \sim N(\tilde{\beta}^*, B^{-1}), \quad (12)$$

where

$$\tilde{\beta}^* = B^{-1} \left\{ B_0 \beta_0 + \begin{pmatrix} \sum_{i \in N_1} x'_{i1} z_{i1} \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \sum_{i \in N_2} X'_{iN_2} \Omega_{N_2}^{-1} Z_{iN_2}^* \\ 0 \end{pmatrix} + \sum_{i \in N_3} X'_{iN_3} \Omega_{N_3}^{-1} Z_{iN_3}^* \right\},$$

$$B = B_0 + \begin{pmatrix} \sum_{i \in N_1} x'_{i1} x_{i1} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} \sum_{i \in N_2} X'_{iN_2} \Omega_{N_2}^{-1} X_{iN_2} & 0 \\ 0 & 0 \end{pmatrix} + \sum_{i \in N_3} X'_{iN_3} \Omega_{N_3}^{-1} X_{iN_3}.$$

The conditional posterior can be derived following a similar procedure discussed by [Chib and Greenberg \(1998\)](#) for multivariate probit models.

3.4.2 Conditional posterior of γ^*

The conditional posterior of the vector of transformed threshold parameters, $(\gamma_2^*, \dots, \gamma_{J-2}^*)'$, is (up to a normalizing constant)

$$p(\gamma_2^*, \dots, \gamma_{J-2}^* | \beta^*, \Sigma^*, Z) \propto I(0 < \gamma_2^* < \dots < \gamma_{J-2}^* < 1). \quad (13)$$

We employ a sampling strategy described by [Li and Tobias \(2006\)](#), who proposed to marginalise the conditional posterior of the transformed threshold parameters over the transformed latent outcome corresponding ordered-outcome equation. This allows them to obtain a closed form conditional posterior of the transformed threshold parameters. Applying the same strategy to our model, we obtain the conditional posterior of $(\gamma_2^*, \dots, \gamma_{J-2}^*)'$ given by

$$p(\gamma_2^*, \dots, \gamma_{J-2}^* | \beta^*, \Omega^*) \propto \prod_{i \in N_3} \left[\Phi\left(\frac{\gamma_{y_{i3}}^* - \mu_{z_{i3}}}{\sigma_{z_{i3}}}\right) - \Phi\left(\frac{\gamma_{y_{i3}-1}^* - \mu_{z_{i3}}}{\sigma_{z_{i3}}}\right) \right],$$

where and $\Phi(\cdot)$ is the cumulative density function (CDF) of the standard normal distribution, and $\mu_{z_{i3}}$ and $\sigma_{z_{i3}}$ are given by (10) and (11), respectively. An independent Metropolis-Hastings algorithm with a Dirichlet proposal density is used to sample the distances between two successive reparameterised threshold parameters, which are $q_j = \gamma_{j+1}^* - \gamma_j^*$, for $j = 1, \dots, J-2$. Due to the fact that $\gamma_1^* = 0$ and $\gamma_{J-1}^* = 1$, the sum of q_j , for $j = 1, \dots, J-2$, is one.

Let \mathbf{q} denote the vector of q_j , for $j = 1, \dots, J-2$, and also denote its current state. A candidate denoted as $\mathbf{q}^{(1)}$ is drawn from the Dirichlet distribution parameterised by $\alpha_j m_j + 1$, for $j = 1, \dots, J-2$, where α_j , for $j = 1, \dots, J-2$, are parameters in the proposal density and pre-determined by users, and $m_j = \sum_{i \in N_3} I(y_{i3} = j)$, for $j = 1, \dots, J-2$.

The probability of accepting this candidate is the minimum value between 1 and

$$r = \left\{ \frac{\prod_{i \in N_3} \left[\Phi \left(\frac{\gamma_{y_{i3}}^{*(1)} - \mu_{z_{i3}}^*}{\sigma_{z_{i3}}^*} \right) - \Phi \left(\frac{\gamma_{y_{i3}-1}^{*(1)} - \mu_{z_{i3}}^*}{\sigma_{z_{i3}}^*} \right) \right]}{\prod_{i \in N_3} \left[\Phi \left(\frac{\gamma_{y_{i3}}^* - \mu_{z_{i3}}^*}{\sigma_{z_{i3}}^*} \right) - \Phi \left(\frac{\gamma_{y_{i3}-1}^* - \mu_{z_{i3}}^*}{\sigma_{z_{i3}}^*} \right) \right]} \right\} \times \left\{ \prod_{j=1}^{J-2} \left(\frac{q_j}{q_j^{(1)}} \right)^{\alpha_j m_j} \right\},$$

where $\gamma_{y_{i3}}^{*(1)}$ is threshold value transformed back from $\mathbf{q}^{(1)}$ with its components given by

$$\gamma_2^{*(1)} = q_1^{(1)}, \gamma_3^{*(1)} = q_2^{(1)} + \gamma_2^{*(1)}, \dots, \gamma_{J-2}^{*(1)} = \gamma_{J-3}^{*(1)} + q_{J-3}^{(1)}.$$

3.4.3 Conditional posterior of ψ

As the prior of ψ denoted as $p(\psi)$, is assumed to be the inverse Gamma density denoted as $IG(a_0/2, b_0/2)$, the conditional posterior of ψ is (up to a normalizing constant)

$$\begin{aligned} p(\psi | Z^*, \beta^*, \lambda_1, \lambda_2, \lambda_3) &\propto p(\psi) \prod_{i \in N_3} \psi^{-1/2} \exp \left\{ -\frac{1}{2} (Z_{N_3} - \mu_{N_3})' \Omega_{N_3}^{-1} (Z_{N_3} - \mu_{N_3}) \right\} \\ &\propto \left(\frac{1}{\psi} \right)^{-a_0/2+1} \exp \left\{ -\frac{b_0}{2\psi} \right\} \left(\frac{1}{\psi} \right)^{-n_3/2} \\ &\exp \left\{ -\frac{1}{2\psi(1-\lambda_1^2)} \sum_{i \in N_3} [(z_{i1} - x'_{i1}\beta_1)(\lambda_2 - \lambda_1\lambda_3) + (z_{i2} - x'_{i2}\beta_2)(\lambda_3 - \lambda_1\lambda_2) - (z_{i3}^* - x'_{i3}\beta_3^*)(1-\lambda_1^2)]^2 \right\}, \end{aligned}$$

where n_3 is the number of individuals in N_3 . After a little algebra operation, we find that the conditional posterior of ψ is also an inverse Gamma density given by

$$\psi | Z^*, \beta^*, \lambda \sim IG(a_1/2, b_1/2), \quad (14)$$

where $a_1 = a_0 + n_3$, and

$$b_1 = b_0 + (1 - \lambda_1^2)^{-1} \sum_{i \in N_3} [(z_{i1} - x'_{i1} \beta_1)(\lambda_2 - \lambda_1 \lambda_3) + (z_{i2} - x'_{i2} \beta_2)(\lambda_3 - \lambda_1 \lambda_2) - (z_{i3}^* - x'_{i3} \beta_3^*)(1 - \lambda_1^2)]^2.$$

3.4.4 Conditional posterior of λ

The conditional posterior of λ_1 is (up to a normalizing constant)

$$p(\lambda_1 | Z^*, \beta^*, \lambda_2, \lambda_3, \psi) \propto p(\lambda_1) \times \prod_{i \in N_2} (1 - \lambda_1^2)^{-1/2} \exp \left\{ -\frac{1}{2} (Z_{N_2} - \mu_{N_2})' \Omega_{N_2}^{-1} (Z_{N_2} - \mu_{N_2}) \right\} \\ \times \prod_{i \in N_3} \exp \left\{ -\frac{1}{2} (Z_{N_3} - \mu_{N_3})' \Omega_{N_3}^{-1} (Z_{N_3} - \mu_{N_3}) \right\}, \quad (15)$$

from which we sample λ_1 using the random-walk Metropolis algorithm with a normal proposal density. The tuning parameter is chosen so that the acceptance rate is between 20% and 30%.

