



MONASH University

Australia

Department of Econometrics and Business Statistics

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

Parameter estimation for a discrete-response model with double
rules of sample selection: A Bayesian approach

Rong Zhang, Brett A. Inder, Xibin Zhang

February 2012

Working Paper 05/12

Parameter estimation for a discrete-response model with double rules of sample selection: A Bayesian approach

Rong Zhang, Brett A. Inder, Xibin Zhang¹

Department of Econometrics and Business Statistics, Monash University

February 2012

Abstract: We present a Bayesian sampling approach to parameter estimation in a discrete-response model with double rules of selectivity, 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 labor-market outcomes, namely labor force participation, employment and occupational skill level. Full information maximum likelihood (FIML) estimation often encounters convergence problems in numerical optimization. The contribution of our investigation is to present a sampling algorithm through a new reparameterization strategy. We conduct Monte Carlo simulation studies and find that the numerical optimization of FIML fails for more than half of the simulated samples. Our Bayesian method performs as well as FIML for the simulated samples where FIML works. Moreover, for the simulated samples where FIML fails, Bayesian works as well as it does for the simulated samples where FIML works. We apply the proposed sampling algorithm to the double-selection model of labor-force participation, employment and occupational skill level. We derive the 95% Bayesian credible intervals for marginal effects of the explanatory variables on the three labor-force outcomes. In particular, the marginal effects of mental health factors on these three outcomes are discussed.

Key words: Bayesian sampling; conditional posterior; marginal effects; mental illness; reparameterization.

JEL Classification: C35, C11

¹Address: 900 Dandenong Road, Caulfield East, Victoria 3145, Australia. Telephone: +61-3-99032130. Fax: +61-3-99032007. Email: 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 to illustrate female labor supply and health expenditures ([Amemiya, 1985](#); [Cameron and Trivedi, 2005](#), among others). Although Heckman's model has the most outstanding impact on empirical studies in economics, some researchers argue that unemployed individuals are misspecified as non-participants in this model ([Blundell, Ham, and Meghir, 1987](#)). This has motivated various studies related to unemployed individuals who want to work at the market wage but cannot find a job. One specification uses double selections, 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 investigate parameter estimation for a double-selection model, which involves three discrete response variables under two layers of sample selection.

Our investigation is motivated by an empirical study involving three labor-market outcomes, which are the labor-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 labor force; and the intensive labor outcomes such as income and occupation, can only be observed after the individual is employed. Moreover, 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 dominates the selection rule of employment in our model. The same model was discussed by [Smith \(2003\)](#) to illustrate the computation of likelihood under the Archimedean copula, and an application of this model was recently studied by [Cornwell,](#)

[Forbes, Inder, and Meadows \(2009\)](#).

Models with simple selectivity could be extended to other realistic models with complex selection rules, and the issue of parameter estimation was discussed by [Maddala \(1983\)](#) and [Vella \(1998\)](#). The most commonly used approach is [Heckman's \(1979\)](#) two-step method, which corrects selection bias by including an inverse Mills ratio as an additional regressor. This method is only suitable for a single-selection model with a continuous outcome in the main equation. Nonetheless, one might sacrifice a certain degree of estimation accuracy for the convenience of using the two-step method in double-selection models. In terms of the model that is of our interest, [Cornwell et al. \(2009\)](#) used such 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.

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, although a full likelihood can be obtained under the normality assumption of the error terms.

This paper aims to provide a Bayesian sampling approach to parameter estimation in the three-equation model with double rules of sample selection. In the literature of sample selection models, [van Hasselt \(2011\)](#) presented a Bayesian sampling algorithm to estimate parameters in a single selection model, where the second equation is a Tobit model conditional on the binary outcomes resulted from selection in the first equation. [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 (MH) 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 MH algorithm. The reparameterization is certainly necessary because the derivation of such conditional posteriors can speed up the convergence of the resulting sampling procedure. In the literature of Bayesian sampling for discrete-response models, [Cowles \(1996\)](#) found that a 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 \(2006\)](#) reparameterized 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 \(2000\)](#). Therefore, 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 performs 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 labor force. We derive the point and interval estimates of the marginal effect of each explanatory variable on its associate response.

The rest of this paper is organized 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 labor-force outcomes in [Section 5](#). The last section concludes the paper.

2 A discrete-response model with double selections

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 that is 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 forms:

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

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, this individual's status is missing and y_{i1} is assigned a zero value; otherwise, the second equation's response variable 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 assigned a zero value. The ordered outcomes are characterized 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

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

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 reparameterized, 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 the probit models, while the last is an ordered probit model. Thus, it is required that the variance of the error term in each equation be one for identification reasons. Moreover, the errors of the three equations are correlated with each other. In the labor market, the decision to participate and the possibility of finding jobs are driven by unobservable factors; and people who target professional jobs may be less likely to be employed than those who target other jobs. Therefore, the assumption of correlated error terms is realistic. Thus, we assume that $(\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3}) \sim N(0, \Omega)$ with

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

3 A Bayesian sampling algorithm

In this paper, we derive the posterior of all parameters in (2) and develop a sampling algorithm to sample these parameters. 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 divide each equation by the largest threshold parameter. As our third equation has an ordered outcome, we propose to divide the latent equation, which is the third equation in (1), by the largest threshold parameter. Therefore, the transformed threshold parameters can be sampled through the Metropolis-Hastings (MH) algorithm with a Dirichlet proposal density.

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 $(\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3}^*)' \sim N(0, \Omega^*)$ with

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

Then 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 $\gamma^* = \gamma/\gamma_{J-1} = (\gamma_2^*, \dots, \gamma_{J-2}^*)'$.

We also reparameterize the parameters in the variance-covariance matrix by generalizing [McCulloch, Polson, and Rossi's \(2000\)](#) reparameterization method from their 2×2 matrix to the 3×3 matrix in our model given by (2). As a consequence, 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 reparameterized variance-covariance matrix is

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

Let's define some notations: $\beta^* = (\beta'_1, \beta'_2, \beta'_3)^t$; $\theta^* = (\beta^{*t}, \gamma^{*t}, \psi, \lambda_1, \lambda_2, \lambda_3)^t$; $Z_i^* = (z_{i1}, z_{i2}, z_{i3}^*)^t$; and $\mu_i^* = (x'_{i1}\beta_1, x'_{i2}\beta_2, x'_{i3}\beta_3^*)^t$. Let \mathbf{y} denote the collection of observed y_1 , y_2 and y_3 , $p(\theta^*)$ the joint prior of θ^* , and $L(\mathbf{y}|\theta^*, Z^*)$ the likelihood for given θ^* and Z^* . The posterior of the

reparameterized latent variables and parameters is (up to a normalizing constant)

$$\begin{aligned}
p(\theta^*, Z^* | Y) &\propto p(\theta^*)p(Z^* | \theta^*)L(\mathbf{y} | \theta^*, Z^*) \\
&= p(\theta^*) \prod_{i=1}^n \phi_3(Z_i^*; \mu_i^*, \Omega^*) \left[I(z_{i1} \leq 0)I(y_{i1} = 0)I(y_{i2} = 0)I(y_{i3} = 0) \right. \\
&\quad + I(z_{i1} > 0)I(y_{i1} = 1)I(z_{i2} \leq 0)I(y_{i2} = 0)I(y_{i3} = 0) \\
&\quad \left. + I(z_{i1} > 0)I(y_{i1} = 1)I(z_{i2} > 0)I(y_{i2} = 1) \sum_{j=1}^J I(y_{i3} = j)I(\gamma_{j-1}^* < z_{i3}^* < \gamma_j^*) \right], \tag{6}
\end{aligned}$$

where $\phi_3(Z_i^*; \mu_i^*, \Omega^*) \propto \psi^{-1/2} \exp\{-\frac{1}{2}(Z_i^* - \mu_i^*)' \Omega^{*-1} (Z_i^* - \mu_i^*)\}$.

