Identification in a Generalization of Bivariate Probit Models with Endogenous Regressors

Sukjin Han
Department of Economics
University of Texas at Austin
sukjin.han@austin.utexas.edu

Edward J. Vytlacil
Department of Economics
New York University
vytlacil@nyu.edu

First Draft: September 4, 2013
This Draft: November 24, 2013

Abstract

This paper provides identification results for a class of models specified by a triangular system of two equations with binary endogenous variables. The joint distribution of the latent error terms is specified through a parametric copula structure that satisfies a particular dependence ordering that is related to the degree of the first-order stochastic dominance, while the marginal distributions are allowed to be arbitrary but known. This class of models is broad and includes bivariate probit models as a special case. The paper demonstrates that having an exclusion restriction is necessary and sufficient for globally identification in a model without common exogenous covariates, where the excluded variable is allowed to be binary. Having an exclusion restriction is sufficient but not necessary in models with common exogenous covariates that are present in both equations.

Keywords: Identification, triangular threshold crossing model, bivariate probit model, endogenous variables, binary response, copula, exclusion restriction.

JEL Classification Numbers: C35, C36.

1 Introduction

This paper examines the identification of a parametric class of bivariate threshold crossing models that nests bivariate probit models as a special case. The bivariate probit model
was introduced in Heckman (1978) as one specification of simultaneous equations models for latent variables, and is commonly used in applied studies, such as Evans and Schwab (1995), Neal (1997), Goldman et al. (2001) and Bhattacharya et al. (2006). Although the model has drawn much attention in the literature, relatively little research has been done to analyze the identification even in this restricted model.¹

The two papers in the literature which have studied identification of bivariate probit models are Freedman and Sekhon (2010) and Wilde (2000). Freedman and Sekhon (2010) provide formal identification results for bivariate probit models, though they assume (and their proof strategy critically relies upon the assumption) that one of the exogenous regressors has large support. The large support condition is restrictive and limits the applicability of their analysis. Wilde (2000) also considers the identification of bivariate probit models. His identification analysis is limited to simply counting the number of unknown parameters and number of informative non-redundant probabilities in the likelihood function, i.e., the number of equations. His analysis only establishes a necessary condition for global identification since there may still exist multiple solutions in a system of nonlinear equations where the number of equations is at least as large as the number of unknown parameters.

In this paper, we derive identification results for a class of models specified by a triangular system of two equations with binary endogenous variables, where we generalize the bivariate normality assumption on the latent error terms of a bivariate probit model through the use of copulas. In particular, instead of requiring that the joint distribution of latent error terms be bivariate normal, we allow the marginal distributions to be arbitrary but known, while restricting their dependence structure through a broad class of parametric copulas that includes the normal copula as a special case. All results derived in this paper also apply to the special case of bivariate probit models. We first provide identification results in a model without common exogenous regressors, showing that in such a model, having a valid exclusion restriction (i.e., instrument) is necessary and sufficient for global identification of the model. Unlike Freedman and Sekhon (2010), this result does not require a full support condition, and holds even if the instrument is binary. Unlike Wilde (2000), we show that a bivariate normal distribution is not necessary for identification. We then extend the result to a model where we allow for the possibility of exogenous covariates

¹Heckman (1978) discusses identification via a maximum likelihood estimation framework in a model where one of the latent dependent variables is observed in the simultaneous equations model. In a framework where both are not observed; however, identification analysis through calculating the second derivative of a maximum likelihood criterion function is problematic since it is analytically hard to solve.
that enter both equations and the possibility of instruments $Z$ being vector valued without requiring any element of $Z$ to be binary.

While having an exclusion restriction is necessary and sufficient for identification in a model without common exogenous variables, it is sufficient but not necessary in models with common exogenous covariates. We show that identification can still be achieved in a model without exclusion restrictions (i.e., instruments) if there are common exogenous regressors in both equations. This result is reminiscent of identification of the Heckman (1979) sample selection model under normality, where the function form restriction implied by the normal selection model allows identification without excluded instruments. The structure that we impose on the copula enables us to exploit the nonlinear aspect of the model to achieve identification. As other related works, Altonji et al. (2005, 2008) consider identification and estimation in a triangular system similar to that in this paper, where excluded instruments are absent. Their strategy is to assume that the degree of selection on observables can be informative about the degree of selection on unobservables. Another example of analysis providing identification of models with endogenous regressors without the use of instruments is Klein and Vella (2010), who establish identification in a linear triangular simultaneous equations models with only common exogenous regressors by exploiting heteroscedasticity restrictions on the error terms.

Our use of copulas to characterize the joint distribution of the latent error terms allows us to separate the error terms’ dependence structure from their marginal distributions. Our analysis shows that identification is obtained through a condition on the copula, with the shape of the marginal distributions playing no role in the analysis. The condition we impose on the copula is that it satisfies a particular dependence ordering with respect to a single dependence parameter. The condition is that the copula is indexed by a dependence parameter that is informative about the degree of dependence in the sense of the first-order stochastic dominance “FOSD.” We show that this condition is satisfied by a broad range of single-parameter copulas including the normal copula. Thus, the assumption used in the literature that the latent variables follow a bivariate normal distribution is not critical in deriving identification results in this type of models. Our use of copulas is related to Lee (1983), who uses a normal copula to generalize normal selection models. Chiburis (2010) is also related to our analysis. He introduces a normal copula to characterize the joint distribution of latent variables in a similar setting as in this paper, although no rigorous identification analysis is conducted for our class of models.

The paper is organized as follows. In the next section, we introduce the model and
preliminary assumptions. Section 3 introduces a dependence ordering and related concepts that are used to define the class of models we analyze. Section 4 shows identification of a simple, special case of our model. Section 5 extends the identification analysis to the full model, which also nests the model without excluded instruments. Section 6 concludes.

2 The Model

Let \( Y \) denote the binary outcome variable and \( D \) the observed binary endogenous treatment variable. Let \( X \equiv (1, X_1, ..., X_k)' \) denote the vector of regressors that determine both \( Y \) and \( D \), and let \( Z \equiv (1, Z_1, ..., Z_l)' \) denote a vector of regressors that directly effect \( D \) but not \( Y \) (variables excluded from the model for \( Y \), i.e., instruments for \( D \)). We consider a bivariate triangular system for \((Y, D)\):

\[
Y = 1[X'\beta + \delta_1 D - \varepsilon \geq 0], \\
D = 1[X'\alpha + Z'\gamma - \nu \geq 0],
\]

(2.1)

where \( \alpha \equiv (\alpha_0, \alpha_1, ..., \alpha_k)' \), \( \beta \equiv (\beta_0, \beta_1, ..., \beta_k)' \), and \( \gamma \equiv (\gamma_1, \gamma_2, ..., \gamma_l)' \).

We will maintain the following assumptions.

Assumption 1 \((X, Z) \perp (\varepsilon, \nu)\), where “\( \perp \)” denotes statistical independence.

Assumption 2 \( F_\varepsilon \) and \( F_\nu \) are known marginal distributions of \( \varepsilon \) and \( \nu \), respectively, that are strictly increasing, are absolutely continuous with respect to Lebesgue measure, and such that \( E[\varepsilon] = E[\nu] = 0 \) and \( \text{Var}(\varepsilon) = \text{Var}(\nu) = 1 \).

Assumption 3 \((\varepsilon, \nu)' \sim F_\varepsilon \nu(\varepsilon, \nu) = C(F_\varepsilon(\varepsilon), F_\nu(\nu); \rho)\) where \( C(\cdot, \cdot; \rho) \) is a copula known up to scalar parameter \( \rho \in \Omega \) such that \( C : (0, 1)^2 \to (0, 1) \).

Assumption 1 imposes that \( X \) and \( Z \) are exogenous. Assumption 2 characterizes the restrictions imposed on the marginal distributions of \( \varepsilon \) and \( \nu \). The moment restrictions are merely normalizations as long as the second moments of \( \varepsilon \) and \( \nu \) are finite. While we assume that the marginal distributions of \( \varepsilon \) and \( \nu \) are known, the restrictions placed on these marginal distributions are weak. In Assumption 3, the copula associated with the joint distribution is unique by Sklar’s theorem. This assumption specifies that the joint dependence between \( \varepsilon \) and \( \nu \) is fully characterized by a scalar parameter \( \rho \). In a special
case of a bivariate normal distribution discussed below, \(\rho\) is the usual correlation coefficient of \((\varepsilon, \nu)\) with \(\Omega = [-1, 1]\). Note that the endogeneity of \(D\) comes from allowing \(\rho\) to be nonzero. Let \(\tilde{\Psi}\) be the parameter space of \(\tilde{\psi} \equiv (\alpha', \beta', \delta_1, \gamma', \rho)'\).

