Bayesian Analysis of Non-linear Multivariate Econometric Models

A thesis submitted for the degree of

Doctor of Philosophy

by

Rong Zhang

Department of Econometrics and Business Statistics

Monash University

Australia

January 2011

Contents

List of Tables

List of Figures

Abstract

This thesis aims to investigate Bayesian sampling techniques for estimating parameters of three nonlinear models with different levels of endogeneity and sample selection. These models include a bivariate probit model with an endogenous dummy regressor, an ordered probit model with sample selection, and an ordered probit model with double hurdles of sample selection. We developed Bayesian sampling algorithms to sample parameters in each of these models, and the resulting posterior estimates of parameters were compared with those obtained through a few classical estimation methods such as maximum likelihood estimate (MLE) and a two-step method. Monte Carlo simulations were conducted to check the performance of different estimators for each model.

In the bivariate probit model with an endogenous dummy regressor, we discussed the identification conditions especially the effect of exclusion restrictions. The Monte Carlo study reveals that exclusion restrictions are not essential for model identification. However, the existence of exclusion restrictions will make the estimation much easier for all estimators. Moreover, model identification can be improved by increasing the variation of explanatory variables and the number of exogenous regressors. In terms of the performance of the three estimators, MLE is often accurate and efficient except for occasional convergence failures. The Bayesian method can always produce an estimate for each simulated sample and is most efficient. However, it shows same small bias when the correlation coefficient between errors is large. The inconsistent two-step method has less convergence problems than MLE, but has quite large biases when the correlation coefficient between errors is large.

In terms of the ordered probit model with binary selection, we used a reparameterization to derive a Gibbs sampler, such that conditional posteriors can be obtained. We also

propose a likelihood-based two-step method in a way similar to the derivation of the concentrated likelihood function. The two proposed methods were compared with the full information maximum likelihood (FIML) method and another established two-step method. Monte Carlo results show that the Bayesian method and the likelihood-based two-step method can be alternative methods to FIML, while the other two-step method is not acceptable in models with large error correlation. The absence of exclusion restrictions does not cause big problems for the model estimation. With the FIML and the Bayesian methods, we used the ordered probit model with binary selectivity to model the effect of mental illness on employment and job categories, where exclusion restrictions do not exist.

The ordered probit model with double-hurdle selection is an extension of the above model with one additional level of sample selection. We found that FIML has encountered severe convergence-failure problems as the model becomes more complicated. As such, the proposed Bayesian sampling method is of great value because it always produces an estimate of the parameter vector. We propose two Bayesian samplers, one obtained through a standard process currently available in the literature, while the other involved reparameterization. In the Monte Carlo study, we found that both samplers and the FIML provide unbiased and efficient estimates. However, FIML fails to converge for more than half of the simulated samples, while Bayesian samplers can always produce estimates for each simulated sample. The reparameterization-based sampler shows better convergence than the other sampler. We applied the three estimators to the estimation of the double-hurdle ordered probit model investigating the effect of mental illness on labor market outcomes. We found that reparameterization-based sampler is the only estimator that did not encounter convergence problems. The resulting estimates of parameters were used for analyzing marginal effects of mental health variables.

Declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma in any university or equivalent institution, and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Rong Zhang

Acknowledgements

First of all, I would like to give my big thanks to my main supervisor, Professor Brett Inder for his greatest support during my PhD study. Not only has he introduced me into the micro-econometric world, he has also shared his brilliant ideas with me without any reservation. Also, I thank him for his patience in explaining the economic theories which have helped me to make sense and to better understand my own research, and all of these have greatly broadened my mind. Above all, without his encouragement and crucial pieces of advice given from time to time, I do not think that I could have gone through so much difficult and stressful times during the whole duration of a long, long study.

I would also like to thank my associate supervisor Dr. Xibin Zhang for his extraordinary support in the past four years helping with solving technical issues of my research study. In fact, he is the one who offered me the opportunity to study at Monash University, and the one who led me into the Bayesian world by teaching me how to use those complicated MCMC algorithms. I actually feel like both supervisors have placed me under their wings and trained me as a tutor or as a research assistant so that I could gain practical experience from these different works.

I am especially grateful to Xiaohui Zhang and Nan Qu who are always available for discussions and giving me their insightful suggestions. My big thanks should also go to Dr. Katy Cornwell, she is the one who not only provided me with valuable empirical data sets but also spent a lot of time explaining the variables in the data to me.

In addition to those who have helped me make progress academically, I would like to thank those who have provided me with a strong support network too. Special thanks go to all my colleagues and friends, especially Bing Li and Lachlan Macquarie who have been accompany me in the same office for nearly four years. I also owe many thanks to administration staff in our department such as Liz Moon, Elke Blaedel and Anna Vanderbom, you guys will be remembered in the years to come!

Of course, I must express my deepest gratitude to my parents for their unconditional love, support and belief in me. No words are enough to thank my husband Xiaoming Chen for being such a powerful companion to keep me away from computer games and internet surfing, so I can finally finish my research :).

Chapter 1

Introduction

1.1 Motivation

Econometric models where the sample is non-randomly selected or where certain regressors are endogenous have been widely discussed in microeconometrics. Most models are the extensions of two classical models. One is the initial sample selection model (Heckman 1976) in which a linear model is partly observed after selection via a binary indicator. The other is the model with endogenous treatment effects (Heckman 1978) where we are interested in the treatment effects captured by a binary variable and the dummy is an endogenous explanatory variable in a linear equation. In both cases, the main equations of interest contain continuous outcomes, while either selectivity or endogeneity arises because the latent variable driving the binary outcomes is correlated with the variable producing the continuous outcomes of interest. The two models are applied a great deal in practice to various circumstances, with some variations from the original model specifications. As a result, an increasing number of studies have extended the initial models to more complicated models with nonlinear forms, especially to multivariate models with probit or ordered probit main equations under situations of sample selection or endogeneity.

Among those nonlinear models, our main interests in this thesis are three types of models, a bivariate probit model with an endogenous dummy regressor, a sample selection model with ordered probit outcomes, and a double-hurdle model with ordered outputs determined by a partly observed selection equation. Some research has already been done on the bivariate probit model with an endogenous dummy regressor. There are controversial arguments about the effects of exclusion restrictions on the identification conditions of endogenous models. Most empirical researchers believe that exclusion restrictions are necessary components to guarantee model identification, while others argue that models without exclusion restrictions can be identified by the nonlinearity of the conditional mean function (Greene 2002). Thus, we intend to use different estimators and to test whether exclusion restrictions are important. In addition, an ordered variable with sample selection can be frequently encountered in empirical work. Similarly, a double-hurdle model, which also fits certain specifications in empirical research, can be treated as an extension of the ordered model with selectivity. Available estimation methods may have no problem estimating a two-equation system, but some are not appropriate for estimating multivariate models especially with sample selection. This motivates our pursuit in this thesis of some alternative estimation methods for multivariate models.

There are well established techniques that can correct for sample selection bias or inconsistency caused by endogenous variables. The two-step method proposed by Heckman (1979) is the most commonly used estimation method to correct for sample selection bias and is consistent for standard linear models. Another consistent estimator is the full information maximum likelihood (FIML) estimator contributed by Greene (2006). For those nonlinear models in which main equations have probit or ordered probit forms, however, the two-step method is no longer consistent and may lead to misleading estimations, while FIML has some convergence problems, a problem that gets worse as the number of equations increases. In this thesis, we will develop estimators for such bivariate and trivariate nonlinear models with sample selection and endogeneity in the Bayesian framework, in the hope these can deal with some of the problems with two-step and FIML estimators.

Bayesian analysis has become popular in recent years. It was first criticized for its computational complexity, but is booming due to the increase of the computational capacity of computers and software. The Bayesian method is quite different from classic estimation methods in how it provides a way to estimate the potential latent variables, so one may have a clearer understanding of the model structure. Basically, Markov chain Monte Carlo (MCMC) algorithms are applied to mimic the joint posterior distributions of model parameters and all latent variables. Furthermore, the convergent draws of the model parameters from the posterior distributions can be used to evaluate other useful distributions like marginal effects, which cannot be conducted by classic methods. The other issue that Bayesian estimation can address is the frequent failure of FIML algorithms to converge to valid estimates. So there are several reasons for using the Bayesian method in this thesis, with derivation of the Bayesian samplers and comparison to classical methods for each specific model.

1.2 Outline of the Thesis

Chapter 2 reviews some relevant literature the development of models with sample selection and endogeneity, together with some estimation methods. Both the classic sample selection model and the endogenous treatment effect model contain a continuous main equation, with a binary indicator deciding the mechanism of sample selection or endogeneity. Both models have similar properties and share the same estimation methods like the two-step method and FIML estimator. The continuous main equation is then replaced by different types of nonlinear equations such as a binary choice equation and an ordered probit equation, to form more complex models matching various empirical specifications. Moreover, a two-equation system can be easily extended to a trivariate case by adding one more equation which can either be simultaneous to other equations or contain one more level of sample selection to form the double-hurdle models. Applications of all those models have been discussed in detail in this chapter.

Chapter 3 considers the identification conditions (with or without exclusion restrictions) around estimating bivariate probit models with an endogenous dummy regressor. In those models, a regressand with binary outcomes becomes one of the regressors in the latent equation which determines the other binary choice equation. Exclusion restrictions are often included to make sure models are identified for empirical analysis. However, counterparts show that identification problems arise from experimental design (Leung &

Yu 1996), not from the absence of exclusion restrictions. As a result, Monte Carlo experiments are applied in different circumstances to check whether exclusion restrictions are essential for model identification, and to compare three methods including maximum likelihood estimation (MLE), the two-step method, and the Bayesian method. Some basic theories about Bayesian analysis have been introduced in this chapter including sampling algorithms like Gibbs and Metropolis-Hastings (MH) algorithms, and the simulation inefficient factor (SIF) which is used to evaluate the convergence of the Bayesian methods. The Bayesian sampler for this particular type of model is a simpler case of Chib & Greenberg (1998)'s multivariate example, but our derivation is somehow different by decomposing the joint posterior distribution into one-dimensional conditional posterior densities. What we are also interested in is whether the inconsistency of the two-step estimator creates major finite sample problems, and whether the Bayesian sampler has better performance than MLE. The Monte Carlo results reveal that each estimator has its own benefits in estimating the model. Exclusion restrictions can help improving performance of all estimators, but they are not that essential to identifying the model. We also discuss a few ways to improve model identification when exclusion restrictions are not available.

In Chapter 4, we model an ordered variable with sample selection. More specifically, the ordered outcomes are only observed when a binary indicator returns the value one. An efficient Gibbs sampler is proposed to estimate this model with the idea of data augmentation and reparameterization, so conjugate conditional posteriors can be obtained with certain prior design. We also propose a likelihood-based two-step method which is based on the form of the concentrated likelihood function. Then a Monte Carlo study compares the two methods with FIML and a two-step method extended from the classic two-step method. Results indicate that both the Bayesian method and the likelihood-based two-step method perform as well as FIML. Moreover, the Bayesian credible intervals can be used for interval estimation and a way to describe the distributions of model parameters, while the likelihood-based two-step method has less convergence problem than FIML. The extension of the classic two-step method does not perform well with strong error correlation due to the inconsistency. Effects of exclusion restrictions are also detected in the simulation study. Once again, results show exclusion restrictions are not essential in determining identification for this model, although they can reduce the difficulties of estimation for all methods. Finally, the model is applied to empirical data studying the effects of mental illness on occupational skill categories and labour force employment. Marginal effects of mental illness are derived based on the estimates of FIML and the Bayesian method. The Bayesian method displays the superiorities in evaluating the distributions of marginal effects.

The double-hurdle model in Chapter 5 is designed to better describe the three levels of labour market outcomes including participation, employment and occupational skill categories. It can be treated as an extension of the model in Chapter 4 by adding one extra level of sample selection about participation. As the two-step method cannot guarantee accurate estimation in the two-equation system, we focus on comparison of FIML and the Bayesian method for this trivariate model. Two Gibbs samplers are proposed here. One is derived in a more standard procedure, while the other uses the same ideas of data reparameterization in Chapter 4 to construct conditional posterior distributions. Both samplers can offer accurate and efficient estimation in a Monte Carlo study, when FIML fails to converge for half the samples but gives reliable estimation for the remaining samples. In addition, an empirical application to mental health and labour market outcomes is analyzed on this double-hurdle model estimated by the three estimators. Although the Monte Carlo experiments do not show much difference in convergence between the two Bayesian samplers, the sampler using reparameterization provides much better convergence in the empirical study. Meanwhile, FIML cannot give reliable estimation for this empirical data. That is why the analysis of model estimation and corresponding marginal effects is concentrated on one valid Bayesian sampler.

Chapter 6 concludes the thesis. It summarizes the main findings of the whole thesis, and includes a few limitations and some potential extensions.

Chapter 2

Literature Review

2.1 Introduction

This chapter systematically reviews the literature related to the thesis, which is mainly about different types of models with endogeneity and non-random selection, as well as some existing estimation methods. We will first review relevant models with non-random selection in Section 2.2, including the traditional sample selection model with a continuous main equation and extended models with nonlinear main equations, such as binary choice models, ordered probit models, and models with double hurdles. Then we review some literature focused on models with endogeneity in Section 2.3. We start from the endogenous treatment model as it is a linear regression model with an endogenous dummy regressor, followed by a discussion of bivariate probit models with endogenous regressors and some extensions to multivariate cases with endogeneity. Applications of such models are introduced as some of the commonly used estimation methods. Section 2.4 concludes.

2.2 Models with Non-random Selection

Selection bias usually arises because of non-random samples. While part of the reason for selection into the sample is observed, the problem of bias arises when the non-observable part is correlated with factors determining outcomes. Mathematically, sample selection occurs as the data is only observed when a binary indicator takes the value one for the relevant outcome. Firstly, we will introduce a traditional sample selection model with continuous data observed after selection, as well as available estimation methods used to estimate this model. Secondly, the development of nonlinear models with selection is outlined, because the models that will be discussed in Chapter 4 and 5 are sample selection models with non-linear forms. Thirdly, models with at least two levels of sample selection are discussed, since Chapter 5 models a specific double-hurdle model for labor market outcomes. This review introduces the main essential estimation methods applied in the literature to estimate the models of interest in Chapter 4 and 5, methods that will later be compared to our proposed Bayesian samplers.

2.2.1 Conventional Sample Selection Models

Most current models of sample selection are based on the brilliant work of Heckman (1976). He introduces the standard structure of models of sample selection, motivated by the discussion of Gronau (1974) about the truncation problem of the wage offer and labour force participation. The model is built for a continuous variable z_{i2} which is only observed when an associated variable z_{i1} is positive.

$$\begin{cases} z_{i1} = x_{i1}'\beta_1 + \epsilon_{i1} \\ z_{i2} = x_{i2}'\beta_2 + \epsilon_{i2} \qquad 1 \le i \le n \end{cases}$$

The disturbances ϵ_{i1} and ϵ_{i2} are assumed to have a bivariate normal distribution with zero means, variance 1 and σ^2, and the crucial coefficient of correlation ρ. The observable outcomes are displayed by

$$\begin{cases} y_{i1} = I(z_{i1} > 0) \\ y_{i2} = z_{i2} \text{ if } y_{i1} = 1, \text{ and missing otherwise.} \end{cases}$$

Because

$$E(y_{i2}|x_{i1}, x_{i2}, y_{i1}) = x_{i2}'\beta_2 + \rho E(\epsilon_{i2}|x_{i1}, x_{i2}, y_{i1}),$$

selection bias can be interpreted the same as omitted variable bias with the omitted variable being $E(\epsilon_{i2}|x_{i1}, x_{i2}, y_{i1})$. The bias will only occur when $\rho \neq 0$.

Heckman proposes an estimator sometimes called **Heckit** (Wooldridge 2002, p.564) by adding an estimated value for the omitted variable as a regressor to remove the selection bias. Section 4.4.2 contains more specific formulations of the two-step method. This simple estimator is easy to implement and is used by Heckman to estimate a model of female labour supply and wages. The consistency properties of the two stage least squares (2SLS) method is confirmed and Heckman (1979) shows how to adjust standard errors to take account of the two-step estimation approach. This method is the most widely used method to estimate models with non-random selection.

Based on Heckman's model, Vella (1998) derives the bias caused by the sample selection and some estimation methods including parametric and semi-nonparametric methods in a likelihood framework. With the assumption that disturbances ϵ_{i1} and ϵ_{i2} are bivariate normally distributed, the likelihood function can be expressed as

$$L = \prod_{i=1}^{n} \left\{ \left[\int_{-x'_{i1}\beta_1}^{\infty} \phi(y_{i2} - x'_{i2}\beta_2, \epsilon_{i2}) d\epsilon_{i2} \right]^{y_{i1}} \times \left[\int_{-\infty}^{-x'_{i1}\beta_1} \int_{-\infty}^{\infty} \phi(\epsilon_{i1}, \epsilon_{i2}) d\epsilon_{i1} d\epsilon_{i2} \right]^{(1-y_{i1})} \right\},$$

where ϕ denotes the probability density function for the bivariate normal distribution. Vella shows the formulation of various parametric and semi-parametric two-step methods, together with some other models which have alternative censoring rules.

Other estimation methods for estimating Heckman's model include Bayesian analysis developed by Omori (2007) who tries to accelerate the convergence of MCMC algorithms in sampling the corresponding posterior distribution. His Gibbs sampler has improved estimation efficiency and reduces the sample autocorrelations, in comparison with a standard Gibbs sampler. However, this method may still be computational burdensome in some ways.

Winship & Mare (1992) provides an early survey of models for sample selection bias. In addition, Manski (1989) derives two kinds of prior restrictions to identify the parameters in a sample selection regression model. One weak restriction is a bound on the conditional expectation, while the other type is a separability restriction derived from latent variable models. Lee (1982) suggests a way of transforming random variables with certain non-normal distributions of the error terms to a bivariate normal distribution, so the two stage method can be applied without the normality assumption of the error terms. Further, the Tobit model proposed by Tobin (1958) is usually considered as a special case of the sample selection model with a specific censoring rule. It can be estimated by MLE, the two-step method, and a Bayesian method proposed by Chib (1992).

2.2.2 Probit Models with Sample Selection

Heckman's approach is designed for models where the dependent variable is continuous and there is non-random selection. However, it is common to be modeling discrete data in various model specifications, such as probit and ordered probit models with selection. The following analysis will focus on bivariate probit models with sample selection.

Different levels of partial observability are investigated in bivariate probit models with sample selection. For example, Meng & Schmidt (1985) have presented bivariate probit models with different levels of observability, in order to measure the loss of asymptotic efficiency of the parameter estimates in comparison with full observability. Poirier (1980) develops a particular bivariate probit model in which only one of the four possible outcomes is observed. He points out that MLE will be inefficient compared to those cases in which fully observed choices are obtained. Identification issues are discussed, and he also mentions that it is essential to identify the model with sufficient variation of exogenous variables over the sample. Besides that, the most commonly discussed case is the one where binary outcomes are censored only when the dependent variable in the selection equation takes the value zero.

Various empirical studies have been done on the bivariate probit model with sample selection. We give just a few examples here to illustrate diversity of applications. For instance, Van de Ven & Van Praag (1981) have analyzed the choice of a health insurance, with a sample which is censored because of questionnaire design. They have estimated their model with Heckman's two-step method. Boyes et al. (1989) apply a bivariate 'censored probit' model to investigate the bank credit scoring problem. They combine a choice-based estimator with partial observability to estimate the relationship between credit card lending and the loan earnings process. Standard Heckman procedure is no longer applicable, so a 'weighted' likelihood function is maximized to obtain estimates of the model. The same credit scoring model for loan approvals is also derived by Greene

(1992) to incorporate the loan default probability with expected profit. Sartori (2003) applies this model in interpreting political phenomena about judicial dependence in electoral democracies, and provides a new maximum-likelihood estimator. In health economics, Belkar et al. (2006) use this model to analyze the relationship between awareness of Pap tests and the choice to screen for cervical cancer. Holm & Jæger (2008) have studied educational transition models, with the first transition from elementary school to high school and the second transition from high school to tertiary study. Selection bias has been discussed in the context of this specification of the bivariate probit model. This model is applied on empirical data to illustrate how sample selection influences the effect of family background on the probability of making educational transitions. A Monte Carlo study is utilized by Belkar & Fiebig (2008) to compare the censoring bivariate probit model with single-equation probit models, and to highlight the necessaries to consider censoring in modeling choices.

Here we will focus on different estimation methods which are commonly applied to estimate bivariate probit models with partial observability. In the literature, many researchers have tried to extend Heckman's approach to estimate such nonlinear models with sample selection. Nicoletti & Peracchi (2001) aim to estimate the bias of two-step estimators with some Monte Carlo experiments. They have shown that two-step estimators approximate the true parameter well, and the accuracy is closely related to the size of the correlation of the errors. Dubin & Rivers (1989) also discuss both the two-step method and a maximum likelihood estimator in the context that Heckman's framework is extended to probit models. However, Greene (2001) criticizes the two-step method as inappropriate for a probit model with sample selection. Greene (2006) points out that three problems will show up in any nonlinear model, so one cannot generalize the two-step method by dropping the inverse Mills ratio into the model. He provides a general maximum likelihood approach to incorporate sample selection in nonlinear models. These two classical methods will be used in estimating the models of interest in this thesis.

2.2.3 Ordered Probit Models with Sample Selection

In addition to binary responses, an ordered multinomial response is commonly found in empirical literature. Given a latent variable y^* which is determined by

$$y_i^* = x_i'\beta + \varepsilon_i \qquad \varepsilon_i \sim N(0,1),$$

the ordered probit model for y can be expressed as

$$y_i = \begin{cases} 0 & \text{if } y_i^* \leq \gamma_1 \\ 1 & \text{if } \gamma_1 \leq y_i^* \leq \gamma_2 \\ & \vdots \\ J & \text{if } y_i^* \geq \gamma_J, \end{cases}$$

in which the threshold parameters are constrained by $\gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_J$ (Wooldridge 2002, p.505). One can use MLE to estimate the single equation model, or use OLS by treating the ordered outcomes as continuous data.

Daykin & Moffatt (2002) emphasis that the ordered probit model is an appropriate framework for statistical analysis especially when survey responses are ordinal. They apply this model in an empirical example about the extent of pain regarded by physiotherapists. Hausman et al. (1992) also use this model to analyze trade-to-trade price change in discrete increments. MLE is used to estimate the model in order to measure several transaction-related quantities.

The ordered probit model can be extended to bivariate or multivariate equation systems together with probit or continuous equations, according to the specific model implied by the empirical example. Those more complicated models are sometimes estimated by Bayesian methods. For instance, Chib & Hamilton (2000) make the treatment variable ordinal to model the effect of clustered data with non-random selection. In their example, a cluster is defined by a hospital after an individual needs treatment. The observed treatment in a cluster is determined by an ordinal probit model. Munkin & Trivedi (2008) also use the Bayesian method to analyze an ordered probit model with endogenous selection. They are especially interested in the effect of endogenous multinomial choice indicators on an ordinal outcome variable which indicates the degrees of medical service utilization.

2.2.4 Double-hurdle Models

The notion of double-hurdle is first proposed by Cragg (1971) when certain positive values can only be observed after two hurdles. His model is quite like a sample selection model but with one more additional adjustment. His model requires that each individual has to satisfy two conditions to be observed in the sample, thus two levels of censoring should be considered in his model. Since then, the use of double-hurdle models have been widely used in cross-sectional studies, especially when censoring is determined by more than one level of sample selection.

The p-tobit model is proposed by Deaton & Irish (1984) to investigate expenditures in household budgets. The model is constructed as a standard tobit specification with the operation of a binary censoring process. Because the tobit model is usually treated as a sample selection model with specific censoring rule, the binary censoring process can be treated as an additional hurdle. Therefore, the p-tobit model can be counted as a special case of a double-hurdle model. The same idea is adopted by Butler & Moffitt (1982) to understand factors contributing to married women's labour supply, such as household demand for clothing.

Jones (1989) has discussed a double-hurdle model in the study of cigarette consumption, and treats starting and quitting smoking as separate participation decisions which both have sample selection effects on consumption. Although he derives the likelihood function to estimate a trivariate model, the assumption that the three equation errors are independent is critical, and is very unlikely to be true. The zero correlation assumption was made due to the lack of computational software which can accurately conduct numerical optimization at that time. This idea of double-hurdle model is also applied by Labeaga (1999) to estimate the demand for tobacco with panel data, and by Yen (2005) to detect gender differences in cigarette consumption.

Most zero-inflated models have similar structures as double-hurdle models, as the zero observations are usually from two different resources. The hurdles in the double-hurdle models are mostly independent, while the two errors in zero-inflated models are correlated. For instance, Blundell et al. (1987) are interested in the possibility that zero hours of work represent both unemployment and non-participation. Another example is the zero-inflated ordered probit model proposed by Harris & Zhao (2007) with an ordered probit model censoring in a split probit rule. The model is applied to modeling tobacco consumption, when zero consumption are caused by those who do not participate and those who participate but do not consume.

2.3 Models with Endogeneity

An endogenous variable is one which is correlated with the error term. Endogeneity can arise as a result of measurement error, simultaneity, omitted variables, and sample selection. Models with sample selection are closely related to models with endogeneity, as the latent variable of the binary selection equation is correlated with the error term in the main equation. In early literature, it was normally assumed that errors are uncorrelated between the sample selection equation and main equation. Since such an assumption is unrealistic in most empirical work, correlation between the equations has been considered in most models with selection. The correlation between the equations in sample selection models is a special form of endogeneity. Vella (1998) shows that it is possible to view the sample selection model as a model with a censored endogenous regressor where the parameters are the same for each subsample. Both types of models have quite similar properties, share similar restrictions for identification and can be estimated by the same estimators most of the time. The effects of a binary endogenous variable are of particular interest and it is generally viewed as a mean shift parameter in the equation.

2.3.1 Endogenous Treatment Models

Historically, endogeneity has mostly been discussed in the context of a linear regression model with a binary endogenous regressor, which is often employed as a 'treatment effects' model. It has the following form,

$$\begin{cases} d_i = I(x_{i1}'\beta_1 + u_i > 0) & 1 \le i \le n \\ y_i = x_{i2}'\beta_2 + \alpha d_i + v_i & 1 \le i \le n \end{cases},$$

in which the random terms u_i and v_i are assumed to be normal distributed with zero means and covariance matrix

$$\Sigma_d = \begin{bmatrix} 1 & \sigma\rho \\ \sigma\rho & \sigma \end{bmatrix},$$

for potential endogeneity of d_i. Heckman (1978) first presents this so-called 'classical simultaneous equations model with structural shift' and shows certain restrictions imposed on the model in order to generate a sensible statistical structure. He applies maximum likelihood estimation and proposes a two-stage least squares estimator to evaluate the structural parameters.

2.3.2 Bivariate Probit Models with Endogeneity

The above model can be easily extended to bivariate probit models with endogeneity, when y_i cannot be observed directly but is treated as propensity deciding an indicator function. Chapter 3 is mainly about estimation of bivariate probit models with an endogenous regressor, in which the probit model contains an endogenous binary regressor. One can refer to Section 3.2.1 for model specifications in detail.

Estimation techniques for endogenous treatment models in Section 2.3.1 can also be utilized in evaluating bivariate probit models with an endogenous regressor. As an example, Wooldridge (2002, p478) attempts to use some 'seemingly obvious' two-step procedure by mimicking two-stage least squares, but finds out it does not produce consistent estimators because of the nonlinearity. Burnett (1997) uses the two-step method discussed by Rivers & Vuong (1988), in order to estimate recursive simultaneous probit models of performance in gender-related economics courses as well as women's studies programs in the undergraduate, liberal arts curriculum. The same model and data are then re-estimated using full information maximum likelihood technique by Greene (1998), as he argues that the two-step procedure has several shortcomings including lack of consistency and inefficient estimation. As other examples, the MLE approach is also employed to study the effectiveness of attending a Catholic high school on the possibilities of attending college by Evans & Schwab (1995), to describe the effect of supplemental insurance ownership on health demand in Switzerland (Holly et al. 1998), and to investigate the relationship between offenders and victimization (Deadman & MacDonald 2004).

2.3.3 Multivariate Models with Endogeneity

The bivariate probit model with endogeneity can be considered as a special case of multivariate models with endogeneity. As we have mentioned, non-random selection may be treated as a type of endogenous regressor, so we may find some similar properties of multivariate models with endogeneity to the trivariate model with two hurdles of sample selection which will be discussed in Chapter 5.

In a multivariate model with endogeneity, the multiple equations may have various forms such as probit, ordered probit and logit. Notice that equations can have different forms in one model, especially in a number of empirical studies. For example, Li & Tobias (2007) study a trivariate simultaneous model with endogeneity. The treatment decision which is shown by a binary indicator has effects on two ordered probit equations, while the three disturbances in potential latent variables are assumed to be correlated. Maddala & Lee (1976) have also discussed recursive models with qualitative endogenous variables under a logit framework.

Some common estimation methods are used to estimate those models. Arendt & Holm (2006) introduce a least-squares approximation in a two-step procedure in the bivariate case and then extend it to a trivariate model to study the effect of trust on voting behavior. They also apply such models and estimation methods on another empirical study about the impact of physician advice on physical activity (Arendt & Holm 2007). To demonstrate the limitations of a two-step procedure, Bhattacharya et al. (2006) conduct a Monte Carlo exercise to compare the performance of two types of two-step estimators and FIML, and also give an empirical example examining the effect of private and public insurance coverage on the mortality of HIV patients. In addition, Kim (2006) use a simple two-step method to estimate a simultaneous equations model with a common endogenous dummy variable under a selection framework. As an alternative, Bayesian analysis is applied by Li & Tobias (2007) to estimate the treatment effects of a binary indicator in a model with two simultaneous ordered outcomes.

Such models and estimation methods are used in a wide range of applications. For instance, Angrist (2001) discusses issues of identifying causal effects with empirical work on the effect of childbearing on employment status and hours of work. Bryson et al. (2004) have done some research on an extension of the recursive ordered probit model studying the effect of union membership on job satisfaction in labor markets. Zhang et al. (2009) give an application of multivariate probit models with endogeneity, in examining the effect of some chronic diseases on the probability of labour force participation.

2.4 Conclusion

This chapter provides a selected survey of the development of models with sample selection and endogeneity, while some estimation methods are briefly introduced. We initially examine the sample selection model and the two-step method proposed by Heckman, followed by some nonlinear models with selection and double hurdle models. Then, the endogenous treatment model is shown with a few estimation methods. Lastly, we have a discussion about bivariate probit models and multivariate models with endogeneity.

We are particularly interested in some properties of bivariate probit models with an endogenous regressor as discussed in Section 2.3.2. There are a number of unresolved issues like the effects of exclusion restrictions on estimator performance. Even though the model can be identified without exclusion restrictions, most empirical studies impose restrictions as they are seen as important for reliable estimation. This issue will be explored further in Chapter 3.

The model examined in Chapter 4 is about an ordered variable, which is observed when a binary indicator returns the value 1. Such a specification can be treated as an extension of the models in Section 2.2. Two types of estimation methods, MLE and the two-step method, are most commonly used to estimate models with selection and endogeneity. The two methods are easy to implement and can be reliable for some simple models like the traditional sample selection model and the endogenous treatment model, but they may have different problems in estimating nonlinear models with sample selection and endogeneity. MLE sometimes fails to converge, while the two-step method is no longer consistent for those nonlinear models. The Bayesian method applied to similar models in Section 2.2.3 can also be applied to the ordered probit model with non-random selection. Therefore, the performance of different estimation methods will be discussed in Chapter 4.

The above model with an additional hurdle of sample selection, which will be estimated in Chapter 5, becomes an example of double-hurdle models discussed in Section 2.2.4. This type of model fits well in some empirical specifications. However, they are not investigated much in the literature, so efficient estimation methods are quite poorly understood. MLE and the two-step method have some problems in estimating the model in Chapter 4. It is even more difficult for these two methods to estimate more complicated models like the double hurdle model in Chapter 5. That is why we will introduce the Bayesian method to estimate those models of interest. Monte Carlo experiments will be used to compare MLE and the Bayesian method, and to examine the performance of the inconsistent two-step method.

Chapter 3

Estimating Bivariate Probit Models with an Endogenous Dummy Regressor

3.1 Introduction

The probit model is the most commonly used discrete choice model with underlying latent variables that manifest binary outcomes through a threshold specification. It can be naturally extended to bivariate or multivariate probit models with correlated error terms for latent variables. Sometimes, the regressand of one equation can be part of the set of regressors in another equation. A model of this form is termed a recursive probit model with discrete endogenous variables and it can be named as a bivariate probit model with an endogenous dummy regressor.

In multiple-equation structural models, exclusion restrictions are often imposed in the literature to guarantee model identification. However, some researchers point out that exclusion restrictions are not theoretically necessary in nonlinear models like the probit. It is important to resolve whether the model parameters can be reliably estimated when exclusion restrictions are not imposed. We will examine this question in this chapter. In addition, we are also trying to identify other factors which may improve parameter precision, including degree of variation in explanatory variables and different types of explanatory variables. More specifically, this chapter will examine the bivariate probit model with an endogenous dummy regressor utilizing Monte Carlo simulations on three estimation methods to check whether exclusion restrictions are important in providing reliable parameter estimates. Further, results will be produced that examine how different types of explanatory variables affect estimator performance in cases with or without exclusion restrictions.

A second aim of the chapter is to compare three methods to check which one will perform best in certain circumstances. Those estimation methods include two commonly used classical methods, maximum likelihood estimation (MLE) and a two-step method, and a Bayesian method. It is widely believed that MLE always gives quite good estimation and it is usually treated as some kind of benchmark, but in multivariate models, there are occasionally convergence problems. That is why the two other methods are considered. On one hand, the two-step method is especially preferred by some applied econometricians for simpler implementation than MLE, but there is no consistent two-step estimator for this particular model. So comparison between these two methods will indicate how this inconsistent estimator performs. On the other hand, Bayesian analysis provides another framework to estimate models. The Bayesian sampler provided in this chapter is similar to Chib & Greenberg (1998)'s sampler on multivariate probit models but it is a simpler bivariate case here. Bayesian inference in this chapter is implemented by first substituting an error term with a transformation function of another error term. The main difference between the sampler proposed here and the Chib & Greenberg's method is that here some complex conditional posterior densities are decomposed into densities with lower dimensions. For example, latent variables are all sampled from univariate truncated normal distributions. The sampler is utilized on a bivariate probit model with an endogenous dummy variable to check whether it is a better approach than MLE with respect to different exclusion restrictions and types of exogenous variables.

This chapter is organized as follows. Section 3.2 describes the bivariate probit model with an endogenous dummy regressor in detail as well as discussing the effects of exclusion restrictions. Section 3.3 introduces the two classical estimation methods, while Section

3.4 shows how the Bayesian method works including some basic theory, sampling algorithms, the specific sampler of this model and the index for convergence diagnosis. After some general Monte Carlo design is described in Section 3.5, Section 3.6 analyzes the Monte Carlo results, while Section 3.7 and 3.8 present results for partial structural equation estimation with or without exclusion restrictions respectively. Section 3.9 concludes the chapter.

3.2 The Model and Exclusion Restrictions

3.2.1 The Model

The relevant model formulation has been proposed by Maddala & Lee (1976), where we have n pairs of latent variables expressed as,

$$\begin{cases} z_{i1} = x_{i1}'\beta_1 + u_i & 1 \le i \le n \\ z_{i2} = x_{i2}'\beta_2 + \alpha y_{i1} + v_i & 1 \le i \le n \end{cases} \tag{3.2.1}$$

in which $E(u_i) = E(v_i) = 0$, $Var(u_i) = Var(v_i) = 1$ and $Cov(u_i, v_i) = \rho$. $\beta_1 \in R^{k_1}$ and $\beta_2 \in R^{k_2}$ are unknown parameters, while x_{i1} and x_{i2} are exogenous variables including one constant respectively. Let $\beta' = (\beta_1', \beta_2') \in R^k$, $k = k_1 + k_2$. Here, z_{i1} and z_{i2} are latent variables which determine the observed binary outcomes of y_{i1} and y_{i2} such that

$$\begin{cases} y_{i1} = I(z_{i1} > 0) \\ y_{i2} = I(z_{i2} > 0). \end{cases} \tag{3.2.2}$$

Note that in this setup z_{i2} is a function of y_{i1}, the chosen binary outcome, not z_{i1}, the propensity. Error terms must have zero mean and unit variance for identification reasons. Take the single equation probit model as an example. Suppose $y_{i1} = I(z_{i1} > 0)$ has a latent variable with mean μ and variance σ^2. This equation is identical to the model $y_{i1} = I((x_{i1}'\beta_1 - \mu)/\sigma > 0)$ in which latent variable has zero mean and unit variance. Thus, non-zero mean becomes part of the constant, while the probit model cannot distinguish β_1 and σ, but obtain β_1/σ instead.

A transformation is used by assuming $v_i = \rho u_i + \sqrt{1-\rho^2}\eta_i$ in which $E(\eta_i) = 0$, $Var(\eta_i) = 1$ and $Cov(u_i, \eta_i) = 0$. Hence the bivariate probit model can be expressed as

$$\begin{cases} z_{i1} = x_{i1}'\beta_1 + u_i \\ z_{i2} = x_{i2}'\beta_2 + \alpha y_{i1} + \rho(z_{i1} - x_{i1}'\beta_1) + \sqrt{1-\rho^2}\eta_i. \end{cases} \tag{3.2.3}$$

This transformation is mainly utilized to improve the Bayesian analysis discussed in detail in Section 3.4.1.

This model is similar to a 'treatment effects' model as the endogenous binary regressor y_{i1} indicates the presence or absence of some treatment. In a traditional treatment effects model, the dummy variable y_{i1} is also included as part of the explanatory regressors of z_{i2}, but z_{i2} is directly observed instead of y_{i2}. The treatment effect for this endogenous bivariate probit model can be estimated as $E[y_{i2}|y_{i1} = 1, x_{i1}, x_{i2}] - E[y_{i2}|y_{i1} = 0, x_{i1}, x_{i2}]$, and Angrist (2001) shows the estimation of causal effects presents no special challenges whether y_{i2} is binary or continuously distributed.

3.2.2 Exclusion Restrictions and the Identification Problem

The jargon of exclusion restrictions first arises in simultaneous equations models. To show the general idea, assume latent variables z_{i1} and z_{i2} can be directly observed and the binary variable y_{i1} is replaced by z_{i1}, then equations (3.2.3) become a linear system. The system can be shown as

$$\begin{cases} z_{i1} = x_{i1}'\beta_1 + u_i \\ z_{i2} = x_{i2}'\beta_2 + \alpha z_{i1} + v_i \qquad 1 \le i \le n \end{cases}$$

Apparently, β_1 can always be identified by the first equation. When $x_{i1} = x_{i2}$, the second equation becomes

$$z_{i2} = x_{i1}'\beta_2 + \alpha(x_{i1}'\beta_1 + u_i) + v_i = x_{i1}'(\beta_2 + \alpha\beta_1) + \alpha u_i + v_i.$$

The sum $\beta_2 + \alpha\beta_1$ can be estimated through this equation, but β_2 and α cannot be isolated. Thus, this linear system cannot be identified.

As a result, exclusion restrictions are always required by allowing at least one exogenous variable in x_{i1} which does not appear in x_{i2}. For example, let $x_{i1} = [x_{i2}, \omega]$ with respective coefficients $\beta_1 = [\beta_x, \beta_\omega]$. Then the linear system becomes

$$\begin{cases} z_{i1} = x'_{i2}\beta_x + \omega'\beta_\omega + u_i \\ z_{i2} = x'_{i2}(\beta_2 + \alpha\beta_x) + \omega'(\alpha\beta_\omega) + \alpha u_i + v_i. \end{cases}$$

β_x and β_ω can still be estimated by the first equation, while $\beta_2 + \alpha\beta_x$ and $\alpha\beta_\omega$ can be estimated by the second equation. α is available once β_ω and $\alpha\beta_\omega$ are obtained. Then, β_2 can be calculated from $\beta_2 + \alpha\beta_x$ when α and β_x are known. Thus, the linear system is successfully identified with exclusion restrictions.

Now consider the case we are interested in here, where z_{i1} and z_{i2} are not observed, and y_{i1} and y_{i2} are observed as nonlinear functions of z_{i1} and z_{i2}. Maddala (1983, p123) especially analyzes a case in which the only explanatory variables are constants and concludes that the parameters are not identified for all recursive models if disturbances are correlated and x_{i2} includes all the variables in x_{i1}. According to Maddala's analysis, a contingency table of probabilities can be obtained if the occurrence of y_{i2} is a precondition for y_{i1}. The entry for $y_{i1} = 1, y_{i2} = 0$ will be identically zero, while the other three probabilities will be available. But three probabilities will not be enough for identifying four parameters including two constants, the coefficient of the endogenous dummy variable and the cross-equation error correlation.

On the other hand, Wilde (2000) argues that Maddala's statement is only valid in his special case when x_{i1} and x_{i2} are constants, and the classic identification problem does not exist in general models with more than four parameters when there is sufficient variation in exogenous regressors. He especially considers a simple case of one exogenous regressor with only two different values and concludes that there is sufficient variation in the data to identify the six parameters. Moreover, Greene (2002, p.E21-115) mentions that the conventional rules for identification in simultaneous equations models do not apply in 'treatment effects' models. Because of the nonlinearity of the conditional mean function, it is not necessary for there to be variables excluded from either equation. However, exclusion restrictions might help in making the estimation results more robust to distributional mis-specification (Monfardini & Radice 2008).

From a practical point of view, some researchers believe that it is essential to estimate the models with exclusion restrictions, although these are not always available in practice. Meanwhile, other researchers point out that it is actually not necessary to include exclusion restrictions in these nonlinear models. In this chapter, a Monte Carlo simulation study has been done to examine the effect of exclusion restrictions on estimation of bivariate probit models with an endogenous dummy regressor.

3.3 Classical Estimation Methods

3.3.1 Maximum Likelihood Estimation

Greene (2003, p715) argues that the endogenous nature of the variables on the right-hand side of the equation can surprisingly be ignored in formulating the log-likelihood. He explains this fact using $Pr(y_{i1} = 1, y_{i2} = 1) = Pr(y_{i2} = 1|y_{i1} = 1)Pr(y_{i1} = 1)$ which enters the log-likelihood. Since the marginal probability is just $\Phi(x'_{i1}\beta_1)$ and the conditional probability is $\Phi_2(\cdots)/\Phi(x'_{i1}\beta_1)$, the product returns $\Phi_2(\cdots)$, where Φ and Φ_2 are respectively the univariate and bivariate cumulative normal distribution functions. Therefore, the likelihood function obtained according to the formulation in Section 3.2.1 is almost the same as that for the reduced form bivariate probit model. Since the probability function of (y_1, y_2) can be written as

$$\begin{cases} P_{11} = Pr(y_{i1} = 1, y_{i2} = 1|x_{i1}, x_{i2}) = \Phi_2(x'_{i1}\beta_1, x'_{i2}\beta_2 + \alpha, \rho) \\ P_{10} = Pr(y_{i1} = 1, y_{i2} = 0|x_{i1}, x_{i2}) = \Phi_2(x'_{i1}\beta_1, -x'_{i2}\beta_2 - \alpha, -\rho) \\ P_{01} = Pr(y_{i1} = 0, y_{i2} = 1|x_{i1}, x_{i2}) = \Phi_2(-x'_{i1}\beta_1, x'_{i2}\beta_2, -\rho) \\ P_{00} = Pr(y_{i1} = 0, y_{i2} = 0|x_{i1}, x_{i2}) = \Phi_2(-x'_{i1}\beta_1, -x'_{i2}\beta_2, \rho), \end{cases}$$

the log-likelihood function to be maximized becomes

$$ln\ L \quad = \quad \sum_{i=1}^{n} [y_{i1}y_{i2}ln(P_{11}) + y_{i1}(1 - y_{i2})ln(P_{10})$$

$$+(1-y_{i1})y_{i2}ln(P_{01})+(1-y_{i1})(1-y_{i2})ln(P_{00})].$$

On maximizing this function, maximum likelihood estimates are obtained.

3.3.2 Two-step Method

Besides MLE, the two-step method is another method frequently used in estimating treatment effects models. It would give consistent estimates and appropriate asymptotic standard errors if y_{i2} was a continuous variable but not when it is a dummy outcome. However, Wooldridge (2002, p478) tries to mimic a two-stage least squares estimator on a bivariate probit model with an endogenous regressor using $P(y_{i2}=1|x_{i2})=E(y_{i2}|x_{i2})=E(I\left[x_{i2}'\beta_2+\alpha y_{i1}+v_i>0\right])$. But he finds that the expected value cannot be passed through, since the indicator function is nonlinear. So it is no longer appropriate to insert the predicted values of the sample estimates of $E[y_{i1}|x_{i1}]$ in the second equation for the dummy outcome.

As a result, Greene (1998) recommends MLE techniques because nonlinear least squares is inefficient in estimating the model, although two-stage least squares is relatively easy to conduct. In addition, he summarizes several drawbacks of two-step procedure and points out that it is potentially inefficient as it does not account for the possible correlation between the disturbances in the two equations.

Nevertheless, some applied econometricians prefer the two-step method for its computational simplicity. All of this suggests a two-step MLE procedure based on the pioneer work of Heckman (1978). First, β_1 can be estimated with the maximum likelihood technique on the single equation probit model using

$$L_1=\prod_{i=1}^{n}\left[\Phi(x_{i1}'\beta_1)^{y_{i1}}\times\Phi(-x_{i1}'\beta_1)^{(1-y_{i1})}\right].$$

At the second step, Arendt & Holm (2006) suggest approximating $E(y_{i2}|x_{i2})$ by $E(z_{i2}|x_{i2})$ which are based on the following conditional expectations:

$$E(z_{i2}|x_{i2},y_{i1}=1) \quad = \quad x_{i2}'\beta_2+\alpha y_{i1}+\rho\frac{\phi(x_{i1}'\beta_1)}{\Phi(x_{i1}'\beta_1)}$$

$$E(z_{i2}|x_{i2}, y_{i1} = 0) \quad = \quad x_{i2}'\beta_2 + \alpha y_{i1} - \rho \frac{\phi(x_{i1}'\beta_1)}{1 - \Phi(x_{i1}'\beta_1)},$$

where z_{i1} and z_{i2} are defined below:

$$\begin{cases} z_{i1} = x_{i1}'\beta_1 + u_i & 1 \leq i \leq n \\ z_{i2} = x_{i2}'\beta_2 + \alpha y_{i1} + v_i & 1 \leq i \leq n. \end{cases}$$

The second step can be conducted by maximizing

$$\begin{aligned} L_2 = \prod_{i=1}^{n} \quad & \left\{ \Phi\left[x_{i2}'\beta_2 + \alpha y_{i1} + \rho \frac{\phi(x_{i1}'\beta_1)}{\Phi(x_{i1}'\beta_1)}\right]^{y_{i1}y_{i2}} \times \right. \\ & \Phi\left[-x_{i2}'\beta_2 - \alpha y_{i1} - \rho \frac{\phi(x_{i1}'\beta_1)}{\Phi(x_{i1}'\beta_1)}\right]^{y_{i1}(1-y_{i2})} \times \\ & \Phi\left[x_{i2}'\beta_2 + \alpha y_{i1} - \rho \frac{\phi(x_{i1}'\beta_1)}{1 - \Phi(x_{i1}'\beta_1)}\right]^{(1-y_{i1})y_{i2}} \times \\ & \left. \Phi\left[-x_{i2}'\beta_2 - \alpha y_{i1} + \rho \frac{\phi(x_{i1}'\beta_1)}{1 - \Phi(x_{i1}'\beta_1)}\right]^{(1-y_{i1})(1-y_{i2})} \right\}. \end{aligned}$$

Because $E(y_{i2}|x_{i2}) \neq E(z_{i2}|x_{i2})$, this two-step estimator does not provide consistent estimates. However, the approximation will be good when ρ is close to zero.

3.4 Bayesian Method

3.4.1 Bayesian Analysis with Data Augmentation

Bayesian econometrics is originally derived from Bayes' rule by treating the parameter vector θ as a random variable. Given the data vector Y,

$$p(\theta|Y) = \frac{p(\theta)L(Y|\theta)}{p(Y)}.$$

As θ is the parameter of interest, $p(Y)$ is ignored since it contains no θ. $p(Y|\theta)$ is equivalent to the likelihood function $L(Y|\theta)$, while $p(\theta)$ is referred to as a prior. Then the former formulae can be written as

$$p(\theta|Y) \propto p(\theta)L(Y|\theta), \tag{3.4.1}$$

where $p(\theta|Y)$ is denoted as the posterior distribution for θ given the data.

A brief general discussion of such notation has been provided by Koop (2003). The prior $p(\theta)$ contains non-data information about θ before seeing the data. The likelihood function $L(Y|\theta)$ is the distribution of the data conditional on the parameters of the model, which is closely related to the data generation process. The posterior, $p(\theta|Y)$, summarizes all the information about θ after seeing the data.

From the likelihood function in Section 3.3.1, evaluation of the bivariate probit model involves conducting double normal integration. However, the idea of augmentation has provided another strategy to evaluate the bivariate probit model by reducing one integration. In stead of the posterior density of all parameters, a joint posterior distribution of the parameters and latent variables (Z) is considered. From Bayes' theorem,

$$p(\theta, Z|Y) \propto p(\theta, Z)L(Y|Z, \theta) = p(\theta)p(Z|\theta)L(Y|Z, \theta), \tag{3.4.2}$$

where $p(\theta)$ is the prior and $p(Z|\theta)$ is a function about multivariate normal density function, while it is assumed that $L(Y|Z, \theta)$ is straightforward to analyze. Based on the joint posterior distribution $p(\theta, Z|Y)$, an iterative algorithm has been derived to calculate $p(\theta|Y)$ under a data augmentation scheme in which multiple integration is avoided by sampling the latent variables. This algorithm will be discussed later in the following subsection.

According to equation (3.2.3), $Y = \{(y_{i1}, y_{i2}) : i = 1, \cdots, n\}$ contains all the binary outcomes and $\theta = (\beta_1', \beta_2', \alpha, \rho)$ includes all the model parameters. Let $Z = (Z_1, Z_2)$ where $Z_1 = \{z_{i1} : i = 1, \cdots, n\}$ and $Z_2 = \{z_{i2} : i = 1, \cdots, n\}$. Substitute equation $p(Z|\theta) = p(Z_1|\theta)p(Z_2|Z_1, \theta)$ into equation (3.4.2) to obtain

$$p(\theta, Z|Y) \propto p(\theta)p(Z_1|\theta)p(Z_2|Z_1, \theta)L(Y|Z, \theta), \tag{3.4.3}$$

where

$$p(Z_1|\theta) = \left(\frac{1}{\sqrt{2\pi}}\right)^n exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(z_{i1} - x_{i1}'\beta_1)^2\right\},$$

$$p(Z_2|Z_1,\theta) = \left(\frac{1}{\sqrt{2\pi(1-\rho^2)}}\right)^n exp\left\{-\frac{1}{2(1-\rho^2)}\sum_{i=1}^{n}(z_{i2} - x'_{i2}\beta_2 - \alpha y_{i1} - \rho z_{i1} + \rho x'_{i1}\beta_1)^2\right\},$$

$$L(Y|Z,\theta) = \prod_{i=1}^{n}\{I(z_{i1} > 0)I(y_{i1} = 1) + I(z_{i1} \leq 0)I(y_{i1} = 0)\} *$$
$$\{I(z_{i2} > 0)I(y_{i2} = 1) + I(z_{i2} \leq 0)I(y_{i2} = 0)\}.$$

Usually prior independence is assumed between β_1, β_2, α and ρ. Once the posterior density is available, Markov chain Monte Carlo (MCMC) algorithms can be used to make point estimates of parameters or any transformation of parameters, such as treatment effects.

3.4.2 Markov Chain Monte Carlo Methods

In this subsection, we will first show how to use the posterior distribution for point estimation of parameters. Secondly, the Gibbs sampling process is explained, including the use of data augmentation. Thirdly, the Metropolis-Hastings (MH) algorithm is discussed as an supplementary algorithm.

Substantial computation is involved in evaluating posterior densities $p(\theta|Y)$. As an example, suppose the mean of the posterior is used as a point estimator of θ. A integral of θ on the posterior distribution is calculated as

$$E(\theta|Y) = \int \theta p(\theta|Y)d\theta.$$

Monte Carlo integration is usually applied to obtain the integral no matter how complex the formula is. More specifically, a random sample $\theta^{(i)}$ for $i = 1, \cdots, M$ is drawn from the posterior $p(\theta|Y)$, and the average

$$\hat{\theta} = \frac{1}{M}\sum_{i=1}^{M}\theta^{(i)} \tag{3.4.4}$$

converges to $E(\theta|Y)$ as M goes to infinity from the weak law of large numbers. In general, let $f(\theta)$ be the function which is of fundamental interest. Then

$$\hat{f} = \frac{1}{M}\sum_{i=1}^{M} f(\theta^{(i)})$$

can be used as the approximation of the expectation of $f(\theta)$ with respect to the posterior.

From the above analysis, sampling from the posterior function is an essential part in the Bayesian method. In practice, the posterior distribution can have a complicated form which may make the sampling process difficult. As a result, applications of MCMC methods have been widespread in Bayesian inference. In particular, sampling based methods such as Gibbs sampling, data augmentation and MH algorithm will be discussed as they are the main methods applicable in the whole thesis.

Gibbs sampling is one of the most popular MCMC methods. The basic idea is to break $p(\theta|Y)$ into some lower dimensional conditional distributions which have standard forms such as the normal, beta, gamma or inverse-gamma. Let $\theta = (\theta_1, \theta_2, \cdots, \theta_k)$. From the Hammersley-Clifford theorem discussed by Besag (1974), the joint density $p(\theta|Y)$ is completely determined by the conditional posterior distributions $p(\theta_1|\theta_2, \theta_3, \cdots, \theta_k, Y)$, $p(\theta_2|\theta_1, \theta_3, \cdots, \theta_k, Y)$, $\cdots$, and $p(\theta_k|\theta_1, \theta_2, \cdots, \theta_{k-1}, Y)$. Since such conditional distributions can be directly sampled from standard distributions, the Gibbs sampler is implemented in the following way. Given an arbitrary initial values of $(\theta_2^{(0)}, \theta_3^{(0)}, \cdots, \theta_k^{(0)})$, for the ith iteration:

1. Draw $\theta_1^{(i)}$ from $p(\theta_1|\theta_2^{(i-1)}, \theta_3^{(i-1)}, \cdots, \theta_k^{(i-1)}, Y)$;

2. Draw $\theta_j^{(i)}$ from $p(\theta_j|\theta_1^{(i)}, \cdots, \theta_{j-1}^{(i)}, \theta_{j+1}^{(i-1)}, \cdots, \theta_k^{(i-1)}, Y)$ for $j = 2, \cdots, k-1$;

3. Draw $\theta_k^{(i)}$ from $p(\theta_k|\theta_1^{(i)}, \theta_2^{(i)}, \cdots, \theta_{k-1}^{(i)}, Y)$.

The Markov chain will be simulated after repeating the three steps several times. Once the Markov chain settles to a stationary distribution, whose states are positive recurrent, the sampler converges to the posterior $p(\theta|Y)$.

According to data augmentation, the joint posterior $p(\theta, Z|Y)$ is obtained while $p(\theta|Y)$ is the distribution one is interested in. Based on the above Gibbs sampling algorithm,

$p(\theta, Z|Y)$ is decomposed into two conditional posterior distributions. Then, the pair (θ, Z) can be generated in two steps:

1. Draw θ from $p(\theta|Z, Y)$;

2. Draw Z from $p(Z|\theta, Y)$.

The amazing part is that this mechanism automatically provides a chain of values of θ drawn from its marginal distribution $p(\theta|Y)$.

In many applications, a closed form expression is not available for some conditional posterior distributions. As a result, alternative algorithms have been devised to sample from distributions without an analytic form. Metropolis et al. (1953) first proposed the Metropolis algorithm, while Hastings (1970) provides a generalization. This generalized algorithm is referred to as the Metropolis-Hastings algorithm, and has become the most commonly used alternative to the Gibbs sampler.

Suppose one of the conditional posterior $\pi(\theta_j) = p(\theta_j|\theta_1, \cdots, \theta_{j-1}, \theta_{j+1}, \cdots, \theta_k, Y)$ has no closed form, a Gibbs sampler step will be replaced by the following two stage procedure of MH algorithm in the ith iteration:

1. Draw $\theta_j^{(i)}$ from the proposal density $q(\theta_j^{(i)}|\theta_j^{(i-1)})$;

2. Accept $\theta_j^{(i)}$ with probability $\alpha(\theta_j^{(i)}, \theta_j^{(i-1)})$,

where

$$\alpha(\theta_j^{(i)}, \theta_j^{(i-1)}) = min\left(\frac{\pi(\theta_j^{(i)})q(\theta_j^{(i-1)}|\theta_j^{(i)})}{\pi(\theta_j^{(i-1)})q(\theta_j^{(i)}|\theta_j^{(i-1)})}, 1\right).$$

For simplicity, the proposal density may follow a random walk process, $\theta_j^{(i)} = \theta_j^{(i-1)} + \varepsilon$, where ε is usually generated from a symmetric density, like the standard normal distribution or uniform distribution. Therefore, the random-walk MH algorithm becomes:

1. Draw $\theta_j^{(i)}$ from $\theta_j^{(i)} = \theta_j^{(i-1)} + \varepsilon$;

2. Accept $\theta_j^{(i)}$ with probability $\alpha(\theta_j^{(i)}, \theta_j^{(i-1)})$,

where

$$\alpha(\theta_j^{(i)}, \theta_j^{(i-1)}) = min\left(\frac{\pi(\theta_j^{(i)})}{\pi(\theta_j^{(i-1)})}, 1\right).$$

Gibbs sampling is a special case of MH algorithm where the proposal density is just the posterior distribution and the acceptance probability is always one. That is why Metropolis-Hastings and Gibbs can be mixed together to generate samples from the joint posterior distribution.