In order to obtain draws of λ_2 and λ_3 , we conceptualise a univariate regression of ε_{i3}^* on $(\varepsilon_{i1}, \varepsilon_{i2})'$, for $i \in N_3$. This conceptual regression has the similar meaning and purpose as [McCulloch, Polson, and Rossi's \(2000\)](#) conceptualization of a multivariate regression of the error terms in the second to the last equations on the error term in the first equation. In their variance-covariance matrix, the first diagonal element is one, while in our variance-covariance matrix, the first two diagonal elements are unity. As a consequence, their method ends up with a multivariate regression with one regressor for multiple-equation models, while ours involves a univariate regression with two regressors for the three-equation model.

As $(\lambda_2, \lambda_3)' = (E(\varepsilon_{i3}^* \varepsilon_{i1}), E(\varepsilon_{i3}^* \varepsilon_{i2}))'$ and $E(\varepsilon_{i3}^2) = (\psi + \lambda_2^2 + \lambda_3^2 - 2\lambda_1 \lambda_2 \lambda_3) / (1 - \lambda_1^2)$, we have $(\varepsilon_{i1}, \varepsilon_{i2})' \sim N(\mathbf{0}_2, \Omega_{N_2})$ and

$$\varepsilon_{i3} | (\varepsilon_{i1}, \varepsilon_{i2})' \sim N \left((\lambda_2, \lambda_3) \Omega_{N_2}^{-1} (\varepsilon_{i1}, \varepsilon_{i2})', \psi \right),$$

where $\mathbf{0}_2$ is a column vector of two zeros. Therefore, we can conceptualise a univariate regression as

$$\varepsilon_{i3} = (\varepsilon_{i1}, \varepsilon_{i2}) \Omega_{N_2}^{-1} (\lambda_2, \lambda_3)' + \eta_i,$$

where η_i , for $i \in N_3$, are independent and follow $N(0, \psi)$. Thus, λ_2 and λ_3 are simply the slope coefficients of the model.

Let $(\varepsilon_1, \varepsilon_2)' = \{(\varepsilon_{i1}, \varepsilon_{i2})' : i \in N_3\}$ be an $n_3 \times 2$ matrix and $\varepsilon_3 = (\varepsilon_{i3} : i \in N_3)$ be an $n_3 \times 1$ vector. As the prior of $(\lambda_2, \lambda_3)'$, which is $N((\lambda_{02}, \lambda_{03})', C_2^{-1})$, is a conjugate prior, we obtain the conditional posterior of $(\lambda_2, \lambda_3)'$:

$$(\lambda_2, \lambda_3)' | Z^*, \beta^*, \lambda_1, \psi \sim N(\mu_\lambda, \Omega_\lambda), \quad (16)$$

where $\Omega_\lambda = \left(\psi^{-1} \Omega_{N_2}^{-1} (\varepsilon_1, \varepsilon_2)' (\varepsilon_1, \varepsilon_2) \Omega_{N_2}^{-1} + C_2 \right)^{-1}$ and $\mu_\lambda = \Omega_\lambda \psi^{-1} \Omega_{N_2}^{-1} (\varepsilon_1, \varepsilon_2)' \varepsilon_3 + C_2 (\lambda_{02}, \lambda_{03})'$.

The final step involves solving back to the original parameters. After $\theta^* = (\beta^{*'}, \gamma^{*'}, \psi, \lambda_1, \lambda_2, \lambda_3)'$ is sampled from the above-given conditional posteriors, the original parameters are calculated as follows: $\gamma_{J-1} = ((1 - \lambda_1^2) / (\psi + \lambda_2^2 + \lambda_3^2 - 2\lambda_1\lambda_2\lambda_3))^{1/2}$, $\beta_3 = \beta_3^* \gamma_{J-1}$, $\rho_1 = \lambda_1$, $\rho_2 = \lambda_2 \gamma_{J-1}$, $\rho_3 = \lambda_3 \gamma_{J-1}$ and $\gamma = \gamma^* \gamma_{J-1}$.

4 Monte Carlo simulation studies

In this section, we conduct a Monte Carlo simulation study with 1,000 simulated samples to demonstrate the performance of our proposed Bayesian estimation method with a sample size of $n = 1,000$.

4.1 Full information maximum likelihood (FIML) estimation

If $y_{i1} = 0$, the joint distribution of y_{i1} , y_{i2} and y_{i3} is simply $\Pr\{y_{i1} = 0 | x_{i1}\}$ denoted by P_0 , because y_{i2} and y_{i3} are not observable. Hence, $P_0 = \Phi(-x'_{i1} \beta_1)$. If $y_{i1} = 1$ and $y_{i2} = 0$, we cannot observe y_{i3} , and the joint distribution of y_{i1} , y_{i2} and y_{i3} is simply $\Pr\{y_{i1}, y_{i2} | x_{i1}, x_{i2}\}$ denoted by P_{10} . Therefore, we have

$$P_{10} = \Phi_2(x'_{i1} \beta_1, -x'_{i2} \beta_2, -\rho_1),$$

where $\Phi_2(\cdot)$ is the CDF of a bivariate normal distribution with variances one and correlation given by its third argument.

If $y_{i1} = 1$ and $y_{i2} = 1$, the observations of y_{i3} can be collected as ordered values from 1 to J . The joint distribution of y_{i1} , y_{i2} and y_{i3} denoted as $P_{11,j}$ is $\Pr\{y_{i1} = 1, y_{i2} = 1, y_{i3} = j | x_{i1}, x_{i2}, x_{i3}\}$,

which is expressed explicitly as follows:

$$\begin{aligned}
P_{11,1} &= \Phi_3(x'_{i1}\beta_1, x'_{i2}\beta_2, -x'_{i3}\beta_3; \rho_1, -\rho_2, -\rho_3), \\
P_{11,j} &= \Phi_3(x'_{i1}\beta_1, x'_{i2}\beta_2, \gamma_j - x'_{i3}\beta_3; \rho_1, -\rho_2, -\rho_3) \\
&\quad - \Phi_3(x'_{i1}\beta_1, x'_{i2}\beta_2, \gamma_{j-1} - x'_{i3}\beta_3; \rho_1, -\rho_2, -\rho_3), \text{ for } j = 2, \dots, J-1, \\
P_{11,J} &= \Phi_3(x'_{i1}\beta_1, x'_{i2}\beta_2, x'_{i3}\beta_3 - \gamma_{J-1}; \rho_1, \rho_2, \rho_3),
\end{aligned} \tag{17}$$

where $\Phi_3(d_1, d_2, d_3; r_1, r_2, r_3)$ is the CDF of a trivariate normal density with variances one, r_1 is the correlation coefficient between d_1 and d_2 , r_2 is the correlation coefficient between d_1 and d_3 , and r_3 is the correlation coefficient between d_2 and d_3 .

The FIML estimation method is to maximise the likelihood given by

$$L_{\text{FIML}}(\mathbf{y}|\theta^*) = \prod_{i=1}^n \left\{ P_0^{(1-y_{i1})} P_{10}^{y_{i1}(1-y_{i2})} \left(\sum_{j=1}^J I(y_{i3} = j) P_{11,j} \right)^{y_{i1}y_{i2}} \right\}, \tag{18}$$

with respect to θ^* .

