CIRJE-F-882

An Econometric Analysis of Insurance Markets with Separate Identification for Moral Hazard and Selection

Shinya Sugawara
Japan Society of Promotion of Science and University of Tokyo

Yasuhiro Omori
University of Tokyo

March 2013

An econometric analysis of insurance markets with separate identification for moral hazard and selection problems

Shinya Sugawara *   Yasuhiro Omori †

Abstract

This paper proposes a simple microeconometric framework that can separately identify moral hazard and selection problems in insurance markets. Our econometric model is equivalent to the approach that is utilized for entry game analyses. We employ a Bayesian estimation approach that avoids a partial identification problem. Due to the standard identification, we propose a statistical model selection method to detect an information structure that consumers face. Our method is applied to the dental insurance market in the United States. In this market, we find not only standard moral hazard but also advantageous selection, which has an intuitive interpretation in the context of dental insurance.

1 Introduction

During the course of the previous decade, empirical studies have been rapidly catching up with highly developed economic theories of asymmetric information. This paper proposes a new microeconometric framework to analyze the information problem in insurance markets. Our main contribution to the extant literature is that we can separately identify moral hazard and selection problems. This is a clear advantage over a traditional method that represents information asymmetry as as single parameter.

*Research Fellow of Japan Society of Promotion of Science and University of Tokyo, Tokyo 113-0033, Japan E-mail: sugawara@e.u-tokyo.ac.jp

†Faculty of Economics, University of Tokyo, Tokyo 113-0033, Japan. E-mail:omori@e.u-tokyo.ac.jp.

In particular, the conventional methodology uses a bivariate probit model for consumer data, as discussed by Chiappori and Salanié (2000). The two dependent variables of this model are the purchase of insurance and the occurrence of an accident. In this approach, the standard asymmetric information problem can be detected as a positive correlation between these two dependent variables. Specifically, moral hazard implies that consumers increase their riskiness after they have purchased insurance, while adverse selection implies that riskier consumers have a stronger demand for insurance.

In addition to computational simplicity, this bivariate probit model has a nice property in data availability. This model does not require detailed contract information, which may be hard to obtain, but instead relies on variables that are commonly obtained from general household surveys. Thus, bivariate probit analysis has rapidly become a popular technique that has been applied to various insurance markets. However, most empirical investigations have failed to detect a significantly positive correlation between two dependent variables; in fact, some studies have even found a negative correlation.

In response to these unexpected results, De Meza and Webb (2001) proposed an important alternative theory to adverse selection. This theory, which is known as advantageous selection, states that less risky individuals have a stronger demand for insurance, due to their risk aversive preference. In contrast to adverse selection, advantageous selection produces a negative effect of risk on an insurance purchase. Thus, the conventional bivariate probit approach may not be applicable if both moral hazard and advantageous selection are present because the distinct and conflicting effects of these two factors cannot be captured by a single correlation parameter. To assess the validity of such a theory, one must separately identify the moral hazard and the selection problems(which mean either adverse selection or advantageous selection). This task is beyond the scope of the bivariate probit model.

In this paper, we present a new approach that permits the separate identification of moral hazard and selection problems. We begin our econometric modeling with adding two terms to the bivariate probit model. These terms allow for moral hazard and selection problems to be measured as two distinct coefficient parameters rather than as a correlation between two dependent variables. Despite the simple appearance of our approach, our model specification encounters statistical difficulty due to the existence of mutual dependencies between the dependent variables. It is shown that we cannot separately identify two coefficients without an additional assumption.

Thus, to overcome this identification problem, we assume the simultaneous determination of the two dependent variables. This simultaneity assumption introduces an econometric problem that has been analyzed in the literature of nonlinear simultaneous equation models with limited dependent variables, such as Amemiya (1975), Heckman (1978) and Gouriéroux and Monfort (1979). These works yielded a statistical problem that is referred to as incoherency by Gouriéroux et al. (1980); this problem involves the fact that probabilistic models are not well-defined without a restrictive assumption on parameter values.

To handle the incoherency problem, we adopt an approach that was proposed by Tamer (2003) and Ciliberto and Tamer (2009) in the literature on empirical analyses of entry games. These studies introduced a latent variable that formulates a well-defined probabilistic model. This variable is called a selection rule, because it characterizes players' choices among multiple Nash equilibria. However, this approach creates the incidental parameter problem (Lancaster, 2000), because it is difficult to construct a consistent estimator for parameters that does not depend on the sample-specific selection rule. To overcome the incidental parameter problem, Ciliberto and Tamer (2009) constructed a moment inequality using boundary conditions of the selection rule. Because the resulting econometric model is partially identified, they employed a set estimation method of Chernozhukov et al. (2007).

This paper adopts a Bayesian approach that we introduced in a prior study (Sugawara and Omori, 2012). A clear distinction between our Bayesian approach and previous classical methods is that we explicitly estimate the sample-specific variable. Because the Bayesian estimation approach can work with finite samples, the sample-specific selection rule is estimable. The lack of the incidental parameter problem enables us to construct a standard likelihood function rather than a moment inequality. Thus, our methodology achieves standard identification, and therefore permits the use of standard inference techniques.

To demonstrate the advantage of our Bayesian methodology, we propose a way to answer the empirically important question of what type of an information structure consumers may face. We show that distinct information structures correspond to non-nested statistical models. This finding indicates that this question is to be answered by statistical model selection. However, there has not yet been invented a classical model selection procedure for our model. On the other hand, our Bayesian estimation is accompanied with standard model selection techniques.

To provide an empirical application of our approach, we analyze the US dental insurance market. Our model selection results indicate that moral hazard and advantageous selection are both present in this market. These findings are intuitively interpreted as an indication that for dental care, early preventive concerns both reduce risk and stimulate insurance demand. The detection of advantageous selection reveals an advantage of our methodology because this result cannot be derived from conventional bivariate probit analysis.

Our study relates to three literatures; the econometrics of insurance markets, Bayesian statistics and empirical analyses of dental care. First, with respect to econometric analyses of insurance markets, this paper supplements two recent streams of research. One of these research streams discusses the structural estimation approach, such as Cardon and Hendel (2001) and Einav et al. (2010b), and was summarized by Einav et al. (2010a). In contrast to the reduced form approach of the bivariate probit model, the structural approach explicitly models moral hazard and selection problems to achieve the separate identification. This body of literature continues to grow because many variations of structural models can exist.

Another growing insurance-related literature is reduced form analyses that are specialized in the examination of advantageous selection. These studies generally involve a binary choice analysis in which advantageous selection is detected as an effect of individual's risk aversion on the insurance purchase dummy. To measure this risk aversion, researchers have employed a variety of explanatory variables, such as subjective mortality rate by Cawley and Philipson (1999), seat belt use by Finkelstein and McGarry (2006), and health status and schooling levels, which serves as proxies for cognitive ability and financial numeracy, by Fang et al. (2008). No consensus has yet been reached with respect to the appropriate choice of an explanatory variable among the many candidates.

We believe that our study can complement these recent literatures. Our methodology proposes a simple model using commonly available data, while the structural approaches analyze more sophisticated models using detailed but costly data. Furthermore, our model requires only data regarding the occurrence of an accident, which is commonly available information, as an explanatory variable to characterize the selection problem. Therefore, one can employ a preliminary analysis using our simple method, to make an appropriate choice among enormous candidates of models and explanatory variables.

Second, with respect to Bayesian statistics, there is a growing literature on partially identified models. Theoretically, Moon and Schorfheide (2012) and Kitagawa (2012) provided theoretical comparisons of the classical and Bayesian estimators. For the technical concern, Liao and Jiang (2010) proposed an estimation procedure that applies the Bayesian version of the method of moments in Kim (2002) and Chernozhukov and Hong (2003). Unlike the quasi-Bayesian approach of Liao and Jiang (2010), our methodology is based on a standard likelihood function that is available at the cost of a distributional assumption. Because the classical set estimation also requires the distributional assumption for our model, this assumption seems harmless.

Third, there have been several empirical studies about the information problem in the US dental insurance market. Previous studies have consistently detected moral hazard in various datasets from this market, such as Mueller and Monheit (1988) fro a surveyed dataset and Manning et al. (1985) for an experimental dataset. On the other hand, selection problems have been regarded as a non-negligible but a difficult concept to be identified, due to the incoherency problem (Sintonen and Linnosmaa, 2000). An exceptional study is Munkin and Trivedi (2008), which we deeply owe in our construction of an empirical study.

The organization of this paper is as follows. In Section 2, we describe relevant econometric models. Section 3 considers the corresponding inferential frameworks. The proposed method is applied to the US dental care insurance market in Section 4. Section 5 concludes the paper.

2 The modeling of asymmetric information in insurance markets

This section presents econometric models for information asymmetry in insurance markets using consumer data. We begin with the conventional methodology which utilizes the bivariate probit model. Then we proceed to describe our own framework which allow for the separate identification of moral hazard and selection problems.

2.1 A conventional modeling: The bivariate probit model

Conventionally, information asymmetry is captured with a bivariate probit model (Poirier, 1980). This line of studies targets a market of a simple insurance policy that covers an accident in the following simple reduced-form framework.

The sample consists of N consumers who are indexed as $i = 1, 2, ..., N$. The dependent variables are two observable binary variables $\boldsymbol{y}_i = (y_{i1}, y_{i2})'$, in which y_{i1} and y_{i2} represent the purchase of insurance and the occurrence of an accident, respectively, for the ith consumer. Further, let y^*_{ij} denote the corresponding latent variable for the ith consumer. We assume that y_{ij} takes unity if y^*_{ij} is non-negative and that it takes zero otherwise. These latent variables are assumed to be linear functions of K_j-dimensional observable regressors $\boldsymbol{x}_{ij}$ with coefficients $\boldsymbol{\beta}_j$ and an error term ϵ_{ij}.

Bivariate Probit Model:

$$y^*_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta}_j + \epsilon_{ij},\ j = 1, 2, \tag{2.1}$$

$$y_{ij} = I[y^*_{ij} \geq 0]. \tag{2.2}$$

In this bivariate probit model, we can detect standard asymmetric information as a positive correlation, namely $\rho > 0$, between y^*_{i1} and y^*_{i2} conditional on $\boldsymbol{x}_i = (\boldsymbol{x}'_{i1}, \boldsymbol{x}'_{i2})'$. Specifically, moral hazard indicates that consumers become riskier after they purchase insurance, while adverse selection implies that riskier consumers are more likely to purchase insurance. The above framework incorporates an implicit assumption that only asymmetric information is the source of ρ and no omitted variable affects both latent variables.

2.2 A new modeling to separately identify moral hazard and selection problems

2.2.1 A basic parametrization and its lack of identification

Next, we present our original econometric methodology, which separately identifies the effects of moral hazard and selection problems. To achieve the separate identification for these two sources of information asymmetry, we introduce two additional elements into the bivariate probit model in (2.1) and (2.2). One element measures the effect of the insurance purchase y_{i1} on the accident occurrence y_{i2}, and another element measures the effect of y_{i2} on y_{i1}. The former element indicates moral hazard, while the latter element represents selection problems. For simplicity, these elements are modeled as linearly additive terms with coefficient parameters α_2 and α_1. In other words,

$$y^*_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta}_j + \alpha_j y_{ik} + u_{ij} \quad j = 1, 2, \ k \neq j, \tag{2.3}$$

$$y_{ij} = I[y^*_{ij} \geq 0]. \tag{2.4}$$

The information structure is now formulated in terms of the signs of α_1 and α_2. A positive α_2 corresponds to the existence of moral hazard, while a positive or negative α_1 indicates the existence of adverse or advantageous selection, respectively.