3.4.3 Sampler for the Endogenous Bivariate Probit Model

With that general introduction to the key Bayesian sampling techniques, we now turn to the model of interest in this chapter. First, the conditional posteriors of z_{i1} and z_{i2} are considered. From the joint posterior function,

$$\begin{aligned} p(z_{i1}|\beta,\rho,z_{i2},y_{i1}) \propto \quad & exp\left\{-\frac{1}{2(1-\rho^2)}\left[z_{i1}-(x'_{i1}\beta_1+\rho z_{i2}-\rho x'_{i2}\beta_2-\rho\alpha y_{i1})\right]^2\right\}* \\ & \{I(z_{i1}>0)I(y_{i1}=1)+I(z_{i1}\leq 0)I(y_{i1}=0)\} \end{aligned}$$

which is a univariate normal distribution with mean $x'_{i1}\beta_1+\rho z_{i2}-\rho(x'_{i2}\beta_2+\alpha y_{i1})$ and variance $1-\rho^2$ truncated to the region B_{i1} where B_{i1} is the interval $(0,\infty)$ if $y_{i1}=1$ and the interval $(-\infty,0]$ if $y_{i1}=0$.

$$\begin{aligned} p(z_{i2}|\beta,\rho,z_{i1},y_{i2}) \propto \quad & exp\left\{-\frac{1}{2(1-\rho^2)}(z_{i2}-x'_{i2}\beta_2-\alpha y_{i1}-\rho(z_{i1}-x'_{i1}\beta_1))^2\right\}* \\ & \{I(z_{i2}>0)I(y_{i2}=1)+I(z_{i2}\leq 0)I(y_{i2}=0)\} \end{aligned}$$

which is also a univariate normal distribution with mean $x'_{i2}\beta_2+\alpha y_{i1}+\rho(z_{i1}-x'_{i1}\beta_1)$ and variance $1-\rho^2$ truncated to the region B_{i2} where B_{i2} is the interval $(0,\infty)$ if $y_{i2}=1$ and the interval $(-\infty,0]$ if $y_{i2}=0$.

In order to draw a parameter from a univariate truncated normal distribution, Geweke (1991) has discussed the classical c.d.f. inversion technique as follows. Suppose that x is a truncated normal (TN) random variable with location μ, scale σ and truncation $a<x<b$. Then we have $\frac{x-\mu}{\sigma}\sim TN(\frac{a-\mu}{\sigma},\frac{b-\mu}{\sigma})$. To draw x, we need to draw $u\sim U[\Phi(\frac{a-\mu}{\sigma}),\Phi(\frac{b-\mu}{\sigma})]$ which is a uniform distribution. As a result, we use the following three steps to obtain a random number from the truncated normal distribution.

1. Draw U from a uniform (0,1) random variable.
2. Calculate $u=\Phi\left(\frac{a-\mu}{\sigma}\right)+U\left(\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)\right)$.
3. Calculate $x=\mu+\sigma\Phi^{-1}(u)$.

However, Robert (1995) suggests this c.d.f. inversion technique may be quite inefficient when $a-\mu$ is large and he also gives an exponential accept-reject algorithm. Here we only introduce one-sided truncation assuming without loss of generality that $\mu = 0$ and $\sigma^2 = 1$. When $a > 0$, the optimal exponential accept-reject algorithm is given by

1. Calculate the optimal scale factor $\lambda = \frac{a+\sqrt{a^2+4}}{2}$;

2. Draw $z \sim Exp(\lambda, a)$ which is the translated exponential distribution with density $f(z) = \lambda exp(-\lambda(z-a))I_{z\leq a}$, where I is an indicator function;

3. Compute $r = exp\left\{-(z-\lambda)^2\right\}$;

4. Draw u from a uniform (0,1) random variable.

5. Take $x = z$ if $u \leq r$; otherwise go back to step 2.

When $a \leq 0$, we use normal rejection sampling, in which x is drawn from N(0,1) and accepted if $x > a$. In our case, we need to generate $z \sim N(\mu, \sigma^2)I_{z\leq 0}$ when $y = 1$. Here we draw $x \sim N(0,1)I_{x\leq -\frac{\mu}{\sigma}}$ first according to the algorithms mentioned above and get z from $z = \mu + \sigma * x$. When $y = 0$, we first draw $x \sim N(0,1)I_{x\leq \frac{\mu}{\sigma}}$, then return z from $z = \mu - \sigma * x$.

Second, given independent priors $p(\beta_1) \sim N(\beta_0, B_0^{-1})$ and $p((\beta_2', \alpha)') \sim N(\beta_*, B_*^{-1})$, conditional posteriors of β_1, β_2 and α are obtained as follows:

$$
\begin{aligned}
f(\beta_1|\beta_2,\rho,Z) \propto \quad & exp\left\{-\tfrac{1}{2}(\beta_1-\beta_0)'B_0(\beta_1-\beta_0)\right\} * exp\left\{-\tfrac{1}{2}\textstyle\sum_{i=1}^n (x_{i1}'\beta_1 - z_{i1})^2\right\} * \\
& exp\left\{-\frac{1}{2(1-\rho^2)}\textstyle\sum_{i=1}^n (z_{i2} - x_{i2}'\beta_2 - \alpha y_{i1} - \rho z_{i1} + \rho x_{i1}'\beta_1))^2\right\},
\end{aligned}
$$

$$
\begin{aligned}
f(\beta_2,\alpha|\beta_1,\rho,Z) \propto \quad & exp\left\{-\tfrac{1}{2}((\beta_2',\alpha)'-\beta_*)'B_*((\beta_2',\alpha)'-\beta_*)\right\} * \\
& exp\left\{-\frac{1}{2(1-\rho^2)}\textstyle\sum_{i=1}^n (z_{i2} - x_{i2}'\beta_2 - \alpha y_{i1} - \rho(z_{i1} - x_{i1}'\beta_1))^2\right\}.
\end{aligned}
$$

These are Gaussian densities with

$$\beta_1|\beta_2,\rho,Z \sim N(\hat{\beta_1}, B_1^{-1})$$

where $\hat{\beta_1} = B_1^{-1}\left[B_0\beta_0 + \frac{1}{1-\rho^2}\sum_{i=1}^n x_{i1}(z_{i1} + \rho x_{i2}'\beta_2 + \rho\alpha y_{i1} - \rho z_{i2})\right]$ and $B_1 = B_0 + \frac{1}{1-\rho^2}\sum_{i=1}^n x_{i1}x_{i1}'$;

$$\beta_2,\alpha|\beta_1,\rho,Z \sim N(\hat{\beta_2}, B_2^{-1})$$

in which $\hat{\beta_2} = B_2^{-1}\left[B_*\beta_* + \frac{1}{1-\rho^2}\sum_{i=1}^{n}(x'_{i2}, y_{i1})'(z_{i2} + \rho x'_{i1}\beta_1 - \rho z_{i1})\right]$ and $B_2 = B_* + \frac{1}{1-\rho^2}\sum_{i=1}^{n}(x'_{i2}, y_{i1})'(x'_{i2}, y_{i1})$.

Finally, the conditional posterior of ρ is analyzed given the prior $\pi(\rho)$,

$$f(\rho|\beta, Z, Y) \propto \pi(\rho)(1-\rho^2)^{-\frac{n}{2}} exp\left\{-\frac{1}{2(1-\rho^2)}\sum_{i=1}^{n}\left[\rho(z_{i1} - x'_{i1}\beta_1) + x'_{i2}\beta_2 + \alpha y_{i1} - z_{i2}\right]^2\right\}$$

where $\rho \in (-1, 1)$. A transformation is used here by letting $\rho^* = ln\left(\frac{1+\rho}{1-\rho}\right)$ and $\rho^* \in (-\infty, \infty)$. Thus, $\rho = \frac{exp(\rho^*)-1}{exp(\rho^*)+1}$ and the Jacobian term is $(1-\rho^2)$. Then unconstrained ρ^* is sampled instead of ρ from the posterior

$$\begin{aligned} f(\rho^*|\beta, Z, Y) \propto & \quad (1-\rho^2)^{1-\frac{n}{2}}\pi(\rho) exp\left\{-\tfrac{1}{2(1-\rho^2)}\textstyle\sum_{i=1}^{n}\left[\rho(z_{i1} - x'_{i1}\beta_1) + x'_{i2}\beta_2 + \alpha y_{i1} - z_{i2}\right]^2\right\} \\ = & \quad \left\{\tfrac{4exp(\rho^*)}{[exp(\rho^*)+1]^2}\right\}^{1-\frac{n}{2}} * \pi\left[\tfrac{exp(\rho^*)-1}{exp(\rho^*)+1}\right] * \\ & \quad exp\left\{-\tfrac{[exp(\rho^*)+1]^2}{8exp(\rho^*)}\textstyle\sum_{i=1}^{n}\left[\left(\tfrac{exp(\rho^*)-1}{exp(\rho^*)+1}\right)(z_{i1} - x'_{i1}\beta_1) + x'_{i2}\beta_2 + \alpha y_{i1} - z_{i2}\right]^2\right\}. \end{aligned}$$

Because this density is not standard, random walk MH algorithm is applied in sampling ρ^* which will then be transformed back to ρ.

3.4.4 Simulation Inefficiency Factor

For estimation to work well, it is necessary that both the Markov chain and the MCMC algorithm converge. Since most Markov chains produced by the MCMC method should converge geometrically to the stationary posterior distribution (Roberts 1996), the convergence performance of our MCMC algorithm will be considered here.

The sequence sampled from the posterior $p(\theta|Y)$ is denoted by $\left\{\theta^i : i = 1, 2, \cdots, M\right\}$. The point estimation of parameters $\hat{\theta}$ is the ergodic average calculated in equation (3.4.4). The central limit theorem of ergodic averages shows:

$$\sqrt{M}\left(\hat{\theta} - E(\theta|Y)\right) \rightarrow N(0, \sigma_f^2)$$

for some positive constant σ_f, as $M \rightarrow \infty$. However, algorithms can also be inefficient, if σ_f is too large comparing to the variance of θ under posterior density.

Roberts (1996) summarizes a well-known batch mean method to measure the efficiency of the Markov chain for estimating targeted distribution. The idea of batch mean is also applied in another statistic named simulation inefficiency factors(SIF) which provides another way to evaluate the convergence of the Markov Chain and Gibbs sampler, introduced by Kim et al. (1998). This index

$$SIF = \frac{\sigma_f^2}{var(\theta)}$$

is used to measure the efficiency of the Markov chain for estimating $E[\theta|Y]$. Here $var(\theta)$ is the variance of the sample mean with independent draws from the posterior, using the formula

$$var(\theta) = \frac{1}{M-1}\Sigma_{i=1}^{M}\left[\theta^i - \hat{\theta}\right]^2.$$

In order to estimate σ_f^2, Roberts(1996) discusses a batch mean method as follows. Let

$$\theta_k = \frac{1}{n}\Sigma_{i=(k-1)n+1}^{kn}\theta^i$$

for $k = 1, 2, \cdots, m$. Therefore σ_f^2 can be estimated by (Tse et al. 2004)

$$\hat{\sigma}_f^2 = \frac{n}{m-1}\Sigma_{k=1}^{m}(\theta_k - \overline{\theta})^2.$$

This estimator should perform well for sufficiently large n, where $M = m*n$. This SIF will be used as an important index to evaluate the convergence rate of the MCMC algorithm. A smaller value of SIF usually indicates better convergence.

Notice that none of the available methods can guarantee 100% that the Gibbs sampler under study has converged for all applications (Tsay 2005), so one must take care in a real application to ensure that there is no obvious violation of the convergence requirement. The most straightforward way to check if the chain is stationary is to view the sampled path as well as its auto correlation function(ACF). This is a very useful approach for single cases, but too burdensome for simulation work with large amount of iterations. As a result, SIF is the main index to evaluate the convergence rate in this thesis.

3.5 General Monte Carlo Design

A Monte Carlo experiment is reported here whose aim is to test the effects of exclusion restrictions on bivariate probit models with an endogenous dummy regressor estimating the parameters via structural and partial structural systems in different circumstances. Structural equations are derived from economic theory and each equation purports to describe a particular aspect of the economy or individual behaviors (Greene 2003). We will use the term structural model to refer to the case in which x_{i1} does not include the same set of variables as x_{i2}. MLE applied on this system is usually named full information maximum likelihood (FIML) estimation. In the Monte Carlo design, x_{i2} has a constant and one explanatory variable. In the full structural system, we set up the equations with the same number of variables by including a different variable in the first equation, so $x_{i1} \neq x_{i2}$.

When the equation for y_1 is not our main concern, it is common to estimate a reduced form equation for y_1. In this case, all variables in x_{i2} are also included in x_{i1}. This system is called a partial structural system, as it includes a structural equation for y_2 and a reduced form equation for y_1. If x_{i1} has at least one extra variable not in x_{i2}, it becomes a reduced form equation with exclusion restrictions. Section 3.6 reports the Monte Carlo results with a structural equation system, while Sections 3.7 and 3.8 discuss the results for partial structural models.

In any given sample there is some possibility that optimization methods fail to find the maximum of the likelihood function and this possibility may be quite large especially when exclusion restrictions are not available. Although this probability becomes small as sample size becomes arbitrarily large (Heckman 1978, p950), there are various circumstances where sample size is limited. Thus, a moderate sample size 1,000 is adopted in all the following experiments as it is not small and also not large enough to eliminate certain finite sample properties.

For simplicity, each independent variable x_{i1} and x_{i2} contains one constant and no more than two variables. Hence, β_1 and β_2 are 2×1 or 3×1 vectors. In empirical work, explanatory variables can be continuous or discrete, and dummy variables are especially in common use. For this reason, some simulations contain only continuous random variables in x_{i1} and x_{i2}, some with only Bernoulli distributions and some with a mixture of continuous and discrete dummy variables. The disturbances u_i and v_i are generated from standard bivariate normal distributions with correlation ρ which is chosen to be 0, 0.5 and 0.8 respectively.

Once true values of parameters are chosen, 1,000 samples are generated based on each specific model. Three estimation methods are used to estimate each sample. For each method, statistics for different sets of samples are summarized and then compared across methods in order to see which method provides the best estimates. MLE is applied here to estimate the model since it is the most frequently used estimation method. And two alternative approaches, the Bayesian method and two-step method, will also be included to allow comparison with MLE.

Optimization in both MLE and two-step methods is obtained using the CML package in GAUSS 8.0. Most starting values are set to zero for all three methods. Occasionally other starting values are used to check whether estimator is sensitive to different starting points. After that, BFGS (Broyden, Fletcher, Goldfarb, Shanno) method is selected as optimization method, but DFP (Davidon, Fletcher, Powell), NEWTON (Newton-Raphson) and BHHH (Berndt, Hall, Hall, Hausman) methods are applied as alternative algorithms where the function cannot be evaluated at initial parameter values. At the same time, the inverse of the Hessian is calculated to compute covariance matrix of parameter estimates.

The code for Bayesian estimation is programmed using GAUSS 8.0 according to the MCMC samplers specified in Section 3.4.3, with the zero vector as starting values. Based on the MCMC algorithms, ρ is sampled using the random-walk MH algorithm and the acceptance rate is always controlled to be between 0.2 and 0.3. 1,000 samples are discarded as burn-in period and 10,000 replications are recorded after the burn-in period. The mean values of the 10,000 replications are then treated as the point estimates of the model parameters.

While three methods are available to estimate the parameters for the 1,000 samples, not all methods produce estimates for each sample. According to MLE process, it is most likely that the estimation is no longer accurate when Hessian matrix fails to invert in program execution. The number of occurrences of inaccurate estimation has been recorded together with the cases that MLE fails to converge and results in the programme terminating directly. To make for valid comparisons of accuracy of the three methods, results of the estimation using MCMC method and two-step method are also summarized with only those samples which can be estimated by MLE. Such results of partial samples are respectively represented by MCMC* and TS* in each table. The MCMC approach never fails to produce estimates, even when MLE fails in estimating certain data sets. So summary statistics for the MCMC method are also calculated based on all samples. Very occasionally, the Hessian matrix fails to invert in a single equation probit estimation. Thus, results of successful estimation of the two-step method are summarized and denoted by TS**.

3.6 Structural Equation Estimation

To start with, we generate data with $x_{i1} \neq x_{i2}$, namely, explanatory variables in both x_{i1} and x_{i2} are totally different. In this case, even the linear system described in Section 3.2.2 can be totally identified. As a result, experiments are all designed without identification problems in this section. Since estimation may vary depending on the types of explanatory variables, different data generation process are considered including generation of continuous random numbers and binary variables which are the most frequently used discrete variables. Continuous explanatory variables are generated from standard normal distributions in Table 3.1, 3.2 and 3.3 with different ρ values, while binary variables are generated from Bernoulli densities with success probability 0.7 in Table 3.4, 3.5 and 3.6. Only one explanatory variable is considered in each equation for computational simplicity. Thus, six parameters are to be estimated, namely $\beta_1 = (\beta_{11}, \beta_{12})$, $\beta_2 = (\beta_{21}, \beta_{22})$, α and ρ.

We will first contrast results for different types of data generation processes with continuous data in Table 3.1 and with discrete data in Table 3.4, both without correlation between errors. Mean values suggest very limited bias exists in all estimators, but standard deviations in Table 3.4 are a little larger than those in Table 3.1. Because mean square errors are the sum of variance and square of bias, larger standard deviations will result in bigger mean square errors in Table 3.4 when biases are similar. In terms of mean absolute errors (MAEs), they are also relatively larger in Table 3.4 than in Table 3.1. In addition, Table 3.4 contains less extreme values of β_1 but more extreme values of β_2, α and ρ. These results are consistent with the mixing performance of the MCMC method: SIF values of β_1 are smaller and SIF values of β_2, α and ρ are larger in Table 3.4. Similar patterns can be found in the other four tables when comparing across types of X data with the same ρ values. These small differences may be caused by the difference in degree of variation in the explanatory variables. To confirm this, other experiments have been done with continuous explanatory variables generated from normal distribution with variance 0.21 which is also the variance of the Bernoulli distribution we use here. Tables of such experiments are not shown in the thesis to avoid too many tables. The results indicate that all estimators perform a little worse with variance 0.21 than those with unit variance, which is in our expectation as smaller variance means less information in X variables. The bottom line is that the difference in results whether the explanatory variables are continuous or dummy is relatively small, when comparing both specifications with the same variance 0.21. Thus, the following discussion will focus on the continuous cases in Table 3.1, 3.2 and 3.3.

Now consider the issue around when estimators fail to converge. MLE fails to converge in some cases and it is more likely to fail when there is less correlation between error terms, as 823, 905 and 935 are the number of samples with converged estimations in 1,000 iterations with $\rho = 0, 0.5, 0.8$ respectively. If comparing MLE and the MCMC method, each based on different samples (e.g. 823 and 1000 when $\rho = 0$), differences in performance may be due to the fact that they use different samples. To overcome this problem, MCMC is applied to just the same 823 samples as MLE, with results indicated by MCMC* in Table 3.1. Note that results for MCMC* vs MCMC are virtually the same in each table, so the difference between MLE and MCMC approaches is not due to their use of different samples. For the same reason, TS* is the notation of estimation of the two-step method with same samples as MLE. Because two step method also fails in some cases which may be different from those when MLE fails, TS** means the estimation of the samples when it does not fail. The two step method tends to fail much less often than MLE because it is a combination of estimation of two univariate probit models. For instance, it only fails 56 times in Table 3.1 and never fails in Table 3.2 and 3.3. Once again, results of TS* vs TS** look almost the same, although they are a little different for some specific measures.

Now differences in results as ρ varies will be discussed. We will start with bias and precision analysis by comparing mean values across three methods. Mean values from MLE estimation are quite accurate for all values of ρ with tiny bias and small standard deviations. The MCMC method also provides precise estimation although there is a small bias compared to MLE, especially in estimating ρ. Estimation of ρ is more difficult than estimation of other parameters, as shown by the mean SIF value for ρ, which is much larger than for other parameters in each table. The mean absolute errors (MAEs) and mean squared errors (MSEs) of the two estimators are very close in each table. At the same time, estimates using the two-step estimator are almost the same compared to other methods when there is no correlation. When $\rho = 0.5$, the performance of the two-step method is still acceptable, although mean values of β_2, α and ρ show more bias and standard deviations of such parameters are a little larger than other methods. Some maximum values are relatively extreme, e.g. the maximum ρ value is 0.927 comparing to 0.834 obtained by MLE. As a result, the two-step method tends to have larger MAEs and MSEs than other methods. When ρ increases to 0.8, bigger problems appear using this inconsistent estimator. More specifically, estimates of β_2, α and ρ are too big, with biases of at least 0.1 in the second step. Furthermore, larger standard deviations reveal that this method is quite inefficient with quite large MAEs and MSEs. Take ρ as an example. The mean of estimates of ρ using TS* shows bias of 0.107 when the true value is 0.8, with standard deviation 0.096 which is larger than 0.064 for the MLE estimator. The maximum and minimum values are also a little more extreme than others, and the mean absolute error is twice as big as MLE while mean square error is five times that of MLE. This bigger MAEs and MSEs are due mostly to bias, as there is little difference between the standard deviations. But note the two-step method always converges for $\rho = 0.8$, while MLE has a 6.5% failure rate.

As well as standard deviation, the mean absolute error and mean square error are treated as measures of efficiency, while minimum and maximum values are used as measures of robustness. Overall, the values of standard deviation, mean absolute error and mean square error are decreasing as ρ increases. When $\rho = 0$, standard deviations of β_{21}, α and ρ are relatively larger than that of other parameters. When $\rho = 0.8$, such differences are not as obvious as standard deviations of β_{21}, α and ρ fall dramatically while these for the other parameters drop a little. In the mean time, MAEs and MSEs of β_{21}, α and ρ are also bigger than that of other parameters. Based on maximum and minimum values, extreme values are not obvious, because the most extreme ones can easily result in failed inversion of the Hessian matrix and then cannot be counted in the successful estimations. When correlation is stronger, mixing performance of the MCMC algorithm becomes worse as the SIF values get bigger and bigger.

We summarize by focusing on difference between methods. The difference between the precision of MLE and MCMC method is quite small no matter how large ρ is. Although MLE occasionally fails to converge, it provides slightly better precision in estimation with relatively less computation comparing to MCMC. Considering its computational complexity, MCMC method can be used as the alternative estimation when MLE cannot estimate certain data sets. The two-step method should only be an alternative choice when error correlation in the model is not too strong.

3.7 Partial Structural Equation Estimation with Exclusion Restrictions

Moving on from the structural equation system, partial structural equation systems with $x_{i2} \subset x_{i1}$ will be considered in this section. In other words, we estimate the y_1 equation in reduced form, where exclusion restrictions will be present when x_{i1} contains variables which are not included in x_{i2}. As in the previous section, explanatory variables are generated from both continuous and discrete densities. Once again, Table 3.7, 3.8 and 3.9 present results with continuous exogenous variables while Table 3.10, 3.11 and 3.12 include binary exogenous variables only. For simplicity, only a constant and one other variable are included in x_{i2}. Because of exclusion restrictions, one more variable is included in x_{i1}, so one more parameter β_{13} is estimated in the six tables.

For the same ρ value, the results suggest that parameters are more difficult to estimate when regressors are binary rather than continuous. This may be caused by smaller variance of dummy independent variables. Take Table 3.7 and 3.10 as examples. Obviously, Table 3.10 has relatively larger standard deviations, MAEs and MSEs than Table 3.7. It also has much bigger mean SIF values of β_2, α and ρ but smaller mean SIF values of β_1. Once again, the relative performance of three estimators does not vary with the type of explanatory variables. So the following discuss will focus on the tables with continuous explanatory variables.

There are few differences when statistics are summarized according to the different number of samples for both MCMC and the two-step method. For example, in Table 3.7, reliable estimates can be obtained by MLE for 785 samples. Statistics for these 785 samples represented by MCMC* reveal minor differences to statistics for all 1000 samples by MCMC. Except α, mean MCMC values are a little less biased than mean MCMC* values, while MCMC standard deviations are all a bit smaller than that of MCMC*. So the MAEs and MSEs of MCMC will be smaller than that of MCMC*. Note the differences are very small. According to results of the two-step method, the 943 reliable two-step estimates produce less bias and smaller standard deviations than the estimates of 785 samples in Table 3.7. However, such differences in results between samples are eliminated as ρ increases. Although mean and standard deviations for different sample sizes are not identical, MAEs and MSEs are quite similar in Table 3.8, while Table 3.9 shows only tiny differences between statistics with different samples.

We now compare the performance of three estimation methods as ρ changes. In Table 3.7, MLE is always accurate and reasonably efficient except for convergence problems with certain samples. Nevertheless, the Bayesian method has slightly more accurate estimates than MLE as shown by smaller standard deviations, while the two-step method gives similar outputs to MLE. The Bayesian method is able to provide reliable estimation on all samples, while the two-step method has some convergence problems as results can be obtained for 943 samples. When $\rho = 0.5$, mean values of MCMC show a larger bias than MLE but has it similar standard deviations. At the same time, the two-step method has greater bias than the others, larger MAEs and larger MSEs for β_2, α and ρ. When ρ is 0.8, mean values of MCMC show more biased than MLE, with similar variation. The mean absolute errors and mean square errors reveal the Bayesian method is as efficient as MLE except for estimating ρ, while the two-step method is not as efficient in estimating β_2, α and ρ.

In summary, all three methods perform well in estimating the partial structural equation model when exclusion restrictions exist. MLE usually returns unbiased and efficient estimates. In contrast, the Bayesian approach can offer quite good results even when MLE fails to converge. In addition, MCMC provides more efficient estimation than MLE when the error correlation is small, but cannot give good estimation of ρ when the correlation is strong. While the two-step method fails to converge much less frequently than MLE, and is as efficient as MLE when $\rho = 0$, it becomes very unreliable for larger ρ values.

3.8 Partial Structural Equation Estimation without Exclusion Restrictions

Now, we consider the partial equation system with $x_{i1} = x_{i2}$, in other words, in a model with no exclusion restrictions. Even though exclusion restrictions are not required in theory, it is likely to make estimation more difficult. When there are no exclusion restrictions, it turns out that estimators perform somewhat differently according to whether the explanatory variables are continuous or dummy and how large their variance is. So the model with one exogenous variable in each equation has been designed in the next three subsections, with three further subsections discussing a model with two explanatory variables in each equation. More specifically, Table 3.13, 3.14 and 3.15, discussed in Section 3.8.1, show estimation results with explanatory variables generated from standard normal densities. Such tables are then compared with Table 3.1, 3.2 and 3.3 to show the effect of exclusion restrictions on models with continuous explanatory variables. Next, Section 3.8.2 gives examples of one dummy variable in each equation which are indicated in Table 3.16, 3.17, 3.18, 3.19, 3.20 and 3.21. Results for models with one continuous variable sampled from normal distribution with variance 0.21 are given in Section 3.8.3. Estimation results for such models are displayed in Table 3.22, 3.23 and 3.24 and are compared with results from Section 3.8.2. A model with two binary exogenous variables in each equation is discussed in Table 3.25, 3.26, 3.27, 3.28, 3.29 and

3.30 to check if an increase in the number of explanatory variables will affect identification conditions. More complicated models with combinations of continuous and dummy variables are considered in Section 3.8.5 and Section 3.8.6.

3.8.1 Models with One Continuous Explanatory Variable (Unit Variance)

In this subsection, Table 3.13, 3.14 and 3.15 are compared with Table 3.1, 3.2 and 3.3 because they have exactly the same number of parameters and all explanatory variables are generated from standard normal distributions. The only difference is that the first Table 3.1, 3.2 and 3.3 are obtained when $x_{i1} \neq x_{i2}$, while Table 3.13, 3.14 and 3.15 are with $x_{i1} = x_{i2}$.

Firstly, the tables are compared across parameters. Estimation of β_1 remain almost the same with or without exclusion restrictions for all three methods, when mean values of β_1 are close to true values and standard deviations are all controlled to be less than 0.1. However, without exclusion restrictions, it becomes much more difficult to estimate β_2, α and ρ with all three estimators. Extreme values appear frequently, which results in standard deviations that are double those previously, as well as larger MAEs and MSEs. For instance, α values estimated by MLE vary from -2.324 to -0.081 with standard deviation 0.406 in Table 3.13, comparing to smaller range from -1.713 to -0.721 with smaller standard deviation 0.155 in Table 3.1. Meanwhile, the mean absolute error of α increases from 0.125 to 0.335 while the mean square error jumps from 0.024 to 0.165.

Similar patterns can also be found in the mixing performance of MCMC algorithms which is significantly different between the models with and without exclusion restrictions. Mean SIF values for β_1 remain quite small, always less than 50 even when the true ρ value is large. However, mean SIF values of β_2, α and ρ have increased dramatically in the model without exclusion restrictions. For example, when $\rho = 0$, mean SIF values of β_2 and α in Table 3.13 are at least double those in Table 3.1, while the mean SIF value of ρ has also doubled in Table 3.13. Such slow mixing performance is most likely caused by the weak identification that exists when the model is estimated without exclusion restrictions.

Next, we will check the impact of exclusion restrictions with continuous independent variables, comparing across methods. MLE tends to fail a little more frequently, when the numbers of available estimates are less than 800 (742, 759, 779 respectively) out of 1000, and much smaller than the cases with $x_{i1} \neq x_{i2}$. But MLE still gives the smallest biases, regardless of the strength of error correlation. Although mean values of β_{21}, α and ρ estimated by the MCMC approach in Table 3.13, 3.14 and 3.15 are biased from true values more than the two other methods, biases are still quite small. Meanwhile, standard deviations obtained by MCMC are somehow smaller than that obtained by MLE and the two-step method in Table 3.13 and 3.14, so are the mean absolute error and mean square error. When $\rho = 0.8$, standard deviations obtained by MCMC are a little bit bigger than MLE but smaller than two-step method, while the mean absolute errors and mean squared errors are relatively larger than for the other two methods except that of β_{22}. The two-step method performs quite well in the model without exclusion restrictions. Firstly, it fails to converge much less often than MLE with 944, 989 and 998 available estimations for the three ρ values respectively. Secondly, it provides quite small biases even when $\rho = 0.8$, although standard deviations, MAEs and MSEs of β_2, α and ρ are relatively larger than the other methods. However, maximum and minimum values are more extreme especially for ρ, so the two-step method can give quite unrealistic estimates occasionally.

To summarize, one continuous variable in each equation of the bivariate endogenous probit model will be enough for identification even if $x_{i1} = x_{i2}$, although the identification is somehow weaker than the case with exclusion restrictions. MLE can obtain reliable estimation except that it often fails to converge. Bayesian method is the most efficient when the correlation is not high and it is also able to estimate all samples. In addition, the two-step method has less convergence issues than MLE and is quite good choice with low correlation. Both Bayesian and two-step methods are not quite reliable when the correlation is strong. But more frequent failure of convergence of MLE and much slower mixing performance of MCMC reveal that exclusion restrictions can be important to improve estimator performance.

3.8.2 Models with One Binary Explanatory Variable

When it comes to models in which x_{i1} and x_{i2} are binary and exclusion restrictions are not imposed, MLE and two-step approaches have quite serious convergence problems and become very sensitive to starting values. Therefore, models are estimated with two different sets of starting values. Table 3.16, 3.17 and 3.18 are estimated results with zero starting values for all parameters using the three methods. Then in Table 3.19, 3.20 and 3.21, starting values of β_1, β_2 and α are (0.892, -1.452, 0.738, -0.943, -1.166) chosen from the estimation of a univariate probit model for one specific sample, while 0 is the starting value for ρ as it is constrained between -1 and 1.

Results of MLE and two-step methods are discussed with different starting values. When starting values are zero, less than two thirds of samples can be estimated by MLE. Only 497, 465 and 610 samples can be obtained by MLE. Furthermore, both MLE and the two-step method cannot provide accurate estimation of β_2, α and ρ at all no matter what value ρ takes, although estimation of β_1 is still fine. Mean values are biased a lot from true values with huge standard deviations, MAEs and MSEs. When starting values are close to the true values, MLE does not show much bias, although standard deviations are relatively large compared to the ones in Table 3.13, 3.14, and 3.15 with continuous variables. Both MLE and the two-step method are less likely to fail to converge when improved starting values are used. In Table 3.19, 3.20 and 3.21, numbers of available samples are limited to 541, 649 and 769 for MLE which is fewer than with continuous explanatory variables. At the same time, two-step method converges only 590, 734 and 875 times, still smaller than continuous cases. It gives unbiased estimation when $\rho = 0$ but mean values indicate bias when $\rho = 0.5$ and 0.8. This method always has larger standard deviations, more extreme values, bigger MAEs and MSEs of β_2, α and ρ comparing to MLE.

The MCMC approach is hardly influenced by starting values when the burn-in period is long enough. The mixing performance of the Bayesian method is worse than continuous cases as shown by the large SIF values of β_2, α and ρ. When there is no correlation in the model, MCMC can provide relatively accurate estimation, see Table 3.16 and 3.19. But it cannot give unbiased estimation at all with different starting values when $\rho = 0.5$ or 0.8.

The kind of results reported here may be caused by a relatively flat or multimodal likelihood function or very small variance of independent variables. For a flat or multimodal likelihood, on one hand, MLE may converge to one of the points with local maximum likelihood value which varies depending on the starting values. On the other hand, MCMC walks along all the possible points of the likelihood, so the average of the sampling path will be somewhere between the points. Sivia & Skilling (2006) give an example of a bimodal posterior distribution $p(X)$ with modes around $X = 10$ and $X = -10$. The problem is that the expectation of the posterior distribution of X is around 0, a value which the posterior indicates is very improbable. Another possible explanation for the poor performance is with the lack of variability in explanatory variables. Notice that the dummy explanatory variables are generated from Bernoulli distribution with success possibility 0.7, so its variance is 0.21 which is rather small comparing to the variance of standard normal distribution used in the continuous case. Consequently, information contained in explanatory variables is much weaker here when they are dummy variables, compared to the continuous case that was used in Section 3.8.1.

3.8.3 Models with One Continuous Explanatory Variable (Variance 0.21)

It may be argued that the difference in previous two subsections is mainly a result of different amount of signal in the explanatory variables and not related to whether the variables are discrete or continuous. More variation in explanatory variables indicates more information about the relationship between the regressand and regressors, which will result in more precise estimation. Similar comments about effects of variation in explanatory variables have been discussed in Section 3.6 when comparing Table 3.1 and Table 3.4. To deal with this issue, three other experiments are designed with continuous explanatory variables which are generated from normal distributions with variance 0.21. This variance is exactly the same variance used in generating the dummy regressors in Section 3.8.2. Results estimated with zero starting values are shown in Table 3.22, 3.23 and 3.24.

Comparing those tables with the ones in Section 3.8.1, it it clear that smaller variance in the explanatory variables will lead to less precise and less efficient estimation of most parameters except sometimes in estimating β_{11} or β_{12}. Take Table 3.15 and 3.24 as examples. On one hand, estimated mean values of β_1 estimates in 3.15 are similar to the ones in Table 3.24. But mean values of other parameters are closer to true values in Table 3.15 than in Table 3.24. On the other hand, standard deviations, mean values of absolute error and mean values of square error of all parameters except β_{11} are smaller in Table 3.15 than that in Table 3.24 for all estimation methods. In addition, the number of successful estimations of MLE is at least 10% more in Table 3.15 than in Table 3.24, so MLE has more trouble in convergence when the explanatory variable has less information.

Next, we will compare the models of one continuous independent variable in each equation with the models of one binary independent variable in each equation, when such variables have variance 0.21. Above all, zero starting values no longer cause too many troubles in estimations when explanatory variables are continuous in comparison with dummy regressors. Thus, we will ignore those results with zero starting values in Table 3.16-3.18, and compare Table 3.19, 3.20 and 3.21 with Table 3.22, 3.23 and 3.24. When comparing Table 3.19 with Table 3.22 with $\rho = 0$, mean values are all close to true values, although standard deviations of β_{12}, β_{21}, in Table 3.19 are smaller than in Table 3.22. The differences are relatively small, especially when Table 3.20 and 3.21 are compared with Table 3.23 and 3.24. Investigating the six tables across methods, MLE and two-step methods can provide quite good estimation, while results of MCMC are much less biased with a continuous independent variable than with a dummy variable. More specifically, mean values of MLE and two-step methods are all around the true values in the six tables. Estimated mean values of MCMC show some bias when error correlation is strong in Table 3.23 and 3.24, but such bias is not as serious as the one in Table 3.20 and 3.21. When $\rho = 0$ or 0.5, two-step method provides the greatest standard deviations, MAEs and MSEs. Meanwhile, MCMC is the most efficient estimator as it has the smallest standard deviations, MAEs and MSEs. When $\rho = 0.8$, two-step method has the largest standard deviations, but MCMC gets biggest MAEs and MSEs among the three methods.

In this setting, small variance in explanatory variables is not the only reason causing identification problems, when larger variance will enhance estimation precision and efficiency in some ways. Moreover, continuous regressors are less likely to lead estimation difficulties than discrete explanatory variables.

3.8.4 Models with Two Binary Explanatory Variables

Section 3.8.2 reveals that it can be difficult to estimate precisely the parameters of a bivariate probit model with endogeneity, when the model has a single binary explanatory variable. In this section, two binary explanatory variables are introduced into each equation to check how changing the number of dummy variables affects the estimators. Once again, three tables are created with zero starting values, while three other tables show results with starting points close to true values which are also MLE estimates on univariate probit models assuming no cross equation correlation. Statistics of estimation of β_1 values are always similar according to all three methods, so the following analysis will concentrate on other parameters.

We will start with Table 3.25 and compare across methods when there is no error correlation. Mean values of β_{21}, β_{22} and α estimated by MLE are quite biased from the true values while their standard deviations, MAEs and MSEs are relatively large. Especially for α, the estimated mean is -0.782 with standard deviation 0.703 when the true value is -1.2. This parameter can tend to have quite extreme estimates varying from -2.617 to 0.819. The Bayesian method indicates some good results without improvement in mixing performance and with mean values close to true values, although estimates of β_{21} and β_{22} are a little biased. Standard deviations of MCMC are smallest among the three methods, and it also has the smallest mean absolute errors and mean square errors. Using the two-step method, mean values are acceptable but there is some bias in estimating β_{21} and α. Standard deviations of the two parameters are extremely large in comparison to other standard deviations. Then it is not surprising to find that quite extreme maximum and minimum values appear in two-step method with largest MAEs and MSEs.

Now look at Table 3.26 and 3.27, when error correlation exists. Table 3.26 indicates that bias and variation of MLE and two-step methods don't improve much, while MAEs and MSEs of the two methods are relatively smaller in contrast to that in Table 3.25. The two-step method still has the highest mean absolute errors and mean square errors among three methods. Meanwhile, Bayesian estimations cannot give unbiased estimates when $\rho = 0.5$ although vary less than other methods. When ρ increases to 0.8 in Table 3.27, it seems correlation is strong enough for MLE to identify the model. Thus, results estimated by MLE are unbiased with relatively small variance, MAEs and MSEs. However, the Bayesian method has some problems in dealing with this data. Mean values of β_{21} and α have significant bias with quite large standard deviations, which also results in big MAEs and MSEs. At the same time, two-step method works fine but not better than MLE because of some obvious bias and large variance.

With starting points close to true values, results in Table 3.28, 3.29 and 3.30 are discussed across methods as follows. Both precision and efficiency have been improved a lot with new starting values for MLE at each level of error correlation. The Bayesian method is less reliable when $\rho = 0.5$ or 0.8, although it is unbiased and efficient without correlation in the error terms. Estimation of two-step method is no longer influenced by starting values, as statistics in these tables are almost the same as the ones in Table 3.25, 3.26 and 3.27. It suggests the two-step method produces moderately biased results with relatively large variance and errors.

In summary, increasing the number of variables in the model may improve estimator performance a little. MLE is still sensitive to starting values when error correlation is not strong, but it performs quite well when $\rho = 0.8$. MCMC approach works fine when $\rho = 0$, but cannot be relied on with strong error correlation. Results show the two-step method is generally acceptable, although it is biased and has some loss of efficiency.

3.8.5 Explanatory Variables are Continuous (Unit Variance) and Binary

The identification condition is much stronger with one continuous and one binary explanatory variables in each equation than with two dummy variables. And the circumstance looks similar to the cases with only one continuous variable in Table 3.31, 3.32 and 3.33. Only zero starting values are discussed here, because MLE and two-step method work fine even when starting points are far from true values. Numbers of successful estimations are 765, 830 and 764 for MLE, while two-step method is applicable 953, 973 and 984 times respectively. According to SIF values, the mixing performance of MCMC algorithms has improved in comparison to models with two binary independent variables, but does not change much in comparison to the case with one continuous explanatory variable.

First, we discuss the performance of the three estimators shown in Table 3.31. Available estimates for MLE indicate that mean values are quite close to true values, but vary too much for β_{21}, α and ρ. In particular, the standard deviation of α estimates is larger than that for other parameters, since for MLE, several extreme estimates of α occur. Once again, the Bayesian method performs best when $\rho = 0$, with accurate mean values, small variance, least extreme values, smallest MAEs and smallest MSEs of β_2, α and ρ among three methods. Furthermore, statistics for the two-step method are quite similar to the ones for MLE. This method has the largest maximum values of β_2, α and ρ, and smallest minimum values of β_{21}, β_{22}, α and ρ.

We now discuss the performance of each method when $\rho = 0.5$ shown in Table 3.32. We see relatively large standard deviations for MLE estimates of β_{21} and α, while other statistics show this method is quite reliable. Although the MCMC method shows a little bias in estimation of β_{21} and ρ, its estimates of β_2, α and ρ have smaller variance than MLE, which results in the least MAEs and MSEs of such parameters among the three methods. The two-step method has estimated mean values quite close to the true values, except the mean of β_{23} which shows bias of more than 0.1, but with quite small variance of this parameter. Overall, the two-step method is not as efficient as the other methods, because its MAEs and MSEs of most parameters are largest.

Table 3.33 shows estimation results when $\rho = 0.8$ and such results are compared across methods. It indicates that MLE performs best among all methods. Even though its mean values of estimates of β_2, α and ρ reveal a small bias, the estimates do not vary as much as the Bayesian and two-step methods. MLE also gains the most efficiency as shown by quite small mean of absolute and square errors. The performance of Bayesian and two-step methods is almost the same, although neither of them produce accurate estimation of certain parameters. MCMC works especially badly in estimating α, with large bias, MAEs and MSEs. Meanwhile, the two-step method gives relatively large errors in estimating β_{22} and β_{23}.

To sum up, there are fewer near identification problems by introducing one continuous variable with unit variance and one binary variable in each equation. Starting points are no longer an issue for MLE and two-step methods and all three methods can produce acceptable estimates.

3.8.6 Explanatory Variables are Continuous (Variance 0.21) and Binary

The previous subsections have shown that systems including one continuous variable with unit variance and one binary variable in each equation can be estimated reasonably well, while models with two dummy variables in each equation perform worse. What will happen if the continuous variable has a small variance, to make a fairer comparison with the binary regressor case? Table 3.34, 3.35 and 3.36 are the results of the model which includes one continuous variable with variance 0.21 and one binary variable in each equation, when starting values are zeros. It seems estimator performance in this case is better in some ways and worse in other ways than the case with two dummy variables in each equation. Comparison will be made based only on estimates that use zero starting values.

Firstly, we start by comparing Table 3.34 and 3.25 with $\rho = 0$ according to different estimators. Generally speaking, MLE fails much less often as it successfully estimates 691 samples in Table 3.34 on contrast with 539 samples in the other Table. It also gives a little less biased estimates and slightly smaller standard deviations except for β_{13}. At the same time, statistics of MCMC in both cases are very similar, which means MCMC always performs best when $\rho = 0$ since it is unbiased and is most efficient among three estimators. For the two-step method, results are much less biased and more efficient in Table 3.34 than in Table 3.25. As a result, for $\rho = 0$, all three methods perform better with a continuous explanatory variable rather than with just binary variables, even when their variance is the same.

Next, Table 3.35 is compared with Table 3.26 where ρ increases to 0.5. MLE can estimate 644 samples in the former table and 590 samples in the other one. It provides better estimation of some parameters but worse estimation of others, but there are no big differences in results between the two tables. For MCMC, mean values show less bias in Table 3.26 except for estimating α. Meanwhile, some standard deviations are smaller in this table, but the difference is quite small, so one can hardly tell which model MCMC works better on. When it comes to the two-step method, results in Table 3.35 look much better than the ones in Table 3.26. Most parameters have much less bias as well as smaller standard deviations in Table 3.35, although MAEs and MSEs give some inconsistent information. For instance, the MAE and MSE of estimates of β_{12} is slightly more in Table 3.35. In all, the MLE and MCMC estimators show similar performance in both cases, while the two-step method works better with one continuous variable.

In contrast with Table 3.27, Table 3.36 reveals that the three methods all lose accuracy and precision when $\rho = 0.8$. Even though MLE performs better than the other methods, mean values of estimates of β_{21} and β_{23} are far away from true values and standard deviations are all larger in Table 3.36 than in Table 3.27, except for estimates of β_{11}. Estimation using MCMC is much worse in Table 3.36, although both tables contain very large SIF values for some parameters. The two-step method produces very biased estimation of β_{21}, β_{23} and α and is less accurate in Table 3.36.

In summary, a model with one continuous variable can be well estimated when $\rho = 0$, but has some identification problems when $\rho = 0.8$. Considering the good behavior when the continuous variable has unit variance in Section 3.8.5, it seems such problems are mainly caused by small variance of independent variable. Therefore, information in explanatory variables as represented by variance of X, can play an important role in identifying models.

3.9 Conclusions

This chapter has estimated a bivariate probit model with an endogenous binary variable using MLE, Bayesian and two-step methods. Some Monte Carlo experiments are used to study model estimation problems related to the use of exclusion restrictions. The main finding are summarized as follows.

All three estimators work very well when the model is totally structural or a partial structural system with exclusion restrictions. A structural model includes $x_{i1} \neq x_{i2}$, while a partial structural model with exclusion restrictions are characterized by $x_{i2} \subset x_{i1}$. MLE is quite accurate and efficient except for some convergence failures that occur across a range of ρ values. The Bayesian method is even more efficient when there is no error correlation or with moderate error correlation when a partial structural system is estimated. Although there is a little bias when ρ is large, the Bayesian method can work for every sample. A two-step method can be an alternative choice when correlation is low, but much less reliable when ρ is closer to 1.

When the model does not impose exclusion restrictions in partial equation estimation, all three estimators perform much worse especially when explanatory variables are all binary. One specification of only one binary independent variable in each equation in section 3.8.2 is equivalent to Wilde (2000)'s example where the explanatory variable takes just two different values. Although he shows this model can be identified in theory, results in section 3.8.4 reveal that estimator performance is quite poor. When the number of binary exogenous variables increases to two in Section 3.8.4, the identification conditions have been improved.

Wilde's comments about sufficient variation which will avoid identification problems are confirmed in the model with at least one continuous variable, as shown in Section 3.8.1 and 3.8.5. With at least one continuous independent variable, the estimators can perform well without imposing exclusion restrictions in most cases. All three methods can show small bias and precise estimates. The simulation study also reveals that the model is much more difficult to estimate than the cases with exclusion restrictions: when no exclusion restrictions are imposed, we see much frequent failure of MLE and the two-step method, and slower mixing of MCMC algorithms.

The relative performances of different estimation methods in different contexts have been discussed in this chapter. Each one is superior in some ways but has flaws in other ways. The benchmark method, MLE, is always accurate and efficient when it can produce estimates. The only drawback is that it often fails to converge especially in models without exclusion restrictions. MCMC is the only method which can estimate each sample. It normally perform best when there is very small error correlation, but tends to give biased estimation when ρ is large. The two-step method is the least time consuming among the three methods and has much less convergence failure than MLE. However, it usually has large standard deviations and is biased when ρ is large.

In conclusion, exclusion restrictions are not essential for accurate estimation of bivariate probit models with endogenous dummy variables. However, introduction of exclusion restrictions is the most easy way to reduce the difficulties in estimating our model. When exclusion restrictions are not available in certain economic contexts, the ways to reinforce model identification include increasing both the number of independent variables, and finding variables with plenty of variation.

Estimator Performance in Structural Models
Table 3.1: *ρ=0, Continuous Explanatory Variables, $x_{i1} \neq x_{i2}$*

	β_{11}	β_{12}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	0.800	-1.000	-1.200	0.000
Mean						
MLE*	1.001	-1.501	0.806	-1.001	-1.209	0.004
MCMC	1.004	-1.503	0.810	-1.003	-1.213	0.006
MCMC*	1.003	-1.504	0.809	-1.003	-1.213	0.007
TS**	1.002	-1.500	0.810	-1.004	-1.212	0.002
TS*	1.001	-1.501	0.808	-1.004	-1.212	0.003
Standard Deviation						
MLE*	0.069	0.092	0.122	0.063	0.155	0.122
MCMC	0.068	0.092	0.120	0.062	0.153	0.112
MCMC*	0.070	0.093	0.119	0.063	0.152	0.115
TS**	0.068	0.092	0.123	0.062	0.158	0.123
TS*	0.069	0.092	0.122	0.063	0.156	0.123
Maximum Values						
MLE*	1.347	-1.246	1.178	-0.824	-0.721	0.374
MCMC	1.345	-1.249	1.174	-0.827	-0.562	0.361
MCMC*	1.345	-1.249	1.174	-0.827	-0.740	0.361
TS**	1.349	-1.247	1.171	-0.826	-0.574	0.382
TS*	1.349	-1.247	1.170	-0.826	-0.734	0.382
Minimum Values						
MLE*	0.824	-1.893	0.452	-1.277	-1.713	-0.388
MCMC	0.823	-1.890	0.352	-1.285	-1.688	-0.384
MCMC*	0.823	-1.890	0.486	-1.285	-1.688	-0.384
TS**	0.825	-1.897	0.361	-1.281	-1.688	-0.418
TS*	0.825	-1.897	0.452	-1.281	-1.688	-0.418
Mean Absolute Error						
MLE*	0.055	0.074	0.098	0.049	0.125	0.101
MCMC	0.054	0.074	0.096	0.049	0.122	0.090
MCMC*	0.055	0.074	0.096	0.049	0.122	0.095
TS**	0.053	0.073	0.098	0.049	0.126	0.101
TS*	0.055	0.074	0.098	0.049	0.125	0.101
Mean Squared Error						
MLE*	0.005	0.009	0.015	0.004	0.024	0.015
MCMC	0.005	0.009	0.014	0.004	0.024	0.013
MCMC*	0.005	0.009	0.014	0.004	0.023	0.013
TS**	0.005	0.008	0.015	0.004	0.025	0.015
TS*	0.005	0.009	0.015	0.004	0.024	0.015
SIF for all MCMC outputs						
Mean	14	19	34	8	40	75
Standard Deviation	3	4	7	2	8	11

* 823 samples included; ** 944 samples included; 1000 samples included in MCMC

Estimator Performance in Structural Models
Table 3.2: *ρ=0.5, Continuous Explanatory Variables, $x_{i1} \neq x_{i2}$*

	β_{11}	β_{12}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	0.800	-1.000	-1.200	0.500
Mean						
MLE*	1.003	-1.501	0.805	-1.002	-1.206	0.505
MCMC	1.005	-1.506	0.791	-1.006	-1.186	0.480
MCMC*	1.006	-1.507	0.792	-1.006	-1.188	0.480
TS**	1.002	-1.500	0.845	-1.049	-1.262	0.527
TS*	1.003	-1.501	0.847	-1.050	-1.264	0.528
Standard Deviation						
MLE*	0.067	0.091	0.107	0.062	0.138	0.103
MCMC	0.067	0.091	0.106	0.063	0.135	0.100
MCMC*	0.068	0.092	0.107	0.063	0.137	0.100
TS**	0.068	0.092	0.127	0.065	0.164	0.124
TS*	0.068	0.092	0.127	0.065	0.166	0.124
Maximum Values						
MLE*	1.319	-1.247	1.178	-0.821	-0.711	0.834
MCMC	1.320	-1.250	1.165	-0.822	-0.676	0.816
MCMC*	1.320	-1.250	1.165	-0.822	-0.676	0.816
TS**	1.349	-1.247	1.316	-0.869	-0.694	0.927
TS*	1.349	-1.247	1.316	-0.869	-0.694	0.927
Minimum Values						
MLE*	0.829	-1.850	0.440	-1.245	-1.721	0.175
MCMC	0.830	-1.854	0.414	-1.246	-1.703	0.124
MCMC*	0.830	-1.854	0.414	-1.246	-1.703	0.160
TS**	0.825	-1.897	0.427	-1.279	-1.907	0.128
TS*	0.825	-1.897	0.427	-1.265	-1.907	0.160
Mean Absolute Error						
MLE*	0.053	0.073	0.086	0.050	0.110	0.082
MCMC	0.053	0.073	0.085	0.050	0.108	0.080
MCMC*	0.053	0.073	0.085	0.050	0.109	0.079
TS**	0.053	0.073	0.107	0.065	0.141	0.101
TS*	0.054	0.073	0.107	0.065	0.143	0.102
Mean Squared Error						
MLE*	0.005	0.008	0.012	0.004	0.019	0.011
MCMC	0.005	0.008	0.011	0.004	0.018	0.010
MCMC*	0.005	0.008	0.011	0.004	0.019	0.010
TS**	0.005	0.008	0.018	0.007	0.031	0.016
TS*	0.005	0.008	0.018	0.007	0.032	0.016
SIF for all MCMC outputs						
Mean	17	26	38	13	46	99
Standard Deviation	5	7	9	4	10	17

* 905 samples included; ** 1000 samples included; 1000 samples included in MCMC

Estimator Performance in Structural Models
Table 3.3: *ρ=0.8, Continuous Explanatory Variables, $x_{i1} \neq x_{i2}$*

	β_{11}	β_{12}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	0.800	-1.000	-1.200	0.800
Mean						
MLE*	1.001	-1.499	0.805	-1.003	-1.206	0.804
MCMC	1.007	-1.512	0.787	-1.011	-1.180	0.773
MCMC*	1.006	-1.511	0.787	-1.011	-1.182	0.774
TS**	1.002	-1.500	0.907	-1.131	-1.351	0.906
TS*	1.001	-1.500	0.908	-1.132	-1.353	0.907
Standard Deviation						
MLE*	0.066	0.089	0.091	0.061	0.109	0.064
MCMC	0.066	0.090	0.091	0.062	0.110	0.066
MCMC*	0.066	0.090	0.092	0.062	0.110	0.066
TS**	0.068	0.092	0.118	0.067	0.147	0.096
TS*	0.068	0.091	0.118	0.067	0.147	0.096
Maximum Values						
MLE*	1.309	-1.252	1.085	-0.854	-0.864	0.977
MCMC	1.309	-1.270	1.116	-0.835	-0.829	0.948
MCMC*	1.309	-1.270	1.116	-0.860	-0.829	0.948
TS**	1.349	-1.247	1.293	-0.927	-0.917	1.000
TS*	1.349	-1.247	1.268	-0.971	-0.917	1.000
Minimum Values						
MLE*	0.831	-1.849	0.502	-1.232	-1.574	0.573
MCMC	0.830	-1.845	0.473	-1.242	-1.596	0.553
MCMC*	0.830	-1.845	0.473	-1.242	-1.596	0.553
TS**	0.825	-1.897	0.523	-1.400	-1.814	0.546
TS*	0.825	-1.897	0.523	-1.400	-1.814	0.546
Mean Absolute Error						
MLE*	0.051	0.072	0.072	0.049	0.086	0.050
MCMC	0.052	0.073	0.073	0.050	0.088	0.055
MCMC*	0.052	0.072	0.073	0.050	0.089	0.055
TS**	0.053	0.073	0.131	0.132	0.176	0.125
TS*	0.053	0.073	0.131	0.132	0.177	0.126
Mean Squared Error						
MLE*	0.004	0.008	0.008	0.004	0.012	0.004
MCMC	0.004	0.008	0.008	0.004	0.013	0.005
MCMC*	0.004	0.008	0.009	0.004	0.013	0.005
TS**	0.005	0.008	0.025	0.022	0.044	0.021
TS*	0.005	0.008	0.026	0.022	0.045	0.021
SIF for all MCMC outputs						
Mean	30	49	42	27	53	141
Standard Deviation	9	14	11	10	13	17

* 935 samples included; ** 1000 samples included; 1000 samples included in MCMC

Estimator Performance in Structural Models
Table 3.4: *ρ=0, Binary Explanatory Variables, $x_{i1} \neq x_{i2}$*

	β_{11}	β_{12}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	0.800	-1.000	-1.200	0.000
Mean						
MLE*	1.001	-1.501	0.805	-1.002	-1.202	0.006
MCMC	1.001	-1.500	0.806	-1.001	-1.207	0.008
MCMC*	1.000	-1.497	0.806	-1.001	-1.206	0.010
TS**	1.003	-1.504	0.808	-1.008	-1.206	0.003
TS*	1.001	-1.501	0.807	-1.008	-1.205	0.004
Standard Deviation						
MLE*	0.087	0.100	0.127	0.095	0.204	0.139
MCMC	0.089	0.102	0.117	0.095	0.179	0.118
MCMC*	0.088	0.101	0.123	0.095	0.193	0.129
TS**	0.088	0.101	0.127	0.095	0.202	0.138
TS*	0.087	0.100	0.128	0.095	0.207	0.141
Maximum Values						
MLE*	1.308	-1.140	1.177	-0.727	-0.642	0.551
MCMC	1.307	-1.135	1.168	-0.650	-0.644	0.473
MCMC*	1.307	-1.135	1.168	-0.727	-0.644	0.473
TS**	1.309	-1.141	1.260	-0.729	-0.631	0.525
TS*	1.309	-1.141	1.260	-0.729	-0.631	0.525
Minimum Values						
MLE*	0.766	-1.871	0.389	-1.278	-1.860	-0.380
MCMC	0.762	-1.873	0.414	-1.278	-1.834	-0.351
MCMC*	0.762	-1.873	0.414	-1.278	-1.834	-0.351
TS**	0.767	-1.871	0.396	-1.286	-1.986	-0.424
TS*	0.767	-1.871	0.396	-1.286	-1.986	-0.424
Mean Absolute Error						
MLE*	0.068	0.080	0.102	0.076	0.165	0.115
MCMC	0.070	0.081	0.094	0.076	0.141	0.093
MCMC*	0.069	0.081	0.099	0.076	0.156	0.107
TS**	0.070	0.081	0.103	0.077	0.161	0.114
TS*	0.068	0.080	0.103	0.077	0.167	0.117
Mean Squared Error						
MLE*	0.008	0.010	0.016	0.009	0.041	0.019
MCMC	0.008	0.010	0.014	0.009	0.032	0.014
MCMC*	0.008	0.010	0.015	0.009	0.037	0.017
TS**	0.008	0.010	0.016	0.009	0.041	0.019
TS*	0.008	0.010	0.016	0.009	0.043	0.020
SIF for all MCMC outputs						
Mean	4	4	35	4	64	82
Standard Deviation	1	1	8	1	11	13

