CIRJE-F-849

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

Shinya Sugawara
Graduate School of Economics, University of Tokyo

Yasuhiro Omori
University of Tokyo

April 2012

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

Shinya Sugawara * Yasuhiro Omori †

April 2012

Abstract

This paper proposes a simple econometric framework that can identify moral hazard and selection problems separately in insurance markets. Although our methodology requires behavioral assumptions on the consumer's optimization, we show that these assumptions are necessary for the separate identification of the two sources of information asymmetry. Our method is applied to the dental insurance market in the United States. In addition to standard moral hazard, we find advantageous selection, which is not detected by a conventional methodology.

1 Introduction

In the last decade, empirical studies have been rapidly catching up with highly-developed economic theories of asymmetric information. This paper proposes a new econometric framework to analyze the information problem in insurance markets. Our main contribution is that we can identify moral hazard and selection problems separately. This contribution is an extension of a traditional method that formulates information asymmetry as one parameter.

The conventional methodology uses a bivariate probit model as discussed in Chiappori and Salanié (2000). For dependent variables, this pioneering work considered an insurance purchase and an accident occurrence of consumers. The standard asymmetric information problem can be detected as a positive correlation between these two variables. Specifically, moral hazard implies that consumers increase their risk after purchasing insurance, while adverse selection implies that riskier consumers have a stronger demand for insurance. The conventional methodology has two advantages:

*Graduate School of Economics, University of Tokyo, Tokyo 113-0033, Japan E-mail: 9474267911@mail.ecc.u-tokyo.ac.jp

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

computational simplicity and no need for detailed contract information, which is difficult to obtain. The bivariate probit analysis has been popular and applied to various insurance markets. However, most empirical studies fail to detect a significantly positive correlation between two dependent variables, and some studies even found a negative correlation.

In response to these unexpected results, De Meza and Webb (2001) proposed an important counter theory called advantageous selection. This theory states that less risky individuals have a stronger demand for insurance, due to their risk-aversion. In contrast to adverse selection, advantageous selection produces a negative effect of risk on an insurance purchase. Then researchers have considered that the positive correlation test might fail when both moral hazard and advantageous selection are present, because their distinct effects cannot be captured by only one correlation parameter. To test the validity of this theory, one needs to identify the moral hazard and the selection problems, meaning adverse selection and advantageous selection, separately.

Recently, two active streams of new researches have arisen in this field. One is the structural estimation approach, as in Cardon and Hendel (2001) and Einav et al. (2010b), where moral hazard and selection problems are explicitly modeled as distinct mechanisms. Another approach is a reduced form analysis for advantageous selection, which detects selection as the effects of the explanatory variables representing an individual's risk preference on the insurance purchase. To measure the risk preference, researchers have used a variety of variables, such as subjective mortality rate in Cawley and Philipson (1999), seat belt use in Finkelstein and McGarry (2006), and health status and schooling levels, which proximate cognitive ability and financial numeracy, in Fang et al. (2008). No consensus has been reached with respect to appropriateness of these variables.

In this paper, we present a new approach to identify the moral hazard and selection problems separately using an econometric model which is much simpler than the previous structural studies. Further, our model requires only the accident occurrence as an explanatory variable to characterize the selection problem.

Specifically, this paper proposes an inference framework which consists of only two equations, similar to the bivariate probit model. We parametrize moral hazard and selection problems not as a correlation but as two distinct coefficients. Despite a simple appearance, this specification has a statistical difficulty due to the mutual dependency between the dependent variables. Therefore, we assume simultaneous determination of these variables. This assumption, however, produces an econometric problem that is equivalent to the endogenous effect of the reflection problem in peer effects models (Manski, 1993).

The simultaneity assumption induces a well-known econometric problem that has

been analyzed in the literature on simultaneous equation models with limited dependent variables, for example in Amemiya (1975), Heckman (1978) and Gouriéroux and Monfort (1979). These types of work yielded a statistical problem called incoherency by Gouriéroux et al. (1980), which states that these models are not well-defined without a restrictive assumption on parameters.

To handle the incoherency problem, we adopt the modeling technique proposed by Tamer (2003) and Ciliberto and Tamer (2009) in the literature on the empirical analysis of entry games. These authors introduced a latent variable that formulates well-defined choice probabilities. This variable is called a selection rule in the literature on entry games, because the variable determines a Nash equilibrium among multiple equilibria. In this paper, by taking a Bayesian approach, we estimate the parameters of the model with the selection rule and conduct model selection to consider the information structure that fits the data.

Our simultaneity assumption produces another difficulty. Unlike the reduced form analysis in the bivariate probit model, our methodology requires a behavioral assumption in the form of an equilibrium condition for a consumer's optimization where consumers can precisely predict the occurrence of their future accidents. However, using theorems of spatial statistics, it is shown that we cannot distinguish the moral hazard and selection problems without the simultaneity assumption. To check the robustness of our specification, we additionally consider econometric models with an alternative parametrization under the simultaneity assumption.

As an empirical application of our methodology, we analyze the US dental insurance market. Previous studies have consistently detected moral hazard in this market. The incoherency problem has been an obstacle to analyzing selection problems. Our results indicate that moral hazard and advantageous selection are both present. This is a more informative result than that derived from the conventional bivariate probit analysis.

The organization of this paper is as follows. In Section 2, we describe our basic econometric framework. Section 3 considers the feasibility of alternative specifications. The proposed method is applied to the US dental care insurance market in Section 4. Section 5 concludes the paper.

2 Econometric models

2.1 The bivariate probit model: Conventional methodology

This section proposes our basic inference scheme for information asymmetry in insurance markets. We begin by reviewing the conventional methodology, the bivariate probit model (Poirier, 1980). This model considers a market of a simple insurance that covers an accident in the following simple reduced-form econometric model.

The sample consists of N consumers 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 has an implicit assumption that only asymmetric information is the source of ρ and that there is no omitted variable that affects both latent variables.

2.2 Mutually dependent dummy variable models

Next, we present our original econometric methodology that identifies the effects of moral hazard and selection problems separately. Our basic inferential scheme consists of seven distinct econometric models and model selection. These models share a parametrization, but are distinct in support of the parameters. Due to the method of the parametrization, we classify these models as mutually dependent dummy variable models.

To achieve the separate identification for the 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 measures the effect of y_{i2} on y_{i1}. The former term indicates moral hazard, while the latter indicates selection problems. For simplicity, these effects are modeled as linearly additive terms, whose amounts are measured by coefficient parameters α_2 and α_1.

Mutually dependent dummy variable model:

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

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

where the new error term u_{ij} is assumed to be independent of u_{ik}, given $\boldsymbol{x}_i$. This conditional independence is a consequence of the above implicit assumption of no omitted variables in the bivariate probit model. Thus our model does not allow endogeneity of u_{ij} and y_{ik} in the sense that u_{ij} can affect y_{ik} only via y_{ij}. As a result, a positive α_2 corresponds to the existence of moral hazard, while a positive or negative α_1 implies the existence of adverse or advantageous selection.

Despite a simple appearance, we cannot employ a straight-forward estimation procedure for model parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \alpha_1, \alpha_2)'$, because of the mutual dependency between y_{i1} and y_{i2}: 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 for y_{i1}.