3 Dependence Orderings for Copulas

In order to obtain meaningful results in identification, we impose additional dependence structure on the copula function of Assumption 3. We show that this structure is embodied in many well-known copulas, including a normal copula. In order to state our condition, we first define the following dependence ordering properties. See Joe (1997) for further discussions on various dependence ordering properties of multivariate distributions or copulas.

Definition 3.1 (Stochastically Increasing “SI”) For r.v.’s \(W_1\) and \(W_2\), \(W_2\) is SI in \(W_1\) or the conditional distribution \(F_{2|1}(w_2|w_1)\) is SI if \(\Pr[W_2 > w_2|W_1 = w_1] = 1 - F_{2|1}(w_2|w_1)\) is increasing in \(w_1\) for all \(w_2\).

The SI property is a positive dependence condition as \(W_2\) is more likely to take on larger values as \(W_1\) increases. This condition is equivalent to the first-order stochastic dominance “FOSD” in the literature. For negative dependence, stochastically decreasing “SD” property can be defined analogously, where \(\Pr[W_2 > w_2|W_1 = w_1]\) is decreasing in \(w_1\). In the following, we define a concept of dependence ordering between two distributions where one is more SI (or less SD) than the other.

Definition 3.2 (Strictly More SI or Less SD) Let \(F_{2|1}(w_2|w_1)\) and \(\tilde{F}_{2|1}(w_2|w_1)\) be respective conditional distributions of the second r.v. given the first. Suppose that \(F_{2|1}(w_2|w_1)\) and \(\tilde{F}_{2|1}(w_2|w_1)\) are continuous in \(w_2\) for all \(w_1\). Then \(\tilde{F}_{2|1}\) is strictly more SI (or less SD) than \(F_{2|1}\) if \(\psi(w_1, w_2) \equiv \tilde{F}^{-1}(F(w_2|w_1)|w_1)\) is strictly increasing in \(w_1\),\(^2\) which is denoted as \(F_{2|1} \prec_S \tilde{F}_{2|1}\).

This ordering is equivalent to having a ranking in FOSD. Let \((W_1, W_2) \sim F\) and \((\tilde{W}_1, \tilde{W}_2) \sim \tilde{F}\). When \(\tilde{F}_{2|1}\) is strictly more SI (less SD) than \(F_{2|1}\), then \(\Pr[\tilde{W}_2 > w_2|\tilde{W}_1 = w_1]\) increases even more than \(\Pr[W_2 > w_2|W_1 = w_1]\) as \(w_1\) increases. More formally, if \(\psi(w_1, w_2)\) is a solution to \(\Pr[\tilde{W}_2 > \psi(w_1, w_2)|\tilde{W}_1 = w_1] = \Pr[W_2 > w_2|W_1 = w_1]\), then

\(^2\)Note that \(\psi(w_1, w_2)\) is increasing in \(w_2\) by definition.
\( \psi(w_1, w_2) \) takes a larger value to compensate that \( W_2 \) is even more likely to take on larger values with \( \tilde{F} \) than it is with \( F \) as \( w_1 \) increases. The SI dependence ordering has been called the (strictly) “more regression dependent” or “more monotone regression dependent” ordering in the statistics literature. Let \( C(\cdot; \rho) \) be the conditional copula of \( C(\cdot; \rho) \).

**Assumption 4 (SI/SD Ordering in \( \rho \))** The copula \( C(u_1, u_2; \rho) \) of Assumption 3 is twice differentiable in \( u_1, u_2 \) and \( \rho \), and satisfies

\[
C(u_1 | u_2; \rho_1) \prec_{S} C(u_1 | u_2; \rho_2) \text{ for any } \rho_1 < \rho_2.
\] (3.1)

Assumption 4 defines a class of copulas within which we derive identification results. It requires that the copula satisfies the SI/SD ordering or equivalently FOSD ordering with respect to the dependence parameter \( \rho \).

The following lemma provides a useful result for our identification analysis in models with excluded instruments. Let \( C_1(\cdot; \cdot; \rho) \) and \( C_\rho(\cdot; \cdot; \rho) \) denote the derivatives of \( C(\cdot; \cdot; \rho) \) with respect to the first argument and \( \rho \), respectively.

**Lemma 3.1** Assumption 4 implies that, for any \( u_1 \in (0, 1) \) and \( \rho \in \Omega \),

\[
\frac{C_\rho(u_1, u_2; \rho)}{C_1(u_1, u_2; \rho)} \text{ is strictly monotonic in } u_2 \in (0, 1).
\] (3.2)

The proofs of this lemma and other results below are found in the Appendix.

Interestingly, with the normal copula defined below, \( \frac{C_\rho(u_1, u_2; \rho)}{C_1(u_1, u_2; \rho)} \) becomes a function of the inverse Mill’s ratio and (3.2) is easily verified; see the Appendix. In the following, we define another concept of dependence ordering called concordance ordering, which is more general than the SI/SD ordering.

**Definition 3.3 (Strictly More Concordant)** Let \( F(w_1, w_2) \) and \( \tilde{F}(w_1, w_2) \) be bivariate distributions. Then \( \tilde{F} \) is strictly more concordant than \( F \) if \( F(w_1, w_2) \prec F(w_1, w_2) \) for any \( w_1 \) and \( w_2 \), which is denoted as \( F \prec_{C} \tilde{F} \).

In the next lemma, we show that the SI/SD ordering implies the concordance ordering. This result is useful for our identification analysis in models without excluded instruments.

**Lemma 3.2** Assumption 4 implies that, for any \( (u_1, u_2) \in (0, 1) \times (0, 1) \) and \( \rho \in \Omega \),

\[
C_\rho(u_1, u_2; \rho) > 0.
\] (3.3)

\(^3\)Note that this assumption implicitly assume that the copula of Assumption 3 is SI.
The SI/SD ordering is not symmetric in arguments in general, but is symmetric for symmetric copulas, i.e., copulas that satisfy $C(u_1, u_2) = C(u_2, u_1)$. In this case, we simply write (3.1) as “$C$ is increasing in $\prec_S$.” There are many well-known symmetric single-parameter copulas that satisfy Assumption 4, i.e., that are increasing in $\prec_S$. By Lemma 3.2, these copulas are also increasing in $\prec_C$. The following lemma provides likely the most interesting case, which involves the structure of a bivariate standard normal distribution function, $\Phi(\cdot, \cdot, \rho)$ with $\rho \in [-1, 1]$. Let $\Phi(\cdot)$ be a marginal standard normal distribution function.

**Proposition 3.1 (Joe (1997), p. 140)** Assumption 4 holds for the normal copula, i.e., for

$$C(u_1, u_2; \rho) = \Phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho).$$

In the special case of a normal copula and, additionally, normal marginal distributions, $F_\varepsilon(\cdot) = F_\nu(\cdot) = \Phi(\cdot)$, the model of equation (2.1) becomes a bivariate probit model. We list four other single-parameter copulas that satisfy Assumption 4.