4.2 Monte Carlo simulation with exclusion restrictions

We generated samples with sample size $n=1,000$ through (1) and (2), where the true parameter values are $\beta_1 = (\beta_{11}, \beta_{12})' = (0.6, -1.2)'$, $\beta_2 = (\beta_{21}, \beta_{22}, \beta_{23})' = (1, -1.5, -1)'$, $\beta_3 = (\beta_{31}, \beta_{32})' = (-0.4, 1.5)'$ and $\gamma = (\gamma_2, \gamma_3)' = (0.8, 1.6)'$. We considered three sets of values for the correlation parameters. The first set is $(\rho_1, \rho_2, \rho_3) = (0, 0, 0)$, which means no correlation between any pair of the three error terms. The second set is $(\rho_1, \rho_2, \rho_3) = (0.25, 0.25, 0.5)$, which reflects a medium level of correlation. The last set is $(\rho_1, \rho_2, \rho_3) = (0.5, 0.8, 0.7)$, which represents strong correlation among the three error terms.

Note that x_{i1} and x_{i3} are 2×1 vectors, while x_{i2} is a 3×1 vector. To allow for the presence of an intercept in each equation, the first elements of x_{i1} , x_{i2} and x_{i3} were set to be one. The second elements of x_{i1} and x_{i2} were randomly generated from the standard normal distribution. The third element of x_{i2} and the second element of x_{i3} were independently generated from the Bernoulli distributions with success probability 0.7. The vector of three error terms was

generated from the trivariate normal distribution with its mean being a vector of zeros and variance-covariance matrix given by (3). Latent variables were calculated and then used to decide the values of y_{i1} , y_{i2} and y_{i3} according to (1) and (2).

For each set of true parameter values of ρ_1 , ρ_2 and ρ_3 , 1,000 samples were generated according to the above-mentioned procedure. In each of the three situations, the relative frequency of $\{y_{i1} = 1\}$ is 65%. The relative frequencies of $\{y_{i2} = 1 | y_{i1} = 1\}$ are 56.3%, 58.2% and 60.2% respectively, under the situations of no correlation, medium correlation and strong correlation. In the situation of no correlation, the relative frequencies that y_{i3} takes values of 1, 2, 3 and 4 conditional on $y_{i2} = 1$ are respectively, 29.1%, 24.1%, 24.5% and 22.4%. In the situation of medium level correlation, these relative frequencies are respectively, 22.2%, 22.3%, 26.1% and 29.5%. In the situation of strong correlation, these relative frequencies are 16.2%, 20.2%, 27.2% and 36.4%, respectively.

We used both FIML and the proposed Bayesian sampling to estimate parameters for the double-selection model based on each generated sample. With the proposed Bayesian sampling, the starting values of parameters are all zero, except for $\gamma_2^* = 0.5$ and $\psi = 1$; the burn-in period contains the first 2,000 draws with the following 10,000 draws recorded; and each parameter is estimated by the corresponding arithmetic mean of 10,000 draws from the corresponding conditional posterior. As to the FIML method, the starting values of parameters are all zero, except for $\gamma_2 = 1$ and $\gamma_3 = 2$ because these threshold values have to be larger than zero.

4.3 Hyperparameter choices and convergence of the sampler

For the Bayesian sampling approach, we choose the hyperparameters of priors as follows: $\beta_0 = \mathbf{0}_7$ which is a vector of 7 zeros, $B_0^{-1} = 1000I_7$ with I_7 being the seven-dimensional identity matrix, $\lambda_{02} = 0$, $C_2 = 1$, $\lambda_{03} = 0$, $C_3 = 1$, $a_0 = 2$ and $b_0 = 0.01$. When sampling $\mathbf{q} = (q_1, \dots, q_{J-2})'$, we choose the tuning parameters of Dirichlet proposal densities to be 0.5, 0.4 and 0.2 under the situations of no correlation, medium correlation and strong correlation, respectively. The resulting acceptance rates are respectively, 55%, 45% and 15%.

When simulating λ_1 , we used the standard normal density as the proposal density. The tuning parameters are respectively, 0.1 in the situation of no correlation and 0.07 in the situation of medium correlation, and thus, the acceptance rates in both situations are between 20% and 30%. In the situation of high correlation, the tuning parameter was chosen either 0.15 or 0.04 so as to control the resulting acceptance rate between 20% and 30% for most generated samples.

As the proposed sampling algorithm for the discrete-response model with double selection is new, one might be interested in the mixing performance, or loosely speaking, the convergence of the sampling algorithm. In the literature on posterior simulation for binary and discrete-response models, it is known that the simulated chains of the correlation parameters usually exhibit slow convergence. In this case, it is likely to be the consequence of the absence of full information conveyed through the observed sample due to sample selections. Unobserved individuals also contain information on the correlations between error terms, but the unobserved individuals contribute nothing to the estimation.

We monitored the convergence status of simulated chains through the simulation inefficiency factors (SIF) (see for example, [Roberts, 1996](#); [Kim, Shepherd, and Chib, 1998](#); [Zhang, Brooks, and King, 2009](#)). The computation of SIF requires us to calculate the batch-mean standard deviation of each simulated chain, which was calculated using 100 batches of the chain with 100 draws in each batch. The SIF value can be approximately explained as the number of draws that are required to produce independent draws. For example, a SIF value of 50 means that we should keep one draw of every 50 draws, and the retained draws are approximately independent. Usually, the smaller a SIF value is, the better the convergence of the simulated chain.² The mean and standard deviation of the SIF values obtained through the 1,000 generated samples are computed, and they would indicate the overall convergence of the proposed Bayesian sampler.

²To our understanding, SIF is not an information criterion or a test statistic. Therefore, it is impossible to find a threshold value to decide whether the mixing is acceptable or not. In our experience, a SIF value of no more than double digits indicates a very good mixing performance, although some researchers may accept a value above 100.

4.4 Results

The simulation results averaged over 1,000 generated samples, are given in Tables 1-3. The FIML estimator sometimes fails to produce meaningful results due to the problem that the Hessian matrix fails to invert. In each table, the notation “FIML*” means that the reported statistics were summarised based on the simulated samples where the numerical optimization of FIML produced meaningful results with the corresponding Hessian matrices invertible. Meanwhile, the proposed Bayesian sampler always produces meaningful results, and the reported statistics were summarised based on all 1,000 generated samples. For comparison purposes, we also report the results obtained from Bayesian sampling based on those generated samples, for which the FIML results were summarised. The summaries of such results are marked as “Bayesian*” and are compared to those marked by “FIML*”.

The numerical optimization of FIML failed to achieve convergence in more than half of the simulated samples; it was able to produce meaningful results in only 420, 453 and 458 simulated samples in the three situations of different levels of error correlation. Obviously, the complexity of the model and the high dimension of its parameter vector have resulted in difficulties in the numerical optimization required by FIML. In contrast, the overall convergence of our sampler is reasonable, even though the simulated chain of ρ_1 exhibits slow convergence, but is acceptable.

Table 1 presents the results of the two estimation methods when samples were simulated under the situation of no correlation among three error terms. The SIF values of the parameters in mean equations are quite small, indicating that the simulated chains of these parameters have achieved very good mixing performance. The largest mean SIF value is 94.7 for ρ_1 , which indicates an acceptable mixing status.