3.2 Conditional posteriors of latent variables

We start with sampling from the conditional posterior of reparameterized latent variables expressed as

$$\begin{aligned}
p(Z_i^* | \theta^*, Y_i) &\propto \phi_3(Z_i^*; \mu_i^*, \Sigma^*) \\
&\times \left[I(z_{i1} \leq 0)I(y_{i1} = 0)I_{(y_{i2}=0)}I(y_{i3} = 0) \right. \\
&\quad + I(z_{i1} > 0)I(y_{i1} = 1)I(z_{i2} \leq 0)I(y_{i2} = 0)I(y_{i3} = 0) \\
&\quad \left. + I(z_{i1} > 0)I(y_{i1} = 1)I(z_{i2} > 0)I(y_{i2} = 1) \sum_{j=1}^J I(y_{i3} = j)I(\gamma_{j-1}^* < z_{i3}^* < \gamma_j^*) \right],
\end{aligned}$$

which is a truncated normal (TN) density. We use the Gibbs sampler discussed by [Robert \(1995\)](#) to sample Z_i^* from this condition posterior. In fact, these latent variables are sampled sequentially. The conditional posterior of z_{i1} is given as

$$z_{i1} | z_{i2}, z_{i3}^* \sim \begin{cases} \text{TN}(\mu_{z_{i1}}^*, \sigma_{z_{i1}}^{*2}) |_{(0, +\infty)}, & \text{if } y_{i1} = 1, \\ \text{TN}(\mu_{z_{i1}}^*, \sigma_{z_{i1}}^{*2}) |_{(-\infty, 0]}, & \text{if } y_{i1} = 0, \end{cases} \tag{7}$$

where the mean and variance in the two univariate truncated normal distributions are

$$\mu_{z_{i1}}^* = x'_{i1} \beta_1 + \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}' \begin{pmatrix} 1 & \lambda_3 \\ \lambda_3 & (\psi + \lambda_2^2 + \lambda_3^2 - 2\lambda_1\lambda_2\lambda_3) / (1 - \lambda_1^2) \end{pmatrix}^{-1} \begin{pmatrix} z_{i2} - x'_{i2} \beta_2 \\ z_{i3}^* - x'_{i3} \beta_3^* \end{pmatrix}, \tag{8}$$

$$\sigma_{z_{i1}}^{*2} = 1 - \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}' \begin{pmatrix} 1 & \lambda_3 \\ \lambda_3 & (\psi + \lambda_2^2 + \lambda_3^2 - 2\lambda_1\lambda_2\lambda_3) / (1 - \lambda_1^2) \end{pmatrix}^{-1} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}. \tag{9}$$

The posterior of each latent variable in the second component is

$$z_{i2}|z_{i1}, z_{i3}^* \sim \begin{cases} \text{TN}(\mu_{z_{i2}}^*, \sigma_{z_{i2}}^{*2})|_{(0, \infty)}, & \text{if } y_{i1} = 1 \text{ and } y_{i2} = 1, \\ \text{TN}(\mu_{z_{i2}}^*, \sigma_{z_{i2}}^{*2})|_{(-\infty, 0]}, & \text{if } y_{i1} = 1 \text{ and } y_{i2} = 0, \\ N(\mu_{z_{i2}}^*, \sigma_{z_{i2}}^{*2}), & \text{if } y_{i1} = 0, \end{cases} \quad (10)$$

where the mean and variance in the above normal distributions are

$$\mu_{z_{i2}}^* = x'_{i2}\beta_2 + \begin{pmatrix} \lambda_1 \\ \lambda_3 \end{pmatrix}' \begin{pmatrix} 1 & \lambda_2 \\ \lambda_2 & (\psi + \lambda_2^2 + \lambda_3^2 - 2\lambda_1\lambda_2\lambda_3) / (1 - \lambda_1^2) \end{pmatrix}^{-1} \begin{pmatrix} z_{i1} - x'_{i1}\beta_1 \\ z_{i3}^* - x'_{i3}\beta_3^* \end{pmatrix}, \quad (11)$$

$$\sigma_{z_{i2}}^{*2} = 1 - \begin{pmatrix} \lambda_1 \\ \lambda_3 \end{pmatrix}' \begin{pmatrix} 1 & \lambda_2 \\ \lambda_2 & (\psi + \lambda_2^2 + \lambda_3^2 - 2\lambda_1\lambda_2\lambda_3) / (1 - \lambda_1^2) \end{pmatrix}^{-1} \begin{pmatrix} \lambda_1 \\ \lambda_3 \end{pmatrix}. \quad (12)$$

In the third equation, the conditional posterior of each latent variable is

$$z_{i3}^*|z_{i1}, z_{i2} \sim \begin{cases} \text{TN}(\mu_{z_{i3}}^*, \sigma_{z_{i3}}^{*2})|_{(\gamma_{j-1}^*, \gamma_j^*)}, & \text{if } y_{i3} = j, \\ N(\mu_{z_{i3}}^*, \sigma_{z_{i3}}^{*2}), & \text{if } y_{i3} = 0, \end{cases} \quad (13)$$

which is a univariate truncated normal distribution for $y_{i3} \neq 0$ and a normal distribution for $y_{i3} = 0$ with their means and variances given by

$$\mu_{z_{i3}}^* = x'_{i3}\beta_3^* + \begin{pmatrix} \lambda_2 \\ \lambda_3 \end{pmatrix}' \begin{pmatrix} 1 & \lambda_1 \\ \lambda_1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} z_{i1} - x'_{i1}\beta_1 \\ z_{i2} - x'_{i2}\beta_2 \end{pmatrix}, \quad (14)$$

$$\sigma_{z_{i3}}^{*2} = \frac{\psi + \lambda_2^2 + \lambda_3^2 - 2\lambda_1\lambda_2\lambda_3}{1 - \lambda_1^2} - \begin{pmatrix} \lambda_2 \\ \lambda_3 \end{pmatrix}' \begin{pmatrix} 1 & \lambda_1 \\ \lambda_1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \lambda_2 \\ \lambda_3 \end{pmatrix}. \quad (15)$$

3.3 Conditional posterior of each parameter

Let $Z^* = \{Z_i^* = (z_{i1}, z_{i2}, z_{i3})' : 1 \leq i \leq n\}$ and

$$X_i = \begin{pmatrix} x'_{i1} & 0 & 0 \\ 0 & x'_{i2} & 0 \\ 0 & 0 & x'_{i3} \end{pmatrix}.$$

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 conditional posterior of β^* is

$$\beta^*|Z^*, \Sigma^* \sim N(\hat{\beta}^*, B^{-1}), \quad (16)$$

where $\hat{\beta}^* = B^{-1} (B_0 \beta_0 + \sum_{i=1}^n X_i' \Omega^{*-1} Z_i^*)$ with $B = B_0 + \sum_{i=1}^n X_i' \Omega^{*-1} X_i$.