**Example 3.1** The Ali-Mikhail-Haq family: For $\rho \in [-1, 1)$,

$$C(u_1, u_2; \rho) = \frac{u_1 u_2}{1 - \rho(1 - u_1)(1 - u_2)}.$$

**Example 3.2** The Farlie-Gumbel-Morgenstern family: For $\rho \in [-1, 1]$,

$$C(u_1, u_2; \rho) = u_1 u_2 + \rho u_1 u_2 (1 - u_1)(1 - u_2).$$

**Example 3.3** The Gumbel-Barnett family: For $\rho \in (0, 1]$,

$$C(u_1, u_2; \rho) = u_1 u_2 \exp(-\rho \ln u_1 \ln u_2).$$

**Example 3.4** The Frank family: For $\rho \in (-\infty, \infty) \setminus \{0\}$,

$$C(u_1, u_2; \rho) = -\frac{1}{\rho} \ln \left\{ \frac{(e^{-\rho u_1} - 1)(e^{-\rho u_2} - 1)}{e^{-\rho} - 1} \right\}.$$ 

\[^4\text{Nelsen (1999, p. 68, pp. 96 - 97). Also see Joe (1997, pp. 140-142) for other copulas that satisfy Assumption 4.}\]
One can readily verify that Assumption 4 holds for these copulas.\footnote{The copulas in Examples 3.3 and 3.4 have distinct ranges of the dependence parameter $\rho$. The Gumbel-Barnett copula only allows positive dependence. The Frank copula is suitable to model variables with strong positive or negative dependence. See, e.g., Trivedi and Zimmer (2007) for detailed features of the copulas listed in this paper.}

## 4 Identification in a Stylized Model

We first consider a simple stylized bivariate threshold crossing model of a triangular system with no common regressors (i.e., no $X$ covariates) and only one excluded covariate (i.e., $Z$ is scalar), so that

\[
Y = 1[\beta_0 + \delta_1 D - \varepsilon \geq 0],
\]
\[
D = 1[\alpha_0 + \gamma_1 Z - \nu \geq 0].
\]  

(4.1)

Let $\Psi$ be the parameter space of $\psi \equiv (\alpha_0, \gamma_1, \beta_0, \delta_1, \rho)\text{'}$. For this simple stylized model, we further assume that $Z$ is a binary variable, namely, $Z \in \text{supp}(Z) = \{0, 1\}$, where $\text{supp}(\cdot)$ denotes the support of its argument. In the following sections, we show how the results for this simple stylized model are readily generalized to the full model of equation (2.1) with possibly vector valued $X$ and $Z$ and without requiring that any element of $Z$ be binary.

### 4.1 Local Identification

Let $U_1 \equiv F_{\varepsilon}(\varepsilon)$ and $U_2 \equiv F_{\nu}(\nu)$. Using Assumption 1, one can derive expressions for all possible fitted probabilities implied from the model of equation (4.1). For instance, $\Pr[Y = 1, D = 1|Z = 0]$ can be expressed as

\[
\Pr[Y = 1, D = 1|Z = 0] = \Pr[\varepsilon \leq \beta_0 + \delta_1, \nu \leq \alpha_0; \rho] = \Pr[U_1 \leq F_{\varepsilon}(\beta_0 + \delta_1), U_2 \leq F_{\nu}(\alpha_0); \rho] = C(F_{\varepsilon}(\beta_0 + \delta_1), F_{\nu}(\alpha_0); \rho),
\]

where the first equality is using Assumption 1. For notational simplicity, we transform $\psi = (\alpha_0, \gamma_1, \beta_0, \delta_1, \rho)\text{'}$. The transformation reduces complications that appear in our proofs. Let

\[
(\alpha_0, \gamma_1, \beta_0, \delta_1)\text{'} \mapsto (a_0, a_1, b_0, b_1)\text{'}
\]  

(4.2)
denote a mapping such that,
\[ a_0 \equiv F_\nu(a_0), \]
\[ a_1 \equiv F_\nu(a_0 + \gamma_1), \]
\[ b_0 \equiv F_\varepsilon(\beta_0), \]
\[ b_1 \equiv F_\varepsilon(\beta_0 + \delta_1), \]
and note that the mapping is one-to-one since \( F_\nu \) and \( F_\varepsilon \) are strictly increasing. Let \( \theta \equiv (a_0, a_1, b_0, b_1, \rho)' \) denote the new parameter vector in a parameter space \( \Theta \subseteq (0, 1)^4 \times \Omega \). Also, write \( p_{y,d,z} \equiv \Pr[Y = y, D = d|Z = z] \) for \( (y, d, z) \in \{0, 1\}^3 \). Now, the fitted probabilities can be written as follows:

\[
\begin{align*}
p_{11,0} &= C(b_1, a_0; \rho), \\
p_{11,1} &= C(b_1, a_1; \rho), \\
p_{10,0} &= b_0 - C(b_0, a_0; \rho), \\
p_{10,1} &= b_0 - C(b_0, a_1; \rho), \\
p_{01,0} &= a_0 - C(b_1, a_0; \rho), \\
p_{01,1} &= a_1 - C(b_1, a_1; \rho).
\end{align*}
\]

Equation (4.3) contains the maximal set of probabilities that are not superfluous, since these probabilities imply the values of \( p_{y,d,z} = \Pr[Y = y, D = d|Z = z] \) for \( (y, d, z) \in \{(0, 0, 1), (0, 0, 0)\} \). Let \( \hat{\pi} \equiv (p_{11,0}, p_{10,0}, p_{11,1}, p_{10,1}, p_{01,0}, p_{01,1})' \) be a reduced form parameter in a parameter space \( \hat{\Pi} \subseteq (0, 1)^6 \), which is trivially identified as \( p_{y,d,z}'s \) are the distributions of the data. Our (local) identification problem is a question of whether we can uniquely recover the true structural parameter \( \theta^0 \equiv (a_0^0, a_1^0, b_0^0, b_1^0, \rho^0)' \) given true reduced form parameter \( \pi^0 \). In this section, we assume a substantially weaker version of Assumption 4 for local identification by introducing a local version of the condition (3.2). Let \( N^0 \) be the union of local neighborhoods of \((b_0^0, a_0^0), (b_0^1, a_0^0), (b_1^0, a_0^0), \) and \((b_1^0, a_1^0)\), respectively, and \( \Omega^0 \) be a local neighborhood of \( \rho^0 \). Also, let \( \lambda(u_1, u_2; \rho) \equiv \frac{C_\nu[u_1, u_2; \rho]}{C_1[u_1, u_2; \rho]} \) for notational simplicity.

**Assumption 4’** The copula \( C(u_1, u_2; \rho) \) of Assumption 3 is twice differentiable in the neighborhood \( N^0 \times \Omega^0 \), and satisfies that \( a_0^0 \neq a_1^0 \) if and only if \( \lambda(b_1^0, a_0^0, \rho^0) \neq \lambda(b_1^0, a_1^0, \rho^0) \) or \( \lambda(b_1^0, a_0^0; \rho^0) \neq \lambda(b_1^0, a_1^0; \rho^0) \).
Define $G : \Theta \subseteq (0, 1)^4 \times \Omega \rightarrow \tilde{\Pi} \subseteq (0, 1)^6$ as

$$G(\theta) \equiv \begin{bmatrix}
C(b_1, a_0; \rho) \\
C(b_1, a_1; \rho) \\
b_0 - C(b_0, a_0; \rho) \\
b_0 - C(b_0, a_1; \rho) \\
a_0 - C(b_1, a_0; \rho) \\
a_1 - C(b_1, a_1; \rho)
\end{bmatrix},$$

and write

$$\tilde{\pi}^0 = G(\theta^0). \tag{4.4}$$

Then $\theta^0$ is (locally) identifiable if and only if, from equation (4.4), $\tilde{\pi}^0$ uniquely determines $\theta^0$ in the neighborhood of $\theta^0$. Let

$$J(\theta) \equiv \frac{\partial G(\theta)}{\partial \theta'}, \tag{4.5}$$

be the Jacobian matrix of $G(\theta)$. Then by Theorem 6 of (Rothenberg, 1971, p. 585), the following proposition holds:

**Proposition 4.1** Assume that there exists an open neighborhood of $\theta^0$ in which $J(\theta)$ has constant rank. Then $\theta^0$ is locally identifiable if and only if $J(\theta^0)$ has rank equal to $\dim(\theta)$. The result of this proposition is essentially due to the implicit function theorem. Now, we find the condition under which the matrix $J(\theta^0)$ has rank equal to $\dim(\theta)$, or equivalently, has full column rank. By conducting elementary row and column operations on the Jacobian matrix $J(\theta)$ for a given value of $\theta$ (see the Appendix) which preserves the rank, it is easy to see that the matrix is full column rank if and only if either

$$\frac{C_\rho(b_1, a_0; \rho)}{C_1(b_1, a_0; \rho)} - \frac{C_\rho(b_1, a_1; \rho)}{C_1(b_1, a_1; \rho)} \neq 0,$$

or

$$\frac{C_\rho(b_0, a_1; \rho)}{1 - C_1(b_0, a_1; \rho)} - \frac{C_\rho(b_0, a_0; \rho)}{1 - C_1(b_0, a_0; \rho)} \neq 0.$$

See the Appendix for the actual expression of $J(\theta)$. 