* 799 samples included; ** 853 samples included; 1000 samples included in MCMC

Estimator Performance in Structural Models
Table 3.5: *ρ=0.5, Binary Explanatory Variables, $x_{i1} \neq x_{i2}$*

	β_{11}	β_{12}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	0.800	-1.000	-1.200	0.500
Mean						
MLE*	1.003	-1.503	0.803	-1.002	-1.200	0.502
MCMC	1.006	-1.506	0.789	-1.007	-1.166	0.473
MCMC*	1.006	-1.506	0.789	-1.007	-1.167	0.473
TS**	1.003	-1.503	0.862	-1.081	-1.279	0.527
TS*	1.003	-1.503	0.862	-1.081	-1.280	0.528
Standard Deviation						
MLE*	0.088	0.101	0.101	0.094	0.155	0.112
MCMC	0.088	0.101	0.100	0.094	0.151	0.107
MCMC*	0.088	0.101	0.100	0.094	0.150	0.107
TS**	0.088	0.101	0.128	0.098	0.203	0.136
TS*	0.088	0.101	0.128	0.098	0.203	0.135
Maximum Values						
MLE*	1.308	-1.146	1.142	-0.657	-0.738	0.849
MCMC	1.313	-1.152	1.134	-0.660	-0.702	0.785
MCMC*	1.313	-1.152	1.134	-0.660	-0.702	0.785
TS**	1.309	-1.141	1.275	-0.725	-0.741	0.961
TS*	1.309	-1.141	1.275	-0.725	-0.741	0.961
Minimum Values						
MLE*	0.772	-1.873	0.478	-1.294	-1.661	0.098
MCMC	0.776	-1.868	0.465	-1.297	-1.630	0.054
MCMC*	0.776	-1.868	0.465	-1.297	-1.630	0.092
TS**	0.767	-1.871	0.483	-1.392	-2.027	0.049
TS*	0.767	-1.871	0.483	-1.392	-2.027	0.090
Mean Absolute Error						
MLE*	0.070	0.081	0.080	0.075	0.124	0.090
MCMC	0.070	0.080	0.079	0.076	0.123	0.087
MCMC*	0.070	0.080	0.079	0.076	0.123	0.087
TS**	0.070	0.081	0.114	0.103	0.173	0.110
TS*	0.070	0.081	0.114	0.103	0.173	0.110
Mean Squared Error						
MLE*	0.008	0.010	0.010	0.009	0.024	0.013
MCMC	0.008	0.010	0.010	0.009	0.024	0.012
MCMC*	0.008	0.010	0.010	0.009	0.024	0.012
TS**	0.008	0.010	0.020	0.016	0.047	0.019
TS*	0.008	0.010	0.020	0.016	0.048	0.019
SIF for all MCMC outputs						
Mean	8	10	31	13	75	106
Standard Deviation	4	5	8	7	14	19

* 991 samples included; ** 1000 samples included; 1000 samples included in MCMC

Estimator Performance in Structural Models
Table 3.6: *ρ=0.8, Binary Explanatory Variables, $x_{i1} \neq x_{i2}$*

	β_{11}	β_{12}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	0.800	-1.000	-1.200	0.800
Mean						
MLE*	1.004	-1.503	0.801	-1.002	-1.197	0.799
MCMC	1.013	-1.517	0.787	-1.021	-1.153	0.761
MCMC*	1.014	-1.518	0.787	-1.021	-1.152	0.761
TS**	1.003	-1.503	0.965	-1.241	-1.388	0.907
TS*	1.004	-1.504	0.965	-1.240	-1.387	0.907
Standard Deviation						
MLE*	0.088	0.101	0.078	0.090	0.117	0.076
MCMC	0.088	0.101	0.080	0.090	0.120	0.078
MCMC*	0.088	0.101	0.080	0.090	0.120	0.078
TS**	0.088	0.101	0.112	0.094	0.169	0.107
TS*	0.088	0.101	0.113	0.094	0.169	0.107
Maximum Values						
MLE*	1.291	-1.162	1.031	-0.759	-0.714	0.999
MCMC	1.312	-1.185	1.022	-0.769	-0.671	0.944
MCMC*	1.312	-1.185	1.022	-0.769	-0.671	0.944
TS**	1.309	-1.141	1.279	-0.986	-0.757	1.000
TS*	1.309	-1.141	1.279	-0.986	-0.757	1.000
Minimum Values						
MLE*	0.776	-1.868	0.531	-1.294	-1.511	0.517
MCMC	0.781	-1.883	0.511	-1.299	-1.467	0.488
MCMC*	0.781	-1.883	0.511	-1.299	-1.467	0.488
TS**	0.767	-1.871	0.581	-1.514	-1.763	0.516
TS*	0.767	-1.871	0.581	-1.514	-1.763	0.516
Mean Absolute Error						
MLE*	0.070	0.081	0.062	0.073	0.093	0.060
MCMC	0.070	0.081	0.064	0.074	0.101	0.067
MCMC*	0.070	0.081	0.064	0.074	0.101	0.067
TS**	0.070	0.081	0.173	0.241	0.220	0.135
TS*	0.070	0.081	0.173	0.240	0.219	0.135
Mean Squared Error						
MLE*	0.008	0.010	0.006	0.008	0.014	0.006
MCMC	0.008	0.010	0.006	0.009	0.017	0.008
MCMC*	0.008	0.010	0.006	0.008	0.017	0.008
TS**	0.008	0.010	0.040	0.067	0.064	0.023
TS*	0.008	0.010	0.040	0.067	0.064	0.023
SIF for all MCMC outputs						
Mean	18	23	22	35	89	146
Standard Deviation	7	9	9	14	16	18

* 987 samples included; ** 1000 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.7: *ρ=0, Continuous Explanatory Variables, $x_{i2} \subset x_{i1}$*

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	-1.000	0.800	-1.000	-1.200	0.000
Mean							
MLE*	1.010	-1.515	-1.006	0.794	-1.003	-1.194	-0.004
MCMC	1.011	-1.517	-1.006	0.796	-1.005	-1.197	0.000
MCMC*	1.014	-1.522	-1.008	0.798	-1.006	-1.200	0.001
TS**	1.008	-1.511	-1.004	0.791	-1.005	-1.191	-0.009
TS*	1.010	-1.515	-1.006	0.795	-1.006	-1.197	-0.007
Standard Deviation							
MLE*	0.074	0.102	0.077	0.151	0.080	0.205	0.148
MCMC	0.073	0.100	0.077	0.141	0.078	0.192	0.131
MCMC*	0.074	0.103	0.078	0.144	0.079	0.195	0.137
TS**	0.072	0.099	0.076	0.151	0.081	0.206	0.147
TS*	0.074	0.102	0.077	0.152	0.081	0.207	0.149
Maximum Values							
MLE*	1.301	-1.203	-0.776	1.284	-0.754	-0.672	0.504
MCMC	1.315	-1.207	-0.778	1.269	-0.763	-0.689	0.479
MCMC*	1.315	-1.207	-0.778	1.269	-0.763	-0.689	0.479
TS**	1.298	-1.206	-0.777	1.321	-0.762	-0.647	0.489
TS*	1.298	-1.206	-0.777	1.321	-0.762	-0.705	0.489
Minimum Values							
MLE*	0.819	-1.913	-1.303	0.380	-1.276	-1.943	-0.380
MCMC	0.822	-1.920	-1.312	0.392	-1.277	-1.922	-0.364
MCMC*	0.822	-1.920	-1.312	0.392	-1.261	-1.922	-0.364
TS**	0.821	-1.904	-1.306	0.358	-1.298	-2.010	-0.383
TS*	0.821	-1.904	-1.306	0.402	-1.277	-2.010	-0.381
Mean Absolute Error							
MLE*	0.057	0.082	0.061	0.120	0.065	0.163	0.121
MCMC	0.058	0.080	0.061	0.112	0.062	0.152	0.104
MCMC*	0.058	0.083	0.062	0.115	0.063	0.155	0.111
TS**	0.056	0.079	0.060	0.120	0.064	0.163	0.120
TS*	0.057	0.082	0.062	0.121	0.065	0.164	0.122
Mean Squared Error							
MLE*	0.006	0.011	0.006	0.023	0.006	0.042	0.022
MCMC	0.006	0.010	0.006	0.020	0.006	0.037	0.017
MCMC*	0.006	0.011	0.006	0.021	0.006	0.038	0.019
TS**	0.005	0.010	0.006	0.023	0.007	0.042	0.022
TS*	0.006	0.011	0.006	0.023	0.007	0.043	0.022
SIF for all MCMC outputs							
Mean	18	24	19	55	31	60	93
Standard Deviation	5	6	5	11	7	11	13

* 785 samples included; ** 943 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.8: *ρ=0.5, Continuous Explanatory Variables, $x_{i2} \subset x_{i1}$*

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	-1.000	0.800	-1.000	-1.200	0.500
Mean							
MLE*	1.005	-1.509	-1.002	0.793	-1.003	-1.195	0.494
MCMC	1.012	-1.518	-1.010	0.768	-0.995	-1.158	0.459
MCMC*	1.010	-1.517	-1.007	0.771	-0.997	-1.163	0.462
TS**	1.008	-1.511	-1.004	0.826	-1.039	-1.254	0.508
TS*	1.006	-1.509	-1.002	0.829	-1.042	-1.259	0.511
Standard Deviation							
MLE*	0.072	0.097	0.076	0.122	0.066	0.159	0.117
MCMC	0.072	0.098	0.076	0.121	0.066	0.157	0.112
MCMC*	0.072	0.097	0.076	0.120	0.066	0.155	0.111
TS**	0.073	0.099	0.076	0.146	0.079	0.198	0.139
TS*	0.073	0.099	0.076	0.146	0.079	0.197	0.139
Maximum Values							
MLE*	1.285	-1.205	-0.774	1.136	-0.806	-0.628	0.839
MCMC	1.299	-1.211	-0.781	1.125	-0.807	-0.610	0.783
MCMC*	1.299	-1.211	-0.781	1.125	-0.807	-0.610	0.783
TS**	1.298	-1.206	-0.777	1.304	-0.805	-0.618	0.985
TS*	1.298	-1.206	-0.777	1.304	-0.805	-0.618	0.985
Minimum Values							
MLE*	0.810	-1.890	-1.304	0.402	-1.200	-1.621	0.044
MCMC	0.809	-1.916	-1.301	0.392	-1.195	-1.604	0.047
MCMC*	0.809	-1.916	-1.301	0.392	-1.195	-1.604	0.047
TS**	0.821	-1.904	-1.306	0.398	-1.319	-1.951	0.044
TS*	0.821	-1.904	-1.306	0.398	-1.319	-1.951	0.044
Mean Absolute Error							
MLE*	0.056	0.077	0.060	0.098	0.052	0.126	0.092
MCMC	0.057	0.078	0.060	0.099	0.052	0.129	0.094
MCMC*	0.057	0.077	0.060	0.098	0.052	0.127	0.092
TS**	0.057	0.079	0.061	0.117	0.070	0.162	0.109
TS*	0.057	0.079	0.060	0.118	0.071	0.163	0.109
Mean Squared Error							
MLE*	0.005	0.009	0.006	0.015	0.004	0.025	0.014
MCMC	0.005	0.010	0.006	0.016	0.004	0.026	0.014
MCMC*	0.005	0.010	0.006	0.015	0.004	0.025	0.014
TS**	0.005	0.010	0.006	0.022	0.008	0.042	0.019
TS*	0.005	0.010	0.006	0.022	0.008	0.042	0.019
SIF for all MCMC outputs							
Mean	22	29	27	60	24	65	112
Standard Deviation	6	8	8	12	6	12	17

* 860 samples included; ** 1000 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.9: *ρ=0.8, Continuous Explanatory Variables, $x_{i2} \subset x_{i1}$*

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	-1.000	0.800	-1.000	-1.200	0.800
Mean							
MLE*	1.007	-1.510	-1.004	0.803	-1.004	-1.205	0.801
MCMC	1.015	-1.523	-1.018	0.770	-0.997	-1.162	0.761
MCMC*	1.015	-1.523	-1.018	0.773	-0.999	-1.167	0.763
TS**	1.008	-1.511	-1.004	0.899	-1.113	-1.389	0.909
TS*	1.008	-1.510	-1.004	0.902	-1.115	-1.395	0.911
Standard Deviation							
MLE*	0.070	0.094	0.076	0.104	0.056	0.129	0.073
MCMC	0.072	0.096	0.075	0.104	0.057	0.130	0.074
MCMC*	0.071	0.095	0.075	0.103	0.057	0.129	0.073
TS**	0.073	0.099	0.076	0.128	0.072	0.168	0.102
TS*	0.071	0.098	0.076	0.129	0.072	0.168	0.102
Maximum Values							
MLE*	1.281	-1.208	-0.778	1.192	-0.821	-0.711	0.982
MCMC	1.314	-1.214	-0.783	1.167	-0.813	-0.697	0.934
MCMC*	1.314	-1.214	-0.783	1.167	-0.813	-0.697	0.934
TS**	1.298	-1.206	-0.777	1.335	-0.903	-0.789	1.000
TS*	1.298	-1.206	-0.777	1.335	-0.903	-0.789	1.000
Minimum Values							
MLE*	0.813	-1.821	-1.311	0.433	-1.196	-1.697	0.544
MCMC	0.820	-1.828	-1.317	0.422	-1.195	-1.664	0.479
MCMC*	0.820	-1.828	-1.317	0.422	-1.195	-1.664	0.479
TS**	0.821	-1.904	-1.306	0.478	-1.325	-1.956	0.548
TS*	0.821	-1.904	-1.306	0.478	-1.325	-1.956	0.548
Mean Absolute Error							
MLE*	0.054	0.075	0.060	0.082	0.045	0.102	0.058
MCMC	0.058	0.077	0.061	0.084	0.046	0.106	0.065
MCMC*	0.056	0.077	0.061	0.083	0.045	0.104	0.063
TS**	0.057	0.079	0.061	0.136	0.116	0.217	0.133
TS*	0.056	0.079	0.061	0.138	0.118	0.221	0.135
Mean Squared Error							
MLE*	0.005	0.009	0.006	0.011	0.003	0.017	0.005
MCMC	0.005	0.010	0.006	0.012	0.003	0.018	0.007
MCMC*	0.005	0.010	0.006	0.011	0.003	0.018	0.007
TS**	0.005	0.010	0.006	0.026	0.018	0.064	0.022
TS*	0.005	0.010	0.006	0.027	0.018	0.066	0.023
SIF for all MCMC outputs							
Mean	37	47	51	65	18	69	146
Standard Deviation	11	13	15	15	6	15	17

* 882 samples included; ** 1000 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.10: *ρ=0, Binary Explanatory Variables, $x_{i2} \subset x_{i1}$*

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	-1.000	0.800	-1.000	-1.200	0.000
Mean							
MLE*	1.001	-1.504	-1.001	0.791	-0.992	-1.187	0.000
MCMC	1.003	-1.508	-1.000	0.803	-1.001	-1.205	0.010
MCMC*	0.999	-1.506	-0.997	0.799	-0.998	-1.199	0.009
TS**	1.006	-1.506	-1.005	0.818	-1.019	-1.225	0.008
TS*	1.001	-1.503	-1.002	0.812	-1.014	-1.217	0.005
Standard Deviation							
MLE*	0.109	0.106	0.104	0.243	0.202	0.369	0.225
MCMC	0.110	0.107	0.105	0.199	0.172	0.294	0.172
MCMC*	0.109	0.106	0.104	0.209	0.178	0.312	0.187
TS**	0.109	0.107	0.104	0.235	0.197	0.356	0.216
TS*	0.109	0.106	0.103	0.240	0.199	0.367	0.224
Maximum Values							
MLE*	1.323	-1.136	-0.691	1.515	0.002	0.462	0.574
MCMC	1.378	-1.139	-0.684	1.511	-0.433	-0.128	0.538
MCMC*	1.323	-1.139	-0.684	1.511	-0.433	-0.232	0.538
TS**	1.378	-1.135	-0.688	1.587	-0.495	-0.264	0.611
TS*	1.322	-1.135	-0.688	1.587	-0.507	-0.264	0.611
Minimum Values							
MLE*	0.682	-1.907	-1.307	-0.388	-1.594	-2.111	-0.860
MCMC	0.677	-1.912	-1.304	0.076	-1.591	-2.056	-0.542
MCMC*	0.677	-1.912	-1.304	0.076	-1.591	-2.056	-0.542
TS**	0.684	-1.907	-1.309	0.185	-1.666	-2.343	-0.612
TS*	0.684	-1.907	-1.309	0.185	-1.666	-2.343	-0.612
Mean Absolute Error							
MLE*	0.086	0.085	0.082	0.200	0.164	0.308	0.189
MCMC	0.088	0.085	0.083	0.160	0.137	0.237	0.138
MCMC*	0.087	0.085	0.082	0.172	0.144	0.260	0.158
TS**	0.087	0.085	0.083	0.194	0.161	0.295	0.179
TS*	0.086	0.084	0.081	0.198	0.163	0.307	0.188
Mean Squared Error							
MLE*	0.012	0.011	0.011	0.059	0.041	0.136	0.050
MCMC	0.012	0.011	0.011	0.040	0.029	0.086	0.030
MCMC*	0.012	0.011	0.011	0.044	0.031	0.097	0.035
TS**	0.012	0.011	0.011	0.055	0.039	0.127	0.047
TS*	0.012	0.011	0.011	0.058	0.040	0.135	0.050
SIF for all MCMC outputs							
Mean	5	4	6	92	74	110	129
Standard Deviation	2	1	3	18	19	17	16

* 755 samples included; ** 871 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.11: *ρ=0.5, Binary Explanatory Variables, $x_{i2} \subset x_{i1}$*

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	-1.000	0.800	-1.000	-1.200	0.500
Mean							
MLE*	1.007	-1.506	-1.007	0.795	-0.999	-1.188	0.496
MCMC	1.009	-1.509	-1.008	0.748	-0.965	-1.110	0.439
MCMC*	1.011	-1.510	-1.010	0.749	-0.966	-1.110	0.439
TS**	1.006	-1.505	-1.005	0.896	-1.096	-1.342	0.551
TS*	1.007	-1.506	-1.007	0.896	-1.097	-1.341	0.551
Standard Deviation							
MLE*	0.109	0.106	0.104	0.146	0.127	0.219	0.147
MCMC	0.109	0.106	0.103	0.147	0.128	0.217	0.142
MCMC*	0.109	0.106	0.103	0.142	0.125	0.210	0.137
TS**	0.109	0.106	0.104	0.211	0.181	0.324	0.192
TS*	0.109	0.106	0.104	0.204	0.175	0.313	0.186
Maximum Values							
MLE*	1.372	-1.126	-0.680	1.185	-0.506	-0.330	0.870
MCMC	1.372	-1.126	-0.698	1.189	-0.548	-0.351	0.874
MCMC*	1.372	-1.126	-0.698	1.189	-0.552	-0.351	0.859
TS**	1.378	-1.135	-0.688	1.496	-0.571	-0.366	1.000
TS*	1.378	-1.135	-0.688	1.496	-0.612	-0.366	1.000
Minimum Values							
MLE*	0.685	-1.889	-1.306	0.274	-1.368	-1.687	-0.083
MCMC	0.685	-1.899	-1.309	0.266	-1.340	-1.681	-0.058
MCMC*	0.685	-1.899	-1.309	0.266	-1.340	-1.598	-0.058
TS**	0.684	-1.907	-1.309	0.276	-1.625	-2.194	-0.046
TS*	0.684	-1.907	-1.309	0.276	-1.625	-2.131	-0.046
Mean Absolute Error							
MLE*	0.088	0.084	0.082	0.117	0.100	0.175	0.119
MCMC	0.087	0.085	0.081	0.123	0.105	0.185	0.122
MCMC*	0.087	0.085	0.082	0.120	0.102	0.181	0.118
TS**	0.087	0.084	0.082	0.185	0.163	0.282	0.161
TS*	0.088	0.084	0.082	0.181	0.159	0.276	0.157
Mean Squared Error							
MLE*	0.012	0.011	0.011	0.021	0.016	0.048	0.022
MCMC	0.012	0.011	0.011	0.024	0.018	0.055	0.024
MCMC*	0.012	0.011	0.011	0.023	0.017	0.052	0.022
TS**	0.012	0.011	0.011	0.053	0.042	0.125	0.040
TS*	0.012	0.011	0.011	0.051	0.040	0.118	0.037
SIF for all MCMC outputs							
Mean	8	5	13	81	55	107	135
Standard Deviation	5	2	9	17	16	16	18

* 955 samples included; ** 1000 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.12: *ρ=0.8, Binary Explanatory Variables, $x_{i2} \subset x_{i1}$*

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	-1.000	0.800	-1.000	-1.200	0.800
Mean							
MLE*	1.014	-1.501	-1.021	0.783	-0.987	-1.169	0.783
MCMC	1.021	-1.510	-1.026	0.758	-0.973	-1.119	0.742
MCMC*	1.029	-1.507	-1.040	0.744	-0.960	-1.101	0.731
TS**	1.006	-1.505	-1.005	1.016	-1.206	-1.516	0.952
TS*	1.014	-1.503	-1.021	1.003	-1.193	-1.500	0.946
Standard Deviation							
MLE*	0.108	0.105	0.096	0.100	0.096	0.142	0.089
MCMC	0.107	0.105	0.100	0.110	0.105	0.156	0.096
MCMC*	0.107	0.106	0.094	0.105	0.099	0.152	0.094
TS**	0.109	0.106	0.104	0.134	0.131	0.175	0.091
TS*	0.109	0.107	0.097	0.132	0.128	0.176	0.094
Maximum Values							
MLE*	1.357	-1.140	-0.741	1.065	-0.611	-0.616	1.000
MCMC	1.364	-1.139	-0.717	1.106	-0.572	-0.511	0.942
MCMC*	1.364	-1.139	-0.725	1.057	-0.572	-0.511	0.942
TS**	1.378	-1.135	-0.688	1.395	-0.659	-0.699	1.000
TS*	1.378	-1.135	-0.740	1.357	-0.659	-0.699	1.000
Minimum Values							
MLE*	0.701	-1.894	-1.307	0.359	-1.327	-1.545	0.468
MCMC	0.731	-1.913	-1.327	0.301	-1.300	-1.489	0.411
MCMC*	0.731	-1.913	-1.327	0.301	-1.289	-1.482	0.411
TS**	0.684	-1.907	-1.309	0.407	-1.624	-1.915	0.526
TS*	0.684	-1.907	-1.303	0.407	-1.624	-1.915	0.526
Mean Absolute Error							
MLE*	0.088	0.084	0.077	0.080	0.077	0.115	0.072
MCMC	0.088	0.084	0.081	0.094	0.086	0.138	0.087
MCMC*	0.089	0.085	0.081	0.094	0.084	0.144	0.091
TS**	0.087	0.084	0.082	0.225	0.213	0.332	0.167
TS*	0.088	0.085	0.078	0.213	0.201	0.318	0.163
Mean Squared Error							
MLE*	0.012	0.011	0.010	0.010	0.009	0.021	0.008
MCMC	0.012	0.011	0.011	0.014	0.012	0.031	0.013
MCMC*	0.012	0.011	0.010	0.014	0.011	0.033	0.014
TS**	0.012	0.011	0.011	0.065	0.060	0.130	0.031
TS*	0.012	0.011	0.010	0.058	0.054	0.121	0.030
SIF for all MCMC outputs							
Mean	20	8	35	62	36	104	154
Standard Deviation	10	4	16	18	14	18	17

* 823 samples included; ** 1000 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.13: *ρ=0, Continuous Explanatory Variables, $x_{i1} = x_{i2}$*

	β_{11}	β_{12}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	0.800	-1.000	-1.200	0.000
Mean						
MLE*	1.008	-1.510	0.813	-1.003	-1.213	0.023
MCMC	1.005	-1.508	0.844	-1.015	-1.254	0.054
MCMC*	1.008	-1.511	0.845	-1.016	-1.255	0.056
TS**	1.005	-1.507	0.801	-1.005	-1.202	-0.001
TS*	1.008	-1.510	0.797	-1.003	-1.197	-0.004
Standard Deviation						
MLE*	0.067	0.088	0.301	0.134	0.406	0.262
MCMC	0.069	0.092	0.250	0.117	0.335	0.212
MCMC*	0.068	0.088	0.258	0.118	0.346	0.220
TS**	0.069	0.092	0.304	0.138	0.414	0.261
TS*	0.067	0.088	0.309	0.139	0.420	0.265
Maximum Values						
MLE*	1.241	-1.251	1.663	-0.615	-0.081	0.814
MCMC	1.352	-1.236	1.632	-0.622	-0.171	0.828
MCMC*	1.239	-1.253	1.632	-0.622	-0.171	0.828
TS**	1.349	-1.237	1.782	-0.632	0.161	0.759
TS*	1.240	-1.253	1.782	-0.619	0.161	0.759
Minimum Values						
MLE*	0.820	-1.844	-0.039	-1.376	-2.324	-0.588
MCMC	0.820	-1.901	0.026	-1.367	-2.233	-0.535
MCMC*	0.820	-1.851	0.026	-1.366	-2.233	-0.535
TS**	0.818	-1.897	-0.230	-1.464	-2.402	-0.885
TS*	0.818	-1.844	-0.230	-1.411	-2.402	-0.885
Mean Absolute Error						
MLE*	0.054	0.071	0.248	0.109	0.335	0.216
MCMC	0.055	0.073	0.206	0.096	0.277	0.175
MCMC*	0.054	0.071	0.214	0.097	0.288	0.184
TS**	0.054	0.073	0.249	0.112	0.339	0.214
TS*	0.054	0.071	0.253	0.112	0.344	0.218
Mean Squared Error						
MLE*	0.005	0.008	0.091	0.018	0.165	0.069
MCMC	0.005	0.009	0.065	0.014	0.115	0.048
MCMC*	0.005	0.008	0.069	0.014	0.122	0.052
TS**	0.005	0.008	0.092	0.019	0.171	0.068
TS*	0.005	0.008	0.096	0.019	0.177	0.070
SIF for all MCMC outputs						
Mean	16	21	128	102	130	151
Standard Deviation	5	6	18	17	18	17

* 742 samples included; ** 944 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.14: *ρ=0.5, Continuous Explanatory Variables, $x_{i1} = x_{i2}$*

	β_{11}	β_{12}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	0.800	-1.000	-1.200	0.500
Mean						
MLE*	1.005	-1.508	0.790	-0.993	-1.180	0.494
MCMC	1.008	-1.513	0.694	-0.960	-1.056	0.406
MCMC*	1.009	-1.514	0.696	-0.960	-1.059	0.408
TS**	1.005	-1.506	0.776	-1.017	-1.193	0.470
TS*	1.006	-1.508	0.776	-1.017	-1.192	0.470
Standard Deviation						
MLE*	0.067	0.090	0.274	0.108	0.348	0.235
MCMC	0.068	0.091	0.227	0.096	0.291	0.186
MCMC*	0.067	0.090	0.221	0.095	0.284	0.181
TS**	0.069	0.092	0.283	0.130	0.384	0.231
TS*	0.067	0.091	0.288	0.132	0.391	0.236
Maximum Values						
MLE*	1.240	-1.258	1.373	-0.676	-0.097	0.999
MCMC	1.349	-1.233	1.328	-0.679	-0.020	0.911
MCMC*	1.243	-1.266	1.296	-0.679	-0.074	0.850
TS**	1.349	-1.237	1.606	-0.705	-0.045	1.000
TS*	1.240	-1.253	1.528	-0.680	-0.045	1.000
Minimum Values						
MLE*	0.816	-1.843	-0.020	-1.296	-1.890	-0.255
MCMC	0.819	-1.907	-0.071	-1.219	-1.833	-0.276
MCMC*	0.819	-1.842	-0.035	-1.219	-1.753	-0.264
TS**	0.818	-1.897	-0.062	-1.426	-2.267	-0.306
TS*	0.818	-1.844	-0.062	-1.426	-2.266	-0.294
Mean Absolute Error						
MLE*	0.053	0.073	0.226	0.087	0.285	0.194
MCMC	0.054	0.073	0.197	0.083	0.254	0.165
MCMC*	0.054	0.073	0.192	0.082	0.247	0.160
TS**	0.055	0.073	0.229	0.105	0.307	0.188
TS*	0.053	0.073	0.233	0.106	0.314	0.191
Mean Squared Error						
MLE*	0.004	0.008	0.075	0.012	0.121	0.055
MCMC	0.005	0.008	0.063	0.011	0.106	0.043
MCMC*	0.005	0.008	0.060	0.011	0.101	0.041
TS**	0.005	0.008	0.080	0.017	0.148	0.054
TS*	0.004	0.008	0.083	0.018	0.153	0.057
SIF for all MCMC outputs						
Mean	20	26	135	97	136	158
Standard Deviation	7	9	18	18	17	17

* 759 samples included; ** 989 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.15: *ρ=0.8, Continuous Explanatory Variables, $x_{i1} = x_{i2}$*

	β_{11}	β_{12}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	0.800	-1.000	-1.200	0.800
Mean						
MLE*	1.006	-1.507	0.779	-0.992	-1.173	0.781
MCMC	1.017	-1.524	0.619	-0.948	-0.981	0.661
MCMC*	1.018	-1.524	0.623	-0.947	-0.987	0.663
TS**	1.005	-1.506	0.733	-1.057	-1.209	0.809
TS*	1.006	-1.506	0.738	-1.058	-1.217	0.813
Standard Deviation						
MLE*	0.068	0.088	0.203	0.078	0.238	0.138
MCMC	0.067	0.089	0.209	0.082	0.251	0.144
MCMC*	0.068	0.088	0.203	0.080	0.244	0.141
TS**	0.068	0.091	0.233	0.110	0.313	0.182
TS*	0.068	0.090	0.229	0.108	0.308	0.181
Maximum Values						
MLE*	1.348	-1.247	1.182	-0.639	-0.175	0.992
MCMC	1.358	-1.227	1.139	-0.619	0.025	0.994
MCMC*	1.358	-1.227	1.139	-0.619	-0.274	0.977
TS**	1.349	-1.237	1.211	-0.646	0.044	1.000
TS*	1.349	-1.237	1.211	-0.646	-0.281	1.000
Minimum Values						
MLE*	0.816	-1.903	0.063	-1.205	-1.626	0.155
MCMC	0.832	-1.916	-0.158	-1.206	-1.570	0.088
MCMC*	0.832	-1.916	0.053	-1.178	-1.570	0.173
TS**	0.818	-1.897	-0.173	-1.335	-1.799	0.075
TS*	0.818	-1.897	0.073	-1.335	-1.799	0.222
Mean Absolute Error						
MLE*	0.054	0.070	0.159	0.062	0.186	0.107
MCMC	0.055	0.073	0.221	0.078	0.263	0.159
MCMC*	0.055	0.072	0.214	0.076	0.255	0.156
TS**	0.054	0.073	0.190	0.102	0.259	0.154
TS*	0.054	0.071	0.186	0.102	0.255	0.154
Mean Squared Error						
MLE*	0.005	0.008	0.042	0.006	0.058	0.019
MCMC	0.005	0.009	0.076	0.009	0.111	0.040
MCMC*	0.005	0.008	0.073	0.009	0.105	0.039
TS**	0.005	0.008	0.059	0.015	0.098	0.033
TS*	0.005	0.008	0.056	0.015	0.095	0.033
SIF for all MCMC outputs						
Mean	29	37	138	82	137	167
Standard Deviation	13	16	18	21	18	14

* 779 samples included; ** 998 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.16: *ρ=0, Binary Explanatory Variables, $x_{i1} = x_{i2}$*

	β_{11}	β_{12}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	0.800	-1.000	-1.200	0.000
Mean						
MLE*	1.003	-1.507	0.206	-0.576	-0.492	-0.352
MCMC	1.005	-1.506	0.737	-0.924	-1.124	0.010
MCMC*	1.004	-1.509	0.751	-0.934	-1.140	0.020
TS**	1.006	-1.507	0.199	-0.620	-0.484	-0.421
TS*	1.002	-1.506	0.117	-0.569	-0.385	-0.480
Standard Deviation						
MLE*	0.083	0.097	0.440	0.295	0.488	0.324
MCMC	0.086	0.098	0.373	0.267	0.424	0.244
MCMC*	0.082	0.096	0.375	0.270	0.429	0.247
TS**	0.088	0.101	0.804	0.524	0.951	0.563
TS*	0.082	0.096	0.645	0.426	0.762	0.454
Maximum Values						
MLE*	1.306	-1.215	2.149	0.677	1.043	0.990
MCMC	1.300	-1.209	1.721	0.300	0.767	0.894
MCMC*	1.300	-1.214	1.721	0.300	0.767	0.801
TS**	1.306	-1.213	2.729	0.207	0.735	1.000
TS*	1.306	-1.213	2.556	0.207	0.657	1.000
Minimum Values						
MLE*	0.779	-1.786	-1.152	-1.925	-2.528	-0.994
MCMC	0.758	-1.805	-0.845	-1.560	-2.399	-0.869
MCMC*	0.780	-1.781	-0.845	-1.555	-2.399	-0.869
TS**	0.775	-1.805	-0.938	-2.353	-3.234	-1.000
TS*	0.779	-1.786	-0.919	-2.353	-3.164	-1.000
Mean Absolute Error						
MLE*	0.066	0.078	0.714	0.492	0.834	0.454
MCMC	0.069	0.079	0.293	0.213	0.333	0.188
MCMC*	0.065	0.078	0.290	0.211	0.332	0.191
TS**	0.071	0.082	0.951	0.606	1.130	0.666
TS*	0.065	0.077	0.897	0.572	1.069	0.631
Mean Squared Error						
MLE*	0.007	0.009	0.546	0.266	0.740	0.228
MCMC	0.007	0.010	0.143	0.077	0.185	0.059
MCMC*	0.007	0.009	0.143	0.077	0.187	0.061
TS**	0.008	0.010	1.008	0.418	1.415	0.494
TS*	0.007	0.009	0.882	0.367	1.243	0.436
SIF for all MCMC outputs						
Mean	5	4	175	163	179	185
Standard Deviation	3	2	13	20	12	10

* 497 samples included; ** 551 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.17: *ρ=0.5, Binary Explanatory Variables, $x_{i1} = x_{i2}$*

	β_{11}	β_{12}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	0.800	-1.000	-1.200	0.500
Mean						
MLE*	1.005	-1.504	0.080	-0.567	-0.359	-0.014
MCMC	1.005	-1.506	0.167	-0.604	-0.456	0.053
MCMC*	1.007	-1.506	0.139	-0.582	-0.429	0.034
TS**	1.007	-1.507	0.236	-0.688	-0.543	0.080
TS*	1.005	-1.503	0.100	-0.596	-0.393	-0.012
Standard Deviation						
MLE*	0.085	0.097	0.436	0.256	0.480	0.333
MCMC	0.086	0.098	0.390	0.264	0.442	0.264
MCMC*	0.085	0.097	0.405	0.275	0.457	0.269
TS**	0.085	0.097	0.756	0.483	0.896	0.531
TS*	0.085	0.097	0.625	0.404	0.742	0.441
Maximum Values						
MLE*	1.233	-1.206	1.531	0.003	0.607	0.997
MCMC	1.308	-1.202	1.482	0.685	1.398	0.958
MCMC*	1.234	-1.202	1.212	0.685	1.398	0.809
TS**	1.306	-1.206	1.931	0.498	1.462	1.000
TS*	1.231	-1.206	1.815	0.454	1.483	1.000
Minimum Values						
MLE*	0.755	-1.776	-0.742	-1.461	-1.846	-0.608
MCMC	0.752	-1.804	-1.482	-1.376	-1.786	-0.951
MCMC*	0.752	-1.779	-1.482	-1.217	-1.607	-0.951
TS**	0.756	-1.786	-1.509	-1.829	-2.490	-1.000
TS*	0.756	-1.776	-1.509	-1.829	-2.294	-1.000
Mean Absolute Error						
MLE*	0.070	0.079	0.817	0.479	0.937	0.598
MCMC	0.069	0.079	0.651	0.408	0.760	0.461
MCMC*	0.070	0.079	0.673	0.426	0.781	0.475
TS**	0.068	0.078	0.894	0.541	1.057	0.642
TS*	0.070	0.079	0.898	0.543	1.050	0.646
Mean Squared Error						
MLE*	0.007	0.009	0.707	0.253	0.938	0.375
MCMC	0.007	0.010	0.553	0.227	0.749	0.269
MCMC*	0.007	0.009	0.600	0.250	0.803	0.290
TS**	0.007	0.009	0.887	0.330	1.233	0.457
TS*	0.007	0.009	0.880	0.325	1.200	0.456
SIF for all MCMC outputs						
Mean	5	5	176	166	180	185
Standard Deviation	3	3	13	19	11	9

* 465 samples included; ** 560 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.18: *ρ=0.8, Binary Explanatory Variables, $x_{i1} = x_{i2}$*

	β_{11}	β_{12}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	0.800	-1.000	-1.200	0.800
Mean						
MLE*	1.010	-1.512	0.420	-0.797	-0.781	0.549
MCMC	1.007	-1.509	-0.102	-0.495	-0.190	0.180
MCMC*	1.013	-1.516	-0.090	-0.500	-0.213	0.197
TS**	1.005	-1.504	0.371	-0.844	-0.782	0.525
TS*	1.009	-1.511	0.213	-0.743	-0.604	0.423
Standard Deviation						
MLE*	0.085	0.097	0.528	0.273	0.570	0.362
MCMC	0.087	0.098	0.475	0.304	0.534	0.319
MCMC*	0.086	0.098	0.453	0.291	0.507	0.306
TS**	0.085	0.094	0.645	0.414	0.758	0.447
TS*	0.085	0.097	0.577	0.369	0.680	0.405
Maximum Values						
MLE*	1.268	-1.207	1.177	0.052	0.882	0.993
MCMC	1.307	-1.204	1.087	0.657	1.675	0.951
MCMC*	1.284	-1.204	1.057	0.657	1.675	0.920
TS**	1.306	-1.260	1.427	0.511	2.000	1.000
TS*	1.268	-1.206	1.418	-0.005	0.592	1.000
Minimum Values						
MLE*	0.757	-1.805	-0.986	-1.260	-1.507	-0.381
MCMC	0.754	-1.820	-1.718	-1.165	-1.484	-0.948
MCMC*	0.754	-1.820	-1.718	-1.128	-1.381	-0.948
TS**	0.779	-1.805	-1.925	-1.625	-1.957	-1.000
TS*	0.756	-1.805	-0.933	-1.562	-1.896	-0.316
Mean Absolute Error						
MLE*	0.069	0.080	0.501	0.263	0.531	0.336
MCMC	0.070	0.079	0.909	0.511	1.016	0.624
MCMC*	0.070	0.081	0.894	0.503	0.990	0.606
TS**	0.068	0.076	0.655	0.401	0.761	0.453
TS*	0.069	0.080	0.732	0.417	0.825	0.499
Mean Squared Error						
MLE*	0.007	0.010	0.423	0.115	0.500	0.193
MCMC	0.008	0.010	1.039	0.348	1.305	0.487
MCMC*	0.008	0.010	0.997	0.335	1.231	0.457
TS**	0.007	0.009	0.600	0.195	0.749	0.275
TS*	0.007	0.010	0.677	0.202	0.817	0.306
SIF for all MCMC outputs						
Mean	6	5	178	166	182	187
Standard Deviation	4	3	15	24	13	9

* 610 samples included; ** 602 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.19: *ρ=0, Binary Explanatory Variables, $x_{i1} = x_{i2}$*

Starting values are (0.892, -1.452, 0.738, -0.943, -1.166, 0)

	β_{11}	β_{12}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	0.800	-1.000	-1.200	0.000
Mean						
MLE*	1.005	-1.504	0.804	-0.983	-1.195	0.041
MCMC	1.005	-1.506	0.739	-0.925	-1.126	0.011
MCMC*	1.005	-1.504	0.746	-0.931	-1.130	0.013
TS**	1.005	-1.504	0.757	-0.978	-1.148	-0.033
TS*	1.002	-1.501	0.753	-0.975	-1.139	-0.037
Standard Deviation						
MLE*	0.089	0.098	0.446	0.279	0.519	0.374
MCMC	0.087	0.098	0.371	0.264	0.420	0.241
MCMC*	0.089	0.099	0.374	0.266	0.423	0.243
TS**	0.083	0.096	0.852	0.550	1.010	0.606
TS*	0.088	0.098	0.710	0.460	0.842	0.505
Maximum Values						
MLE*	1.266	-1.237	1.947	0.101	0.466	0.991
MCMC	1.300	-1.210	1.710	0.136	0.412	0.838
MCMC*	1.260	-1.239	1.710	-0.025	0.209	0.768
TS**	1.306	-1.213	2.729	0.252	0.735	1.000
TS*	1.260	-1.237	2.729	0.149	0.735	1.000
Minimum Values						
MLE*	0.775	-1.800	-0.702	-1.760	-2.477	-0.722
MCMC	0.759	-1.805	-0.616	-1.583	-2.244	-0.758
MCMC*	0.775	-1.788	-0.467	-1.583	-2.140	-0.601
TS**	0.757	-1.805	-0.921	-2.286	-3.234	-1.000
TS*	0.775	-1.779	-0.887	-2.286	-3.234	-1.000
Mean Absolute Error						
MLE*	0.071	0.080	0.307	0.204	0.354	0.239
MCMC	0.070	0.080	0.297	0.215	0.338	0.190
MCMC*	0.071	0.081	0.303	0.220	0.344	0.194
TS**	0.066	0.077	0.603	0.396	0.716	0.429
TS*	0.071	0.080	0.468	0.314	0.554	0.332
Mean Squared Error						
MLE*	0.008	0.010	0.198	0.078	0.269	0.141
MCMC	0.008	0.010	0.141	0.075	0.182	0.058
MCMC*	0.008	0.010	0.142	0.076	0.184	0.059
TS**	0.007	0.009	0.727	0.303	1.021	0.367
TS*	0.008	0.010	0.505	0.212	0.712	0.256
SIF for all MCMC outputs						
Mean	5	4	175	163	178	185
Standard Deviation	2	2	13	20	12	9

* 541 samples included; ** 590 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.20: *ρ=0.5, Binary Explanatory Variables, $x_{i1} = x_{i2}$*

Starting values are (0.892, -1.452, 0.738, -0.943, -1.166, 0)

	β_{11}	β_{12}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	0.800	-1.000	-1.200	0.500
Mean						
MLE*	1.004	-1.505	0.833	-1.001	-1.219	0.549
MCMC	1.004	-1.506	0.164	-0.602	-0.452	0.051
MCMC*	1.005	-1.506	0.194	-0.620	-0.489	0.076
TS**	1.002	-1.504	0.936	-1.132	-1.381	0.575
TS*	1.003	-1.504	0.994	-1.168	-1.454	0.620
Standard Deviation						
MLE*	0.085	0.096	0.436	0.246	0.482	0.338
MCMC	0.087	0.098	0.387	0.258	0.438	0.261
MCMC*	0.086	0.097	0.381	0.254	0.431	0.258
TS**	0.086	0.098	0.835	0.537	0.989	0.583
TS*	0.085	0.096	0.732	0.472	0.867	0.513
Maximum Values						
MLE*	1.244	-1.213	1.531	0.009	0.730	0.997
MCMC	1.301	-1.202	1.482	0.576	1.398	0.958
MCMC*	1.250	-1.212	1.322	0.215	1.119	0.958
TS**	1.306	-1.213	1.931	0.500	1.493	1.000
TS*	1.246	-1.213	1.861	0.493	1.493	1.000
Minimum Values						
MLE*	0.757	-1.783	-0.763	-1.450	-1.874	-0.591
MCMC	0.752	-1.800	-1.396	-1.371	-1.786	-0.951
MCMC*	0.761	-1.791	-1.087	-1.354	-1.691	-0.761
TS**	0.757	-1.805	-1.509	-1.886	-2.490	-1.000
TS*	0.757	-1.779	-1.484	-1.886	-2.334	-1.000
Mean Absolute Error						
MLE*	0.069	0.078	0.378	0.202	0.414	0.302
MCMC	0.070	0.079	0.654	0.409	0.764	0.463
MCMC*	0.069	0.079	0.622	0.390	0.727	0.438
TS**	0.069	0.079	0.724	0.475	0.868	0.506
TS*	0.069	0.078	0.658	0.435	0.791	0.460
Mean Squared Error						
MLE*	0.007	0.009	0.191	0.060	0.233	0.116
MCMC	0.008	0.010	0.554	0.225	0.752	0.270
MCMC*	0.007	0.009	0.512	0.209	0.691	0.246
TS**	0.007	0.010	0.715	0.305	1.009	0.345
TS*	0.007	0.009	0.573	0.251	0.814	0.277
SIF for all MCMC outputs						
Mean	5	5	177	167	180	185
Standard Deviation	3	3	13	17	11	9

* 649 samples included; ** 734 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.21: *ρ=0.8, Binary Explanatory Variables, $x_{i1} = x_{i2}$*

Starting values are (0.892, -1.452, 0.738, -0.943, -1.166, 0)

	β_{11}	β_{12}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	0.800	-1.000	-1.200	0.800
Mean						
MLE*	1.006	-1.507	0.733	-0.964	-1.124	0.757
MCMC	1.007	-1.509	-0.117	-0.484	-0.173	0.169
MCMC*	1.009	-1.512	-0.102	-0.493	-0.195	0.181
TS**	1.001	-1.501	0.922	-1.194	-1.441	0.914
TS*	1.006	-1.507	0.916	-1.191	-1.437	0.911
Standard Deviation						
MLE*	0.084	0.095	0.317	0.175	0.327	0.206
MCMC	0.087	0.098	0.480	0.309	0.542	0.325
MCMC*	0.084	0.096	0.462	0.300	0.524	0.315
TS**	0.086	0.098	0.412	0.276	0.478	0.278
TS*	0.084	0.095	0.385	0.257	0.445	0.259
Maximum Values						
MLE*	1.268	-1.213	1.240	0.138	1.013	0.975
MCMC	1.307	-1.210	1.184	0.657	1.675	0.946
MCMC*	1.277	-1.216	0.954	0.649	1.641	0.897
TS**	1.268	-1.206	1.427	0.590	1.978	1.000
TS*	1.268	-1.213	1.418	0.585	1.978	1.000
Minimum Values						
MLE*	0.779	-1.805	-1.151	-1.278	-1.548	-0.459
MCMC	0.757	-1.813	-1.718	-1.237	-1.484	-0.941
MCMC*	0.782	-1.813	-1.627	-1.176	-1.324	-0.910
TS**	0.756	-1.805	-1.937	-1.625	-1.957	-1.000
TS*	0.779	-1.805	-1.937	-1.562	-1.896	-1.000
Mean Absolute Error						
MLE*	0.067	0.077	0.220	0.125	0.215	0.140
MCMC	0.070	0.079	0.924	0.521	1.033	0.634
MCMC*	0.069	0.078	0.904	0.510	1.006	0.620
TS**	0.069	0.079	0.301	0.282	0.428	0.232
TS*	0.067	0.077	0.294	0.272	0.419	0.227
Mean Squared Error						
MLE*	0.007	0.009	0.105	0.032	0.112	0.044
MCMC	0.008	0.010	1.071	0.362	1.349	0.504
MCMC*	0.007	0.009	1.026	0.347	1.284	0.482
TS**	0.007	0.010	0.184	0.114	0.286	0.090
TS*	0.007	0.009	0.161	0.102	0.254	0.079
SIF for all MCMC outputs						
Mean	6	5	179	166	182	187
Standard Deviation	4	3	14	22	12	9

* 769 samples included; ** 875 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.22: *ρ=0, Continuous Explanatory Variables (Variances 0.21), $x_{i1} = x_{i2}$*

	β_{11}	β_{12}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	0.800	-1.000	-1.200	0.000
Mean						
MLE*	1.003	-1.504	0.775	-0.971	-1.163	0.017
MCMC	1.003	-1.501	0.838	-1.000	-1.241	0.057
MCMC*	1.004	-1.504	0.847	-1.004	-1.254	0.067
TS**	1.003	-1.505	0.820	-1.009	-1.225	0.012
TS*	1.003	-1.504	0.775	-0.992	-1.170	-0.019
Standard Deviation						
MLE*	0.054	0.125	0.502	0.227	0.624	0.395
MCMC	0.054	0.125	0.337	0.174	0.419	0.252
MCMC*	0.054	0.126	0.347	0.174	0.431	0.262
TS**	0.053	0.122	0.558	0.259	0.703	0.414
TS*	0.054	0.125	0.584	0.262	0.735	0.436
Maximum Values						
MLE*	1.185	-1.156	1.874	-0.268	0.296	0.985
MCMC	1.194	-1.150	1.869	-0.349	0.072	0.937
MCMC*	1.194	-1.150	1.869	-0.349	0.035	0.937
TS**	1.193	-1.153	2.423	-0.264	0.612	1.000
TS*	1.193	-1.153	2.282	-0.264	0.576	1.000
Minimum Values						
MLE*	0.839	-1.949	-0.425	-1.568	-2.491	-0.733
MCMC	0.839	-1.944	-0.229	-1.625	-2.518	-0.742
MCMC*	0.839	-1.944	-0.206	-1.625	-2.518	-0.577
TS**	0.835	-1.925	-0.645	-1.812	-3.086	-1.000
TS*	0.835	-1.948	-0.645	-1.812	-3.059	-1.000
Mean Absolute Error						
MLE*	0.043	0.100	0.420	0.187	0.524	0.324
MCMC	0.042	0.100	0.272	0.138	0.339	0.207
MCMC*	0.043	0.100	0.284	0.138	0.356	0.221
TS**	0.042	0.098	0.450	0.207	0.567	0.338
TS*	0.043	0.100	0.479	0.212	0.604	0.360
Mean Squared Error						
MLE*	0.003	0.016	0.252	0.052	0.390	0.156
MCMC	0.003	0.016	0.115	0.030	0.177	0.067
MCMC*	0.003	0.016	0.122	0.030	0.188	0.073
TS**	0.003	0.015	0.311	0.067	0.495	0.171
TS*	0.003	0.016	0.341	0.069	0.541	0.190
SIF for all MCMC outputs						
Mean	8	9	156	113	158	170
Standard Deviation	3	4	14	20	14	13

* 658 samples included; ** 872 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.23: *ρ=0.5, Continuous Explanatory Variables (Variances 0.21), $x_{i1} = x_{i2}$*

	β_{11}	β_{12}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	0.800	-1.000	-1.200	0.500
Mean						
MLE*	1.006	-1.500	0.766	-0.972	-1.145	0.481
MCMC	1.005	-1.507	0.555	-0.899	-0.894	0.321
MCMC*	1.008	-1.504	0.584	-0.909	-0.928	0.341
TS**	1.003	-1.502	0.810	-1.042	-1.234	0.501
TS*	1.005	-1.498	0.817	-1.043	-1.242	0.505
Standard Deviation						
MLE*	0.054	0.123	0.429	0.172	0.510	0.318
MCMC	0.054	0.124	0.321	0.149	0.387	0.233
MCMC*	0.054	0.123	0.309	0.143	0.370	0.221
TS**	0.054	0.124	0.478	0.220	0.600	0.345
TS*	0.054	0.123	0.480	0.220	0.599	0.344
Maximum Values						
MLE*	1.190	-1.187	1.508	-0.330	0.725	0.999
MCMC	1.196	-1.152	1.356	-0.409	0.543	0.904
MCMC*	1.196	-1.184	1.334	-0.409	0.462	0.904
TS**	1.193	-1.153	1.811	-0.350	0.819	1.000
TS*	1.193	-1.186	1.811	-0.350	0.740	1.000
Minimum Values						
MLE*	0.839	-1.925	-0.772	-1.402	-1.966	-0.512
MCMC	0.839	-1.935	-0.609	-1.351	-1.795	-0.374
MCMC*	0.839	-1.930	-0.565	-1.351	-1.778	-0.374
TS**	0.835	-1.948	-0.839	-1.651	-2.445	-0.609
TS*	0.835	-1.925	-0.749	-1.651	-2.445	-0.609
Mean Absolute Error						
MLE*	0.042	0.097	0.347	0.133	0.409	0.258
MCMC	0.043	0.099	0.318	0.141	0.386	0.233
MCMC*	0.043	0.097	0.289	0.132	0.350	0.212
TS**	0.042	0.100	0.384	0.179	0.484	0.278
TS*	0.042	0.098	0.389	0.181	0.488	0.278
Mean Squared Error						
MLE*	0.003	0.015	0.185	0.030	0.262	0.101
MCMC	0.003	0.015	0.163	0.032	0.244	0.086
MCMC*	0.003	0.015	0.142	0.029	0.210	0.074
TS**	0.003	0.015	0.229	0.050	0.361	0.119
TS*	0.003	0.015	0.231	0.050	0.360	0.118
SIF for all MCMC outputs						
Mean	9	10	158	103	159	170
Standard Deviation	4	4	15	24	14	13

* 639 samples included; ** 995 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.24: *ρ=0.8, Continuous Explanatory Variables (Variances 0.21), $x_{i1} = x_{i2}$*

	β_{11}	β_{12}	β_{21}	β_{22}	α	ρ
True Values	1.000	-1.500	0.800	-1.000	-1.200	0.800
Mean						
MLE*	1.007	-1.511	0.675	-0.958	-1.058	0.724
MCMC	1.010	-1.521	0.409	-0.884	-0.752	0.551
MCMC*	1.012	-1.523	0.398	-0.878	-0.743	0.548
TS**	1.003	-1.501	0.737	-1.102	-1.222	0.845
TS*	1.006	-1.504	0.701	-1.084	-1.179	0.824
Standard Deviation						
MLE*	0.054	0.123	0.337	0.125	0.375	0.201
MCMC	0.053	0.122	0.319	0.130	0.367	0.198
MCMC*	0.054	0.121	0.306	0.121	0.350	0.189
TS**	0.053	0.125	0.320	0.163	0.398	0.227
TS*	0.054	0.124	0.339	0.163	0.419	0.238
Maximum Values						
MLE*	1.176	-1.201	1.220	-0.475	0.380	0.992
MCMC	1.190	-1.143	1.222	-0.366	0.514	0.972
MCMC*	1.181	-1.194	1.160	-0.366	0.491	0.972
TS**	1.193	-1.153	1.218	-0.429	0.646	1.000
TS*	1.173	-1.186	1.198	-0.429	0.646	1.000
Minimum Values						
MLE*	0.858	-1.947	-0.591	-1.350	-1.668	-0.133
MCMC	0.855	-1.962	-0.672	-1.264	-1.688	-0.187
MCMC*	0.862	-1.962	-0.672	-1.227	-1.547	-0.187
TS**	0.835	-1.948	-0.776	-1.509	-1.808	-0.238
TS*	0.857	-1.948	-0.776	-1.439	-1.766	-0.238
Mean Absolute Error						
MLE*	0.043	0.097	0.263	0.100	0.288	0.153
MCMC	0.043	0.098	0.414	0.140	0.471	0.261
MCMC*	0.043	0.097	0.418	0.139	0.474	0.261
TS**	0.042	0.100	0.231	0.162	0.311	0.194
TS*	0.043	0.098	0.252	0.154	0.328	0.202
Mean Squared Error						
MLE*	0.003	0.015	0.129	0.017	0.160	0.046
MCMC	0.003	0.015	0.255	0.030	0.335	0.101
MCMC*	0.003	0.015	0.255	0.030	0.331	0.099
TS**	0.003	0.016	0.106	0.037	0.159	0.053
TS*	0.003	0.015	0.125	0.033	0.176	0.057
SIF for all MCMC outputs						
Mean	12	13	163	88	163	175
Standard Deviation	7	7	14	28	14	11

* 655 samples included; ** 960 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models

Table 3.25: *ρ=0, Two More Binary Explanatory Variables, $x_{i1} = x_{i2}$*

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	α	ρ
True Values	1.000	-1.500	-1.000	0.800	-1.000	1.500	-1.200	0.000
Mean								
MLE*	1.006	-1.510	-1.008	0.445	-0.739	1.535	-0.782	-0.197
MCMC	1.008	-1.509	-1.006	0.697	-0.908	1.483	-1.087	-0.041
MCMC*	1.008	-1.513	-1.010	0.741	-0.937	1.466	-1.138	-0.013
TS**	1.004	-1.508	-1.004	1.027	-1.153	1.445	-1.477	0.148
TS*	1.005	-1.510	-1.007	0.994	-1.134	1.451	-1.438	0.126
Standard Deviation								
MLE*	0.100	0.105	0.104	0.599	0.377	0.238	0.703	0.399
MCMC	0.098	0.103	0.103	0.375	0.266	0.146	0.434	0.218
MCMC*	0.099	0.105	0.103	0.384	0.266	0.155	0.443	0.230
TS**	0.095	0.105	0.101	0.951	0.603	0.321	1.143	0.628
TS*	0.099	0.105	0.102	0.934	0.591	0.318	1.120	0.617
Maximum Values								
MLE*	1.319	-1.237	-0.667	2.113	0.420	2.007	0.819	0.824
MCMC	1.333	-1.240	-0.673	2.078	0.169	1.924	0.666	0.720
MCMC*	1.333	-1.240	-0.679	2.078	0.169	1.888	0.666	0.720
TS**	1.329	-1.240	-0.657	2.904	0.277	2.293	0.830	1.000
TS*	1.329	-1.240	-0.679	2.760	0.134	2.268	0.830	1.000
Minimum Values								
MLE*	0.689	-1.808	-1.341	-0.897	-1.792	0.722	-2.617	-0.961
MCMC	0.699	-1.832	-1.340	-0.761	-1.895	0.864	-2.552	-0.792
MCMC*	0.699	-1.814	-1.340	-0.761	-1.822	0.864	-2.552	-0.792
TS**	0.696	-1.811	-1.301	-0.987	-2.717	0.779	-3.626	-1.000
TS*	0.696	-1.811	-1.328	-0.944	-2.423	0.779	-3.528	-1.000
Mean Absolute Error								
MLE*	0.079	0.083	0.082	0.657	0.423	0.201	0.774	0.418
MCMC	0.079	0.083	0.081	0.313	0.225	0.116	0.360	0.178
MCMC*	0.079	0.084	0.081	0.314	0.217	0.124	0.359	0.186
TS**	0.077	0.084	0.080	0.837	0.526	0.278	1.008	0.555
TS*	0.079	0.083	0.081	0.813	0.510	0.273	0.977	0.537
Mean Squared Error								
MLE*	0.010	0.011	0.011	0.484	0.210	0.058	0.668	0.198
MCMC	0.010	0.011	0.011	0.151	0.079	0.022	0.201	0.049
MCMC*	0.010	0.011	0.011	0.151	0.074	0.025	0.199	0.053
TS**	0.009	0.011	0.010	0.955	0.386	0.106	1.381	0.416
TS*	0.010	0.011	0.011	0.909	0.367	0.103	1.308	0.396
SIF for all MCMC outputs								
Mean	5	5	5	163	148	110	169	178
Standard Deviation	2	2	2	16	24	30	14	11