The mean estimate of each parameter obtained through each method is close to the true value of the corresponding parameter. This indicates that both methods can provide largely unbiased estimates when the three error terms are not correlated with each other. We also report the sampling standard deviation of the posterior mean of each parameter. Note that this deviation measure reflects sampling variation of posterior means across 1,000 samples

and has nothing to do with the posterior standard deviation of the posterior mean. We found that the estimates of threshold parameters have less variation than those of the mean-equation parameters, and that the estimates of correlation parameters have larger variations than those of the mean-equation parameters and threshold parameters. This phenomenon is likely to be the consequence of information loss due to double selection. Another way to think about this is the identification issue and the curvature of likelihood function, in which any undesirable feature is likely to increase the sampling variation of the posterior estimates.

Table 2 presents the results obtained under the situation of $(\rho_1, \rho_2, \rho_3)' = (0.25, 0.25, 0.5)'$. An increase of the correlation level among the three error terms slightly increases the mean SIF values for the threshold parameters and mean-equation parameters, but such differences are marginal. However, the increase of such correlations increases the mean SIF values for the correlation parameters with the largest mean SIF being 110.2 for ρ_1 . Thus, the existence of medium level correlation among the three error terms results in slow convergence of the simulated chains of the correlation parameters.

In terms of the mean estimate of each parameter, both methods lead to largely unbiased estimates of parameters. The largest mean bias under Bayesian sampling is 0.017 (or 0.021 based on all simulated samples) for ρ_2 , and this represents 6.8% departure below its true value (or 8.4% based on all simulated samples). The largest mean bias under FIML is 0.011 for ρ_3 , and this represents 2.2% departure below its true value. The sampling standard deviation of the estimated values of each parameter under the situation of medium correlation are respectively, similar to that under the situation of no correlation.

Table 3 presents the results derived through the two estimation methods based on samples simulated under the situation of $(\rho_1, \rho_2, \rho_3) = (0.5, 0.8, 0.7)$. With Bayesian sampling, the mean SIF values are respectively, larger than those derived under the situation of medium level correlation, and the largest mean SIF value is 140.8 for ρ_1 . This indicates that the simulated chains, especially those of the correlation parameters, have a slow mixing speed when the correlation among the three error terms are strong. We believe that the three SIF values indicate acceptable mixing

status of the proposed sampler, considering the fact that data containing information on such correlations could only be partly observed due to sample selection.

Under the situation of strong correlation among error terms, the mean estimates of almost all parameters are very close to their corresponding true values, except for ρ_1 , whose mean estimate is less than the true value by 0.06 or 9.4% (based on all simulated samples).

With Bayesian sampling, the average value of standard deviations for each mean-equation parameter or each threshold parameter obtained under this situation, are similar to that obtained under the situation of no correlation or medium level correlation. However, such a variation measure under the situation of strong correlated errors is clearly smaller than that under the situation of no correlation or medium level correlation. The average standard deviation values for ρ_1 , ρ_2 and ρ_3 are reduced by 11%, 50.6% and 22.8%, respectively. Meanwhile, the results obtained through FIML reveal a similar finding on the variation of parameter estimates.

4.5 Findings revealed from the simulation study

The numerical optimization of FIML fails to achieve convergence for more than half of simulated samples due to the problem that Hessian matrix cannot be inverted. In contrast, our proposed Bayesian sampling approach can always produce meaningful results even for those samples where FIML fails. In this sense, the numerical optimization of FIML cannot compete against our proposed Bayesian sampling method. For the simulated samples where FIML works, the estimate of each parameter is on average, very close to the corresponding true value, and the variation of the estimated values derived across different simulated samples are reasonable.

The proposed Bayesian method is comparable with FIML in terms of the mean and variation measures of parameter estimates for the simulated samples where FIML works. Moreover, for the simulated samples where FIML fails, the Bayesian method performs as well as it does for the simulated samples where FIML works. The only limitation of this sampling approach is the slow mixing of the simulated chain of the correlation parameter between the error terms in the first and second equations. We think it is probably a consequence of information loss due to

sample selection. Also, the mixing status of the simulated chain of this correlation parameter is affected by the relative frequency of individuals who pass the first selection rule. A practical remedy to the problem of slow mixing is to use the posterior mode, rather than the commonly used posterior mean, as an estimate of each correlation parameter after a posterior sample is simulated through the sampling procedure.

For each simulated sample, we calculated the mode of each correlation parameter based on its simulated chain. A kernel estimator of the density of each parameter is obtained based on the corresponding simulated chain with bandwidth selected through likelihood cross-validation. The mode estimate of the parameter is approximated through 10,000 grid points. Table 4 summarises the mean and standard deviation of the mode estimates of each correlation parameter over the same 1,000 samples simulated under each situation of the correlation level. Averaging over the 1,000 simulated samples, we calculated the mean and standard deviation of the mode estimates of each parameter. The two measures for the mean-equation parameters and threshold parameters derived through posterior mode are similar to those derived through posterior mean.

In comparison to the use of posterior mean, the use of posterior mode makes no obvious difference in the mean estimate of each correlation parameter, but slightly increase the mean value of the standard deviations of the mode estimate of each correlation parameter. Nonetheless, posterior mode is more robust than posterior mean with respect to different values of the simulated chain. Therefore, when a simulated chain results in a large SIF value, we recommend using the mode of the simulated chain as an estimate of the corresponding parameter.

5 The impact of mental illness on labour market outcomes

It is widely acknowledged that mental health is an important factor in determining labour market outcomes. Mental illness can affect not only individuals' chances of being employed, but also their capacity to work, the occupational skill levels at which they work, and their earnings. Australian nationwide mental health surveys provide us with the opportunity to examine the relationship between labour market outcomes and mental health factors. In this empirical study,

we look at the impact of mental illness on an individual's chances of participating in the labour force and being employed; and for the employed, we look at the impact of mental illness on an individual's occupational skill level. It is usually expected that an individual's mental illness would hinder her/his chances of participating in the labour force and finding a job. Also, people with mental illness usually work in occupational skill categories that are at a lower level that they would otherwise work in, if they did not suffer from such an illness. This application investigates whether there exists empirical evidence supporting these effects.

5.1 Data

The data are from the National Survey of Mental Health and Wellbeing of Adults in 1997 at Australia. This survey collects information about normal demographic factors and various mental health indicators from 10,614 individuals, where 6,928 individuals participated into the labour force and 6490 were employed. Double rules of sample selection exist because we could only observe occupational levels for individuals who participated in the labour market and then were employed. This data set was first analyzed by [Cornwell et al. \(2009\)](#), who seek to explain labour market outcomes by mental health status. However, if there exists causality in the other direction, the mental health variables would be endogenous, and the resulting parameter estimates would be biased. Therefore, they dealt with endogeneity by the use of temporal information in the data to make sure that mental illness could not have been caused by unemployment experience. Previous research has shown a clear effect of mental illness on employment and occupational skill level.

We used the following specification of exogenous regressors in the model. These include binary indicators of mental illness, categorised into substance use disorders, anxiety disorder and affective disorder. The other regressors include age, gender, education, geographic location indicators, and a socioeconomic index for area (SEIFA). The participation equation contains two more regressors, which are the number of children in the household and a binary variable indicating whether the individual is currently studying. These variables are not included in the

employment equation. Providing identifying variables via these exclusion restrictions generally makes estimation easier than it would be otherwise. These variables are clearly likely to have an effect on participation in the labour force — study and parenthood are two common reasons why people choose not to participate. However, there is no reason to expect that they would have a direct effect on the likelihood of employment or level of occupation (assuming no discrimination in the labour market based on family or other responsibilities).