As a consequence of the assumption with the bivariate probit that no omitted variables exist given $\boldsymbol{x}_{ij}$, the new error term u_{ij} is assumed to be independent of u_{ik}, given $\boldsymbol{x}_i$. This assumption indicates that we do not allow for the endogeneity of u_{ij} and y_{ik}, and u_{ij} can affect y_{ik} only via y_{ij}. This assumption is reconsidered in the later empirical portion of this paper to assess the robustness of the robustness of the proposed approach.

Despite a simple appearance, we cannot employ a straight-forward estimation procedure for model parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \alpha_1, \alpha_2)'$. The problem is intuitively summarized as follows: In (2.3) and (2.4), a dependent variable y_{i1} appears on the right-hand side of the equation for y_{i2}, while y_{i2} appears in the right-hand side of the equation for y_{i1}. Despite these mutual dependencies, we wish to identify two-way interactions, α_1 and α_2, using only one observed pair of dependent variables. Because of this essential lack of observations, the separate identification cannot be achieved without an additional assumption. This intuitive argument was formally proved in the spatial statistics literature Besag (1974), and I summarize this proof in Appendix A.

2.2.2 A modeling with a simultaneity assumption

To deal with the mutual dependencies, this paper assumes that equations (2.3) and (2.4) simultaneously hold. Under this assumption, simple calculations yield the following relationships between outcomes and error terms:

$$z_i = 1 \quad \Leftrightarrow \quad \big\{u_{i1} < -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1, \quad u_{i2} < -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2\big\}, \tag{2.5}$$

$$z_i = 2 \quad \Leftrightarrow \quad \big\{u_{i1} \geq -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1, \quad u_{i2} < -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 - \alpha_2\big\}, \tag{2.6}$$

$$z_i = 3 \quad \Leftrightarrow \quad \big\{u_{i1} < -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 - \alpha_1, \quad u_{i2} \geq -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2\big\}, \tag{2.7}$$

$$z_i = 4 \quad \Leftrightarrow \quad \big\{u_{i1} \geq -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 - \alpha_1, \quad u_{i2} \geq -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 - \alpha_2\big\}. \tag{2.8}$$

For notational simplicity, we define a scalar variable $z_i = 1$, 2, 3 or 4 that corresponds to $(y_{i1}, y_{i2}) = (0, 0)$, $(1, 0)$, $(0, 1)$ or $(1, 1)$, respectively.

The simultaneity assumption can overcome the mutual dependencies of dependent variables, but this assumption introduces another econometric problem that Gouriéroux et al. (1980) referred to as incoherency. In our situation, the incoherency problem indicates that the probabilistic model is not well-defined unless there exists j such that $\alpha_j = 0$. To observe this issue, in Figure 1, we illustrate the above correspondence (2.5) - (2.8) with assuming that $\alpha_j \neq 0$ for neither $j = 1$ nor $j = 2$ on the coordinates of (u_{i1}, u_{i2}).

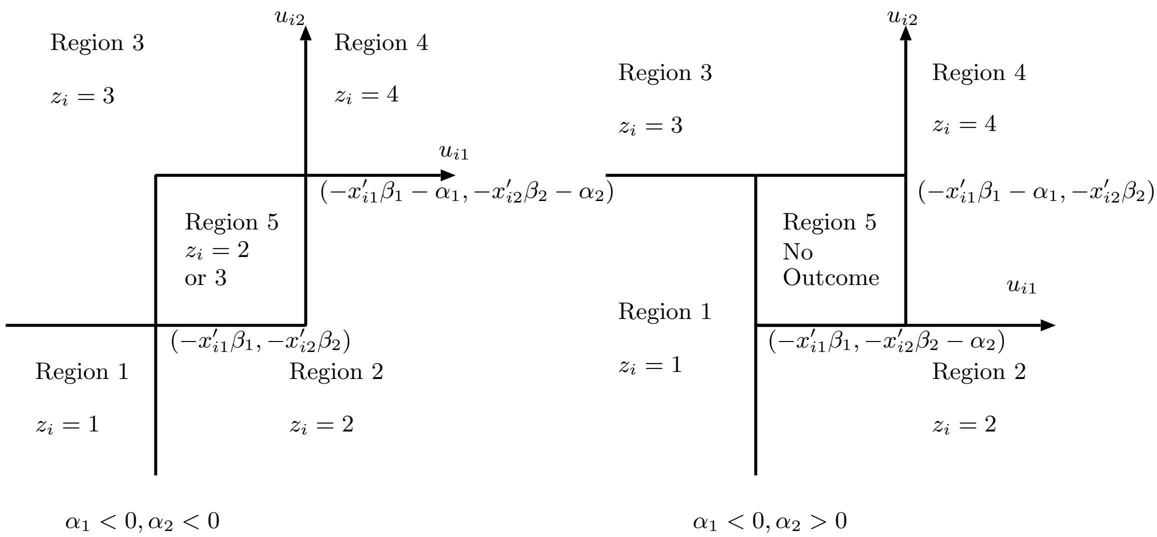

Figure 1: Data generating process for basic parametrization in (u_{i1}, u_{i2}) coordinates

Figure 1 represents two cases, namely ($\alpha_1 < 0$, $\alpha_2 < 0$) in the left side of the figure and ($\alpha_1 < 0$, $\alpha_2 > 0$) in the right side of the figure. The remaining two sign conditions, ($\alpha_1 > 0$, $\alpha_2 > 0$) and ($\alpha_1 > 0$, $\alpha_2 < 0$), can yield similar econometric models and are therefore not shown in the figure. In both of the depicted cases, Regions 1 thorough 4 have a unique pair of outcomes, but Region 5 does not. In particular, for ($\alpha_1 < 0$, $\alpha_2 < 0$), there are two candidate solutions $(y_{i1}, y_{i2}) = (1, 0)$ or $(0, 1)$, while for ($\alpha_1 < 0$, $\alpha_2 > 0$), there is no possible pair of outcomes. Because Region 5 does not produce unique outcomes, it is impossible to obtain well-defined joint choice probabilities unless Region 5 vanishes by assuming the coherency condition, namely

$\exists j, \alpha_j = 0$.

Recently, this incoherency problem has gathered a new attention, because (2.3) and (2.4) are equivalent to the best response functions for a general class of entry games. To address the incoherency problem, researchers who study entry games explicitly modeled a data generating process in Region 5 using sample-specific parameters $\boldsymbol{p}_i = (p_{i1}, p_{i2}, p_{i3}, p_{i4}) \in [0,1]^4$ with $\sum_{l=1}^{4} p_{il} = 1$, which are called selection rules. Each selection rule p_{il} represents a proportion of Region 5 within which each pair of outcomes is realized. Consequently, the choice probability can be written as follows:

$$\Pr[z_i = l | \boldsymbol{x}_i, \boldsymbol{\theta}, p_{il}] = P_l(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{il} P_5(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{2.9}$$

where $P_l(\boldsymbol{x}_i, \boldsymbol{\theta})$ measures an area of Region l in Figure 1. Because the choice probabilities for z_i = 1, 2, 3 and 4 sums to unity, the suggested approach has successfully created a well-defined probabilistic model.

2.2.3 The justification of the simultaneity assumption

The simultaneity assumption requires additional restrictions on the underlying economic model. A natural candidate of for this type of behavioral models is the following individual optimization process.

Before consumers actually make an insurance purchase, they seek their best behavior using a feedback loop. To begin with, these consumers receive hypothetical draws for the error terms u_{i1} and u_{i2}. They determine a hypothetical value for y_{i1}, whether to purchase insurance, given either value of y_{i2}. Then, these consumers evaluate their hypothetical predictions for the future occurrence of an accident, y_{i2}, based on this insurance purchase. Furthermore, they reconsider their purchasing decision based on this updated prediction. This feedback loop continues until consumers reach a steady state in which equations (2.3) and (2.4) are simultaneously satisfied. This equilibrium behavior is realized as consumers' final decisions to purchase insurance.

In entry game models, the incoherency problem is caused by the possible lack of a uniqueness of Nash equilibrium. In our context, the incoherency can be interpreted as an issue that corresponds to the non-uniqueness of the best action. If one draws hypothetical values of u_{i1} and u_{i2} that are located in Region 5, there is no unique steady state for the feedback loop. In this region, the case of $(\alpha_1 < 0,\ \alpha_2 < 0)$ involves multiple possible combinations of the insurance purchase decision and the corresponding accident occurrence, while there is no best action for the case of $(\alpha_1 < 0,\ \alpha_2 > 0)$.

3 Bayesian inferential technique

This section is concerned with a Bayesian inferential procedure. We implement our estimation using the Markov chain Monte Carlo(MCMC) algorithm, whose technical details are provided by Sugawara and Omori (2012).

3.1 Properties of Bayesian estimation

Our Bayesian methodology differs from the previous approach to the identification status. In the previous classical approach, the sample specific selection rule produces partial identification. On the other hand, in our Bayesian approach, the selection rule parameter is explicitly estimated. This approach is feasible because Bayesian estimation can work with finite samples. Thus, our method can avoid both the incidental parameter problem and the partial identification problem.

In spite of the clear difference in identification status between the classical and Bayesian approaches, underlying statistical setup of the Bayesian approach is similar to the setup for a moment inequality. To express the lack of an economic theory regarding the selection rule, we put a uniform prior distribution over $[0, 1]$ to the selection rule parameter. Besides, the likelihood function can only provide information from one sample with respect to this sample specific variable. Due to this small sample size, we should obtain a posterior distribution that might vary greatly from the uniform prior for the selection rule. Thus, the Bayesian approach should employ a similar setup to the moment inequality, in which the selection rule is equal likely to be any value on $[0, 1]$.

A disadvantage of the Bayesian method is that it requires a distributional assumption. Unlike the classical estimation which is based only on moment conditions, Bayesian estimation requires a distributional assumption to ensure the existence of a well-defined likelihood function[1]. However, the classical estimation approach of Ciliberto and Tamer (2009) also requires a distributional assumption because their estimator was defined via simulations. Therefore, this assumption does not appear to be greatly harmful to the proposed approach in our situation.

3.2 Model selection for assessing an information structure

An advantage of our method is the flexibility of inferences. Because our approach features standard identification, out Bayesian estimation is accompanied with standard inference techniques; several of these techniques have not yet been created in the classical set estimation literature. Specifically, we use model selection in a key empirical analysis of this paper.

In empirical analyses of insurance markets, it is important to detect the information structure that consumers face. In (2.3) and (2.4), we formulate the information structure as sign conditions for (α_1, α_2). However, it is difficult to construct a t-type statistics to detect the sign conditions, because distinct sign conditions for (α_1, α_2) produce non-nested statistical models.

