



Identification in a generalization of bivariate probit models with dummy endogenous regressors[☆]



Sukjin Han^{a,*}, Edward J. Vytlačil^b

^a Department of Economics, University of Texas at Austin, United States

^b Department of Economics, Yale University, United States

ARTICLE INFO

Article history:

Received 4 September 2013

Received in revised form 25 December 2016

Accepted 1 April 2017

Available online 27 April 2017

Keywords:

Identification

Triangular threshold crossing model

Bivariate probit model

Dummy endogenous regressors

Binary response

Copula

Exclusion restriction

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, 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 global identification in a model without common exogenous covariates, where the excluded variable is allowed to be binary. Having an exclusion restriction is sufficient in models with common exogenous covariates that are present in both equations. The paper then extends the identification analyses to a model where the marginal distributions of the error terms are unknown.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

This paper examines the identification of a 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), Altonji et al. (2005), Bhattacharya et al. (2006), and Rhine et al. (2006), to name a few. 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

There are three papers in the literature that have studied identification of bivariate probit models: Freedman and Sekhon (2010), Wilde (2000), and Meango and Mourifié (2014). 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 fact, Meango and Mourifié (2014) show that, using as many equations as the number of parameters, there can be multiple solutions in a bivariate probit model where there are common binary exogenous regressors but no excluded instruments.²

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,

[☆] We thank the coeditor, the associate editor and three anonymous referees for useful comments and suggestions.

* Corresponding author.

E-mail addresses: sukjin.han@Austin.utexas.edu (S. Han), edward.vytlacil@yale.edu (E.J. Vytlačil).

¹ Heckman (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.

² Building upon Meango and Mourifié (2014) and the present paper, Han and Lee (2017) show that the solution is not unique even when exploiting the full set of equations implied by the model. These results demonstrate that Wilde's (2000) counting exercise is not sufficient for identification analysis.

while restricting their dependence structure by imposing that their copula function belongs to a broad class of parametric copulas that includes the normal copula as a special case. We then extend the results to a model where the marginal distributions are unknown. All results derived in this paper also apply to the special parametric 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. While [Wilde \(2000\)](#) restricts his analysis to bivariate probit models, we show that a bivariate normal distribution is not necessary for our identification strategy to work as long as a certain dependence structure is maintained. We extend the result to allow for the possibility of exogenous covariates that enter both equations and the possibility of instruments Z being vector valued without requiring any element of Z to be binary. Having an exclusion restriction is sufficient for identification in this context.³ In this full model, we also provide identification results without assuming that the marginal distributions of the error terms are known. The structural parameters are shown to be identified under similar conditions as in the known-marginal case and the marginal distributions are shown to be additionally identified under a stronger support condition.

We make use of copulas to characterize the joint distribution of the latent error terms, which 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. Specifically, the condition is that the copula is ordered 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.⁴

We also introduce a novel dependence ordering concept that characterizes minimal structure on the copula that is required for our identification results. This ordering is more general than the FOSD ordering but slightly less interpretable.

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. To facilitate their inference procedure in a censored linear quantile regression model, [Fan and Liu \(2015\)](#) introduce one-parameter ordered families of Archimedean copulas in characterizing dependence between the dependent variable and censoring variable, but the ordering concept which defines their class of copulas differs from ours. Copulas have also been used to model the joint distributions of error terms in switching regime models ([Fan and Wu, 2010](#)) or the joint distribution of

potential outcomes in randomized experiment settings ([Fan and Park, 2010](#)), where bounds on the distribution of treatment effects are derived. There are also recent papers that generalize a bivariate probit model using a copula structure ([Winkelmann, 2012](#)) or using nonparametric index functions instead of linear functions ([Marra and Radice, 2011](#)), or both ([Radice et al., 2015](#)), but all of these papers rely on the counting exercise for identification analysis.

The paper is organized as follows. In Section 2, we introduce the model and preliminary assumptions. Section 3 introduces dependence orderings 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, which is useful for subsequent analyses. Section 5 extends the identification analysis to the full model. Section 6 extends the results of the previous section to the case of nonparametric marginal distributions. Section 7 concludes with discussions on estimation and inference.

2. The model

Let Y denote the binary outcome variable and D the observed binary endogenous treatment variable. Let $X \equiv (1, X_1, \dots, X_k)'$ denote the vector of regressors that determine both Y and D , and let $Z \equiv (Z_1, \dots, Z_l)'$ denote a vector of regressors that directly affects 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) :

$$\begin{aligned} Y &= \mathbf{1}[X'\beta + \delta_1 D - \varepsilon \geq 0], \\ D &= \mathbf{1}[X'\alpha + Z'\gamma - \nu \geq 0], \end{aligned} \quad (2.1)$$

where $\alpha \equiv (\alpha_0, \alpha_1, \dots, \alpha_k)'$, $\beta \equiv (\beta_0, \beta_1, \dots, \beta_k)'$, and $\gamma \equiv (\gamma_1, \gamma_2, \dots, \gamma_l)'$. As an example of this model, Y might be an employment status or voting decision, D an indicator for having a bachelor degree, and Z college tuition. As another example, Y could be an indicator for patient death, D a medical treatment, and Z some randomization scheme. In these examples, X represents other individual characteristics.