One approach to the mutual dependency is to assume that equations (2.3) and (2.4) hold simultaneously. To justify this simultaneity assumption, we need to assign additional restrictions on the underlying economic model. A natural candidate of such a behavioral model is the following optimization process: first, consumers consider y_{i1}, whether to purchase insurance, taking account of y_{i2}, the occurrence of a future accident. Second, consumers re-evaluate their prediction of their future accident occurrence based on this insurance purchase. Third, consumers re-consider their purchasing decision, considering this updated prediction. These feedback loops continue until consumers reach a steady state, where equations (2.3) and (2.4) are simultaneously satisfied. This equilibrium behavior is realized as consumers' final decisions to purchase insurance.

More precisely, we make two assumptions here. The first is the above behavioral model with an infinite feedback system. The second is an implicit assumption that the consumer can predict y_{i2} precisely. Perfect foresight is guaranteed if we assume that the error term, u_{i2}. is completely known to the consumers. The necessity of such strong assumptions is a clear disadvantage to the bivariate probit model, which does not include a behavioral assumption. We consider the possibility of alternative assumptions in the next section.

Under the simultaneity assumption, a straight-forward calculation shows the following correspondences between the outcomes and error terms:

$$z_i = 1 \quad \Leftrightarrow \quad \left\{u_{1i} < -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1, \quad u_{2i} < -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2\right\}, \tag{2.5}$$

$$z_i = 2 \quad \Leftrightarrow \quad \left\{u_{1i} \geq -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1, \quad u_{2i} < -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 - \alpha_2\right\}, \tag{2.6}$$

$$z_i = 3 \quad \Leftrightarrow \quad \left\{u_{1i} < -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 - \alpha_1, \quad u_{2i} \geq -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2\right\}, \tag{2.7}$$

$$z_i = 4 \quad \Leftrightarrow \quad \left\{u_{1i} \geq -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 - \alpha_1, \quad u_{2i} \geq -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 - \alpha_2\right\}, \tag{2.8}$$

where we define a scalar variable $z_i = 1$, 2, 3 and 4 which corresponds to $(y_{i1}, y_{i2}) = (0,0)$, $(1,0)$, $(0,1)$ and $(1,1)$ for notational simplicity.

The simultaneity assumption can manage mutual dependencies, but it induces an econometric problem that Gouriéroux et al. (1980) called the incoherency problem. For our model, the incoherency problem means that the statistical model is not well-defined unless there exists j such that $\alpha_j = 0$.

We illustrate the above correspondence (2.5) - (2.8) on the coordinates of (u_{i1}, u_{i2}). We must consider models separately according to the sign conditions of (α_1, α_2). Specifically, we define four distinct models as follows:

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

with (2.3) and (2.4).

Figure 1 is here

Figure 1 shows two representative models, Models NN and NP, and the remaining Models PP and PN can be analyzed similarly. In both representative Models NN and NP, Regions 1 thorough 4 have a unique pair of outcomes, but Region 5 does not: For Model NN, there are two candidates $(y_{i1}, y_{i2}) = (1, 0)$ and $(0, 1)$, while for Model NP, there is no possible pair of outcomes. Due to this non-uniqueness of outcomes in Region 5, it is not possible to obtain well-defined joint choice probabilities, unless Region 5 vanishes by assuming the coherency condition, namely $\exists i, \alpha_i = 0$.

Recently, such incoherent models have gathered new attention from entry game researchers, because (2.3) and (2.4) are equivalent to the best response functions of a general class of entry games. To manage the incoherency problem, researchers explicitly modeled a data generating system 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 in which each pair of outcomes is realized. Consequently, the choice probability in both Models NN and NP can be written as:

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

where $P_l(\boldsymbol{x}_i, \boldsymbol{\theta})$ measures an area of Region l in Figure 1. Note that Regions 1 through 5 are differently defined in Models NN and NP in Figure 1. Thus parameters $\boldsymbol{\beta_1}, \boldsymbol{\beta_2}$ in Model NN and NP have distinct effects on choice probabilities, and hence on the likelihood function. Therefore, Models NN and NP are indeed different statistical models and not nested.

Because the summation of these choice probabilities with respect to $z_i = 1, 2, 3$ and 4 is unity, the model is now well-defined. However, it is difficult to construct a consistent estimator for $\boldsymbol{\theta}$ which does not depend on a sample-specific nuisance parameter p_{il}. In other words, the problem shifts from the incoherency problem to the incidental parameter problem (Lancaster, 2000).

We can interpret the incoherency problem in this context as follows: in entry game models, the incoherency is caused by a possible non-uniqueness of Nash equilibria. In our setting, the incoherency also corresponds to the non-uniqueness of the best action. Model NN has multiple possible combinations of the insurance purchase decision and the corresponding accident occurrence, while in Model NP, there is no best action.

In addition to Models NN, NP, PN and PP, we have three models with a coherency restriction, namely $\alpha_j = 0$ for some j. Specifically, three models are in this category:

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

with (2.3) and (2.4). CI, MH and AS stand for complete information, moral hazard and "adverse or advantageous" selection, respectively. For these models, it is possible to adopt standard probit estimation because we do not have problems caused by mutual dependencies. We note that the bivariate probit model with zero correlation reduces to Model CI.

Next, we provide a computational methodology for the statistical inference of the above scheme. The main purpose of this study is to characterize the information problem in insurance markets. It is difficult to construct a statistical test for this purpose, because the models are not nested. Instead, we conduct model selection for Models NN, NP, PN, PP, CI, MH and AS to determine a suitable information structure.

We employ a Bayesian procedure which enables us both to overcome the incidental parameter problem for Models NN, NP, PN and PP and to conduct model selection. The procedure uses the Markov chain Monte Carlo (MCMC) method, which we presented for entry game models in Sugawara and Omori (2012). In short, our methodology estimates sample-specific selection rule parameters. Posterior distributions of the selection rule parameters may depend heavily on the prior distribution due to the small sample size, because we can use only one sample for estimation of the selection rule parameter. However, there is no problem of the statistical inference because Bayesian estimation can work with finite samples. Thus, we can use standard estimation and model selection techniques. In addition, Models CI, MH and AS are

estimated using standard Gibbs samplers, as in see Koop et al. (2007).

3 Alternative approaches

This section investigates alternative approaches to the above mutually dependent dummy variable models. For notational simplicity, we drop the subscript for individuals i in this section.

We pursue two directions in this section. First, we examine alternative assumptions regarding the mutual dependency of dependent variables. This analysis is motivated by the simultaneity assumption's requirement of a restrictive behavioral assumption of a consumer's optimization. Our discussion shows that it is not possible to achieve the separate identification of moral hazard and selection problems without the simultaneity assumption. This analysis additionally yields an intuitive explanation for the meaning of an identification problem in this topic. Second, under the simultaneity assumption, we consider a parametrization that is different to mutually dependent dummy variable models. This parametrization induces an alternative econometric model that also consists of seven submodels.

3.1 Inevitability of the simultaneity assumption

To investigate an alternative assumption for the mutual dependency among the dependent variables, we refer to spatial statistics, which has a long history of this problem. This literature has revealed that there are two methodologies to model such a situation. In addition to the simultaneous specification method of the previous section, there is another method called the conditional specification (Besag, 1974).