To observe this fact, we demonstrate that choice probabilities have different functional forms under distinct sign conditions. To prove this argument, we concentrate on the differences in choice probabilities for $z_i = 2$ and $z_i = 3$ in the two cases of $(\alpha_1 < 0, \alpha_2 < 0)$ and $(\alpha_1 < 0, \alpha_2 > 0)$. From (2.9) and Figure 1, for $(\alpha_1 < 0, \alpha_2 < 0)$,

[1]In their comparison of classical and Bayesian estimators for partially identified models, Moon and Schorfheide (2012) adopt a distributional assumption only for Bayesian estimation

we obtain the following expressions:

$$\begin{aligned}
\Pr(z_i = 2|\boldsymbol{\theta}, \boldsymbol{p}_i) &= P_2(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i2} P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) \leq P_2(\boldsymbol{x}_i, \boldsymbol{\theta}) + P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) \\
&= \Pr(u_{i1} \geq -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1,\ u_{i2} < -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 - \alpha_2), \qquad (3.1) \\
\Pr(z_i = 3|\boldsymbol{\theta}, \boldsymbol{p}_i) &= P_3(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i3} P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) \leq P_3(\boldsymbol{x}_i, \boldsymbol{\theta}) + P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) \\
&= \Pr(u_{i1} < -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 - \alpha_1,\ u_{i2} \geq -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2), \qquad (3.2)
\end{aligned}$$

while for $(\alpha_1 < 0, \alpha_2 > 0)$, we have the following expressions:

$$\begin{aligned}
\Pr(z_i = 2|\boldsymbol{\theta}, \boldsymbol{p}_i) &= P_2(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i2} P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) \geq P_2(\boldsymbol{x}_i, \boldsymbol{\theta}) \\
&= \Pr(u_{i1} \geq -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1,\ u_{i2} < -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 - \alpha_2), \qquad (3.3) \\
\Pr(z_i = 3|\boldsymbol{\theta}, \boldsymbol{p}_i) &= P_3(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i3} P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) \geq P_3(\boldsymbol{x}_i, \boldsymbol{\theta}) \\
&= \Pr(u_{i1} < -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 - \alpha_1,\ u_{i2} \geq -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2). \qquad (3.4)
\end{aligned}$$

Inequalities (3.1) and (3.3) imply that $\Pr(z_i = 2|\boldsymbol{\theta}, \boldsymbol{p}_i)$ can be equal for these two cases only if $p_{i2} = 1$ for $(\alpha_1 < 0, \alpha_2 < 0)$ and $p_{i2} = 0$ for $(\alpha_1 < 0, \alpha_2 > 0)$. However, these conditions indicates $p_{i3} = 0$ for $(\alpha_1 < 0, \alpha_2 < 0)$ and $p_{i3} = 1$ for $(\alpha_1 < 0, \alpha_2 > 0)$. Thus, given inequalities (3.2) and (3.4), these conditions imply that $\Pr(z_i = 3|\boldsymbol{\theta}, \boldsymbol{p}_i)$ can not be equal in the two cases. This reasoning indicates that choice probabilities are differently formulated under distinct sign conditions. Distinct choice probabilities produce different functional forms of likelihood functions, and hence different statistical models. Thus, the models are not nested, and we need model selection to determine an appropriate information structure.

With respect to model selection, we have the following candidates of models. First, we have the following four models which require a selection rule approach:

$$\begin{aligned}
&\textit{Model NN}: && \alpha_1 < 0, \alpha_2 < 0, \\
&\textit{Model NP}: && \alpha_1 < 0, \alpha_2 > 0, \\
&\textit{Model PN}: && \alpha_1 > 0, \alpha_2 < 0, \\
&\textit{Model PP}: && \alpha_1 > 0, \alpha_2 > 0,
\end{aligned}$$

with (2.3) and (2.4). Additionally, there are following three models with coherent conditions:

$$\begin{aligned}
&\textit{Model CI}: && \alpha_1 = 0, \alpha_2 = 0, \\
&\textit{Model MH}: && \alpha_1 = 0, \alpha_2 \in \mathbb{R}, \\
&\textit{Model AS}: && \alpha_1 \in \mathbb{R}, \alpha_2 = 0,
\end{aligned}$$

with (2.3) and (2.4). In the above expressions, CI, MH and AS represent complete information, moral hazard and "adverse or advantageous" selection, respectively. Models CI, MH and AS are estimated using standard Gibbs samplers, as seen in Koop et al. (2007). Furthermore, we note that the bivariate probit model with zero correlation reduces to Model CI.

Although Models NN, NP, PN and PP are not mutually nested, each of these models is nested within the coherent models CI, MH and AS. Thus, if prior knowledge could narrow our focus to a hypothesis regarding whether a specific information structure exists, we could conduct not only model selection but also a t-type testing. However, without such powerful knowledge, model selection is necessary to detect the information structure that consumers face.

3.3 Identification

For identification, we depend on a distributional assumption regarding error terms (u_{i1}, u_{i2}). By combining this distributional assumption with functional forms in (2.3) and (2.4), we can calculate a closed-form expression of $P_l(\boldsymbol{x}_i, \boldsymbol{\theta})$ for each l. Therefore, we can obtain a well-defined likelihood function and achieve identification. On the other hand, Ciliberto and Tamer (2009) derived a nonparametric identification condition using an argument of identification at infinity, unlike our parametric identification.

In our Bayesian approach, the distributional assumption is required not only for identification, but also for estimation. Besides, as mentioned before, a nonparametric estimation method has not yet been created within this context. Thus, from a practical perspective, our parametric identification appears to be relatively innocuous to our approach.

4 An empirical analysis of the US dental insurance market

In this section, we apply our methodology to the US dental care insurance market. This market is ideal for our research because in this market, the dependent variables take stable and modest values. In the U.S., both the rate of dental insurance coverage and the proportion of individuals who experience at least one dental visit per year, which we define as the occurrence of an accident, are consistently at approximately 50 % (Manski et al., 2002). These numbers might indicate the robustness of our dependent variables to unobserved shocks. We note that in the context of health insurance, moral hazard can produce positive impacts both for both an individual and for society, because frequent visits to a doctor might help identify illness in its early stages.

Our empirical study is influenced by Munkin and Trivedi (2008). They examined the American dental insurance market by a switching regression model using a Bayesian approach. Their analysis adopted insurance purchase and the number of dental visits as the "treatment" and the main outcome variables, respectively. This approach had two parameters to indicate information asymmetry. One parameter was the coefficient parameter of the effect of the insurance purchase on the accident occurrence. This specification is similar to our parametrization for moral hazard. Another parameter in Munkin and Trivedi (2008) was a correlation, which could be affected by both moral hazard and selection problems. Our study has an advantage that selection problems are explicitly represented by a coefficient parameter that is independent from influences of moral hazard. With the other topics that are common between the current investigation and the work of Munkin and Trivedi (2008), our estimation results generally correspond to their persuasive findings. This correspondence provides supporting evidence of the validity of our methodology.

4.1 Data

We construct a dataset in a similar manner to Munkin and Trivedi (2008). We use data from the 2004 wave of the Medical Expenditure Panel Survey(MEPS). The MEPS is a rotating panel dataset that consists of five interviews during a year for each sample individual. We restrict our samples to privately-employed workers who are between 25 to 65 years of age. Self-employed individuals and governmental workers are eliminated because these two groups typically feature peculiar insurance coverage statuses. Our sample size is 5090.

Two dependent dummy variables for this study are defined as follows. First, the accident occurrence dummy takes unity if a sampled individual obtains a dental care at least once during the survey period. Second, the insurance purchase dummy takes unity if a consumer possesses dental insurance at the time of the first interview. Although the MEPS is panel data, we only use information from a respondent's first interview and ignore the remaining interviews to avoid a confusion due to time inconsistency. We also adopt this approach for the time-variant explanatory variables that are defined below.

We choose our explanatory variables from the four categories: demographic, health-related, geographic and insurance-purchase-specific considerations. Table 1 indicates descriptive statistics for these variables. Our sample is similar to that of Munkin and Trivedi (2008), who used the MEPS data from 1996 to 2000. Several notes regarding the definitions of explanatory variables are provided in the following

	Mean	S.D.
Insurance purchase (y_1)	0.537	(0.499)
1 or more Dental visit (y_2)	0.550	(0.498)
3 or more Dental visits	0.294	(0.456)
$30 \leq$ Age < 35	0.140	(0.347)
$35 \leq$ Age < 40	0.153	(0.360)
$40 \leq$ Age < 45	0.156	(0.363)
$45 \leq$ Age < 50	0.146	(0.354)
$50 \leq$ Age < 55	0.119	(0.324)
$55 \leq$ Age < 60	0.085	(0.278)
$60 \leq$ Age < 65	0.034	(0.181)
Afro-American	0.133	(0.339)
Hispanic	0.222	(0.416)
Married	0.634	(0.482)
Family Size	3.106	(1.540)
Schooling years	12.910	(3.113)
Income	38.654	(30.810)
Female	0.487	(0.500)
Age $\times$ Female	2.054	(2.228)
Very good health	0.328	(0.469)
Good health	0.275	(0.446)
Fair or poor health	0.105	(0.306)
# Chronic conditions	0.485	(0.768)
Northeast	0.158	(0.365)
Midwest	0.203	(0.402)
South	0.400	(0.490)
MSA	0.827	(0.379)
Firm size	13.617	(17.988)
N	5090	

Table 1: Descriptive Statistics

paragraph.

With respect to the demographic variables, consumer ages are represented by dummy variables that correspond to five-year categories. One category, "29 years old or younger", is excluded from the set of explanatory variables as a reference. For health variables, self-reported health status is included in terms of the following categorical dummies: "Very good", "good" and "fair or poor". "Excellent" is the reference category. In addition, we include a respondent's number of chronic conditions, namely diabetes, asthma, high blood pressure, coronary heart disease, emphysema and arthritis. There are two types of geographic variables. One group consists of location dummies in which "west" is the reference category, while another geographical variable is a dummy variable that takes unity if a consumer lives in a metropolitan statistical area. Finally, the firm size, which might have an effect on available insurance plans but not on dental visits, is adopted as an explanatory variable only for the insurance purchase.

4.2 Empirical results

We employ the Bayesian methodology using the MCMC method. Detailed procedures regarding this approach is found in Sugawara and Omori (2012). We also estimate the bivariate probit model to allow for comparisons between my approach and the conventional methodology.

We need a distributional assumption for error terms. For the bivariate probit model, we adopt a standard assumption of the normal error terms with unit variances and a correlation ρ. For the other models of this study, we assume the standard normal distribution for error terms. Distributions other than the normal distribution can be utilized, because the only requirement for the distributional assumption is that it must define the area probability $P_k(\boldsymbol{x}_i, \boldsymbol{\theta})$.

We use the same hyperparameters for distinct schemes in the following manner. The prior covariance matrix for $\boldsymbol{\theta}$ is set as an orthogonal matrix. Its diagonal elements, or prior variances, are assumed to be 10 for both $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, reflecting the lack of prior information regarding these parameters. The prior means for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are set to zero. The prior means of α_j depends on the domain of this variable in each model. In particular, for models without constraints for α_j, the prior mean is set to zero. For models with constraints of $\alpha_j < 0$ or $\alpha_j > 0$, the prior means are set to -1 or 1, respectively. For the selection rule $\boldsymbol{p}$, we set hyperparameters such that the prior distribution is the uniform distribution. For the correlation parameter in the bivariate probit model, we performed the Fisher transformation and utilize the normal prior distribution with mean 0 and variance 10 for the transformed value.

For all models, we generate 10,000 posterior samples after discarding 5,000 initial samples as the burn-in period. In the posterior sampling for Models NN, NP, PN and PP, because the dimensions of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are large, we separately generate each element of the coefficient vectors.