However, in the second hurdle, the variables that have an effect on employment are all likely to affect occupational skill levels. Thus, regressors are exactly the same for the second and third equations. This means that there exist no exclusion restrictions in the second and third equations. [Greene \(2002, pp. E21-115\)](#) pointed out that because of the nonlinearity of conditional mean functions, it is not necessary to impose exclusion restrictions to achieve identification. While we accept that in practice, relying on nonlinearity for identification may result in imprecise estimation due to only weak identification of parameters, this does provides a set-up that will test the performance of the estimation technique we are proposing in a challenging but strictly identified model setting.

5.2 Estimation

We applied the FIML and Bayesian methods to the estimation of parameters in this three-equation model. The numerical optimization of FIML was carried out through the CML package in GAUSS 9.0 and ended with a failure to derive the variance-covariance matrix, even though we tried all available numerical optimization methods provided by this package. Therefore, FIML is not practically applicable.

We applied the proposed Bayesian sampling algorithm to the estimation of all parameters. The hyperparameters of priors were chosen as follows: $\beta_0 = \mathbf{0}_{65}$ which is a vector of 65 zeros, $B_0^{-1} = 1000I_{65}$ with I_{65} being the 65-dimensional identity matrix, $\lambda_{01} = 0$, $C_1 = 1$, $\lambda_{02} = 0$, $C_2 = 1$, $\lambda_{03} = 0$, $C_3 = 1$, $a_0 = 2$ and $b_0 = 0.01$. The burn-in period contains 10,000 iterations, and the following 100,000 iterations were recorded to calculate the mean of each simulated chain. The

SIF was used to monitor the mixing performance of each simulated chain. The posterior-mean estimates of the mean-equation parameters in the first and second equations, as well as their corresponding SIF values and the 95% Bayesian credible intervals, are reported in the left panels of Tables 5 and 6, respectively. For the third equation, posterior-mean estimates of the mean-equation parameters and their associated SIF values are reported in the left panel of Table 7, and their corresponding 95% Bayesian credible intervals are given in the second column of Table 8.

The estimated correlation parameters are respectively, 0.109 (41.1), -0.199 (21.8) and -0.602 (7.8) with their associated SIF values given in parentheses. The 95% Bayesian credible intervals of the three correlation coefficients are respectively, (0.070, 0.147), (-0.231 , -0.166) and (-0.621 , -0.581). These estimates suggest some sizeable correlations between unobservables across equations.

The estimated threshold parameters are respectively, 0.397 (42.2), 0.617 (25.0) and 0.808 (5.4) with their SIF values given in parentheses. The 95% Bayesian credible intervals of the three threshold parameters are respectively, (0.384, 0.410), (0.602, 0.633) and (0.791, 0.825). All the simulated chains of mean-equation parameters, correlation parameters and threshold parameters have achieved very good mixing performance. Considering the fact that numerical optimization of FIML fails to achieve convergence, we believe that our sampling algorithm has done remarkably well in achieving convergence.

One may have concerns about the two negative values of the estimated correlation parameters. We believe that a negative correlation is possible. For example, the negative correlation between the errors of the employment and the occupational skill equations means that there are unobservables that make some individuals more likely to get a job, but also more likely to have a low-skilled job. This phenomenon could reflect the fact that some might have financial or personal circumstances that make them desperate to gain employment, so they would have a higher chance of participating and gaining employment. In turn, in their desperation for a job, they might take jobs below the skill level of which they are capable. In this case, the individual would have an error in the employment equation that has negative correlation with the error in

the skill-level equation.

5.3 Marginal effects

An important benefit of this sampling approach to parameter estimation is that it can facilitate the computation of point and interval estimates of the marginal effect of any explanatory variable on the probability of a certain response value that is of interest. The probability of participating the labour force is formulated as

$$Pr \{y_{i1} = 1|x_{i1}\} = \Phi(x'_{i1}\beta_1). \quad (19)$$

The probability of being employed conditional on participation is described by

$$Pr \{y_{i2} = 1|y_{i1} = 1, x_{i1}, x_{i2}\} = \frac{\Phi_2(x'_{i1}\beta_1, x'_{i2}\beta_2, \rho_1)}{\Phi(x'_{i1}\beta_1)}. \quad (20)$$

The probability of working in each occupational skill category conditional on employment can be calculated from

$$Pr \{y_{i3} = j|y_{i1} = 1, y_{i2} = 1, x_{i1}, x_{i2}, x_{i3}\} = \frac{P_{11,j}}{\Phi_2(x'_{i1}\beta_1, x'_{i2}\beta_2, \rho_1)}, \quad (21)$$

for $j = 1, 2, \dots, 5$, where $P_{11,j}$ is given by (17).

If the covariate of interest is continuous, its marginal effect is calculated from $\partial Pr(y|x)/\partial x$, in which only the covariate of interest is varied, while the remaining covariates are fixed at their sample mean values. If the covariate of interest is binary, we use the method proposed by Greene (2003) to calculate its marginal effect. Denote the covariate of interest as x_d , then its marginal effect is evaluated by

$$Pr(y|\bar{x}_{(d)}, x_d = 1) - Pr(y|\bar{x}_{(d)}, x_d = 0),$$

where $\bar{x}_{(d)}$ denotes the sample means of the other covariates in the model. Note that in this application, all covariates are binary, except for mental health variables.

With our proposed Bayesian sampling procedure, we are able to derive the point estimate and 95% Bayesian credible interval of the marginal effect of each regressor on its corresponding probability given by (19)-(21).

Table 5 presents the point estimate and 95% Bayesian credible interval of the marginal effect for each explanatory variable on the probability of participation.

Individuals aged between 25 and 44 are 41% more likely to participate in the labour force than 18-24 year-olds, with 45-64 year-olds also significantly more likely to participate. Males, those with higher education and socioeconomic levels, those currently studying and those from rural areas are all more likely to participate. The instrument measuring number of children is clearly relevant, and shows lower participation rates as number of children increase. In terms of health effects, the results suggest that physical illness would decrease the possibility of an individual's participation. The mental health problem of anxiety and affective disorders has no obvious effect on participation, shown by the fact that the 95% Bayesian credible interval of the corresponding marginal effect covers zero. However, individuals with substance use disorders are 10% more likely to participate than those without this type of disorder.

The marginal effects of explanatory variables on the probability of being employed conditional on participation are summarised in the last column of Table 6. The results suggest that employment prospects improve with age, education and socioeconomic status, but do not differ between males and females. There is some evidence of regional differences once socioeconomic status is taken into account. Although physical illness and anxiety disorder have no effect on the probability of being employed, the other two mental disorders would all reduce the possibility of being employed.

We also studied the marginal effect of each explanatory variable on the probability of working in each occupational skill category conditional on employment. Table 7 presents parameter estimates and their corresponding marginal effects, while Table 8 presents the 95% Bayesian credible intervals of the marginal effects. The results around age, education, socioeconomic status and geographic area are all quite sensible: older workers are much more likely to be in the higher skill categories, as are males and those with tertiary education. For the highest 20% using the socioeconomic indicator, compared to the lowest 10% there is an 11.1% greater probability of being in the highest occupational category (professional), but little evidence of

difference in the impact of socioeconomic status on other occupational categories. Individuals with physical illness are less likely to work as professionals and more likely to work at lower levels. The marginal effects of anxiety disorders and affective disorders suggest no clear impact on the level of skill category. However, substance use disorders would slightly decrease the chance of being employed in intermediate and advanced levels.