In the conditional specification, we begin by defining the conditional distributions y_j given y_k, namely $y_j|y_k$. As noted 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 main difficulty of the conditional specification is that we do not always have a corresponding joint distribution for given conditional distributions. To have a well-defined multivariate statistical model, we must recover the joint distribution for dependent variables from conditional distributions. This recoverability is called compatibility and has been actively studied (Arnold et al., 1999).

For the analysis of the compatibility, we follow the methods presented by Besag (1974) and summarized by Cressie (1993). It is difficult to check the compatibility directly in general distributions. Instead, we first assume that we have compatible conditionals and then consider the necessary conditions. Specifically, we begin our discussion assuming that we have conditional distributions for $y_1|y_2$ and $y_2|y_1$, and

there exists a corresponding joint distribution for (y_1, y_2). We do not make an assumption of distributional forms, but assume $\boldsymbol{y} = \mathbf{0}$ occurs with a positive probability. Let us define a negpotential function 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:

$$\frac{Pr(y_j|y_k)}{Pr(y_j = 0|y_k)} = \frac{Pr(\boldsymbol{y})}{Pr(y_j = 0, y_k)} = \exp[Q(\boldsymbol{y}) - Q(y_j = 0, y_k)]. \tag{3.1}$$

Theorem 2
Q can be expanded as:

$$Q(\boldsymbol{y}) = y_1 G_1(y_1) + y_2 G_2(y_2) + y_1 y_2 G_{12}(y_1, y_2), \tag{3.2}$$

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 displays the relationship among the conditional distribution, the joint distribution and the negpotential function, while Theorem 2 shows a unique decomposition of the negpotential function. In our context, Theorem 2 specifies the unique negpotential function as:

$$Q(\boldsymbol{y}) = y_1 G_1(1) + y_2 G_2(2) + y_1 y_2 G_{12}(1, 1). \tag{3.3}$$

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

$$Q(\boldsymbol{y}) - Q(y_j = 0, y_k) = \mu_k y_k + \gamma y_1 y_2. \tag{3.4}$$

Applying Theorem 1, we obtain conditional distributions explicitly as:

$$Pr(y_j = 0|y_k) = \frac{1}{1 + \exp[\mu_j + \gamma y_k]}, \tag{3.5}$$

$$\rightarrow Pr(y_k = 1|y_j) = \frac{\exp[\mu_j + \gamma y_k]}{1 + \exp[\mu_j + \gamma y_k]}. \tag{3.6}$$

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

$$Pr(y_j|y_k) = \frac{y_j\{\exp[\boldsymbol{x}_j' \boldsymbol{\beta_j} + \gamma y_k]\}}{1 + \exp[\boldsymbol{x}_j' \boldsymbol{\beta_j} + \gamma y_k]}, \quad j = 1, 2, \ k \neq j. \tag{3.7}$$

Because γ does not depend on j nor 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. Because this is a direct consequence of Besag's theorems for

any compatible system, it is a necessary condition for compatibility in our context. Therefore, conditional specification cannot distinguish the effects of moral hazard and selection problems because they are to be measured by the same parameter γ. In other word, the simultaneity assumption is required for the separate identification for the two sources of asymmetric information.

We now provide an intuitive interpretation for identification problems in the two specifications. For both specifications, we want to estimate the relationship between two dependent variables; the insurance purchase and the accident occurrence. However, the econometric analysis uses only one pair of outcomes, namely (y_{i1}, y_{i2}), for each sample. Because it is difficult to distinguish the two-way interactions without an additional assumption, the conditional specification can identify the interactions only as one coefficient parameter. The same is true for the bivariate probit model, which identifies the interactions as one correlation parameter. On the other hand, the simultaneous specification achieves the separate identification of two effects in the price of behavioral assumptions.

3.2 Alternative parametrization

Next, we consider an alternative parameterization, keeping the simultaneity assumption. In the mutually dependent dummy variable models presented in Section 2.2, selection problems are measured as an effect of the accident occurrence y_{i2} on the insurance purchase, y_{i1}. Apparently, this parametrization seems to exhibit an reverse causality in that an accident, which should occur after the insurance purchase, affects the purchase decision. However, this is actually not a logical fallacy under the assumption that the error terms are known, because consumers can make a precise prediction.

Alternately, we can consider an alternative parameterization by changing the measurement of the selection problem. For this purpose, we assume that the latent variable y_{i2}^* measures accident risk. Then we define the selection problem as an effect of the accident risk on the insurance purchase, while everything else is unchanged. We call this setting the riskiness models. Specifically:

Riskiness Models:

$$y_{i1}^* = \boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 + \alpha_1 y_{i2}^* + u_{i1}. \tag{3.8}$$

$$y_{i2}^* = \boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 + \alpha_2 y_{i1} + u_{i2}. \tag{3.9}$$

$$y_{ij} = I[y_{ij}^* \geq 0]. \tag{3.10}$$

We need to separate seven models according to the support of (α_1, α_2). First, we define the following four models:

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

with (3.8), (3.9) and (3.10).

This alternative parametrization still has the mutual dependency between the two dependent variables. Adopting the simultaneity assumption, we again have the incoherency problem[1]. To address this problem, we consider the following correspondences between outcomes and error terms:

$$
\begin{aligned}
z_i = 1 \quad &\Leftrightarrow \quad \left\{u_{i1} + \alpha_1 u_2 \leq -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 - \alpha_1 \boldsymbol{x}_{i2}'\boldsymbol{\beta}_2, \quad u_{i2} \leq -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2\right\}, && (3.11) \\
z_i = 2 \quad &\Leftrightarrow \quad \left\{u_{i1} + \alpha_1 u_2 \geq -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 - \alpha_1 \boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 - \alpha_1\alpha_2, \quad u_{i2} \leq -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 - \alpha_2\right\}, && (3.12) \\
z_i = 3 \quad &\Leftrightarrow \quad \left\{u_{i1} + \alpha_1 u_2 \leq -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 - \alpha_1 \boldsymbol{x}_{i2}'\boldsymbol{\beta}_2, \quad u_{i2} \geq -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2\right\}, && (3.13) \\
z_i = 4 \quad &\Leftrightarrow \quad \left\{u_{i1} + \alpha_1 u_2 \geq -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 - \alpha_1 \boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 - \alpha_1\alpha_2, \quad u_{i2} \geq -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 - \alpha_2\right\}. \\
&\quad . && (3.14)
\end{aligned}
$$

Figure 2 is here

Figure 2 illustrates the above correspondences (3.11) - (3.14) on the coordinates of (u_{i2}, u_{i1}). We again concentrate on Models NN and NP. In Figure 2, the four lines are defined as:

$$
\begin{aligned}
A \quad &: \quad u_{i1} + \alpha_1 u_{i2} = -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 - \alpha_1 \boldsymbol{x}_{i2}'\boldsymbol{\beta}_2, && (3.15) \\
B \quad &: \quad u_{i1} + \alpha_1 u_{i2} = -\boldsymbol{x}_{i1}'\boldsymbol{\beta}_1 - \alpha_1 \boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 - \alpha_1\alpha_2, && (3.16) \\
C \quad &: \quad u_{i2} = -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2, && (3.17) \\
D \quad &: \quad u_{i2} = -\boldsymbol{x}_{i2}'\boldsymbol{\beta}_2 - \alpha_2. && (3.18)
\end{aligned}
$$