4.2.1 The model selection results

	(1)		(2)		(3)	
	DIC	Likelihood	DIC	Likelihood	DIC	Likelihood
NN	12406	-6154.4	11596.	-5749.0	12407	-6154.4
NP	12289	-6093.0	11509.	-5703.7	12296	-6095.5
PN	12319	-6109.1	11559.	-5729.6	12441	-6177.5
PP	12299	-6099.3	11538.	-5718.3	12299	-6099.4
CI	12403	-6152.4	11592.	-5747.1		
AS	12327	-6113.1	11557.	-5728.6	12334	-6117.4
MH	12307	-6103.9	11535.	-5717.4		
Bivariate Probit	12335	-6117.8	11559.	-5730.1		

Table 2: Model selection via DIC and likelihood at posterior mean
(1) is our basic result and (2) and (3) are results for robustness check. (2) uses an alternative definition for the accident dummy, and (3) adopts the alternative parametrization for selection problems.

The first two columns of Table 2 illustrate the model selection results. The first column of Table 2 presents the deviance information criterion (DIC) of Spiegelhalter et al. (2002). In the second column, we present values of the likelihood function evaluated at the posterior means in the second column, to indicate the goodness of fit of the examined models.

Both of these criteria indicate that Model NP is the most appropriate model for the assessed situation. This selected model exhibits the existence of moral hazard and advantageous selection. The detection of moral hazard is consistent to the findings of previous studies. On the other hand, our important contribution is the discovery of advantageous selection. This advantageous selection is intuitively interpreted in the context of the dental insurance in the following manner. Dental illness is often a result of the chronic accumulation of damage to teeth. Thus, individuals might decide to purchase dental insurance before they actually need serious dental care. In other words, people with early preventive concerns, individuals who purchase dental insurance may be concerned about early preventive issues. These individuals will have a lower risk of dental problems than individuals who may less vigilant about their dental health.

The uniqueness of our finding is reinforced when we compare our results with the results from the conventional method. In the bivariate probit estimation, the posterior mean for the correlation parameter ρ is 0.129 and its 95% credible interval is [0.089, 0.168]. Thus, the posterior probability of $\rho > 0$ is greater than 0.95. This result implies that the conventional methodology found evidence of standard asymmetric information, moral hazard and/or adverse selection. However, the conventional approach cannot detect advantageous selection that was found by our methodology. Given the bivariate probit model is outperformed by Model NP in Table 2, it appears likely that our dataset actually contains advantageous selection. In summary, our methodology has the potential power to shed a new light on information asymmetry which have remained hidden in the previous investigations.

4.2.2 Coefficient estimation results

Based on the above model selection result, we present detailed estimation results for Model NP. Figures 3, 4 and 5 in Appendix C depict the paths of the posterior samples that mix well and are stable. The convergences of α_1 and α_2 appear to be relatively slow, but their inefficiency factors, which measures the sampling efficiency, as discussed in Chib (2001), are 228 and 229. This result indicates that we obtained 43 (10,000 / 229 = 43.67) hypothetical uncorrelated samples to conduct statistical inferences. We have sufficiently large acceptance rates of more than 0.95 for all the model parameters to guarantee the efficiency of our Metropolis-Hastings algorithm.

	Insurance purchase: y_1			Dental Visit: y_2		
	Mean	S.D.	95% interval	Mean	S.D.	95% interval
Constant	-0.856	(0.150)	[-1.153, -0.565]	-1.390	(0.153)	[-1.701, -1.105]
Very good	0.042	(0.049)	[-0.055, 0.137]	0.027	(0.051)	[-0.072, 0.129]
Good	-0.126	(0.053)	[-0.231, -0.023]	-0.015	(0.056)	[-0.122, 0.097]
Fair and poor	-0.292	(0.076)	[-0.439, -0.145]	-0.159	(0.078)	[-0.311, -0.001]
# Chronic condition	0.110	(0.028)	[0.056, 0.165]	0.027	(0.030)	[-0.033, 0.084]
Family Size	-0.039	(0.015)	[-0.069, -0.009]	-0.059	(0.015)	[-0.089, -0.028]
$30 \leq$ Age < 35	0.116	(0.071)	[-0.025, 0.254]	0.002	(0.075)	[-0.148, 0.147]
$35 \leq$ Age < 40	0.074	(0.072)	[-0.068, 0.215]	0.027	(0.073)	[-0.119, 0.171]
$40 \leq$ Age < 45	0.128	(0.076)	[-0.021, 0.275]	0.117	(0.077)	[-0.035, 0.266]
$45 \leq$ Age <50	0.076	(0.082)	[-0.084, 0.239]	0.179	(0.084)	[0.014, 0.345]
$50 \leq$ Age < 55	0.210	(0.094)	[0.025, 0.395]	0.278	(0.094)	[0.092, 0.464]
$55 \leq$ Age <60	0.084	(0.102)	[-0.112, 0.287]	0.275	(0.108)	[0.064, 0.485]
$60 \leq$ Age < 65	-0.077	(0.133)	[-0.333, 0.186]	0.220	(0.140)	[-0.059, 0.493]
Schooling years	0.050	(0.009)	[0.032, 0.068]	0.060	(0.008)	[0.044, 0.077]
Income	0.007	(0.001)	[0.005, 0.008]	0.003	(0.001)	[0.002, 0.005]
Female	0.476	(0.166)	[0.135, 0.787]	0.754	(0.172)	[0.363, 1.067]
Age $\times$ Female	-0.066	(0.037)	[-0.135, 0.010]	-0.110	(0.040)	[-0.185, -0.022]
Afro-American	0.037	(0.061)	[-0.083, 0.156]	-0.222	(0.063)	[-0.349, -0.099]
Hispanic	-0.484	(0.058)	[-0.598, -0.372]	-0.104	(0.065)	[-0.228, 0.026]
Married	0.394	(0.047)	[0.305, 0.486]	0.073	(0.055)	[-0.037, 0.179]
Northeast	-0.115	(0.064)	[-0.241, 0.010]	0.006	(0.066)	[-0.124, 0.138]
Midwest	-0.055	(0.061)	[-0.176, 0.063]	-0.028	(0.063)	[-0.153, 0.096]
South	-0.326	(0.055)	[-0.434, -0.218]	-0.199	(0.055)	[-0.306, -0.089]
MSA	0.178	(0.052)	[0.076, 0.278]	0.046	(0.057)	[-0.066, 0.158]
Firm size	0.018	(0.001)	[0.015, 0.020]			
α	-0.708	(0.198)	[-1.119, -0.340]	1.128	(0.210)	[0.743, 1.568]
N	5090					

Table 3: Estimation results for coefficients of Model NP with basic parametrization: Posterior means, standard deviations and 95% credible intervals

Table 3 shows the estimation results for the coefficient parameters of Model NP. The effects of the health status variables are generally compatible to the effects that were found by Munkin and Trivedi (2008). In particular, self-reported health variables, "Fair and poor" and "Good" status, produce negative effects on insurance purchase compared to the reference category of excellent health. This result supports the notion of advantageous selection, which states that healthier people are more likely to purchase dental insurance. However, this type of reasoning does not hold for chronic illnesses, which have a positive effect on dental insurance purchase. With respect to the dental visit, all of the examined health variables have only minor impacts. This result might indicate that dental health is not strongly correlated with general health status.

An examination of the demographic variables reveals that age categories have minor effects on the insurance purchase. People who are older than 45 years of age receive dental care more frequently than younger people. This phenomenon may reflect the fact that dental illness is generally caused by damages that has accumulated over time. The positive effects of education both on the insurance purchase and on dental visits implies that a positive relationship exists between cognitive ability and each of these behaviors. Being female has positive effects on both the insurance purchase and dental visits, while an interaction term between gender and age have negative effects on dental visits. These results indicate the presence of gender-based differences in consumer behaviors. For racial variables, we observe the same interesting results as in Munkin and Trivedi (2008) where Hispanics are the least likely to purchase dental insurance while African-Americans are the least likely to visit a dentist. Furthermore, employees of larger firms are more likely to have dental insurance, which might indicate the generosity of larger firms with respect to the provision of benefits.

4.3 Robustness check

4.3.1 The exclusion of checkups: An alternative definition of accident occurrence

In our empirical analysis, we define the accident occurrence dummy such that it takes unity if a consumer visits a dentist at least once annually. However, there is a possibility that this definition reflects checkups, which are offered for free by many dental insurance contracts. If these free checkups provided a direct motivation to visit a doctor, our economic model would not be directly related to the information problem.

To resolve this issue, we introduce an alternative variable definition to exclude a checkup visit. Specifically, we redefine the accident dummy to take unity if a respondent engages in at least three dentist visits during a year. We then conduct model selection using this new variable. The third and fourth columns of Table 2 show the result of this model selection. Both criteria again select the model with moral hazard and advantageous selection. This result supports the robustness of our findings.

4.3.2 Endogeneity in error terms

In section 2.2, we assume that there is no endogeneity in the error terms on equations (2.3) and (2.4). Although this assumption directly corresponds to the bivariate probit model's assumption that no omitted variables exist, we can relax this condition by introducing a correlation between the error terms (u_{i1}, u_{i2}). The ability to relax this condition is an advantage of our methodology relative to the bivariate probit model, which strictly eliminate any correlation pattern other than information asymmetry.

In practice, we assume that (u_{i1}, u_{i2}) follows a bivariate normal distribution with unit variances and the correlation parameter $\rho_u \in [-1, 1]$. Except for the inclusion of this correlation term, we utilize the same setting that were used for our basic estimation. The MCMC sampling for the new parameter ρ_u is implemented for its Fisher transformation to enlarge the support for this parameter to the the set of all real numbers. The prior distribution for this transformed version of ρ_u is set to be a normal distribution $N(0, 10)$.

Using this setting, we employ an additional estimation for the Model NP, which was selected in our basic analysis. The resulting 95% credible interval for the estimates for ρ_u is $[-0.536, 0.084]$. Because this credible interval contains zero, this result provides supporting evidence for our basic framework, which lacks the endogeneity between the error terms.

4.3.3 An alternative parametrization for selection problems

This subsection provides an alternative parametrization to characterize the selection problems. In (2.3) and (2.4), selection problems are measured in terms of an effect of the accident occurrence y_{i2} on the insurance purchase, y_{i1}. Apparently, this specification appears to exhibit an reverse causality, because an accident, which should occur after the insurance purchase, affects a purchase decision. Alternately, we can define selection problems as an effect of the latent variable y_{i2}^* on the insurance purchase, leaving all other aspects of the proposed approach unaltered.

Alternative parametrization:

$$y_{i1}^* = \boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 + \alpha_1 y_{i2}^* + u_{i1}. \tag{4.1}$$

$$y_{i2}^* = \boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 + \alpha_2 y_{i1} + u_{i2}. \tag{4.2}$$

$$y_{ij} = I[y_{ij}^* \geq 0]. \tag{4.3}$$

This modeling appears to be more natural than the basic parametrization. However, under the behavioral assumption of Section 2.2.3, the apparent reverse causality is not truly a logical fallacy under the simultaneity assumption. Because the feedback loop is considered only in consumers' mind, the accident occurrence which is used for their decision making is not actual but instead represent hypothetically predicted values. Thus, variables y_{i2} and y_{i2}^* contain the same amount of information. In particular, these variables represent hypothetical draws of (u_{i1}, u_{i2}). This reasoning indicates that the basic and alternative parameterizations of the suggested approach are based on the same economic model for describing consumer behavior.