* 539 samples included; ** 739 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.26: *ρ=0.5, Two More Binary Explanatory Variables, $x_{i1} = x_{i2}$*

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	α	ρ
True Values	1.000	-1.500	-1.000	0.800	-1.000	1.500	-1.200	0.500
Mean								
MLE*	1.007	-1.508	-1.009	0.566	-0.863	1.586	-0.916	0.332
MCMC	1.008	-1.510	-1.006	0.322	-0.733	1.698	-0.622	0.166
MCMC*	1.009	-1.512	-1.011	0.412	-0.777	1.651	-0.726	0.220
TS**	1.004	-1.503	-1.006	1.067	-1.233	1.543	-1.517	0.670
TS*	1.005	-1.506	-1.008	1.028	-1.207	1.549	-1.472	0.643
Standard Deviation								
MLE*	0.096	0.104	0.101	0.528	0.280	0.306	0.635	0.364
MCMC	0.098	0.103	0.103	0.432	0.268	0.206	0.509	0.274
MCMC*	0.096	0.104	0.101	0.437	0.271	0.206	0.510	0.276
TS**	0.097	0.104	0.103	0.698	0.436	0.265	0.828	0.459
TS*	0.095	0.104	0.100	0.698	0.433	0.266	0.824	0.463
Maximum Values								
MLE*	1.318	-1.240	-0.674	1.652	0.665	2.176	1.484	0.984
MCMC	1.329	-1.242	-0.662	1.625	0.240	2.202	1.232	0.951
MCMC*	1.329	-1.242	-0.689	1.625	0.240	2.113	1.232	0.951
TS**	1.329	-1.240	-0.679	2.183	0.426	2.613	1.645	1.000
TS*	1.329	-1.240	-0.679	2.183	0.008	2.454	1.144	1.000
Minimum Values								
MLE*	0.692	-1.811	-1.330	-1.578	-1.532	0.890	-2.420	-0.958
MCMC	0.706	-1.817	-1.335	-1.233	-1.558	1.006	-2.328	-0.824
MCMC*	0.706	-1.816	-1.335	-1.233	-1.558	1.015	-2.328	-0.824
TS**	0.696	-1.819	-1.326	-1.589	-1.954	1.042	-2.606	-1.000
TS*	0.696	-1.819	-1.328	-1.205	-1.947	1.042	-2.606	-1.000
Mean Absolute Error								
MLE*	0.077	0.082	0.080	0.475	0.250	0.274	0.572	0.329
MCMC	0.079	0.083	0.082	0.534	0.306	0.243	0.638	0.361
MCMC*	0.077	0.083	0.080	0.466	0.274	0.214	0.557	0.317
TS**	0.078	0.083	0.082	0.690	0.450	0.207	0.825	0.457
TS*	0.077	0.082	0.080	0.697	0.443	0.220	0.830	0.466
Mean Squared Error								
MLE*	0.009	0.011	0.010	0.333	0.097	0.101	0.483	0.161
MCMC	0.010	0.011	0.011	0.415	0.143	0.081	0.593	0.187
MCMC*	0.009	0.011	0.010	0.341	0.123	0.065	0.484	0.155
TS**	0.009	0.011	0.011	0.558	0.244	0.072	0.785	0.240
TS*	0.009	0.011	0.010	0.538	0.230	0.073	0.752	0.234
SIF for all MCMC outputs								
Mean	5	6	5	162	138	129	169	180
Standard Deviation	2	2	2	20	36	26	17	11

* 590 samples included; ** 774 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.27: *ρ=0.8, Two More Binary Explanatory Variables, $x_{i1} = x_{i2}$*

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	α	ρ
True Values	1.000	-1.500	-1.000	0.800	-1.000	1.500	-1.200	0.800
Mean								
MLE*	1.007	-1.512	-1.005	0.733	-0.980	1.549	-1.119	0.756
MCMC	1.009	-1.515	-1.009	0.383	-0.835	1.776	-0.697	0.524
MCMC*	1.012	-1.520	-1.011	0.367	-0.828	1.783	-0.680	0.514
TS**	1.004	-1.507	-1.002	0.890	-1.275	1.835	-1.331	0.921
TS*	1.007	-1.510	-1.005	0.836	-1.241	1.847	-1.265	0.884
Standard Deviation								
MLE*	0.095	0.101	0.100	0.308	0.156	0.242	0.354	0.174
MCMC	0.097	0.103	0.103	0.479	0.252	0.283	0.559	0.293
MCMC*	0.096	0.102	0.101	0.448	0.242	0.262	0.523	0.273
TS**	0.097	0.102	0.102	0.353	0.235	0.179	0.409	0.216
TS*	0.096	0.101	0.100	0.398	0.259	0.185	0.474	0.253
Maximum Values								
MLE*	1.311	-1.246	-0.686	1.454	0.387	2.389	1.594	0.993
MCMC	1.320	-1.232	-0.656	1.383	0.682	2.582	1.905	0.985
MCMC*	1.320	-1.251	-0.693	1.383	0.628	2.554	1.905	0.956
TS**	1.329	-1.240	-0.679	1.710	-0.515	2.441	0.164	1.000
TS*	1.329	-1.241	-0.679	1.659	-0.270	2.608	0.634	1.000
Minimum Values								
MLE*	0.707	-1.807	-1.321	-1.637	-1.399	1.028	-1.755	-0.837
MCMC	0.705	-1.817	-1.332	-1.986	-1.388	1.060	-1.667	-0.907
MCMC*	0.705	-1.817	-1.332	-1.916	-1.388	1.092	-1.631	-0.907
TS**	0.696	-1.819	-1.328	-0.358	-1.842	1.317	-2.011	0.054
TS*	0.696	-1.819	-1.328	-0.853	-1.808	1.317	-1.895	-0.150
Mean Absolute Error								
MLE*	0.077	0.081	0.079	0.214	0.118	0.181	0.237	0.104
MCMC	0.078	0.083	0.082	0.467	0.217	0.317	0.555	0.292
MCMC*	0.077	0.082	0.080	0.467	0.216	0.314	0.556	0.294
TS**	0.078	0.081	0.081	0.284	0.328	0.337	0.349	0.229
TS*	0.077	0.081	0.079	0.313	0.318	0.348	0.386	0.243
Mean Squared Error								
MLE*	0.009	0.010	0.010	0.099	0.025	0.061	0.132	0.032
MCMC	0.010	0.011	0.011	0.403	0.090	0.156	0.565	0.162
MCMC*	0.009	0.011	0.010	0.387	0.088	0.148	0.544	0.157
TS**	0.009	0.010	0.010	0.133	0.130	0.144	0.185	0.061
TS*	0.009	0.010	0.010	0.159	0.125	0.155	0.229	0.071
SIF for all MCMC outputs								
Mean	7	9	8	151	103	139	161	183
Standard Deviation	4	5	5	32	53	25	26	11

* 784 samples included; ** 920 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.28: *ρ=0, Two More Binary Explanatory Variables, $x_{i1} = x_{i2}$*

Starting values are (0.835,-1.419,-0.820,0.816,-0.959,1.374,-1.158,0)

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	α	ρ
True Values	1.000	-1.500	-1.000	0.800	-1.000	1.500	-1.200	0.000
Mean								
MLE*	1.008	-1.506	-1.007	0.949	-1.068	1.404	-1.389	0.128
MCMC	1.008	-1.509	-1.006	0.701	-0.912	1.484	-1.091	-0.039
MCMC*	1.010	-1.509	-1.008	0.754	-0.944	1.464	-1.154	-0.003
TS**	1.004	-1.507	-1.003	1.014	-1.145	1.450	-1.461	0.136
TS*	1.007	-1.505	-1.006	1.074	-1.183	1.430	-1.534	0.181
Standard Deviation								
MLE*	0.097	0.100	0.100	0.521	0.326	0.228	0.630	0.361
MCMC	0.098	0.103	0.103	0.362	0.255	0.145	0.419	0.211
MCMC*	0.097	0.100	0.100	0.379	0.262	0.149	0.444	0.227
TS**	0.098	0.101	0.102	0.956	0.605	0.322	1.153	0.639
TS*	0.097	0.100	0.099	0.908	0.575	0.307	1.096	0.603
Maximum Values								
MLE*	1.320	-1.241	-0.683	2.242	0.419	1.910	0.819	0.955
MCMC	1.333	-1.239	-0.673	1.975	-0.161	1.924	0.071	0.734
MCMC*	1.333	-1.239	-0.679	1.975	-0.161	1.834	0.068	0.734
TS**	1.329	-1.240	-0.679	2.904	0.277	2.293	0.830	1.000
TS*	1.329	-1.240	-0.679	2.904	0.277	2.293	0.792	1.000
Minimum Values								
MLE*	0.704	-1.803	-1.327	-0.896	-2.086	0.765	-2.857	-0.961
MCMC	0.699	-1.832	-1.346	-0.333	-1.801	0.954	-2.509	-0.602
MCMC*	0.699	-1.806	-1.327	-0.333	-1.801	0.954	-2.509	-0.574
TS**	0.696	-1.819	-1.328	-0.987	-2.717	0.779	-3.626	-1.000
TS*	0.696	-1.804	-1.328	-0.919	-2.717	0.779	-3.626	-1.000
Mean Absolute Error								
MLE*	0.078	0.080	0.079	0.423	0.261	0.191	0.515	0.297
MCMC	0.079	0.083	0.081	0.304	0.219	0.115	0.349	0.173
MCMC*	0.079	0.080	0.079	0.309	0.216	0.121	0.361	0.183
TS**	0.080	0.081	0.080	0.797	0.505	0.271	0.967	0.537
TS*	0.078	0.080	0.079	0.752	0.478	0.259	0.911	0.501
Mean Squared Error								
MLE*	0.010	0.010	0.010	0.293	0.111	0.061	0.432	0.146
MCMC	0.010	0.011	0.011	0.141	0.073	0.021	0.187	0.046
MCMC*	0.010	0.010	0.010	0.145	0.072	0.024	0.199	0.051
TS**	0.010	0.010	0.010	0.958	0.387	0.106	1.396	0.426
TS*	0.009	0.010	0.010	0.899	0.364	0.099	1.311	0.396
SIF for all MCMC outputs								
Mean	5	5	5	163	148	111	169	178
Standard Deviation	2	2	2	16	24	29	14	11

* 636 samples included; ** 720 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.29: *ρ=0.5, Two More Binary Explanatory Variables, $x_{i1} = x_{i2}$*

Starting values are (0.835,-1.419,-0.820,0.816,-0.959,1.374,-1.158,0)

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	α	ρ
True Values	1.000	-1.500	-1.000	0.800	-1.000	1.500	-1.200	0.500
Mean								
MLE*	1.006	-1.507	-1.005	0.834	-1.007	1.463	-1.244	0.521
MCMC	1.008	-1.510	-1.006	0.332	-0.739	1.694	-0.634	0.172
MCMC*	1.008	-1.510	-1.007	0.401	-0.776	1.662	-0.716	0.217
TS**	1.005	-1.507	-1.003	1.154	-1.287	1.514	-1.630	0.737
TS*	1.005	-1.506	-1.004	1.209	-1.320	1.493	-1.695	0.773
Standard Deviation								
MLE*	0.099	0.103	0.105	0.335	0.192	0.223	0.395	0.214
MCMC	0.098	0.103	0.103	0.432	0.272	0.206	0.507	0.273
MCMC*	0.099	0.103	0.105	0.416	0.264	0.202	0.487	0.263
TS**	0.098	0.103	0.104	0.613	0.386	0.237	0.731	0.405
TS*	0.098	0.103	0.105	0.530	0.341	0.209	0.634	0.350
Maximum Values								
MLE*	1.318	-1.237	-0.657	1.703	0.017	2.156	0.764	0.984
MCMC	1.327	-1.241	-0.655	1.586	0.459	2.202	1.268	0.940
MCMC*	1.327	-1.241	-0.655	1.586	0.459	2.202	1.268	0.940
TS**	1.329	-1.241	-0.657	2.183	0.426	2.613	1.645	1.000
TS*	1.329	-1.240	-0.657	2.183	0.374	2.549	1.645	1.000
Minimum Values								
MLE*	0.705	-1.813	-1.330	-0.813	-1.543	0.890	-2.420	-0.616
MCMC	0.698	-1.820	-1.335	-1.301	-1.534	0.976	-2.269	-0.888
MCMC*	0.698	-1.820	-1.335	-1.301	-1.534	0.976	-2.269	-0.888
TS**	0.696	-1.819	-1.328	-1.589	-1.954	1.042	-2.606	-1.000
TS*	0.696	-1.819	-1.328	-1.567	-1.954	1.042	-2.606	-1.000
Mean Absolute Error								
MLE*	0.080	0.082	0.084	0.263	0.146	0.182	0.310	0.169
MCMC	0.079	0.083	0.081	0.526	0.302	0.241	0.627	0.355
MCMC*	0.080	0.083	0.084	0.467	0.272	0.218	0.556	0.315
TS**	0.080	0.082	0.083	0.618	0.419	0.174	0.746	0.415
TS*	0.079	0.082	0.083	0.597	0.412	0.159	0.722	0.403
Mean Squared Error								
MLE*	0.010	0.011	0.011	0.113	0.037	0.051	0.158	0.046
MCMC	0.010	0.011	0.011	0.406	0.142	0.080	0.578	0.182
MCMC*	0.010	0.011	0.011	0.332	0.120	0.067	0.471	0.149
TS**	0.010	0.011	0.011	0.501	0.232	0.056	0.719	0.220
TS*	0.010	0.011	0.011	0.448	0.219	0.044	0.647	0.197
SIF for all MCMC outputs								
Mean	5	6	5	161	137	129	169	179
Standard Deviation	2	2	2	21	36	26	17	12

* 844 samples included; ** 887 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.30: *ρ=0.8, Two More Binary Explanatory Variables, $x_{i1} = x_{i2}$*

Starting values are (0.835,-1.419,-0.820,0.816,-0.959,1.374,-1.158,0)

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	α	ρ
True Values	1.000	-1.500	-1.000	0.800	-1.000	1.500	-1.200	0.800
Mean								
MLE*	1.002	-1.507	-1.001	0.795	-1.003	1.509	-1.192	0.796
MCMC	1.009	-1.514	-1.009	0.396	-0.842	1.770	-0.712	0.532
MCMC*	1.007	-1.515	-1.007	0.398	-0.840	1.767	-0.717	0.535
TS**	1.005	-1.506	-1.003	0.963	-1.318	1.810	-1.418	0.973
TS*	1.004	-1.506	-1.002	0.967	-1.317	1.805	-1.426	0.978
Standard Deviation								
MLE*	0.096	0.104	0.103	0.229	0.133	0.201	0.252	0.109
MCMC	0.097	0.103	0.103	0.467	0.244	0.282	0.549	0.289
MCMC*	0.097	0.104	0.103	0.453	0.239	0.274	0.535	0.281
TS**	0.097	0.103	0.102	0.254	0.184	0.158	0.269	0.126
TS*	0.097	0.104	0.103	0.227	0.172	0.151	0.232	0.096
Maximum Values								
MLE*	1.311	-1.214	-0.660	1.454	-0.510	2.194	-0.072	1.000
MCMC	1.321	-1.232	-0.659	1.317	0.564	2.605	1.755	0.965
MCMC*	1.321	-1.232	-0.659	1.317	0.564	2.605	1.755	0.965
TS**	1.329	-1.240	-0.657	1.711	-0.233	2.577	0.897	1.000
TS*	1.329	-1.240	-0.657	1.711	-0.438	2.435	0.201	1.000
Minimum Values								
MLE*	0.707	-1.805	-1.322	-0.187	-1.399	0.962	-1.759	0.207
MCMC	0.709	-1.816	-1.329	-1.666	-1.327	1.088	-1.626	-0.924
MCMC*	0.709	-1.816	-1.329	-1.666	-1.327	1.088	-1.626	-0.924
TS**	0.696	-1.819	-1.328	-0.899	-1.843	1.317	-2.011	-0.308
TS*	0.696	-1.819	-1.328	-0.440	-1.843	1.317	-2.011	0.116
Mean Absolute Error								
MLE*	0.077	0.083	0.082	0.180	0.106	0.159	0.198	0.083
MCMC	0.078	0.083	0.082	0.456	0.209	0.312	0.542	0.285
MCMC*	0.078	0.084	0.082	0.448	0.207	0.307	0.534	0.281
TS**	0.078	0.082	0.081	0.236	0.333	0.311	0.287	0.203
TS*	0.078	0.083	0.081	0.228	0.326	0.307	0.279	0.198
Mean Squared Error								
MLE*	0.009	0.011	0.011	0.052	0.018	0.040	0.064	0.012
MCMC	0.009	0.011	0.011	0.382	0.085	0.153	0.539	0.155
MCMC*	0.009	0.011	0.011	0.367	0.083	0.146	0.519	0.149
TS**	0.009	0.011	0.010	0.091	0.135	0.121	0.120	0.046
TS*	0.009	0.011	0.011	0.079	0.130	0.116	0.105	0.041
SIF for all MCMC outputs								
Mean	8	9	8	150	101	138	160	182
Standard Deviation	5	6	5	32	53	25	26	11

* 919 samples included; ** 983 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.31: *ρ=0, Continuous and Binary Explanatory Variables (A), $x_{i1} = x_{i2}$*

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	α	ρ
True Values	1.000	-1.500	-1.000	0.800	-1.000	1.500	-1.200	0.000
Mean								
MLE*	1.004	-1.514	-1.000	0.779	-0.987	1.484	-1.167	-0.016
MCMC	1.007	-1.514	-1.004	0.819	-1.007	1.478	-1.223	0.014
MCMC*	1.006	-1.518	-1.002	0.823	-1.010	1.478	-1.229	0.015
TS**	1.006	-1.512	-1.004	0.805	-1.011	1.514	-1.208	-0.006
TS*	1.004	-1.515	-1.000	0.811	-1.015	1.513	-1.217	-0.003
Standard Deviation								
MLE*	0.109	0.090	0.127	0.393	0.186	0.173	0.533	0.326
MCMC	0.108	0.088	0.125	0.280	0.137	0.148	0.367	0.215
MCMC*	0.109	0.090	0.126	0.288	0.141	0.149	0.380	0.224
TS**	0.107	0.088	0.124	0.386	0.180	0.173	0.517	0.321
TS*	0.108	0.090	0.126	0.396	0.185	0.174	0.533	0.330
Maximum Values								
MLE*	1.382	-1.293	-0.605	1.963	-0.457	1.996	0.242	0.889
MCMC	1.388	-1.301	-0.617	1.673	-0.511	1.951	0.002	0.690
MCMC*	1.388	-1.301	-0.617	1.673	-0.511	1.951	0.002	0.690
TS**	1.387	-1.303	-0.616	2.057	-0.499	2.114	0.546	0.986
TS*	1.387	-1.303	-0.616	2.057	-0.499	2.114	0.546	0.986
Minimum Values								
MLE*	0.707	-1.851	-1.457	-0.326	-1.500	0.920	-2.753	-0.770
MCMC	0.679	-1.857	-1.464	-0.230	-1.431	0.988	-2.369	-0.684
MCMC*	0.679	-1.857	-1.464	-0.230	-1.431	0.988	-2.369	-0.684
TS**	0.683	-1.849	-1.470	-0.432	-1.632	0.963	-2.784	-1.000
TS*	0.683	-1.849	-1.470	-0.432	-1.632	0.963	-2.689	-1.000
Mean Absolute Error								
MLE*	0.089	0.072	0.101	0.327	0.152	0.135	0.442	0.274
MCMC	0.087	0.071	0.100	0.225	0.110	0.119	0.295	0.174
MCMC*	0.089	0.073	0.101	0.233	0.114	0.119	0.309	0.183
TS**	0.087	0.070	0.099	0.312	0.143	0.136	0.417	0.263
TS*	0.088	0.072	0.100	0.322	0.148	0.135	0.433	0.272
Mean Squared Error								
MLE*	0.012	0.008	0.016	0.154	0.035	0.030	0.284	0.106
MCMC	0.012	0.008	0.016	0.079	0.019	0.022	0.135	0.047
MCMC*	0.012	0.008	0.016	0.083	0.020	0.023	0.145	0.050
TS**	0.012	0.008	0.015	0.149	0.033	0.030	0.267	0.103
TS*	0.012	0.008	0.016	0.157	0.035	0.030	0.284	0.109
SIF for all MCMC outputs								
Mean	9	17	8	122	99	66	130	154
Standard Deviation	3	5	2	19	20	31	18	15

* 765 samples included; ** 953 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.32: *ρ=0.5, Continuous and Binary Explanatory Variables (A), $x_{i1} = x_{i2}$*

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	α	ρ
True Values	1.000	-1.500	-1.000	0.800	-1.000	1.500	-1.200	0.500
Mean								
MLE*	1.008	-1.511	-1.005	0.770	-0.986	1.501	-1.154	0.471
MCMC	1.011	-1.519	-1.008	0.633	-0.947	1.595	-0.977	0.354
MCMC*	1.012	-1.518	-1.010	0.647	-0.951	1.586	-0.993	0.365
TS**	1.006	-1.511	-1.004	0.807	-1.054	1.623	-1.235	0.519
TS*	1.007	-1.510	-1.005	0.804	-1.052	1.623	-1.229	0.516
Standard Deviation								
MLE*	0.107	0.087	0.123	0.343	0.132	0.237	0.437	0.261
MCMC	0.107	0.088	0.124	0.277	0.116	0.195	0.352	0.206
MCMC*	0.107	0.087	0.124	0.273	0.114	0.192	0.344	0.200
TS**	0.107	0.088	0.125	0.356	0.166	0.174	0.472	0.289
TS*	0.108	0.087	0.124	0.363	0.166	0.174	0.479	0.292
Maximum Values								
MLE*	1.372	-1.309	-0.644	1.535	-0.509	2.160	0.624	0.932
MCMC	1.384	-1.309	-0.626	1.387	-0.573	2.190	0.369	0.863
MCMC*	1.384	-1.312	-0.633	1.387	-0.573	2.190	0.369	0.863
TS**	1.387	-1.303	-0.616	1.744	-0.614	2.276	0.562	1.000
TS*	1.387	-1.304	-0.630	1.744	-0.600	2.192	0.562	1.000
Minimum Values								
MLE*	0.681	-1.856	-1.447	-0.602	-1.348	0.781	-2.067	-0.650
MCMC	0.687	-1.851	-1.463	-0.431	-1.313	0.932	-1.892	-0.400
MCMC*	0.687	-1.851	-1.463	-0.431	-1.313	0.932	-1.892	-0.400
TS**	0.683	-1.849	-1.470	-0.522	-1.537	1.091	-2.284	-0.450
TS*	0.683	-1.849	-1.470	-0.522	-1.537	1.091	-2.284	-0.450
Mean Absolute Error								
MLE*	0.087	0.070	0.099	0.267	0.102	0.190	0.338	0.198
MCMC	0.087	0.071	0.100	0.257	0.101	0.175	0.326	0.196
MCMC*	0.087	0.071	0.100	0.248	0.098	0.169	0.313	0.188
TS**	0.087	0.070	0.100	0.291	0.141	0.169	0.386	0.236
TS*	0.087	0.070	0.100	0.295	0.140	0.169	0.390	0.237
Mean Squared Error								
MLE*	0.011	0.008	0.015	0.118	0.018	0.056	0.193	0.069
MCMC	0.012	0.008	0.016	0.105	0.016	0.047	0.174	0.063
MCMC*	0.012	0.008	0.015	0.098	0.015	0.044	0.161	0.058
TS**	0.012	0.008	0.016	0.126	0.030	0.045	0.224	0.084
TS*	0.012	0.008	0.015	0.132	0.030	0.045	0.230	0.085
SIF for all MCMC outputs								
Mean	10	20	9	126	83	101	133	160
Standard Deviation	3	6	3	19	24	25	18	14

* 830 samples included; ** 973 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.33: *ρ=0.8, Continuous and Binary Explanatory Variables (A), $x_{i1} = x_{i2}$*

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	α	ρ
True Values	1.000	-1.500	-1.000	0.800	-1.000	1.500	-1.200	0.800
Mean								
MLE*	1.008	-1.519	-1.008	0.719	-0.989	1.574	-1.105	0.752
MCMC	1.013	-1.525	-1.013	0.548	-0.957	1.724	-0.891	0.628
MCMC*	1.016	-1.531	-1.018	0.537	-0.955	1.732	-0.881	0.622
TS**	1.004	-1.512	-1.002	0.613	-1.084	1.935	-1.049	0.770
TS*	1.007	-1.516	-1.007	0.611	-1.082	1.932	-1.048	0.768
Standard Deviation								
MLE*	0.104	0.085	0.120	0.297	0.095	0.265	0.355	0.168
MCMC	0.106	0.087	0.124	0.303	0.105	0.251	0.373	0.187
MCMC*	0.104	0.086	0.121	0.283	0.098	0.244	0.345	0.169
TS**	0.107	0.087	0.125	0.322	0.146	0.179	0.420	0.239
TS*	0.106	0.087	0.122	0.322	0.144	0.180	0.419	0.237
Maximum Values								
MLE*	1.340	-1.301	-0.637	1.363	-0.543	2.705	0.663	0.980
MCMC	1.392	-1.310	-0.638	1.324	-0.298	2.751	1.194	0.979
MCMC*	1.392	-1.310	-0.638	1.137	-0.529	2.751	0.598	0.968
TS**	1.387	-1.303	-0.616	1.282	-0.463	2.525	0.865	1.000
TS*	1.387	-1.303	-0.616	1.282	-0.531	2.795	0.470	1.000
Minimum Values								
MLE*	0.662	-1.867	-1.427	-0.496	-1.278	0.988	-1.872	-0.330
MCMC	0.674	-1.889	-1.479	-0.768	-1.236	1.069	-1.789	-0.628
MCMC*	0.674	-1.889	-1.479	-0.560	-1.229	1.069	-1.694	-0.287
TS**	0.683	-1.849	-1.470	-0.801	-1.424	1.441	-1.802	-0.254
TS*	0.683	-1.849	-1.470	-0.454	-1.424	1.441	-1.802	-0.194
Mean Absolute Error								
MLE*	0.083	0.069	0.096	0.221	0.074	0.207	0.259	0.116
MCMC	0.086	0.072	0.099	0.301	0.088	0.268	0.366	0.190
MCMC*	0.084	0.072	0.097	0.300	0.085	0.271	0.364	0.190
TS**	0.087	0.070	0.100	0.280	0.138	0.435	0.340	0.193
TS*	0.085	0.070	0.097	0.283	0.135	0.432	0.342	0.194
Mean Squared Error								
MLE*	0.011	0.008	0.015	0.094	0.009	0.076	0.135	0.030
MCMC	0.011	0.008	0.016	0.155	0.013	0.113	0.234	0.064
MCMC*	0.011	0.008	0.015	0.149	0.012	0.113	0.221	0.060
TS**	0.012	0.008	0.016	0.138	0.028	0.221	0.199	0.058
TS*	0.011	0.008	0.015	0.139	0.027	0.219	0.198	0.057
SIF for all MCMC outputs								
Mean	16	31	14	131	65	123	137	174
Standard Deviation	8	14	8	24	28	22	2	12

* 764 samples included; ** 984 samples included; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.34: *ρ=0, Continuous and Binary Explanatory Variables (B), $x_{i1} = x_{i2}$*

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	α	ρ
True Values	1.000	-1.500	-1.000	0.800	-1.000	1.500	-1.200	0.000
Mean								
MLE*	0.996	-1.500	-0.993	0.547	-0.854	1.522	-0.878	-0.161
MCMC	1.003	-1.504	-1.000	0.837	-1.009	1.435	-1.239	0.037
MCMC*	1.000	-1.505	-0.996	0.829	-1.010	1.442	-1.232	0.033
TS**	1.001	-1.503	-0.998	0.882	-1.051	1.482	-1.299	0.051
TS*	0.998	-1.503	-0.994	0.823	-1.025	1.506	-1.229	0.009
Standard Deviation								
MLE*	0.093	0.110	0.106	0.519	0.291	0.236	0.651	0.380
MCMC	0.092	0.114	0.105	0.342	0.216	0.175	0.411	0.231
MCMC*	0.093	0.111	0.106	0.339	0.218	0.172	0.409	0.230
TS**	0.093	0.114	0.105	0.689	0.384	0.280	0.861	0.510
TS*	0.093	0.111	0.106	0.684	0.382	0.277	0.859	0.511
Maximum Values								
MLE*	1.339	-1.067	-0.721	2.025	-0.140	2.018	0.524	0.935
MCMC	1.339	-1.065	-0.718	1.804	-0.287	1.945	0.123	0.793
MCMC*	1.339	-1.065	-0.718	1.804	-0.287	1.945	0.097	0.793
TS**	1.335	-1.068	-0.725	2.517	-0.007	2.248	0.810	1.000
TS*	1.335	-1.068	-0.706	2.517	-0.007	2.248	0.769	1.000
Minimum Values								
MLE*	0.769	-1.835	-1.303	-0.663	-1.634	0.502	-2.596	-0.888
MCMC	0.766	-1.854	-1.308	-0.347	-1.589	0.562	-2.347	-0.669
MCMC*	0.766	-1.854	-1.308	-0.347	-1.589	0.734	-2.347	-0.669
TS**	0.766	-1.854	-1.308	-0.969	-1.985	0.738	-3.208	-1.000
TS*	0.767	-1.854	-1.308	-0.969	-1.963	0.838	-3.208	-1.000
Mean Absolute Error								
MLE*	0.074	0.087	0.085	0.509	0.279	0.186	0.639	0.360
MCMC	0.074	0.090	0.084	0.274	0.173	0.144	0.327	0.184
MCMC*	0.074	0.088	0.084	0.272	0.176	0.141	0.326	0.185
TS**	0.074	0.090	0.084	0.569	0.317	0.229	0.710	0.423
TS*	0.074	0.088	0.084	0.563	0.315	0.226	0.705	0.420
Mean Squared Error								
MLE*	0.009	0.012	0.011	0.333	0.106	0.056	0.527	0.170
MCMC	0.009	0.013	0.011	0.118	0.047	0.035	0.170	0.054
MCMC*	0.009	0.012	0.011	0.116	0.047	0.033	0.168	0.054
TS**	0.009	0.013	0.011	0.480	0.150	0.079	0.750	0.262
TS*	0.009	0.012	0.011	0.468	0.146	0.076	0.737	0.261
SIF for all MCMC outputs								
Mean	6	6	5	155	119	115	160	173
Standard Deviation	2	2	2	17	23	37	15	12

* 691 samples included; ** 892 samples included ; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models

Table 3.35: *ρ=0.5, Continuous and Binary Explanatory Variables (B), $x_{i1} = x_{i2}$*

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	α	ρ
True Values	1.000	-1.500	-1.000	0.800	-1.000	1.500	-1.200	0.500
Mean								
MLE*	1.008	-1.507	-1.005	0.516	-0.866	1.625	-0.846	0.294
MCMC	1.005	-1.508	-1.002	0.401	-0.817	1.707	-0.706	0.204
MCMC*	1.009	-1.511	-1.007	0.429	-0.833	1.689	-0.741	0.228
TS**	1.001	-1.504	-0.999	0.771	-1.040	1.664	-1.187	0.489
TS*	1.004	-1.504	-1.002	0.778	-1.051	1.661	-1.196	0.498
Standard Deviation								
MLE*	0.091	0.114	0.105	0.536	0.260	0.346	0.642	0.380
MCMC	0.093	0.114	0.106	0.354	0.205	0.218	0.422	0.239
MCMC*	0.092	0.115	0.105	0.360	0.202	0.226	0.426	0.240
TS**	0.092	0.114	0.104	0.619	0.337	0.269	0.765	0.456
TS*	0.091	0.115	0.104	0.641	0.343	0.281	0.791	0.469
Maximum Values								
MLE*	1.353	-1.156	-0.731	1.533	0.030	2.467	1.087	0.915
MCMC	1.348	-1.077	-0.705	1.471	0.041	2.441	1.034	0.933
MCMC*	1.348	-1.161	-0.728	1.471	-0.091	2.441	0.799	0.811
TS**	1.335	-1.068	-0.725	1.774	0.241	2.756	1.569	1.000
TS*	1.335	-1.154	-0.735	1.774	0.241	2.756	1.569	1.000
Minimum Values								
MLE*	0.758	-1.834	-1.305	-1.075	-1.468	0.731	-2.031	-0.833
MCMC	0.774	-1.861	-1.311	-0.964	-1.422	0.947	-2.077	-0.717
MCMC*	0.775	-1.861	-1.311	-0.924	-1.422	0.947	-1.991	-0.562
TS**	0.766	-1.854	-1.308	-1.559	-1.886	1.058	-2.358	-1.000
TS*	0.766	-1.854	-1.308	-1.559	-1.743	1.058	-2.358	-1.000
Mean Absolute Error								
MLE*	0.073	0.092	0.083	0.487	0.228	0.311	0.587	0.342
MCMC	0.074	0.090	0.084	0.441	0.222	0.252	0.539	0.317
MCMC*	0.074	0.092	0.084	0.423	0.210	0.243	0.514	0.297
TS**	0.074	0.090	0.083	0.500	0.278	0.236	0.620	0.369
TS*	0.073	0.092	0.083	0.515	0.285	0.235	0.640	0.379
Mean Squared Error								
MLE*	0.008	0.013	0.011	0.367	0.086	0.135	0.538	0.187
MCMC	0.009	0.013	0.011	0.285	0.075	0.090	0.422	0.145
MCMC*	0.008	0.013	0.011	0.267	0.068	0.087	0.392	0.132
TS**	0.008	0.013	0.011	0.384	0.115	0.099	0.585	0.208
TS*	0.008	0.013	0.011	0.411	0.120	0.105	0.624	0.219
SIF for all MCMC outputs								
Mean	6	6	5	158	110	133	163	175
Standard Deviation	2	2	2	16	27	30	15	12

* 644 samples included; ** 920 samples included ; 1000 samples included in MCMC

Estimator Performance in Partial Structural Models
Table 3.36: *ρ=0.8, Continuous and Binary Explanatory Variables (B), $x_{i1} = x_{i2}$*

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	α	ρ
True Values	1.000	-1.500	-1.000	0.800	-1.000	1.500	-1.200	0.800
Mean								
MLE*	1.005	-1.512	-1.001	0.509	-0.909	1.739	-0.869	0.630
MCMC	1.005	-1.513	-1.004	0.072	-0.783	2.092	-0.351	0.361
MCMC*	1.005	-1.516	-1.003	0.094	-0.787	2.077	-0.382	0.378
TS**	0.999	-1.502	-0.997	0.347	-1.008	2.175	-0.740	0.628
TS*	1.000	-1.504	-0.997	0.350	-1.002	2.172	-0.748	0.631
Standard Deviation								
MLE*	0.090	0.111	0.104	0.539	0.210	0.459	0.627	0.317
MCMC	0.092	0.113	0.105	0.449	0.207	0.342	0.523	0.267
MCMC*	0.091	0.112	0.105	0.443	0.199	0.339	0.517	0.261
TS**	0.093	0.114	0.106	0.582	0.309	0.279	0.720	0.415
TS*	0.091	0.112	0.104	0.575	0.304	0.275	0.715	0.413
Maximum Values								
MLE*	1.306	-1.182	-0.727	1.422	0.160	2.852	1.789	0.982
MCMC	1.331	-1.077	-0.703	1.284	-0.071	2.880	1.369	0.956
MCMC*	1.331	-1.168	-0.730	1.152	-0.123	2.880	1.369	0.949
TS**	1.335	-1.068	-0.706	1.258	0.204	3.223	2.188	1.000
TS*	1.335	-1.154	-0.725	1.252	0.204	3.223	2.188	1.000
Minimum Values								
MLE*	0.766	-1.858	-1.312	-1.564	-1.350	0.770	-1.862	–0.856
MCMC	0.767	-1.869	-1.311	-1.270	-1.320	1.089	-1.731	-0.544
MCMC*	0.767	-1.869	-1.311	-1.270	-1.320	1.089	-1.605	-0.544
TS**	0.766	-1.830	-1.308	-1.911	-1.660	1.523	-1.774	-1.000
TS*	0.767	-1.854	-1.308	-1.911	-1.660	1.596	-1.774	-1.000
Mean Absolute Error								
MLE*	0.073	0.088	0.084	0.444	0.169	0.405	0.504	0.238
MCMC	0.074	0.090	0.084	0.744	0.248	0.611	0.864	0.444
MCMC*	0.073	0.089	0.084	0.720	0.241	0.595	0.831	0.427
TS**	0.075	0.091	0.085	0.525	0.246	0.675	0.604	0.336
TS*	0.073	0.089	0.084	0.522	0.244	0.672	0.602	0.337
Mean Squared Error								
MLE*	0.008	0.013	0.011	0.375	0.052	0.268	0.502	0.129
MCMC	0.009	0.013	0.011	0.731	0.090	0.467	0.994	0.263
MCMC*	0.008	0.013	0.011	0.694	0.085	0.448	0.936	0.246
TS**	0.009	0.013	0.011	0.543	0.095	0.533	0.730	0.202
TS*	0.008	0.013	0.011	0.532	0.093	0.527	0.714	0.199
SIF for all MCMC outputs								
Mean	8	8	7	166	100	149	170	182
Standard Deviation	4	4	4	17	33	28	15	10

* 764 samples included; ** 984 samples included; 1000 samples included in MCMC

Chapter 4

Modeling an Ordered Variable with Sample Selection

4.1 Introduction

This chapter develops a Bayesian estimation method for a specific sample selection model with endogeneity. This sample selection model is a two-equation system in which the first equation has a binary choice form deciding the sample selection mechanism, while the categorical data outcome variable is observed in the second equation only for censored selected observations. In this model, endogeneity is represented by the correlation between the unobservables driving selection and those explaining the categorical outcome. This correlation is captured in a latent variable structure.

Even though this model is not particularly complex, classical estimation via maximum likelihood is not easy, with estimates often not converging. This motivates introduction of a Bayesian method. In broad terms, data augmentation is utilized with the latent variable structure of the model to construct posterior distributions and avoid the double integration required in FIML. After reparameterization and designing special priors, an efficient Gibbs sampler is set up that uses conjugate conditional posteriors. Then a numerical study is conducted to evaluate the performance of MCMC estimates.

The Monte Carlo experiment compares the Bayesian method with three other applicable methods including FIML, a two-step method derived from Heckman (1979)'s two-step estimation and a likelihood-based two-step method based on the form of the concentrated likelihood function. The effects of exclusion restrictions on estimator performance are explored during the comparison. It is found that our Gibbs sampler provides accurate and efficient estimates, while the likelihood-based two-step method appears a little superior to FIML.

An application to mental illness and labor market indicates that the Bayesian method can work well. In this application, we are interested in the relationship between mental illness and occupational skill levels which are categories only observed when people are employed. Potential selection bias arises when unobservable factors affect both possibilities of employment (and hence selection into the sample) and occupational skill level. The Gibbs sampler is used to estimate both the parameters of the model and marginal effects.

The chapter is organized as follows. Section 4.2 presents the latent variables structure of this sample selection model. Section 4.3 shows Bayesian analysis including how to reparameterize the model and how to set priors to get conjugate conditional posterior distributions, as well as a simulation example with convergence diagnosis. In section 4.4, three other estimation methods are described in detail and a numerical study indicates performance of the four methods in models that include with exclusion restrictions and models that do not. The estimation technique is applied to the analysis of mental illness and labor market outcomes in section 4.5. Finally, we conclude in section 4.6.

4.2 The Model

Our model is an extension of Heckman (1979)'s sample selection model with a two-equation system containing endogenous variables. The first variable y_{i1} is a 0/1 binary choice response, which determines sample selection. The second variable, however, instead of the continuous variable in the main equation of Heckman's model, is discrete and ordinal with $y_{i2} \in (1, 2, \ldots, J)$. To fully describe the features of each response, we assume y_{i1} and y_{i2} are generated from the following latent variable forms respectively:

$$\begin{cases} z_{i1} = x_{i1}'\beta_1 + \epsilon_{i1} \\ z_{i2} = x_{i2}'\beta_2 + \epsilon_{i2} \qquad 1 \le i \le n \end{cases} \tag{4.2.1}$$

where x_{i1} and x_{i2} are sets of explanatory variables while β_1 and β_2 are the final parameters we are interested in.

To allow for selection on unobservables in this model, ϵ_{i1} and ϵ_{i2} are assumed to be correlated with each other. Together with the normality assumption, the error terms are described as follows:

$$\begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right].$$

As with the bivariate probit model, unit variance of error terms is assumed in order to guarantee identification.

Selection into the sample on which the second equation is estimated is decided by the binary choice indicator y_{i1} related to the latent variable z_{i1}. At the same time, the latent variable z_{i2} is assumed to determine ordered outcomes of y_{i2} which are only observed when $y_{i1} = 1$, while x_{i2} are always observed. Otherwise, y_{i2} is missing when $y_{i1} = 0$, so this null is represented by a zero value. Our observed dependent variables are defined by:

$$\begin{cases} y_{i1} = I(z_{i1} > 0) \\ y_{i2} = j \times y_{i1} \quad \text{if } \gamma_{j-1} \le z_{i2} \le \gamma_j \quad 1 \le j \le J \end{cases} \tag{4.2.2}$$

where $I(.)$ denotes the indicator function. $\left\{\gamma_0, \gamma_1, \gamma_2, \cdots, \gamma_J\right\}$ are threshold parameters, mapping latent variables z_{i2} into categories. For identification, and without loss of generality, let $\gamma_0 = -\infty$, $\gamma_1 = 0$, $\gamma_J = +\infty$. To show that setting $\gamma_1 = 0$ imposes no effective restriction, suppose $\gamma_1 \neq 0$. Then the second equation can be written as $y_{i2} = j \times y_{i1}$ if $\gamma_{j-1} - \gamma_1 \le z_{i2} - \gamma_1 \le \gamma_j - \gamma_1$ $(1 \le j \le J)$ and $z_{i2} - \gamma_1 = x_{i2}'\beta_2 - \gamma_1 + \epsilon_{i2}$. Let the coefficient of the constant term in the vector β_2 be β_{20}. Thus, $\beta_{20} - \gamma_1$ is estimated instead of β_{20}. That is why γ_1 can be fixed at 0 without loss of generality as it simply locates the intercept parameter β_{20}. In addition, we separate the largest threshold parameter to be estimated, γ_{J-1}, from $\gamma = (\gamma_2, \cdots, \gamma_{J-2})'$ for reasons to do with how the Gibbs sampler is performed. Specifically, γ_{J-1} is not directly sampled in the upcoming Bayesian analysis, but is used to reparameterize the model.

4.3 Bayesian Analysis

4.3.1 Introduction

For Bayesian analysis, we first describe a joint posterior function involving the latent variables. This procedure is also adapted from Chib & Greenberg (1998), who estimate multivariate probit models, and builds on the approach taken in Chapter 3. Let $\theta = (\beta_1', \beta_2', \gamma', \gamma_{J-1}, \rho)'$ and $Z = \{(z_{i1}, z_{i2}) : i = 1, 2, \cdots, n\}$. Given the prior $p(\theta)$, we get the posterior distribution of the model parameters and latent variables from Bayes Theorem:

$$\begin{aligned} p(\theta, Z|Y) &\propto p(\theta)p(Z|\theta)L(Y|\theta, Z) \\ &= p(\theta)\prod_{i=1}^{n} p(z_{i1}, z_{i2}|\theta)L(y_{i1}|\theta, z_{i1}, z_{i2})L(y_{i2}|y_{i1}, \theta, z_{i1}, z_{i2}). \end{aligned} \tag{4.3.1}$$

From equations (4.2.1) and (4.2.2), it is obvious that

$$p(z_{i1}, z_{i2}|\theta) = \phi_2\left[\begin{pmatrix} z_{i1} \\ z_{i2} \end{pmatrix}; \begin{pmatrix} x_{i1}'\beta_1 \\ x_{i2}'\beta_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right],$$

where ϕ_2 denotes the bivariate normal distribution. Let $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. From equation (4.2.2), we obtain

$$L(y_{i1}|\theta, z_{i1}, z_{i2}) = I(z_{i1} > 0)I(y_{i1} = 1) + I(z_{i1} \leq 0)I(y_{i1} = 0).$$

Since $y_{i2} = j \times y_{i1}$, we get

$$\begin{aligned} L(y_{i2}|y_{i1} = 1, \theta, z_{i1}, z_{i2}) &= \Sigma_{j=1}^{J} I(y_{i2} = j)I(\gamma_{j-1} < z_{i2} < \gamma_j) \\ L(y_{i2}|y_{i1} = 0, \theta, z_{i1}, z_{i2}) &= I(y_{i2} = 0). \end{aligned}$$

Therefore, equation (4.3.1) can be fully displayed as follows:

$$p(\theta, Z|Y) \propto p(\theta)\prod_{i=1}^{n}\phi_2\left[\begin{pmatrix} z_{i1} \\ z_{i2} \end{pmatrix};\begin{pmatrix} x'_{i1}\beta_1 \\ x'_{i2}\beta_2 \end{pmatrix},\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right]$$
$$\times\Big[I(z_{i1} > 0)I(y_{i1} = 1)\Sigma_{j=1}^{J} I(y_{i2} = j)I(\gamma_{j-1} < z_{i2} < \gamma_j)$$
$$+\ I(z_{i1} \leq 0)I(y_{i1} = 0)I(y_{i2} = 0)\Big]. \tag{4.3.2}$$

4.3.2 Reparameterization

A reparameterization process is introduced before posterior computation to deal with the problem of a slow mixing Gibbs sampler. In research on the ordered probit one-equation model, Cowles (1996) finds that such slow mixing is caused by high correlation between the estimated threshold and latent data. Nandram & Chen (1996) put forward a strategy of reparameterization to solve this problem. Then, Li & Tobias (2006) apply this strategy in an ordered probit two-equation system, which greatly improves the performance of posterior simulation. In addition, Li & Tobias (2007) summarize in detail three benefits and one drawback of such reparameterization.

In order to speed up the convergence rate of our Gibbs algorithm, we also use reparameterization in our selectivity modeling. However, our strategy is somewhat different from that used by Li & Tobias (2006). In their model, reparameterization can be adopted in both equations because both are ordinal. Then the conditional posterior of parameters in the covariance matrix is specifically designed to be Inverse-Wishart Density. However, only one equation is ordinal in our model, which results in one diagonal element in the covariance matrix being fixed at one after such reparameterization. Therefore, Li & Tobias (2006)'s method cannot be used to produce a standard form for the whole covariance matrix.

To use as efficient an algorithm as possible for the Gibbs sampling, we adopt the parameterization utilized by McCulloch et al. (2000), whose covariance matrix has one as diagonal elements. As a result, we are aiming to transform the covariance matrix into the form in McCulloch et al. (2000) by letting

$$\beta_2^* = \beta_2/\gamma_{J-1},\ z_{i2}^* = z_{i2}/\gamma_{J-1},\ \epsilon_{i2}^* = \epsilon_{i2}/\gamma_{J-1},$$

$$\gamma^* = \gamma/\gamma_{J-1},\ \lambda = \rho/\gamma_{J-1},\ \psi = (1-\rho^2)/\gamma_{J-1}^2.$$

These transformations benefit posterior computation a great deal, an issue we will discuss further in the next section. Using these transformations, equations (4.2.1) and (4.2.2) are transformed into the following system:

$$\begin{cases} z_{i1} = x_{i1}'\beta_1 + \epsilon_{i1} & 1 \le i \le n \\ z_{i2}^* = x_{i2}'\beta_2^* + \epsilon_{i2}^* & 1 \le i \le n, \end{cases}$$

$$\begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2}^* \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \lambda \\ \lambda & \psi+\lambda^2 \end{pmatrix}\right],$$

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

After those transformations, the error terms $(\epsilon_{i1}, \epsilon_{i2}^*)$ have exactly the same form as in McCulloch et al. (2000). But notice that we are primarily interested in the initial model specified by equations (4.2.1) and (4.2.2). That is why we have to add a Jacobian term into the posterior function for the new parameters to complete our algorithm. After calculating posteriors for the new parameters, we then solve back for initial model parameters.

4.3.3 Prior Specification

To produce a standard Gibbs sampler, we specially design priors for our model parameters in the following forms:

$$\beta_1 \sim N_{k_1}(0, B_1) \tag{4.3.3}$$

$$\beta_2|\gamma_{J-1} \sim N_{k_2}(0, \gamma_{J-1}^2 B_2) \tag{4.3.4}$$

$$p(\gamma_2,\gamma_3,\cdots,\gamma_{J-2})|\gamma_{J-1} \propto \gamma_{J-1}^{J-3} I[0<\gamma_2<\gamma_3<\cdots<\gamma_{J-2}<\gamma_{J-1}] \tag{4.3.5}$$

$$\gamma_{J-1}^2|\rho \sim IG(-\frac{n_0}{2}+3, \frac{C_0\rho^2}{2}) \tag{4.3.6}$$

$$(1-\rho^2)|\gamma_{J-1} \sim IG(\frac{n_0}{2}, \frac{D_0\gamma_{J-1}^2}{2}),\ |\rho|<1 \tag{4.3.7}$$

where I is an indicator function and IG denotes an Inverse Gamma distribution. Priors from (4.3.3) to (4.3.7) have similar structures to those proposed by Li & Tobias (2006). Priors (4.3.6) and (4.3.7) are created to ensure that conjugate posteriors exist for λ and ψ.

From the transformation

$$z_{i2} = \frac{z_{i2}^*}{\sqrt{\psi+\lambda^2}}, \beta_2 = \frac{\beta_2^*}{\sqrt{\psi+\lambda^2}}, \gamma = \frac{\gamma^*}{\sqrt{\psi+\lambda^2}}, \gamma_{J-1} = \frac{1}{\sqrt{\psi+\lambda^2}} \text{ and } \rho = \frac{\lambda}{\sqrt{\psi+\lambda^2}},$$

we can show that the Jacobian of the transformation from $(z_2, \beta_2', \gamma', \gamma_{J-1}, \rho)$ to $(z_2^*, \beta_2^{*\prime}, \gamma^{*\prime}, \lambda, \psi)$ is $0.5(\psi+\lambda^2)^{-\frac{1}{2}(n+k_2+J+1)}$, where k_2 is the dimension of β_2.

Let $\theta^* = [\beta', \psi, \lambda, \gamma^{*\prime}]'$, $\beta = (\beta_1', \beta_2^{*\prime})'$ and

$$\Sigma^* = \begin{pmatrix} 1 & \lambda \\ \lambda & \psi+\lambda^2 \end{pmatrix}.$$

Using the change of variables and equation (4.3.2), the posterior of the new parameters can be fully displayed as follows:

$$\begin{aligned} p(\theta^*, z_1, z_2^*|Y) \propto p(\theta^*) \prod_{i=1}^n \phi_2 &\left[\begin{pmatrix} z_{i1} \\ z_{i2}^* \end{pmatrix}; \begin{pmatrix} x_{i1}'\beta_1 \\ x_{i2}'\beta_2^* \end{pmatrix}, \begin{pmatrix} 1 & \lambda \\ \lambda & \psi+\lambda^2 \end{pmatrix}\right] \\ &\times \Big[I(z_{i1}>0)I(y_{i1}=1)\Sigma_{j=1}^J I(y_{i2}=j)I(\gamma_{j-1}^*<z_{i2}^*<\gamma_j^*) \\ &+ I(z_{i1}\le 0)I(y_{i1}=0)I(y_{i2}=0)\Big] \end{aligned} \tag{4.3.8}$$

and $\phi_2 = \frac{1}{2\pi\sqrt{\psi}} exp\left\{-\frac{\psi+\lambda^2}{2\psi}\left[(z_{i1}-x_{i1}'\beta_1)^2 + \frac{(z_{i2}^*-x_{i2}'\beta_2^*)^2}{\psi+\lambda^2} - 2\psi(z_{i1}-x_{i1}'\beta_1)(z_{i2}^*-x_{i2}'\beta_2^*)\right]\right\}$ where the priors in $\pi(\theta^*)$ are assumed to be independent of each other with forms as follows:

$$\beta \sim N_k(0, B_0^{-1}) \tag{4.3.9}$$

$$p(\gamma_2^*,\gamma_3^*,\cdots,\gamma_{J-2}^*) \propto I[0<\gamma_2^*<\gamma_3^*<\cdots<\gamma_{J-2}^*<1] \tag{4.3.10}$$

$$\lambda \sim N(0, C_0^{-1}) \tag{4.3.11}$$

$$\psi \sim IG(\frac{n_0}{2}, \frac{D_0}{2}) \tag{4.3.12}$$

and $B_0^{-1} = \begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix}$. The above derivation shows that priors of original parameters in equations (4.3.3) to (4.3.7) are equivalent to the priors of transformed parameters in equations (4.3.9) to (4.3.12).

4.3.4 Conditional Posteriors

From the joint posterior (4.3.8) with special priors (4.3.9)-(4.3.12), we now infer conditional posteriors and implement the Gibbs sampler . We start with sampling the conditional posterior of z_{i1} from

$$\begin{aligned} p(z_{i1}, z_{i2}^*|\theta^*, y_{i1}, y_{i2}) \propto \phi_2 &\left[\begin{pmatrix} z_{i1} \\ z_{i2}^* \end{pmatrix}; \begin{pmatrix} x_{i1}'\beta_1 \\ x_{i2}'\beta_2^* \end{pmatrix}, \begin{pmatrix} 1 & \lambda \\ \lambda & \psi + \lambda^2 \end{pmatrix}\right] \\ &\times \Big[I(z_{i1} > 0)I(y_{i1} = 1)\Sigma_{j=1}^J I(y_{i2} = j)I(\gamma_{j-1}^* < z_{i2}^* < \gamma_j^*) \\ &+ I(z_{i1} \leq 0)I(y_{i1} = 0)I(y_{i2} = 0)\Big]. \end{aligned}$$

This gives,

$$z_{i1}|\theta^*, z_{i2}^*, y_{i1}, y_{i2} \sim \begin{cases} TN(\mu_{zi1}, \sigma_{zi1}^2)|_{(0,+\infty)}, & \text{if } y_{i1} = 1 \\ TN(\mu_{zi1}, \sigma_{zi1}^2)|_{(-\infty,0]}, & \text{if } y_{i1} = 0 \end{cases} \tag{4.3.13}$$

where TN denotes a univariate truncated normal distribution with $\mu_{zi1} = x_{i1}'\beta_1 + \lambda(z_{i2}^* - x_{i2}'\beta_2^*)$ and $\sigma_{zi1}^2 = \psi/(\psi + \lambda^2)$. When $y_{i1} = 1$, it is truncated into $(0, +\infty)$. Otherwise, it is truncated into $(-\infty, 0]$.

In a similar way to posterior inference on z_{i1}, we can get a result for z_{i2}^* when $y_{i1} = 1$ and $y_{i2} = j$, which is also a univariate normal distribution truncated to the region (γ_{j-1}, γ_j). When $y_{i1} = 0$ and $y_{i2} = 0$, however, there is no constraint for z_{i2}^*, so it can be drawn directly from a normal distribution. As a result, z_{i2}^* has the following conditional posterior distribution:

$$z_{i2}^*|\theta^*, z_{i1}, y_{i1}, y_{i2} \sim \begin{cases} TN(\mu_{zi2}, \sigma_{zi2}^2)|_{(\gamma_{j-1}, \gamma_j)}, & \text{if } y_{i1} = 1 \text{ and } y_{i2} = j \\ N(\mu_{zi2}, \sigma_{zi2}^2), & \text{otherwise} \end{cases} \tag{4.3.14}$$

where $\mu_{zi2} = x_{i2}'\beta_2^* + \lambda(z_{i1} - x_{i1}'\beta_1)$ and $\sigma_{zi2}^2 = \psi$.

To sample β, we use the prior $p(\beta) = \phi_k(\beta|\beta_0, B_0^{-1})$, and let $X_i = (x_{i1}' \ 0, \ 0 \ x_{i2}')$, $Z_i = (z_{i1}, z_{i2}^*)'$ and $Z^* = \left\{(z_{i1}, z_{i2}^*)' : i = 1, \cdots, n\right\}$. We get the conditional posterior function of β which is a Gaussian density,

$$\beta|Z^*, \Sigma^* \sim N_k(\hat{\beta}, B^{-1}), \tag{4.3.15}$$

where $\hat{\beta} = B^{-1}(B_0\beta_0 + \Sigma_{i=1}^n X_i'\Sigma^{*-1}Z_i)$ and $B = B_0 + \Sigma_{i=1}^n X_i'\Sigma^{*-1}X_i$. Thus we can simulate β in a straight forward manner.