For both Models NN and NP, Regions 1, 2, 3, 7, 8 and 9 have a unique pair of outcomes, while Regions 4, 5 and 6 do not. We call these regions without a unique pair of outcomes are nonsingular. The coherency condition is $\exists i, \alpha_i = 0$, which makes the two lines A and B be equivalent. If this condition is not satisfied, we again introduce selection rules. When Region k is nonsingular, the data generating process

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

can be modeled using $\boldsymbol{p}_{ik} = (p_{ik1}, p_{ik2}, p_{ik3}, p_{ik4}) \in [0,1]^4$ with $\sum_{l=1}^{4} p_{ikl} = 1$. Now p_{ikl} represents a proportion of Region k for the outcome $z_i = l$. We obtain the choice probability for $z_i = l$ as:

$$Pr[z_i = l|\boldsymbol{x}_i, \boldsymbol{\theta}, \boldsymbol{p}_i] = \sum_{k:z_i=l \text{ uniquely}} P_k(\boldsymbol{x}_i, \boldsymbol{\theta}) + \sum_{k: \text{ nonsingular}} p_{ikl} P_k(\boldsymbol{x}_i, \boldsymbol{\theta}), \quad (3.19)$$

where $P_k(\boldsymbol{x}_i, \boldsymbol{\theta})$ measures an area of Region k in Figure 2 and $\boldsymbol{p}_i$ is a vector whose element is p_{ikl} for all the nonsingular k and $l = 1, 2, 3$ and 4. We present a closed form expression of this choice probability in Appendix A.1.

For models with a coherency restriction, not all the riskiness models are different from those in the mutually dependent dummy variable models. Specifically, we have the same models for Models CI and MH, while Model AS is different:

$$\textit{Riskiness Model AS}\text{: } \alpha_1 \in \mathbb{R}, \alpha_2 = 0, \text{ (3.8), (3.9), and (3.10)}$$

The compatibility analysis induces the same conditional choice probability as in Equation (3.7) in Section 3.1. This compatibility condition 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. Consequently, the separate identification of moral hazard and selection problems is not achieved by the conditional specification in the riskiness models.

4 Empirical analysis for the US dental insurance market

4.1 Data

In this section, we apply our methodology to the US dental care insurance market. This is an ideal market for our research because the dependent variables takes stable and modest figures. Both the rate of insurance coverage and the ratio of those have at least one dental visit per year, which we interpret as the accident occurrence, are consistently about 50 % (Manski et al., 2002). These numbers might indicate the robustness of our dependent variables to unobserved shocks.

There have been several studies about the information problem in this market. Moral hazard has been consistently detected. Mueller and Monheit (1988) moral hazard in a surveyed dataset and Manning et al. (1985) found it in an experimental dataset. Moral hazard in this context means a positive effect of insurance coverage on dental visits, which does not always have a negative implication. For example, a positive effect can imply a successful preventive efforts encouraged by insurance companies or HMOs.

On the other hand, the selection problems have been considered as non-negligible but difficult to identify due to the incoherency problem (Sintonen and Linnosmaa, 2000). One exception is Munkin and Trivedi (2008), who implemented Bayesian estimation for a switching regression model. They adopted insurance purchase and the number of dental visits as the "treatment" and the main outcome variables, respectively. Their approach has two parameters to indicate information asymmetry. One parameter is a coefficient of the insurance purchase on the accident occurrence, similar to our parameter for moral hazard. Another parameter is a sample-selection correlation, which can be affected by both moral hazard or selection problems.

We construct a dataset similar to Munkin and Trivedi (2008). We use 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 25 to 65 years old. Self-employed and governmental workers are eliminated due to their peculiar insurance coverage. Our sample size is 5090.

Two dependent dummy variables are defined as follows. First, the accident occurrence dummy takes unity if the sample individual visits a dental care at least once in the survey period. Second, the insurance purchase dummy takes unity if the consumer has dental insurance at the time of the first interview. Despite having panel data, we only use information from the first interview and ignore the rest of interviews to avoid confusion due to time inconsistency. We adopt this manner also for time-variant explanatory variables below.

[Table 1 is here]

We choose our explanatory variables from four categories: demographic, health, geographic and insurance-purchase specific variables. Table 1 shows descriptive statistics. Our sample is similar to that of Munkin and Trivedi (2008), which is made by MEPS from 1996 to 2000. There are several comments on the definition of explanatory variables as follows.

For the demographic variables, the consumer's ages are 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 dummy. For health variables, self-reported health status is included as categorical dummies; "very good","good" and "fair or poor". "Excellent" is the reference category. In addition, we include the number of chronic conditions, namely diabetes, asthma, high blood pressure, coronary heart disease, emphysema and arthritis. The geographic variables have two groups. One consists of location dummies in which "west" is the reference category, and another is a dummy variable which takes unity if the consumer lives in a metropolitan statistical area. Finally, the firm size, which might have an effect on available in-

surance plans but not on dental visits, is adopted as an explanatory variable for the insurance purchase only.

4.2 Model selection and estimation results

We employ the Bayesian methodology using the MCMC method. Detailed procedures for the mutually dependent dummy variable models and the riskiness models are presented in Sugawara and Omori (2012) and Appendix A, respectively. We further estimate the bivariate probit model to make a comparison with the conventional methodology. The estimation procedure for the bivariate probit model is described in Appendix B.

In our implementation of MCMC samplers, we make a distributional assumption for error terms. In the mutually dependent dummy variable models and the riskiness models, we assumed the standard normal distribution for error terms. Other distributions beside the normal distribution can work, because the only requirement is to define the area probability $P_k(\boldsymbol{x}_i, \boldsymbol{\theta})$, as discussed in Sugawara and Omori (2012) and Appendix A. For the bivariate probit model, we adopt a standard assumption of the normal error terms with unit variances and a correlation ρ.

We use the same hyperparameters upon distinct schemes as follows. The prior covariance matrix for $\boldsymbol{\theta}$ is set as orthogonal. Its diagonal elements, or prior variances, are assumed to be 10 both for $\boldsymbol{\beta}$ 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 at zero. The prior means of α_j depends on its domain in each model. For models without constraint for α_j, the prior mean is set at zero. For models with constraints $\alpha_j < 0$ or $\alpha_j > 0$, the prior means are set at -1 or 1, respectively. For the selection rule $\boldsymbol{p}$, we set hyperparameters such that the prior distribution becomes the uniform. For a parameter that is specific to the bivariate probit model, namely δ, we set the normal prior distribution with mean 0 and variance 10.

We generate different numbers of posterior samples for distinct models so that each Markov chain mixes well. First, for all models in the mutually dependent variable models, Model AS in the riskiness models and the bivariate probit model, we generate 10,000 posterior samples after discarding 5,000 initial samples as the burn-in period. Second, for Models NN and PP in the riskiness models, we run 20,000 iterations after a burn-in of 20,000 iterations. Third, for Models NP and PN in the riskiness models, we run our sampler with 50,000 samples after 20,000 burn-in iterations. In the posterior sampling for Models NN, NP, PN and PP in both mutually dependent dummy variable models and riskiness models, 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 Model selection results

[Table 2 is here]

Table 2 shows the result of model selection using the deviance information criterion (DIC) of Spiegelhalter et al. (2002). The first and the third rows of Table 2 presents the DICs of the models. Additionally, we present values of the likelihood function evaluated at the posterior means in the second and fourth columns, to indicate the goodness of fit of models.