Similar to the basic parametrization, the alternative parametrization has mutual dependencies between its two dependent variables. It still requires an additional assumption to identify moral hazard and selection problems separately, as shown in Appendix A. We again adopt the simultaneity assumption and address the incoherency problem[2]. The correspondences between outcomes and error terms are as follows.

$$z_i = 1 \quad \Leftrightarrow \quad \left\{u_{i1} + \alpha_1 u_{i2} \leq -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 - \alpha_1 \boldsymbol{x}_{i2}'\boldsymbol{\beta}_2, \quad u_{i2} \leq -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2\right\}, \tag{4.4}$$

$$z_i = 2 \quad \Leftrightarrow \quad \left\{u_{i1} + \alpha_1 u_{i2} \geq -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 - \alpha_1 \boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 - \alpha_1\alpha_2, \quad u_{i2} \leq -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 - \alpha_2\right\}, \tag{4.5}$$

$$z_i = 3 \quad \Leftrightarrow \quad \left\{u_{i1} + \alpha_1 u_{i2} \leq -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 - \alpha_1 \boldsymbol{x}_{i2}'\boldsymbol{\beta}_2, \quad u_{i2} \geq -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2\right\}, \tag{4.6}$$

$$z_i = 4 \quad \Leftrightarrow \quad \left\{u_{i1} + \alpha_1 u_{i2} \geq -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 - \alpha_1 \boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 - \alpha_1\alpha_2, \quad u_{i2} \geq -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 - \alpha_2\right\}. \tag{4.7}$$

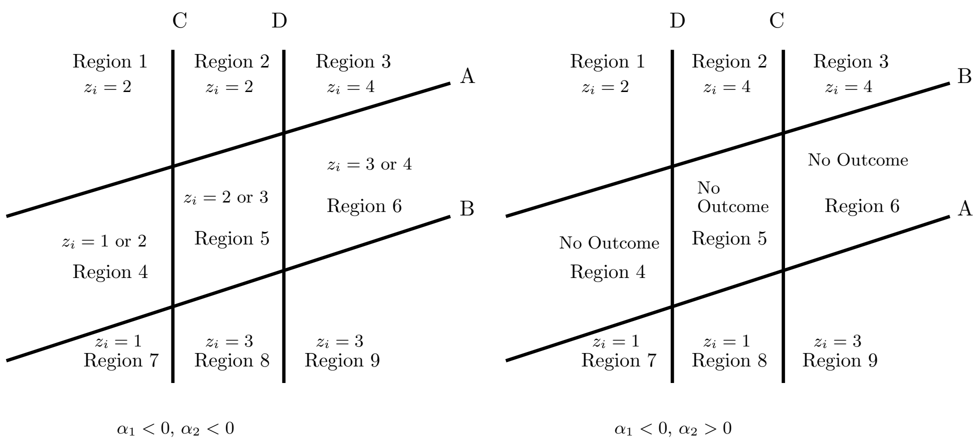

Figure 2: Data generating process for alternative parametrization in (u_{i2}, u_{i1}) coordinates

Figure 2 illustrates the above correspondences (4.4) - (4.7) on the coordinates of

[2]This is listed by Maddala (1983, p.119) as an example of incoherent econometric models.

(u_{i2}, u_{i1}). We concentrate on the two cases $(\alpha_1 < 0,\ \alpha_2 < 0)$ and $(\alpha_1 < 0,\ \alpha_2 > 0)$. The four lines in Figure 2 are defined as:

$$A \ : \ u_{i1} + \alpha_1 u_{i2} = -\boldsymbol{x}'_{i1}\boldsymbol{\beta}_1 - \alpha_1 \boldsymbol{x}'_{i2}\boldsymbol{\beta}_2, \tag{4.8}$$

$$B \ : \ u_{i1} + \alpha_1 u_{i2} = -\boldsymbol{x}'_{i1}\boldsymbol{\beta}_1 - \alpha_1 \boldsymbol{x}'_{i2}\boldsymbol{\beta}_2 - \alpha_1\alpha_2, \tag{4.9}$$

$$C \ : \ u_{i2} = -\boldsymbol{x}'_{i2}\boldsymbol{\beta}_2, \tag{4.10}$$

$$D \ : \ u_{i2} = -\boldsymbol{x}'_{i2}\boldsymbol{\beta}_2 - \alpha_2. \tag{4.11}$$

For both cases, Regions 1, 2, 3, 7, 8 and 9 have a unique pair of outcomes, while Regions 4, 5 and 6 lack this trait. We refer to regions without a unique pair of outcomes as nonsingular regions. The coherency condition is $\exists j, \alpha_j = 0$, which make the two lines A and B to be equivalent. If this condition is not satisfied, we again introduce selection rules. If Region k is nonsingular, the data generating process can be modeled using $\boldsymbol{p}_{ik} = (p_{ik1}, p_{ik2}, p_{ik3}, p_{ik4}) \in [0,1]^4$ with $\sum_{l=1}^{4} p_{ikl} = 1$. In this expression, p_{ikl} represents a proportion of Region k for the outcome $z_i = l$. Thus, we obtain the following choice probability for $z_i = l$:

$$\Pr[z_i = l | \boldsymbol{x}_i, \boldsymbol{\theta}, \boldsymbol{p}_i] = \sum_{k: z_i = l \text{ uniquely}} P_k(\boldsymbol{x}_i, \boldsymbol{\theta}) + \sum_{k: \text{ nonsingular}} p_{ikl} P_k(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{4.12}$$

where $P_k(\boldsymbol{x}_i, \boldsymbol{\theta})$ measures an area of Region k in Figure 2 and $\boldsymbol{p}_i$ is a vector with elements p_{ikl} for all the nonsingular k and $l = 1, 2, 3$ and 4. Appendix B provides further details regarding the Bayesian estimation for this alternative parametrization.

For model selection, certain models with a coherency restriction are the same in the basic and alternative parameterizations. Specifically, these two parameterizations share the same Models CI and MH but have different Model AS. Accordingly, the alternative parametrization produces five new models:

$$\begin{aligned}
&\textit{Model NN, alternative parametrization:} && \alpha_1 < 0, \alpha_2 < 0, \\
&\textit{Model NP, alternative parametrization:} && \alpha_1 < 0, \alpha_2 > 0, \\
&\textit{Model PN, alternative parametrization:} && \alpha_1 > 0, \alpha_2 < 0, \\
&\textit{Model PP, alternative parametrization:} && \alpha_1 > 0, \alpha_2 > 0, \\
&\textit{Model AS, alternative parametrization:} && \alpha_1 \in \mathbb{R}, \alpha_2 = 0,
\end{aligned}$$

with (4.1), (4.2) and (4.3).

A clear disadvantage of the alternative parametrization is its computational burden relative to the basic parametrization. The models of the alternative parametrization require many selection rules, which slow down the convergence of our estimation algorithm. We generate different numbers of posterior samples for distinct models to ensure that each Markov chain mixes well. For Model AS, we generate 10,000 posterior samples after discarding 5,000 initial samples as the burn-in period. For Models NN and PP of the alternative parametrization, we run 20,000 iterations after a burn-in of 20,000 iterations. For Models NP and PN with the alternative parametrization, we generate 50,000 posterior samples after 20,000 burn-in iterations.

The fifth and sixth columns of Table 2 show the results of model selection. For both model selection criteria, Model NP with the basic parametrization outperforms all of the models that are generated with the alternative selection mechanism. Moreover, Model NP remains the best of the models of the alternative parametrization. This result indicates the robustness of our model selection result for the information structure of the U.S. dental insurance market.

5 Conclusion

This paper has proposed an econometric methodology to analyze moral hazard and selection problems separately in insurance markets. We adopted a Bayesian approach to detect the information structure that consumers face. Our empirical study has shown that moral hazard and advantageous selection are present in the US dental insurance market.

Our method can have a good synergy effect in combination with structural approaches. For example, our results indicate that we can concentrate on structural models with moral hazard and advantageous selection for the purpose of analyzing the U.S. dental insurance market.

References

Amemiya, T. (1975) "Qualitative response models," *Annals of Economic and Social Measurement*, Vol. 4, pp. 363-372.

Anselin, L. (2003) "Spatial externalities, spatial multipliers and spatial economietrics," *International Regional Science Review*, Vol. 26, No. 2, pp. 153-166.

Arnold, B. C., E. Castillo, and J. M. Sarabia (1999) *Conditional specification of statistical models*, New York, NY: Springer-Verlag.

Besag, J. (1974) "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society, Series B*, Vol. 36, pp. 192-236.

Cardon, J. H. and I. Hendel (2001) "Asymmetric information in health insurance: Evidence from the National Medical Expenditure Survey," *RAND Journal of Economics*, Vol. 32, No. 3, pp. 408-27.

Cawley, J. and T. Philipson (1999) "An empirical examination of information barriers to trade in insurance," *American Economic Review*, Vol. 89, No. 4, pp. 827-846.

Chernozhukov, V., H. Hong, and E. Tamer (2007) "Estimation and confidence regions for parameter sets in econometric models," *Econometrica*, Vol. 75, No. 5, pp. 1243-1284.

Chernozhukov, V. and H. Hong (2003) "An MCMC approach to classical estimation," *Journal of Econometrics*, Vol. 115, pp. 293-346.

Chiappori, P.-A. and B. Salanié (2000) "Testing for asymmetric information on insurance markets," *Journal of Political Economy*, Vol. 108, No. 1, pp. 56-78.

Chib, S. (2001) "Markov chain Monte Carlo methods: computation and inference," in Heckman, J. J. and E. Leamer eds. *Handbook of Econometrics 5*, Amsterdam: North Holland.

Ciliberto, F. and E. Tamer (2009) "Market strucuture and multiple equilibria in airline markets," *Econometrica*, Vol. 77, No. 6, pp. 1791-1828.

Cressie, N. (1993) *Statistics for Spatial data*, New York: Wiley.

De Meza, D. and D. C. Webb (2001) "Advantageous selection in insurance markets," *RAND Journal of Economics*, Vol. 32, No. 2, pp. 249-262.

Einav, L., A. Finkelstein, and J. Levin (2010a) "Beyond testing: Empirical models of insurance markets," *Annual Review of Economics*, Vol. 2, pp. 311-326.

Einav, L., A. Finkelstein, and P. Schrimpf (2010b) "Optimal mandates and the welfare cost of asymmetric information: Evidence from the U. K. Annuity market," *Econometrica*, Vol. 78, No. 3, pp. 1031-1092.

Fang, H., M. P. Keane, and D. Silverman (2008) "Sources of advantageous selection: Evidence from the Medigap market," *Journal of Political Economy*, Vol. 116, No. 2, pp. 303-50.

Finkelstein, A. and K. McGarry (2006) "Multiple Dimensions of Private Information: Evidence from the long-term care insurance market," *American Economic Review*, Vol. 96, pp. 938-58.

Gouriéroux, C., J. J. Lafont, and A. Monfort (1980) "Coherency conditions in simultaneous linear equation models with endogenous switching regimes," *Econometrica*, Vol. 48, No. 3, pp. 675-696.

Gouriéroux, C. and A. Monfort (1979) "On the characterization of a joint probability distribution by conditional distributions," *Journal of Econometrics*, Vol. 10, pp. 115-118.

Heckman, J. (1978) "Dummy endogenous variables in a simultaneous equation system," *Econometrica*, Vol. 46, pp. 931-959.

Kim, J. Y. (2002) "Limited information likelihood and Bayesian analysis," *Journal of Econometrics*, Vol. 107, pp. 175-193.

Kitagawa, T. (2012) "Estimation and inference for set-identified parameters using posterior lower probability." Working paper.

Koop, G., D. J. Poirier, and J. L. Tobias (2007) *Bayesian Econometric Methods*, New York: Cambridge University Press.