6 Conclusion

This paper has presented a Bayesian sampling approach to parameter estimation for a discrete-response model with double rules of sample selection, where we have presented a new reparameterization strategy to facilitate the derivation of conditional posteriors. An alternative estimation method is the full information maximum likelihood (FIML) estimation. However, the numerical optimization used in FIML estimation often encounters convergence problems. We have carried out a Monte Carlo simulation study and found that the FIML failed for more than half of the simulated samples due to the problem that the Hessian matrix failed to invert. However, for those simulated samples where FIML fails, our Bayesian method works as well as it does for the simulated samples where FIML works. Moreover, the reparameterization strategy presented here could be used in a range of multiple-equation models involving ordered responses, where numerical optimization used by FIML estimation is likely to struggle to achieve convergence.

We employed the three-equation model with double selection rules to model people's participation in the labour force, employment status and occupational skill levels, where the correlation between any pair of three Gaussian error terms is specified. Applying the proposed sampling algorithm to this model, we derived the estimates of all parameters, as well as their corresponding 95% Bayesian credible intervals. This Bayesian approach also allows us to derive the 95% Bayesian credible interval for the marginal effect of any explanatory variable on its corresponding response. Therefore, we can evaluate whether such a marginal effect is clear-cut or not. The mental health problem of anxiety and affective disorders has no clear effect on participation, but individuals with substance use disorders are more likely to participate than those without

this type of disorder. Anxiety disorder has no effect on the probability of being employed, but affective and substance use disorders all reduce the possibility of being employed. Anxiety and affective disorders have no clear impact on the level of skill category, whereas substance use disorders slightly decrease the chance of being employed in intermediate and advanced levels.

The proposed sampling algorithm can easily be modified for a wide range of multivariate nonlinear models that involve single or double selectivity and are difficult to estimate by other means, and this may include for example, double selection model with count response or continuous response. We leave such modifications for future research should a motivation arise in empirical studies.

Acknowledgements

We wish to thank Murray Smith and Rodney Strachan for their very insightful comments on an early version of this paper, which was a main chapter of the first author's PhD thesis. This research was supported under Australian Research Council's *Discovery Projects* funding scheme (project number DP1095838).

References

- Amemiya, T. (1985). *Advanced Econometrics*, Harvard University Press, New York.
- Blundell, R., J. Ham, and C. Meghir (1987). Unemployment and female labour supply. *The Economic Journal* 97, 44–64.
- Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: Methods and Applications*, Cambridge University Press, London.
- Caudill, S. B. and S. L. Oswald (1993). A sequential selectivity model of the decisions of arbitrators. *Managerial and Decision Economics* 14(3), 261–267.
- Chib, S. and E. Greenberg (1998). Analysis of multivariate probit models. *Biometrika* 85(2), 347–361.

- Chib, S., E. Greenberg, and I. Jeliazkov (2009). Estimation of semiparametric models in the presence of endogeneity and sample selection. *Journal of Computational and Graphical Statistics* 18(2), 321–348.
- Cornwell, K., C. Forbes, B. Inder, and G. Meadows (2009). Mental illness and its effects on labour market outcomes. *The Journal of Mental Health Policy and Economics* 12(3), 107–118.
- Cowles, M. K. (1996). Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing* 6(2), 101–111.
- Greene, W. H. (2002). *LIMDEP Version 8.0: Econometric Modeling Guide*, Econometric Software, New York.
- Greene, W. H. (2003). *Econometric analysis*. Prentice Hall, Upper Saddle River, NJ.
- Greene, W. H. (2006). A general approach to incorporating selectivity in a model. Working paper, Department of Economics, Stern School of Business, New York University.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica* 47(1), 153–161.
- Henneberger, F. and A. Sousa-Poza (1998). Estimating wage functions and wage discrimination using data from the 1995 Swiss labour force survey: A double-selectivity approach. *International Journal of Manpower* 19(7), 486–506.
- Kim, S., N. Shepherd, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65(3), 361–393.
- Li, M. and J. L. Tobias (2006). Bayesian analysis of structural effects in an ordered equation system. *Studies in Nonlinear Dynamics and Econometrics* 10(4), 1–24.
- Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, London.
- Maumbe, B. M. and S. M. Swinton (2003). Adoption of cotton IPM in Zimbabwe: The role of technology awareness and pesticide-related health risks. *Journal of Sustainable Development in Africa* 5(2), 60–86.

- McCulloch, R. and P. E. Rossi (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics* 64(1), 207–240.
- McCulloch, R. E., N. G. Polson, and P. E. Rossi (2000). A Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics* 99(1), 173–193.
- Mohanty, M. S. (2001). Testing for the specification of the wage equation: Double selection approach or single selection approach. *Applied Economics Letters* 8(8), 525–529.
- Nandram, B. and M. H. Chen (1996). Reparameterizing the generalized linear model to accelerate Gibbs sampler convergence. *Journal of Statistical Computation and Simulation* 54, 129–144.
- Robert, C. P. (1995). Simulation of truncated normal variables. *Statistics and Computing* 5, 121–125.
- Roberts, G. O. (1996). Markov chain concepts related to sampling algorithms. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*, pp. 45–57. Chapman & Hall, London.
- Smith, M. D. (2003). Modelling sample selection using Archimedean copulas. *Econometrics Journal* 6(1), 99–123.
- van Hasselt, M. (2011). Bayesian inference in a sample selection model. *Journal of Econometrics* 165, 221–232.
- Vella, F. (1998). Estimating models with sample selection bias: A survey. *Journal of Human Resources* 33, 127–172.
- Zhang, X., R. D. Brooks, and M. L. King (2009). A Bayesian approach to bandwidth selection for multivariate kernel regression with an application to state–price density estimation. *Journal of Econometrics* 153(1), 21–32.

Table 1: A summary of parameter estimates with samples simulated under no correlation among the three error terms

	β_{11}	β_{12}	β_{21}	β_{22}	β_{23}	β_{31}	β_{32}	γ_2	γ_3	ρ_1	ρ_2	ρ_3
True values	0.6	-1.2	1.0	-1.5	-1.0	-0.4	1.5	0.8	1.6	0.0	0.0	0.0
Mean												
FIML*	0.605	-1.213	0.995	-1.504	-0.989	-0.396	1.499	0.791	1.598	0.001	0.003	-0.006
Bayesian	0.606	-1.206	1.001	-1.510	-1.002	-0.399	1.485	0.784	1.571	0.007	-0.003	-0.001
Bayesian*	0.608	-1.211	0.998	-1.504	-0.991	-0.399	1.484	0.777	1.571	0.003	0.004	-0.004
Standard deviation												
FIML*	0.053	0.076	0.149	0.107	0.149	0.143	0.140	0.075	0.096	0.193	0.164	0.145
Bayesian	0.053	0.073	0.149	0.110	0.149	0.141	0.143	0.074	0.095	0.175	0.164	0.140
Bayesian*	0.054	0.076	0.146	0.108	0.148	0.142	0.139	0.074	0.095	0.182	0.167	0.145
SIF												
Mean	6.6	11.4	25.6	17.8	8.0	9.1	1.9	3.7	5.7	94.7	20.4	15.7
Standard deviation	1.7	3.0	8.2	5.6	2.4	2.6	1.1	1.3	2.0	17.1	6.9	4.3

Note: The symbol * indicates that the corresponding summaries are based on 420 simulated samples, for which numerical optimization of FIML achieved convergence. Standard deviation denotes the sampling standard deviation of the FIML point estimates or the Bayesian posterior means or SIF values.