Assume that the prior of λ_1 denoted as $p(\lambda_1)$, is the density of $N(\lambda_{01}, C_1^{-1})$. The conditional posterior of λ_1 is (up to a normalizing constant)

$$p(\lambda_1 | Z^*, \beta^*, \lambda_2, \lambda_3, \psi) \propto p(\lambda_1) \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (Z_i^* - \mu_i^*)' \Omega^{*-1} (Z_i^* - \mu_i^*) \right\}, \quad (17)$$

from which we sample λ_1 using the random-walk Metropolis algorithm.

Assuming that the prior of λ_2 denoted as $p(\lambda_2)$, is the density of $N(\lambda_{02}, C_2^{-1})$, we derive the conditional posterior of λ_2 as (up to a normalizing constant)

$$p(\lambda_2 | Z^*, \beta^*, \lambda_1, \lambda_3, \psi) \propto p(\lambda_2) \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (Z_i^* - \mu_i^*)' \Omega^{*-1} (Z_i^* - \mu_i^*) \right\}.$$

It follows that

$$\lambda_2 | Z^*, \beta^*, \lambda_1, \lambda_3, \psi \sim N(\mu_{\lambda_2}, \sigma_{\lambda_2}^2), \quad (18)$$

where

$$\mu_{\lambda_2} = \sigma_{\lambda_2}^2 \left\{ \lambda_{02} C_2 + \sum_{i=1}^n \left[\frac{\lambda_3 (\mu_{i1} \lambda_1 - \mu_{i2})}{\psi (1 - \lambda_1^2)} + \frac{\mu_{i3}}{\psi} \right] (\mu_{i1} - \lambda_1 \mu_{i2}) \right\}, \quad (19)$$

$$\sigma_{\lambda_2}^2 = \left[C_2 + \frac{1}{\psi (1 - \lambda_1^2)} \sum_{i=1}^n (\mu_{i1} - \mu_{i2} \lambda_1)^2 \right]^{-1}. \quad (20)$$

The prior of λ_3 denoted as $p(\lambda_3)$, is assumed to be the density of $N(\lambda_{03}, C_3^{-1})$, and the conditional posterior of λ_3 is (up to a normalizing constant)

$$p(\lambda_3 | Z^*, \beta^*, \lambda_1, \lambda_2, \psi) \propto p(\lambda_3) \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (Z_i^* - \mu_i^*)' \Omega^{*-1} (Z_i^* - \mu_i^*) \right\}.$$

It turns out that

$$\lambda_3 | Z^*, \beta^*, \lambda_1, \lambda_2, \psi \sim N(\mu_{\lambda_3}, \sigma_{\lambda_3}^2), \quad (21)$$

where

$$\mu_{\lambda_3} = \sigma_{\lambda_3}^3 \left\{ \lambda_{03} C_3 + \sum_{i=1}^n \left[\frac{\lambda_2 (\mu_{i2} \lambda_1 - \mu_{i1})}{\psi (1 - \lambda_1^2)} + \frac{\mu_{i3}}{\psi} \right] (\mu_{i2} - \lambda_1 \mu_{i1}) \right\},$$

$$\sigma_{\lambda_3}^3 = \left[C_3 + \frac{1}{\psi (1 - \lambda_1^2)} \sum_{i=1}^n (\mu_{i2} - \mu_{i1} \lambda_1)^2 \right]^{-1}.$$

Therefore, λ_3 can be directly sampled from $N(\mu_{\lambda_3}, \sigma_{\lambda_3}^2)$.

We assume that the prior of ψ denoted as $p(\psi)$, is the inverse Gamma (IG) density denoted as $\text{IG}(a_0/2, b_0/2)$. Then the conditional posterior of ψ is (up to a normalizing constant)

$$\begin{aligned} p(\psi|Z^*, \beta^*, \lambda) &\propto p(\psi) \prod_{i=1}^n \phi_3(Z_i^*; \mu_i^*, \Omega^*) \\ &\propto \left(\frac{1}{\psi}\right)^{-a_0/2+1} \exp\left\{-\frac{b_0}{2\psi}\right\} \left(\frac{1}{\psi}\right)^{-n/2} \exp\left\{-\frac{1}{2\psi} \sum_{i=1}^n \frac{[\mu_{i1}(\lambda_2 - \lambda_1\lambda_3) + \mu_{i2}(\lambda_3 - \lambda_1\lambda_2) - \mu_{i3}(1 - \lambda_1^2)]^2}{1 - \lambda_1^2}\right\}. \end{aligned}$$

A little algebra shows that the conditional posterior of ψ is also an IG density:

$$\psi|Z^*, \beta^*, \lambda \sim \text{IG}(a_1/2, b_1/2), \quad (22)$$

where $a_1 = a_0 + n$ and $b_1 = b_0 + (1 - \lambda_1^2)^{-1} \sum_{i=1}^n [\mu_{i1}(\lambda_2 - \lambda_1\lambda_3) + \mu_{i2}(\lambda_3 - \lambda_1\lambda_2) - \mu_{i3}(1 - \lambda_1^2)]^2$.

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

$$p(\gamma_2^*, \dots, \gamma_{j-2}^* | \beta^*, \Omega^*, Z) \propto \prod_{i=1}^n \left\{ \Phi\left((\gamma_{y_{i3}}^* - \mu_{z_{i3}}^*)/\sigma_{z_{i3}}^*\right) - \Phi\left((\gamma_{y_{i3}-1}^* - \mu_{z_{i3}}^*)/\sigma_{z_{i3}}^*\right) \right\}, \quad (23)$$

where $\mu_{z_{i3}}^*$ and $\sigma_{z_{i3}}^*$ are given by (14) and (15), and $\Phi(\cdot)$ is the cumulative density function (CDF) of the standard normal distribution. We sample the threshold parameters from the conditional posterior given by (23) using the MH algorithm with a Dirichlet proposal density. This sampling strategy is described in details in [Nandram and Chen \(1996\)](#) and [Li and Tobias \(2006\)](#).

After the reparameterized parameter vector $\theta^* = (\beta^{*'}, \gamma^{*'}, \psi, \lambda_1, \lambda_2, \lambda_3)'$ 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 compare the performance of our proposed Bayesian estimation method with that of the FIML estimation method, where the sample size is $n = 1000$.