Both criteria choose Model NP in the mutually dependent dummy variable models. This selected model exhibits the existence of moral hazard and advantageous selection. Further, if we consider model selection only among the riskiness models, Model NP is still selected. This result implies the robustness of our model selection.

While the detection of moral hazard is compatible to the previous studies, our important contribution is the discovery of advantageous selection. The adverse selection indicates risk averse preferences of consumers for dental insurance. Because dental illness is often a result of the chronic accumulation of damage to teeth, people might decide to purchase insurance before they need serious dental care. That is, early preventive concerns, which lower risk, would result in the insurance purchase.

The uniqueness of our finding is reinforced when we make a comparison with 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 $[0.089,\ 0.168]$, which indicates 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 or adverse selection. On the other hand, the conventional method cannot detect advantageous selection, which our methodology found. Considering that the bivariate probit model is outperformed by Model NP in the mutually dependent dummy variable models in Table 2, our dataset likely contains advantageous selection. In summary, our methodology has the potential power to shed a new light on information asymmetry which hidden in previous studies.

4.2.2 Estimation results

Based on the above model selection result, we present detailed estimation results for Model NP in the mutually dependent dummy variable models. Figures 3, 4 and 5 in Appendix C show the paths of the posterior samples where they mix well and are stable. The convergences of α_1 and α_2 seem 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, namely more than 0.95 for all the model parameters, to guarantee

the efficiency of our Metropolis-Hastings algorithm.

[Table 3 is here]

Table 3 shows estimation results for the coefficient parameters of Model NP in the mutually dependent dummy variable models. The effects of the health status variables are generally compatible to those of Munkin and Trivedi (2008). For self-reported health, "Fair and poor" and "Good" status, have a negative relationship with insurance purchase compared to the reference category of excellent health. This result supports advantageous selection, where healthier people are more likely to purchase dental insurance. However, this story does not hold for chronic illness, which has a positive effect on insurance purchase. These health variables have only minor impacts on the dental visit, which might indicate that general health status is not strongly correlated with dental health.

For the demographic variables, the age categories have minor effects on the insurance purchase. People older than 45 years of age receive a dental care more frequently than younger people, possibly because dental illness is often caused by accumulated damages. The positive effects of education both on the insurance purchase and a dental visit implies positive relationship between these behaviors and cognitive ability. Being female has positive effects both on insurance purchase and a dental visit, while an interaction term between gender and age has negative effects on a dental visit. These results indicate a gender difference on the consumer's behavior. For racial variables, we observe the same interesting results as in Munkin and Trivedi (2008) where Hispanics are less likely to purchase dental insurance while African-Americans are less likely to visit a dentist. Further, employees of larger firms are more likely to have dental insurance, which might indicate the generosity of larger firms in provision of benefits.

5 Conclusion

This paper has proposed an econometric methodology to analyze moral hazard and selection problems separately in insurance markets. We need to impose a behavioral assumption on the consumer's optimization to ensure the separate identification of the two sources of asymmetric information. Our applied study has shown that moral hazard and advantageous selection are present in the US dental insurance market. This result is more informative than that of the conventional methodology.

As summarized in Einav et al. (2010a), a literature of structural estimation for empirical insurance analysis has emerged. Our simple econometric model can be used to determine an appropriate model for an intensive structural analysis. For example, our result indicates that we can concentrate on structural models with moral hazard and advantageous selection to analyze the US dental insurance market.

Acknowledgment

Authors would like to thank Yuji Genda, Hidehiko Ichimura, Satoshi Kitatoka, Sayaka Nakamura, Teruo Nakatsuma, Ryo Okui, and Dale Poirier for helpful comments. This work is supported by the Research Fellowship(DC2) from the Japanese Society for the Promotion of Science. The computational results are obtained by using Ox version 6(Doornik, 2011).

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.

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 J. J. Heckman and E. Leamer eds. *Handbook of Econometrics 5*, Amsterdam: North Holland.

Chib, S. and E. Greenberg (1998) "Analysis of multivariate probit models," *Binometrica*, Vol. 85, No. 2, pp. 347–361.

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.

Doornik, J. A. (2011) *Object-Oriented Matrix Programming Using Ox*, London: Timberlake Consultants Press and Oxford, 6th edition.

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.

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.

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, C. (1993) "Identification of endogenous social effects: The refrection problem," *Review of Economic Studies*, Vol. 60, No. 3, pp. 531–542.

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.

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 T. B. Fomby 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 J. Newhouse 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*. forthcoming.

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

A Estimation details for riskiness models

A.1 Choice probabilities

This appendix presents detailed analysis for the riskiness models, which is described in Section 3.2. First, we provide closed forms of the choice probabilities (3.19). For notational simplicity, we define the following variables:

$$
\begin{aligned}
u'_{i1} &= u_{i1} + \alpha_1 u_{i2}, && \text{(A.1)}\\
W_{i1} &= -\boldsymbol{x}'_{i1}\boldsymbol{\beta}_1 - \alpha_1 \boldsymbol{x}'_{i2}\boldsymbol{\beta}_2 && \text{(A.2)}\\
W_{i2} &= -\boldsymbol{x}'_{i1}\boldsymbol{\beta}_1 - \alpha_1 \boldsymbol{x}'_{i2}\boldsymbol{\beta}_2 - \alpha_1\alpha_2 && \text{(A.3)}\\
W_{i3} &= -\boldsymbol{x}'_{i2}\boldsymbol{\beta}_2 && \text{(A.4)}\\
W_{i4} &= -\boldsymbol{x}'_{i2}\boldsymbol{\beta}_2 - \alpha_2. && \text{(A.5)}
\end{aligned}
$$

Using these variables, we have simple expressions for the lines A, B, C and D which are defined in Section 3.2 as

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

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

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:

$$
\begin{aligned}
Pr(z_i = 1|\boldsymbol{\theta}, \boldsymbol{p}_i) &= P_7(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i4} P_4(\boldsymbol{x}_i, \boldsymbol{\theta}), && \text{(A.8)}\\
Pr(z_i = 2|\boldsymbol{\theta}, \boldsymbol{p}_i) &= P_1(\boldsymbol{x}_i, \boldsymbol{\theta}) + P_2(\boldsymbol{x}_i, \boldsymbol{\theta}) + (1 - p_{i4}) P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i5} P_5(\boldsymbol{x}_i, \boldsymbol{\theta}), && \text{(A.9)}\\
Pr(z_i = 3|\boldsymbol{\theta}, \boldsymbol{p}_i) &= P_8(\boldsymbol{x}_i, \boldsymbol{\theta}) + P_9(\boldsymbol{x}_i, \boldsymbol{\theta}) + (1 - p_{i5}) P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) + p_{i6} P_6(\boldsymbol{x}_i, \boldsymbol{\theta}), && \text{(A.10)}\\
Pr(z_i = 4|\boldsymbol{\theta}, \boldsymbol{p}_i) &= P_3(\boldsymbol{x}_i, \boldsymbol{\theta}) + (1 - p_{i6}) P_6(\boldsymbol{x}_i, \boldsymbol{\theta}), && \text{(A.11)}
\end{aligned}
$$