Given prior $p(\lambda) \sim N(\lambda_0, C_0^{-1})$, we then sample λ from $p(\lambda|Z^*, \beta, \psi) \propto p(\lambda)p(Z^*|\beta, \Sigma^*)$, where

$$p(Z^*|\beta, \Sigma^*) = |\Sigma^*|^{-n/2} exp\left\{-\frac{1}{2}\Sigma_{i=1}^n (Z_i - X_i\beta)'\Sigma^{*-1}(Z_i - X_i\beta)\right\}.$$

This conditional posterior function of λ is a Gaussian density

$$\lambda|Z^*, \beta, \psi \sim N(\hat{\lambda}, C^{-1}) \tag{4.3.16}$$

where $\hat{\lambda} = C^{-1}\left[\frac{1}{\psi}\sum_{i=1}^n (z_{i1} - x_{i1}'\beta_1)(z_{i2}^* - x_{i2}'\beta_2^*)\right]$ and $C = C_0 + \frac{1}{\psi}\sum_{i=1}^n (z_{i1} - x_{i1}'\beta_1)^2$.

After that, given prior $p(\psi) \sim IG(\frac{n_0}{2}, \frac{D_0}{2})$, ψ is drawn from a conjugate conditional posterior:

$$\begin{aligned} &p(\psi|Z^*, \beta, \lambda) \propto p(\psi)p(Z^*|\beta, \Sigma^*) \\ &\propto (\frac{1}{\psi})^{\frac{n_0}{2}+1} exp(-\frac{D_0}{2\psi})(\frac{1}{\psi})^{\frac{n}{2}} exp\left\{-\frac{1}{2\psi}\Sigma_{i=1}^n \left[\lambda(z_{i1} - x_{i1}'\beta_1) - (z_{i2}^* - x_{i2}'\beta_2^*)\right]^2\right\}. \end{aligned}$$

This conditional posterior function of ψ is an Inverse-Gamma density

$$\psi|Z^*, \beta, \lambda \sim IG(\frac{n_1}{2}, \frac{D}{2}) \tag{4.3.17}$$

where $n_1 = n_0 + n$ and $D = D_0 + \Sigma_{i=1}^n \left[\lambda(z_{i1} - x_{i1}'\beta_1) - (z_{i2}^* - x_{i2}'\beta_2^*)\right]^2$.

Finally, we sample the threshold parameters $\left\{\gamma_j^*\right\}_{j=2}^{J-2}$ from its conditional posterior distribution marginalized over all z_{i2}^*:

$$p(\left\{\gamma_j^*\right\}_{j=2}^{J-2}|\beta,\Sigma^*,Z) \propto \prod_{i=1}^{n}\left\{\Phi\left[(\gamma_{y_{i2}}^*-\mu_\gamma)/\sqrt{\psi}\right]-\Phi\left[(\gamma_{y_{i2}-1}^*-\mu_\gamma)/\sqrt{\psi}\right]\right\} \tag{4.3.18}$$

where $\mu_\gamma = x_{i2}'\beta_2^* + \lambda(z_{i1} - x_{i1}'\beta_1)$. In order to draw this particular conditional distribution, the Metropolis-Hasting algorithm is utilized here with a Dirichlet proposal density. This technique is described in detail by Nandram & Chen (1996) and Li & Tobias (2006).

4.3.5 A Simulation Example with MCMC Convergence Diagnostics

We illustrate our proposed Gibbs sampler using a simple simulation study that allows us to focus on convergence diagnosis. As shown in the previous chapter, the SIF is a very useful index with which to examine the convergence rate of algorithms. However, Tsay (2005) mentions that none of the available methods can guarantee 100% that the Gibbs sampler under study has converged for all applications. So SIF may also fail to indicate the true convergence in certain circumstances. That is why we will also use visual inspection by sample path and autocorrelation function (ACF) values, to check whether the MCMC algorithm has converged in this particular case.

The sample size is set to 1,000 and true values are set as $\beta_1 = (\beta_{11},\beta_{12})' = (0.8,-1)'$, $\beta_2 = (\beta_{21},\beta_{22})' = (1.5,2)'$, $\gamma_2 = 2$ and $\rho = 0.5$. Both x_{i1} and x_{i2} are 2×1 vectors with their first rows fixed at one to make constant terms for each equation. The two other rows in x_{i1} and x_{i2} are independently generated from standard normal distributions. ϵ_{i1} and ϵ_{i2} are then generated from bivariate normal density with zero mean and variance-covariance matrix (1 0.5 , 0.5 1). Latent variables are calculated based on equation (4.2.1) and about half the data is censored after selection. With two threshold values 0 and 2, the uncensored part is cut into three categories as follows:

$$y_{i2} = \begin{cases} 1 \text{ if } z_{i2} < 0 \\ 2 \text{ if } 0 < z_{i2} < 2 \\ 3 \text{ if } 2 < z_{i2}. \end{cases}$$

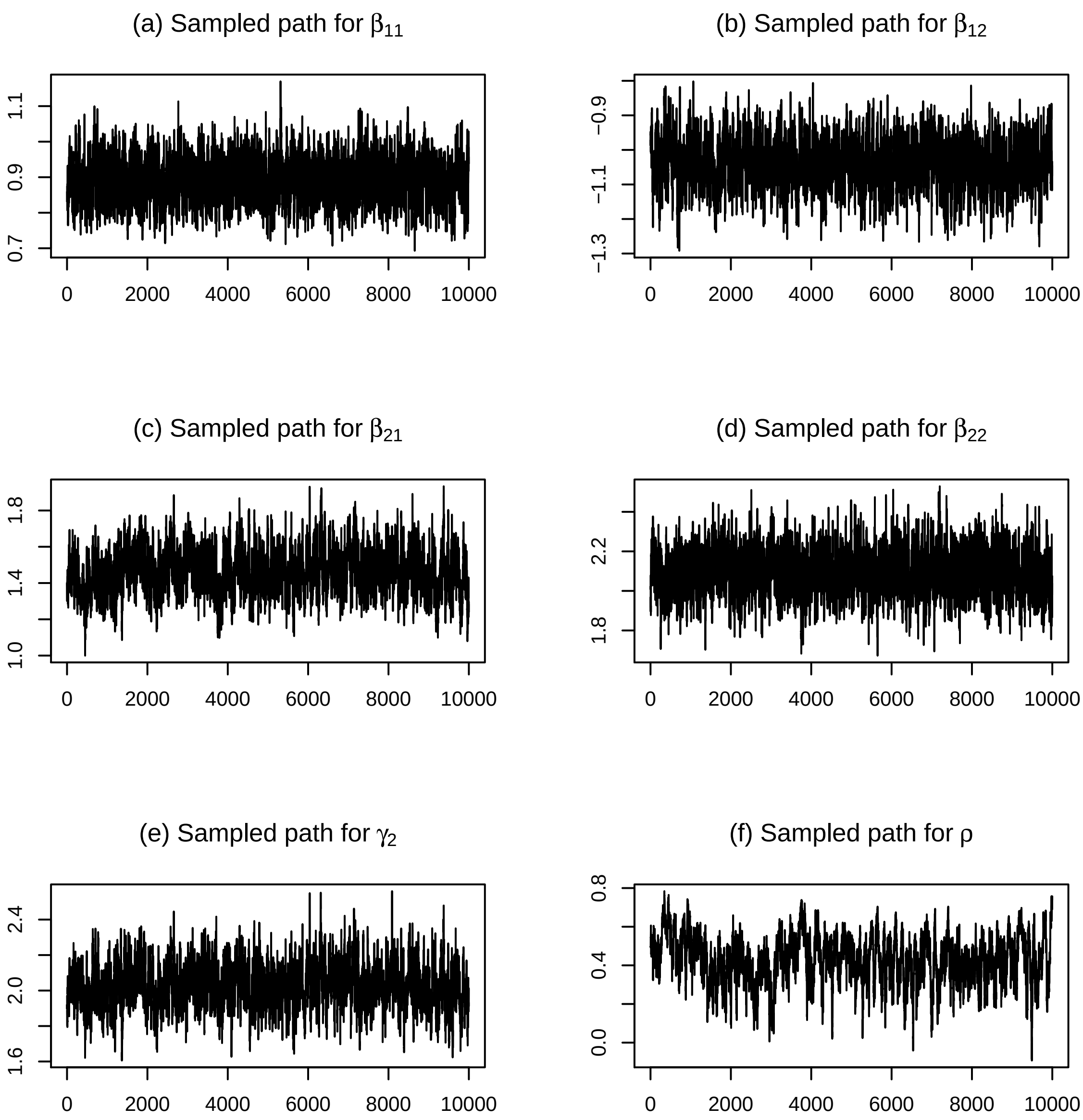
After we generate a series of explanatory variables and discrete outcomes y_{i1} and y_{i2} with true parameters, scalars for priors are set in advance in order to conduct the Gibbs sampler for sampling each parameter for this specific dataset. We set $B_0^{-1} = 1000I_k$, $C_0^{-1} = 1000$, $n_0 = 2$ and $D_0 = 0.01$ in our MCMC simulation procedure to obtain draws from the posterior. These priors are flat enough not to dominate the posterior, and also strong enough not to cause a slow convergence problem. After that, we follow the conditional posterior sampling process to estimate the parameters. In this MCMC simulation, 3,000 initial draws are discarded as the burn-in period, and the next 10,000 iterations are recorded.

Because results of the MCMC simulation in obtaining draws from the posterior can only be accepted when the algorithm is convergent, we will focus on convergence diagnosis of this MCMC simulation result and we do this by analyzing the convergence properties through examining sample paths and autocorrelation functions. 10,000 iterations of each parameter are plotted in Figure 4.1, while their autocorrelation functions are shown in Figure 4.2. From Figure 4.1, all sample paths seem to have no obvious patterns, although that for ρ tends to contain small cyclical fluctuations. As a result, all the paths are believed to be stationary in an acceptable range. Figure 4.2 indicates that each Markov Chain mixes quite well, as autocorrelations of all parameters decay fast in 80 lags although small bumps occasionally appear after that. Estimation of the threshold parameter is mixing especially well with other parameters which proves reparameterization greatly reduces the correlation between threshold parameter and latent variables. However, it is not surprising to find that the autocorrelation status of the sample path for ρ is much worse than that for β_1, β_2 and γ_2, as estimation of ρ is usually the most difficult in sample selection models and in simultaneous equation models. One possible reason is that the effect of correlation between error terms in two equations can be partially eliminated by the explanatory variables generation process. Therefore, some of the effect of ρ is included in that of β_1 and β_2 during estimation, making ρ more difficult to estimate.

Summary statistics on the 10,000 iterations are shown in Table 4.1, including SIF values. Means of all conditional posterior densities provide quite accurate point estimates of the true values while standard deviations are quite small, indicating accurate estimation.

Figure 4.1: *Sample Paths*

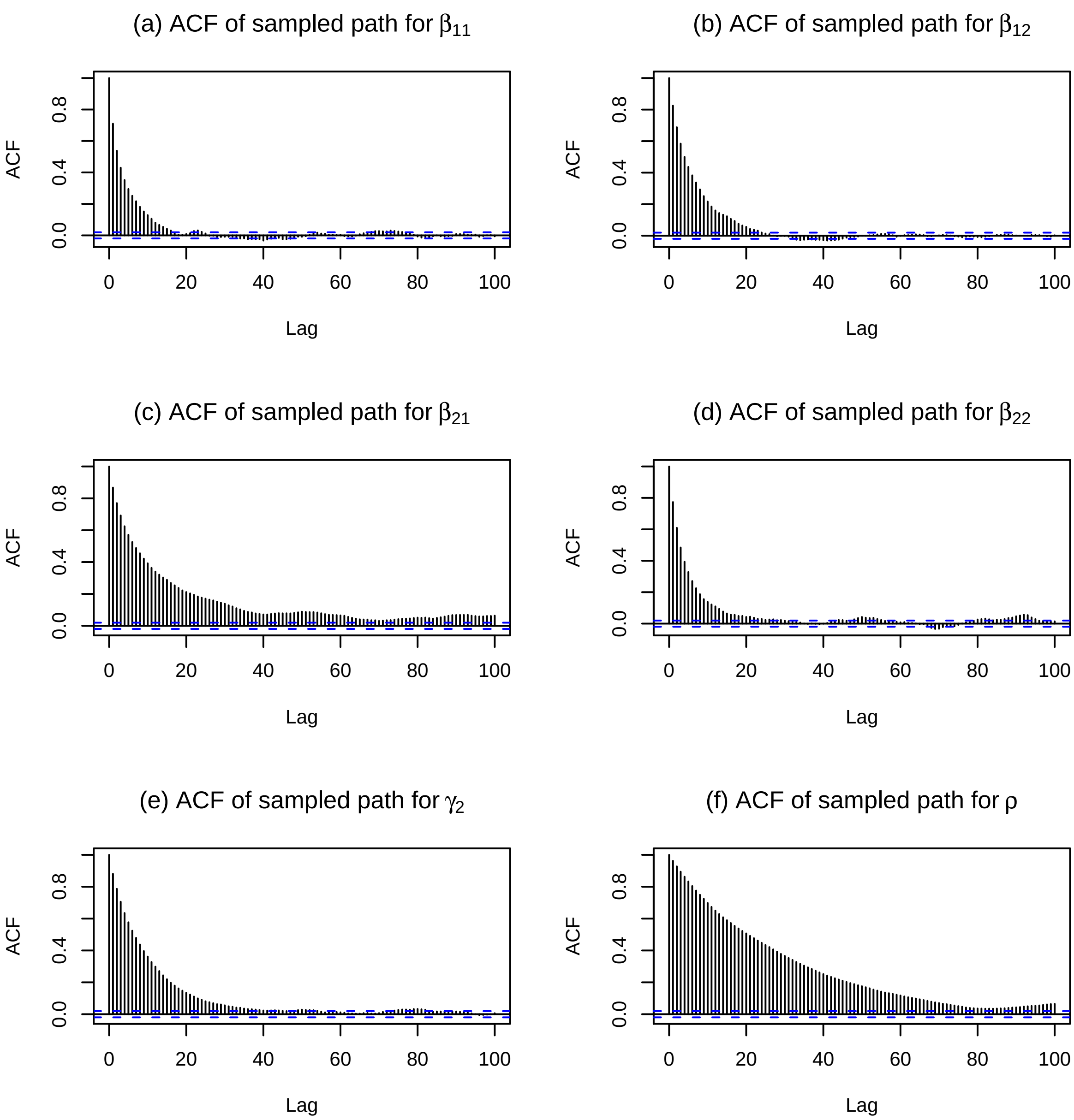
(The x-axis represents iterations;
The y-axis represents the recorded value of a parameter.)



Notice that standard deviations of β_{11} and β_{12} are smaller than that of other parameters, which indicates that β_{11} and β_{12} vary in a relatively smaller range. Figure 4.2 also provides evidence of this conclusion, because β_{11} and β_{12} have relatively faster convergence rates than all other parameters since their autocorrelation functions decay more rapidly.

Figure 4.2: *Sample Autocorrelation Functions*

(The x-axis represents lags;
The y-axis represents the autocorrelation coefficient.)



A Bayesian credible interval can be used for interval estimation, a great benefit of the Bayesian framework. All true parameter values are contained in the respective 95% credible intervals, which once again supports the view that our estimation is accurate. The SIF values in Table 4.1 are consistent with the convergence rates revealed in Figure 4.2. β_{11} and β_{12} have the smallest two SIFs while the largest SIF value 44.7 for ρ confirms that convergence of ρ is much slower than other parameters.

Table 4.1: *Bayesian Estimation Results (One sample only)*

	True	Posterior Density			SIF
		Mean	St. Dev.	95% Credible Interval	
β_{11}	0.8	0.892	0.058	(0.781, 1.009)	8.9
β_{12}	-1	-1.040	0.070	(-1.183, -0.907)	13.4
β_{21}	1.5	1.460	0.119	(1.234, 1.703)	27.5
β_{22}	2	2.090	0.111	(1.876, 2.311)	11.2
γ_2	2	2.022	0.126	(1.779, 2.276)	23.4
ρ	0.5	0.429	0.127	(0.160, 0.662)	44.7

4.4 Monte Carlo Experiments Comparing Estimation Methods

Even though we have a sample selection problem, our model is quite like a simultaneous-equations model with reduced form specification, where the relevant equations are for binary and ordered categorical endogenous variables. As discussed in earlier chapters, the structural form model is automatically identified even when $x_{i1} = x_{i2}$ due to the non-linearity of the models. Sample selection effects can be represented in the same ways as endogenous treatment effects because they have quite similar properties. As we have discussed in Chapter 3, such models cannot be identified when both x_{i1} and x_{i2} are constants, but has no serious identification problem when there are more explanatory variables with sufficient variation. Having said that, while exclusion restrictions are not essential to identify the model, they can be helpful in estimating parameters. To examine these issues further, the Monte Carlo simulation is divided into two categories, cases with and without exclusion restrictions. Three other estimation methods are provided in this section to compare with the Bayesian method.

4.4.1 Full Information Maximum Likelihood Estimation

One natural way to estimate the model is full information maximum likelihood (FIML) estimation. Vella (1998) mentions that maximum likelihood estimation is a benchmark against which to examine the efficiency loss of alternative estimators. Since it is often the case that the selection decision has an effect on y_{i2} through correlation of unobserved ϵ_{i2}

with ϵ_{i1}, the full likelihood function is obtained from the joint density of both censored and uncensored data.

First when y_{i1} equals one, the observed y_{i2} is discrete data from 1 to J. For $j \in (1, 2, \cdots, J)$, we seek $Pr(y_{i1} = 1, y_{i2} = j|x_{i1}, x_{i2})$. Under the bivariate normal assumption for the errors, it follows:

$$\begin{cases} Pr(y_{i1} = 1, y_{i2} = 1|x_{i1}, x_{i2}) & = & \Phi_2(x'_{i1}\beta_1, -x'_{i2}\beta_2, -\rho) \\ Pr(y_{i1} = 1, y_{i2} = j|x_{i1}, x_{i2}) & = & \Phi_2(x'_{i1}\beta_1, \gamma_j - x'_{i2}\beta_2, -\rho) - \\ & & \Phi_2(x'_{i1}\beta_1, \gamma_{j-1} - x'_{i2}\beta_2, -\rho) \\ & & (j = 2, \cdots, J-1) \\ Pr(y_{i1} = 1, y_{i2} = J|x_{i1}, x_{i2}) & = & \Phi_2(x'_{i1}\beta_1, x'_{i2}\beta_2 - \gamma_{J-1}, \rho) \end{cases} \tag{4.4.1}$$

where $\Phi_2(a, b; \rho)$ is the cumulative distribution function of the standardized bivariate normal density with correlation coefficient ρ. When y_{i1} equals zero, $Pr(y_{i1}, y_{i2}|x_{i1}, x_{i2})$ is just $Pr(y_{i1} = 0|x_{i1})$ because y_{i2} cannot be observed. Therefore,

$$Pr(y_{i1} = 0|x_{i1}) = \Phi(-x'_{i1}\beta_1),$$

where Φ is the cumulative distribution function of univariate standard normal distribution.

Thus, it is simple to estimate the model parameters by maximizing the following likelihood function:

$$L = \prod_{i=1}^{n} \left\{ \left[\sum_{j=1}^{J} I(y_{i2} = j) Pr(y_{i1} = 1, y_{i2} = j|x_{i1}, x_{i2}) \right]^{y_{i1}} [Pr(y_{i1} = 0|x_{i1})]^{(1-y_{i1})} \right\} \tag{4.4.2}$$

In this chapter, the CML package in GAUSS is used to obtain FIML estimates.

4.4.2 Two-step Method

A common alternative to FIML is some kind of two-step procedure. The traditional sample selection model discussed by Heckman (1976) arose from interest in modeling female wage income, taking account of the selection decision via a labor supply equation. His model has the following form,

$$\begin{cases} y_{i1} = I(x'_{i1}\beta_1 + \epsilon_{i1} > 0) \\ y_{i2} = y_{i1} \times (x'_{i2}\beta_2 + \epsilon_{i2}) \qquad 1 \leq i \leq n. \end{cases}$$

The first equation models whether a woman is employed or not, while the second equation models market wages for working women.

Heckman (1979) provides a consistent two-step method to estimate his model. This method works from

$$\begin{aligned} E(y_{i2}|x_{i2}, y_{i1} = 1) &= x'_{i2}\beta_2 + E(\epsilon_{i2}|y_{i1} = 1) \\ &= x'_{i2}\beta_2 + E(\epsilon_{i2}|\epsilon_{i1} > -x'_{i1}\beta_1) \\ &= x'_{i2}\beta_2 + \rho \times h(x'_{i1}\beta_1), \end{aligned}$$

in which ρ is still the correlation between the error terms while ϵ_{i1} can have a non-unit variance. The inverse Mills ratio $h(x'_{i1}\beta_1)) = \phi(x'_{i1}\beta_1)/\Phi(x'_{i1}\beta_1)$ enters the equation for y_{i2} as an additional regressor, so the equation can be estimated by ordinary least squares, with an estimated h based on a first step estimates.

Since our selection equation is probit like Heckman's model, probit maximum likelihood estimation can be used for consistent estimation of the first step. Heckman's two-step method obtains estimation for the second step from the OLS regression on the selected sample with the addition of the inverse Mills ratio. However, our second equation is an ordered probit equation, not the continuous equation specified in Heckman's original model. One possible way to reduce computational burden is using single integration in the second step under certain rules. The basic idea is to create a variable which reflects the effect of the selection equation and plug it into the second equation.

In the first step, MLE is applied, so the parameter vector β_1 is estimated by maximizing the likelihood function L_1 which can be expressed as

$$L_1 = \prod_{i=1}^{n} \Big[\Phi(x'_{i1}\beta_1)^{y_{i1}} \times \Phi(-x'_{i1}\beta_1)^{(1-y_{i1})}\Big]. \tag{4.4.3}$$

In the second step, we are trying to treat the second equation of latent variables as a continuous equation as in Heckman's model and introduce the inverse Mills ratio $h(x'_{i1}\beta_1)$ in the second step. But notice that an expression for $E(\epsilon_{i2}|\epsilon_{i1} > -x'_{i1}\beta_1)$ cannot be found, as our equation is an non-linear equation. As a result, a two-step procedure will be inconsistent. Despite this, the two-step procedure will be simple to implement, so we include it in the analysis for comparison purposes. To describe the estimation procedure, we write the second function of latent variables as

$$z_{i2} = x'_{i2}\beta_2 + \rho^* h(x'_{i1}\beta_1),$$

where ρ^* is one parameter evaluating the effects of selectivity but not quite equivalent to the correlation coefficient ρ. The selected sample still follows the ordered probit structure,

$$y_{i2} = j \quad \text{if } \gamma_{j-1} \le z_{i2} \le \gamma_j \quad 1 \le j \le J.$$

After substituting the consistent estimator of β_1 from the first step into the inverse Mills ratio, we simply compute each response probability:

$$\begin{cases} Pr(y_{i2}=1) &= \Phi\left[-x'_{i2}\beta_2 - \rho^* h(x'_{i1}\beta_1))\right] \\ Pr(y_{i2}=j) &= \Phi\left[\gamma_j - x'_{i2}\beta_2 - \rho^* h(x'_{i1}\beta_1))\right] - \Phi\left[\gamma_{j-1} - x'_{i2}\beta_2 - \rho^* h(x'_{i1}\beta_1))\right] \\ & \quad (j = 2, \cdots, J-1) \\ Pr(y_{i2}=J) &= \Phi\left[x'_{i2}\beta_2 + \rho^* h(x'_{i1}\beta_1)) - \gamma_{J-1}\right] \end{cases}$$

Thus, the model parameters β_2, γ and ρ^* are estimated by maximizing the following likelihood function:

$$L_2 = \prod_{i=1}^{n}\left[\sum_{j=1}^{J} I(y_{i2}=j)Pr(y_{i2}=j)\right] \tag{4.4.4}$$

The problem with this two-step method is that the ML estimation of y_2 on x_2 including inverse Mills ratio generally leads to inconsistent estimation of β_2. This is because the inverse Mills ratio is not the proper form to evaluate the effect of ρ, so ρ^* is estimated instead. However, it may still provide quite accurate estimation of the parameters of interest. Its accuracy will be explored in the Monte Carlo experiments.

4.4.3 Likelihood-based Two-step Method

The likelihood-based two-step method is supported by the idea of a concentrated likelihood function. The parameter vector θ can be partitioned into two parts as $\theta = (\beta_1', \alpha')'$ in which $\alpha = (\beta_2', \gamma', \rho)'$, so the likelihood function in Section 4.4.1 can be referred as $L = L(\beta_1, \alpha)$. Define

$$L^* = L(\hat{\beta_1}(\alpha), \alpha),$$

where $\hat{\beta_1}(\alpha)$ is a root of

$$\left.\frac{\partial lnL}{\partial \beta_1}\right|_{\hat{\beta_1}} = 0,$$

and define $\hat{\alpha}$ as a root of

$$\left.\frac{\partial lnL^*}{\partial \alpha}\right|_{\hat{\alpha}} = 0.$$

$L^*(\alpha)$ is called the concentrated likelihood function of α and it is sometimes easier to maximize L in two steps than to maximize L simultaneously. This two-step likelihood estimator is proved to be consistent by Amemiya (1985) under certain conditions.

However, the analytic form of $\partial lnL/\partial \beta_1$ will definitely include α, so we cannot obtain the root $\hat{\beta_1}(\alpha)$ directly from this function. Thus we replace $\hat{\beta_1}(\alpha)$ with an approximation, $\hat{\beta_1}$ which is estimated by maximizing the likelihood

$$L_1 = \prod_{i=1}^{n} \left[\Phi(x_{i1}'\beta_1)^{y_{i1}} \times \Phi(-x_{i1}'\beta_1)^{(1-y_{i1})} \right].$$

In other words, the first step here is identical to the first step in Section 4.4.2. Secondly, we maximize L^* with respect to α after insert $\hat{\beta_1}$ into L. This likelihood-based two-step procedure may be a little less computationally complex than FIML.

4.4.4 General Monte Carlo Design

Monte Carlo experiments are performed to compare the Bayesian method with three other estimators of the sample selection model with ordered outcomes. 1,000 sets of data are generated and parameters are estimated with the four estimation methods.

Each sample of data is generated from the original model in a similar way to that outlined in Section 4.3.5. In order to simplify the Monte Carlo process, two explanatory variables are included in each equation, plus a constant. The 3×1 vectors, x_{i1} and x_{i2} contain one as the first element, a continuous variable and a binary variable. The continuous random variable is sampled from the standard normal distribution, while the last element is a binary variable generated from Bernoulli distributions with success probability 0.7. True values of parameters are set to $\beta_1 = (\beta_{11}, \beta_{12}, \beta_{13})' = (1, -1.5, -1)'$, $\beta_2 = (\beta_{21}, \beta_{22}, \beta_{23})' = (-0.8, -1, 1.5)'$ and $\gamma_2 = 1.2$. The sample size is 1,000. Error terms ϵ_{i1} and ϵ_{i2} are then generated from a standard bivariate normal density with correlation ρ. ρ takes three values, 0, 0.5 and 0.8. After latent variables are calculated from equation (4.2.1), y_{i1} and y_{i2} are determined according to equation (4.2.2).

All estimation methods are applied in each case using starting values of zero for parameters, except for γ_2. Because γ_2 is assumed to be larger than zero, its starting value is set at one. The MCMC estimator is obtained as the mean of the posterior density based on 10,000 draws after discarding 1,000 initial draws. The Monte Carlo process is repeated 1,000 times, giving 1,000 estimates for each estimation method.

Maximum likelihood optimization methods may fail to converge while the Bayesian method gives estimates in all 1,000 cases. To eliminate the possible effects of different samples on relative estimator performance, statistics for the four methods are summarized with respect to the samples which FIML can successfully estimate. In each result table, FIML* represents the statistics of available estimates obtained by FIML after removing those where the Hessian matrix fails to invert. Meanwhile, MCMC*, TS* and LBTS* are denoted as the three other estimators with exactly the same set of samples as FIML*. The Hessian matrix may also fail to invert when applying the two-step method and likelihood-based two-step method. TS** denotes the results for the two-step estimator in the cases where it converges. Meanwhile, LBTS*** reveals the results when the likelihood-based two-step method can get reliable estimates. MCMC in the tables represents summary statistics from the Bayesian method with all 1,000 samples.

4.4.5 Results when $x_{i1} \neq x_{i2}$

Table 4.2, 4.3 and 4.4 present the results for all four estimation methods as $\rho = 0$, 0.5 and 0.8 respectively. No identification problem arises in those tables as $x_{i1} \neq x_{i2}$. It is quite similar to the structural equations we have discussed in Chapter 3, which can be identified even if the system was linear.

We start with analyzing the number of times each method is able to provide valid estimates, considering convergence of the three optimization methods and the SIF values for the MCMC method. Firstly, FIML can estimate 804, 873 and 899 samples in the three tables respectively. In other words, 10%-20% of samples cannot be estimated due to numerical issues such as the covariance matrix failing to invert, although this situation may be improved by changing starting values. Secondly, the two-step method only fails to converge 27 times when $\rho = 0$ and can estimate all 1000 samples when $\rho = 0.5$ or 0.8. Thirdly, the likelihood-based two-step method tends to converge more frequently than FIML but less often than the two-step method, as it can estimate 866, 915 and 940 samples in these tables. The convergence rate of the MCMC algorithm is quite good especially when error correlation is weak, but it becomes more and more difficult to estimate ρ when the actual ρ value is large. The mean SIF value of ρ is only 37 when the true ρ value is 0, but jumps to 70 when true ρ value is 0.8. The mean SIF values of other parameters are all less than 50. The convergence rate of the MCMC algorithm is generally fast when $x_{i1} \neq x_{i2}$ in comparison with the cases we discuss in the next section.

Next we compare across methods, referring first to Table 4.2, where $\rho = 0$. Statistics for different samples reveal only small differences, so we will concentrate on the ones derived from 804 samples. All four methods can give unbiased estimates with small variance and estimation errors. The Bayesian method is a little less biased in estimating β_{22}, β_{23} and ρ, but a little more biased in estimating other parameters than the other methods. The four methods have almost the same efficiency, as standard deviations, mean of absolute errors and mean of squared errors are close across methods for each parameter. Outlier

Table 4.2: *Estimator Performance when ρ=0 and $x_{i1} \neq x_{i2}$*

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ	ρ
True Values	1.000	-1.500	-1.000	-0.800	-1.000	1.500	1.200	0.000
Mean								
FIML*	1.007	-1.507	-1.006	-0.809	-1.008	1.509	1.200	-0.002
MCMC	1.012	-1.514	-1.010	-0.815	-0.999	1.497	1.167	0.000
MCMC*	1.013	-1.512	-1.010	-0.817	-1.000	1.498	1.165	0.000
TS**	1.005	-1.508	-1.005	-0.809	-1.010	1.512	1.205	-0.002
TS*	1.006	-1.507	-1.005	-0.811	-1.011	1.513	1.204	-0.002
LBTS***	1.008	-1.507	-1.010	-0.806	-1.006	1.506	1.201	-0.001
LBTS*	1.006	-1.507	-1.005	-0.809	-1.008	1.509	1.200	-0.001
Standard Deviation								
FIML*	0.111	0.087	0.128	0.125	0.067	0.137	0.081	0.134
MCMC	0.111	0.089	0.128	0.126	0.067	0.137	0.083	0.126
MCMC*	0.111	0.088	0.128	0.124	0.067	0.136	0.081	0.128
TS**	0.110	0.088	0.127	0.126	0.069	0.138	0.083	0.133
TS*	0.110	0.087	0.127	0.125	0.068	0.137	0.081	0.134
LBTS***	0.108	0.090	0.125	0.127	0.068	0.139	0.082	0.132
LBTS*	0.110	0.087	0.127	0.125	0.067	0.137	0.081	0.134
Maximum Values								
FIML*	1.370	-1.251	-0.582	-0.405	-0.786	1.935	1.452	0.508
MCMC	1.390	-1.257	-0.582	-0.403	-0.778	1.923	1.415	0.478
MCMC*	1.377	-1.257	-0.582	-0.415	-0.778	1.923	1.415	0.478
TS**	1.379	-1.250	-0.590	-0.403	-0.786	1.937	1.452	0.406
TS*	1.374	-1.250	-0.590	-0.403	-0.786	1.937	1.452	0.406
LBTS***	1.379	-1.250	-0.654	-0.405	-0.841	1.936	1.435	0.502
LBTS*	1.374	-1.250	-0.590	-0.405	-0.786	1.936	1.452	0.502
Minimum Values								
FIML*	0.631	-1.900	-1.420	-1.283	-1.223	1.123	0.946	-0.473
MCMC	0.632	-1.909	-1.450	-1.292	-1.227	1.108	0.912	-0.447
MCMC*	0.632	-1.909	-1.433	-1.292	-1.212	1.113	0.912	-0.447
TS**	0.637	-1.903	-1.440	-1.292	-1.240	1.122	0.948	-0.444
TS*	0.637	-1.903	-1.415	-1.292	-1.233	1.129	0.948	-0.444
LBTS***	0.690	-1.903	-1.440	-1.282	-1.238	1.118	0.969	-0.472
LBTS*	0.637	-1.903	-1.415	-1.282	-1.223	1.123	0.946	-0.472
Mean Absolute Error								
FIML*	0.087	0.068	0.099	0.100	0.054	0.108	0.065	0.111
MCMC	0.087	0.071	0.099	0.101	0.054	0.108	0.072	0.101
MCMC*	0.088	0.069	0.100	0.100	0.054	0.107	0.071	0.106
TS**	0.087	0.069	0.099	0.101	0.055	0.110	0.067	0.108
TS*	0.087	0.068	0.099	0.100	0.054	0.109	0.065	0.111
LBTS***	0.086	0.071	0.097	0.101	0.054	0.110	0.066	0.108
LBTS*	0.087	0.068	0.099	0.100	0.054	0.108	0.065	0.110
Mean Squared Error								
FIML*	0.012	0.008	0.016	0.016	0.005	0.019	0.007	0.018
MCMC	0.012	0.008	0.016	0.016	0.005	0.019	0.008	0.016
MCMC*	0.013	0.008	0.017	0.016	0.004	0.018	0.008	0.016
TS**	0.012	0.008	0.016	0.016	0.005	0.019	0.007	0.018
TS*	0.012	0.008	0.016	0.016	0.005	0.019	0.007	0.018
LBTS***	0.012	0.008	0.016	0.016	0.005	0.019	0.007	0.017
LBTS*	0.012	0.008	0.016	0.016	0.005	0.019	0.007	0.018
SIF for all MCMC outputs								
Mean	7	15	8	16	6	6	25	37
Standard Deviation	2	4	2	4	1	1	5	7

* 804 samples included; ** 973 samples included; *** 866 samples included; 1000 samples included in MCMC

Table 4.3: *Estimator Performance when ρ=0.5 and $x_{i1} \neq x_{i2}$*

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ	ρ
True Values	1.000	-1.500	-1.000	-0.800	-1.000	1.500	1.200	0.500
Mean								
FIML*	1.008	-1.509	-1.007	-0.807	-1.009	1.511	1.208	0.503
MCMC	1.013	-1.515	-1.011	-0.810	-1.005	1.505	1.180	0.480
MCMC*	1.015	-1.517	-1.013	-0.810	-1.005	1.505	1.180	0.483
TS**	1.006	-1.508	-1.006	-0.845	-1.058	1.584	1.266	0.527
TS*	1.008	-1.510	-1.008	-0.846	-1.059	1.585	1.267	0.531
LBTS***	1.005	-1.509	-1.005	-0.806	-1.009	1.511	1.208	0.499
LBTS*	1.008	-1.510	-1.008	-0.806	-1.010	1.511	1.208	0.502
Standard Deviation								
FIML*	0.108	0.087	0.125	0.114	0.070	0.137	0.086	0.110
MCMC	0.109	0.089	0.126	0.113	0.069	0.135	0.084	0.106
MCMC*	0.109	0.088	0.126	0.114	0.069	0.136	0.085	0.106
TS**	0.110	0.088	0.127	0.127	0.072	0.143	0.088	0.136
TS*	0.110	0.087	0.127	0.129	0.072	0.143	0.089	0.137
LBTS***	0.110	0.089	0.127	0.113	0.069	0.137	0.084	0.109
LBTS*	0.110	0.087	0.127	0.114	0.070	0.137	0.086	0.109
Maximum Values								
FIML*	1.382	-1.251	-0.566	-0.469	-0.792	1.946	1.518	0.783
MCMC	1.394	-1.258	-0.570	-0.485	-0.786	1.933	1.480	0.770
MCMC*	1.394	-1.258	-0.570	-0.485	-0.786	1.933	1.480	0.760
TS**	1.379	-1.250	-0.590	-0.499	-0.842	2.092	1.554	0.931
TS*	1.379	-1.250	-0.590	-0.499	-0.842	2.092	1.554	0.931
LBTS***	1.379	-1.250	-0.590	-0.486	-0.819	1.946	1.518	0.797
LBTS*	1.379	-1.250	-0.590	-0.467	-0.790	1.946	1.518	0.783
Minimum Values								
FIML*	0.621	-1.910	-1.439	-1.214	-1.212	1.124	0.966	0.181
MCMC	0.624	-1.919	-1.449	-1.220	-1.204	1.127	0.938	0.164
MCMC*	0.624	-1.919	-1.449	-1.220	-1.204	1.127	0.938	0.164
TS**	0.637	-1.903	-1.440	-1.304	-1.279	1.212	1.023	0.173
TS*	0.637	-1.903	-1.440	-1.304	-1.279	1.212	1.023	0.173
LBTS***	0.637	-1.903	-1.440	-1.214	-1.212	1.124	0.966	0.181
LBTS*	0.637	-1.903	-1.440	-1.214	-1.212	1.124	0.966	0.181
Mean Absolute Error								
FIML*	0.086	0.068	0.098	0.090	0.057	0.109	0.069	0.088
MCMC	0.087	0.070	0.098	0.090	0.056	0.107	0.070	0.086
MCMC*	0.087	0.070	0.099	0.091	0.057	0.108	0.071	0.086
TS**	0.087	0.070	0.099	0.106	0.076	0.131	0.089	0.112
TS*	0.087	0.069	0.099	0.107	0.076	0.131	0.089	0.113
LBTS***	0.087	0.070	0.098	0.090	0.056	0.109	0.067	0.088
LBTS*	0.087	0.069	0.099	0.090	0.057	0.108	0.069	0.088
Mean Squared Error								
FIML*	0.012	0.008	0.016	0.013	0.005	0.019	0.007	0.012
MCMC	0.012	0.008	0.016	0.013	0.005	0.018	0.008	0.012
MCMC*	0.012	0.008	0.016	0.013	0.005	0.019	0.008	0.012
TS**	0.012	0.008	0.016	0.018	0.009	0.027	0.012	0.019
TS*	0.012	0.008	0.016	0.019	0.009	0.028	0.012	0.020
LBTS***	0.012	0.008	0.016	0.013	0.005	0.019	0.007	0.012
LBTS*	0.012	0.008	0.016	0.013	0.005	0.019	0.007	0.012
SIF for all MCMC outputs								
Mean	10	20	11	18	8	7	30	48
Standard Deviation	3	5	3	4	2	2	7	11

* 873 samples included; ** 1000 samples included; *** 915 samples included; 1000 samples included in MCMC

Table 4.4: *Estimator Performance when ρ=0.8 and $x_{i1} \neq x_{i2}$*

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ	ρ
True Values	1.000	-1.500	-1.000	-0.800	-1.000	1.500	1.200	0.800
Mean								
FIML*	1.003	-1.507	-1.004	-0.806	-1.009	1.512	1.208	0.803
MCMC	1.016	-1.523	-1.017	-0.807	-1.015	1.521	1.199	0.766
MCMC*	1.014	-1.522	-1.015	-0.807	-1.015	1.521	1.197	0.767
TS**	1.006	-1.508	-1.006	-0.909	-1.143	1.712	1.370	0.910
TS*	1.003	-1.508	-1.004	-0.909	-1.144	1.713	1.369	0.911
LBTS***	1.006	-1.507	-1.006	-0.806	-1.009	1.513	1.210	0.800
LBTS*	1.003	-1.508	-1.004	-0.805	-1.010	1.512	1.209	0.800
Standard Deviation								
FIML*	0.104	0.086	0.120	0.102	0.069	0.127	0.085	0.070
MCMC	0.106	0.087	0.121	0.103	0.068	0.125	0.083	0.064
MCMC*	0.106	0.086	0.121	0.103	0.069	0.127	0.083	0.064
TS**	0.110	0.088	0.127	0.123	0.076	0.141	0.094	0.104
TS*	0.110	0.088	0.127	0.123	0.077	0.143	0.095	0.103
LBTS***	0.110	0.088	0.127	0.102	0.069	0.125	0.085	0.069
LBTS*	0.110	0.088	0.127	0.102	0.070	0.127	0.085	0.070
Maximum Values								
FIML*	1.351	-1.219	-0.554	-0.477	-0.812	2.006	1.506	0.978
MCMC	1.367	-1.238	-0.562	-0.489	-0.817	2.011	1.499	0.896
MCMC*	1.367	-1.238	-0.562	-0.489	-0.817	2.011	1.499	0.896
TS**	1.379	-1.250	-0.590	-0.552	-0.919	2.172	1.685	1.000
TS*	1.379	-1.250	-0.590	-0.552	-0.919	2.172	1.685	1.000
LBTS***	1.379	-1.250	-0.590	-0.471	-0.813	2.005	1.515	0.971
LBTS*	1.379	-1.250	-0.590	-0.471	-0.813	2.005	1.515	0.971
Minimum Values								
FIML*	0.611	-1.908	-1.438	-1.214	-1.231	1.176	0.968	0.543
MCMC	0.616	-1.920	-1.457	-1.219	-1.231	1.202	0.962	0.517
MCMC*	0.616	-1.920	-1.457	-1.219	-1.231	1.202	0.962	0.517
TS**	0.637	-1.903	-1.440	-1.289	-1.358	1.340	1.118	0.543
TS*	0.637	-1.903	-1.440	-1.289	-1.358	1.340	1.118	0.543
LBTS***	0.637	-1.903	-1.440	-1.213	-1.233	1.180	0.969	0.536
LBTS*	0.637	-1.903	-1.440	-1.213	-1.233	1.180	0.969	0.536
Mean Absolute Error								
FIML*	0.083	0.068	0.093	0.082	0.057	0.100	0.068	0.056
MCMC	0.085	0.070	0.095	0.082	0.056	0.100	0.067	0.055
MCMC*	0.085	0.069	0.095	0.082	0.057	0.101	0.067	0.055
TS**	0.087	0.070	0.099	0.133	0.145	0.220	0.172	0.135
TS*	0.087	0.069	0.099	0.134	0.145	0.220	0.171	0.136
LBTS***	0.087	0.069	0.099	0.082	0.056	0.099	0.068	0.055
LBTS*	0.087	0.069	0.099	0.082	0.057	0.101	0.068	0.055
Mean Squared Error								
FIML*	0.011	0.007	0.014	0.010	0.005	0.016	0.007	0.005
MCMC	0.011	0.008	0.015	0.011	0.005	0.016	0.007	0.005
MCMC*	0.011	0.008	0.015	0.011	0.005	0.016	0.007	0.005
TS**	0.012	0.008	0.016	0.027	0.026	0.065	0.038	0.023
TS*	0.012	0.008	0.016	0.027	0.027	0.065	0.037	0.023
LBTS***	0.012	0.008	0.016	0.010	0.005	0.016	0.007	0.005
LBTS*	0.012	0.008	0.016	0.010	0.005	0.016	0.007	0.005
SIF for all MCMC outputs								
Mean	17	36	18	26	15	13	45	70
Standard Deviation	5	10	6	7	4	4	11	13

* 899 samples included; ** 1000 samples included; *** 940 samples included; 1000 samples included in MCMC

estimates are somewhat different (as seen by comparing the maximum and minimum values across methods) but have no big effect on overall estimation.

Table 4.3 when $\rho = 0.5$, shows that FIML, the Bayesian method and likelihood-based two-step method can provide good estimators, although the two-step method performs a bit worse in some dimensions. FIML is reliable with unbiased mean values and small standard deviations and estimation errors. In contrast, the MCMC method also offers good estimation with slightly smaller standard deviations than FIML for some parameters and larger standard deviations for other parameters. There is little difference between methods in mean of absolute errors and mean of squared errors. When comparing the likelihood-based two-step method with FIML, results indicate no big difference between the methods, no matter according to bias, variance or estimation errors. For the two-step method, mean values can show bias for certain parameters such as β_{22}, β_3 and γ, where the bias is more than 0.05. It also has the largest standard deviations and estimation errors in comparison with other methods, however, differences are still small.

When $\rho = 0.8$, there is virtually difference in performance of FIML, the Bayesian method and the likelihood-based two-step method, while the two-step method performs worst among the four estimators. Mean values of FIML and the likelihood-based two-step method are very close to the true values. The difference of mean values between the two methods are smaller than 0.003. Most mean values estimated by the MCMC method show a little more bias than those estimated by FIML except for γ. However, the bias of each parameter is normally less than 0.04 which is a very small amount. Difference in standard deviations and estimation errors are quite small among the three methods. For the two-step method, bias of β_2, γ and ρ is much more obvious as ρ increases. Standard deviation of such parameters are also a little larger than those of other methods, while mean of absolute errors and mean of squared errors of β_2, γ and ρ are as twice big as those estimated by other methods. As mentioned in Section 4.4.2, ρ is replaced by ρ^* and they are not identical for the two-step method, so it has some difficulty in estimating ρ accurately when true ρ value is large. That is why the two-step method gives a little more biased and inefficient estimation in the second step.

In all, the Bayesian method and likelihood-based two-step method can provide as accurate and efficient estimation as FIML when $x_{i1} \neq x_{i2}$, while the two-step method performs a little worse with large error correlation. FIML is quite reliable except for some convergence problems. The Bayesian method needs more computation than other methods, but can produce reliable estimates for each sample. The likelihood-based two-step method is less likely to fail to converge than FIML, while the two-step method fails less frequently than the likelihood-based two-step method.

4.4.6 Comparison when $x_{i1} = x_{i2}$

We are interested in the effects of exclusion restrictions, when the model has sample selection issues. We would expect the sample selection model to have similar properties as models with endogenous treatment effects, where exclusion restrictions can influence the precision of estimators.

Therefore, tables of results with exclusion restrictions are compared to tables without exclusion restrictions based on the same error correlation. Let us compare Table 4.2 with Table 4.5 first. Although mean values appear slightly different, the estimates for all methods are still unbiased in Table 4.5. The standard deviations of β_1 remain in the same level, but those of β_{21}, β_{23} and γ are a little larger while those of β_{22} and ρ have doubled in Table 4.5 in comparison with standard deviations in Table 4.2. Mean of absolute errors and mean of squared error of β_{22} and ρ are also much larger in Table 4.5 than those in Table 4.2. Meanwhile, maximum and minimum values are much more extreme, when the model is without exclusion restrictions. FIML, the two-step and likelihood-based two-step methods also fail more frequently. Mean SIF values of β_1 are still quite small, but those of other parameters jump a lot especially for ρ whose mean SIF value has increased from 37 in Table 4.2 to 120 in Table 4.5. Similar patterns can be found when comparing Table 4.3 and Table 4.6. When $\rho = 0.8$, however, differences in standard deviations and estimation errors between Table 4.4 and Table 4.7 are not as obvious as differences when ρ is smaller. What we conclude is that, although models without exclusion restrictions can still be identified, they are more difficult to estimate than those with exclusion restrictions.

Table 4.5: *Estimator Performance when ρ=0 and $x_{i1} = x_{i2}$*

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ	ρ
True Values	1.000	-1.500	-1.000	-0.800	-1.000	1.500	1.200	0.000
Mean								
FIML*	1.008	-1.511	-1.005	-0.805	-1.006	1.481	1.187	0.019
MCMC	1.008	-1.515	-1.005	-0.816	-1.004	1.461	1.151	0.034
MCMC*	1.011	-1.515	-1.007	-0.818	-1.007	1.457	1.150	0.040
TS**	1.007	-1.512	-1.005	-0.812	-1.014	1.505	1.205	0.008
TS*	1.008	-1.511	-1.005	-0.816	-1.020	1.500	1.206	0.019
LBTS***	1.008	-1.513	-1.006	-0.805	-1.004	1.485	1.188	0.018
LBTS*	1.008	-1.511	-1.005	-0.808	-1.009	1.479	1.187	0.025
Standard Deviation								
FIML*	0.107	0.088	0.125	0.167	0.155	0.146	0.085	0.278
MCMC	0.108	0.088	0.125	0.144	0.129	0.133	0.083	0.208
MCMC*	0.107	0.088	0.125	0.146	0.131	0.134	0.081	0.213
TS**	0.107	0.088	0.124	0.173	0.164	0.146	0.084	0.292
TS*	0.106	0.088	0.124	0.173	0.163	0.147	0.083	0.292
LBTS***	0.108	0.088	0.124	0.163	0.153	0.146	0.087	0.272
LBTS*	0.106	0.088	0.124	0.167	0.155	0.146	0.085	0.277
Maximum Values								
FIML*	1.392	-1.304	-0.632	-0.290	-0.486	1.941	1.442	0.867
MCMC	1.395	-1.302	-0.603	-0.251	-0.574	1.887	1.451	0.586
MCMC*	1.395	-1.302	-0.632	-0.379	-0.574	1.887	1.399	0.582
TS**	1.387	-1.303	-0.616	-0.132	-0.505	1.964	1.484	0.975
TS*	1.387	-1.303	-0.630	-0.259	-0.532	1.964	1.461	0.975
LBTS***	1.387	-1.303	-0.630	-0.188	-0.499	1.941	1.504	0.859
LBTS*	1.387	-1.303	-0.630	-0.283	-0.499	1.941	1.460	0.859
Minimum Values								
FIML*	0.738	-1.849	-1.478	-1.355	-1.465	0.891	0.934	-0.751
MCMC	0.695	-1.847	-1.479	-1.306	-1.377	0.995	0.909	-0.614
MCMC*	0.737	-1.847	-1.479	-1.306	-1.377	0.995	0.909	-0.614
TS**	0.683	-1.849	-1.470	-1.315	-1.603	1.047	0.954	-1.000
TS*	0.736	-1.849	-1.470	-1.305	-1.603	1.047	0.954	-1.000
LBTS***	0.683	-1.849	-1.470	-1.354	-1.470	0.903	0.934	-0.750
LBTS*	0.736	-1.849	-1.470	-1.354	-1.470	0.903	0.934	-0.750
Mean Absolute Error								
FIML*	0.087	0.070	0.100	0.133	0.125	0.114	0.068	0.226
MCMC	0.088	0.071	0.100	0.114	0.103	0.109	0.077	0.168
MCMC*	0.087	0.071	0.101	0.116	0.106	0.109	0.076	0.177
TS**	0.087	0.071	0.099	0.137	0.130	0.115	0.067	0.234
TS*	0.086	0.070	0.100	0.137	0.130	0.115	0.066	0.234
LBTS***	0.087	0.070	0.100	0.129	0.122	0.115	0.070	0.221
LBTS*	0.086	0.070	0.100	0.132	0.125	0.115	0.068	0.225
Mean Squared Error								
FIML*	0.011	0.008	0.016	0.028	0.024	0.022	0.007	0.078
MCMC	0.012	0.008	0.016	0.021	0.017	0.019	0.009	0.044
MCMC*	0.012	0.008	0.016	0.022	0.017	0.020	0.009	0.047
TS**	0.012	0.008	0.015	0.030	0.027	0.021	0.007	0.085
TS*	0.011	0.008	0.015	0.030	0.027	0.021	0.007	0.086
LBTS***	0.012	0.008	0.016	0.027	0.023	0.022	0.008	0.074
LBTS*	0.011	0.008	0.015	0.028	0.024	0.022	0.007	0.077
SIF for all MCMC outputs								
Mean	8	17	10	79	97	58	37	120
Standard Deviation	2	4	3	17	18	22	13	18

* 720 samples included; ** 952 samples included; *** 830 samples included; 1000 samples included in MCMC

Table 4.6: *Estimator Performance when ρ=0.5 and $x_{i1} = x_{i2}$*

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ	ρ
True Values	1.000	-1.500	-1.000	-0.800	-1.000	1.500	1.200	0.500
Mean								
FIML*	1.006	-1.512	-1.003	-0.800	-0.994	1.502	1.192	0.488
MCMC	1.014	-1.522	-1.011	-0.763	-0.949	1.543	1.185	0.382
MCMC*	1.015	-1.523	-1.011	-0.771	-0.954	1.540	1.183	0.392
TS**	1.005	-1.511	-1.003	-0.838	-1.044	1.586	1.274	0.519
TS*	1.006	-1.512	-1.002	-0.847	-1.049	1.584	1.274	0.531
LBTS***	1.006	-1.512	-1.005	-0.795	-0.992	1.500	1.192	0.485
LBTS*	1.006	-1.512	-1.002	-0.800	-0.994	1.503	1.193	0.488
Standard Deviation								
FIML*	0.107	0.087	0.125	0.128	0.109	0.167	0.096	0.209
MCMC	0.108	0.087	0.125	0.121	0.106	0.139	0.084	0.169
MCMC*	0.108	0.087	0.126	0.121	0.101	0.140	0.083	0.165
TS**	0.107	0.087	0.124	0.162	0.156	0.143	0.086	0.267
TS*	0.107	0.088	0.125	0.161	0.151	0.145	0.085	0.263
LBTS***	0.107	0.088	0.123	0.127	0.112	0.165	0.095	0.207
LBTS*	0.107	0.088	0.125	0.128	0.109	0.166	0.096	0.207
Maximum Values								
FIML*	1.327	-1.293	-0.616	-0.132	-0.408	2.044	1.548	0.893
MCMC	1.411	-1.307	-0.618	-0.245	-0.528	2.009	1.488	0.718
MCMC*	1.328	-1.310	-0.618	-0.245	-0.528	2.009	1.488	0.718
TS**	1.387	-1.303	-0.616	-0.286	-0.593	2.046	1.549	1.000
TS*	1.327	-1.304	-0.616	-0.286	-0.593	2.046	1.549	1.000
LBTS***	1.327	-1.303	-0.616	-0.138	-0.412	2.044	1.549	0.884
LBTS*	1.327	-1.304	-0.616	-0.138	-0.412	2.044	1.549	0.884
Minimum Values								
FIML*	0.717	-1.846	-1.396	-1.255	-1.281	0.902	0.937	-0.647
MCMC	0.683	-1.848	-1.506	-1.199	-1.401	0.982	0.935	-0.436
MCMC*	0.723	-1.848	-1.397	-1.199	-1.242	0.982	0.935	-0.436
TS**	0.683	-1.849	-1.470	-1.371	-1.571	1.045	1.016	-0.373
TS*	0.708	-1.849	-1.383	-1.371	-1.420	1.045	1.016	-0.373
LBTS***	0.683	-1.849	-1.383	-1.249	-1.445	0.899	0.935	-0.639
LBTS*	0.708	-1.849	-1.383	-1.249	-1.282	0.899	0.935	-0.639
Mean Absolute Error								
FIML*	0.086	0.070	0.100	0.100	0.084	0.131	0.077	0.161
MCMC	0.088	0.071	0.101	0.101	0.091	0.113	0.069	0.155
MCMC*	0.088	0.072	0.102	0.098	0.086	0.114	0.068	0.147
TS**	0.087	0.070	0.099	0.134	0.129	0.132	0.092	0.217
TS*	0.087	0.070	0.101	0.134	0.127	0.132	0.091	0.214
LBTS***	0.087	0.071	0.098	0.099	0.087	0.129	0.076	0.161
LBTS*	0.087	0.070	0.101	0.100	0.083	0.131	0.077	0.160
Mean Squared Error								
FIML*	0.011	0.008	0.016	0.016	0.012	0.028	0.009	0.044
MCMC	0.012	0.008	0.016	0.016	0.014	0.021	0.007	0.043
MCMC*	0.012	0.008	0.016	0.015	0.012	0.021	0.007	0.039
TS**	0.012	0.008	0.015	0.028	0.026	0.028	0.013	0.072
TS*	0.012	0.008	0.016	0.028	0.025	0.028	0.013	0.070
LBTS***	0.012	0.008	0.015	0.016	0.013	0.027	0.009	0.043
LBTS*	0.012	0.008	0.016	0.016	0.012	0.027	0.009	0.043
SIF for all MCMC outputs								
Mean	9	20	12	65	84	70	50	113
Standard Deviation	3	5	3	19	22	16	14	18

* 803 samples included; ** 992 samples included; *** 872 samples included; 1000 samples included in MCMC

Table 4.7: *Estimator Performance when ρ=0.8 and $x_{i1} = x_{i2}$*

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ	ρ
True Values	1.000	-1.500	-1.000	-0.800	-1.000	1.500	1.200	0.800
Mean								
FIML*	1.006	-1.511	-1.003	-0.804	-1.002	1.510	1.200	0.797
MCMC	1.032	-1.538	-1.033	-0.753	-0.946	1.642	1.262	0.643
MCMC*	1.031	-1.537	-1.033	-0.755	-0.948	1.645	1.264	0.644
TS**	1.005	-1.511	-1.003	-0.869	-1.086	1.762	1.419	0.849
TS*	1.004	-1.510	-1.002	-0.872	-1.088	1.766	1.421	0.851
LBTS***	1.003	-1.510	-1.000	-0.801	-0.999	1.510	1.201	0.794
LBTS*	1.004	-1.510	-1.002	-0.803	-1.001	1.512	1.201	0.795
Standard Deviation								
FIML*	0.106	0.084	0.124	0.102	0.079	0.159	0.112	0.098
MCMC	0.108	0.086	0.125	0.103	0.084	0.140	0.093	0.097
MCMC*	0.109	0.085	0.126	0.101	0.083	0.142	0.093	0.096
TS**	0.107	0.087	0.124	0.140	0.131	0.145	0.099	0.195
TS*	0.108	0.086	0.126	0.140	0.130	0.147	0.099	0.195
LBTS***	0.107	0.086	0.125	0.103	0.079	0.158	0.112	0.100
LBTS*	0.108	0.086	0.126	0.102	0.079	0.159	0.111	0.099
Maximum Values								
FIML*	1.395	-1.289	-0.638	-0.338	-0.574	2.094	1.633	0.959
MCMC	1.433	-1.317	-0.661	-0.362	-0.580	2.134	1.612	0.803
MCMC*	1.433	-1.317	-0.661	-0.362	-0.583	2.134	1.612	0.803
TS**	1.387	-1.303	-0.616	-0.365	-0.593	2.265	1.776	1.000
TS*	1.387	-1.304	-0.616	-0.365	-0.615	2.265	1.776	1.000
LBTS***	1.387	-1.303	-0.616	-0.337	-0.573	2.093	1.634	0.958
LBTS*	1.387	-1.304	-0.616	-0.337	-0.573	2.093	1.634	0.958
Minimum Values								
FIML*	0.680	-1.840	-1.490	-1.220	-1.231	1.099	0.902	-0.137
MCMC	0.704	-1.856	-1.534	-1.170	-1.241	1.260	1.027	-0.088
MCMC*	0.704	-1.856	-1.534	-1.170	-1.177	1.260	1.027	-0.088
TS**	0.683	-1.849	-1.470	-1.425	-1.438	1.353	1.119	-0.051
TS*	0.683	-1.849	-1.470	-1.425	-1.438	1.353	1.119	-0.051
LBTS***	0.683	-1.773	-1.470	-1.219	-1.225	1.100	0.910	-0.141
LBTS*	0.683	-1.849	-1.470	-1.219	-1.225	1.100	0.924	-0.141
Mean Absolute Error								
FIML*	0.086	0.067	0.100	0.080	0.063	0.125	0.088	0.072
MCMC	0.090	0.075	0.103	0.090	0.079	0.163	0.087	0.157
MCMC*	0.091	0.074	0.104	0.089	0.077	0.166	0.087	0.156
TS**	0.087	0.070	0.100	0.128	0.132	0.266	0.219	0.174
TS*	0.087	0.069	0.101	0.128	0.132	0.270	0.221	0.174
LBTS***	0.086	0.069	0.100	0.080	0.063	0.124	0.088	0.073
LBTS*	0.087	0.069	0.101	0.080	0.063	0.125	0.087	0.073
Mean Squared Error								
FIML*	0.011	0.007	0.015	0.010	0.006	0.025	0.012	0.010
MCMC	0.013	0.009	0.017	0.013	0.010	0.040	0.013	0.034
MCMC*	0.013	0.009	0.017	0.012	0.010	0.041	0.013	0.033
TS**	0.012	0.008	0.015	0.024	0.025	0.090	0.058	0.041
TS*	0.012	0.008	0.016	0.025	0.024	0.092	0.059	0.040
LBTS***	0.011	0.008	0.016	0.011	0.006	0.025	0.013	0.010
LBTS*	0.012	0.008	0.016	0.010	0.006	0.025	0.012	0.010
SIF for all MCMC outputs								
Mean	13	27	17	49	62	63	62	98
Standard Deviation	4	7	5	15	21	15	13	20

* 871 samples included; ** 1000 samples included; *** 934 samples included; 1000 samples included in MCMC

We are also interested in how different methods perform when they are used to estimate sample selection models without exclusion restrictions. Thus, each table is analyzed across methods again. In Table 4.5, all mean values suggest very small biases. The largest bias appears in the mean value of γ estimated by the MCMC method which is biased by around 0.05 from the true value. Mean values of β_{23} and γ estimated by the two-step method are closer to the true values than those of the other three methods. The MCMC method tends to work more efficiently as most values of standard deviations, mean of absolute error and mean of squared error are smaller than those estimated by other methods. In contrast, the two-step method has the largest standard deviations and estimation errors of β_{21}, β_{22} and ρ. Statistics are quite close across FIML and the likelihood-based two-step method. Overall, the differences in summary statistics are very small across methods in this table. Thus, all four estimators work fine when $\rho = 0$.