---

10
But by Assumption 4', this condition is true for $\theta^0$ if and only if $a_0^0 \neq a_1^0$ (that is, $\gamma_1^0$, the coefficient on $Z$, is nonzero). The following theorem summarizes this identification result after rephrasing it in terms of the original parameters:

**Theorem 4.1** In model (4.1), let Assumptions 1 - 3 and 4' hold. Then $(\alpha_0^0, \gamma_1^0, \beta_0^0, \delta_1^0, \rho^0) \in \Psi$ is locally identified if and only if $\gamma_1^0 \neq 0$.

The identification condition is the exclusion restriction that the coefficient on the instrument $Z$ is nonzero. This condition implies that the excluded instrument plays a key role in identifying the parameters of the stylized model.

### 4.2 Global Identification

We now extend the local identification result to global identification, replacing Assumption 4' with Assumption 4. For global identification, we apply the global inverse function theorem by Hadamard (1906a,b), which is stated as follows:

**Proposition 4.2 (Hadamard)** Let $A$ and $B$ be subsets of $\mathbb{R}^k$ and let $f : A \to B$ be a continuously differentiable function. If (i) $f$ is proper, (ii) the Jacobian of $f$ vanishes nowhere, and (iii) $B$ is simply connected, then $f$ is a homeomorphism, and hence, one-to-one and onto.

A mapping $f : A \to B$ is proper if whenever $K \subset B$ is compact then $f^{-1}(K) \subset A$ is compact. A topological space is simply connected if it is path-connected and any simple closed curve can be shrunk to a point. Note that, for example, any convex subset of $\mathbb{R}^k$ and its half spaces are simply connected.

Let $G^* : \Theta \subseteq (0,1)^4 \times \Omega \to \Pi \subseteq (0,1)^5$ be defined as

$$G^*(\theta) \equiv \begin{bmatrix} C(b_1, a_0; \rho) \\ C(b_1, a_1; \rho) \\ b_0 - C(b_0, a_0; \rho) \\ a_0 - C(b_1, a_0; \rho) \\ a_1 - C(b_1, a_1; \rho) \end{bmatrix},$$

Rothenberg (1971) also derives a global identification results, but the condition he imposes on the Jacobian matrix is difficult to verify in our setting.

Refer to Rudin (1986, p. 222) for a technical definition of the simple connectedness.

A half-space is either of the two parts into which a hyperplane divides a space.
where one of the elements in the vector $G(\theta)$ is dropped because one of the rows in $J(\theta)$ in equation (4.5) can be shown to be redundant.\(^{10}\) The reduced form parameter $\pi$ and its parameter space $\Pi$ are defined accordingly. Then, the Jacobian matrix $J^* \equiv J^*(\theta) \equiv \frac{\partial G^*(\theta)}{\partial \theta}$ becomes a square matrix. Next, since $\lambda(b, \cdot; \rho)$ is continuous, Assumption 4 holds if and only if the following statement holds: $a_0 \neq a_1$ if and only if $\lambda(b, a_0; \rho) \neq \lambda(b, a_1; \rho)$ for any $b$ and $\rho$. Therefore, the determinant of $J^*$ is nonzero (that is, $J^*$ has full rank) if and only if $a_0 \neq a_1$ (or equivalently, $\gamma_1 \neq 0$) by a similar argument as in the proof of Theorem 4.1.

Define $\Theta_c \subseteq (0, 1)^4 \times \Omega$ to be a 5-dimensional finite open set such that its half spaces, $\Theta_{c1} \equiv \{ \theta \in \Theta_c : a_0 < a_1 \}$ and $\Theta_{c2} \equiv \{ \theta \in \Theta_c : a_0 > a_1 \}$ are simply connected. Define $\Pi_{c1} \equiv G^*(\Theta_{c1})$ and $\Pi_{c2} \equiv G^*(\Theta_{c2})$. Also, define $G^*|_{\Theta_{c1}} : \Theta_{c1} \rightarrow \Pi_{c1}$ and $G^*|_{\Theta_{c2}} : \Theta_{c2} \rightarrow \Pi_{c2}$ to be the function $G^*(\cdot)$ on its restricted domains.

$G^*|_{\Theta_{c1}}(\cdot)$ and $G^*|_{\Theta_{c2}}(\cdot)$ are continuous and therefore the pre-image of a closed set under $G^*|_{\Theta_{c1}}(\cdot)$ and $G^*|_{\Theta_{c2}}(\cdot)$ is closed. Also, since $\Theta_{c1}$ and $\Theta_{c2}$ are bounded, the pre-image of a bounded set is bounded. Therefore, $G^*|_{\Theta_{c1}}(\cdot)$ and $G^*|_{\Theta_{c2}}(\cdot)$ are proper. Also, by the fact that (i) $\Theta_{c1}$ and $\Theta_{c2}$ are simply connected, (ii) $G^*|_{\Theta_{c1}}(\theta)$ and $G^*|_{\Theta_{c2}}(\theta)$ are continuous on $\Theta_{c1}$ and $\Theta_{c2}$, respectively, and (iii) the sign of the determinant of $J^*$ is preserved on $\Theta_{c1}$ and $\Theta_{c2}$, it follows that $\Pi_{c1}$ and $\Pi_{c2}$ are also simply connected.\(^{11}\)

Lastly, $J^*$ has full rank over $\Theta_{c1}$ and $\Theta_{c2}$ by Assumption 4. Therefore, $G^*|_{\Theta_{c1}} : \Theta_{c1} \rightarrow \Pi_{c1}$ and $G^*|_{\Theta_{c2}} : \Theta_{c2} \rightarrow \Pi_{c2}$ are one-to-one and onto by Proposition 4.2. This result implies that there exists $G^*|_{\Theta_{c1}}^{-1}(\cdot)$ such that, for any given $\pi \in \Pi_{c1},$

$$G^*|_{\Theta_{c1}}^{-1}(\pi) = \theta \in \Theta_{c1}.$$

This proves that the parameter $\theta$ is globally identified in $\Theta_{c1}$. Similarly, $\theta$ is globally identified in $\Theta_{c2}$.

Let $\Theta_1 \equiv \{ \theta \in \Theta : a_0 < a_1 \}$ and $\Theta_2 \equiv \{ \theta \in \Theta : a_0 > a_1 \}$. Then this conclusion is true for any given subset of $\Theta_1$ or $\Theta_2$ that is a finite simply connected set. Assume that the original parameter space $\Psi$ is open and $\Psi_1 \equiv \{ \psi \in \Psi : \gamma_1 > 0 \}$ and $\Psi_2 \equiv \{ \psi \in \Psi : \gamma_1 > 0 \}$ are simply connected. By the continuous monotone map defined in (4.2), the transformed parameter space $\Theta$ is open and $\Theta_1$ and $\Theta_2$ are open, simply connected. Then $\Theta_1$ and $\Theta_2$ can be represented by a countable union of finite open simply connected sets. For example, we have $\Theta_1 = \bigcup_{i=1}^{\infty} \Theta_{1i}$, where $\{\Theta_{1i}\}_{i=1}^{\infty}$ is a sequence of finite open simply connected sets.

\(^{10}\)The result of the proof still holds no matter which element is dropped from $G(\theta)$.

\(^{11}\)See, e.g., results in Arnold (2009, p. 33).
in $\Theta_1$ such that $\Theta_{11} \subset \Theta_{12} \subset \cdots \subset \Theta_1$. Also, let $G^*(\Theta_1) \equiv \Pi_i$ for $i \in \{1, 2, 3, \ldots\}$, so that $\Pi_1 = G^*(\Theta_1) = G^*(\cup_{i=1}^{\infty} \Theta_{1i}) = \cup_{i=1}^{\infty} G^*(\Theta_{1i}) = \cup_{i=1}^{\infty} \Pi_i$ and $\Pi_{11} \subset \Pi_{12} \subset \cdots \subset \Pi_1$. Then, for any given $\pi \in \Pi_1$, we have that $\pi \in \Pi_i$ for all $i \geq q$ (for some $q$), then $G^*(-1)_{\Theta_{1i}}(\pi) \in \Theta_{1i}$ for all $i \geq q$ from the previous result, and hence $G^*_{-1}(\pi) = G^*_{-1}|_{\cup_{i=q}^{\infty} \Theta_{1i}}(\pi) \in \cup_{i=q}^{\infty} \Theta_{1i} = \Theta_1$. Therefore, $\theta$ is globally identified in $\Theta_1$. Then, we can conclude that $\psi$ is globally identified in $\Psi_1$. By similar arguments, $\psi$ is globally identified in $\Psi_2$. Thus, since $\gamma_1$ is identified by $\gamma_1 = F^*_\nu^{-1}(\Pr[D = 1|X = x, Z = z]) - F^*_\nu^{-1}(\Pr[D = 1|Z = 0])$, $\psi$ is globally identified in $\Psi$ if $\gamma_1 \neq 0$.