Table 2: A summary of parameter estimates with samples simulated under medium level of correlation among the three error terms

	β_{11}	β_{12}	β_{21}	β_{22}	β_{23}	β_{31}	β_{32}	γ_2	γ_3	ρ_1	ρ_2	ρ_3
True values	0.6	-1.2	1.0	-1.5	-1.0	-0.4	1.5	0.8	1.6	0.25	0.25	0.5
Mean												
FIML*	0.599	-1.211	1.002	-1.515	-1.005	-0.391	1.511	0.805	1.610	0.255	0.245	0.489
Bayesian	0.606	-1.206	1.013	-1.513	-1.004	-0.397	1.486	0.786	1.578	0.238	0.229	0.501
Bayesian*	0.603	-1.208	1.011	-1.514	-1.006	-0.394	1.495	0.790	1.584	0.243	0.233	0.488
Standard deviation												
FIML*	0.052	0.072	0.153	0.106	0.149	0.123	0.131	0.085	0.101	0.166	0.151	0.112
Bayesian	0.054	0.073	0.160	0.111	0.154	0.126	0.133	0.081	0.100	0.173	0.162	0.114
Bayesian*	0.053	0.072	0.152	0.107	0.149	0.123	0.130	0.084	0.100	0.164	0.161	0.113
SIF												
Mean	7.9	13.7	37.5	27.2	12.8	13.2	4.6	8.6	10.5	110.2	36.2	26.3
Standard Deviation	4.6	8.0	14.6	13.8	9.1	7.0	4.6	5.8	8.0	22.6	19.8	13.2

Note: The symbol * indicates that the corresponding summaries are based on 453 simulated samples, for which numerical optimization of FIML achieved convergence. Standard deviation denotes the sampling standard deviation of the FIML point estimates or the Bayesian posterior means or SIF values.

Table 3: A summary of parameter estimates with samples simulated under strong level of correlation among the three error terms

	β_{11}	β_{12}	β_{21}	β_{22}	β_{23}	β_{31}	β_{32}	γ_2	γ_3	ρ_1	ρ_2	ρ_3
True values	0.6	-1.2	1.0	-1.5	-1.0	-0.4	1.5	0.8	1.6	0.5	0.8	0.7
Mean												
FIML*	0.601	-1.207	1.013	-1.514	-1.009	-0.394	1.515	0.810	1.614	0.500	0.793	0.699
Bayesian	0.607	-1.205	1.037	-1.520	-1.014	-0.408	1.491	0.789	1.581	0.453	0.793	0.692
Bayesian*	0.605	-1.203	1.040	-1.518	-1.014	-0.405	1.497	0.793	1.586	0.455	0.791	0.691
Standard deviation												
FIML*	0.053	0.070	0.158	0.111	0.147	0.109	0.124	0.090	0.115	0.147	0.079	0.084
Bayesian	0.053	0.072	0.163	0.112	0.148	0.103	0.124	0.085	0.111	0.154	0.080	0.088
Bayesian*	0.053	0.071	0.162	0.111	0.149	0.110	0.123	0.088	0.114	0.152	0.080	0.087
SIF												
Mean	17.3	34.7	52.8	48.3	24.7	28.9	24.7	33.5	43.9	140.8	80.5	70.6
Standard deviation	9.1	18.9	17.1	19.7	14.2	14.3	15.7	19.7	25.2	23.0	26.8	28.5

Note: The symbol * indicates that the corresponding summaries are based on 458 simulated samples, for which numerical optimization of FIML achieved convergence. Standard deviation denotes the sampling standard deviation of the FIML point estimates or the Bayesian posterior means or SIF values.

Table 4: A summary of the mode statistic for the correlation parameters

	No correlation			Medium correlation			Strong correlation		
	ρ_1	ρ_2	ρ_3	ρ_1	ρ_2	ρ_3	ρ_1	ρ_2	ρ_3
True values	0.0	0.0	0.0	0.25	0.25	0.5	0.5	0.8	0.7
Mean	0.008	-0.004	-0.002	0.269	0.243	0.521	0.440	0.815	0.710
Standard deviation	0.208	0.174	0.147	0.213	0.174	0.117	0.204	0.079	0.088

Table 5: Parameter estimates and marginal effects on the probability of participation

Variable	Coefficient			Marginal effect	
	Estimate	SIF	95% credible interval	Mean	95% credible interval
Age 25-44	1.255	3.2	(1.194, 1.316)	0.410	(0.392, 0.428)
Age 45-64	0.936	3.1	(0.875, 0.998)	0.292	(0.275, 0.308)
Male	0.053	3.2	(-0.011, 0.118)	0.019	(-0.004, 0.042)
Secondary school	0.207	3.0	(0.154, 0.260)	0.073	(0.055, 0.092)
Higher education	0.367	4.5	(0.286, 0.448)	0.121	(0.096, 0.145)
Vocational education	0.178	3.6	(0.095, 0.262)	0.061	(0.033, 0.088)
From a regional center	0.022	2.8	(-0.047, 0.089)	0.008	(-0.017, 0.031)
From a rural area	0.143	3.3	(0.080, 0.206)	0.050	(0.028, 0.071)
SEIFA: 2nd decile	0.193	3.3	(0.088, 0.297)	0.065	(0.031, 0.099)
SEIFA: 3rd decile	0.279	3.0	(0.175, 0.382)	0.093	(0.060, 0.124)
SEIFA: 4th decile	0.283	3.3	(0.176, 0.390)	0.094	(0.060, 0.126)
SEIFA: 5th decile	0.348	2.8	(0.243, 0.454)	0.114	(0.082, 0.145)
SEIFA: 6th decile	0.386	3.1	(0.277, 0.496)	0.125	(0.093, 0.156)
SEIFA: 7th decile	0.388	2.8	(0.285, 0.491)	0.126	(0.095, 0.155)
SEIFA: 8th decile	0.478	3.0	(0.371, 0.585)	0.151	(0.121, 0.179)
SEIFA: 9th & 10th deciles	0.439	3.1	(0.348, 0.529)	0.146	(0.117, 0.173)
Number of children	-0.259	3.6	(-0.280, -0.237)	-0.092	(-0.100, -0.084)
Currently studying	0.360	2.5	(-0.032, 0.756)	0.109	(-0.011, 0.208)
Physical illness	-0.514	3.0	(-0.563, -0.466)	-0.186	(-0.203, -0.168)
Anxiety disorder	-0.080	2.8	(-0.172, 0.012)	-0.029	(-0.063, 0.004)
Affective disorder	0.003	3.0	(-0.094, 0.101)	0.001	(-0.034, 0.035)
Substance use disorder	0.302	3.9	(0.208, 0.397)	0.100	(0.070, 0.128)

Table 6: Parameter estimates and marginal effects on the probability of being employed

Variable	Coefficient			Marginal effect	
	Estimate	SIF	95% credible interval	Mean	95% credible interval
Age 25-44	0.349	8.9	(0.241, 0.455)	0.036	(0.022, 0.050)
Age 45-64	0.383	8.9	(0.261, 0.505)	0.040	(0.026, 0.054)
Male	0.010	9.8	(-0.071, 0.091)	0.001	(-0.010, 0.012)
Secondary school	0.216	9.9	(0.126, 0.306)	0.027	(0.015, 0.038)
Higher education	0.527	17.0	(0.381, 0.677)	0.052	(0.040, 0.063)
Vocational education	0.260	12.9	(0.112, 0.413)	0.028	(0.013, 0.042)
From a regional center	-0.102	7.4	(-0.211, 0.008)	-0.015	(-0.032, 0.001)
From a rural area	0.115	10.0	(0.004, 0.226)	0.013	(-0.001, 0.026)
SEIFA: 2nd decile	0.139	6.2	(-0.027, 0.306)	0.015	(-0.005, 0.033)
SEIFA: 3rd decile	0.303	7.0	(0.134, 0.472)	0.032	(0.015, 0.047)
SEIFA: 4th decile	0.316	6.8	(0.140, 0.492)	0.033	(0.015, 0.048)
SEIFA: 5th decile	0.319	7.9	(0.147, 0.491)	0.033	(0.016, 0.048)
SEIFA: 6th decile	0.562	9.9	(0.376, 0.750)	0.051	(0.038, 0.063)
SEIFA: 7th decile	0.434	7.8	(0.262, 0.606)	0.043	(0.028, 0.056)
SEIFA: 8th decile	0.363	8.2	(0.194, 0.534)	0.036	(0.020, 0.050)
SEIFA: 9th & 10th deciles	0.500	8.8	(0.349, 0.651)	0.053	(0.038, 0.067)
Physical illness	-0.051	8.4	(-0.139, 0.038)	-0.003	(-0.015, 0.009)
Anxiety disorder	-0.129	6.7	(-0.273, 0.018)	-0.019	(-0.042, 0.003)
Affective disorder	-0.241	6.7	(-0.386, -0.094)	-0.038	(-0.066, -0.013)
Substance use disorder	-0.280	5.8	(-0.397, -0.161)	-0.048	(-0.071, -0.027)