Next, we consider how the four estimators perform when $\rho = 0.5$. Table 4.6 reveals that most mean values suggest the estimators are still unbiased. However, the mean value of the estimator of ρ estimated by the MCMC method is biased by more than 0.1, while mean values of β_{23} and γ evaluated by the two-step method are biased by around 0.08. Standard deviations of each parameter in β_1 are close across all methods. Standard deviations of estimates of β_2, γ and ρ estimated by the MCMC method are the smallest in comparison to those estimated by other methods, while standard deviations of β_{21}, β_{22} and ρ estimated by the two-step method are largest. According to estimation errors, the two-step method shows more extreme values of β_{21}, β_{22}, γ and ρ while other statistics are close across methods. In short, all four methods perform reasonably well, except a little bias for the MCMC method and the two-step method.

Table 4.7 presents summary statistics with $\rho = 0.8$ and we will focus on the difference across methods. Mean values of β_{23} and ρ estimated by the MCMC method are biased from the true values by around 0.15, while estimates of β_{23} and γ from the two-step method are biased by more than 0.2. Standard deviations of estimates of β_1 are almost the same for all methods. According to standard deviations of other parameters, the MCMC method can provide relatively more precise estimates of β_{23} and γ, while the two-step method is more inaccurate in estimating β_{21}, β_{22} and ρ. The mean absolute error and mean squared error from FIML and the likelihood-based two-step method are almost the same, while those from the MCMC method are slightly larger. However, estimation errors for estimates of β_2, γ and ρ from the two-step method are the largest among all methods. Maximum and minimum values are a little different for the different methods, but such differences are not big.

In summary, FIML has small bias and is efficient, but has more than one tenth convergence failure rate. The Bayesian method can estimate all samples, but has small bias on certain parameters when ρ is large while the rest of the parameters are estimated efficiently and with small bias. The two-step method only fails to converge a few times when $\rho = 0$. However, some bias and inefficiency arise with strong error correlation. Finally, estimation based on a two-step maximum likelihood method is always close to that of FIML, but fails to converge much less often than FIML.

4.5 An Application to Mental Illness and Labor Market Employment

It is widely acknowledged that mental health is an important factor in determining labor market outcomes. Mental illness can not only affect people's chances of finding employment, but also influence people's capacity to work, the occupational skill levels at which they work, and their earnings. Nationwide mental health surveys provide us with the opportunity to examine the relationship between mental health factors and labor market outcomes. In this empirical application we look at the relationship between mental illness and an individual's chances of being employed, and for the employed, the impact of mental illness on a person's occupational skill level. It might be suggested that mental illness would hinder the chances of finding employment, and might also mean people work in lower skilled jobs than they would otherwise have if they did not suffer from an illness. This application will look at whether there is empirical support for these effects, using the Australian National Survey of Mental Health and Wellbeing of Adults for 1997.

Substantial research suggests that the unemployment or employment in lower-level jobs can cause mental health problems. For example, Flatau et al. (2000) show that there are negative associations between mental health and unemployment. Artazcoz et al. (2004) believe that the financial strain of unemployment can cause poor mental health, while unemployment can also be associated with poor mental health as a result of the absence of nonfinancial benefits provided by one's job, such as social status, self-esteem, physical and mental activity, and use of one's skills. Some research explores the endogeneity of the relationship between mental health and employment. For example, Hamilton et al. (1997) apply a simultaneous equation generalized probit model to estimate jointly the determinants of an individual's latent index of employability and their mental health, and find that employability leads to improved mental health and that stronger mental health improves employability.

The relationship between occupational levels and mental health is not clear. It is possible that some poor mental health among the employed may be due to a recent proliferation of low-skill jobs, to the take-up of jobs not matched to the skill background of the individual, or to job insecurity (Flatau et al. 2000). Employment in lower-level jobs usually involves high demand and a lack of control over decision making, thus may cause an accumulation of stress and can result in anxiety, depression and psychosomatic illness as well as physical disorders (Dockery 2006). However, some argue that high-stress senior management jobs can lead to increased incidence of mental health conditions, especially depression and anxiety (Cornwell et al. 2009).

This example will investigate the effects of mental health on the relationship between employment and occupational skill levels. Sample selection exists in this case because we can only observe occupational levels for people who are employed. The data set we use here is the same as that used by Cornwell et al. (2009). They seek to explain labor market outcomes by mental health states, but if there is causality in the other direction, the mental health variable will be endogenous, and biased estimates will result. So they deal with endogeneity by use of temporal information in the data to make sure the mental illness cannot have been caused by the unemployment experience. It is revealed that there is a strongly significant effect of mental illness on employment and clear evidence of reduced occupational skill level.

The estimation method used by Cornwell et al. (2009) is the two-step method discussed in Section 4.4.2. However, their model has a three-equation system, so the two-step method is applied twice. After the first equation is estimated as a probit model, an estimated inverse Mills ratio is plugged into the second equation as a regressor. Then parameters in the second equation are estimated and the estimates are used to form a second correction factor to add into the third equation. As we have commented before, this two-step method is inconsistent when the second equation is non-linear, so FIML and the Bayesian method will be utilized to estimate the parameters in this empirical example.

4.5.1 Features of the Data

The data we use for the model are from the 1997 National Survey of Mental Health and Wellbeing of Adults (NSMHWB) in Australia. This survey collects information about normal demographic factors and various mental health indicators, involving a total of more than 10,000 participants. Since effects of participation in labor market are not particularly considered in this section, a sample of 6,928 observations is used to estimate our model after removal of respondents who do not participate in the labor market.

The probability of employment can be driven by a number of socioeconomic factors like gender, age, education levels and geographic factors. Health concerns including mental and physical conditions are also considered in the equation to capture the influence of health status on the labor market. Since all factors which determine employment can possibly also affect the choice of occupations, the explanatory variables are the same in both binary choice and ordered probit equations.

Table 4.8: *Summary Statistics of Number of Mental Health Disorders (A)*

(with less than 12 months)

Number of Mental Health Disorders with less than 12 months	Frequency	Percent
0	6,805	98.2
1	114	1.7
2	8	0.1
3	1	0.0
Total	6,928	100.0

Table 4.8 and Table 4.9 describe some summary statistics about number of mental health disorders. Table 4.8 shows that a small proportion of respondents (1.78%) have short term mental health disorders whose onset was less than 12 months ago, while most of them have only one disorder. Table 4.9, however, shows that 17.93% of respondents

Table 4.9: *Summary statistics of Number of Mental Health Disorders (B)*

(with more than 12 months)

Number of Mental Health Disorders with more than 12 months	Frequency	Percent
0	5,686	82.1
1	608	8.8
2	299	4.3
3	171	2.5
4	101	1.5
5	43	0.6
6	10	0.1
7	4	0.1
8	3	0.0
9	3	0.0
Total	6,928	100.0

suffer long term mental health problems while 4.84% suffer more than two mental health disorders. 17.93% is not a small percentage, but is not surprising because people tend to have more and more pressure in modern society, especially when they are living in a fast life style.

4.5.2 Estimates

This specific model has almost the same latent structure of Equation (4.2.1),

$$\begin{cases} z_{i1} = x'_{i1}\beta_1 + \epsilon_{i1} \\ z_{i2} = x'_{i1}\beta_2 + \epsilon_{i2} \qquad 1 \leq i \leq n. \end{cases}$$

Notice that exclusion restrictions are not imposed here, so independent variables are exactly the same in the two equations, namely $x_{i1} = x_{i2}$. The regressors include individual characteristics that are generally seen to impact labour market outcomes - gender, age, level of education and indices indicating a person's socio-economic status. In addition, two mental health indicators and one physical health indicator are also included in the set of regressors to test their effects. All remaining effects are assumed to be contained in error terms. If the errors across the two equations are correlated, then the equation for y_{i2} cannot be estimated without taking account of the sample selection given by the equation for y_{i1}. The correlation introduces a selection bias in the occupational skill category equation. Therefore, ϵ_{i1} and ϵ_{i2} are defined in the same way as in Equation (4.2.1) with correlation ρ.

The individual is employed if $z_{i1} > 0$, so $y_{i1} = 1$. Otherwise, $z_{i1} < 0$, so $y_{i1} = 0$. Occupational skill levels are shown as y_{i2} which are categorical data. Although respondents answer the question about most recent occupation even when they are not employed, we believe that long term unemployment will cause possible bias in evaluating their true occupational levels. As a result, occupational skill categories are assumed missing when respondents are not employed, which makes this data set consistent with our model. The missing cases are given a zero value, so the occupational equation can be written as

$$y_{i2} = \begin{cases} 0 \text{ missing if } y_{i1} = 0 \\ 1 \text{ if } y_{i1} = 1 \text{ and } z_{i2} < 0 \\ 2 \text{ if } y_{i1} = 1 \text{ and } 0 < z_{i2} < \gamma_2 \\ 3 \text{ if } y_{i1} = 1 \text{ and } \gamma_2 < z_{i2} < \gamma_3 \\ 4 \text{ if } y_{i1} = 1 \text{ and } \gamma_3 < z_{i2} < \gamma_4 \\ 5 \text{ if } y_{i1} = 1 \text{ and } \gamma_4 < z_{i2}. \end{cases}$$

More specifically, $y_{i2} = 0$ if the individual is not employed. $y_{i2} = 1$ if the individual is employed as elementary clerical, sales and service workers, laborers and related workers. $y_{i2} = 2$ represents employment as intermediate clerical, sales, service, production and transport workers. $y_{i2} = 3$ if one is employed as trades persons and related workers, advanced clerical and service workers. $y_{i2} = 4$ represents associate professionals, while $y_{i2} = 5$ means managers, administrators and professionals.

Two estimation methods, FIML and the Bayesian method are applied on this empirical data. In order to get reliable FIML estimates, the data is estimated by two programs. The CML package in GAUSS fails to invert the covariance matrix of the parameter estimates, no matter what the starting values are, so such outputs cannot be relied on. Therefore, we use another commercial software NLOGIT 4.0, which also applies FIML to estimate parameters for this particular type of model (Greene 2002). We find that values of parameters estimated by both software packages are very similar even with different starting values, except for ρ. The ρ value estimated by GAUSS can be very sensitive to starting values while that obtained by NLOGIT is not significant.

Table 4.10: *Parameter Estimation Using FIML*

Variable	Coefficient	St.Er.	$P[\lvert Z\rvert > z]$	Mean of X
Employment Equation				
Constant	1.159	0.072	0.000	
Male	-0.063	0.050	0.207	0.512
Age 25-44	0.359	0.064	0.000	0.557
Age 45-64	0.425	0.073	0.000	0.295
Secondary School	0.249	0.057	0.000	0.500
Higher Education	0.481	0.088	0.000	0.195
Vocational Education	0.249	0.094	0.008	0.113
Rural Area	-0.148	0.068	0.028	0.136
Regional Center	0.098	0.067	0.145	0.182
Number of Mental Health Disorders with less than 12 months	-0.370	0.133	-2.784	0.019
Number of Mental Health Disorders with more than 12 months	-0.161	0.020	-8.020	0.358
Has Physical Illness	-0.018	0.055	0.741	0.289
Skill Equation				
Constant	0.139	0.120	0.245	
Male	0.210	0.032	0.000	0.512
Age 25-44	0.393	0.046	0.000	0.557
Age 45-64	0.456	0.051	0.000	0.295
Secondary School	0.208	0.036	0.000	0.500
Higher Education	1.549	0.043	0.000	0.195
Vocational Education	0.771	0.045	0.000	0.113
Rural Area	-0.099	0.044	0.025	0.136
Regional Center	0.225	0.035	0.000	0.182
Number of Mental Health Disorders with less than 12 months	-0.196	0.103	0.056	0.019
Number of Mental Health Disorders with more than 12 months	-0.054	0.026	0.039	0.358
Has Physical Illness	-0.026	0.030	0.395	0.289
γ_2	0.802	0.047	0.000	
γ_3	1.328	0.067	0.000	
γ_4	1.730	0.078	0.000	
ρ	0.404	0.476	0.397	

Table 4.10 presents FIML results estimated by NLOGIT, including estimated coefficients, standard error, p-value and mean of X. We discuss various factors. If 5% is chosen as the significance level, gender does not seem to have an effect on employment as its p-value is 0.207, but can impact on occupational skill levels. Age groups and education play quite important roles in determining both employment and occupational skill levels, as

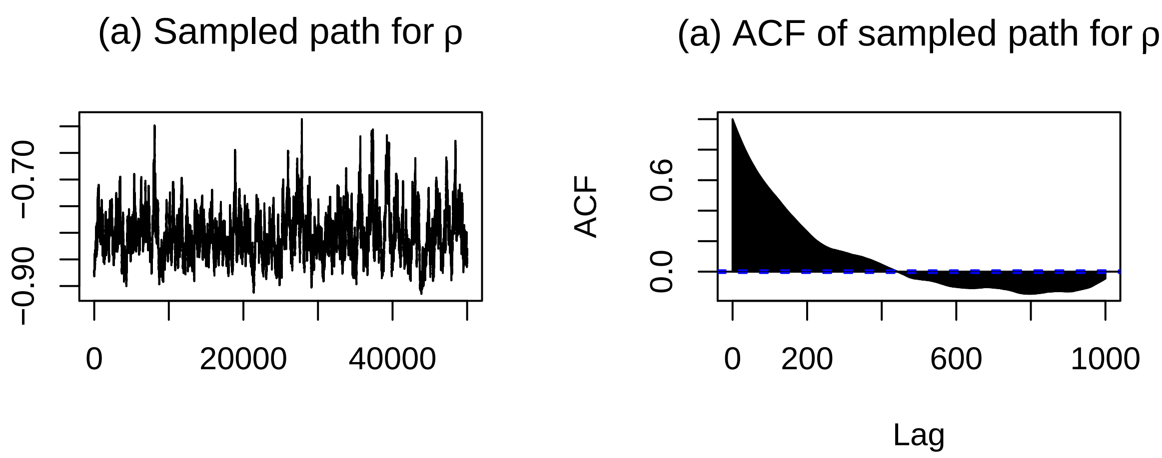
corresponding dummy variables are all significant in both equations. The dummy variable about whether the one is from regional center is not significant in the employment equation but significant in the other equation. Having a physical illness is not significant in either equation, as the p-values are extremely large. P-values of the number of mental disorders with less than 12 months is slightly above 5%, which suggest mental disorders whose onset was less than 12 months age do not have much effect on occupational outcomes. The effects of this variable may be too difficult to detect as only 1.78 % respondents have short term mental illness. The estimate of ρ is not significant, thus no correlation exists between the two equations.

Table 4.11: *Parameter Estimation Using the MCMC Method*

	Mean	St. Dev	BM St. Dev	SIF	95% Credible Interval
Employment Equation					
Constant	1.182	0.112	0.012	12	(0.968 , 1.408)
Male	-0.032	0.070	0.010	21	(-0.167 , 0.107)
Age 25-44	0.263	0.082	0.013	24	(0.104 , 0.426)
Age 45-64	0.361	0.085	0.014	27	(0.198 , 0.532)
Secondary School	0.242	0.091	0.014	25	(0.070 , 0.428)
Higher Education	0.520	0.136	0.032	57	(0.283 , 0.812)
Vocational Education	0.282	0.122	0.020	28	(0.060 , 0.539)
Rural Area	-0.145	0.085	0.011	17	(-0.313 , 0.022)
Regional Center	0.184	0.083	0.013	26	(0.026 , 0.352)
Mental Health Disorders with less than 12 months	-0.360	0.173	0.021	15	(-0.687 , -0.004)
Mental Health Disorders with more than 12 months	-0.145	0.029	0.004	18	(-0.204 , -0.088)
Has Physical Illness	0.007	0.061	0.008	17	(-0.114 , 0.127)
Skill Equation					
Constant	0.396	0.042	0.004	9	(0.314 , 0.476)
Male	0.209	0.021	0.001	4	(0.169 , 0.250)
Age 25-44	0.260	0.027	0.004	22	(0.208 , 0.315)
Age 45-64	0.311	0.027	0.004	26	(0.259 , 0.364)
Secondary School	0.140	0.025	0.002	8	(0.092 , 0.188)
Higher Education	1.448	0.037	0.006	28	(1.376 , 1.520)
Vocational Education	0.689	0.031	0.004	13	(0.630 , 0.751)
Rural Area	-0.047	0.027	0.002	5	(-0.100 , 0.005)
Regional Center	0.201	0.024	0.001	4	(0.155 , 0.248)
Mental Health Disorders with less than 12 months	-0.061	0.065	0.005	7	(-0.189 , 0.065)
Mental Health Disorders with more than 12 months	0.004	0.012	0.001	11	(-0.019 , 0.027)
Has Physical Illness	-0.017	0.019	0.001	5	(-0.054 , 0.020)
γ_2	0.836	0.015	0.003	51	(0.807 , 0.864)
γ_3	1.170	0.020	0.005	51	(1.129 , 1.210)
γ_4	1.671	0.029	0.007	51	(1.613 , 1.728)
ρ	-0.805	0.044	0.021	215	(-0.878 , -0.704)

When it comes to the MCMC procedure, the priors are set in the same way in Section 4.3.5 and an arbitrary set of starting values are used. β_1, β_2 and ρ start with zeros and three threshold parameters are given values, 1, 2 and 3. A long burn-in period is necessary because the sample size is quite large and many variables are included in the model. As a result, we discard the first 20,000 iterations and record the following 50,000 iterations to evaluate statistics needed for the output. Table 4.11 shows some estimated coefficients and summary statistics applying the Bayesian method, including mean, standard deviation, batch mean standard deviation (BM St. Dev), simulation inefficient factors (SIF) and 95% Bayesian credible interval.

Figure 4.3: *Convergence of* ρ



Once estimates are obtained, MCMC convergence diagnosis is necessary using SIF values. As shown in Table 4.11, most SIF values vary from 4 to 50 which is an acceptable range, except the SIF value for the coefficient of correlation ρ. Considering the simulation results in Section 4.4.6, ρ also has the largest SIF values because it is the most difficult parameter to estimate in this model. Besides that, the ACF in Figure 4.3 indicates that ρ is slowly convergent after long lags. As a result, we still treat 215 as a convergent result although it is a quite large number. One should notice that the estimated ρ value, -0.805, is totally different from the one from FIML which is 0.404 but not significant. It is a puzzle that the two methods give such different estimates of ρ. The Monte Carlo experiments in Section 4.4.6 show that estimating ρ is relatively difficult than estimating other variables. The difficulty with estimating ρ in this application suggests the likelihood is probably bimodal in this case and it is hard to know where the true parameter lies. Meanwhile, negative error correlation is not common in empirical studies. But still possible. For instance, an individual who due to specific circumstances is especially keen to be employed may also be willing to accept a job for which they are over-qualified. In that case they may have a positive error in the employment equation, and a negative error in the skill category equation. At the other extreme, an individual who has enough savings and decides to only take on a position that matches his skills closely, may have a lower chance of employment and a higher skill level propensity.

A 5% significance level is applied in the Bayesian credible interval which is used to test significance in Table 4.11. Gender is not significant in determining employment status while males still dominate higher occupational levels. Obviously, age and education can influence both employment and occupational skill categories. The significance remain the same for gender, age and education variables, although the coefficients of those variables are somehow different across methods. The remaining coefficients have quite different significance across methods, while their magnitudes are similar for the two methods. With geographic factors, the variable indicating a person lives in a rural area is not significant in both equations. It may be caused by the fact that certain proportion of people who live in rural area travel to work in city or regional areas, so the effect of rural area variable is eliminated in some aspects. According to mental health variables, it seems both short term and long term mental disorders can determine employment, but have no effect on occupations because zero values are contained in the 95% credible intervals. Moreover, physical conditions have no affect on both equations. It is possible that physical conditions can only determine people's decision to participate. Therefore, physical conditions cannot help much in explaining labor market once people participate in labor market.

4.5.3 Marginal Effects

Marginal effects are useful in interpreting the model. If a variable is continuous, marginal effects can be obtained from $\partial Pr[y|\mathbf{x}]/\partial \mathbf{x}$, in which $Pr[y|\mathbf{x}]$ can be any probability of interest, such as expectation of y. This expression is evaluated at the sample means of the data. Greene (2003) suggests an appropriate marginal effect for a binary independent variable, say d, would be

$$\text{Marginal Effect} = Pr(y|\overline{x}_d, d = 1) - Pr(y|\overline{x}_d, d = 0), \tag{4.5.1}$$

where $\overline{x}_d$ denotes the means of all the other variables in the model. Notice that all variables are binary in this empirical work, except mental health variables.

We are particularly interested in the marginal effects on the probabilities for an individual to join certain occupational categories after being employed. Such probabilities can be written as

$$Pr(y_{i2} = 1|y_{i1} = 1) = Pr(y_{i1} = 1, y_{i2} = 1|x_{i1}, x_{i2})/Pr(y_{i1} = 1) \tag{4.5.2}$$

$$Pr(y_{i2} = 2|y_{i1} = 1) = Pr(y_{i1} = 1, y_{i2} = 2|x_{i1}, x_{i2})/Pr(y_{i1} = 1) \tag{4.5.3}$$

$$Pr(y_{i2} = 3|y_{i1} = 1) = Pr(y_{i1} = 1, y_{i2} = 3|x_{i1}, x_{i2})/Pr(y_{i1} = 1) \tag{4.5.4}$$

$$Pr(y_{i2} = 4|y_{i1} = 1) = Pr(y_{i1} = 1, y_{i2} = 4|x_{i1}, x_{i2})/Pr(y_{i1} = 1) \tag{4.5.5}$$

$$Pr(y_{i2} = 5|y_{i1} = 1) = Pr(y_{i1} = 1, y_{i2} = 5|x_{i1}, x_{i2})/Pr(y_{i1} = 1) \tag{4.5.6}$$

in which the joint probabilities are defined in equation (4.4.1) and $Pr(y_{i1} = 1) = \Phi(x'_{i1}\beta_1)$.

Table 4.12: *Marginal Effects Using FIML*

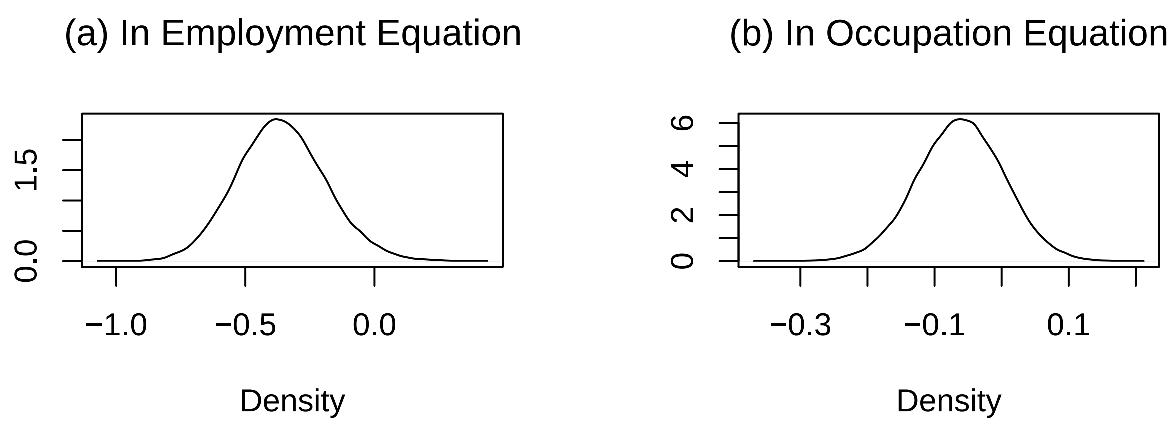
Variable	Elementary Skill	Intermediate Skill	Advanced Skill	Associate Professionals	Professionals
Male	-0.046	-0.037	-0.002	0.014	0.071
Age 25-44	-0.076	-0.063	-0.004	0.023	0.121
Age 45-64	-0.078	-0.077	-0.014	0.020	0.150
Secondary School	-0.038	-0.034	-0.003	0.011	0.064
Higher Education	-0.188	-0.241	-0.114	-0.010	0.553
Vocational Education	-0.109	-0.138	-0.050	0.013	0.283
Rural Area	0.018	0.015	0.001	-0.005	-0.029
Regional Center	-0.042	-0.040	-0.006	0.011	0.076
Number of Mental Health Disorders with less than 12 months	0.033	0.030	0.004	-0.010	-0.057
Number of Mental Health Disorders with more than 12 months	0.008	0.008	0.001	-0.002	-0.015
Has Physical Illness	0.005	0.004	0.000	-0.002	-0.008

After substituting the point estimates from FIML into the joint probabilities of five occupational categories, marginal effects of the number of mental health disorders are calculated by taking derivatives and marginal effects of other variables are evaluated by equation (4.5.1). As shown in Table 4.12, males are more likely to be employed in higher skill occupations than females, as are older people. Higher education increases the possibilities of entering the professional occupational category by 55 percentage points. People living in rural area are less likely to be employed as associate professionals or professionals, which may be due to less professional jobs available in such area. Short term mental illness reduces the chances of entering the higher occupational levels. Meanwhile, an individual with long term disorders or physical problems are no more or less likely to enter any skill categories, because marginal effects of the number of mental health disorders with more than 12 months and physical condition are all quite small.

The 50,000 draws from the Bayesian method can form the distributions of the model parameters, while mean values of such draws are used for the point estimates of the Bayesian method in Table 4.11. After substituting each draw into equations (4.5.3)-(4.5.6), marginal effects can be obtained in the same way we get those from FIML estimates. Thus, 50,000 marginal effects will be available for each parameter, then they can be used to construct the distribution of the marginal effect for that parameter. The mean value of 50,000 marginal effects is used as the point estimate of marginal effect for each parameter and a 95% credible interval is shown as an interval estimation in Table 4.13.

Figure 4.4: *Distributions of the Coefficients of*

Number of Mental Health Disorders with less than 12 months



The effects of mental illness are what we are mainly interested in. The mean marginal effects of the number of mental disorders with more than 12 months are all less than 1 percentage points. Therefore, having a long term mental disorder does not lead to a sizeable increase or decline in the possibility of joining any of the occupational categories. This

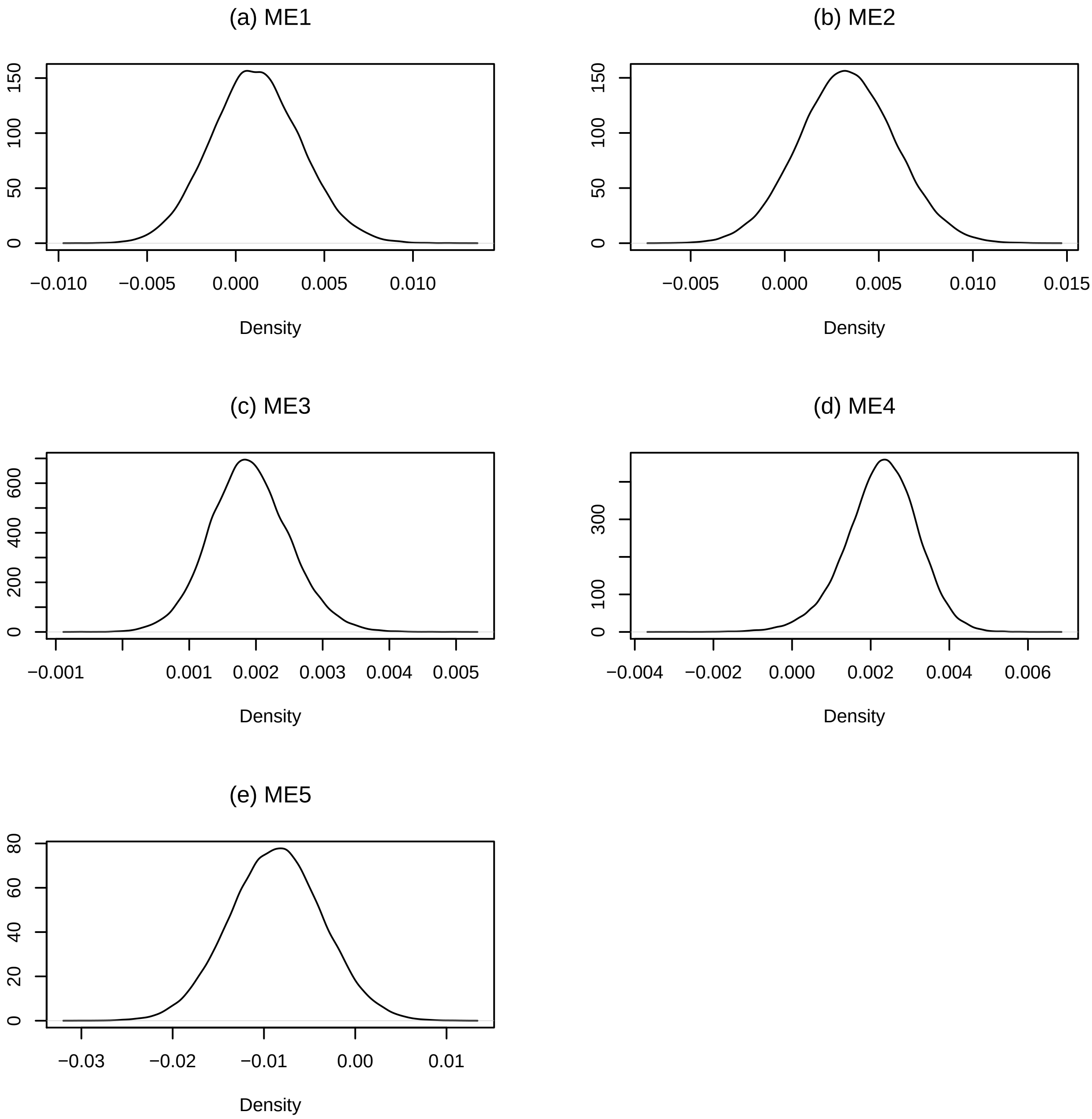
Table 4.13: *Marginal Effect Using the MCMC Method*

(Mean and 95% Credible Interval)

Variable	Elementary Skill	Intermediate Skill	Advanced Skill	Associate Professionals	Professionals
Male	-0.043(-0.051,-0.034)	-0.038(-0.047,-0.029)	-0.005 (-0.008 , -0.001)	0.011 (0.007 , 0.016)	0.074 (0.055 , 0.093)
Age 25-44	-0.059(-0.073,-0.046)	-0.056(-0.068,-0.044)	-0.009 (-0.013 , -0.006)	0.011 (0.004 , 0.019)	0.113 (0.091 , 0.136)
Age 45-64	-0.063(-0.076,-0.052)	-0.068(-0.080,-0.056)	-0.015 (-0.019 , -0.011)	0.006 (0.000 , 0.014)	0.139 (0.117 , 0.162)
Secondary School	-0.032(-0.043,-0.022)	-0.033(-0.044,-0.022)	-0.007 (-0.010 , -0.004)	0.004 (-0.001 , 0.009)	0.068 (0.046 , 0.090)
Higher Education	-0.184(-0.205,-0.164)	-0.252(-0.264,-0.240)	-0.086 (-0.093 , -0.077)	-0.049 (-0.066 , -0.029)	0.570 (0.546 , 0.595)
Vocational Education	-0.103(-0.117,-0.091)	-0.137(-0.151,-0.125)	-0.040 (-0.046 , -0.033)	-0.008 (-0.019 , 0.004)	0.288 (0.261 , 0.314)
Rural Area	0.012(0.001, 0.025)	0.013(0.002, 0.025)	0.003 (0.000 , 0.006)	-0.001 (-0.005 , 0.004)	-0.028 (-0.052 , -0.004)
Regional Center	-0.040(-0.050,-0.031)	-0.043(-0.054,-0.033)	-0.009 (-0.013 , -0.006)	0.005 (0.000 , 0.010)	0.088 (0.067 , 0.110)
Number of Mental Health Disorders with less than 12 months	0.017(-0.009, 0.044)	0.022(-0.004, 0.047)	0.006 (0.001 , 0.012)	0.002 (-0.007 , 0.011)	-0.048 (-0.099 , 0.006)
Number of Mental Health Disorders with more than 12 months	0.001(-0.004, 0.006)	0.003(-0.002, 0.008)	0.002 (0.001 , 0.003)	0.002 (0.000 , 0.004)	-0.009 (-0.019 , 0.001)
Has Physical Illness	0.003(-0.005, 0.011)	0.003(-0.005, 0.011)	0.000 (-0.002 , 0.002)	-0.001 (-0.004 , 0.002)	-0.006 (-0.022 , 0.011)

Figure 4.5: *Distributions of the Marginal Effects of*

Number of Mental Health Disorders with less than 12 months



ME1: Marginal effect on $Pr(y_{i2} = 1|y_{i1} = 1)$
ME2: Marginal effect on $Pr(y_{i2} = 2|y_{i1} = 1)$
ME3: Marginal effect on $Pr(y_{i2} = 3|y_{i1} = 1)$
ME4: Marginal effect on $Pr(y_{i2} = 4|y_{i1} = 1)$
ME5: Marginal effect on $Pr(y_{i2} = 5|y_{i1} = 1)$

finding is quite consistent with the results shown by FIML. The mean marginal effects of the five categories are 0.017, 0.022, 0.006, 0.002 and -0.048 respectively, which suggests that short term mental illnesses slightly increase the possibility of being employed in lower skill occupations and decrease the chances to be in the professional category. However, the 95% credible intervals reveal that effects of mental disorders with less than 12 months are not quite significant except the one for the advanced skill category. Figure 4.4 indicates the distributions of coefficients of mental disorders with less than 12 months in both equations, and Figure 4.5 contains five marginal effects densities. The coefficient densities looks quite close to normally distributed with mean -0.360 and -0.061. All marginal effect distributions tend to have very slight skewness.

4.6 Conclusion

This chapter has provided a Bayesian approach for a specific model with binary selection and ordered outcome observations. In this approach, a reparameterization is introduced to improve algorithm convergence rate, while other computational techniques are combined with reparameterization to refine the Gibbs sampler. In particular, special priors are set in advance to obtain conjugate conditional posteriors. After that, a numerical study is designed to illustrate Gibbs sampling results including MCMC convergence diagnostic analysis.

The Monte Carlo study compares the Bayesian approach with three other estimation methods including FIML, the two-step method and likelihood-based two-step method. It also discusses how the four estimators react under the effect of exclusion restrictions. The study shows that exclusion restrictions are not necessary in obtaining valid estimates of this particular model, but these restrictions do reduce the difficulty in estimation. Although the results do not show that the MCMC method is better than FIML in this case, it can be treated as an alternative method to FIML. On the other hand, it gives some benefits, such as Bayesian credible intervals which can be used for interval estimation. Besides that, it provides distributions of all parameters or other quantities of interest like marginal effects, so potentially provides more information from estimation results. The results show that the two-step method can give estimates for each sample, but has some bias and inefficiency when error correlation is strong. The likelihood-based two-step method works as accurately and efficiently as FIML. Moreover, it has less convergence problems than FIML.

In the application section, the Bayesian method is applied to data on mental health and labor market from a nationwide survey in Australia. Within the spirit of posterior distributions, it is not surprising that Bayesian methods can provide some efficient way to estimate parameters as well as marginal effects. In particular, the Bayesian method shows a more complete picture of the distributions of quantities of interest like marginal effects, compared to FIML.

Chapter 5

A Bayesian Approach to Estimating Double-hurdle Models

5.1 Introduction

In models of labor market outcomes, individuals must pass two separate hurdles before their labour income or occupational level is observed. Specifically, individuals have to choose to participate in the market before they can be employed, and then they have to be successful in finding employment. Generally, models of labour income are of most interest. However, in some cases other variables of interest might be the focus of modeling. Another example with two hurdles is about medical trials. In the first hurdle, people make their decision to apply to participate in a trial. They are selected after some pre-testing of applicant in the second hurdle. Then, the selected persons take the trial, with outcomes being some measure of the effect of the trials which is perhaps a binary outcome, or a continuous measure, or an ordered outcome like severity.

The two-stage sample selection motivates the use of the term "double-hurdle", which is first proposed by Cragg (1971). Some double-hurdle models involve two simultaneous hurdles. In other words, one hurdle does not influence the other. For instance, Jones (1989) has suggested a trivariate model of cigarette which is consumed by both current smokers and ex-smokers. He believes it is not possible to estimate the full trivariate model. What we are interested in are two hurdles in which one can result in non-random selection of the other. Although the model specification of interest is different from Jones', our estimation techniques for trivariate models can still be applied to those similar situations.

Since selection equations have discrete outcomes, auxiliary information evaluated by latent variables is used to construct the double-hurdle model, and also for facilitating estimation of the trivariate model, with two stages of sample selection and a third equation that could be continuous, binary or ordered. In this chapter, we are particularly interested in a double-hurdle model where a third equation has ordered outcomes. This double-hurdle model can be viewed as an extension of the model in the previous chapter by adding one extra stage of sample selection.

Classical MLE method is preferred by a lot of empirical researchers. However, convergence problems always exist even with one stage of sample selection, as has been discussed in the previous chapter. Multiple equations and non-linear forms may result in more serious convergence problems. For this double-hurdle model, the full likelihood function can be obtained through the normality assumption of error terms, when there are correlations between the three equations. Although computation of cumulative trivariate normal distributions is available nowadays, numerical optimization can still fail if the process is too complicated.

Therefore, we apply a Bayesian method to estimate the parameters of this trivariate model. Two Gibbs samplers are proposed in this chapter. One is mainly based on standard Bayesian inference to get posterior distributions. And the other develops the idea of reparameterization, so a Dirichlet proposal density can be applied to estimate threshold parameters, while two coefficients of error correlation can be drawn from standard distributions.

A Monte Carlo study is performed to compare FIML and the two Bayesian samplers. FIML fails to offer valid estimates more than half the time, although in the remaining cases, estimates are quite accurate. Meanwhile, both Gibbs samplers can give estimates for each sample and their overall performance is quite accurate and efficient. Moreover, the sampler using reparameterization manages to give more accurate estimation on the coefficients of error correlation than the standard sampler, especially when there is a strong correlation between the error terms. Simulation results also show that reparameterization can greatly speed up the convergence rate for several parameters, although the convergence rate is still poor in estimating one particular coefficient of correlation.

Later in this chapter, the three estimators are applied on an empirical study about mental health and its effects on labor market outcomes such as participation, employment and occupational skill categories. In the previous chapter, we only discussed the relationship between employment and occupational skill categories. Sample selection bias could arise without considering the impact of participation, as people must decide to participate in the job market before they can be employed. The empirical data and the double-hurdle model have been applied by Cornwell et al. (2009), but they use a two-step method twice to estimate the model. As we have discussed in the previous chapter, the two-step method is not consistent when the main equation is non-linear. That is why we are providing a more accurate and efficient estimation methodology like the Bayesian method to estimate the double-hurdle model.

This chapter is structured as follows. The double-hurdle model is built with latent variables in Section 5.2. Section 5.3 presents one Gibbs sampler derived from a standard way, while Section 5.4 analyzes another sampler constructed with the idea of reparameterization. After the full information likelihood estimator is derived, Section 5.5 presents a brief Monte Carlo study to compare FIML and the two Bayesian samplers. Section 5.6 applies those methods on an empirical study of the labor market. The final section concludes.

5.2 The Model

This model can be treated as an extension of the model specification in Chapter 4 with an additional hurdle of sample selection. The first equation is presented by a probit equation,

$$\begin{cases} z_{i1} = x'_{i1}\beta_1 + \epsilon_{i1} & 1 \le i \le n \\ y_{i1} = I(z_{i1} > 0). \end{cases} \tag{5.2.1}$$

The indicator function gives 1 as the outcome when the individual i participates in the labor market and 0 otherwise.

The second hurdle of sample selection is also decided by a probit equation but conditional on the outcome of the first equation,

$$\begin{cases} z_{i2} = x'_{i2}\beta_2 + \epsilon_{i2} & 1 \le i \le n \\ y_{i2} = I(z_{i2} > 0) \times y_{i1}. \end{cases} \tag{5.2.2}$$

When an individual does not go through the first hurdle, his or her status is missing and given a zero value. Otherwise, the individual is selected with $y_{i2} = 1$ when $z_{i2} > 0$, and with $y_{i2} = 0$ when $z_{i2} < 0$.

Non-random selection also arises when the third level can be observed after the two hurdles of sample selection. The third equation has ordered outcomes conditional on the output of the second equation,

$$\begin{cases} z_{i3} = x'_{i3}\beta_3 + \epsilon_{i3} & 1 \le i \le n \\ y_{i3} = j \times y_{i2} & \text{if } \gamma_{j-1} \le z_{i3} \le \gamma_j \text{ and } 1 \le j \le J. \end{cases} \tag{5.2.3}$$

The ordered outcomes can only be observed when an individual passes the second hurdle, and are missing otherwise. The threshold parameters $\left\{\gamma_0, \gamma_1, \gamma_2, \cdots, \gamma_J\right\}$ divide the latent variable z_{i3} into categories, with $\gamma_0 = -\infty$, $\gamma_1 = 0$, and $\gamma_J = +\infty$ for identification. The largest threshold parameter γ_{J-1} will be used for reparameterization, so it is separated from the other threshold parameters $\gamma = (\gamma_2, \cdots, \gamma_{J-2})'$.

The properties of probit models and ordered probit models determine that the variance of error term in each equation must be one for identification reasons. In addition, the errors of the three stages may be correlated. In a labour market example, unobservables may drive decision to participate and likelihood of finding employment, or people targeting professional jobs may be less likely to be employed due to small job supply, a correlation which is represented in error terms. Thus, we assume $(\epsilon_{i1}, \epsilon_{i2}, \epsilon_{i3}) \sim N(0, \Sigma)$ and

$$\Sigma = \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{pmatrix}.$$

Exclusion restrictions can be imposed in the explanatory variables x_{i1}, x_{i2} and x_{i3}, by allowing at least one regressor in each equation to be different from the regressors in other equations. However, exclusion restrictions are often not available in specific empirical specifications. As was discussed in Chapter 4, the lack of exclusion restrictions may increase the difficulty of estimation for all methods. But exclusion restrictions are not 100% necessary for model estimation.

5.3 A Standard Gibbs Sampler (Sampler A)

5.3.1 The Joint Posterior with Latent Variables and Parameters

Bayesian analysis in this chapter involves the joint posterior distributions of parameters and all latent variables in the model. Let $\theta = (\beta_1', \beta_2', \beta_3', \gamma', \gamma_{J-1}, \rho_1, \rho_2, \rho_3)'$ and $Z = \{(z_{i1}, z_{i2}, z_{i3}) : i = 1, 2, \cdots, n\}$. Given the prior $p(\theta)$, the joint posterior function follows

$$p(\theta, Z|Y) \propto p(\theta)p(Z|\theta)L(Y|\theta, Z),$$

where

$$p(Z|\theta) = \prod_{i=1}^{n} p(Z_i|\theta),$$

$$L(Y|\theta, Z) = \prod_{i=1}^{n} L(y_{i1}|\theta, z_{i1}, z_{i2}, z_{i3})L(y_{i2}|y_{i1}, \theta, z_{i1}, z_{i2}, z_{i3})L(y_{i3}|y_{i1}, y_{i2}, \theta, z_{i1}, z_{i2}, z_{i3})$$

for this double-hurdle model.

From the normality assumption of error terms, we can get

$$p(Z_i|\theta) = \phi_3\left[Z_i; \mu_i, \Sigma\right],$$

where ϕ_3 denotes the probability density function of a multivariate normal distribution of a three-dimensional random vector Z_i with mean μ_i and covariance matrix Σ by letting

$$Z_i = \begin{pmatrix} z_{i1} \\ z_{i2} \\ z_{i3} \end{pmatrix}, \ \mu_i = \begin{pmatrix} x'_{i1}\beta_1 \\ x'_{i2}\beta_2 \\ x'_{i3}\beta_3 \end{pmatrix}.$$

From equation (5.2.1), we obtain

$$L(y_{i1}|\theta, z_{i1}, z_{i2}, z_{i3}) = I(z_{i1} > 0)I(y_{i1} = 1) + I(z_{i1} \leq 0)I(y_{i1} = 0).$$

Since $y_{i2} = I(z_{i2} > 0) \times y_{i1}$, we get

$$\begin{aligned} L(y_{i2}|y_{i1} = 1, \theta, z_{i1}, z_{i2}, z_{i3}) &= I(z_{i2} > 0)I(y_{i2} = 1) + I(z_{i2} \leq 0)I(y_{i2} = 0) \\ L(y_{i2}|y_{i1} = 0, \theta, z_{i1}, z_{i2}, z_{i3}) &= I(y_{i2} = 0). \end{aligned}$$

Since $y_{i3} = j \times y_{i2} = j \times I(z_{i2} > 0) \times y_{i1}$, the conditional likelihood function becomes

$$\begin{aligned} L(y_{i3}|y_{i1} = 1, y_{i2} = 1, \theta, z_{i1}, z_{i2}, z_{i3}) &= \Sigma_{j=1}^{J} I(y_{i3} = j)I(\gamma_{j-1} < z_{i3} < \gamma_j) \\ L(y_{i3}|y_{i1}, y_{i2} = 0, \theta, z_{i1}, z_{i2}, z_{i3}) &= I(y_{i3} = 0). \end{aligned}$$

Therefore, the joint probability function of the parameters and latent variables can be fully displayed as follows:

$$\begin{aligned} &p(\theta, Z|Y) \propto \ p(\theta)\prod_{i=1}^{n}\phi_3[Z_i; \mu_i, \Sigma] \times \\ &\{I(z_{i1} \leq 0)I(y_{i1} = 0)I(y_{i2} = 0)I(y_{i3} = 0) + \\ &I(z_{i1} > 0)I(y_{i1} = 1)I(z_{i2} \leq 0)I(y_{i2} = 0)I(y_{i3} = 0) + \\ &I(z_{i1} > 0)I(y_{i1} = 1)I(z_{i2} > 0)I(y_{i2} = 1)\Sigma_{j=1}^{J} I(y_{i3} = j)I(\gamma_{j-1} < z_{i3} < \gamma_j)\Big\} \end{aligned} \tag{5.3.1}$$

5.3.2 Conditional Posteriors of Latent Variables

We start with sampling the conditional posterior of latent variable Z_i from

$$\begin{aligned}
&p(z_{i1}, z_{i2}, z_{i3}|\theta, Y_i) \propto \phi_3[Z_i; \mu_i, \Sigma]\times \\
&[I(z_{i1} \leq 0)I(y_{i1} = 0)I(y_{i2} = 0)I(y_{i3} = 0)+ \\
&I(z_{i1} > 0)I(y_{i1} = 1)I(z_{i2} \leq 0)I(y_{i2} = 0)I(y_{i3} = 0)+ \\
&I(z_{i1} > 0)I(y_{i1} = 1)I(z_{i2} > 0)I(y_{i2} = 1)\Sigma_{j=1}^{J} I(y_{i3} = j)I(\gamma_{j-1} < z_{i3} < \gamma_j)\big],
\end{aligned} \tag{5.3.2}$$

which is a truncated multivariate normal distribution. It involves samples from a trivariate normal distribution subject to linear inequality restrictions,

$$Z_i \sim N(\mu_i, \Sigma), A \leq Z_i \leq B \tag{5.3.3}$$

where

$$A = \begin{pmatrix} -\infty \\ -\infty \\ -\infty \end{pmatrix}, B = \begin{pmatrix} 0 \\ +\infty \\ +\infty \end{pmatrix} \quad \text{if} \begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix};$$

$$A = \begin{pmatrix} 0 \\ -\infty \\ -\infty \end{pmatrix}, B = \begin{pmatrix} +\infty \\ 0 \\ +\infty \end{pmatrix} \quad \text{if} \begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix};$$

$$A = \begin{pmatrix} 0 \\ 0 \\ \gamma_{j-1} \end{pmatrix}, B = \begin{pmatrix} +\infty \\ +\infty \\ \gamma_j \end{pmatrix} \quad \text{if} \begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ j \end{pmatrix}.$$

Geweke (1991) states that the distribution of each element of Z_i, conditional on all of the other elements of Z_i, is still truncated normal. Thus, a Gibbs sampling process applied by Hajivassiliou & McFadden (1998) will be used here to sample truncated multivariate normal distribution. The algorithm of general multivariate cases has been discussed by Robert (1995). More specifically, the conditional posteriors of the three components can be derived from the following distributions.

The first component z_{i1} conditional on z_{i2} and z_{i3} follows,

$$z_{i1}|z_{i2}, z_{i3} \sim \begin{cases} TN(\mu_{zi1}, \sigma^2_{zi1})|_{(0,+\infty)}, & \text{if } y_{i1} = 1 \\ TN(\mu_{zi1}, \sigma^2_{zi1})|_{(-\infty,0]}, & \text{if } y_{i1} = 0 \end{cases}$$

which is a truncated univariate normal distribution with

$$\mu_{zi1} = x'_{i1}\beta_1 + \frac{\rho_1 - \rho_2\rho_3}{1-\rho_3^2}(z_{i2} - x'_{i2}\beta_2) + \frac{\rho_2 - \rho_1\rho_3}{1-\rho_3^2}(z_{i3} - x'_{i3}\beta_3)$$

and

$$\sigma^2_{zi1} = \frac{1-\rho_1^2-\rho_2^2-\rho_3^2+2\rho_1\rho_2\rho_3}{1-\rho_3^2}.$$

The second component is represented by

$$z_{i2}|z_{i1}, z_{i3} \sim \begin{cases} TN(\mu_{zi2}, \sigma^2_{zi2})|_{(0,+\infty)}, & \text{if } y_{i1} = 1 \text{ and } y_{i2} = 1 \\ TN(\mu_{zi2}, \sigma^2_{zi2})|_{(-\infty,0]}, & \text{if } y_{i1} = 1 \text{ and } y_{i2} = 0 \\ N(\mu_{zi2}, \sigma^2_{zi2}), & \text{if } y_{i1} = 0 \end{cases}$$

which is a normal distribution when $y_{i1} = 0$ and a truncated univariate normal distribution otherwise with

$$\mu_{zi2} = x'_{i2}\beta_2 + \frac{\rho_1 - \rho_2\rho_3}{1-\rho_2^2}(z_{i1} - x'_{i1}\beta_1) + \frac{\rho_3 - \rho_1\rho_2}{1-\rho_2^2}(z_{i3} - x'_{i3}\beta_3)$$

and

$$\sigma^2_{zi2} = \frac{1-\rho_1^2-\rho_2^2-\rho_3^2+2\rho_1\rho_2\rho_3}{1-\rho_2^2}.$$

The third component follows

$$z_{i3}|z_{i1}, z_{i2} \sim \begin{cases} TN(\mu_{zi3}, \sigma^2_{zi3})|_{(\gamma_{j-1},\gamma_j)}, & \text{if } y_{i3} = j \text{ and } 1 \le j \le J \\ N(\mu_{zi3}, \sigma^2_{zi3}), & \text{if } y_{i3} = 0 \end{cases}$$

which is a truncated univariate normal distribution when $y_{i3} \neq 0$ and a normal density otherwise with

$$\mu_{zi3} = x'_{i3}\beta_3 + \frac{\rho_2 - \rho_1\rho_3}{1-\rho_1^2}(z_{i1} - x'_{i1}\beta_1) + \frac{\rho_3 - \rho_1\rho_2}{1-\rho_1^2}(z_{i2} - x'_{i2}\beta_2)$$

and

$$\sigma_{zi3}^2 = \frac{1-\rho_1^2-\rho_2^2-\rho_3^2+2\rho_1\rho_2\rho_3}{1-\rho_1^2}.$$

One can draw the three components repeatedly until their paths become stationary. Any Z_i vector in stationary paths can be considered as a random draw from the density given in Equation (5.3.2). Then, such a random draw can join the Gibbs sampler of the joint posterior distribution in Equation (5.3.1). Alternatively, the Gibbs sampling of each conditional univariate distribution for latent variables can become part of whole Gibbs algorithm of the joint posterior distribution in the Equation (5.3.1). One draw for each component, together with one draw for each of the other parameters, is treated as a chain of the whole algorithm. Once the chains become stationary, the random draws can be recorded to get the Bayesian estimates. In this chapter, the latter procedure is used to get the final results.

5.3.3 Conditional Posteriors of Parameters

Given the prior $p(\beta) = \phi_k(\beta|\beta_0, B_0^{-1})$, let $\beta = (\beta_1', \beta_2', \beta_3')'$ and

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

We can directly sample β from the conditional posterior function which is a Gaussian density,

$$\beta|Z,\Sigma \sim N_k(\hat{\beta}, B^{-1}), \tag{5.3.4}$$

where $\hat{\beta} = B^{-1}(B_0\beta_0 + \Sigma_{i=1}^n X_i'\Sigma^{-1}Z_i)$ and $B = B_0 + \Sigma_{i=1}^n X_i'\Sigma^{-1}X_i$.

The conditional density of threshold parameters is given by

$$\prod_{i=1}^{n}\Big[I(y_{i3}=j)I(\gamma_{j-1}<z_{i3}<\gamma_j)+I(y_{i3}=j+1)I(\gamma_j<z_{i3}<\gamma_{j+1})\Big], \tag{5.3.5}$$

which is a uniform distribution on the interval

$$\Big[\max\Big\{\max\{z_{i3}:y_{i3}=j\},\gamma_{j-1}\Big\},\min\Big\{\min\{z_{i3}:y_{i3}=j+1\},\gamma_{j+1}\Big\}\Big].$$

Albert & Chib (1993) provide an example of this for a single equation with ordered outcomes. In the next subsection, it is shown that sampling from uniform distribution is not quite efficient, as the convergence rate is quite poor in sampling the threshold parameters.

Finally, given prior $p(\rho)=\phi_3(\rho|\rho_0,C_0^{-1})$, the conditional posterior density of $(\rho_1,\rho_2,\rho_3)'$ is given by

$$p(\rho)\prod_{i=1}^{n}\phi_3[Z_i;\mu_i,\Sigma].$$

And $\rho=(\rho_1,\rho_2,\rho_3)$ should be constrained as $1-\rho_1^2-\rho_2^2-\rho_3^3>0$ to support $|\Sigma|>0$. Since this conditional distribution has no closed form with respect to the error correlation parameters, Metropolis-Hastings (MH) algorithm is utilized here to sample these three parameters.