The following theorem summarizes the results:

**Theorem 4.2** In model (4.1), let Assumptions 1 - 4 hold. Then $(\alpha_0, \gamma_1, \beta_0, \delta_1, \rho_0) \in \Psi$ is globally identified if (i) $\gamma_1 \neq 0$; (ii) $\Psi$ is open and $\Psi_1$ and $\Psi_2$ are simply connected.

To satisfy (ii), one can simply have $\Psi = \mathbb{R}^4 \times \text{int}(\Omega)$ where $\text{int}(\Omega)$ is the interior of $\Omega$. In fact, any open convex $\Psi$ is sufficient.

## 5 Identification in Full Model

In this section, we conduct identification analysis of the full model of equation (2.1). Thus, we generalize the previous section to allow for the possibility of exogenous regressors $X$ that enter both the equation for $Y$ and the equation for $D$, we allow for the possibility of instruments $Z$ being vector valued without requiring any element of $Z$ to be binary, and we also allow for the possibility that there are no instruments in the model. We present results for global identification of $\tilde{\psi} = (\alpha', \beta', \delta_1, \gamma', \rho)^\prime$ in $\tilde{\Psi}$ here, and local identification results can be obtained similarly which are omitted.

Consider the identification of $(\alpha, \gamma)$ from the $D$ equation. Since $\nu \perp (X, Z)$, it follows that $\Pr[D = 1|X = x, Z = z] = F^*_\nu(x'\alpha + z'\gamma)$, or

$$x'\alpha + z'\gamma = F^*_\nu^{-1}(\Pr[D = 1|X = x, Z = z]). \quad (5.1)$$

Therefore, as long as $(X', Z')$ does not lie in a proper linear subspace of $\mathbb{R}^{k+l}$ a.s., we identify $(\alpha, \gamma)$ from equation (5.1).\textsuperscript{12}

\textsuperscript{12}A proper linear subspace of $\mathbb{R}^{k+l}$ is a linear subspace with a dimension strictly less than $k + l$. The assumption is that, if $M$ is a proper linear subspace of $\mathbb{R}^{k+l}$, then $\Pr[(X', Z') \in M] < 1$. 

13
First, suppose that $\gamma$ is a nonzero vector, i.e., there exists at least a variable in $Z$ with a non-zero coefficient. Then, there exist two values $z$ and $\tilde{z}$ in $\text{supp}(Z)$ such that $z'\gamma \neq \tilde{z}'\gamma$. Suppose not and that $z'\gamma$ is constant for all $z$ in $\text{supp}(Z)$, then it contradicts the assumption that $Z$ does not lie in a proper linear subspace of $\mathbb{R}^l$. Assume that $\text{supp}(X|Z = z) \cap \text{supp}(X|Z = \tilde{z})$ is a nonempty set. Take $(x, z)$ and $(x, \tilde{z})$ for some $x \in \text{supp}(X|Z = z) \cap \text{supp}(X|Z = \tilde{z})$, and write

$$s_0 \equiv F_\nu(x'\alpha + z'\gamma),$$
$$s_1 \equiv F_\nu(x'\alpha + \tilde{z}'\gamma),$$
$$r_0 \equiv F_\varepsilon(x'\beta),$$
$$r_1 \equiv F_\varepsilon(x'\beta + \delta_1).$$

Let $p_{y,d,xz} \equiv \Pr[Y = y, D = d|X = x, Z = z]$ for $(y, d) \in \{0, 1\}^2$. Since $(\epsilon, \nu) \perp (X, Z)$, the fitted probabilities are written as

$$p_{11,xz} = C(r_1, s_0; \rho),$$
$$p_{11,xz} = C(r_1, s_1; \rho),$$
$$p_{10,xz} = r_0 - C(r_0, s_0; \rho),$$
$$p_{10,xz} = r_0 - C(r_0, s_1; \rho),$$
$$p_{01,xz} = s_0 - C(r_1, s_0; \rho),$$
$$p_{01,xz} = s_1 - C(r_1, s_1; \rho).$$

The set of equations has the same form as (4.3) in the previous section. Let $w \equiv (x, z, \tilde{z})$. By pursuing a similar argument as in the previous section, the problem eventually becomes inverting the following:

$$\pi_w = G^*(\theta_w),$$

where $\pi_w$ is a $5 \times 1$ subvector of $\pi_{w'} \equiv (p_{11,xz}, p_{11,xz}, p_{10,xz}, p_{10,xz}, p_{01,xz}, p_{01,xz})'$ in its parameter space $\Pi_w$ and $\theta_w \equiv (s_0, s_1, r_0, r_1, \rho)'$ in $\Theta_w$. Now we proceed similar to the proof of Theorem 4.2. Under Assumption 4, $J^*(\theta_w)$ has full rank for any $w = (x, z, \tilde{z})$ since $z'\gamma \neq \tilde{z}'\gamma$ implies $s_0 \neq s_1$. Assume that the original parameter space $\tilde{\Psi}$ is open and $\tilde{\Psi}_{1,(z-\tilde{z})} \equiv \{\tilde{\psi} \in \tilde{\Psi} : (z - \tilde{z})'\gamma < 0\}$ and $\tilde{\Psi}_{2,(z-\tilde{z})} \equiv \{\tilde{\psi} \in \tilde{\Psi} : (z - \tilde{z})'\gamma > 0\}$ are simply connected. Then $\Theta_{1,w} \equiv \{\theta_w \in \Theta_w : s_0 < s_1\}$ and $\Theta_{2,w} \equiv \{\theta_w \in \Theta_w : s_0 > s_1\}$ are
also open and simply connected, and can be approximated by sequences of open, simply
connected sets. Eventually by Proposition 4.2, it follows that, for any given \( \pi_w \in G^*(\Theta_{1,w}) \),

\[
G^{*-1}(\pi_w) = \theta_w \in \Theta_{1,w},
\]

and \( \theta_w \) is globally identified in \( \Theta_{1,w} \). Similarly, \( \theta_w \) is globally identified in \( \Theta_{2,w} \). Since \( s_0 \)
and \( s_1 \) are known, we can conclude that \( \theta_w \) is globally identified in \( \Theta_w \). Identification of \( \delta_1 \) follows from

\[
\delta_1 = F^{-1}_\epsilon(r_1) - F^{-1}_\epsilon(r_0).
\]

Let

\[
\mathcal{X} = \bigcup_{z \neq \tilde{z}} \text{supp}(X|Z = z) \cap \text{supp}(X|Z = \tilde{z}).
\]

Using the fact that we can recover \( r_0 \) for any \( x \in \mathcal{X} \), identification of \( \beta \) follows from

\[
x'\beta = F^{-1}_\epsilon(r_0),
\]

assuming that \( \mathcal{X} \) does not lie in a proper linear subspace of \( \mathbb{R}^k \) a.s.

Next, suppose that \( \gamma = 0 \), i.e., there does not exist any excluded variable with a non-zero coefficient. We prove that even in this case we identify the parameters \( \tilde{\psi} = (\alpha', \beta', \delta_1, \gamma', \rho) = (\alpha', \beta', \delta_1, 0', \rho) \). Assume that \( \alpha \) is a nonzero vector. Then, since we assume that \( X \) does not lie in a proper linear subspace of \( \mathbb{R}^k \) a.s., there exist two values \( x \) and \( \tilde{x} \) in \( \text{supp}(X) \) such that \( x'\alpha \neq \tilde{x}'\alpha \). If \( x'\beta = \tilde{x}'\beta \) then the previous proof can be applied and identification is achieved. Suppose \( x'\beta \neq \tilde{x}'\beta \). Now with

\[
q_0 \equiv F_\nu(x'\alpha),
q_1 \equiv F_\nu(\tilde{x}'\alpha),
t_0 \equiv F_\epsilon(x'\beta),
t_1 \equiv F_\epsilon(\tilde{x}'\beta),
\]