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. Therefore, $P_0 = \Phi(-x'_{i1}\beta_1)$. 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:

$$\begin{aligned} P_{11,1} &= \Phi_3(x'_{i1}\beta_1, x'_{i2}\beta_2, -x'_{i3}\beta_3; \rho_1, -\rho_2, -\rho_3), \\ P_{11,j} &= \Phi_3(x'_{i1}\beta_1, x'_{i2}\beta_2, \gamma_j - x'_{i3}\beta_3; \rho_1, -\rho_2, -\rho_3) \\ &\quad - \Phi_3(x'_{i1}\beta_1, x'_{i2}\beta_2, \gamma_{j-1} - x'_{i3}\beta_3; \rho_1, -\rho_2, -\rho_3), \text{ for } j = 2, \dots, J-1, \\ P_{11,J} &= \Phi_3(x'_{i1}\beta_1, x'_{i2}\beta_2, x'_{i3}\beta_3 - \gamma_{J-1}; \rho_1, \rho_2, \rho_3), \end{aligned} \quad (24)$$

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 maximize the likelihood given by

$$L(Y_1, Y_2, \dots, Y_n|\theta^*) = \prod_{i=1}^n \left\{ P_0^{(1-y_{i1})} P_{10}^{y_{i1}(1-y_{i2})} \left(\sum_{j=1}^J I(y_{i3} = j) P_{11,j} \right)^{y_{i1}y_{i2}} \right\}, \quad (25)$$

with respect to θ^* .

4.2 Monte Carlo design

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 medium strength of error 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. Due to the existence 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 the true values of (ρ_1, ρ_2, ρ_3) , 1,000 samples were generated from the above process, and both the FIML and Bayesian sampler were conducted to estimate parameters based on each generated sample.

For the proposed Bayesian sampler, the starting values are all zero, except $\gamma_2^* = 0.5$ and $\psi = 1$; the burn-in period contains the first 2,000 draws, and the following 10,000 draws are recorded; and the point estimate of each parameter is the arithmetic mean of 10,000 draws from the corresponding conditional posterior. The starting values are all zero for FIML, except $\gamma_2 = 1$ and $\gamma_3 = 2$ because these threshold values must be larger than zero.

4.3 Hyperparameter choices and convergence of the sampler

Hyperparameters of the priors for the Bayesian sampler are chosen as follows: $\beta_0 = \mathbf{0}$, $B_0^{-1} = 1000I_7$ where I_7 is the seven-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 $b_0 = 0.01$. Whenever the random-walk Metropolis algorithm was used, the acceptance rate was controlled to be between 0.2 and 0.3.

As the proposed sampling algorithm for the discrete-response model with double se-

lections are new, one might be interested in the mixing performance, or loosely speaking, the convergence of the sampler. In the literature on Bayesian sampling for binary- and discrete-response models, it is known that the simulated chains of the correlation parameters usually exhibit slow convergence. This 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 the 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 for every 50 draws, and thus, the retained draws are approximately independent. Usually, the smaller a SIF value is, the better the convergence of the simulated chain. 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. Generally speaking, the convergence of our sampler is acceptable.

4.4 Results

The simulation results 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 summarized based on the simulated samples resulting in meaningful results with the corresponding Hessian matrices invertible. Meanwhile, the proposed Bayesian sampler always produces meaningful results, and the reported statistics were summarized based on all 1,000 generated samples. For comparison

purposes, we also report the summarized results based on the generated samples, at which the reported FIML results were summarized. The summaries of such results are marked as “Bayesian*” and are compared to those marked by “FIML*”.

The FIML failed to converge in more than half of the simulated samples, and it was able to produce meaningful results in 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 results of the two estimation methods when the 1,000 samples were simulated with $(\rho_1, \rho_2, \rho_3) = (0, 0, 0)$. The summary measures for FIML were derived based on 420 simulated samples, for which the numerical optimization of FIML reached convergence. Both estimation methods produced similar results. First, 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.

Second, the standard deviation of the estimated values of ρ_1 is obviously larger than that of any other parameter. Moreover, such standard deviations of ρ_2 and ρ_3 are respectively, larger than those of the parameters in mean equations. Also, standard deviations of β_{21} and β_3 are twice as large as those of β_1 . This phenomenon is likely to be the consequence of information loss due to double selections.

Third, the mean absolute errors (MAEs) of β_{21} , β_3 and (ρ_1, ρ_2, ρ_3) are relatively larger than those of the other parameters. Once again, we tend to believe that this phenomenon is probably due to information loss caused by double selections.

Finally, the SIF values of the parameters in mean equations are quite small, indicating that the simulated chains of these parameters have achieved very reasonably convergence. The

largest mean SIF value across all simulated samples is 77 for ρ_1 , which indicates an acceptable convergence status.

When samples were generated under the situation of $(\rho_1, \rho_2, \rho_3) = (0.25, 0.25, 0.5)$, the simulation results are presented in Table 2. With the Bayesian method, the mean estimate for each parameter is quite close to the corresponding true value, except that for ρ_2 with its bias being 0.02. Meanwhile, based on the 453 simulated samples for which the FIML estimator works properly, FIML produced largely unbiased estimates for all parameters with the largest bias being 0.015.

In terms of the mean and standard deviation of the estimated values of each parameter, the two measures in the situation of medium level correlation among the three error terms are similar to those in situations of no correlation. The MAE for each parameter are almost similar to those in the situation of no correlation among error terms. The largest mean SIF value is 84 for ρ_1 .

Table 3 presents the simulation results derived through the two estimation methods, while the samples were simulated under the situation of $(\rho_1, \rho_2, \rho_3) = (0.5, 0.8, 0.7)$. With the Bayesian method, the mean estimate of ρ_1 is less than the true value by 0.048, while the mean estimates of the other parameters are all very close to the corresponding true values. FIML achieved convergence in only 458 out of 1,000 simulated samples. Based on the 458 simulated samples, the mean estimates of all parameters are very close to the corresponding true values. The other measures such as the standard deviation and MAE, obtained under this situation, are similar to those obtained under the situation of no correlation or medium level correlation. The mean SIF values of the parameters in mean equations are all small, indicating that the sampler achieved reasonable convergence. The largest mean SIF value is 92 for ρ_1 , while the mean SIF values for ρ_2 and ρ_3 are also larger than those under the situation of no correlation or medium level correlation. In our experience, the three SIF values indicate that the convergence status is not too bad and acceptable, considering the fact that

data containing information on such correlations could only be partly observed due to double selections.

4.5 Findings revealed from the simulation study

To conclude, the numerical optimization of FIML fails for more than half of the simulated samples because of the problem that the Hessian matrix cannot be inverted. In contrast, our proposed Bayesian sampling approach can always produce meaningful results even for those samples where FIML fails. Of the simulated samples where FIML works, on average, the estimate of each parameter is very close to the corresponding true value, and the variation measures of the estimated values derived across different simulated samples are reasonable.

The 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 the Bayesian sampling approach is that the simulated chains of the correlation coefficients exhibit slow convergence, which we think, is the consequence of information loss due to double selections. A practical remedy to the problem of slow convergence 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. Table 4 presents a summary of the mode statistic under each situation of the correlation setting, where the simulated samples are exactly the same as those used previously. Summarizing among the 1,000 simulated samples, we found that the mean of the mode statistics for each parameter is very close to its corresponding true value. The standard deviation of the mode statistics for each parameter is slightly larger than that of the corresponding mean statistics previously derived. On average, the mode statistic leads to a similar set of results as the mean statistic, while the former is more robust than the latter 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 Modelling the effect of mental health on labor outcomes

It is widely acknowledged that mental health is an important factor in determining labor market outcomes. Mental illness can affect not only individuals' chances to be 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 labor 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 the labor 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 labor 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 participants, where 6,928 individuals were employed.