Table 7: Parameter estimates and marginal effects on the probability of being employed differer levels of occupation skill category

Variable	Coefficient		Mean of marginal effect				
	Estimate	SIF	Elementary skill	Intermediate skill	Advanced skill	Associate professionals	Professionals
Age 25-44	0.102	2.9	-0.182	-0.030	0.000	0.010	0.202
Age 45-64	0.165	2.1	-0.157	-0.031	-0.004	0.006	0.186
Male	0.115	2.2	-0.052	-0.006	0.001	0.004	0.053
Secondary school	0.076	1.6	-0.057	-0.009	0.001	0.004	0.062
Higher education	1.139	2.7	-0.357	-0.107	-0.037	-0.013	0.514
Vocational education	0.494	1.7	-0.191	-0.043	-0.008	0.004	0.238
From a regional center	-0.013	1.6	0.005	0.000	0.000	0.000	-0.004
From a rural area	0.168	1.8	-0.082	-0.013	0.000	0.005	0.090
SEIFA: 2nd decile	0.054	1.5	-0.043	-0.008	-0.001	0.002	0.049
SEIFA: 3rd decile	-0.034	1.5	-0.019	-0.006	-0.001	0.000	0.026
SEIFA: 4th decile	-0.035	1.5	-0.020	-0.006	-0.001	0.000	0.027
SEIFA: 5th decile	0.042	1.6	-0.054	-0.012	-0.002	0.002	0.066
SEIFA: 6th decile	-0.037	1.6	-0.030	-0.009	-0.002	0.000	0.040
SEIFA: 7th decile	0.051	1.6	-0.063	-0.013	-0.002	0.002	0.076
SEIFA: 8th decile	0.091	1.8	-0.083	-0.018	-0.003	0.003	0.102
SEIFA: 9th & 10th deciles	0.116	1.6	-0.096	-0.018	-0.002	0.004	0.111
Physical illness	0.042	2.1	0.046	0.010	0.001	-0.002	-0.055
Anxiety disorder	-0.046	1.5	0.031	0.003	-0.001	-0.002	-0.031
Affective disorder	0.055	1.5	-0.018	-0.002	0.000	0.001	0.019
Substance use disorder	-0.048	1.4	-0.007	-0.004	-0.002	-0.001	0.014

Table 8: The 95% Bayesian credible intervals of parameters and marginal effects on the probability of being employed differer levels of occupation skill category

Variable	Credible interval of coefficient	Credible interval of marginal effect				
		Elementary skill	Intermediate skill	Advanced skill	Associate professionals	
Age 25-44	(0.034, 0.169)	(-0.212, -0.153)	(-0.033, -0.027)	(-0.002, 0.002)	(0.007, 0.013)	(0.175, 0.229)
Age 45-64	(0.091, 0.238)	(-0.185, -0.129)	(-0.035, -0.027)	(-0.005, -0.002)	(0.004, 0.008)	(0.156, 0.215)
Male	(0.069, 0.160)	(-0.071, -0.033)	(-0.009, -0.004)	(0.001, 0.002)	(0.002, 0.005)	(0.034, 0.073)
Secondary school	(0.027, 0.126)	(-0.078, -0.037)	(-0.011, -0.006)	(0.000, 0.001)	(0.002, 0.005)	(0.041, 0.082)
Higher education	(1.066, 1.213)	(-0.373, -0.342)	(-0.115, -0.098)	(-0.042, -0.031)	(-0.017, -0.009)	(0.488, 0.539)
Vocational Education	(0.421, 0.568)	(-0.212, -0.169)	(-0.052, -0.035)	(-0.012, -0.005)	(0.002, 0.006)	(0.206, 0.270)
From a regional center	(-0.081, 0.055)	(-0.023, 0.032)	(-0.003, 0.003)	(-0.001, 0.001)	(-0.003, 0.002)	(-0.031, 0.024)
From a rural area	(0.107, 0.229)	(-0.104, -0.059)	(-0.018, -0.009)	(-0.001, 0.001)	(0.003, 0.006)	(0.064, 0.116)
SEIFA: 2nd decile	(-0.062, 0.168)	(-0.086, 0.002)	(-0.015, -0.002)	(-0.002, 0.000)	(-0.001, 0.004)	(0.002, 0.097)
SEIFA: 3rd decile	(-0.146, 0.077)	(-0.062, 0.025)	(-0.012, -0.001)	(-0.003, 0.000)	(-0.003, 0.003)	(-0.019, 0.072)
SEIFA: 4th decile	(-0.149, 0.080)	(-0.063, 0.025)	(-0.012, -0.001)	(-0.003, 0.000)	(-0.003, 0.003)	(-0.019, 0.073)
SEIFA: 5th decile	(-0.071, 0.153)	(-0.095, -0.012)	(-0.019, -0.005)	(-0.003, -0.001)	(0.000, 0.004)	(0.020, 0.112)
SEIFA: 6th decile	(-0.149, 0.075)	(-0.073, 0.013)	(-0.015, -0.003)	(-0.003, -0.001)	(-0.002, 0.003)	(-0.005, 0.087)
SEIFA: 7th decile	(-0.058, 0.159)	(-0.102, -0.022)	(-0.021, -0.007)	(-0.003, -0.001)	(0.000, 0.004)	(0.031, 0.121)
SEIFA: 8th decile	(-0.019, 0.201)	(-0.122, -0.044)	(-0.026, -0.011)	(-0.004, -0.002)	(0.001, 0.004)	(0.056, 0.148)
SEIFA: 9th & 10th deciles	(0.020, 0.213)	(-0.131, -0.059)	(-0.024, -0.012)	(-0.003, 0.000)	(0.002, 0.006)	(0.071, 0.151)
Physical illness	(-0.010, 0.093)	(0.024, 0.068)	(0.008, 0.012)	(0.000, 0.002)	(-0.004, 0.000)	(-0.076, -0.034)
Anxiety disorder	(-0.139, 0.047)	(-0.007, 0.071)	(0.000, 0.006)	(-0.003, 0.001)	(-0.006, 0.001)	(-0.067, 0.006)
Affective disorder	(-0.042, 0.153)	(-0.056, 0.021)	(-0.008, 0.002)	(-0.001, 0.001)	(-0.002, 0.004)	(-0.021, 0.059)
Substance use disorder	(-0.128, 0.032)	(-0.039, 0.025)	(-0.008, -0.001)	(-0.003, -0.001)	(-0.003, 0.002)	(-0.018, 0.046)