15
we can write

\[
\begin{align*}
p_{11,x} &= C(F_x(F_x^{-1}(t_0) + \delta_1), q_0; \rho), \\
p_{11,x} &= C(F_x(F_x^{-1}(t_1) + \delta_1), q_1; \rho), \\
p_{10,x} &= t_0 - C(t_0, q_0; \rho), \\
p_{10,x} &= t_1 - C(t_1, q_1; \rho), \\
p_{00,x} &= 1 - t_0 - q_0 + C(t_0, q_0; \rho), \\
p_{00,x} &= 1 - t_1 - q_1 + C(t_1, q_1; \rho).
\end{align*}
\]

Note that this particular set of fitted probabilities reduces complications in this case. Let \( \tilde{w} \equiv (x, \tilde{x}) \), and define \( \theta_{\tilde{w}} \equiv (q_0, q_1, t_0, t_1, \delta_1, \rho) \) in its parameter space \( \Theta_{\tilde{w}} \) and \( \tilde{\pi}_{\tilde{w}} \equiv (\tilde{p}_{11,x}, \tilde{p}_{11,x}, \tilde{p}_{10,x}, \tilde{p}_{10,x}, \tilde{p}_{00,x}, \tilde{p}_{00,x})' \) in \( \tilde{\Pi}_{\tilde{w}} \). Then, by letting \( H(\cdot) \) be a vector-valued function defined using the right hand side expressions of the fitted probabilities similar as before, we have \( \tilde{\pi}_{\tilde{w}} = H(\theta_{\tilde{w}}) \).

Again, we proceed similar to the proof of Theorem 4.2. In the Appendix, we show that the Jacobian \( \tilde{J}(\theta_{\tilde{w}}) \) of \( H(\theta_{\tilde{w}}) \) has full rank for any \( \tilde{w} = (x, \tilde{x}) \) if

\[
C_{\rho}(u_1, u_2; \rho) \neq 0 \text{ for any } u_1, u_2, \text{ and } \rho. \tag{5.2}
\]

The condition (5.2) is closely related to the concordance ordering in Definition 3.3. By Lemma 3.2, Assumption 4 implies (5.2).

Therefore under Assumption 4, similarly as before, we can prove the global identification of \( \theta_{\tilde{w}} \) in subsets of \( \Theta_{\tilde{w}} \) that are implied by \( \tilde{\Psi}_{1,(x-\tilde{x})} \equiv \{ \tilde{\psi} \in \tilde{\Psi} : (x - \tilde{x})' \alpha < 0 \} \) and \( \tilde{\Psi}_{2,(x-\tilde{x})} \equiv \{ \tilde{\psi} \in \tilde{\Psi} : (x - \tilde{x})' \alpha > 0 \} \), respectively. Lastly, using the fact that we can recover \( t_0 \) for any \( x \in \text{supp}(X) \), identification of \( \beta \) follows from \( x' \beta = F_x^{-1}(t_0) \). The following theorem summarizes all the identification results.

**Theorem 5.1** In model (2.1), let Assumptions 1 - 3 and 4 hold. Then \((\alpha', \beta', \delta_1, \gamma, \rho) \in \tilde{\Psi} \) are globally identified if (i) either \( \gamma \) or \( \alpha \) is a nonzero vector; (ii) \((X', Z')\) does not lie in a proper linear subspace of \( \mathbb{R}^{k+1} \) a.s.; (iii) \( X \) does not lie in a proper linear subspace of \( \mathbb{R}^{k} \) a.s.; (iv) \( \tilde{\Psi} \) is open and either \( \tilde{\Psi}_{1,(z-\tilde{z})} \) and \( \tilde{\Psi}_{2,(z-\tilde{z})} \) are simply connected for any \( z, \tilde{z} \) in \( \text{supp}(Z) \), or \( \tilde{\Psi}_{1,(x-\tilde{x})} \) and \( \tilde{\Psi}_{2,(x-\tilde{x})} \) are simply connected for any \( x, \tilde{x} \) in \( \text{supp}(X) \).
Theorem 5.1 nests identification results for a model without excluded instruments. Condition (i) requires that if there is no excluded instrument, then at least one variable in X in the D equation has to be relevant. A sufficient condition for Condition (iii) is that \( \text{supp}(X, Z) = \text{supp}(X) \times \text{supp}(Z) \) and Condition (ii) holds, since \( \mathcal{X} = \text{supp}(X) \) in this case. Also, \( \mathcal{X} = \text{supp}(X) \) holds when \( \gamma = 0 \). Note that Condition (iii) implies that there exists \( z \) and \( \tilde{z} \) in \( \text{supp}(Z) \) such that \( \text{supp}(X|Z = z) \cap \text{supp}(X|Z = \tilde{z}) \) is nonempty. Condition (ii) is the standard full rank condition found in most identification analyses. This rank condition can also be seen as a support condition. For instance, in a case where \( X_i \) is binary for some \( i \in \tilde{I} \), this condition implies that \( \Pr[X_i = x_i] > 0 \) for \( x_i \in \{0, 1\} \). Condition (iv) is satisfied with open convex \( \tilde{\Psi} \).

6 Conclusions

In a generalization of bivariate probit models we derive conditions for global identification of the coefficient parameters as well as of a distributional parameter. In a model with excluded instruments, the condition requires the role of the instruments that accounts for the endogeneity in the model. When there are common exogenous regressors in both equations, identification is still achieved without such excluded instruments. It is worth noting that a bivariate normality assumption of the latent variables is not critical for the identification results we obtain. We substantially relax such an assumption in the literature by introducing a broad class of copulas for their joint distribution and allowing their marginal distributions to be arbitrary.

References


7 Appendix

7.1 SI Ordering and Related Concept

Before we prove Lemma 3.1, we relate the SI/SD ordering concept in Definition 3.2 with a more general concept of dependence ordering that has greater relevance to our identification proof. To the best of our knowledge, no result has been found in the literature that defines the concept and shows connection to the SI/SD ordering. Note that this dependence ordering is not weaker nor stronger than the concordance ordering.

Lemma 7.1 Let $C : (0,1)^2 \rightarrow (0,1)$ and $\tilde{C} : (0,1)^2 \rightarrow (0,1)$ be two distinct copulas. Suppose $\tilde{C}(u_1|u_2)$ is strictly more SI (less SD) than $C(u_1|u_2)$, i.e., $u_1^\dagger(u_1, u_2) = \tilde{C}^{-1}(C(u_1|u_2)|u_2)$ is strictly increasing in $u_2$. Then $u_1^* (u_1, u_2) = \tilde{C}^{-1}(C(u_1, u_2), u_2)$ is strictly increasing in $u_2$.

Proof of Lemma 7.1: We prove that if $u_1^\dagger = u_1^\dagger(u_1, u_2)$ is strictly increasing in $u_2$ with $u_1^\dagger$ being the root of $\tilde{C}(u_1^\dagger|u_2) = C(u_1|u_2)$, then $u_1^* = u_1^*(u_1, u_2)$ is strictly increasing in $u_2$ with $u_1^*$ being the root of $\tilde{C}(u_1^*, u_2) = C(u_1, u_2)$.

Suppose that $u_1^\dagger(u_1, u_2)$ is strictly increasing in $u_2$. Then, for any $u_2' < u_2$, we have $u_1^\dagger(u_1, u_2') < u_1^\dagger(u_1, u_2)$ or, since $\tilde{C}(\cdot|u_2')$ is strictly increasing, $\tilde{C}(u_1^\dagger(u_1, u_2')|u_2') < \tilde{C}(u_1^\dagger(u_1, u_2)|u_2')$. It follows that

$$
\tilde{C}(u_1^\dagger(u_1, u_2), u_2) = \int_0^{u_2} \tilde{C}(u_1^\dagger(u_1, u_2)|u_2') du_2' \\
> \int_0^{u_2} \tilde{C}(u_1^\dagger(u_1, u_2')|u_2') du_2' \\
= \int_0^{u_2} C(u_1|u_2') du_2' \\
= C(u_1, u_2) \\
= \tilde{C}(u_1^*(u_1, u_2), u_2).
$$

Therefore, since $\tilde{C}(\cdot, u_2)$ is strictly increasing, it follows that

$$
u_1^\dagger(u_1, u_2) > u_1^*(u_1, u_2),$$

or

$$
\tilde{C}^{-1}(C(u_1|u_2)|u_2) > u_1^*(u_1, u_2), \tag{7.1}
$$
by the definition of $u_1^\dagger$. Since $\tilde{C}(\cdot|u_2)$ is strictly increasing, (7.1) implies

$$C(u_1|u_2) > \tilde{C}(u_1^\dagger(u_1, u_2)|u_2).$$

(7.2)