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 (u'_{1i}, u_{i2}). Specifically:

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

$$P_4(\boldsymbol{x}_i,\boldsymbol{\theta}) = Pr[(u_{i1},u_{i2}) \le (W_{i1},W_{i3})'] - P_7(\boldsymbol{x}_i,\boldsymbol{\theta}), \tag{A.13}$$

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

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

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

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

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

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

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

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:

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

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

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

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

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[(u'_{i1},u_{i2}) \leq (W_{i1},W_{i4})], && \text{(A.25)}\\
P_4(\boldsymbol{x}_i,\boldsymbol{\theta}) &= Pr[(u'_{i1},u_{2}) \leq (W_{i2},W_{i4})] - P_7(\boldsymbol{x}_i,\boldsymbol{\theta}), && \text{(A.26)}\\
P_1(\boldsymbol{x}_i,\boldsymbol{\theta}) &= Pr[u_{i2} \leq W_{i4}] - P_4(\boldsymbol{x}_i,\boldsymbol{\theta}) - P_7(\boldsymbol{x}_i,\boldsymbol{\theta}), && \text{(A.27)}\\
P_8(\boldsymbol{x}_i,\boldsymbol{\theta}) &= Pr[(u'_{i1},u_{2}) \leq (W_{i1},W_{i3})] - P_7(\boldsymbol{x}_i,\boldsymbol{\theta}), && \text{(A.28)}\\
P_5(\boldsymbol{x}_i,\boldsymbol{\theta}) &= Pr[(u'_{i1},u_{i2}) \leq (W_{i2},W_{i3})] - P_4(\boldsymbol{x}_i,\boldsymbol{\theta}) - P_7(\boldsymbol{x}_i,\boldsymbol{\theta}) - P_8(\boldsymbol{x}_i,\boldsymbol{\theta}), && \text{(A.29)}\\
P_2(\boldsymbol{x}_i,\boldsymbol{\theta}) &= Pr[u_{i2} \leq 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{(A.30)}\\
P_9(\boldsymbol{x}_i,\boldsymbol{\theta}) &= Pr[u'_{i1} \leq W_{i1}] - P_7(\boldsymbol{x}_i,\boldsymbol{\theta}) - P_8(\boldsymbol{x}_i,\boldsymbol{\theta}), && \text{(A.31)}\\
P_6(\boldsymbol{x}_i,\boldsymbol{\theta}) &= Pr[u'_{i1} \leq 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{(A.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{(A.33)}
\end{aligned}
$$

A.2 Bayesian estimation

A.2.1 Additional hierarchical structure and the likelihood function

This appendix provides an estimation procedure for the riskiness models. 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 $\lambda_{ik}|p_{ik} \sim$ Bernoulli(p_{ik}) for $k = 4$, 5 and 6. Second, for Model NP, we adopt $\boldsymbol{\lambda_{ik}} = (\lambda_{ik,1}, \lambda_{ik,2}, \lambda_{ik,3}, \lambda_{ik,4})|\boldsymbol{p}_{ik} \sim$ 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:

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

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

$$Pr(z_i = 1|\boldsymbol{\theta}, \boldsymbol{\lambda}_i) = P_7(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i4} P_4(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{A.35}$$

$$Pr(z_i = 2|\boldsymbol{\theta}, \boldsymbol{\lambda}_i) = P_1(\boldsymbol{x}_i, \boldsymbol{\theta}) + P_2(\boldsymbol{x}_i, \boldsymbol{\theta}) + (1 - \lambda_{i4}) P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) + (1 - \lambda_{i5}) P_5(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{A.36}$$

$$Pr(z_i = 3|\boldsymbol{\theta}, \boldsymbol{\lambda}_i) = P_8(\boldsymbol{x}_i, \boldsymbol{\theta}) + P_9(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i5} P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i6} P_6(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{A.37}$$

$$Pr(z_i = 4|\boldsymbol{\theta}, \boldsymbol{\lambda}_i) = P_3(\boldsymbol{x}_i, \boldsymbol{\theta}) + (1 - \lambda_{6i}) P_6(\boldsymbol{x}_i, \boldsymbol{\theta}). \tag{A.38}$$

For Model NP:

$$Pr(z_i = 1|\boldsymbol{\theta}, \boldsymbol{\lambda}_i) = P_7(\boldsymbol{x}_i, \boldsymbol{\theta}) + P_8(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i4,1} P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i5,1} P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i6,1} P_6(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{A.39}$$

$$Pr(z_i = 2|\boldsymbol{\theta}, \boldsymbol{\lambda}_i) = P_1(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i4,2} P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i5,2} P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i6,2} P_6(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{A.40}$$

$$Pr(z_i = 3|\boldsymbol{\theta}, \boldsymbol{\lambda}_i) = P_9(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i4,3} P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i5,3} P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i6,3} P_6(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{A.41}$$

$$Pr(z_i = 4|\boldsymbol{\theta}, \boldsymbol{\lambda}_i) = P_2(\boldsymbol{x}_i, \boldsymbol{\theta}) + P_3(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i4,4} P_4(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i5,4} P_5(\boldsymbol{x}_i, \boldsymbol{\theta}) + \lambda_{i6,4} P_6(\boldsymbol{x}_i, \boldsymbol{\theta}). \tag{A.42}$$

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

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

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 $P_k(\boldsymbol{x}_i, \boldsymbol{\theta})$.

A.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 $\boldsymbol{a}_{ik} = (a_{ik1}, \ldots, a_{ik4})$:

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

A.2.3 Posterior sampling for Riskiness Model NN

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:

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

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:

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

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

where

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

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.

A.2.4 Posterior sampling for Riskiness Model NP

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

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

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

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

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

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

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

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.

B Bayesian estimation for the bivariate probit model

This appendix presents the detail of our Bayesian estimation procedure for the bivariate probit model. It is similar to that of Chib and Greenberg (1998) for the multivariate probit model, but is simpler because we have only two dependent variables.

In the bivariate probit model (2.1) and (2.2), we place a standard assumption of the normal errors:

$$\begin{pmatrix}\epsilon_{i1}\\ \epsilon_{i2}\end{pmatrix}\sim N(\mathbf{0},\Sigma), \quad \Sigma=\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}. \tag{B.1}$$

For notational simplicity, we let $\boldsymbol{y}_i^*=(y_{i1}^*,y_{i2}^*)'$ and $\boldsymbol{y}^*=[(\boldsymbol{y}_1^*)',(\boldsymbol{y}_2^*)',...,(\boldsymbol{y}_N^*)']'$. We obtain the likelihood function given the latent variables as:

$$f(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\rho|\boldsymbol{y}^*)=(2\pi)^{-N}|\Sigma|^{-N/2}\exp\Big(-\frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{y}_i^*-X_i\boldsymbol{\beta})'\Sigma^{-1}(\boldsymbol{y}_i^*-X_i\boldsymbol{\beta})\Big), \tag{B.2}$$

where

$$X_i=\begin{pmatrix}\boldsymbol{x}_{i1}' & \mathbf{0}\\ \mathbf{0} & \boldsymbol{x}_{i2}'\end{pmatrix}, \quad \boldsymbol{\beta}=\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}. \tag{B.3}$$