5.3.4 A Simulation Example with MCMC Convergence Diagnostics

The standard Gibbs sampler is illustrated by a simulation study on one sample with MCMC convergence diagnostics. The sample size is set to be 1,000 and the true values of parameters are set as $\beta_1=(\beta_{11},\ \beta_{12})'=(0.6,\ -1.2)'$, $\beta_2=(\beta_{21},\ \beta_{22},\ \beta_{23})'=(1,\ -1.5,\ -1)'$, $\beta_3=(\beta_{31},\ \beta_{32})'=(-0.4,\ 1.5)'$, $\gamma=(\gamma_2,\ \gamma_3)'=(0.8,\ 1.6)'$ and $\rho=(\rho_1,\ \rho_2,\ \rho_3)'=(0.25,\ 0.25,\ 0.5)'$. x_{i1} and x_{i3} are designed to be 2×1 vectors, while x_{i2} is a 3×1 vector. As has been shown in Section 4.4 of Chapter 4, it is more difficult to estimate a model with all methods when independent variables are the same for both equations. So exclusion restrictions are imposed here to make estimation easier. Since each equation has an intercept, the first components of vectors x_{i1}, x_{i2} and x_{i3} are set to be one. The second components of x_{i1} and x_{i2} are randomly generated from independent standard normal distributions. Meanwhile, the third component of x_{i2} and the second component of x_{i3} are independently generated from Bernoulli distributions with success probability 0.7. Then error terms are generated from standard trivariate normal distributions with zero mean and covariance matrix (1 0.25 0.25, 0.25 1 0.5, 0.25 0.5 1). Latent variables are then calculated and used to obtain observations y_{i1}, y_{i2} and y_{i3} based on Equations (5.2.1)-(5.2.3). More than 1/3 of the data is censored at the first level of sample selection, and another 1/3 of the data is censored after the second hurdle. After that, the uncensored data is divided into four categories with

$$y_{i3} = \begin{cases} 1 \text{ if } z_{i3} < 0 \\ 2 \text{ if } 0 < z_{i3} < 0.8 \\ 3 \text{ if } 0.8 < z_{i3} < 1.6 \\ 4 \text{ if } 1.6 < z_{i3}. \end{cases}$$

Scalars for the priors of the Gibbs sampler are set as $B_0^{-1} = 1000 I_7$ and $C_0 = I_3$, while β_0 and ρ_0 are zero vectors. Tuning parameters for MH algorithm are set as (0.2, 0.2, 0.2) to make sure that the acceptance rates of the algorithm are around 30%. One cycle of the MCMC algorithm is completed by simulating each conditional distribution one by one. This process is first repeated 2,000 times and these are discarded as the burn-in period, and then continuously repeated 10,000 times to form the full samples of the posterior distributions.

Those 10,000 iterations of each parameter are plotted in Figure 5.1, while the corresponding autocorrelation functions are presented in Figure 5.2. Figure 5.1 shows the sample paths of β parameters have no obvious pattern. However, it indicates a pattern of up-and-down swings for the sample paths of γ and ρ parameters, although there is no clear trend in the graph. Figure 5.2 reveals that paths of autocorrelation function values for β parameters decay quite quickly, but those for γ and ρ decay very slowly. In other words, the convergence rate of γ and ρ is quite slow according to this sampler.

Table 5.1 displays summary statistics of the 10,000 iterations: mean, standard deviation, 95% credible interval and simulation inefficient factor (SIF). Estimated mean value of β_{21}

Figure 5.1: *Sample Paths for Sampler A*

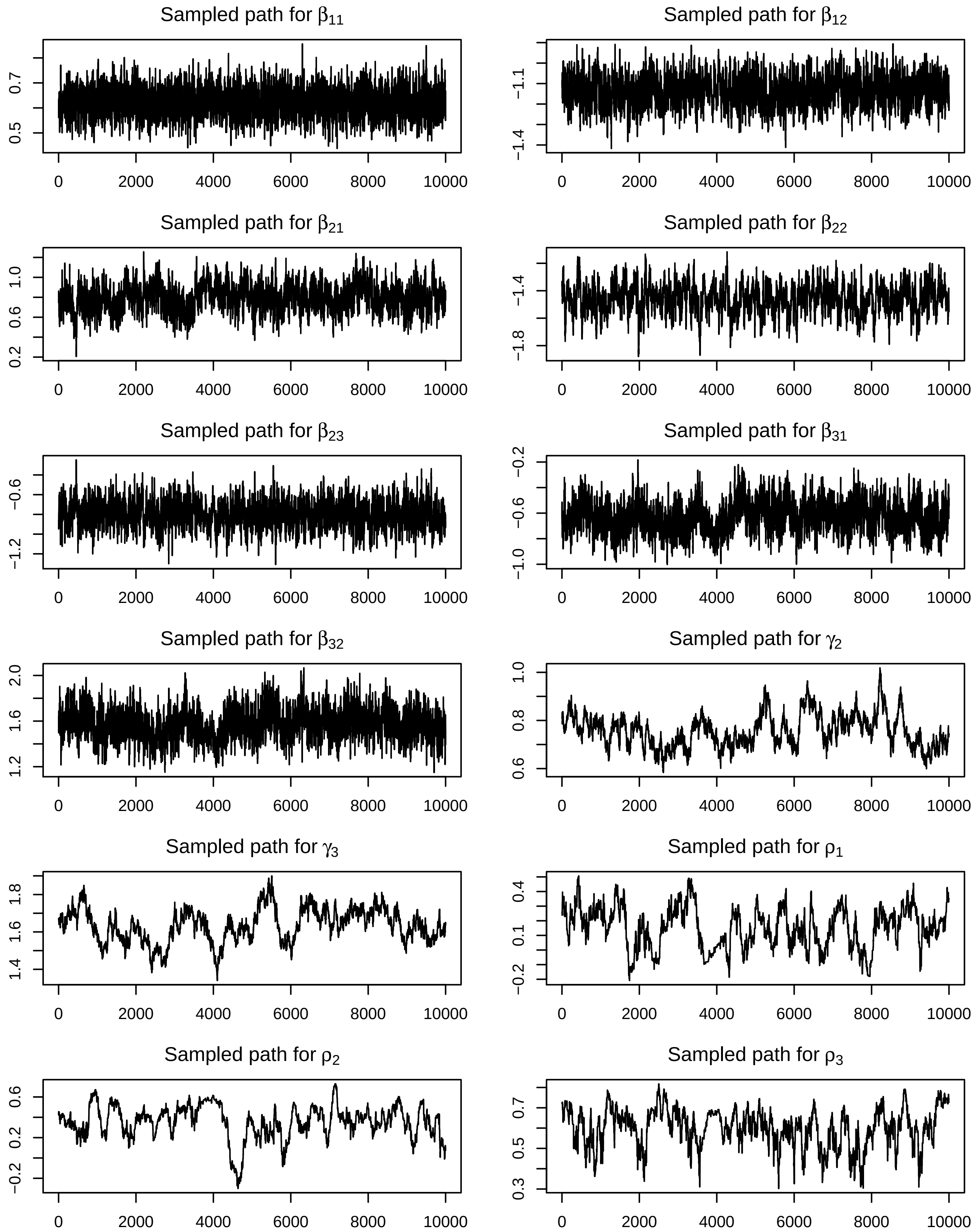


Figure 5.2: *Sample Autocorrelation Functions for Sampler A*

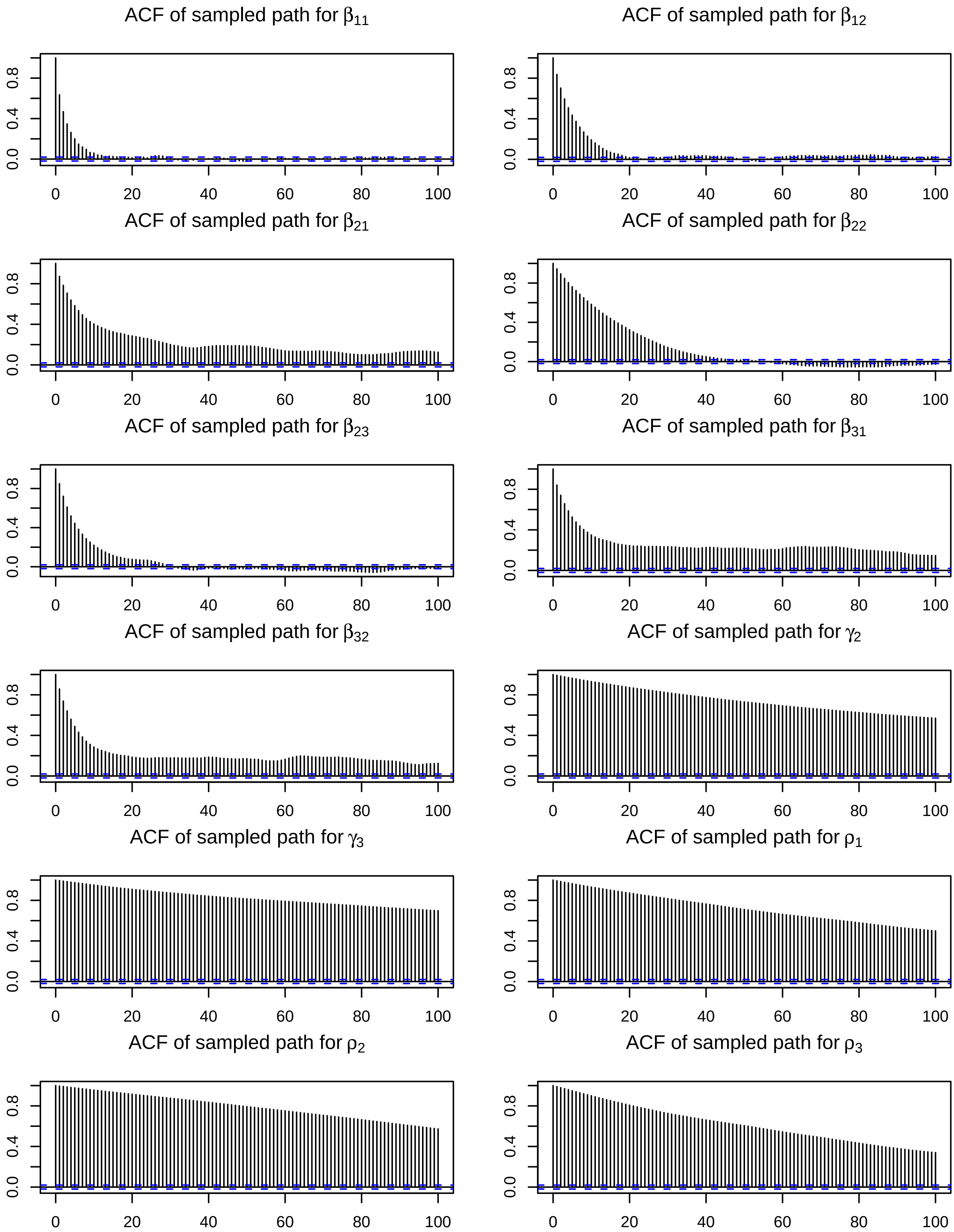


Table 5.1: *One Sample Estimation Results (Sampler A)*

		Posterior Density			
	True	Mean	St. Dev.	95% Credible Interval	SIF
β_{11}	0.600	0.623	0.053	(0.562 , 0.684)	7
β_{12}	-1.200	-1.135	0.071	(-1.217 , -1.053)	13
β_{21}	1.000	0.793	0.131	(0.640 , 0.947)	39
β_{22}	-1.500	-1.461	0.106	(-1.586 , -1.339)	24
β_{23}	-1.000	-0.798	0.133	(-0.950 , -0.645)	8
β_{31}	-0.400	-0.626	0.118	(-0.762 , -0.488)	46
β_{32}	1.500	1.572	0.132	(1.421 , 1.725)	39
γ_2	0.800	0.757	0.069	(0.678 , 0.837)	142
γ_3	1.600	1.635	0.090	(1.529 , 1.737)	159
ρ_1	0.250	0.152	0.140	(-0.026 , 0.315)	139
ρ_2	0.250	0.355	0.163	(0.184 , 0.538)	150
ρ_3	0.500	0.599	0.095	(0.478 , 0.701)	119

is 0.27 less than the true value, while β_{23} are overestimated by 0.22. At the same time, mean values of ρ_1 and ρ_2 are underestimated by about 0.1 from the true values. The standard deviations of those four parameters are also quite large, especially for ρ_1 and ρ_2 considering their magnitudes which are constrained in $(-1,1)$. The 95% credible interval for ρ_1 contains zero, which demonstrates this coefficient is not quite significant. The SIF values for β values are quite small, which illustrates that simulations of β converge quickly. Meanwhile, the SIF values for γ and ρ parameters are around 150, and the poor convergence is consistent with their autocorrelation functions in Figure 5.2.

5.4 A Sampler with Reparameterization (Sampler B)

5.4.1 Reparameterization

The techniques of reparameterization implemented in Chapter 4 provide a way to get conjugate conditional posterior distributions, which may accelerate the convergence of MCMC algorithms. They are based on an approach discussed by Li & Tobias (2006) who divide each equation of multivariate ordered probit models by the largest threshold parameter to form the new models. Then the new threshold parameters are drawn by the MH algorithm with a Dirichlet proposal density. Such techniques are used in our double-hurdle model, since the third equation is an ordered equation whose corresponding latent equation can be also divided by the largest threshold parameter. Let

$$\beta_3^* = \beta_3/\gamma_{J-1}, z_{i3}^* = z_{i3}/\gamma_{J-1}, \epsilon_{i3}^* = \epsilon_{i3}/\gamma_{J-1}.$$

The latent variables can be written as

$$\begin{cases} z_{i1} = x_{i1}'\beta_1 + \epsilon_{i1} \\ z_{i2} = x_{i2}'\beta_2 + \epsilon_{i2} \qquad 1 \le i \le n \\ z_{i3}^* = x_{i3}'\beta_3^* + \epsilon_{i3}^* \end{cases} \tag{5.4.1}$$

where

$$(\epsilon_{i1}, \epsilon_{i2}, \epsilon_{i3}^*) \sim N(0, \Sigma^*)$$

and

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

The model becomes

$$\begin{cases} y_{i1} = I(z_{i1} > 0) \\ y_{i2} = I(z_{i2} > 0) \times y_{i1} \\ y_{i3} = j \times y_{i2} \quad \text{if } \gamma_{j-1}^* \le z_{i3}^* \le \gamma_j^* \quad 1 \le j \le J, \end{cases} \tag{5.4.2}$$

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

Further, a new parameter which is identical to the determinant of covariance matrix,

$$\psi = |\Sigma^*| = (1 - \rho_1^2 - \rho_2^2 - \rho_3^2 + 2\rho_1\rho_2\rho_3)/\gamma_{J-1}^2,$$

is introduced to reparameterize Σ^*, generalizing McCulloch et al. (2000)'s approach in the two equation system. The other new variables are transferred from the old ones by:

$$\lambda_1 = \rho_1, \lambda_2 = \rho_2/\gamma_{J-1}, \lambda_3 = \rho_3/\gamma_{J-1}.$$

Thus, the new covariance matrix has the following form,

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

Let $\beta^* = (\beta_1', \beta_2', \beta_3^{*\prime})'$, $\theta^* = (\beta^{*\prime}, \gamma^{*\prime}, \psi, \lambda_1, \lambda_2, \lambda_3)'$,

$$Z_i^* = \begin{pmatrix} z_{i1} \\ z_{i2} \\ z_{i3}^* \end{pmatrix} \text{ and } \mu_i^* = \begin{pmatrix} x_{i1}'\beta_1 \\ x_{i2}'\beta_2 \\ x_{i3}'\beta_3^* \end{pmatrix}.$$

Given prior $p(\theta^*)$, the posterior distribution of the new latent variables and new model parameters follows,

$$\begin{aligned} & p(\theta^*, Z^*|Y) \propto \\ & p(\theta^*)\prod_{i=1}^{n} \phi_3(Z_i^*; \mu_i^*, \Sigma^*)\,[I(z_{i1} \le 0)I(y_{i1} = 0)I(y_{i2} = 0)I(y_{i3} = 0)+ \qquad (5.4.3) \\ & I(z_{i1} > 0)I(y_{i1} = 1)I(z_{i2} \le 0)I(y_{i2} = 0)I(y_{i3} = 0)+ \\ & I(z_{i1} > 0)I(y_{i1} = 1)I(z_{i2} > 0)I(y_{i2} = 1)\Sigma_{j=1}^{J} I(y_{i3} = j)I(\gamma_{j-1}^* < z_{i3}^* < \gamma_j^*)\big] \end{aligned}$$

where

$$\phi_3(Z_i^*; \mu_i^*, \Sigma^*) \propto \psi^{-\frac{1}{2}} exp\left\{-\frac{1}{2}(Z_i^* - \mu_i^*)'\Sigma^{*-1}(Z_i^* - \mu_i^*)\right\}.$$

5.4.2 Conditional Posteriors of Latent Variables

The conditional posteriors of new latent variables are derived in a similar way to the process in Section 5.3.2 from

$$\begin{aligned} & p(Z_i^*|\theta^*, Y_i) \propto \phi_3\left[Z_i^*; \mu_i^*, \Sigma^*\right]\times \\ & [I(z_{i1} \le 0)I(y_{i1} = 0)I(y_{i2} = 0)I(y_{i3} = 0)+ \\ & I(z_{i1} > 0)I(y_{i1} = 1)I(z_{i2} \le 0)I(y_{i2} = 0)I(y_{i3} = 0)+ \end{aligned}$$

$$I(z_{i1}>0)I(y_{i1}=1)I(z_{i2}>0)I(y_{i2}=1)\Sigma_{j=1}^{J}I(y_{i3}=j)I(\gamma_{j-1}^{*}<z_{i3}^{*}<\gamma_{j}^{*})\Big],$$

which is a truncated multivariate normal distribution.

In order to apply the algorithm for generating a truncated multivariate normal distribution proposed by Robert (1995), we must use

$$\Sigma^{*-1}=\frac{1}{\psi}\begin{pmatrix} \frac{\psi+\lambda_1^2\lambda_3^2+\lambda_2^2-2\lambda_1\lambda_2\lambda_3}{1-\lambda_1^2} & \frac{\lambda_2\lambda_3-\psi\lambda_1-\lambda_1\lambda_2^2-\lambda_1\lambda_3^2+\lambda_1^2\lambda_2\lambda_3}{1-\lambda_1^2} & \lambda_1\lambda_3-\lambda_2 \\ \frac{\lambda_2\lambda_3-\psi\lambda_1-\lambda_1\lambda_2^2-\lambda_1\lambda_3^2+\lambda_1^2\lambda_2\lambda_3}{1-\lambda_1^2} & \frac{\psi+\lambda_1^2\lambda_2^2+\lambda_3^2-2\lambda_1\lambda_2\lambda_3}{1-\lambda_1^2} & \lambda_1\lambda_2-\lambda_3 \\ \lambda_1\lambda_3-\lambda_2 & \lambda_1\lambda_2-\lambda_3 & 1-\lambda_1^2 \end{pmatrix}.$$

Following his algorithm, the conditional posterior of z_{i1} can be generated from

$$z_{i1}|z_{i2},z_{i3}^{*}\sim\begin{cases} TN(\mu_{zi1}^{*},\sigma_{zi1}^{*2})|_{(0,+\infty)}, & \text{if } y_{i1}=1 \\ TN(\mu_{zi1}^{*},\sigma_{zi1}^{*2})|_{(-\infty,0]}, & \text{if } y_{i1}=0 \end{cases}$$

which is a truncated univariate normal distribution with

$$\mu_{zi1}^{*}=x_{i1}'\beta_1+\begin{pmatrix}\lambda_1\\ \lambda_2\end{pmatrix}'\begin{pmatrix}1 & \lambda_3 \\ \lambda_3 & \frac{\psi+\lambda_2^2+\lambda_3^2-2\lambda_1\lambda_2\lambda_3}{1-\lambda_1^2}\end{pmatrix}^{-1}\begin{pmatrix}z_{i2}-x_{i2}'\beta_2\\ z_{i3}^{*}-x_{i3}'\beta_3^{*}\end{pmatrix}$$

and

$$\sigma_{zi1}^{*2}=1-\begin{pmatrix}\lambda_1\\ \lambda_2\end{pmatrix}'\begin{pmatrix}1 & \lambda_3 \\ \lambda_3 & \frac{\psi+\lambda_2^2+\lambda_3^2-2\lambda_1\lambda_2\lambda_3}{1-\lambda_1^2}\end{pmatrix}^{-1}\begin{pmatrix}\lambda_1\\ \lambda_2\end{pmatrix}.$$

The second component follows

$$z_{i2}|z_{i1},z_{i3}^{*}\sim\begin{cases} TN(\mu_{zi2}^{*},\sigma_{zi2}^{*2})|_{(0,+\infty)}, & \text{if } y_{i1}=1 \text{ and } y_{i2}=1 \\ TN(\mu_{zi2}^{*},\sigma_{zi2}^{*2})|_{(-\infty,0]}, & \text{if } y_{i1}=1 \text{ and } y_{i2}=0 \\ N(\mu_{zi2}^{*},\sigma_{zi2}^{*2}), & \text{if } y_{i1}=0 \end{cases}$$

which is a normal distribution when $y_{i1}=0$ and a truncated normal distribution otherwise with

$$\mu_{zi2}^{*}=x_{i2}'\beta_2+\begin{pmatrix}\lambda_1\\ \lambda_3\end{pmatrix}'\begin{pmatrix}1 & \lambda_2 \\ \lambda_2 & \frac{\psi+\lambda_2^2+\lambda_3^2-2\lambda_1\lambda_2\lambda_3}{1-\lambda_1^2}\end{pmatrix}^{-1}\begin{pmatrix}z_{i1}-x_{i1}'\beta_1\\ z_{i3}^{*}-x_{i3}'\beta_3^{*}\end{pmatrix}$$

and

$$\sigma_{zi2}^{*2} = 1 - \begin{pmatrix} \lambda_1 \\ \lambda_3 \end{pmatrix}' \begin{pmatrix} 1 & \lambda_2 \\ \lambda_2 & \frac{\psi+\lambda_2^2+\lambda_3^2-2\lambda_1\lambda_2\lambda_3}{1-\lambda_1^2} \end{pmatrix}^{-1} \begin{pmatrix} \lambda_1 \\ \lambda_3 \end{pmatrix}.$$

The third component is from

$$z_{i3}^*|z_{i1}, z_{i2} \sim \begin{cases} TN(\mu_{zi3}^*, \sigma_{zi3}^{*2})|_{(\gamma_{j-1}^*, \gamma_j^*)}, & \text{if } y_{i3} = j \\ N(\mu_{zi3}^*, \sigma_{zi3}^{*2}), & \text{if } y_{i3} = 0 \end{cases}$$

which is a truncated univariate normal distribution when $y_{i3} \neq 0$ and a normal density otherwise with

$$\mu_{zi3}^* = x_{i3}'\beta_3^* + \begin{pmatrix} \lambda_2 \\ \lambda_3 \end{pmatrix}' \begin{pmatrix} 1 & \lambda_1 \\ \lambda_1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} z_{i1} - x_{i1}'\beta_1 \\ z_{i2} - x_{i2}'\beta_2 \end{pmatrix}$$

and

$$\sigma_{zi3}^{*2} = \frac{\psi + \lambda_2^2 + \lambda_3^2 - 2\lambda_1\lambda_2\lambda_3}{1 - \lambda_1^2} - \begin{pmatrix} \lambda_2 \\ \lambda_3 \end{pmatrix}' \begin{pmatrix} 1 & \lambda_1 \\ \lambda_1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \lambda_2 \\ \lambda_3 \end{pmatrix}.$$

5.4.3 Conditional Posteriors of Parameters

Given the prior $p(\beta^*) = \phi_k(\beta^*|\beta_0, B_0^{-1})$, let $Z^* = \left\{Z_i^* : 1 \leq i \leq n\right\}$ and

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

β^* can be directly simulated from the conditional posterior function which is a Gaussian density,

$$\beta^*|Z^*, \Sigma^* \sim N_k(\hat{\beta}^*, B^{-1}), \tag{5.4.4}$$

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

Given prior $p(\lambda_1) \sim N(\lambda_{01}, C_1^{-1})$, we then sample λ_1 from

$$p(\lambda_1|Z^*, \beta^*, \lambda_2, \lambda_3, \psi) \propto p(\lambda_1) exp\left\{-\frac{1}{2}\Sigma_{i=1}^n (Z_i^* - \mu_i^*)'\Sigma^{*-1}(Z_i^* - \mu_i^*)\right\}. \tag{5.4.5}$$

Let $\mu_{i1} = z_{i1} - x_{i1}'\beta_1$, $\mu_{i2} = z_{i2} - x_{i2}'\beta_2$ and $\mu_{i3} = z_{i3}^* - x_{i3}'\beta_3^*$, so $(Z_i^* - \mu_i^*)'\Sigma^{*-1}(Z_i^* - \mu_i^*)$ can be written as,

$$\frac{1}{\psi}\Big[\mu_{i1}^2 \frac{\psi+\lambda_1^2\lambda_3^2+\lambda_2^2-2\lambda_1\lambda_2\lambda_3}{1-\lambda_1^2} + \mu_{i1}\mu_{i2}\frac{\lambda_2\lambda_3-\psi\lambda_1-\lambda_1\lambda_2^2-\lambda_1\lambda_3^2+\lambda_1^2\lambda_2\lambda_3}{1-\lambda_1^2} + \mu_{i1}\mu_{i3}(\lambda_1\lambda_3 - \lambda_2) +$$
$$\mu_{i1}\mu_{i2}\frac{\lambda_2\lambda_3-\psi\lambda_1-\lambda_1\lambda_2^2-\lambda_1\lambda_3^2+\lambda_1^2\lambda_2\lambda_3}{1-\lambda_1^2} + \mu_{i2}^2\frac{\psi+\lambda_1^2\lambda_2^2+\lambda_3^2-2\lambda_1\lambda_2\lambda_3}{1-\lambda_1^2} + \mu_{i2}\mu_{i3}(\lambda_1\lambda_2 - \lambda_3) +$$
$$\mu_{i1}\mu_{i3}(\lambda_1\lambda_3 - \lambda_2) + \mu_{i2}\mu_{i3}(\lambda_1\lambda_2 - \lambda_3) + \mu_{i3}^2(1 - \lambda_1^2)\Big].$$

This conditional posterior has no standard form in terms of λ_1. Thus, we apply MH algorithm to sample this parameter.

Once given prior $p(\lambda_2) \sim N(\lambda_{02}, C_2^{-1})$, the conditional posterior density of λ_2 is given by

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

which follows a Gaussian density

$$\lambda_2|Z^*, \beta^*, \lambda_1, \lambda_3, \psi \sim N(\mu_{\lambda_2}, \sigma_{\lambda_2}^2) \tag{5.4.6}$$

where

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

and

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

With prior $p(\lambda_3) \sim N(\lambda_{03}, C_3^{-1})$, the conditional posterior density of λ_3 follows

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

which is also a normal distribution

$$\lambda_3|Z^*, \beta^*, \lambda_1, \lambda_2, \psi \sim N(\mu_{\lambda_3}, \sigma_{\lambda_3}^2) \tag{5.4.7}$$

where

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

and

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

Given prior $p(\psi) \sim IG(\frac{n_0}{2}, \frac{D_0}{2})$, the conditional posterior of ψ is a conjugate posterior, as it also follows a Inverse-Gamma distribution:

$$\begin{aligned}
p(\psi|Z^*,\beta^*,\lambda) \propto & p(\psi)\prod_{i=1}^{n}\phi_3(Z_i^*;\mu_i^*,\Sigma^*) \\
\propto & (\frac{1}{\psi})^{\frac{n_0}{2}+1} exp(-\frac{D_0}{2\psi})(\frac{1}{\psi})^{\frac{n}{2}}\times \\
& exp\left\{-\frac{1}{2\psi}\sum_{i=1}^{n}\frac{\left[\mu_{i1}(\lambda_2-\lambda_1\lambda_3)+\mu_{i2}(\lambda_3-\lambda_1\lambda_2)-\mu_{i3}(1-\lambda_1^2)\right]^2}{1-\lambda_1^2}\right\}
\end{aligned}$$

This conditional posterior function of ψ is an Inverse-Gamma density

$$\psi|Z^*,\beta^*,\lambda \sim IG(\frac{n_1}{2},\frac{D}{2}) \tag{5.4.8}$$

where $n_1 = n_0 + n$ and

$$D = D_0 + \frac{1}{1-\lambda_1^2}\sum_{i=1}^{n}\left[\mu_{i1}(\lambda_2-\lambda_1\lambda_3)+\mu_{i2}(\lambda_3-\lambda_1\lambda_2)-\mu_{i3}(1-\lambda_1^2)\right]^2.$$

Finally, the posterior of threshold parameters $\left\{\gamma_j^*\right\}_{j=2}^{J-2}$ is calculated from its conditional posterior distribution marginalized over all z_{i2}^*:

$$p(\left\{\gamma_j^*\right\}_{j=2}^{J-2}|\beta^*,\Sigma^*,Z) \propto \prod_{i=1}^{n}\left\{\Phi\left[(\gamma^*_{y_{i3}} - \mu^*_{z_{i3}})/\sqrt{\sigma_{i3}^{*2}}\right] - \Phi\left[(\gamma^*_{y_{i3}-1} - \mu^*_{z_{i3}})/\sqrt{\sigma_{i3}^{*2}}\right]\right\} \tag{5.4.9}$$

where $\mu^*_{z_{i3}}$ and σ_{i3}^{*2} are respectively identical to the mean and the variance for sampling latent variable z_{i3}^*. Once again, a Dirichlet proposal density is applied in the MH algorithm to sample the threshold parameters.

5.4.4 A Simulation Example with MCMC Convergence Diagnostics

The sample simulated in Section 5.3.4 is directly used in this section and the model is estimated by Sampler B. Scalars of the priors for Sampler B are set as $B_0^{-1} = 1000I_7$, $\beta_0 = \mathbf{0}$, $\lambda_{01} = 0$, $\lambda_{02} = 0$, $\lambda_{03} = 0$, $C_1 = 1$, $C_2 = 1$, $C_3 = 1$, $n_0 = 2$ and $D_0 = 0.01$. The tuning parameter to sample λ_1 is set as 0.2 to guarantee a certain acceptance rate of MH algorithm. Each latent variable and parameter are simulated to draw one cycle of the MCMC algorithm. The burn-in period is still 2,000 iterations, while the following 10,000 iterations are recorded to calculate the estimates of the model.

The sampled paths of the 10,000 iterations are plotted in Figure 5.3 and the autocorrelation function values are shown in Figure 5.4, so visual inspection can help to check convergence. The sampled paths are all randomly distributed in Figure 5.3, except the one for ρ_1 which has some fluctuations although it still looks stationary. Moreover, sampled paths for γ_2, γ_3, ρ_2 and ρ_3 appear to be much more stationary than those in Figure 5.1. Figure 5.4 also illustrates that ρ_1 has the worst convergence rate as its autocorrelation function value decay slowly, while those of other parameters decay much faster.

Table 5.2: *One Sample Estimation Results (Sampler B)*

		Posterior Density			
	True	Mean	St. Dev.	95% Credible Interval	SIF
β_{11}	0.600	0.620	0.052	(0.559 , 0.678)	9
β_{12}	-1.200	-1.134	0.071	(-1.215 , -1.052)	14
β_{21}	1.000	0.804	0.132	(0.654 , 0.958)	55
β_{22}	-1.500	-1.446	0.105	(-1.568 , -1.325)	31
β_{23}	-1.000	-0.791	0.134	(-0.947 , -0.640)	18
β_{31}	-0.400	-0.608	0.124	(-0.752 , -0.463)	37
β_{32}	1.500	1.575	0.138	(1.420 , 1.734)	15
γ_2	0.800	0.755	0.081	(0.664 , 0.848)	29
γ_3	1.600	1.640	0.112	(1.508 , 1.768)	41
ρ_1	0.250	0.094	0.167	(-0.089 , 0.293)	161
ρ_2	0.250	0.298	0.149	(0.117 , 0.463)	81
ρ_3	0.500	0.608	0.091	(0.501 , 0.711)	59

Figure 5.3: *Sample Paths for Sampler B*

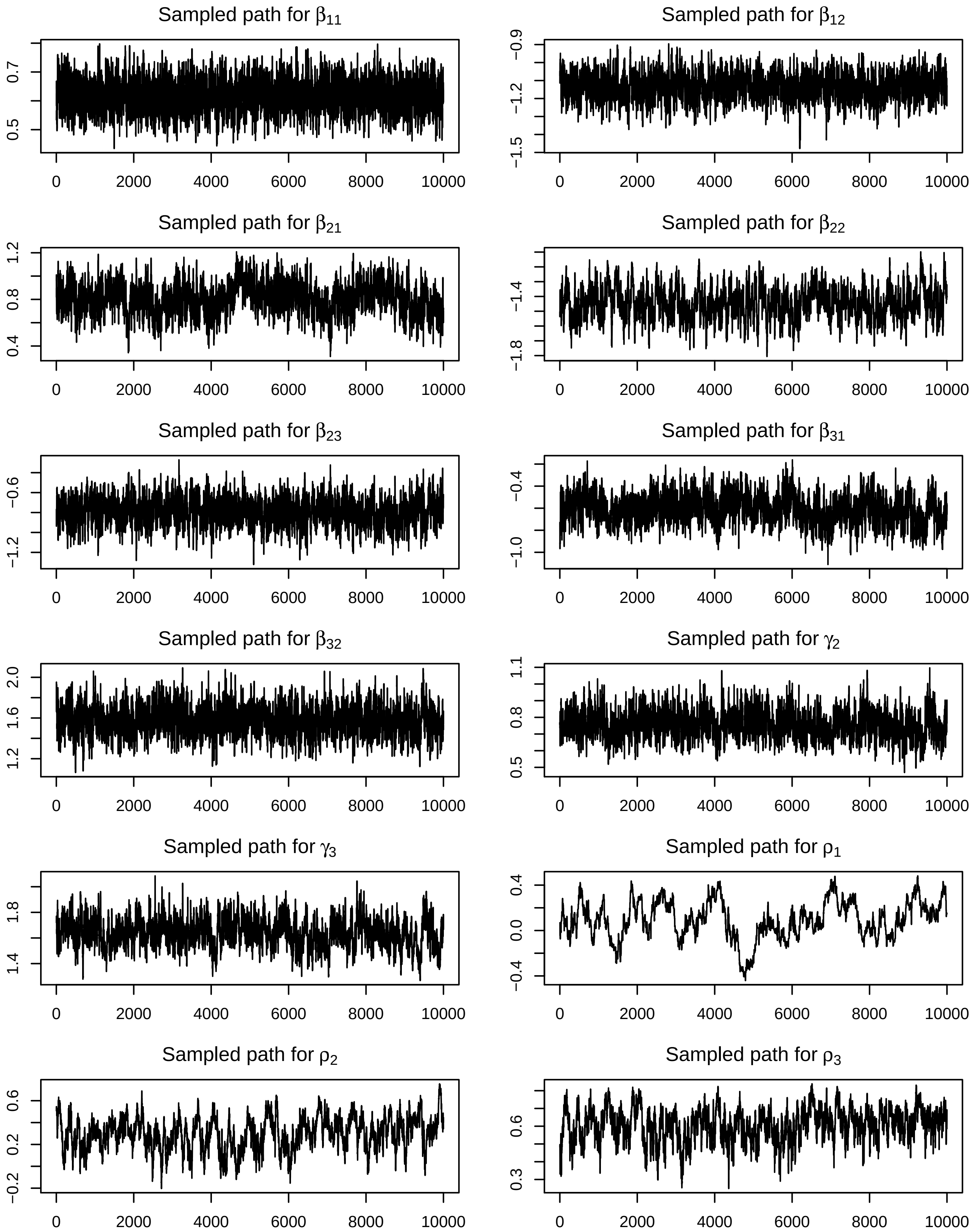
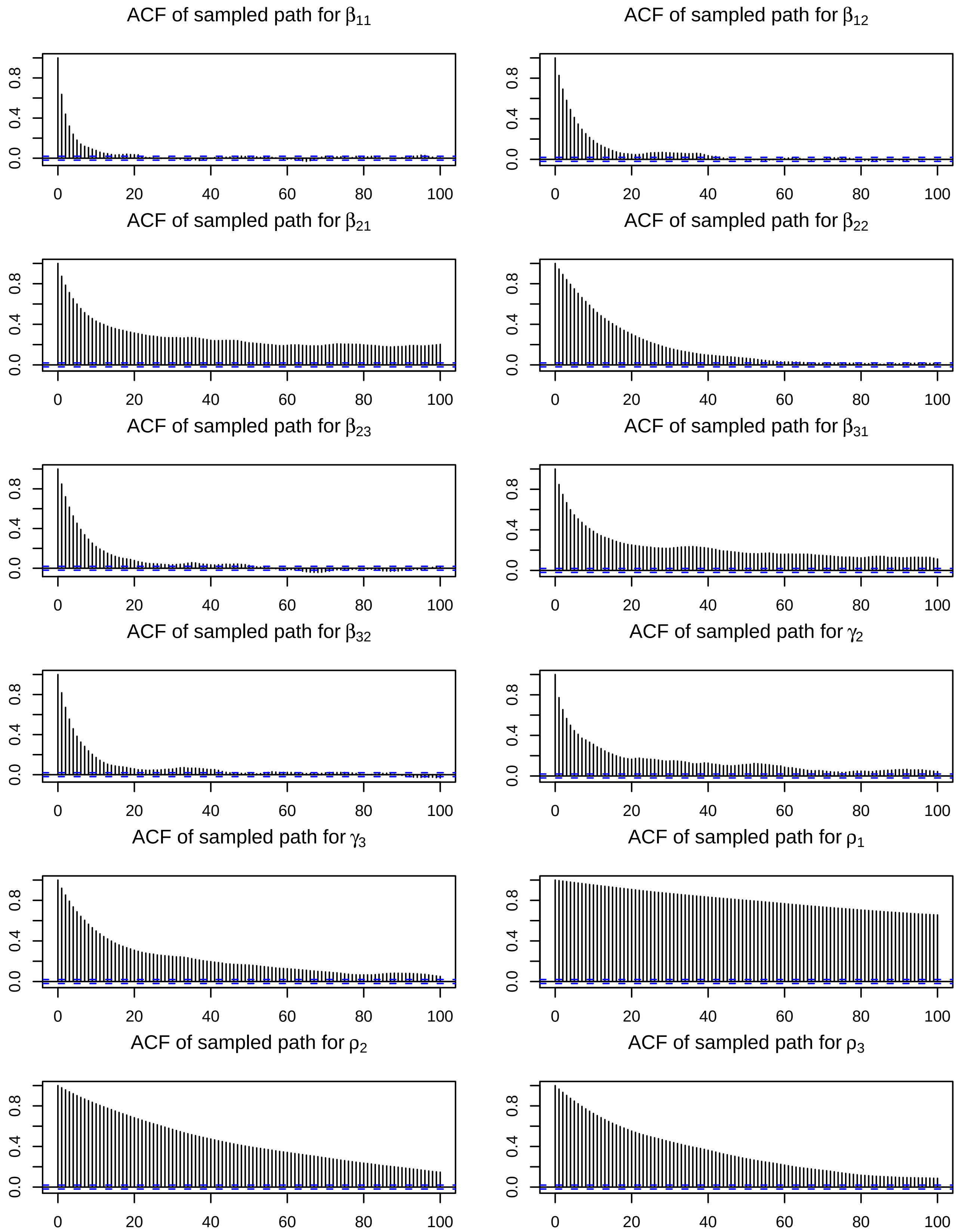


Figure 5.4: *Sample Autocorrelation Functions for Sampler B*



Summary statistics of the model parameters are displayed in Table 5.2. In comparison to Table 5.1, the mean value of ρ_1 in Table 5.2 is less accurate, but this coefficient is also not significant as its 95% credible interval is $(-0.089, 0.293)$ which includes zero. The mean value of ρ_2 seems to be more accurate in Table 5.2. Standard deviations in the two tables are similar, however, those for γ_2 and γ_3 are smaller in Table 5.2. SIF values in Table 5.2 are all smaller than 100 except the one for ρ_1 revealing that ρ_1 is the most difficult to estimate for Sampler B. In contrast, SIF values for γ and ρ parameters are all larger than 100 for Sampler A, as has been shown in Table 5.1. As a result, reparameterization has achieved its goal in improving the convergence of the MCMC algorithms.

5.5 Monte Carlo Experiments

Results of one sample estimation obtained in Section 5.3.4 and 5.4.4 show that Sampler A produces poor convergence in sampling all γ and ρ parameters, while Sampler B can get quite good convergent posteriors except for one parameter ρ_1. In this section, a Monte Carlo study with 1,000 samples will be conducted to systematically check the performance of the two samplers as well as comparing with FIML estimates.

5.5.1 Full Information Maximum Likelihood Estimation

Let Φ, Φ_2 and Φ_3 represent the cumulative distribution function of the standardized univariate, bivariate and trivariate normal density respectively. When $y_{i1} = 0$, the joint distribution of y_{i1}, y_{i2} and y_{i3} is just $Pr(y_{i1} = 0|x_{i1})$ because both y_{i2} and y_{i3} cannot be observed. Therefore,

$$P_0 = Pr(y_{i1} = 0|x_{i1}) = \Phi(-x'_{i1}\beta_1).$$

If $y_{i1} = 1$ and $y_{i2} = 0$, y_{i3} is still unobserved. Thus, $Pr(y_{i1}, y_{i2}, y_{i3}|x_{i1}, x_{i2}, x_{i3})$ is just $Pr(y_{i1} = 1, y_{i2} = 0|x_{i1}, x_{i2})$ and

$$P_{10} = Pr(y_{i1} = 1, y_{i2} = 0|x_{i1}, x_{i2}) = \Phi_2(x'_{i1}\beta_1, -x'_{i2}\beta_2, -\rho_1).$$

When $y_{i1} = 1$ and $y_{i2} = 1$, y_{i3} can be collected as a categorial number from 1 to J. Let

$$P_{11j} = Pr(y_{i1} = 1, y_{i2} = 1, y_{i3} = j | x_{i1}, x_{i2}, x_{i3}) \qquad (j = 1, \cdots, J).$$

The joint distribution of y_{i1}, y_{i2} and y_{i3} becomes

$$\begin{cases} P_{111} &= \Phi_3(x'_{i1}\beta_1, x'_{i2}\beta_2, -x'_{i3}\beta_3, \rho_1, -\rho_2, -\rho_3) \\ P_{11j} &= \Phi_3(x'_{i1}\beta_1, x'_{i2}\beta_2, \gamma_j - x'_{i3}\beta_3, \rho_1, -\rho_2, -\rho_3) - \\ & \quad \Phi_3(x'_{i1}\beta_1, x'_{i2}\beta_2, \gamma_{j-1} - x'_{i3}\beta_3, \rho_1, -\rho_2, -\rho_3) \\ & \quad (j = 2, \cdots, J-1) \\ P_{11J} &= \Phi_3(x'_{i1}\beta_1, x'_{i2}\beta_2, x'_{i3}\beta_3 - \gamma_{J-1}, \rho_1, \rho_2, \rho_3), \end{cases} \tag{5.5.1}$$

with the assumption that error terms follow standard trivariate normal distribution. Notice $\Phi_3(x_1, x_2, x_3, \rho_{12}, \rho_{13}, \rho_{23})$ is the cumulative distribution function of the standardized trivariate Normal density in which x_1, x_2, x_3 are the upper limits of integration for the three variables, ρ_{12} is the correlation coefficient between the two variables $x1$ and $x2$, ρ_{13} is the correlation coefficient between the two variables $x1$ and $x3$, and ρ_{23} is the correlation coefficient between the two variables $x2$ and $x3$.

In order to estimate the model parameters, the likelihood function,

$$L = \prod_{i=1}^{n} \left\{ P_0^{(1-y_{i1})} P_{10}^{y_{i1}(1-y_{i2})} \left[\sum_{j=1}^{J} I(y_{i3} = j) P_{11j} \right]^{y_{i1}y_{i2}} \right\}, \tag{5.5.2}$$

is maximized.

5.5.2 General Monte Carlo Design

For each sample, the sample size is fixed at 1,000 and the true parameter values are set as $\beta_1 = (\beta_{11}, \beta_{12})' = (0.6, -1.2)'$, $\beta_2 = (\beta_{21}, \beta_{22}, \beta_{23})' = (1, -1.5, -1)'$, $\beta_3 = (\beta_{31}, \beta_{32})' = (-0.4, 1.5)'$ and $\gamma = (\gamma_2, \gamma_3)' = (0.8, 1.6)'$. Three sets of true values are used for ρ to detect different strength of error correlation. One set is $(\rho_1, \rho_2, \rho_3) = (0, 0, 0)$ when there is no correlation between error terms. Another is with medium strength correlation such that $(\rho_1, \rho_2, \rho_3) = (0.25, 0.25, 0.5)$. Strong correlation errors use the values

$(\rho_1, \rho_2, \rho_3) = (0.5, 0.8, 0.7)$. Explanatory variables are randomly simulated imposing exclusion restrictions in the same way as they are designed in Section 5.3.4. Latent variables are calculated after the error terms are generated from trivariate normal distributions. Then, the outcomes of y_{i1}, y_{i2} and y_{i3} are obtained from Equations (5.2.1)-(5.2.3). For each set of the true ρ values, 1,000 samples are generated from the above process. Then FIML and the two Bayesian samplers are applied to estimate those samples.

Starting values are quite arbitrary for the three estimators. For FIML, the starting values are all given as zero, except $\gamma_2 = 1$ and $\gamma_3 = 2$ as these threshold parameters must be larger than zero. Using Sampler A, the starting value for γ_2 is 0.5 and that for γ_3 is set at 1, while the other starting values are zero. For Sampler B, the starting values are set at zero, besides $\gamma_2^* = 0.5$ and $\psi = 1$. The MCMC point estimators are obtained by taking the average of 10,000 draws from the joint posterior distribution after discarding 2,000 initial draws.

FIML sometimes fails to provide reliable estimates when the Hessian matrix fails to invert. Meanwhile, the MCMC samplers can always give some output, so summary statistics denoted by SamplerA and SamplerB are obtained by the whole 1,000 samples. In each result table, FIML* indicates the statistics with respect to the samples from which FIML can get reliable results after removing those without Hessian matrix. In order to make sure estimators are compared according to the same samples, SamplerA* and SamplerB* represent the Bayesian estimators on exactly the same samples as FIML*.

5.5.3 Monte Carlo Results

The Monte Carlo results are displayed in Table 5.3 - Table 5.8. First, we will analyze the convergence of the optimization method. FIML fails to converge more than half the time as it can only estimate 420, 453 and 458 samples according to different levels of error correlation, much more often than its performance in estimating the sample selection model which has been discussed in the previous chapter. Obviously, the greater complexity of the model has increased the difficulties in estimation. The may be a result of the increased number of equations or one additional hurdle of sample selection in the model. Notice that these are cases with exclusion restrictions. The performance of FIML will be even worse if exclusion restrictions are not imposed in explanatory variables (see Section 4.4).

Table 5.3 and Table 5.4 present results of the three estimation methods when the 1,000 samples are simulated without error correlation. We will compare statistics across parameters, as statistics of each parameter are quite close across methods. Firstly, all mean values are quite close to the true values, meaning all three estimators can provide unbiased estimation when there is no error correlation. Secondly, ρ_1 has the largest standard deviation for all estimators, which proves that it is the most difficult to estimate. Standard deviations of ρ_2 and ρ_3 are also large considering that they must be constrained in (-1, 1). Standard deviations of β_{21} and β_3 are twice as large as those of β_1, which may be due to the double hurdles of sample selection. Thirdly, extreme values of different samples are not exactly the same. For instance, minimum values of β_{22} obtained from the 420 samples by three methods are all around -1.83, but those got from 1,000 samples by Sampler A and Sampler B are around -2.14. As a result, a few samples FIML fails to estimate may lead to more extreme estimation for the Bayesian samplers. Fourthly, mean absolute errors and mean squared errors of β_{21}, β_3 and ρ are also relatively larger than those of other parameters. Finally, convergence rates evaluated by SIF values are compared between the two Bayesian samplers. SIF values of β are quite small for both samplers in Table 5.3, meaning those parameters converge rapidly. Table 5.4 shows that Sampler A performs quite poorly in simulating threshold parameters whose largest mean SIF value is around 150, and Sampler A also does not produce fast convergent chains in evaluating ρ values. The largest mean SIF value for Sampler B is 124, the one of ρ_1, while the remaining values are not large. Therefore, Sampler B can provide more efficient estimation than Sampler A when there is no error correlation.

Estimation methods are compared according to Table 5.5 and Table 5.6 when there is medium strength error correlation. Once again, FIML gives quite unbiased estimation for all parameters where the largest bias is 0.015. Meanwhile, Sampler A results in a little bias in estimating β_{31} and ρ_2 by around 0.03. Mean values obtained from Sampler B are quite close to the true values, except ρ_2 whose bias is more than 0.02. The standard deviations of ρ_1, ρ_2 and ρ_3 for Sampler A are relatively smaller than those for other methods. The standard deviations of other parameters are slightly different for different

Table 5.3: *Estimates of β without Correlation in Errors*

	β_{11}	β_{12}	β_{21}	β_{22}	β_{23}	β_{31}	β_{32}
True Values	0.600	-1.200	1.000	-1.500	-1.000	-0.400	1.500
Mean							
FIML*	0.605	-1.213	0.995	-1.504	-0.989	-0.396	1.499
SamplerA	0.605	-1.208	1.003	-1.512	-1.003	-0.395	1.497
SamplerA*	0.608	-1.212	1.001	-1.507	-0.993	-0.396	1.497
SamplerB	0.606	-1.206	1.001	-1.510	-1.001	-0.398	1.484
SamplerB*	0.608	-1.211	0.998	-1.504	-0.991	-0.399	1.483
Standard Deviation							
FIML*	0.053	0.076	0.149	0.107	0.149	0.143	0.140
SamplerA	0.053	0.073	0.147	0.110	0.150	0.138	0.145
SamplerA*	0.054	0.076	0.146	0.108	0.149	0.138	0.141
SamplerB	0.054	0.073	0.149	0.109	0.149	0.141	0.143
SamplerB*	0.054	0.076	0.148	0.107	0.148	0.142	0.139
Maximum Values							
FIML*	0.761	-1.013	1.402	-1.257	-0.629	0.080	1.932
SamplerA	0.756	-1.020	1.565	-1.205	-0.616	0.036	1.977
SamplerA*	0.756	-1.021	1.440	-1.258	-0.632	0.036	1.920
SamplerB	0.759	-1.009	1.552	-1.211	-0.615	0.074	1.936
SamplerB*	0.759	-1.017	1.405	-1.249	-0.631	0.074	1.922
Minimum Values							
FIML*	0.471	-1.489	0.518	-1.832	-1.483	-0.819	1.071
SamplerA	0.444	-1.479	0.549	-2.148	-1.605	-0.873	1.003
SamplerA*	0.470	-1.479	0.549	-1.836	-1.484	-0.873	1.081
SamplerB	0.442	-1.475	0.483	-2.139	-1.588	-0.886	0.976
SamplerB*	0.473	-1.475	0.483	-1.830	-1.484	-0.886	1.060
Mean Absolute Error							
FIML*	0.043	0.060	0.119	0.085	0.118	0.112	0.112
SamplerA	0.043	0.058	0.116	0.086	0.119	0.108	0.115
SamplerA*	0.043	0.060	0.115	0.085	0.118	0.107	0.113
SamplerB	0.043	0.058	0.117	0.086	0.118	0.110	0.114
SamplerB*	0.044	0.060	0.117	0.085	0.117	0.110	0.111
Mean Squared Error							
FIML*	0.003	0.006	0.022	0.012	0.022	0.020	0.020
SamplerA	0.003	0.005	0.021	0.012	0.022	0.019	0.021
SamplerA*	0.003	0.006	0.021	0.012	0.022	0.019	0.020
SamplerB	0.003	0.005	0.022	0.012	0.022	0.020	0.021
SamplerB*	0.003	0.006	0.022	0.011	0.022	0.020	0.020
SIF for SamplerA							
Mean	7	12	33	26	13	41	22
Standard Deviation	2	3	9	7	3	9	6
SIF for SamplerB							
Mean	7	12	36	28	13	29	7
Standard Deviation	2	5	10	8	4	7	2

* 420 samples included

Table 5.4: *Estimates of γ and ρ without Correlation in Errors*

	γ_2	γ_3	ρ_1	ρ_2	ρ_3
True Values	0.800	1.600	0.000	0.000	0.000
Mean					
FIML*	0.791	1.598	0.001	0.003	-0.006
SamplerA	0.798	1.597	0.005	-0.003	-0.002
SamplerA*	0.791	1.597	0.002	0.003	-0.004
SamplerB	0.784	1.570	0.008	-0.004	0.000
SamplerB*	0.777	1.570	0.004	0.003	-0.003
Standard Deviation					
FIML*	0.075	0.096	0.193	0.164	0.145
SamplerA	0.077	0.098	0.163	0.143	0.127
SamplerA*	0.079	0.100	0.169	0.146	0.132
SamplerB	0.074	0.095	0.176	0.165	0.140
SamplerB*	0.074	0.096	0.183	0.168	0.144
Maximum Values					
FIML*	1.013	1.847	0.530	0.491	0.345
SamplerA	1.070	1.932	0.439	0.409	0.457
SamplerA*	1.070	1.889	0.439	0.380	0.347
SamplerB	1.031	1.916	0.531	0.473	0.473
SamplerB*	1.002	1.814	0.516	0.473	0.352
Minimum Values					
FIML*	0.596	1.343	-0.470	-0.443	-0.373
SamplerA	0.583	1.329	-0.473	-0.432	-0.395
SamplerA*	0.583	1.329	-0.457	-0.385	-0.366
SamplerB	0.579	1.304	-0.559	-0.495	-0.428
SamplerB*	0.585	1.312	-0.559	-0.451	-0.387
Mean Absolute Error					
FIML*	0.061	0.076	0.159	0.137	0.120
SamplerA	0.062	0.078	0.130	0.116	0.102
SamplerA*	0.064	0.078	0.138	0.122	0.109
SamplerB	0.061	0.079	0.140	0.134	0.113
SamplerB*	0.063	0.080	0.150	0.140	0.119
Mean Squared Error					
FIML*	0.006	0.009	0.037	0.027	0.021
SamplerA	0.006	0.010	0.027	0.020	0.016
SamplerA*	0.006	0.010	0.029	0.021	0.017
SamplerB	0.006	0.010	0.031	0.027	0.020
SamplerB*	0.006	0.010	0.034	0.028	0.021
SIF for SamplerA					
Mean	145	149	115	105	90
Standard Deviation	12	12	15	15	14
SIF for SamplerB					
Mean	10	17	124	61	47
Standard Deviation	3	5	17	13	10

* 420 samples included

methods. Although mean values and standard deviations are very similar with different samples for each method, maximum and minimum values in the two tables show estimation can be affected by different samples. Mean absolute errors and mean squared errors are quite close across methods. In order to evaluate convergence rates, mean and standard deviation of SIF values are compared between the two Bayesian samplers. Five mean SIF values of Sampler A are larger than 100, with the largest value 163. At the same time, the largest mean SIF value is 144, the one of β_1 for Sampler B, while the other mean SIF values are smaller than 100. All standard deviations of SIF values are equal to or less than 20. Once again, Sampler B provides estimation with better overall convergence rates than Sampler A.

Next we consider the performance of the three estimators when correlation of the error terms is high, in Table 5.7 and Table 5.8. All mean values are quite close to the true values for FIML. Sampler A shows a little bias in estimating β_{22}, β_{31}, β_{32}, ρ_1, ρ_2 and ρ_3, where the bias is around 0.05. For Sampler B, only ρ_1 has relatively larger bias (at around 0.05) than other parameters. There is not much difference in statistics like standard deviations, mean absolute errors and mean squared errors across methods. Statistics on SIF are not large in Table 5.7, revealing that β estimates converge fast for both Bayesian samplers. In Table 5.8, mean SIF values of γ are 154 and 173 for Sampler A, much larger than those for Sampler B, which are less than 100. The largest mean SIF value for Sampler B is 170, the one for ρ_1, which is a litter larger than that for Sampler A. Meanwhile, mean SIF values of ρ_2 and ρ_3 for Sampler B are both smaller than those for Sampler A. As a result, it is most difficult to estimate γ and ρ for Sampler A and to estimate ρ_1 for Sampler B.

To sum up, FIML and the Bayesian methods have shown different advantages in estimation. FIML fails to offer reliable estimation half the time, while the Bayesian methods can evaluate all samples. FIML estimations are all accurate and efficient, while both Bayesian samplers are a little less accurate but more efficient in some ways than FIML. When comparing the two Bayesian samplers, sampler B shows better convergence and more efficiency than sampler A. Stronger error correlation will reduce the overall convergence rate for both Bayesian samplers. The convergence rates in estimating γ and ρ are relatively slow for Sampler A, while Sampler B has some difficulty to evaluate ρ_1. Overall, Sampler B is preferred to Sampler A.