Next, differentiating $\tilde{C}(u_1^\dagger, u_2) = C(u_1, u_2)$ w.r.t. $u_2$ yields $\tilde{C}_1(u_1^\dagger, u_2) \cdot \frac{\partial u_1^\dagger}{\partial u_2} + \tilde{C}_2(u_1^\dagger, u_2) = C_2(u_1, u_2)$, or $\tilde{C}_1(u_1^\dagger, u_2) \cdot \frac{\partial u_1^\dagger}{\partial u_2} = C_2(u_1, u_2) - \tilde{C}_2(u_1^\dagger, u_2) = C(u_1|u_2) - \tilde{C}(u_1^\dagger|u_2)$. But since $C_1(u_1^*, u_2) = C(u_2|u_1^*) > 0$ for $u_2 \in (0, 1)$, (7.2) implies that $\frac{\partial u_1^*}{\partial u_2} > 0$. □

### 7.2 Proof of Lemma 3.1

Let $\rho_1 < \rho_2$ and suppose that Assumption 4 holds. By Lemma 7.1, $u_1^* = u_1^\dagger(u_1, u_2)$ is strictly increasing in $u_2$ with $u_1^*$ being the root of

$$C(u_1^*, u_2; \rho_2) = C(u_1, u_2; \rho_1).$$

(7.3)

Differentiating (7.3) w.r.t. $u_2$ yields

$$C_1(u_1^*, u_2; \rho_2) \cdot \frac{\partial u_1^*}{\partial u_2} + C_2(u_1^*, u_2; \rho_2) = C_2(u_1, u_2; \rho_1),$$

(7.4)

where $\frac{\partial u_1^*}{\partial u_2} > 0$. Since $C_1(u_1^*, u_2; \rho_2) > 0$ for $u_2 \in (0, 1)$, it satisfies that $C_1(u_1^*, u_2; \rho_2) \cdot \frac{\partial u_1^*}{\partial u_2} = C_2(u_1, u_2; \rho_1) - C_2(u_1^*, u_2; \rho_2) > 0$. Equivalently, it satisfies that $\frac{\partial}{\partial \rho} C_2(u_1^*(\rho), u_2; \rho) < 0$, or

$$C_{12}(u_1^*(\rho), u_2; \rho) \cdot \frac{\partial u_1^*}{\partial \rho} + C_{\rho 2}(u_1^*(\rho), u_2; \rho) < 0.$$  

(7.5)

Note that $\frac{\partial u_1^*}{\partial \rho} = -\frac{C_{\rho 1}}{C_1}$ by differentiating (7.3) w.r.t $\rho_2$ and by letting $\rho = \rho_2$. Therefore, (7.5) implies $C_{\rho 2}C_1 - C_{\rho 1}C_{12} < 0$ for any $(u_1, u_2, \rho)$, which is equivalent to the condition (3.2). □

### 7.3 Verification of Conditions (3.2) and (3.3) with Normal Copula

Although Proposition 3.1 shows that the normal copula satisfies Assumption 4, it would still be informative for readers to show that the normal copula satisfies the dependence orderings that are directly used in identification analyses. The following proposition provides an interesting result which is useful in verifying the conditions (3.2) and (3.3) with the normal copula.

---

13In general, a copula satisfies that $C_1(u, v) = C(v|u)$ and $C_2(u, v) = C(u|v)$.  

---

21
Proposition 7.1 (Plackett (1954)) Let \( f(w_1, w_2; \rho) \) be a normal density function with a correlation coefficient \( \rho \). Then the following holds:

\[
\frac{\partial}{\partial \rho} f(w_1, w_2; \rho) = \frac{\partial^2}{\partial w_1 \partial w_1} f(w_1, w_2; \rho).
\]

Verification of (3.2): Denote \( W_1 \equiv \Phi^{-1}(U_1) \) and \( W_2 \equiv \Phi^{-1}(U_2) \). By Proposition 7.1, it follows that

\[
\frac{C_1(u_1, u_2; \rho)}{C_1(u_1, u_2; \rho)} = \frac{\phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho)}{\phi(\Phi^{-1}(u_1)) \Phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho)}
\]

\[
= \frac{\phi(w_2|W_1 = w_1; \rho)}{\phi(w_1) \Phi(w_2|W_1 = w_1; \rho)}
\]

\[
= \phi(w_1) \lambda(-w_2|W_1 = w_1; \rho),
\]

where \( \lambda(w_2|W_1 = w_1; \rho) = \frac{\phi(w_2|W_1 = w_1; \rho)}{\Phi(-w_2|W_1 = w_1; \rho)} = \frac{\phi(-w_2|W_1 = w_1; \rho)}{\Phi(-w_2|W_1 = w_1; \rho)} \) is a conditional inverse Mill’s ratio (or hazard function of the conditional standard normal distribution) which is strictly increasing in \( w_2 \). Therefore, \( \phi(w_1) \lambda(-w_2|W_1 = w_1; \rho) \) is strictly decreasing in \( w_2 \) or in \( u_2 = \Phi(w_2) \). □

Verification of (3.3): The proof readily follows from \( \Phi_\rho(w_1, w_1; \rho) = \phi(w_1, w_2; \rho) > 0 \). □
7.4 Jacobian Matrix of $G(\theta)$

Let $C_1(\cdot, \cdot; \rho)$, $C_2(\cdot, \cdot; \rho)$, and $C_\rho(\cdot, \cdot; \rho)$ be the derivatives of $C(\cdot, \cdot; \rho)$ with respect to the first, second arguments and $\rho$, respectively. The Jacobian matrix $J(\theta) = \frac{\partial G(\theta)}{\partial \theta}$ has the following expression:

$$
\begin{bmatrix}
C_2(b_1, a_0; \rho) & 0 & 0 & C_1(b_1, a_0; \rho) & C_\rho(b_1, a_0; \rho) \\
0 & C_2(b_1, a_1; \rho) & 0 & C_1(b_1, a_1; \rho) & C_\rho(b_1, a_1; \rho) \\
-C_2(b_0, a_0; \rho) & 0 & 1 - C_1(b_0, a_0; \rho) & 0 & -C_\rho(b_0, a_0; \rho) \\
0 & -C_2(b_0, a_1; \rho) & 1 - C_1(b_0, a_1; \rho) & 0 & -C_\rho(b_0, a_1; \rho) \\
1 - C_2(b_1, a_0; \rho) & 0 & 0 & -C_1(b_1, a_0; \rho) & -C_\rho(b_1, a_0; \rho) \\
0 & 1 - C_2(b_1, a_1; \rho) & 0 & -C_1(b_1, a_1; \rho) & -C_\rho(b_1, a_1; \rho)
\end{bmatrix}
$$

Pre- and post-multiplying $J(\theta)$ by $E_1$ and $E_2$ defined on the following page, produces the following simplified matrix:

$$
E_1 \cdot J(\theta) \cdot E_2 =
\begin{bmatrix}
0 & 0 & 0 & 0 & \frac{C_\rho(b_1, a_0; \rho)}{C_1(b_1, a_0; \rho)} & \frac{C_\rho(b_1, a_1; \rho)}{C_1(b_1, a_1; \rho)} \\
0 & 0 & 0 & 0 & \frac{C_\rho(b_1, a_0; \rho)}{C_1(b_1, a_0; \rho)} & \frac{C_\rho(b_1, a_1; \rho)}{C_1(b_1, a_1; \rho)} \\
0 & 0 & 0 & 1 & \frac{C_\rho(b_0, a_1; \rho)}{1 - C_1(b_0, a_1; \rho)} & \frac{C_\rho(b_0, a_0; \rho)}{1 - C_1(b_0, a_0; \rho)} \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$
\[
E_1 \equiv \begin{bmatrix}
\frac{1}{C_1(b_1,a_{0};\rho)} - \frac{C_2(b_1,a_{0};\rho)}{C_1(b_1,a_{0};\rho)} - \frac{1}{C_1(b_1,a_{1};\rho)} + \frac{C_2(b_1,a_{1};\rho)}{C_1(b_1,a_{1};\rho)} & 0 & 0 & - \frac{C_2(b_1,a_{0};\rho)}{C_1(b_1,a_{0};\rho)} & C_2(b_1,a_{1};\rho) \\
0 & \frac{1}{C_1(b_1,a_{1};\rho)} - \frac{C_2(b_1,a_{1};\rho)}{C_1(b_1,a_{1};\rho)} & 0 & 0 & 0 & - \frac{C_2(b_1,a_{1};\rho)}{C_1(b_1,a_{1};\rho)} \\
- \frac{C_2(b_0,a_{0};\rho)}{1-C_1(b_0,a_{0};\rho)} & \frac{C_2(b_0,a_{1};\rho)}{1-C_1(b_0,a_{1};\rho)} & \frac{1}{1-C_1(b_0,a_{1};\rho)} & - \frac{1}{1-C_1(b_0,a_{1};\rho)} & - \frac{C_2(b_0,a_{0};\rho)}{1-C_1(b_0,a_{0};\rho)} & C_2(b_0,a_{1};\rho) \\
0 & - \frac{C_2(b_0,a_{1};\rho)}{1-C_1(b_0,a_{1};\rho)} & 0 & \frac{1}{1-C_1(b_0,a_{1};\rho)} & 0 & - \frac{C_2(b_0,a_{1};\rho)}{1-C_1(b_0,a_{1};\rho)} \\
1 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