For the prior distributions, we use the normal distributions for $\boldsymbol{\beta}$:

$$\boldsymbol{\beta}\sim N(\boldsymbol{\mu}_{\beta,0},\Sigma_{\beta,0}). \tag{B.4}$$

On the other hand, for the covariance parameter, we must have the positive definiteness of the covariance matrix, or $|\rho|\leq 1$. To impose this condition, we use the

following Fisher transformation $\delta = (1/2)\ln[(1+\rho)/(1-\rho)]$. Using this transformation, $|\rho| < 1$ corresponds to $\delta \in (-\infty, \infty)$. Then we adopt the unconstrained normal prior distribution for δ:

$$\delta \sim N(\mu_{\delta,0}, \sigma^2_{\delta,0}). \tag{B.5}$$

For the posterior analysis, our MCMC sampler consists of the three parts; (1) Data augmentation for $\boldsymbol{y}^*|\boldsymbol{\beta}, \rho, \boldsymbol{y}$, (2) Posterior sampling for $\boldsymbol{\beta}|\rho, \boldsymbol{y}^*$ and (3) Posterior sampling for $\delta|\boldsymbol{\beta}, \boldsymbol{y}^*$. This subsection describes all the steps.

The first part is the data augmentation for the latent variables $\boldsymbol{y}^*_i$. We adopt the following two-step sequential sampling. First, we generate y^*_{i1} from a marginal distribution. Then, we generate y^*_{i2} from a conditional distribution given the generated value of y^*_{i1}. The distributions are truncated according to the observed values of $\boldsymbol{y}_i$. In other words, for $i = 1, 2, ..., N$:

$$y^*_{i1}|\boldsymbol{\beta}, \rho, \boldsymbol{y}_i \sim \begin{cases} TN_{(-\infty,0]}(\boldsymbol{x}'_{i1}\boldsymbol{\beta}_1, 1) & \text{if} \quad y_{i1} = 0 \\ TN_{(0,\infty)}(\boldsymbol{x}_{i1}\boldsymbol{\beta}_1, 1) & \text{if} \quad y_{i1} = 1 \end{cases}, \tag{B.6}$$

$$y^*_{i2}|\boldsymbol{\beta}, \rho, \boldsymbol{y}_i, y^*_{1i} \sim \begin{cases} TN_{(-\infty,0]}[\boldsymbol{x}'_{i2}\boldsymbol{\beta}_2 + \rho(y^*_{i1} - \boldsymbol{x}'_{i1}\boldsymbol{\beta}_1), 1-\rho^2] & \text{if} \quad y_{i2} = 0 \\ TN_{(0,\infty)}[\boldsymbol{x}'_{i2}\boldsymbol{\beta}_2 + \rho(y^*_{i1} - \boldsymbol{x}'_{i1}\boldsymbol{\beta}_1), 1-\rho^2] & \text{if} \quad y_{i2} = 1 \end{cases}. \tag{B.7}$$

The second part is the posterior sampling for $\boldsymbol{\beta}$. This part can be implemented using the Gibbs sampler from the following conditional posterior distribution:

$$\boldsymbol{\beta}|\rho, \boldsymbol{y} \sim N(\boldsymbol{\mu}_{\beta,1}, \Sigma_{\beta,1}), \tag{B.8}$$

where

$$\Sigma_{\beta,1} = \Big[\sum_{i=1}^{N} X'_i \Sigma^{-1} X_i + \Sigma^{-1}_{\beta,0}\Big]^{-1}, \tag{B.9}$$

$$\boldsymbol{\mu}_{\beta,1} = \Sigma_{\beta,1}\Big[\sum_{i=1}^{N} X'_i \Sigma^{-1} \boldsymbol{y}^*_i + \Sigma^{-1}_{\beta,0}\boldsymbol{\mu}_{\beta,0}\Big]. \tag{B.10}$$

The third part is the posterior analysis for the covariance parameter. In this step, we construct a sampler for the transformed parameter δ. This part can be implemented using the Metropolis-Hastings algorithm for the following conditional posterior density:

$$\pi(\delta|\boldsymbol{\beta}, \boldsymbol{y}^*) \propto |\Sigma(\delta)|^{1-N/2} \exp\left(-\frac{1}{2} \left(\sum_{i=1}^{N} (\boldsymbol{y}_i^* - X_i\boldsymbol{\beta})' \Sigma(\delta)^{-1} (\boldsymbol{y}_i^* - X_i\boldsymbol{\beta}) + \frac{(\delta - \mu_{\delta,0})^2}{\sigma_{\delta,0}^2} \right) \right). \tag{B.11}$$

For the proposal distribution, we use the normal distribution whose mean and variance are the mode and the inverse of the negative Hessian for the logarithm of the conditional posterior density.

C Posterior sample paths for the MCMC samplers

Figure 3 is here

Figure 4 is here

Figure 5 is here

D Tables and Figures

	Mean	S.D.
Insurance purchase (y_1)	0.537	(0.499)
Dental visit (y_2)	0.550	(0.498)
$30 \leq$ 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

	Mutually dependent dummy variable models		Riskiness models	
	DIC	Likelihood	DIC	Likelihood
NN	12406	-6154.4	12407	-6154.4
NP	12289	-6093.0	12296	-6095.5
PN	12319	-6109.1	12441	-6177.5
PP	12299	-6099.3	12299	-6099.4
CI	12403	-6152.4		
AS	12316	-6108.1	12334	-6117.4
MH	12297	-6098.4		
Bivariate Probit	12335	-6117.8		

Table 2: Model selection via DIC and likelihood at posterior mean

	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 Mutually Dependent Dummy Variable Model NP: Posterior means, standard deviations and 95% credible intervals

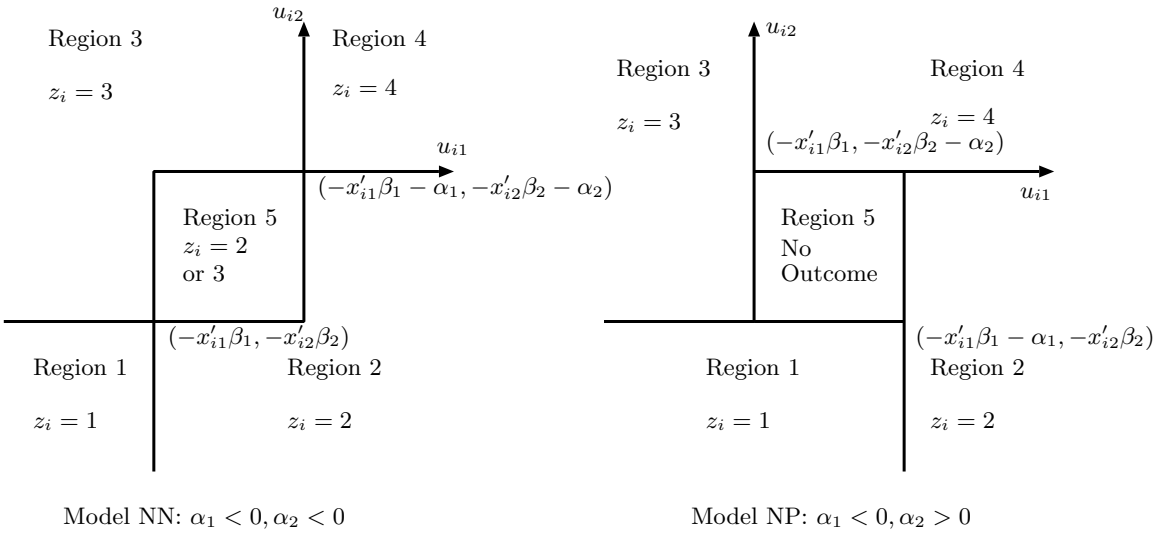

Figure 1: Data generating process for mutually dependent dummy variable models in (u_{i1}, u_{i2}) coordinates