We will maintain the following assumptions.

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

Assumption 2. F_ε and F_ν are known marginal distributions of ε and ν , 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 $C : (0, 1)^2 \rightarrow (0, 1)$ is twice differentiable in its arguments and ρ .

Assumption 4. (X', Z') does not lie in a proper linear subspace of \mathbb{R}^{k+l} a.s.⁵

Assumption 1 imposes that X and Z are exogenous. This assumption, which is commonly imposed in the literature on binary choice models, excludes heteroskedasticity of the error terms. **Assumption 2** characterizes the restrictions imposed on the marginal distributions of ε and ν . The moment restrictions are merely normalizations as long as the second moments of ε and ν are finite. Under these normalizations, the intercept parameter is present in the model and the correlation coefficient is the only

³ As mentioned, the results of [Meango and Mourifié \(2014\)](#) and [Han and Lee \(2017\)](#) show that an exclusion restriction is also necessary for identification when the common exogenous covariates are binary.

⁴ This contrasts with the identification result in a model related to ours, i.e., the sample selection model by [Heckman \(1979\)](#), where identification can be achieved solely by the functional form of the joint normal errors as long as there are common exogenous covariates. Excluded instruments only become necessary for identification in that model once the normality assumption is relaxed, which is not the case in our model.

⁵ 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 $\Pr[(X', Z') \in M] < 1$.

distributional parameter present in the model. While we assume that the marginal distributions of ε and ν are known, the restrictions placed on these marginal distributions are weak. This assumption of known marginal distributions is relaxed in Section 6. In [Assumption 3](#), the copula associated with the joint distribution is unique by Sklar's theorem. This assumption specifies that the joint dependence between ε and ν is fully characterized by a scalar parameter ρ . In the special case of a bivariate normal distribution discussed below, ρ is the usual correlation coefficient of (ε, ν) with $\Omega = (-1, 1)$, excluding the endpoints. Note that the endogeneity of D comes from allowing ρ to be nonzero. [Assumption 4](#) is the standard full rank condition found in most identification analyses. Let $\tilde{\Psi}$ be the parameter space of $\tilde{\psi} \equiv (\alpha', \beta', \delta_1, \gamma', \rho)$.

To keep our identification analyses simple for the case of known marginal (Sections 4 and 5), we assume that the reduced-form parameters (α, γ) are (globally) identified by the standard identification exercise of a single-equation threshold crossing model with known distribution: Since $\nu \perp (X, Z)$, it follows that $\Pr[D = 1|X = x, Z = z] = F_\nu(x'\alpha + z'\gamma)$, or

$$x'\alpha + z'\gamma = F_\nu^{-1}(\Pr[D = 1|X = x, Z = z]). \quad (2.2)$$

Therefore, as long as (X', Z') does not lie in a proper linear subspace of \mathbb{R}^{k+l} a.s. ([Assumption 4](#)), we globally identify (α, γ) from Eq. (2.2).

3. Dependence orderings for copulas

In order to obtain meaningful identification results, 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 the 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*). For r.v.'s W_1 and W_2 , W_2 is SI in W_1 or the conditional distribution $F_{1|2}(w_1|w_2)$ is SI, if $\Pr[W_1 > w_1|W_2 = w_2] = 1 - F_{1|2}(w_1|w_2)$ is increasing in w_2 for all w_1 .

The stochastically increasing “SI” property is a positive dependence condition as W_1 is more likely to take on larger values as W_2 increases. This condition is related to the FOSD in the literature. Specifically, the condition can be equivalently stated as “ $F_{1|2}(w_1|w_2)$ first-order stochastically dominates $F_{1|2}(w_1|w_2')$ for any $w_2 > w_2'$ ”. For negative dependence, stochastically decreasing “SD” property can be defined analogously, where $\Pr[W_1 > w_1|W_2 = w_2]$ is decreasing in w_2 . 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 $F_{1|2}(w_1|w_2)$ and $\tilde{F}_{1|2}(w_1|w_2)$ be respective conditional distributions of the first r.v. given the second that are SI (or SD). Suppose that $F_{1|2}(w_1|w_2)$ and $\tilde{F}_{1|2}(w_1|w_2)$ are continuous in w_1 for all w_2 . Then $\tilde{F}_{1|2}$ is strictly more SI (or less SD) than $F_{1|2}$ if $\psi(w_1, w_2) \equiv \tilde{F}_{1|2}^{-1}(F_{1|2}(w_1|w_2)|w_2)$ is strictly increasing in w_2 ,⁶ which is denoted as $F_{1|2} <_S \tilde{F}_{1|2}$.