In this study, we investigate the effects of various explanatory variables including mental health factors, on the probabilities of participation, employment and occupational skill levels. Double rules of sample selection exist because we could only observe occupational levels for individuals who participated in the labor market and then were employed. This data set was first analyzed by [Cornwell et al. \(2009\)](#), who seek to explain labor market outcomes by

mental health status. However, if there exist 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. It has been found that there exists an obvious effect of mental illness on employment and occupational skill level.

We used the following specification of exogenous regressors. These include mental illness indicators, categorized into substance use disorder, anxiety disorder and affective disorder. They are all binary. The other regressors include age, gender, education, geographic location indicators, and a socio-economic 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, but they are not included in the employment equation. The purpose of such an exclusion restriction is to make estimation easier than it would be otherwise by providing identifying variables. However, in the second hurdle, the factors 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 \(2002, p.E21-115\)](#) mentioned that the conventional rules for identification in simultaneous-equation models do not apply in ‘treatment effects’ models. Because of the nonlinearity of the conditional mean function, it is not necessary to exclude some variables from any equation.

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. In contrast, our proposed Bayesian method is

able to estimate not only the parameters in the three mean equations, but also the correlation parameter between any pair of the three error terms. The estimates of some parameters are presented in Table 5. Importantly, our Bayesian method can also facilitate the computation of point and interval estimates of marginal effects of any explanatory variable.

The hyperparameters of the priors were chosen as follows: $\beta_0 = \mathbf{0}$, $B_0^{-1} = 1000I_{65}$ 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 either the mean or mode of each simulated chain. The SIF was used to monitor the convergence status of each simulated chain. All the simulated chains of regression parameters and threshold parameters have achieved very reasonable convergence, while the simulated chains of ρ_1 and ρ_3 produced large SIF values. Therefore, each of the three correlation parameters was estimated by the mode of the corresponding simulated chain.

The estimated parameters, their 95% Bayesian credible intervals and the corresponding SIF values for the first and second equations are reported in the left panels of Table 5 and Table 6, respectively. For the third equation, the estimated parameters and their associated SIF values are reported in the left panel of Table 7, and the corresponding 95% Bayesian credible intervals are given in the second column of Table 8. The estimated correlation parameters are respectively, 0.0958 (97), 0.3268 (8) and 0.4021 (49) based on the mean of each simulated chain, where the associated SIF values are given in parentheses. They suggest some sizeable correlations between unobservables across equations. With the posterior mode, the three correlation parameters are estimated as 0.1248, 0.3273 and 0.4050.

5.3 Marginal effects

The proposed Bayesian sampling approach also makes a contribution to the computation of the marginal effect of each explanatory variable on the probability of a certain response value

that is of interest. The probability of participating the labor force is formulated as

$$Pr \{y_{i1} = 1|x_{i1}\} = \Phi(x'_{i1}\beta_1). \quad (26)$$

The probability of being employed conditional on participation is described by

$$Pr \{y_{i2} = 1|y_{i1} = 1, x_{i1}, x_{i2}\} = \frac{\Phi_2(x'_{i1}\beta_1, x'_{i2}\beta_2, \rho_1)}{\Phi(x'_{i1}\beta_1)}. \quad (27)$$

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

$$Pr \{y_{i3} = j|y_{i1} = 1, y_{i2} = 1, x_{i1}, x_{i2}, x_{i3}\} = \frac{P_{11,j}}{\Phi_2(x'_{i1}\beta_1, x'_{i2}\beta_2, \rho_1)}, \quad (28)$$

for $j = 1, 2, \dots, 5$, where $P_{11,j}$ is given by (24). 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 (26)–(28).

First, we studied the marginal effect of each explanatory variable on the probability of participation, which is modelled by the first equation. The right-hand-side panel of Table 5 presents the point estimates and 95% Bayesian credible intervals of the marginal effects of all the regressors in the first equation. Individuals aged between 25 and 44 are most likely to participate in the labor force, while those aged between 45 and 64 are 28.9% more likely to participate than those aged between 18 and 24. Males have a larger probability to participate than females. People with higher levels of education are more likely to participate than those with less education.

Coming from a regional center has no obvious effect on participation, because the 95% Bayesian credible interval of its marginal effect covers zero. People from a rural area are 4.7% more likely to participate than those from an urban area. The participation rates in more socio-economically advanced areas are generally higher than those in less advanced areas, except the area in the 8th decile that has the highest participation rate. An increase in the number of children would reduce the chance of an individual participating in the labor force.

People who are currently studying are more likely to seek jobs than those who are not studying. 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 9.5% more likely to participate than those without this type of disorder.

Second, we studied the marginal effect of each explanatory variable in the second equation on the probability of being employed conditional on participation in the labor force given by (27). The parameter estimates and the corresponding values of marginal effects are presented in Table 6. Individuals in the age groups 25–44 and 45–64 are equally likely to be employed, and both groups have higher employment rates than the age group 18–24. Being a male would not obviously increase or decrease his chance of being employed, because the corresponding 95% Bayesian credible interval contains zero. Thus, gender has little impact on employment. People with a secondary school education have the same opportunity to be employed as those with a vocational qualification. At the same time, those who have not completed secondary school are least likely to be employed, while higher education would be most likely to increase the possibility of being employment.

Whether an individual is from a regional center has no obvious effect on the probability of being employed. People in rural areas are 1.2% more likely to be employed than those in urban areas. In terms of socio-economic indices, more advanced areas have higher employment rates than less advanced areas. Although physical illness has no effect on the probability of being employed, the three types of mental disorders would all reduce the possibility of being employed. For example, individuals with substance use disorders would be 5.2% less likely to be employed than those without this type of disorders.

Last, we studied the marginal effect of each explanatory variable in the third equation on the probability of working in each occupational skill category conditional on employment

based on the formulae given by (28). Table 7 presents the estimates of parameters and their corresponding marginal effects, while Table 8 presents the corresponding 95% Bayesian credible intervals. Older people are more likely to be employed in higher levels of skill categories and less likely to be employed in elementary and intermediate levels of skill categories. Males are more likely to get a job as associate professionals and professionals than females, and less likely to be employed in lower-level occupational skill categories. With a marginal effect of 52.4%, people with a tertiary education are much more likely to be employed as professionals, while they are least likely to be employed in lower-level skill categories.