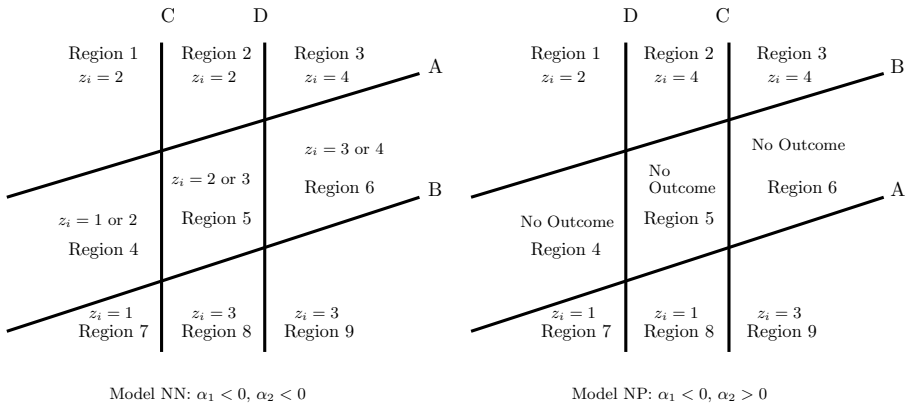

Figure 2: Data generating process for riskiness models in (u_{i2}, u_{i1}) coordinates

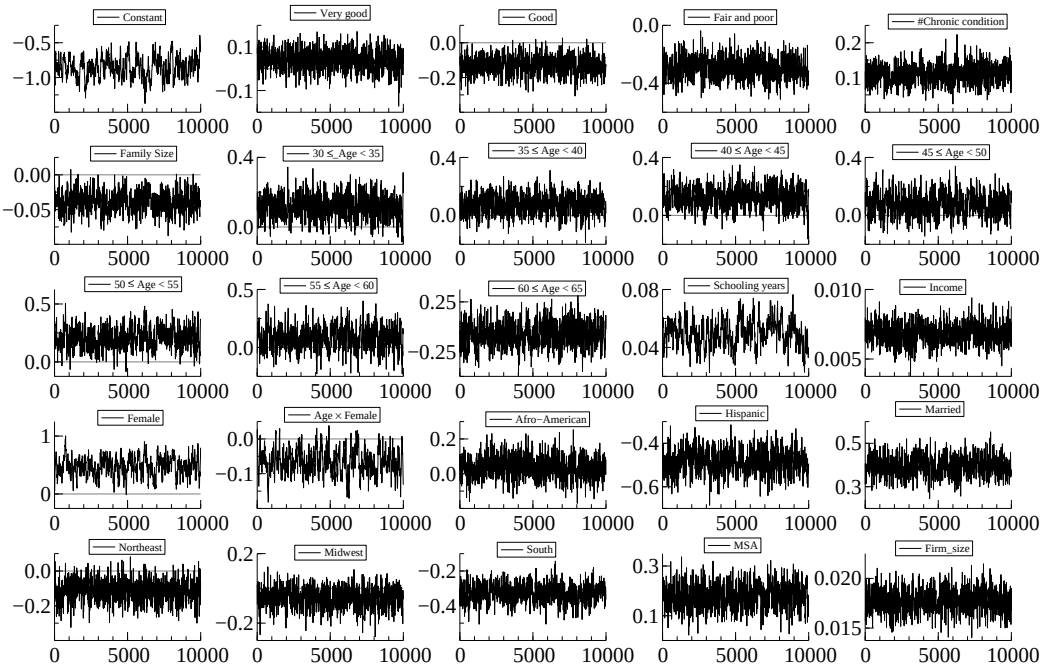

Figure 3: Sample paths of $\boldsymbol{\beta}_1$ for Mutually Dependent Dummy Variable Model NP

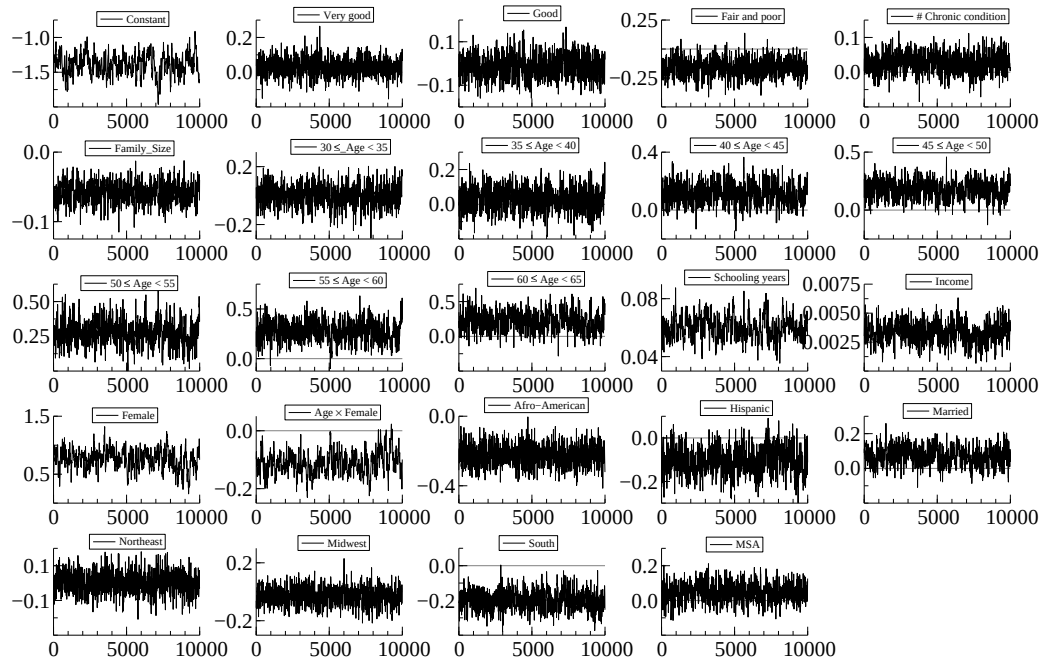

Figure 4: Sample paths of $\boldsymbol{\beta}_2$ for Mutually Dependent Dummy Variable Model NP

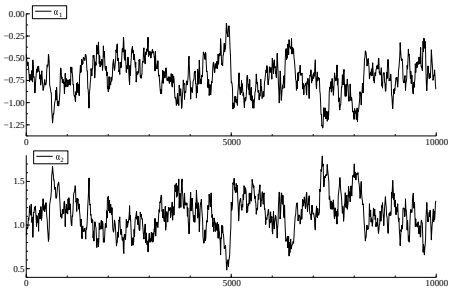

Figure 5: Sample paths of $\boldsymbol{\alpha}$ for Mutually Dependent Dummy Variable Model NP