Lancaster, T. (2000) "The incidental parameter problem since 1948," *Journal of Econometrics*, Vol. 95, pp. 391-413.

Liao, Y. and W. Jiang (2010) "Bayesian analysis in moment inequality models," *Annals of Statistics*, Vol. 38, pp. 275-316.

Maddala, G. S. (1983) *Limited-dependent and qualitative variables in econometrics*, New York, NY: Cambridge University Press.

Manning, W. G., B. Benjamin, H. L. Bailit, and J. P. Newhouse (1985) "The demand for dental care: evidence from a randomized trial in health insurance," *Journal of American Dental Association*, Vol. 110, pp. 895-902.

Manski, R. J., M. D. Macek, and J. F. Moeller (2002) "Private dental coverage: Who has it and how does it influence dental visits and expenditures?" *Journal of American Dental Association*, Vol. 133, No. 11, pp. 1551-1559.

Moon, H. R. and F. Schorfheide (2012) "Bayesian and frequentist inference in partially identified models," *Econometrica*, Vol. 80, No. 2, pp. 755-782.

Mueller, C. D. and A. C. Monheit (1988) "Insurance coverage and the demand for dental care: Results for non-aged white adults," *Journal of Health Economics*, Vol. 7, pp. 59-72.

Munkin, M. K. and P. K. Trivedi (2008) "A Bayesian analysis of the OPES model with a nonparametric component: An application to dental insurance and dental care," in Fomby, T. B. and R. C. Hill eds. *Bayesian Econometrics*, Vol. 23 of Advances in Econometrics: Emerald Group Publishing Limited, pp. 87-114.

Poirier, D. J. (1980) "Partial observability in bivariate probit model," *Journal of Econometrics*, Vol. 12, pp. 209-217.

Sintonen, H. and I. Linnosmaa (2000) "Economics of dental services," in Newhouse, J. and A. Culyer eds. *Handbook of health economics*, Vol. 1B, Amsterdam: North-Holland, pp. 1251-1296.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. v. d. Linde (2002) "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society, Series B*, Vol. 64, pp. 583-640.

Sugawara, S. and Y. Omori (2012) "Duopoly in the Japanese airline market: Bayesian estimation for entry games," *Japanese Economic Review*, Vol. 63, No. 3, pp. 310-332.

Tamer, E. (2003) "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, Vol. 70, pp. 147-165.

A A proof for impossibility of the separate identification without additional assumption

This appendix provides a formal proof for the statement that from equations (2.3) and (2.4) alone, we cannot separately identify α_1 and α_2. We also provide a corresponding proof for the alternative parametrization that is defined by equations (4.1), (4.2) and (4.3) in the later portion of this appendix. To deal with mutual dependencies of dependent variables without an additional assumption, there is a modeling methodology called the conditional specification in spatial statistics. As mentioned by Anselin (2003), the conditional specification handles the exogenous effects of the conditioned variable, which are not explained by an underlying economic model. Accordingly, the conditional specification is suitable for a reduced-form analysis, if feasible.

A conditional specification begins with defining the conditional distributions y_j given y_k, namely $y_j|y_k$, for $j, k = 1, 2$ and $j \neq k$. A main difficulty of the conditional specification is that we do not always have a corresponding joint distribution for given conditional distributions. To obtain a well-defined multivariate statistical model, we must recover the joint distribution for dependent variables from conditional distributions. This recoverability is known as compatibility and has been actively studied (Arnold et al., 1999).

To analyze the compatibility, we follow an argument presented by Besag (1974) and summarized by Cressie (1993). It is difficult to assess the compatibility directly in general distributions. Instead, we first assume that we have compatible conditionals and then consider the necessary conditions for the compatibility. Specifically, we begin our discussion with assuming that we have conditional distributions for $y_1|y_2$ and $y_2|y_1$, and that there exists a corresponding joint distribution for (y_1, y_2). At this stage, the discussion is completely nonparametric, in the sense that it does not requires specific distributional nor functional assumptions but only relies on the fact that the binary variables y_1 and y_2 are mutually dependent. We assume that $\boldsymbol{y} = \mathbf{0}$ occurs with a positive probability. Let us then define a negpotential function Q as $Q(\boldsymbol{y}) = \log[\Pr(\boldsymbol{y})/\Pr(\boldsymbol{y} = \mathbf{0})]$. Besag (1974) provided the following two fundamental theorems:

Theorem 1

For compatible conditional distributions, the following relationship hold:

$$\frac{Pr(y_j|y_k)}{Pr(y_j = 0|y_k)} = \frac{Pr(\boldsymbol{y})}{Pr(y_j = 0, y_k)} = \exp[Q(\boldsymbol{y}) - Q(y_j = 0, y_k)]. \tag{A.1}$$

Theorem 2

Q can be expanded as follows:

$$Q(\boldsymbol{y}) = y_1 G_1(y_1) + y_2 G_2(y_2) + y_1 y_2 G_{12}(y_1, y_2), \tag{A.2}$$

where $G_{\cdot}$ and $G_{\cdot\cdot}$ are uniquely determined if we define $G_j(y_j) = 0$ when $y_j = 0$ and $G_{12}(y_1, y_2) = 0$ when $y_1 = 0$ or $y_2 = 0$.

In words, Theorem 1 indicates the relationship among the conditional distribution, the joint distribution and the negpotential function, while Theorem 2 provides a unique decomposition of the negpotential function. In our context, Theorem 2 specifies the unique negpotential function as follows:

$$Q(\boldsymbol{y}) = y_1 G_1(1) + y_2 G_2(1) + y_1 y_2 G_{12}(1, 1). \tag{A.3}$$

By letting $\mu_j = G_j(1)$ and $\gamma = G_{1,2}(1,1)$, we have:

$$\begin{aligned} Q(y_j = 1, y_k) - Q(y_j = 0, y_k) &= y_k G_k(1) + G_j(1) + y_k G_{12}(1,1) - y_k G_k(1) \\ &= \mu_j + \gamma y_k. \end{aligned} \tag{A.4}$$

Applying Theorem 1, we obtain conditional distributions explicitly as:

$$\begin{aligned} \frac{1 - \Pr(y_j = 0|y_k)}{\Pr(y_j = 0|y_k)} &= \frac{\Pr(y_j = 1|y_k)}{\Pr(y_j = 0|y_k)} \\ &= \exp[Q(y_j = 1, y_k) - Q(y_j = 0, y_k)] \\ &= \exp[\mu_j + \gamma y_k]. \end{aligned} \tag{A.5}$$

These distributions produce the following relationships:

$$\begin{aligned}\Pr(y_j = 0|y_k) &= \frac{1}{1+\exp[\mu_j + \gamma y_k]}, \qquad (A.6)\\ \Pr(y_j = 1|y_k) &= \frac{\exp[\mu_j + \gamma y_k]}{1+\exp[\mu_j + \gamma y_k]}. \qquad (A.7)\end{aligned}$$

Finally, letting $\mu_j = \boldsymbol{x}_j'\boldsymbol{\beta}_j$, we have the following conditional probability:

$$\Pr(y_j|y_k) = \frac{y_j\{\exp[\boldsymbol{x}_j'\boldsymbol{\beta}_{\boldsymbol{j}} + \gamma y_k]\}}{1+\exp[\boldsymbol{x}_j'\boldsymbol{\beta}_{\boldsymbol{j}} + \gamma y_k]}, \quad j = 1, 2,\ k \neq j. \qquad (A.8)$$

Because γ does not depend on either j or k, the above equation implies that the effect of y_k on the conditional distribution for y_j is the same as the effect of y_j on the conditional distribution for y_k. The requirement of such a specific conditional distribution is a direct consequence of Besag's theorems for any compatible system under the completely nonparametric setting. In other words, it is a necessary condition for compatibility in our context.

The above discussion indicates that, for the basic parametrization, the conditional specification cannot distinguish effects of moral hazard and selection problems because they are to be measured by the same parameter γ. For the alternative parametrization, the compatibility condition (A.8) also yields a deterministic relationship between moral hazard and the selection problems, which are measured by γy_{i1} and $\gamma I[y_{i2}^* \geq 0]$, respectively.

B Estimation details for models with the alternative parametrization

B.1 Functional forms of choice probabilities

This appendix presents detailed analysis for models with the alternative parametrization, which is described in Section 4.3.3. First, we provide functional forms of the choice probabilities (4.12). For notational simplicity, we define the following variables:

$$\tilde{u}_{i1} = u_{i1} + \alpha_1 u_{i2}, \tag{B.1}$$

$$W_{i1} = -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 - \alpha_1 \boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 \tag{B.2}$$

$$W_{i2} = -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 - \alpha_1 \boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 - \alpha_1\alpha_2 \tag{B.3}$$

$$W_{i3} = -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 \tag{B.4}$$

$$W_{i4} = -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 - \alpha_2. \tag{B.5}$$

Using these variables, we have simple expressions for the lines A, B, C and D which are defined in Section 4.3.3 as

$$A : \tilde{u}_{i1} = W_{i1}, \qquad B : \tilde{u}_{i1} = W_{i2}, \tag{B.6}$$

$$C : u_{i2} = W_{i3}, \qquad D : u_{i2} = W_{i4}. \tag{B.7}$$

To obtain choice probabilities, we separately consider Models NN ($\alpha_1 < 0,\ \alpha_2 < 0$) and NP ($\alpha_1 < 0,\ \alpha_2 > 0$). For Model NN, Figure 2 shows that each of nonsingular regions has two possible pairs of outcomes, namely $z_i = 1$ or 2 in Region 4, $z_i = 2$ or 3 in Region 5 and $z_i = 3$ or 4 in Region 6. We then need to define three selection rules, each of which distributes a nonsingular region into two pairs of outcomes. Specifically, we assume that p_{i4}, p_{i5}, and p_{i6} represent proportions of $z_i = 1$ in Region 4, $z_i = 2$ in Region 5 and $z_i = 3$ in Region 6, respectively. Then we have the following choice probabilities:

$$\Pr(z_i = 1|\boldsymbol{\theta}, \boldsymbol{p}_i) = P_7(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i4}P_4(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{B.8}$$

$$\Pr(z_i = 2|\boldsymbol{\theta}, \boldsymbol{p}_i) = P_1(\boldsymbol{x}_i, \boldsymbol{\theta}) + P_2(\boldsymbol{x}_i, \boldsymbol{\theta}) + (1 - p_{i4})P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i5}P_5(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{B.9}$$

$$\Pr(z_i = 3|\boldsymbol{\theta}, \boldsymbol{p}_i) = P_8(\boldsymbol{x}_i, \boldsymbol{\theta}) + P_9(\boldsymbol{x}_i, \boldsymbol{\theta}) + (1 - p_{i5})P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i6}P_6(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{B.10}$$

$$\Pr(z_i = 4|\boldsymbol{\theta}, \boldsymbol{p}_i) = P_3(\boldsymbol{x}_i, \boldsymbol{\theta}) + (1 - p_{i6})P_6(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{B.11}$$

where $\boldsymbol{p}_i = (p_{i4}, p_{i5}, p_{i6})$. We define $\boldsymbol{p} = (\boldsymbol{p}_1, ..., \boldsymbol{p}_N)$ for the future reference.