\[
E_2 \equiv \begin{bmatrix}
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & \frac{C_2(b_0,a_{1};\rho)}{1-C_1(b_0,a_{1};\rho)} \\
1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & \frac{C_2(b_1,a_{1};\rho)}{C_1(b_1,a_{1};\rho)} \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]
7.5 Jacobian Matrix of $H(\cdot)$

Let $f_\epsilon(\cdot)$ be the PDF of $\epsilon$. Also, let $\tilde{F}_\epsilon(t) \equiv F_\epsilon(F_\epsilon^{-1}(t) + \delta_1)$ and $\tilde{f}_\epsilon(t) \equiv f_\epsilon(F_\epsilon^{-1}(t) + \delta_1)$ for brevity. The Jacobian matrix $\tilde{J}(\theta) = \frac{\partial H(\theta)}{\partial \theta}$ has the following expression:

$$
\begin{bmatrix}
C_2 \left( \tilde{F}_\epsilon(t_0), q_0; \rho \right) & 0 & C_1 \left( \tilde{F}_\epsilon(t_0), q_0; \rho \right) & \frac{\tilde{f}_\epsilon(t_0)}{\tilde{f}_\epsilon(t_0)} & 0 & C_1 \left( \tilde{F}_\epsilon(t_0), q_0; \rho \right) & \tilde{f}_\epsilon(t_0) & C_\rho \left( \tilde{F}_\epsilon(t_0), q_0; \rho \right) \\
0 & C_2 \left( \tilde{F}_\epsilon(t_1), q_1; \rho \right) & 0 & C_1 \left( \tilde{F}_\epsilon(t_1), q_1; \rho \right) & \frac{\tilde{f}_\epsilon(t_1)}{\tilde{f}_\epsilon(t_1)} & C_1 \left( \tilde{F}_\epsilon(t_1), q_1; \rho \right) & \tilde{f}_\epsilon(t_1) & C_\rho \left( \tilde{F}_\epsilon(t_1), q_1; \rho \right) \\
-2_2 (t_0, q_0; \rho) & 0 & 1 - C_1 (t_0, q_0; \rho) & 0 & 0 & 0 & -C_\rho (t_0, q_0; \rho) \\
0 & -C_2 (t_1, q_1; \rho) & 0 & 1 - C_1 (t_1, q_1; \rho) & 0 & 0 & -C_\rho (t_1, q_1; \rho) \\
-2_2 (t_0, q_0; \rho) & 0 & -1 + C_1 (t_0, q_0; \rho) & 0 & 0 & 0 & C_\rho (t_0, q_0; \rho) \\
0 & -1 + C_2 (t_1, q_1; \rho) & 0 & -1 + C_1 (t_1, q_1; \rho) & 0 & 0 & C_\rho (t_1, q_1; \rho)
\end{bmatrix}
$$

Then it can be readily shown that the determinant

$$
|\tilde{J}(\theta)| = (1 - C_1 (t_0, q_0; \rho)) C_1 (F_\epsilon(F_\epsilon^{-1}(t_0) + \delta_1), q_0; \rho) f_\epsilon(F_\epsilon^{-1}(t_0) + \delta_1)
$$

$$
\times \left[ (1 - C_1 (t_1, q_1; \rho)) C_\rho (F_\epsilon(F_\epsilon^{-1}(t_1) + \delta_1), q_1; \rho) + C_1 (F_\epsilon(F_\epsilon^{-1}(t_1) + \delta_1), q_1; \rho) \frac{f_\epsilon(F_\epsilon^{-1}(t_1) + \delta_1)}{f_\epsilon(t_1)} C_\rho (t_1, q_1; \rho) \right].
$$

Note that $0 < C_1(u_1, u_2; \rho) < 1$ for $u_2 \in (0, 1)$, and $f_\epsilon(\epsilon) > 0$ since $F_\epsilon(\epsilon)$ is strictly increasing. Therefore, $|\tilde{J}(\theta)| \neq 0$ for any $\theta$ if $C_\rho (u_1, u_2; \rho) \neq 0$ for any $(u_1, u_2, \rho)$.
7.6 Proof of Lemma 3.2

The proof of Lemma 3.2 is a slight modification of the proof of Theorem 2.14 of Joe (1997, p. 44). Suppose $C_{2|1} \prec_S \tilde{C}_{2|1}$. Let $(U_1, U_2) \sim C$, $(\tilde{U}_1, \tilde{U}_2) \sim \tilde{C}$, with $U_j \overset{d}{=} \tilde{U}_j$, $j = 1, 2$. By Theorem 2.9 of Joe (1997, p. 40), $(U_1, U_2) \overset{d}{=} (\tilde{U}_1, \psi(U_1, U_2))$ with $\psi(u_1, u_2) = \tilde{C}_{2|1}^{-1}(C_{2|1}(u_2|u_1)|u_1)$. Since $C_{2|1} \prec_S \tilde{C}_{2|1}$, $\psi$ is increasing in $u_1$ and $u_2$. To prove $C \prec_C \tilde{C}$, we consider two cases:

- **Case 1:** Suppose that $u_1$ and $u_2$ are such that $\psi(u_1, u_2) \leq u_2$. Then

$$\tilde{C}(u_1, u_2) = \Pr[\tilde{U}_1 \leq u_1, \tilde{U}_2 \leq u_2]$$
$$= \Pr[\tilde{U}_1 < u_1, \tilde{U}_2 < u_2]$$
$$= \Pr[U_1 < u_1, \psi(U_1, U_2) < u_2]$$
$$\geq \Pr[U_1 < u_1, \psi(u_1, U_2) < u_2]$$
$$> \Pr[U_1 < u_1, U_2 < u_2] = C(u_1, u_2)$$

where the strict inequality holds since $U_2 < u_2$ implies $\psi(u_1, U_2) \leq \psi(u_1, u_2) \leq u_2$ (but not vice versa since $\psi(u_1, U_2) \leq u_2$ and $\psi(u_1, u_2) \leq u_2$ does not necessarily imply $U_2 < u_2$ and $\Pr[\psi(u_1, u_2) < \psi(u_1, U_2)] = \Pr[u_2 < U_2] \neq 0$), and the second last inequality holds since, given $U_1 < u_1$, $\psi(U_1, U_2) \leq \psi(u_1, U_2) < u_2$.

- **Case 2:** Suppose that $u_1$ and $u_2$ are such that $\psi(u_1, u_2) > u_2$. Then

$$u_2 - C(u_1, u_2) = \Pr[U_1 > u_1, U_2 < u_2]$$
$$> \Pr[U_1 > u_1, \psi(u_1, U_2) \leq u_2]$$
$$\geq \Pr[U_1 > u_1, \psi(U_1, U_2) \leq u_2]$$
$$= \Pr[\tilde{U}_1 > u_1, \tilde{U}_2 < u_2] = u_2 - \tilde{C}(u_1, u_2)$$

where the strict inequality holds since $U_2 > u_2$ implies $\psi(u_1, U_2) \geq \psi(u_1, u_2) > u_2$ or $\psi(u_1, U_2) \leq u_2$ implies $U_2 \leq u_2$ (but not vice versa).

Therefore $C(u_1, u_2) < \tilde{C}(u_1, u_2)$ for any $u_1$ and $u_2$. In conclusion, if $C(u_1, u_2; \rho)$ is increasing in $\prec_S$ then it is increasing in $\prec_C$. □