Table 5.5: *Estimates of β with Moderate Correlation in Errors*

	β_{11}	β_{12}	β_{21}	β_{22}	β_{23}	β_{31}	β_{32}
True Values	0.600	-1.200	1.000	-1.500	-1.000	-0.400	1.500
Mean							
FIML*	0.599	-1.211	1.002	-1.515	-1.005	-0.391	1.511
SamplerA	0.605	-1.209	1.023	-1.525	-1.011	-0.367	1.511
SamplerA*	0.602	-1.211	1.021	-1.525	-1.014	-0.364	1.520
SamplerB	0.606	-1.206	1.015	-1.514	-1.004	-0.397	1.486
SamplerB*	0.603	-1.209	1.013	-1.515	-1.007	-0.394	1.496
Standard Deviation							
FIML*	0.052	0.072	0.153	0.106	0.149	0.123	0.131
SamplerA	0.053	0.073	0.157	0.111	0.154	0.125	0.135
SamplerA*	0.053	0.072	0.149	0.106	0.149	0.121	0.130
SamplerB	0.054	0.073	0.159	0.111	0.154	0.126	0.133
SamplerB*	0.052	0.072	0.152	0.106	0.149	0.123	0.129
Maximum Values							
FIML*	0.752	-1.022	1.481	-1.211	-0.588	-0.027	1.978
SamplerA	0.757	-1.023	1.563	-1.210	-0.589	0.015	2.032
SamplerA*	0.757	-1.023	1.514	-1.223	-0.589	-0.017	1.952
SamplerB	0.756	-1.017	1.565	-1.200	-0.589	-0.007	1.951
SamplerB*	0.750	-1.021	1.464	-1.200	-0.589	-0.038	1.951
Minimum Values							
FIML*	0.442	-1.405	0.613	-1.974	-1.515	-0.753	0.976
SamplerA	0.443	-1.465	0.504	-2.033	-1.524	-0.727	0.989
SamplerA*	0.443	-1.406	0.632	-1.967	-1.524	-0.727	0.989
SamplerB	0.447	-1.463	0.476	-2.022	-1.540	-0.777	0.973
SamplerB*	0.447	-1.408	0.602	-1.960	-1.540	-0.772	0.973
Mean Absolute Error							
FIML*	0.043	0.059	0.121	0.084	0.117	0.100	0.103
SamplerA	0.043	0.058	0.123	0.088	0.122	0.104	0.106
SamplerA*	0.043	0.059	0.118	0.084	0.116	0.101	0.104
SamplerB	0.043	0.058	0.125	0.087	0.121	0.101	0.106
SamplerB*	0.043	0.059	0.120	0.084	0.117	0.099	0.102
Mean Squared Error							
FIML*	0.003	0.005	0.023	0.012	0.022	0.015	0.017
SamplerA	0.003	0.005	0.025	0.013	0.024	0.017	0.018
SamplerA*	0.003	0.005	0.023	0.012	0.022	0.016	0.017
SamplerB	0.003	0.005	0.026	0.012	0.024	0.016	0.018
SamplerB*	0.003	0.005	0.023	0.011	0.022	0.015	0.017
SIF for SamplerA							
Mean	7	14	45	38	18	51	32
Standard Deviation	2	4	11	10	5	12	10
SIF for SamplerB							
Mean	8	15	51	42	20	37	12
Standard Deviation	3	6	14	14	9	10	7

* 453 samples included

Table 5.6: *Estimates of γ and ρ with Moderate Correlation in Errors*

	γ_2	γ_3	ρ_1	ρ_2	ρ_3
True Values	0.800	1.600	0.250	0.250	0.500
Mean					
FIML*	0.805	1.610	0.255	0.245	0.489
SamplerA	0.808	1.618	0.220	0.209	0.457
SamplerA*	0.811	1.622	0.223	0.210	0.445
SamplerB	0.787	1.578	0.233	0.230	0.500
SamplerB*	0.790	1.583	0.237	0.234	0.487
Standard Deviation					
FIML*	0.085	0.101	0.166	0.151	0.112
SamplerA	0.084	0.103	0.157	0.137	0.108
SamplerA*	0.086	0.102	0.152	0.135	0.108
SamplerB	0.081	0.100	0.172	0.158	0.113
SamplerB*	0.083	0.100	0.161	0.156	0.114
Maximum Values					
FIML*	1.031	1.965	0.696	0.612	0.780
SamplerA	1.071	1.992	0.621	0.632	0.743
SamplerA*	1.071	1.992	0.621	0.528	0.717
SamplerB	1.011	1.928	0.709	0.650	0.790
SamplerB*	1.011	1.928	0.693	0.595	0.771
Minimum Values					
FIML*	0.552	1.270	-0.335	-0.299	0.144
SamplerA	0.565	1.258	-0.268	-0.268	0.107
SamplerA*	0.565	1.258	-0.268	-0.268	0.107
SamplerB	0.542	1.250	-0.361	-0.338	0.132
SamplerB*	0.542	1.250	-0.196	-0.321	0.142
Mean Absolute Error					
FIML*	0.068	0.079	0.130	0.120	0.089
SamplerA	0.068	0.083	0.127	0.114	0.092
SamplerA*	0.069	0.081	0.123	0.111	0.096
SamplerB	0.066	0.081	0.137	0.128	0.091
SamplerB*	0.067	0.079	0.127	0.123	0.091
Mean Squared Error					
FIML*	0.007	0.010	0.028	0.023	0.013
SamplerA	0.007	0.011	0.025	0.020	0.013
SamplerA*	0.007	0.011	0.024	0.020	0.015
SamplerB	0.007	0.010	0.030	0.025	0.013
SamplerB*	0.007	0.010	0.026	0.025	0.013
SIF for SamplerA					
Mean	151	163	133	123	113
Standard Deviation	13	11	16	17	18
SIF for SamplerB					
Mean	18	28	144	81	64
Standard Deviation	7	10	18	20	16

* 453 samples included

Table 5.7: *Estimates of β with High Correlation in Errors*

	β_{11}	β_{12}	β_{21}	β_{22}	β_{23}	β_{31}	β_{32}
True Values	0.600	-1.200	1.000	-1.500	-1.000	-0.400	1.500
Mean							
FIML*	0.601	-1.207	1.013	-1.514	-1.009	-0.394	1.515
SamplerA	0.605	-1.219	1.048	-1.538	-1.026	-0.358	1.545
SamplerA*	0.603	-1.217	1.050	-1.537	-1.027	-0.357	1.551
SamplerB	0.607	-1.206	1.038	-1.521	-1.014	-0.405	1.493
SamplerB*	0.605	-1.204	1.040	-1.520	-1.015	-0.403	1.499
Standard Deviation							
FIML*	0.053	0.070	0.158	0.111	0.147	0.109	0.124
SamplerA	0.053	0.072	0.157	0.112	0.149	0.107	0.129
SamplerA*	0.053	0.071	0.158	0.110	0.149	0.112	0.127
SamplerB	0.053	0.072	0.161	0.113	0.149	0.104	0.124
SamplerB*	0.053	0.071	0.162	0.112	0.148	0.110	0.123
Maximum Values							
FIML*	0.757	-1.014	1.617	-1.265	-0.636	-0.048	1.930
SamplerA	0.759	-1.027	1.596	-1.251	-0.626	-0.023	1.944
SamplerA*	0.757	-1.027	1.596	-1.284	-0.641	-0.023	1.907
SamplerB	0.768	-0.997	1.691	-1.215	-0.618	-0.043	1.896
SamplerB*	0.768	-0.997	1.691	-1.244	-0.631	-0.043	1.896
Minimum Values							
FIML*	0.457	-1.457	0.539	-1.847	-1.659	-0.693	1.196
SamplerA	0.454	-1.478	0.593	-1.906	-1.644	-0.661	1.190
SamplerA*	0.456	-1.460	0.593	-1.832	-1.644	-0.661	1.190
SamplerB	0.459	-1.461	0.596	-1.904	-1.682	-0.708	1.173
SamplerB*	0.461	-1.461	0.596	-1.857	-1.682	-0.708	1.180
Mean Absolute Error							
FIML*	0.042	0.055	0.124	0.089	0.114	0.087	0.099
SamplerA	0.042	0.059	0.129	0.093	0.119	0.093	0.107
SamplerA*	0.042	0.056	0.129	0.092	0.118	0.096	0.107
SamplerB	0.042	0.057	0.131	0.091	0.118	0.083	0.099
SamplerB*	0.042	0.056	0.130	0.090	0.116	0.088	0.098
Mean Squared Error							
FIML*	0.003	0.005	0.025	0.012	0.022	0.012	0.016
SamplerA	0.003	0.006	0.027	0.014	0.023	0.013	0.019
SamplerA*	0.003	0.005	0.027	0.013	0.023	0.014	0.019
SamplerB	0.003	0.005	0.027	0.013	0.022	0.011	0.015
SamplerB*	0.003	0.005	0.028	0.013	0.022	0.012	0.015
SIF for SamplerA							
Mean	16	34	57	57	27	57	59
Standard Deviation	7	14	14	16	10	16	17
SIF for SamplerB							
Mean	24	52	68	73	39	59	50
Standard Deviation	14	27	21	27	20	19	23

* 458 samples included

Table 5.8: *Estimates of γ and ρ with High Correlation in Errors*

	γ_2	γ_3	ρ_1	ρ_2	ρ_3
True Values	0.800	1.600	0.500	0.800	0.700
Mean					
FIML*	0.810	1.614	0.500	0.793	0.699
SamplerA	0.829	1.653	0.443	0.736	0.648
SamplerA*	0.833	1.656	0.446	0.735	0.648
SamplerB	0.791	1.583	0.452	0.790	0.689
SamplerB*	0.796	1.589	0.455	0.788	0.689
Standard Deviation					
FIML*	0.090	0.115	0.147	0.079	0.084
SamplerA	0.090	0.117	0.143	0.087	0.088
SamplerA*	0.094	0.119	0.142	0.085	0.087
SamplerB	0.085	0.112	0.154	0.081	0.089
SamplerB*	0.089	0.116	0.154	0.082	0.089
Maximum Values					
FIML*	1.063	1.967	0.834	0.958	0.903
SamplerA	1.092	2.013	0.799	0.934	0.875
SamplerA*	1.092	2.013	0.777	0.934	0.875
SamplerB	1.045	1.942	0.843	0.963	0.909
SamplerB*	1.045	1.942	0.828	0.963	0.909
Minimum Values					
FIML*	0.590	1.369	-0.012	0.515	0.394
SamplerA	0.545	1.351	-0.100	0.303	0.289
SamplerA*	0.610	1.351	-0.096	0.472	0.392
SamplerB	0.512	1.279	-0.151	0.357	0.306
SamplerB*	0.577	1.297	-0.151	0.483	0.374
Mean Absolute Error					
FIML*	0.073	0.093	0.115	0.062	0.067
SamplerA	0.075	0.102	0.118	0.083	0.079
SamplerA*	0.080	0.105	0.115	0.083	0.078
SamplerB	0.069	0.092	0.125	0.064	0.070
SamplerB*	0.072	0.094	0.123	0.065	0.070
Mean Squared Error					
FIML*	0.008	0.013	0.021	0.006	0.007
SamplerA	0.009	0.016	0.024	0.012	0.010
SamplerA*	0.010	0.017	0.023	0.012	0.010
SamplerB	0.007	0.013	0.026	0.007	0.008
SamplerB*	0.008	0.013	0.026	0.007	0.008
SIF for SamplerA					
Mean	154	173	165	160	152
Standard Deviation	14	10	15	16	18
SIF for SamplerB					
Mean	61	91	170	122	112
Standard Deviation	23	30	14	23	26

* 458 samples included

5.6 Empirical Example

Cornwell et al. (2009) use the double-hurdle model to discuss the impact of mental illness at three stages of engagement in labour market. They formulate their model in a similar way to what has been discussed at the beginning of this chapter. The biggest difference is that the error correlations between the equations are not specially evaluated, as they apply a two-step method twice to deal with possible sample selection bias. In contrast, we are trying to estimate the model in a whole system even including the possible correlation between the first and third equations. Data used in this empirical study was generously provided by Cornwell. Background details can be found in their article in which selected population-weighted descriptive statistics are shown with mental health disorders. All analysis in their paper is performed with weighting according to provided replicate weights. Deaton (1997) has discussed both the benefit and drawback of weighted estimation. On one hand, the weighting makes the sample look like the population and removes the dependence of the estimates on the sample design. On the other hand, the difference in parameter values across strata is a feature of the population, so a regression on census data is no less problematic than on sample data. Since the arguments about whether weights should be used are quite controversial, the following analysis will not use weights resulting from the survey sampling process.

Certain specification of the variables are utilized as exogenous regressors in this section. The variables indicating mental health are dummy variables for substance use disorders, anxiety disorders and affective disorders. Other variables include the ones for age , gender, education, geography indices, and a socio-economic index for area (SEIFA). In Chapter 4, we pointed out that models with sample selection can be identified through nonlinearity but exclusion restrictions will greatly reduce the difficulty of estimation. In the first hurdle, two variables are included in the participation equation but not included in the employment equation: number of children in the household, and whether the individual is currently studying. We impose this exclusion restriction to make estimation easier for all estimation methods. In the second hurdle, however, the factors which have effect on employment are all likely to influence occupational skill levels. Thus, regressors are exactly the same for the second and third equations. In other words, no exclusion restriction exists in the two equations. After estimating the model, we will discuss whether different categories of mental illness have an effect on labour market outcomes.

5.6.1 Estimation Methodology

Currently, no software package including NLOGIT can directly estimate the double-hurdle model as a whole system. Available estimators include the one applied by Cornwell et al. which is an extension of Heckman's two-step least squares, as well as FIML and the Bayesian samplers discussed in this chapter.

We will start with the description of the estimator used by Cornwell et al. with the idea of Heckman's two-step method applied to estimate each hurdle. After the first probit equation (5.2.1) is estimated for participation, a selection adjustment is added as a regressor into the second equation (5.2.2). Then the second equation can be estimated as a probit model with normality assumption. A second adjustment factor is plugged in the third equation (5.2.3) as an explanatory variable, so the third equation can be estimated as an ordered probit model in a usual way. This estimator attempts to remove selection bias at each hurdle and it is easy to implement. However, the two-step method is inconsistent even in estimating the first hurdle. In addition, the step-by-step method fails to treat the model as a whole system, because it does not fully take into account the effects of error correlations. Although one may argue that the adjustment factors have captured part of the effects of the correlations, the correlation between the first and third equation is certainly not considered by this procedure. In other words, the effect of the first hurdle may influence the evaluation of third equation through the error correlation, while the two-step method cannot clearly identify such effect. Thus, the sample selection effects of the two hurdles are somehow isolated.

As a result, more accurate estimation methods are applied to evaluate the empirical data including FIML and the Bayesian method. FIML is estimated on the empirical data by the CML package in GAUSS 9.0 but fails to give covariance matrix of the parameters, so such estimates are not valid. Although the Bayesian method is not simpler to implement, the two samplers proposed in this chapter are more appropriate for the model specification with double hurdles. In addition, coefficients of error correlation can be estimated and the computation of marginal effects is more efficient.

The two Bayesian samplers are used to estimate the empirical data with 10,000 iterations as burn-in period and the following 100,000 iterations recorded for constructing posterior densities. Convergence of Markov chains must be guaranteed before estimates can be used for further analysis. To examine the convergence rate of all parameters, 5,000 sampled parameter vectors are used (i.e. 1 draw for every 20 draws of the MCMC iterations). All sampled paths and autocorrelation functions of 5,000 vectors have been checked as well as their densities. Only those of six parameters for γ and ρ are displayed in Figure 5.5 and Figure 5.6, because sampled paths for all β parameters converge very fast for both samplers.

As shown in Figure 5.5, Sampler A has some convergence issues. The sampled paths for the three threshold parameters look terrible with very strong cyclical patterns, although other paths stay quite stationary. In the second column, the autocorrelation functions of the three parameters decay quite slow, while their densities are not smooth at all and each has multiple modes. Therefore, we cannot ensure that the Markov Chains for Sampler A converge to the joint posterior distributions and the estimates from this sampler are not acceptable for this empirical work.

According to results in Figure 5.6, all chains converge fast for Sampler B except the one of ρ_1. Sampled paths for γ parameters look quite randomly distributed with ACF values decaying very fast, and their densities are quite normally distributed. However, the sample path of ρ_1 does not converge as well as γ and also has a few fluctuations, although it does hover around the horizontal level. The ACF values of ρ_1 decay slowest in this figure, revealing that it takes much more time to get convergent results for ρ_1 than other parameters. The distribution of ρ_1 shows a little unevenness in the left tail and it is also not quite smooth on the right side. Meanwhile, ρ_2 converge as fast as γ values. Although the fluctuations in the sample path for ρ_1 are quite similar to those for ρ_3, the sampled path of ρ_3 looks very stationary and the distribution of ρ_3 is quite bell-shaped. Ignoring a few extreme draws, the Markov Chains can be accepted for Sampler B.

Figure 5.5: *Results of the Empirical Study for Sampler A*

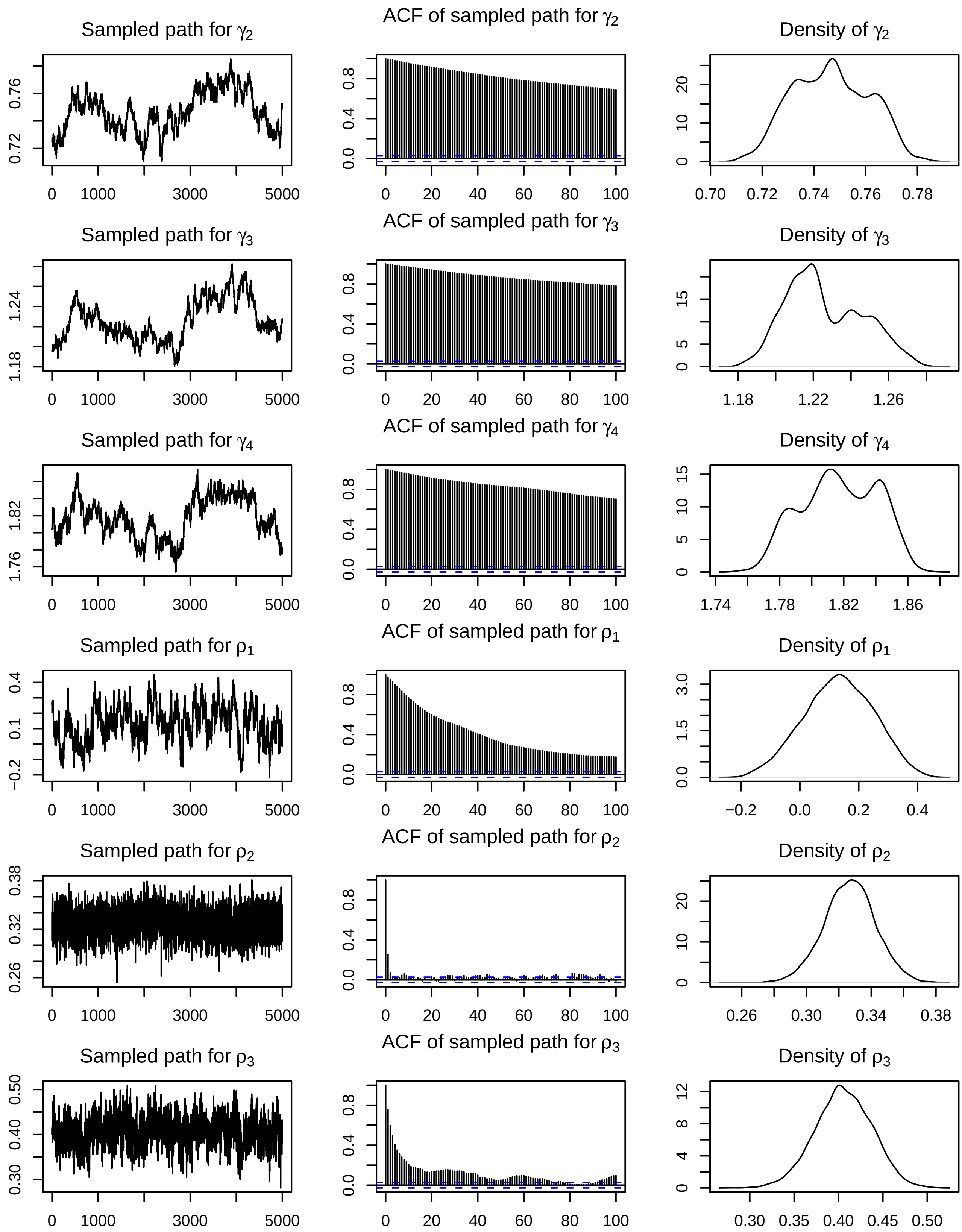
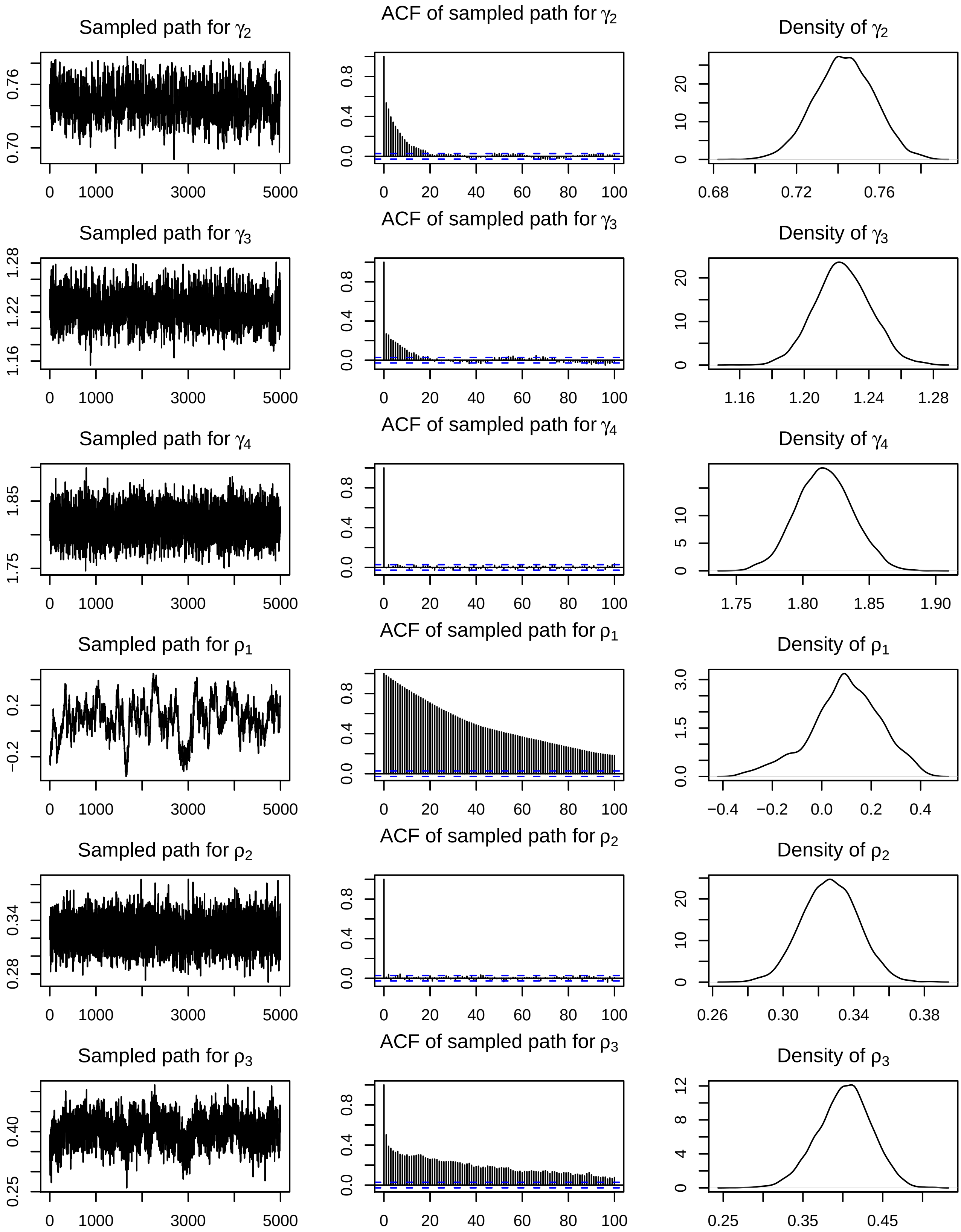


Figure 5.6: *Results of the Empirical Study for Sampler B*



5.6.2 Marginal Effects and Discussion

Chapter 4 has shown the importance of marginal effects and how to obtain appropriate marginal effects according to the features of variables. If the variable is continuous, the marginal effects are obtained from the derivative of the probability with respect to the variable. If it is binary, the marginal effect is calculated from the difference displayed in equation (4.5.1). In this section, the estimates of Sampler B are utilized to estimate the marginal effects of the following probabilities. The probability of participating in the labour market can be shown as

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

The probability of being employed after participation is represented by

$$\begin{aligned} Pr(y_{i1} = 1|y_{i2} = 1, x_{i1}, x_{i2}) &= Pr(y_{i1} = 1, y_{i2} = 1|x_{i1}, x_{i2})/Pr(y_{i1} = 1|x_{i1}) \\ &= \Phi_2(x_{i1}'\beta_1, x_{i2}'\beta_2, \rho_1)/\Phi(x_{i1}'\beta_1). \end{aligned}$$

And the probability of working in each occupational skill category after employment can be calculated from

$$\begin{aligned} Pr(y_{i3} = j|y_{i1} = 1, y_{i2} = 1, x_{i1}, x_{i2}, x_{i3}) &= P_{11j}/Pr(y_{i1} = 1, y_{i2} = 1|x_{i1}, x_{i2}) \\ &= P_{11j}/\Phi_2(x_{i1}'\beta_1, x_{i2}'\beta_2, \rho_1), \end{aligned}$$

referring to equation (5.5.1) for the expressions of joint probability P_{11j}. The estimated coefficients and corresponding marginal effects for such probabilities are all displayed in Table 5.9-5.12 with mean values and 95% Bayesian credible intervals.

First consider the results across variables in Table 5.9. People aged 25-44 are most likely to participate in the labour market, while those aged 45-64 are 28.9 percent points more likely to participate than those aged 18-24. Males have more probability to join in the labour force than females. People with higher levels of education are more likely to participate in the labour market than those with less education. The marginal effect about whether a person is from regional center is insignificant, but those from a rural

Table 5.9: *Modeling the Probability of Participating in the Labour Force*

	Coefficient		Marginal Effect	
Variable	Mean	95% Credible Interval	Mean	95% Credible Interval
Constant	-0.433	(-0.533 , -0.333)		
Age 25-44	1.260	(1.199 , 1.320)	0.409	(0.391 , 0.426)
Age 45-64	0.934	(0.872 , 0.996)	0.289	(0.272 , 0.305)
Male	0.062	(-0.001 , 0.126)	0.022	(0.000 , 0.044)
Competed Secondary School	0.217	(0.164 , 0.270)	0.076	(0.057 , 0.094)
Competed Higher Education	0.390	(0.309 , 0.473)	0.127	(0.103 , 0.151)
Competed Vocational Education	0.170	(0.087 , 0.254)	0.058	(0.030 , 0.085)
From a Regional Center	0.014	(-0.053 , 0.082)	0.005	(-0.019 , 0.029)
From a Rural Area	0.136	(0.072 , 0.199)	0.047	(0.025 , 0.068)
Socio-Economic Index: 2nd Decile	0.203	(0.098 , 0.307)	0.068	(0.034 , 0.101)
Socio-Economic Index: 3rd Decile	0.279	(0.175 , 0.382)	0.092	(0.060 , 0.123)
Socio-Economic Index: 4th Decile	0.285	(0.179 , 0.391)	0.094	(0.061 , 0.126)
Socio-Economic Index: 5th Decile	0.358	(0.254 , 0.463)	0.116	(0.085 , 0.146)
Socio-Economic Index: 6th Decile	0.390	(0.282 , 0.499)	0.125	(0.093 , 0.155)
Socio-Economic Index: 7th Decile	0.388	(0.286 , 0.491)	0.125	(0.095 , 0.154)
Socio-Economic Index: 8th Decile	0.487	(0.380 , 0.593)	0.152	(0.123 , 0.180)
Socio-Economic Index: 9th & 10th Decile	0.448	(0.357 , 0.538)	0.147	(0.120 , 0.174)
Number of Children	-0.250	(-0.271 , -0.229)	-0.088	(-0.096 , -0.081)
Currently Studying	0.434	(0.051 , 0.823)	0.128	(0.018 , 0.217)
Has at least one Physical Illness	-0.511	(-0.560 , -0.462)	-0.183	(-0.201 , -0.166)
Has at least one Anxiety Disorder(s)	-0.079	(-0.171 , 0.013)	-0.028	(-0.062 , 0.005)
Has at least one Affective Disorder(s)	0.011	(-0.087 , 0.110)	0.004	(-0.031 , 0.038)
Has at least one Substance Disorder(s)	0.290	(0.197 , 0.383)	0.095	(0.067 , 0.123)

The base person is a female, aged 18-24, not completed secondary school, from an urban area with the 1st(lowest) SEIFA Decile, with no children, not studying, and no physical or mental illnesses.

area are 4.7 percent points more likely to participate than those from an urban area. The participation rates in more socio-economically advanced areas are generally larger than those in less advanced areas, except areas in the 8th Decile which have the highest rate. Increased number of children will reduce the chances of a person participating in the labour force. People who are currently studying are more eager to seek jobs than those who are not studying. Physical illness will largely decrease the possibility of a person participating, while marginal effects of anxiety and affective disorders are insignificant as zero values are contained in their credible intervals. Therefore, such mental disorders have no effect on labour force participation. However, there is a higher participation rate by 9.5 percent points for people with substance disorders compared to those without this kind of disorders.

Marginal effects in the employment equation will be discussed based on the results in Table 5.10. People in the age group 25-44 and 45-64 are equally likely to be employed, and both groups have higher employment rates than age group 18-24. The credible interval of the variable 'Male' contains zero, thus its marginal effect is not significant and gender can hardly influence employment at the 5% significance level. People with 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 will most greatly increase the possibility of employment. Regional areas have similar employment rate to urban areas, while rural areas have 1.2 percent points higher employment rate. When considering the socio-economic indices, more advanced areas have higher employment rates than less advanced areas. Areas in 6th and 7th deciles have slightly higher employment rates than other areas except those in the highest decile. Although physical problems have no effect on the probability of being employed, three types of mental disorders will all reduce the possibility of employment. Especially, employment rate will decrease by 5.2 percent points for people with substance disorders.

Tables 5.11 shows mean values of estimated coefficients and marginal effects about occupational skill categories, while Table 5.12 displays 95% Credible Intervals for the respective coefficients. These tables indicate that older people are more likely to be employed in

Table 5.10: *Modeling the Probability of being Employed*

	Coefficient		Marginal Effect	
Variable	Mean	95% Credible Interval	Mean	95% Credible Interval
Constant	0.729	(0.400 , 1.065)		
Age 25-44	0.347	(0.178 , 0.513)	0.032	(0.019 , 0.045)
Age 45-64	0.369	(0.219 , 0.518)	0.034	(0.022 , 0.046)
Male	0.004	(-0.094 , 0.103)	0.000	(-0.011 , 0.011)
Competed Secondary School	0.232	(0.135 , 0.329)	0.025	(0.015 , 0.036)
Competed Higher Education	0.515	(0.362 , 0.673)	0.044	(0.033 , 0.054)
Competed Vocational Education	0.262	(0.112 , 0.415)	0.025	(0.011 , 0.037)
From a Regional Center	-0.080	(-0.190 , 0.031)	-0.010	(-0.025 , 0.004)
From a Rural Area	0.120	(0.009 , 0.231)	0.012	(0.000 , 0.023)
Socio-Economic Index: 2nd Decile	0.159	(-0.010 , 0.329)	0.015	(-0.002 , 0.030)
Socio-Economic Index: 3rd Decile	0.300	(0.126 , 0.475)	0.027	(0.013 , 0.040)
Socio-Economic Index: 4th Decile	0.348	(0.168 , 0.528)	0.031	(0.017 , 0.043)
Socio-Economic Index: 5th Decile	0.358	(0.182 , 0.535)	0.032	(0.018 , 0.044)
Socio-Economic Index: 6th Decile	0.582	(0.389 , 0.780)	0.045	(0.034 , 0.055)
Socio-Economic Index: 7th Decile	0.431	(0.254 , 0.609)	0.037	(0.024 , 0.048)
Socio-Economic Index: 8th Decile	0.368	(0.189 , 0.546)	0.032	(0.018 , 0.044)
Socio-Economic Index: 9th & 10th Decile	0.522	(0.362 , 0.679)	0.048	(0.035 , 0.060)
Has at least one Physical Illness	-0.025	(-0.137 , 0.087)	0.000	(-0.010 , 0.011)
Has at least one Anxiety Disorder(s)	-0.151	(-0.297 , -0.004)	-0.020	(-0.042 , 0.000)
Has at least one Affective Disorder(s)	-0.240	(-0.386 , -0.092)	-0.034	(-0.059 , -0.012)
Has at least one Substance Disorder(s)	-0.334	(-0.457 , -0.210)	-0.052	(-0.073 , -0.032)

The base person is a female, aged 18-24, not completed secondary school, from an urban area with the 1st(lowest) SEIFA Decile, with no children, not studying, and no physical or mental illnesses.

Table 5.11: *Modeling the Probability of Working in Occupational Category (A)*

(Mean Values)

Variable	Mean Coefficient	Mean Marginal Effect: Elementary Skill	Intermediate Skill	Advanced Skill	Associate Professionals	Professionals
Constant	-0.268					
Age 25-44	0.331	-0.020	-0.021	-0.005	0.009	0.037
Age 45-64	0.313	-0.030	-0.027	-0.005	0.014	0.049
Male	0.309	-0.072	-0.047	0.000	0.034	0.086
Competed Secondary School	0.219	-0.039	-0.027	-0.001	0.018	0.049
Competed Higher Education	1.543	-0.200	-0.228	-0.105	0.009	0.524
Competed Vocational Education	0.791	-0.129	-0.132	-0.041	0.041	0.261
From a Regional Center	0.032	-0.009	-0.006	0.000	0.004	0.010
From a Rural Area	0.270	-0.053	-0.041	-0.004	0.025	0.073
Socio-Economic Index: 2nd Decile	0.157	-0.025	-0.020	-0.002	0.012	0.035
Socio-Economic Index: 3rd Decile	0.132	-0.014	-0.013	-0.002	0.006	0.022
Socio-Economic Index: 4th Decile	0.181	-0.024	-0.020	-0.003	0.011	0.036
Socio-Economic Index: 5th Decile	0.252	-0.037	-0.030	-0.005	0.017	0.055
Socio-Economic Index: 6th Decile	0.225	-0.027	-0.024	-0.005	0.012	0.043
Socio-Economic Index: 7th Decile	0.308	-0.046	-0.039	-0.006	0.021	0.071
Socio-Economic Index: 8th Decile	0.338	-0.050	-0.042	-0.007	0.022	0.078
Socio-Economic Index: 9th & 10th Decile	0.398	-0.062	-0.051	-0.007	0.028	0.093
Has at least one Physical Illness	-0.062	-0.008	-0.003	0.002	0.004	0.006
Has at least one Anxiety Disorder(s)	-0.093	0.016	0.011	0.000	-0.008	-0.019
Has at least one Affective Disorder(s)	-0.057	0.008	0.006	0.001	-0.004	-0.011
Has at least one Substance Disorder(s)	-0.084	0.024	0.015	-0.001	-0.012	-0.027

The base person is a female, aged 18-24, not completed secondary school, from an urban area with the 1st(lowest) SEIFA Decile, with no children, not studying, and no physical or mental illnesses.

Table 5.12: *Modeling the Probability of Working in Occupational Category (B)*

(95% Credible Interval)

Variable	Coefficient	Marginal Effect Elementary Skill	Intermediate Skill	Advanced Skill	Associate Professionals	Professionals
Constant	(-0.350,-0.185)					
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)
Competed 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)
Competed 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)
Competed 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)
Socio-Economic Index: 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)
Socio-Economic Index: 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)
Socio-Economic Index: 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)
Socio-Economic Index: 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)
Socio-Economic Index: 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)
Socio-Economic Index: 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)
Socio-Economic Index: 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)
Socio-Economic Index: 9th & 10th Decile	(0.322, 0.474)	(-0.078,-0.047)	(-0.064,-0.038)	(-0.010,-0.005)	(0.021, 0.035)	(0.069, 0.117)
Has at least one 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)
Has at least one Anxiety Disorder(s)	(-0.169,-0.018)	(-0.003, 0.036)	(0.000, 0.021)	(-0.002, 0.001)	(-0.017, 0.001)	(-0.038, 0.001)
Has at least one Affective Disorder(s)	(-0.137, 0.023)	(-0.012, 0.028)	(-0.006, 0.018)	(-0.001, 0.002)	(-0.014, 0.006)	(-0.032, 0.011)
Has at least one Substance Disorder(s)	(-0.154,-0.015)	(0.006, 0.042)	(0.005, 0.024)	(-0.002, 0.001)	(-0.021,-0.003)	(-0.044,-0.009)

The base person is a female, aged 18-24, not completed secondary school, from an urban area with the 1st(lowest) SEIFA Decile, with no children, not studying, and no physical or mental illnesses.

higher levels of skill categories and less likely to be employed in elementary and intermediate skill categories. Males are more likely to get a job as associate professionals and professionals, and less likely to be employed in lower skill categories than females. People with tertiary education are most likely to be employed as professionals, with a marginal effect of 52.4 percentage points, and they are least likely to be employed in lower-skilled categories. The coefficient and marginal effects of the variable about 'From a Regional center' are not significant again. Meanwhile, people from rural areas have larger propensities to get jobs in higher-skilled levels and less likely to have lower-skilled jobs. Based on results for the SEIFA indices, more advanced areas normally result in higher possibilities to enter associate professional and professional categories and less probability to work in the other three lower categories. Although the coefficient of physical condition is negative in the third equation, most of its marginal effects are insignificant. The only significant 95% credible interval is the one for advanced skill category which is (0.001, 0.002). As a result, physical conditions have little influence on occupation choice. The coefficient of anxiety disorders is also negative, but most of its credible intervals contain zero except the one for intermediate skill category. Therefore, anxiety disorders do not have much influence on skill categories. For affective disorders, all 95% credible intervals contain zero, so it does not have any effect on skill levels. It seems substance disorders will reduce the probability of being employed in higher occupational levels and increase the chance of working in elementary and intermediate levels.

Most results are consistent with those found by Cornwell et al., but there are differences in the marginal effects of mental illness, especially on their significance levels. Considering participation, Cornwell et al. find a negative effect of anxiety disorders and weaker effect on affective disorders, while that for substance disorders is not significant. In contrast, Table 5.9 indicates that marginal effects of both anxiety and affective disorders are not significant while substance disorders has a positive effect on participating in the labour force. When it comes to employment rate, significant effects of the three mental disorders are shown in both the results of Cornwell et al. and our results, although there is slightly difference in magnitudes. For occupational skill categories, Cornwell et al. argue that coefficients of mental illness are significantly negative. Although Table 5.11 also shows the three coefficients are all negative, 95% credible intervals in Table 5.12 reveal that the coefficient of affective disorders is insignificant as are the marginal effects of this mental illness. Similarly, even though the coefficient of anxiety disorders is significant, most marginal effects of this variable are insignificant. The two-step method applied by Cornwell et al. does not consider much about the correlation between the equations, while the indirect effects from other two equations can be transferred via error correlation to eliminate the direct negative effects from the third equation. This may explain why our results depart from those of Cornwell et al.

5.7 Conclusion

This chapter has discussed a double-hurdle model with sample selection existing in each hurdle. Two Bayesian samplers have been proposed. While Sampler A is derived in a standard way, Sampler B is constructed from reparameterization to improve the convergence rate of Gibbs algorithms. MCMC convergence diagnosis in Section 5.3.4 and 5.4.4 has shown that both samplers can get convergent results for one simple simulation sample, although Sampler B appears superior.

The two samplers are then compared with classic FIML estimation by a simulation study. Monte Carlo results reveal that FIML fails to offer reliable estimation more than half the time, while the Bayesian method can provide estimates for each sample. The results show that both Bayesian samplers can give accurate and efficient estimation. However, the overall performance of Sampler B is better than Sampler A, because Sampler B can produce less biased estimates and faster convergence for the 1,000 samples, especially when the error correlation is large.

When the double-hurdle model is applied on empirical work to discuss the relationship between mental illness and participation, employment and occupational skill categories in the labour market, the Bayesian method has shown great superiority in estimating the model compared to the two-step method and FIML. While the covariance matrix cannot be obtained for FIML and the two-step method cannot give consistent estimates, sampler B can successfully provide convergent results. In addition, more precise marginal effects can be derived since some indirect effects can be possibly evaluated through estimates of coefficients of error correlation.

In summary, the proposed Bayesian samplers can achieve convergent estimation when classic FIML and the two-step method cannot give reliable results. Besides, the Bayesian method provides some very useful information on the analysis of marginal effects, like the distributions of marginal effects, which cannot be done by other estimation methods. Moreover, the Bayesian inference in this chapter can be easily extended to construct samplers for those complicated models with more equations, regardless of whether they are linear or non-linear.

Chapter 6

Conclusion

This thesis has been concerned with nonlinear models that include issues with sample selection and endogeneity which are prevalent in econometric analysis of cross sectional data. We start with a bivariate probit model with an endogenous dummy regressor, followed by an ordered probit model with binary non-random selection. Then, we extend the two-equation system to a double-hurdle model with two levels of sample selection before an ordered outcome is observed. The main methodologies involved are the Bayesian method, FIML and various two-step methods. This thesis focuses on the derivation of Bayesian samplers and the comparison between the Bayesian method and other methods for each specific model. Overall, the Bayesian method is quite competitive and shows great superiorities in estimating the double-hurdle model.

6.1 Key Findings

The main contributions of this thesis can be summarized as follows: (i) the introduction of the Bayesian method to investigate the effects of exclusion restrictions on bivariate probit models with an endogenous dummy regressor; (ii) the development of the idea of reparameterization to derive the Bayesian sampler for ordered probit models with binary sample selection rules; (iii) the proposal of the likelihood-based two-step method which can be an alternative method to FIML in estimating nonlinear models with sample selection; (iv) the proposal of Bayesian samplers to estimate double-hurdle models when other methods cannot offer reliable estimation; and (v) Monte Carlo studies which are used to compare different estimation methods for each model.

Here we will outline our main findings in detail across the chapters. Chapter 2 briefly reviews current literature about models of sample selection and endogeneity. Models are discussed from the initial sample selection model to nonlinear models with selectivity, and extended to double hurdle models with one extra level of sample selection. The endogenous treatment effect model is discussed with further extension to bivariate and multivariate probit models with endogeneity. MLE and the two-step method are two main methods used to estimate such models. Applications of those models are also commented on in this chapter.

Chapter 3 investigates the effects of exclusion restrictions on estimates of bivariate probit models with an endogenous dummy regressor. To our knowledge, no Bayesian method has been included in an investigation of this topic, and the chapter includes a Monte Carlo comparison to other methods. The Bayesian sampler described in this chapter is equivalent to Chib & Greenberg (1998)'s sampler but is derived by decomposition to form an easier sampling process. The research questions are around how the Bayesian method and the inconsistent two-step method perform in contrast to MLE, whether exclusion restrictions are important in providing reliable parameter estimates, and what factors affect model identification in cases with or without exclusion restrictions. To answer the first question, MLE is always accurate and efficient except for encountering some convergence problems. The MCMC method is the only method that can estimate each sample and is most efficient with small error correlation, but is quite biased with large correlation. The straightforward two-step method has less convergence problem than MLE, but is inconsistent when the error correlation is large. For the second question, Monte Carlo results show that estimation in models with exclusion restrictions is much easier than in cases without exclusion restrictions. This is despite the fact that exclusion restrictions are not required for model identification once other factors are satisfied. Wilde (2000) gives an example with only one dummy regressor and argues that it can be identified in theory. However, we show that this model cannot be accurately estimated by any of three methods. Nevertheless, model identification can be improved by increasing the number of dummy exogenous regressors. Meanwhile, Monte Carlo results verify Wilde's other comments about the variation of explanatory variables, which is important to reinforce model identification.

In Chapter 4, an ordered probit model with a binary selection rule is estimated by four estimators: FIML, the two-step method, the likelihood-based two-step method and the Bayesian method. The MCMC sampler is proposed with conjugate conditional posteriors after reparameterization and specially designed priors, while the likelihood-based two-step method is proposed as an alternative estimation method. The effects of exclusion restrictions are also studied on this model with comparison between the four methods. Monte Carlo results reveal that the absence of exclusion restrictions will not result in serious identification problems. Simulation results also show that both FIML and the Bayesian method give unbiased and efficient estimation. Meanwhile, the likelihood-based two-step method has less convergence problem than FIML, while the two-step method causes bias and inefficiency when there is strong error correlation. The model is applied to an empirical study about mental health and labour force employment in which exclusion restrictions are not available. It is estimated by FIML and the Bayesian method, followed by some inference of marginal effects in terms of mental health variables. Convergence diagnosis is emphasized on the Bayesian method to make sure the sampled paths can be used to illustrate distributions of model parameters and marginal effects.

A double-hurdle model of labour market outcomes is discussed in Chapter 5. This particular model is somehow different from the traditional double-hurdle model, because it contains two levels of sample selection in which the level deciding partly observed ordered outcomes is partly observed itself. FIML is used to get model estimates as well as two Bayesian samplers. One sampler is derived from a standard procedure, when the other is an extension of the sampler in Chapter 4 with reparameterization. The Bayesian method displays more advantages in estimating this model than FIML, in both the simulation study and the empirical study. The Monte Carlo study indicates that FIML fails to converge for more than half the cases, although the available estimates are precise and efficient. Meanwhile, each sample can be estimated by both Bayesian samplers whose overall performances are as good as FIML. When comparing the two samplers, the sampler with reparameterization has a better convergence rate for some parameters. And the better convergence comes with more accurate estimation especially when there is a high correlation between the errors. The model as well as the three estimators are used to study an application about effects of mental health on labor market outcomes including participation, employment and occupational skill categories. In this example, FIML cannot give reliable estimation and the standard sampler has some convergence issues. Thus, estimates of the sampler with reparameterization are used for further analysis about marginal effects.

6.2 Limitations

We discuss two limitations with the research reported in this thesis. Firstly, the main issue about MLE is the convergence problem. A reason to prefer Bayesian methods is its superior convergence reliability. That is why more work could have gone into methods of numerical optimization for MLE. For example, we found that CML package in GAUSS and NLOGIT have different convergence performance. When the same model is estimated based on the same data, in some cases, one software package gives convergent results while the other cannot. As a result, the algorithms for numerical optimization in different softwares may need to be considered.

Secondly, the Bayesian analysis with data augmentation cannot be used for weighted estimation, where estimation considers the effects of weights of samples and a weight represents number of individuals with the same features of each sample in the population. For the MLE method, weights can be easily included by adding polynomial terms to the log likelihood function. According to the Bayesian method, a latent variable needs to be sampled for each individual. Including weights will greatly increase the number of potential latent variables. For example, 1000 weights for a single individual will mean 1000 latent variables to be simulated. Therefore, the use of weights will result in some computational burdens to Bayesian algorithms. So future work is needed to make it possible to include weights in the Bayesian method.

6.3 Future Work

Bayesian analysis has been used in this thesis, and compared with classical MLE methods and extensions of the two-step methods. It generally shows great superiorities especially in estimating the double-hurdle models when exclusion restrictions are not available, as it has far fewer difficulties with estimation. The future work we are interested in will be to focus on other applications of the Bayesian methods.

The Bayesian framework can be applied to any multivariate models with sample selection or endogenous issues. In many cases MLE methods tend to have more serious convergence problems and the two-step method has obvious flaws because of inconsistency of the estimator. As in the models discussed in this thesis, the idea of data augmentation can be easily used to estimate latent variables in multivariate models.

Another issue that could be addressed in future is the convergence rate of the Bayesian algorithms. In Chapter 5, we have shown that reparameterization can improve the convergence rate of the Gibbs sampling in some ways. Other approaches may be needed to further improve the overall convergence rate of simulators.

The normality assumption is made for the errors in all three models discussed in this thesis. It would be useful to consider other assumptions. One may also combine the idea of a non-parametric method with the Bayesian method, so there is no need for specific assumptions about the error distributions.

Bibliography

Albert, J. H. & Chib, S. (1993), 'Bayesian analysis of binary and polychotomous response data', *Journal of the American Statistical Association* **88**, 669–679.

Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press, New York.

Angrist, J. (2001), 'Estimation of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical practice', *Journal of Business and Economic Statistics* **19**, 2–16.

Arendt, J. N. & Holm, A. (2006), 'Probit models with binary endogenous regressors', *Discussion Papers on Business and Economics* **4**.

Arendt, J. N. & Holm, A. (2007), Probit models with dummy endogenous regressors, Working paper, University of Southern Denmark.

Artazcoz, L., Benach, J., Borrell, C. & Cortes, I. (2004), 'Unemployment and mental health: understanding the interactions among gender, family roles, and social class', *American Journal of Public Health* **94**, 82.

Belkar, R. & Fiebig, D. G. (2008), 'A Monte Carlo comparison of estimators for a bivariate probit model with selection', *Mathematics and Computers in Simulation* **78**, 250–256.

Belkar, R., Fiebig, D. G., Haas, M. & Viney, R. (2006), 'Why worry about awareness in choice problems? Econometric analysis of screening for cervical cancer', *Health Economics* **15**, 33–48.

Besag, J. (1974), 'Spatial interaction and the statistical analysis of lattice systems', *Journal of the Royal Statistical Society. Series B (Methodological)* **36**, 192–236.

Bhattacharya, J., Goldman, D. & McCaffrey, D. (2006), 'Estimating probit models with self-selected treatments', *Statistics in Medicine* **25**, 389.

Blundell, R., Ham, J. & Meghir, C. (1987), 'Unemployment and female labour supply', *The Economic Journal* **97**, 44–64.

Boyes, W. J., Hoffman, D. L. & Low, S. A. (1989), 'An econometric analysis of the bank credit scoring problem', *Journal of Econometrics* **40**, 3–14.

Bryson, A., Cappellari, L. & Lucifora, C. (2004), 'Does union membership really reduce job satisfaction?', *British Journal of Industrial Relations* **42**, 439–459.

Burnett, N. J. (1997), 'Gender economics courses in liberal arts colleges', *The Journal of Economic Education* **28**, 369–376.

Butler, J. S. & Moffitt, R. (1982), 'A computationally efficient quadrature procedure for the one-factor multinomial probit model', *Econometrica* pp. 761–764.

Chib, S. (1992), 'Bayes inference in the Tobit censored regression model', *Journal of Econometrics* **51**, 79–99.

Chib, S. & Greenberg, E. (1998), 'Analysis of multivariate probit models', *Biometrika* **85**, 347.

Chib, S. & Hamilton, B. (2000), 'Bayesian analysis of cross-section and clustered data treatment models', *Journal of Econometrics* **97**, 25–50.

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**, 107.

Cowles, M. (1996), 'Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models', *Statistics and Computing* **6**, 101–111.

Cragg, J. G. (1971), 'Some statistical models for limited dependent variables with application to the demand for durable goods', *Econometrica* **39**, 829–844.

Daykin, A. R. & Moffatt, P. G. (2002), 'Analyzing ordered responses: A review of the ordered probit model', *Understanding Statistics* **1**, 157–166.

Deadman, D. & MacDonald, Z. (2004), 'Offenders as victims of crime?: an investigation into the relationship between criminal behaviour and victimization', *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **167**, 53–67.

Deaton, A. (1997), *The analysis of household surveys: a microeconometric approach to development policy*, World Bank Publications.

Deaton, A. & Irish, M. (1984), 'Statistical models for zero expenditures in household budgets', *Journal of Public Economics* **23**, 59–80.

Dockery, A. M. (2006), Mental health and labour force status: panel estimates with four waves of HILDA, Working paper, Centre for Labour Market Research.

Dubin, J. A. & Rivers, D. (1989), 'Selection bias in linear regression, logit and probit models', *Sociological Methods and Research* **18**, 360.

Evans, W. N. & Schwab, R. M. (1995), 'Finishing high school and starting college: do Catholic schools make a difference?', *The Quarterly Journal of Economics* **110**, 941–974.

Flatau, P., Galea, J. & Petridis, R. (2000), 'Mental health and wellbeing and unemployment', *Australian Economic Review* **33**, 161–181.

Geweke, J. (1991), Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities, *in* 'Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface', Citeseer, pp. 571–578.

Greene, W. (1992), 'A statistical model for credit scoring', *Stern School of Business, Dept. of Economics, Working Paper* pp. 92–29.

Greene, W. (2003), *Econometric analysis*, Prentice Hall Upper Saddle River, NJ.

Greene, W. (2006), A General Approach to Incorporating Selectivity in a Model, Working paper, Department of Economics, Stern School of Business, New York University.

Greene, W. H. (1998), 'Gender economics courses in liberal arts colleges: further results', *Journal of Economic Education* pp. 291–300.

Greene, W. H. (2001), 'FIML estimation of sample selection models for count data', *Economic Theory, Dynamics and Markets: Essays in Honor of Ryuzo Sato* p. 73.

Greene, W. H. (2002), *LIMDEP version 8.0: Econometric modeling guide*, Econometric Software, Plainview.

Gronau, R. (1974), 'Wage comparisons–a selectivity bias', *the Journal of Political Economy* pp. 1119–1143.

Hajivassiliou, V. A. & McFadden, D. L. (1998), 'The method of simulated scores for the estimation of LDV models', *Econometrica* **66**, 863–896.

Hamilton, V. H., Merrigan, P. & Dufresne, É. (1997), 'Down and out: estimating the relationship between mental health and unemployment', *Health Economics* **6**, 397–406.

Harris, M. N. & Zhao, X. (2007), 'A zero-inflated ordered probit model, with an application to modelling tobacco consumption', *Journal of Econometrics* **141**, 1073–1099.

Hastings, W. K. (1970), 'Monte Carlo sampling methods using Markov chains and their applications', *Biometrika* **57**, 97–109.

Hausman, J. A., Lo, A. W. & MacKinlay, A. C. (1992), 'An ordered probit analysis of transaction stock prices', *Journal of Financial Economics* **31**, 319–379.

Heckman, J. J. (1976), 'The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models', *NBER Chapters* pp. 120–137.

Heckman, J. J. (1978), 'Dummy endogenous variables in a simultaneous equation system', *Econometrica* **46**, 931–959.

Heckman, J. J. (1979), 'Sample selection bias as a specification error', *Econometrica* **47**, 153–161.

Holly, A., Gardiol, L., Domenighetti, G. & Bisig, B. (1998), 'An econometric model of health care utilization and health insurance in Switzerland', *European Economic Review* **42**, 513–522.

Holm, A. & Jæger, M. M. (2008), 'Selection bias in educational transition models: theory and empirical evidence', *Social Policy and Welfare Services Working Paper* **11**, 2008.

Jones, A. M. (1989), 'A double-hurdle model of cigarette consumption', *Journal of Applied Econometrics* **4**, 23–39.

Kim, K. (2006), 'Sample selection models with a common dummy endogenous regressor in simultaneous equations: a simple two-step estimation', *Economics Letters* **91**, 280–286.

Kim, S., Shephard, N. & Chib, S. (1998), 'Stochastic volatility: likelihood inference and comparison with ARCH models', *The Review of Economic Studies* **65**, 361–393.

Koop, G. (2003), *Bayesian econometrics*, J. Wiley Hoboken, NJ.

Labeaga, J. M. (1999), 'A double-hurdle rational addiction model with heterogeneity: estimating the demand for tobacco', *Journal of Econometrics* **93**, 49–72.

Lee, L. F. (1982), 'Some approaches to the correction of selectivity bias', *The Review of Economic Studies* **49**, 355–372.

Leung, S. F. & Yu, S. (1996), 'On the choice between sample selection and two-part models', *Journal of Econometrics* **72**, 197–229.

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.

Li, M. & Tobias, J. L. (2007), Bayesian analysis of treatment effects in an ordered potential outcomesModel, Staff general research papers 12429, Iowa State University, Department of Economics.

Maddala, G. S. (1983), *Limited-dependent and qualitative variables in econometrics*, Cambridge University Press.

Maddala, G. S. & Lee, L. F. (1976), 'Recursive models with qualitative endogenous variables', *NBER Chapters* pp. 168–188.

Manski, C. F. (1989), 'Anatomy of the selection problem', *Journal of Human Resources* **24**, 343–360.

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.

Meng, C. L. & Schmidt, P. (1985), 'On the cost of partial observability in the bivariate probit model', *International Economic Review* **26**, 71–85.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953), 'Equation of state calculations by fast computing machines', *The Journal of Chemical Physics* **21**, 1087.

Monfardini, C. & Radice, R. (2008), 'Testing exogeneity in the bivariate probit model: A Monte Carlo study', *Oxford Bulletin of Economics and Statistics* **70**, 271–282.

Munkin, M. K. & Trivedi, P. K. (2008), 'Bayesian analysis of the ordered probit model with endogenous selection', *Journal of Econometrics* **143**, 334–348.

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.

Nicoletti, C. & Peracchi, F. (2001), Two-step estimation of binary response models with sample selection, Working paper, Faculty of Economics, Tor Vergata University.

Omori, Y. (2007), 'Efficient Gibbs sampler for Bayesian analysis of a sample selection model', *Statistics and Probability Letters* **77**, 1300–1311.

Poirier, D. J. (1980), 'Partial observability in bivariate probit models', *Journal of Econometrics* **12**, 209–217.

Rivers, D. & Vuong, Q. H. (1988), 'Limited information estimators and exogeneity tests for simultaneous probit models', *Journal of Econometrics* **39**, 347–366.

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', *Markov Chain Monte Carlo in Practice* **57**, 45–58.

Sartori, A. E. (2003), 'An estimator for some binary-outcome selection models without exclusion restrictions', *Political Analysis* **11**, 111.

Sivia, D. S. & Skilling, J. (2006), *Data analysis: a Bayesian tutorial*, Oxford University Press, USA.

Tobin, J. (1958), 'Estimation of relationships for limited dependent variables', *Econometrica* **26**, 24–36.

Tsay, R. S. (2005), *Analysis of financial time series*, Wiley-Interscience.

Tse, Y. K., Zhang, X. & Yu, J. (2004), 'Estimation of hyperbolic diffusion using the Markov chain Monte Carlo method', *Quantitative Finance* **4**, 158–169.

Van de Ven, W. & Van Praag, B. M. S. (1981), 'The demand for deductibles in private health insurance', *Journal of Econometrics* **17**, 229–252.

Vella, F. (1998), 'Estimating models with sample selection bias: a survey', *Journal of Human Resources* **33**, 127–172.

Wilde, J. (2000), 'Identification of multiple equation probit models with endogenous dummy regressors', *Economics Letters* **69**, 309–312.

Winship, C. & Mare, R. D. (1992), 'Models for sample selection bias', *Annual Review of Sociology* **18**, 327–350.

Wooldridge, J. M. (2002), 'Inverse probability weighted M-estimators for sample selection, attrition, and stratification', *Portuguese Economic Journal* **1**, 117–139.

Yen, S. T. (2005), 'Zero observations and gender differences in cigarette consumption', *Applied Economics* **37**, 1839–1849.

Zhang, X., Zhao, X. & Harris, A. (2009), 'Chronic diseases and labour force participation in Australia', *Journal of Health Economics* **28**, 91–108.