Next, we provide area probabilities $P_k(\boldsymbol{x}_i, \boldsymbol{\theta})$, for $k = 1, ..., 9$. From the definition of lines A, B ,C and D, these area probabilities are formulated as functions of joint probabilities of $(\tilde{u}_{1i}, u_{i2})$. Specifically:

$$P_7(\boldsymbol{x}_i, \boldsymbol{\theta}) = \Pr(\tilde{u}_{i1} \le W_{i2},\ u_{i2} \le W_{i3}), \tag{B.12}$$

$$P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) = \Pr(\tilde{u}_{i1} \le W_{i1},\ u_{i2} \le W_{i3}) - P_7(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{B.13}$$

$$P_1(\boldsymbol{x}_i, \boldsymbol{\theta}) = \Pr(u_{i2} \le W_{i3}) - P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_7(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{B.14}$$

$$P_8(\boldsymbol{x}_i, \boldsymbol{\theta}) = \Pr(\tilde{u}_{i1} \le W_{i2},\ u_{i2} \le W_{i4}) - P_7(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{B.15}$$

$$P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) = \Pr(\tilde{u}_{i1} \le W_{i1},\ u_{i2} \le W_{i4}) - P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_7(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_8(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{B.16}$$

$$P_2(\boldsymbol{x}_i, \boldsymbol{\theta}) = \Pr(u_{i2} \le W_{i4}) - P_1(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_7(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_8(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{B.17}$$

$$P_9(\boldsymbol{x}_i, \boldsymbol{\theta}) = \Pr(\tilde{u}_{i1} \le W_{i2}) - P_7(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_8(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{B.18}$$

$$P_6(\boldsymbol{x}_i, \boldsymbol{\theta}) = \Pr(\tilde{u}_{i1} \le W_{i1}) - P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_7(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_8\boldsymbol{x}_i, \boldsymbol{\theta}) - P_9(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{B.19}$$

$$\begin{aligned} P_3(\boldsymbol{x}_i, \boldsymbol{\theta}) = {} & 1 - P_1(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_2(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_6(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_7(\boldsymbol{x}_i, \boldsymbol{\theta}) \\ & - P_8(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_9(\boldsymbol{x}_i, \boldsymbol{\theta}). \end{aligned} \tag{B.20}$$

For Model NP, Figure 2 shows that each of nonsingular regions has all four possible pairs of outcomes, $z_i = 1$, 2, 3 and 4. Then we need to define three selection rule vectors each of which distributes a nonsingular region into four pairs. Namely, we assume that for each Region $k = 4$, 5, and 6, $\boldsymbol{p}_{ik} = (p_{ik,1}, p_{ik,2}, p_{ik,3}, p_{ik,4})$ represents the proportion of $z_i = 1$, 2, 3 and 4, respectively. Consequently, we have the following choice probabilities:

$$\Pr(z_i = 1|\boldsymbol{\theta}, \boldsymbol{p}_i) = P_7(\boldsymbol{x}_i, \boldsymbol{\theta}) + P_8(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i4,1}P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i5,1}P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i6,1}P_6(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{B.21}$$

$$\Pr(z_i = 2|\boldsymbol{\theta}, \boldsymbol{p}_i) = P_1(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i4,2}P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i5,2}P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i6,2}P_6(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{B.22}$$

$$\Pr(z_i = 3|\boldsymbol{\theta}, \boldsymbol{p}_i) = P_9(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i4,3}P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i5,3}P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i6,3}P_6(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{B.23}$$

$$\Pr(z_i = 4|\boldsymbol{\theta}, \boldsymbol{p}_i) = P_2(\boldsymbol{x}_i, \boldsymbol{\theta}) + P_3(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i4,4}P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i5,4}P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i6,4}P_6(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{B.24}$$

where $\boldsymbol{p}_i = (\boldsymbol{p}_{i4}, \boldsymbol{p}_{i5}, \boldsymbol{p}_{i6})$. We define $\boldsymbol{p} = (\boldsymbol{p}_1, ..., \boldsymbol{p}_N)$ for the future reference. We have the area probabilities as:

$$
\begin{aligned}
P_7(\boldsymbol{x}_i, \boldsymbol{\theta}) &= \Pr(\tilde{u}_{i1} \le W_{i1},\ u_{i2} \le W_{i4}), && \text{(B.25)}\\
P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) &= \Pr(\tilde{u}_{i1} \le W_{i2},\ u_2 \le W_{i4}) - P_7(\boldsymbol{x}_i, \boldsymbol{\theta}), && \text{(B.26)}\\
P_1(\boldsymbol{x}_i, \boldsymbol{\theta}) &= \Pr(u_{i2} \le W_{i4}) - P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_7(\boldsymbol{x}_i, \boldsymbol{\theta}), && \text{(B.27)}\\
P_8(\boldsymbol{x}_i, \boldsymbol{\theta}) &= \Pr(\tilde{u}_{i1} \le W_{i1},\ u_{i2} \le W_{i3}) - P_7(\boldsymbol{x}_i, \boldsymbol{\theta}), && \text{(B.28)}\\
P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) &= \Pr(\tilde{u}_{i1} \le W_{i2},\ u_{i2} \le W_{i3}) - P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_7(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_8(\boldsymbol{x}_i, \boldsymbol{\theta}), && \text{(B.29)}\\
P_2(\boldsymbol{x}_i, \boldsymbol{\theta}) &= \Pr(u_{i2} \le W_{i3}) - P_1(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_7(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_8(\boldsymbol{x}_i, \boldsymbol{\theta}), \\
& && \text{(B.30)}\\
P_9(\boldsymbol{x}_i, \boldsymbol{\theta}) &= \Pr(\tilde{u}_{i1} \le W_{i1}) - P_7(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_8(\boldsymbol{x}_i, \boldsymbol{\theta}), && \text{(B.31)}\\
P_6(\boldsymbol{x}_i, \boldsymbol{\theta}) &= \Pr(\tilde{u}_{i1} \le W_{i2}) - P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_7(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_8(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_9(\boldsymbol{x}_i, \boldsymbol{\theta}), \\
& && \text{(B.32)}\\
P_3(\boldsymbol{x}_i, \boldsymbol{\theta}) &= 1 - P_1(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_2(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_6(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_7(\boldsymbol{x}_i, \boldsymbol{\theta}) \\
&\quad - P_8(\boldsymbol{x}_i, \boldsymbol{\theta}) - P_9(\boldsymbol{x}_i, \boldsymbol{\theta}). && \text{(B.33)}
\end{aligned}
$$

B.2 Bayesian estimation for models with the alternative parametrization

B.2.1 Additional hierarchical structure and the likelihood function

This appendix provides an estimation procedure for the models with the alternative parametrization. For technical tractability, we insert an additional structure using hierarchical Bayesian modeling. Specifically, we introduce a latent dummy variable λ which takes unity in proportion to the selection rule p. We need different hierarchical structures for Models NN and NP: First, for Model NN, we adopt $\lambda_{ik}|p_{ik} \sim \text{Bernoulli}(p_{ik})$ for $k = 4$, 5 and 6. Second, for Model NP, we adopt $\boldsymbol{\lambda_{ik}} = (\lambda_{ik,1}, \lambda_{ik,2}, \lambda_{ik,3}, \lambda_{ik,4})|\boldsymbol{p}_{ik} \sim \text{Multinomial}(1; \boldsymbol{p}_{ik})$. We let $\boldsymbol{\lambda}_i = (\lambda_{i4}, \lambda_{i5}, \lambda_{i6})$ or $\boldsymbol{\lambda}_i = (\boldsymbol{\lambda}_{i4}, \boldsymbol{\lambda}_{i5}, \boldsymbol{\lambda}_{i6})$ for Models NN and NP, respectively. For both models, we define $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, ..., \boldsymbol{\lambda}_N)$. Using these new structures, the choice probabilities can be represented as:

$$Pr(z_i = l|\boldsymbol{\theta}, \boldsymbol{\lambda}_i) = \int Pr(z_i = l|\boldsymbol{\theta}, \boldsymbol{p})\pi(\boldsymbol{\lambda}_i|\boldsymbol{p}_i)d\boldsymbol{p}_i. \tag{B.34}$$

For Model NN, new representations of the choice probabilities are:

$$\begin{aligned}
\Pr(z_i = 1|\boldsymbol{\theta}, \boldsymbol{\lambda}_i) &= P_7(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i4} P_4(\boldsymbol{x}_i, \boldsymbol{\theta}), && \text{(B.35)}\\
\Pr(z_i = 2|\boldsymbol{\theta}, \boldsymbol{\lambda}_i) &= P_1(\boldsymbol{x}_i, \boldsymbol{\theta}) + P_2(\boldsymbol{x}_i, \boldsymbol{\theta}) + (1 - \lambda_{i4}) P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) + (1 - \lambda_{i5}) P_5(\boldsymbol{x}_i, \boldsymbol{\theta}), && \text{(B.36)}\\
\Pr(z_i = 3|\boldsymbol{\theta}, \boldsymbol{\lambda}_i) &= P_8(\boldsymbol{x}_i, \boldsymbol{\theta}) + P_9(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i5} P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i6} P_6(\boldsymbol{x}_i, \boldsymbol{\theta}), && \text{(B.37)}\\
\Pr(z_i = 4|\boldsymbol{\theta}, \boldsymbol{\lambda}_i) &= P_3(\boldsymbol{x}_i, \boldsymbol{\theta}) + (1 - \lambda_{6i}) P_6(\boldsymbol{x}_i, \boldsymbol{\theta}). && \text{(B.38)}
\end{aligned}$$

For Model NP:

$$\begin{aligned}
\Pr(z_i = 1|\boldsymbol{\theta}, \boldsymbol{\lambda}_i) &= P_7(\boldsymbol{x}_i, \boldsymbol{\theta}) + P_8(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i4,1} P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i5,1} P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i6,1} P_6(\boldsymbol{x}_i, \boldsymbol{\theta}), && \text{(B.39)}\\
\Pr(z_i = 2|\boldsymbol{\theta}, \boldsymbol{\lambda}_i) &= P_1(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i4,2} P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i5,2} P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i6,2} P_6(\boldsymbol{x}_i, \boldsymbol{\theta}), && \text{(B.40)}\\
\Pr(z_i = 3|\boldsymbol{\theta}, \boldsymbol{\lambda}_i) &= P_9(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i4,3} P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i5,3} P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i6,3} P_6(\boldsymbol{x}_i, \boldsymbol{\theta}), && \text{(B.41)}\\
\Pr(z_i = 4|\boldsymbol{\theta}, \boldsymbol{\lambda}_i) &= P_2(\boldsymbol{x}_i, \boldsymbol{\theta}) + P_3(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i4,4} P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i5,4} P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i6,4} P_6(\boldsymbol{x}_i, \boldsymbol{\theta}). && \text{(B.42)}
\end{aligned}$$

Given these choice probabilities, the likelihood function is obtained as:

$$f(\boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{\lambda}) = \prod_{i=1}^{N} \prod_{l=1}^{4} \Pr(z_i = l|\boldsymbol{\theta}, \boldsymbol{\lambda}_i)^{I[z_i = l]}. \tag{B.43}$$

There are two remarks on the above likelihood function. First, because of the hierarchical nature of our setting, the marginal posterior distribution for the parameter

$\boldsymbol{\theta}$ is equivalent whether we use $\boldsymbol{\lambda}$ or $\boldsymbol{p}$. Second, to have the closed form expression for the likelihood function, we need to provide distributional assumptions for error terms (u_{i1}, u_{i2}) to calculate the area probabilities $P_k(\boldsymbol{x}_i, \boldsymbol{\theta})$.