Being based in a regional center has no obvious effect on the levels of occupational skill categories. Individuals from rural areas are more likely to be employed at higher skill levels, and less likely to be employed at lower skill levels than those from urban areas. In terms of the SEIFA indices, people in more advanced areas usually have a high opportunity to be employed as associate professionals and professionals, and a low opportunity to work in the other three categories than people in less advanced areas. Even though the estimated coefficient of physical illness is negative in the third equation, its marginal effects on four different levels of occupational skill category are not obvious, with the exception of the advanced skill category. As a result, physical illness has little impact on levels of occupational choice. The estimated coefficient of anxiety disorders is negative, but its marginal effects on most skill levels are not obvious, except the marginal effect on the intermediate skill category. The marginal effect of affective disorders suggests no clear impact on the level of skill category. However, substance use disorders would reduce the probability of being employed in higher occupational skill levels and increase the chance of being employed in elementary and intermediate 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 presented a new reparam-

eterization strategy to facilitate the derivation of conditional posteriors. A benefit of the proposed Bayesian method is that it allows for specification and estimation of the correlation coefficients between any pair of the three error terms, while this cannot be achieved by the conventional two-step estimation discussed by Heckman (1979). 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 for achieving convergence.

We employed the three-equation model with double selection rules to model people's participation in the labor 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 allows us to derive the 95% Bayesian credible interval for the marginal effect of any explanatory variable on its corresponding response. Consequently, we can evaluate whether such a marginal effect is obvious or not. The results show that although the mental illness of anxiety and affective disorders have no obvious impact on participation, they obviously reduce the chance of being employed. However, conditional on being employed, the two types of mental illness have no obvious effect on the levels of occupational skill. In terms of substance use disorders, individuals with this type of illness are more likely to participate in the labor force, but they are less likely to be employed. Moreover, even if they are employed, this type of illness would

reduce their chance to be employed in higher levels of occupational skill categories.

Acknowledgements

We extend our sincere thanks to Murray Smith and Rodney Strachan for their very insightful comments on an earlier version of this paper, which was a main chapter of the first author's PhD thesis. Thanks also go to Monash Sun Grid for its quality computing facilities.

References

- Amemiya, T., 1985. *Advanced Econometrics*, *Harvard University Press, New York*.
- Blundell, R., Ham, J., Meghir, C., 1987. Unemployment and female labour supply. *The Economic Journal* 97, 44–64.
- Cameron, A. C., Trivedi, P. K., 2005. *Microeconometrics: Methods and Applications*, *Cambridge University Press, London*.
- Chib, S., Greenberg, E., Jeliazkov, I., 2009. Estimation of semiparametric models in the presence of endogeneity and sample selection. *Journal of Computational and Graphical Statistics* 18 (2), 321–348.
- Cornwell, K., Forbes, C., Inder, B., Meadows, G., 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, 101–111.
- Greene, W. H., 2002. *LIMDEP Version 8.0: Econometric Modeling Guide*, *Econometric Software, New York*.
- Heckman, J. J., 1979. Sample selection bias as a specification error. *Econometrica* 47, 153–161.

- Henneberger, F., Sousa-Poza, A., 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., Shepherd, N., Chib, S., 1998. Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65 (3), 361–393.
- Li, M., Tobias, J. L., 2006. Bayesian analysis of structural effects in an ordered equation system. *Studies in Nonlinear Dynamics and Econometrics* 10, 1–24.
- Maddala, G. S., 1983. *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, London.
- McCulloch, R. E., Polson, N. G., Rossi, P. E., 2000. A Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics* 99, 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., Chen, M. H., 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: Gilks, W. R., Richardson, S., Spiegelhalter, D. J. (Eds.), *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London, pp. 45–57.
- 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., Brooks, R. D., King, M. L., 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 | -1.001 | -0.398 | 1.484 | 0.784 | 1.570 | 0.008 | -0.004 | 0.000 |
| Bayesian* | 0.608 | -1.211 | 0.998 | -1.504 | -0.991 | -0.399 | 1.483 | 0.777 | 1.570 | 0.004 | 0.003 | -0.003 |
| 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 | 0.054 | 0.073 | 0.149 | 0.109 | 0.149 | 0.141 | 0.143 | 0.074 | 0.095 | 0.176 | 0.165 | 0.140 |
| Bayesian* | 0.054 | 0.076 | 0.148 | 0.107 | 0.148 | 0.142 | 0.139 | 0.074 | 0.096 | 0.183 | 0.168 | 0.144 |
| Mean absolute error | | | | | | | | | | | | |
| FIML* | 0.043 | 0.060 | 0.119 | 0.085 | 0.118 | 0.112 | 0.112 | 0.061 | 0.076 | 0.159 | 0.137 | 0.120 |
| Bayesian | 0.043 | 0.058 | 0.117 | 0.086 | 0.118 | 0.110 | 0.114 | 0.061 | 0.079 | 0.140 | 0.134 | 0.113 |
| Bayesian* | 0.044 | 0.060 | 0.117 | 0.085 | 0.117 | 0.110 | 0.111 | 0.063 | 0.080 | 0.150 | 0.140 | 0.119 |
| SIF | | | | | | | | | | | | |
| Mean | 7 | 12 | 26 | 25 | 12 | 24 | 7 | 10 | 16 | 77 | 48 | 40 |
| Standard deviation | 2 | 3 | 5 | 5 | 3 | 4 | 2 | 2 | 3 | 6 | 7 | 5 |

Note: The symbol * indicates that the corresponding summaries are based on 420 simulated samples, for which FIML achieved convergence.

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 | 1.015 | -1.514 | -1.004 | -0.397 | 1.486 | 0.787 | 1.578 | 0.233 | 0.230 | 0.500 |
| Bayesian* | 0.603 | -1.209 | 1.013 | -1.515 | -1.007 | -0.394 | 1.496 | 0.790 | 1.583 | 0.237 | 0.234 | 0.487 |
| 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 | 0.159 | 0.111 | 0.154 | 0.126 | 0.133 | 0.081 | 0.100 | 0.172 | 0.158 | 0.113 |
| Bayesian* | 0.052 | 0.072 | 0.152 | 0.106 | 0.149 | 0.123 | 0.129 | 0.083 | 0.100 | 0.161 | 0.156 | 0.114 |
| Mean absolute error | | | | | | | | | | | | |
| FIML* | 0.043 | 0.059 | 0.121 | 0.084 | 0.117 | 0.100 | 0.103 | 0.068 | 0.079 | 0.130 | 0.120 | 0.089 |
| Bayesian | 0.043 | 0.058 | 0.125 | 0.087 | 0.121 | 0.101 | 0.106 | 0.066 | 0.081 | 0.137 | 0.128 | 0.091 |
| Bayesian* | 0.043 | 0.059 | 0.120 | 0.084 | 0.117 | 0.099 | 0.102 | 0.067 | 0.079 | 0.127 | 0.123 | 0.091 |
| SIF | | | | | | | | | | | | |
| Mean | 7 | 14 | 34 | 34 | 18 | 28 | 11 | 16 | 24 | 84 | 58 | 50 |
| Standard deviation | 2 | 4 | 7 | 8 | 6 | 5 | 4 | 4 | 6 | 5 | 8 | 8 |

Note: The symbol * indicates that the corresponding summaries are based on 453 simulated samples, for which FIML achieved convergence.