B.2.2 Prior distributions

For coefficient parameters $\boldsymbol{\theta}$, we assume a normal prior distribution with mean $\boldsymbol{\theta}_0$ and covariance matrix Σ_0 truncated on the region R:

$$\boldsymbol{\theta} \sim TN_R(\boldsymbol{\theta}_0, \Sigma_0).$$

The truncation corresponds to a prescribed sign condition for (α_1, α_2). For example, we take the region $R = (-\infty, \infty)^{K_1+K_2} \times (-\infty, 0) \times (-\infty, 0)$ for Model NN.

For $\boldsymbol{p}$, we use conjugate prior distributions: In Model NN, we assume the beta prior with parameters (a_{ik1}, a_{ik2}) for p_{ik}, $k = 4$, 5 and 6:

$$p_{ik} \sim Beta(a_{ik1}, a_{ik2}).$$

In Model NP, the prior distribution of $\boldsymbol{p}_{ik}$ is assumed to be a Dirichlet distribution with parameter $\boldsymbol{a}_{ik} = (a_{ik1}, \ldots, a_{ik4})$:

$$\boldsymbol{p}_{ik} \sim Dirichlet(\boldsymbol{a}_{ik}).$$

B.2.3 Posterior sampling for Model NN with the alternative parametrization

Here we derive the MCMC sampling procedure for Model NN. Given the likelihood function and the prior distributions, we have the joint posterior density as:

$$\pi(\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{p}|\boldsymbol{z}) \propto f(\boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{\lambda})\pi(\boldsymbol{\theta}) \prod_{i=1}^{N} \prod_{k=4}^{6} p_{ik}^{(\lambda_{ik}+a_{ik1})-1}(1-p_{ik})^{(1-\lambda_{ik}+a_{ik2})-1}, \tag{B.44}$$

where $\pi(\boldsymbol{\theta})$ denotes a probability density function of the truncated normal distribution $TN_R(\boldsymbol{\theta}_0, \Sigma_0)$. The conditional posterior distributions of λ_{ik} and p_{ik} are:

$$\lambda_{ik}|\boldsymbol{\theta}, p_{ik}, z_i \sim Bernoulli(q_{ik}), \tag{B.45}$$

$$p_{ik}|\boldsymbol{\theta}, \lambda_{ik}, z_i \sim Beta(a_{ik1} + \lambda_{ik}, a_{ik2} + 1 - \lambda_{ik}), \tag{B.46}$$

where

$$q_{ik} = \frac{p_{ik}^{a_{ik1}}(1-p_{ik})^{a_{ik2}-1} f(z_i|\boldsymbol{\theta}, \lambda_{ik} = 1, \boldsymbol{\lambda}_{i,(-k)}}{p_{ik}^{a_{ik1}}(1-p_{ik})^{a_{ik2}-1} f(z_i|\boldsymbol{\theta}, \lambda_{ik} = 1, \boldsymbol{\lambda}_{i,(-k)}) + p_{ik}^{a_{ik1}-1}(1-p_{ik})^{a_{ik2}} f(z_i|\boldsymbol{\theta}, \lambda_{h,m} = 0, \boldsymbol{\lambda}_{(-h),m})}, \tag{B.47}$$

and $\boldsymbol{\lambda_{i,(-k)}}$ is a vector which consists of components of $\boldsymbol{\lambda}_i$ except λ_{ik} and $f(z_i|\boldsymbol{\theta}, \boldsymbol{\lambda}_i)$ is the individual i's contribution to the likelihood function.

Because the conditional posterior distributions take familiar forms, we implement Gibbs samplers for λ_{ik} and p_{ik}, for $k = 4, 5$ and 6 and $i = 1, ..., N$. On the other hand, $\boldsymbol{\theta}$ is sampled using the Metropolis-Hastings algorithm.

B.2.4 Posterior sampling for Model NP with alternative parametrization

Next, we describe the MCMC implementation for Model NP. The joint posterior density is:

$$\pi(\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{p}|\boldsymbol{z}) = f(\boldsymbol{z}|\boldsymbol{\theta}, \boldsymbol{\lambda})\pi(\boldsymbol{\theta}) \prod_{i=1}^{N} \prod_{k=4}^{6} \prod_{l=1}^{4} p_{ikl}^{\lambda_{ikl}+a_{ikl}-1}. \tag{B.48}$$

The conditional posterior distributions of $\boldsymbol{\lambda}_{ik}$ and $\boldsymbol{p}_{ik}$ are:

$$\boldsymbol{\lambda}_{ik}|\boldsymbol{\theta}, \boldsymbol{p}_{ik}, z_i \sim \text{Multinomial}(1, \boldsymbol{q}_{ik}), \tag{B.49}$$

$$\boldsymbol{p}_{ik}|\boldsymbol{\theta}, \boldsymbol{\lambda}_{ik}, z_i \sim \text{Dirichlet}(\boldsymbol{a}_{ik} + \boldsymbol{\lambda}_{ik}), \tag{B.50}$$

where $\boldsymbol{q}_{ik} = (q_{ik1}, ..., q_{ik4})$ such that:

$$q_{ikl} = \frac{p_{ikl}^{a_{ikl}} \left(\prod_{j \neq l} p_{ikj}^{a_{ikj}-1} \right) f(z_i | \boldsymbol{\theta}, \lambda_{ikl} = 1, \boldsymbol{\lambda}_{ik \backslash l} = \mathbf{0})}{\sum_{h=1}^{4} p_{ikh}^{a_{ikh}} \left(\prod_{j \neq h} p_{ikj}^{a_{ikj}-1} \right) f(z_i | \boldsymbol{\theta}, \lambda_{ikh} = 1, \boldsymbol{\lambda}_{ik \backslash h} = \mathbf{0})}, \quad l = 1, ..., 4, \quad \text{(B.51)}$$

and $\boldsymbol{\lambda}_{ik \backslash h} = \mathbf{0}$ is a vector which consists of λ_{ikh} and zeros.

Using the above conditional posterior densities, we can implement Gibbs samplers for $\boldsymbol{\lambda}_{ik}$ and $\boldsymbol{p}_{ik}$, for $k = 4, 5$ and 6 and $i = 1, ..., N$. For $\boldsymbol{\theta}$, we implement the Metropolis-Hastings algorithm for the posterior sampling.

C Posterior sample paths for the MCMC samplers

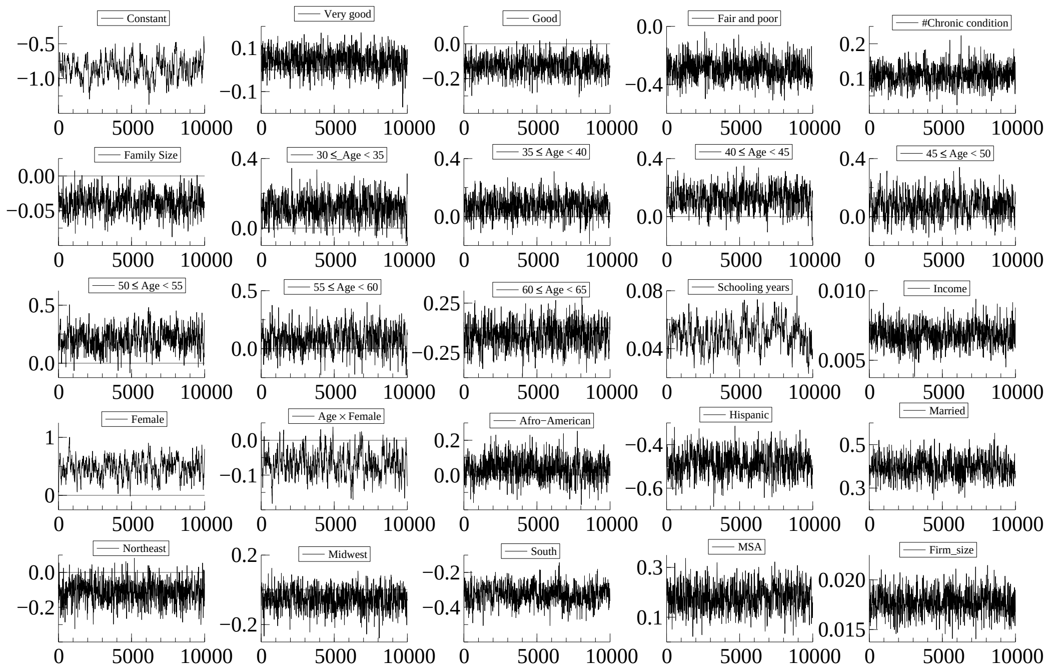

Figure 3: Sample paths of $\boldsymbol{\beta}_1$ for Model NP with basic parametrization

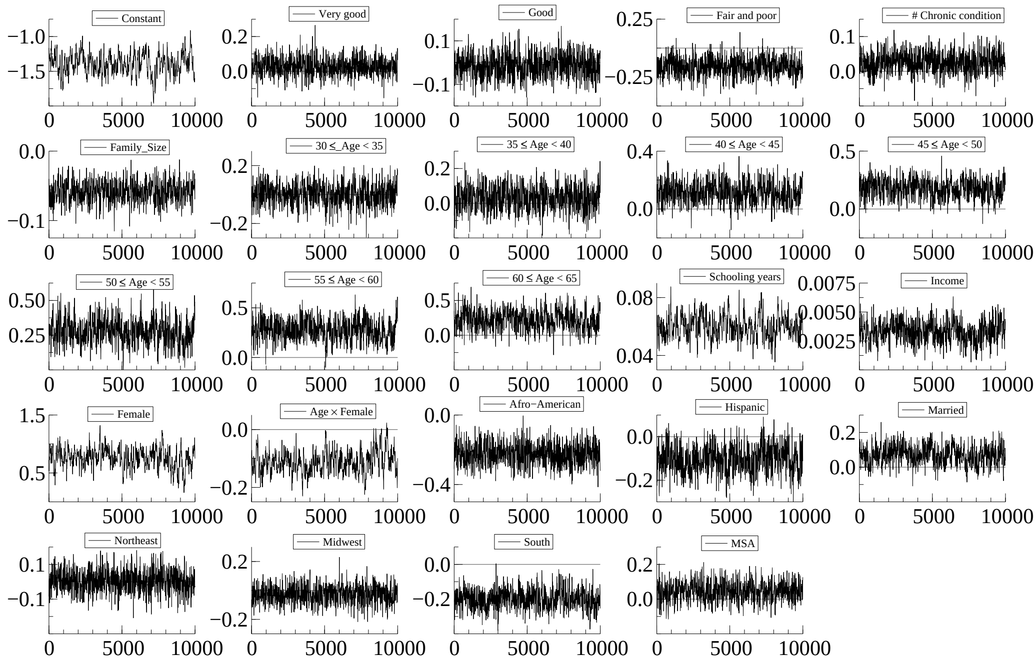

Figure 4: Sample paths of $\boldsymbol{\beta}_2$ for Model NP with basic parametrization

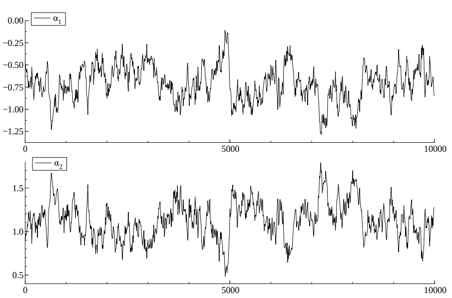

Figure 5: Sample paths of $\boldsymbol{\alpha}$ for Model NP with basic parametrization