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 | -1.206 | 1.038 | -1.521 | -1.014 | -0.405 | 1.493 | 0.791 | 1.583 | 0.452 | 0.790 | 0.689 |
| Bayesian* | 0.605 | -1.204 | 1.040 | -1.520 | -1.015 | -0.403 | 1.499 | 0.796 | 1.589 | 0.455 | 0.788 | 0.689 |
| 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 | 0.161 | 0.113 | 0.149 | 0.104 | 0.124 | 0.085 | 0.112 | 0.154 | 0.081 | 0.089 |
| Bayesian* | 0.053 | 0.071 | 0.162 | 0.112 | 0.148 | 0.110 | 0.123 | 0.089 | 0.116 | 0.154 | 0.082 | 0.089 |
| Mean absolute error | | | | | | | | | | | | |
| FIML* | 0.042 | 0.055 | 0.124 | 0.089 | 0.114 | 0.087 | 0.099 | 0.073 | 0.093 | 0.115 | 0.062 | 0.067 |
| Bayesian | 0.042 | 0.057 | 0.131 | 0.091 | 0.118 | 0.083 | 0.099 | 0.069 | 0.092 | 0.125 | 0.064 | 0.070 |
| Bayesian* | 0.042 | 0.056 | 0.130 | 0.090 | 0.116 | 0.088 | 0.098 | 0.072 | 0.094 | 0.123 | 0.065 | 0.070 |
| SIF | | | | | | | | | | | | |
| Mean | 19 | 39 | 44 | 52 | 30 | 40 | 35 | 42 | 59 | 92 | 75 | 70 |
| Standard deviation | 8 | 15 | 11 | 13 | 12 | 10 | 13 | 11 | 13 | 4 | 8 | 10 |

Note: The symbol * indicates that the corresponding summaries are based on 458 simulated samples, for which FIML achieved convergence.

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

| | No correlation | | | Medium correlation | | | High 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 | 0.009 | -0.005 | -0.002 | 0.262 | 0.243 | 0.521 | 0.448 | 0.819 | 0.712 |
| Standard deviation | 0.220 | 0.180 | 0.149 | 0.225 | 0.173 | 0.119 | 0.204 | 0.083 | 0.091 |
| Mean absolute error | 0.174 | 0.145 | 0.119 | 0.178 | 0.138 | 0.098 | 0.164 | 0.070 | 0.074 |

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 | 1.260 | 4 | (1.199, 1.320) | 0.409 | (0.391, 0.426) |
| Age 45-64 | 0.934 | 3 | (0.872, 0.996) | 0.289 | (0.272, 0.305) |
| Male | 0.062 | 4 | (-0.001, 0.126) | 0.022 | (0.000, 0.044) |
| Secondary school | 0.217 | 3 | (0.164, 0.270) | 0.076 | (0.057, 0.094) |
| Higher education | 0.390 | 4 | (0.309, 0.473) | 0.127 | (0.103, 0.151) |
| Vocational education | 0.170 | 3 | (0.087, 0.254) | 0.058 | (0.030, 0.085) |
| From a regional center | 0.014 | 3 | (-0.053, 0.082) | 0.005 | (-0.019, 0.029) |
| From a rural area | 0.136 | 3 | (0.072, 0.199) | 0.047 | (0.025, 0.068) |
| SEIFA: 2nd decile | 0.203 | 3 | (0.098, 0.307) | 0.068 | (0.034, 0.101) |
| SEIFA: 3rd decile | 0.279 | 3 | (0.175, 0.382) | 0.092 | (0.060, 0.123) |
| SEIFA: 4th decile | 0.285 | 3 | (0.179, 0.391) | 0.094 | (0.061, 0.126) |
| SEIFA: 5th decile | 0.358 | 3 | (0.254, 0.463) | 0.116 | (0.085, 0.146) |
| SEIFA: 6th decile | 0.390 | 3 | (0.282, 0.499) | 0.125 | (0.093, 0.155) |
| SEIFA: 7th decile | 0.388 | 3 | (0.286, 0.491) | 0.125 | (0.095, 0.154) |
| SEIFA: 8th decile | 0.487 | 3 | (0.380, 0.593) | 0.152 | (0.123, 0.180) |
| SEIFA: 9th & 10th deciles | 0.448 | 3 | (0.357, 0.538) | 0.147 | (0.120, 0.174) |
| Number of children | -0.250 | 4 | (-0.271, -0.229) | -0.088 | (-0.096, -0.081) |
| Currently studying | 0.434 | 2 | (0.051, 0.823) | 0.128 | (0.018, 0.217) |
| Physical illness | -0.511 | 3 | (-0.560, -0.462) | -0.183 | (-0.201, -0.166) |
| Anxiety disorder | -0.079 | 3 | (-0.171, 0.013) | -0.028 | (-0.062, 0.005) |
| Affective disorder | 0.011 | 3 | (-0.087, 0.110) | 0.004 | (-0.031, 0.038) |
| Substance use disorder | 0.290 | 4 | (0.197, 0.383) | 0.095 | (0.067, 0.123) |

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 | 0.347 | 66 | (0.178, 0.513) | 0.032 | (0.019, 0.045) |
| Age 45-64 | 0.369 | 46 | (0.219, 0.518) | 0.034 | (0.022, 0.046) |
| Male | 0.004 | 38 | (-0.094, 0.103) | 0.000 | (-0.011, 0.011) |
| Secondary school | 0.232 | 23 | (0.135, 0.329) | 0.025 | (0.015, 0.036) |
| Higher education | 0.515 | 24 | (0.362, 0.673) | 0.044 | (0.033, 0.054) |
| Vocational education | 0.262 | 15 | (0.112, 0.415) | 0.025 | (0.011, 0.037) |
| From a regional center | -0.080 | 11 | (-0.190, 0.031) | -0.010 | (-0.025, 0.004) |
| From a rural area | 0.120 | 15 | (0.009, 0.231) | 0.012 | (0.000, 0.023) |
| SEIFA: 2nd decile | 0.159 | 13 | (-0.010, 0.329) | 0.015 | (-0.002, 0.030) |
| SEIFA: 3rd decile | 0.300 | 14 | (0.126, 0.475) | 0.027 | (0.013, 0.040) |
| SEIFA: 4th decile | 0.348 | 15 | (0.168, 0.528) | 0.031 | (0.017, 0.043) |
| SEIFA: 5th decile | 0.358 | 16 | (0.182, 0.535) | 0.032 | (0.018, 0.044) |
| SEIFA: 6th decile | 0.582 | 18 | (0.389, 0.780) | 0.045 | (0.034, 0.055) |
| SEIFA: 7th decile | 0.431 | 17 | (0.254, 0.609) | 0.037 | (0.024, 0.048) |
| SEIFA: 8th decile | 0.368 | 18 | (0.189, 0.546) | 0.032 | (0.018, 0.044) |
| SEIFA: 9th & 10th deciles | 0.522 | 21 | (0.362, 0.679) | 0.048 | (0.035, 0.060) |
| Physical illness | -0.025 | 47 | (-0.137, 0.087) | 0.000 | (-0.010, 0.011) |
| Anxiety disorder | -0.151 | 10 | (-0.297, -0.004) | -0.020 | (-0.042, 0.000) |
| Affective disorder | -0.240 | 9 | (-0.386, -0.092) | -0.034 | (-0.059, -0.012) |
| Substance use disorder | -0.334 | 16 | (-0.457, -0.210) | -0.052 | (-0.073, -0.032) |

Table 7: Parameter estimates and marginal effects on the probability of being employed differen 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 | 0.331 | 2 | -0.020 | -0.021 | -0.005 | 0.009 | 0.037 |
| Age 45-64 | 0.313 | 2 | -0.030 | -0.027 | -0.005 | 0.014 | 0.049 |
| Male | 0.309 | 2 | -0.072 | -0.047 | 0.000 | 0.034 | 0.086 |
| Secondary school | 0.219 | 2 | -0.039 | -0.027 | -0.001 | 0.018 | 0.049 |
| Higher education | 1.543 | 2 | -0.200 | -0.228 | -0.105 | 0.009 | 0.524 |
| Vocational education | 0.791 | 1 | -0.129 | -0.132 | -0.041 | 0.041 | 0.261 |
| From a regional center | 0.032 | 2 | -0.009 | -0.006 | 0.000 | 0.004 | 0.010 |
| From a rural area | 0.270 | 2 | -0.053 | -0.041 | -0.004 | 0.025 | 0.073 |
| SEIFA: 2nd decile | 0.157 | 2 | -0.025 | -0.020 | -0.002 | 0.012 | 0.035 |
| SEIFA: 3rd decile | 0.132 | 2 | -0.014 | -0.013 | -0.002 | 0.006 | 0.022 |
| SEIFA: 4th decile | 0.181 | 2 | -0.024 | -0.020 | -0.003 | 0.011 | 0.036 |
| SEIFA: 5th decile | 0.252 | 2 | -0.037 | -0.030 | -0.005 | 0.017 | 0.055 |
| SEIFA: 6th decile | 0.225 | 2 | -0.027 | -0.024 | -0.005 | 0.012 | 0.043 |
| SEIFA: 7th decile | 0.308 | 2 | -0.046 | -0.039 | -0.006 | 0.021 | 0.071 |
| SEIFA: 8th decile | 0.338 | 2 | -0.050 | -0.042 | -0.007 | 0.022 | 0.078 |
| SEIFA: 9th & 10th deciles | 0.398 | 2 | -0.062 | -0.051 | -0.007 | 0.028 | 0.093 |
| Physical illness | -0.062 | 2 | -0.008 | -0.003 | 0.002 | 0.004 | 0.006 |
| Anxiety disorder | -0.093 | 2 | 0.016 | 0.011 | 0.000 | -0.008 | -0.019 |
| Affective disorder | -0.057 | 2 | 0.008 | 0.006 | 0.001 | -0.004 | -0.011 |
| Substance use disorder | -0.084 | 2 | 0.024 | 0.015 | -0.001 | -0.012 | -0.027 |

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

| Variable | Credible interval of coefficient | Credible interval of marginal effect | | | | |
|---------------------------|----------------------------------|--------------------------------------|--------------------|-----------------|-------------------------|-----------------|
| | | Elementary skill | Intermediate skill | Advanced skill | Associate professionals | Professionals |
| Age 25-44 | (0.283, 0.380) | (-0.033,-0.008) | (-0.029,-0.013) | (-0.006,-0.004) | (0.003, 0.015) | (0.023, 0.051) |
| Age 45-64 | (0.262, 0.364) | (-0.042,-0.018) | (-0.035,-0.019) | (-0.006,-0.004) | (0.008, 0.019) | (0.034, 0.064) |
| Male | (0.271, 0.346) | (-0.082,-0.063) | (-0.054,-0.041) | (-0.002, 0.001) | (0.030, 0.039) | (0.075, 0.097) |
| Secondary school | (0.178, 0.261) | (-0.049,-0.029) | (-0.034,-0.021) | (-0.003, 0.000) | (0.014, 0.023) | (0.038, 0.061) |
| Higher education | (1.481, 1.604) | (-0.208,-0.192) | (-0.237,-0.218) | (-0.113,-0.096) | (0.000, 0.017) | (0.502, 0.545) |
| Vocational Education | (0.730, 0.853) | (-0.137,-0.121) | (-0.143,-0.121) | (-0.048,-0.035) | (0.037, 0.046) | (0.237, 0.284) |
| From a regional center | (-0.023, 0.087) | (-0.022, 0.004) | (-0.014, 0.003) | (0.000, 0.001) | (-0.002, 0.011) | (-0.005, 0.026) |
| From a rural area | (0.220, 0.320) | (-0.064,-0.043) | (-0.049,-0.032) | (-0.006,-0.002) | (0.020, 0.030) | (0.058, 0.089) |
| SEIFA: 2nd decile | (0.067, 0.246) | (-0.044,-0.005) | (-0.034,-0.005) | (-0.004,-0.001) | (0.002, 0.020) | (0.009, 0.062) |
| SEIFA: 3rd decile | (0.043, 0.220) | (-0.033, 0.007) | (-0.027, 0.001) | (-0.004,-0.001) | (-0.003, 0.015) | (-0.003, 0.048) |
| SEIFA: 4th decile | (0.091, 0.271) | (-0.043,-0.004) | (-0.035,-0.006) | (-0.005,-0.002) | (0.002, 0.020) | (0.010, 0.063) |
| SEIFA: 5th decile | (0.164, 0.341) | (-0.055,-0.018) | (-0.046,-0.016) | (-0.008,-0.002) | (0.008, 0.024) | (0.029, 0.083) |
| SEIFA: 6th decile | (0.136, 0.315) | (-0.045,-0.007) | (-0.039,-0.010) | (-0.007,-0.003) | (0.003, 0.020) | (0.017, 0.070) |
| SEIFA: 7th decile | (0.222, 0.394) | (-0.063,-0.029) | (-0.054,-0.025) | (-0.010,-0.004) | (0.013, 0.028) | (0.044, 0.098) |
| SEIFA: 8th decile | (0.250, 0.426) | (-0.067,-0.033) | (-0.058,-0.028) | (-0.011,-0.004) | (0.015, 0.029) | (0.050, 0.106) |
| SEIFA: 9th & 10th deciles | (0.322, 0.474) | (-0.078,-0.047) | (-0.064,-0.038) | (-0.010,-0.005) | (0.021, 0.035) | (0.069, 0.117) |
| Physical illness | (-0.102,-0.021) | (-0.018, 0.002) | (-0.009, 0.003) | (0.001, 0.002) | (-0.001, 0.009) | (-0.006, 0.017) |
| Anxiety disorder | (-0.169,-0.018) | (-0.003, 0.036) | (0.000, 0.021) | (-0.002, 0.001) | (-0.017, 0.001) | (-0.038, 0.001) |
| Affective disorder | (-0.137, 0.023) | (-0.012, 0.028) | (-0.006, 0.018) | (-0.001, 0.002) | (-0.014, 0.006) | (-0.032, 0.011) |
| Substance use disorder | (-0.154,-0.015) | (0.006, 0.042) | (0.005, 0.024) | (-0.002, 0.001) | (-0.021,-0.003) | (-0.044,-0.009) |