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

w_2 increases. More formally, if $\psi(w_1, w_2)$ is a solution to $\Pr[\tilde{W}_1 > \psi(w_1, w_2)|\tilde{W}_2 = w_2] = \Pr[W_1 > w_1|W_2 = w_2]$, then $\psi(w_1, w_2)$ takes a larger value to compensate that W_1 is even more likely to take on larger values with \tilde{F} than it is with F as w_2 increases. The SI dependence ordering has been called the (strictly) “more regression dependent” or “more monotone regression dependent” ordering in the statistics literature. Using this definition, we assume that the ordering is indexed with respect to ρ for the copula $C(\cdot, \cdot; \rho)$. Let $C(\cdot|\cdot; \rho)$ be the conditional copula of $C(\cdot, \cdot; \rho)$.⁷

Assumption 5 ($<_S$ w.r.t. ρ). The copula $C(u_1, u_2; \rho)$ of [Assumption 3](#) satisfies

$$C(u_1|u_2; \rho_1) <_S C(u_1|u_2; \rho_2) \text{ for any } \rho_1 < \rho_2. \quad (3.1)$$

This assumption states that the copula satisfies the more SI (less SD) ordering, or equivalently FOSD ordering, with respect to the dependence parameter ρ . [Assumption 5](#) defines a class of copulas that is sufficient for us to derive identification results; below we provide a more general class of copulas.

We now introduce a dependence ordering concept that is more general than the more SI (less SD) ordering. Let $F(w_1, w_2)$ and $\tilde{F}(w_1, w_2)$ be bivariate distributions and $F(w_1)$ and $\tilde{F}(w_1)$ be marginal distributions. Also let $D(w_1, w_2) \equiv F(w_1) - \tilde{F}(w_1, w_2)$ and $\tilde{D}(w_1, w_2) \equiv \tilde{F}(w_1) - \tilde{F}(w_1, w_2)$.

Definition 3.3 (*Strictly More SI or Less SD in Joint Distribution*). Suppose that $F(w_1, w_2)$ and $\tilde{F}(w_1, w_2)$ are continuous in (w_1, w_2) . Then \tilde{F} is strictly more SI (or less SD) in joint distribution than F if $\psi^*(w_1, w_2) \equiv \tilde{F}^{-1}(w_1, F(w_1, w_2))$ and $\psi^{**}(w_1, w_2) \equiv \tilde{D}^{-1}(w_1, D(w_1, w_2))$ are strictly increasing in w_2 , which is denoted as $F <_{SJ} \tilde{F}$.

This ordering is a variant of the more SI (less SD) ordering, where joint distributions are used in place of conditional distributions. To the best of our knowledge, no result has been found in the literature that defines this concept and that shows its connection to the more SI (less SD) ordering, which is nontrivial ([Lemma 3.1](#) below). This new ordering concept is important in our context, since it characterizes minimal structure we need on the copula function for identification. Using this definition, we make the following assumption:

Assumption 6 ($<_{SJ}$ w.r.t. ρ). The copula $C(u_1, u_2; \rho)$ of [Assumption 3](#) satisfies

$$C(u_1, u_2; \rho_1) <_{SJ} C(u_1, u_2; \rho_2) \text{ for any } \rho_1 < \rho_2. \quad (3.2)$$

In the next lemma, we establish the connection between $<_S$ and $<_{SJ}$.

Lemma 3.1. *Assumption 5 implies Assumption 6.*

The proofs of this lemma and other results below are found in [Appendix](#).

The orderings $<_S$ and $<_{SJ}$ are not symmetric in arguments in general, but are 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 $<_S$ ” and (3.2) as “ C is increasing in $<_{SJ}$ ”. There are many well-known symmetric single-parameter copulas that satisfy [Assumption 5](#), i.e., that are increasing in $<_S$. By [Lemma 3.1](#), these copulas are also increasing in $<_{SJ}$. We list below well-known single-parameter copulas that satisfy [Assumption 5](#); see [Joe \(1997, pp. 140–142\)](#) for the results. In [Appendix](#), we list other copulas and show that they satisfy [Assumption 5](#) or the implication of it,

⁶ Note that $\psi(w_1, w_2)$ is increasing in w_1 by definition.

⁷ This notation and the terminology are commonly used in the literature; see e.g., [Joe \(1997\)](#) or [Fan and Liu \(2015\)](#).

namely, **Assumption 6**. In each example, Ω is defined as the interior of the parameter space of ρ .⁸

Example 3.1. The normal copula: For $\rho \in [-1, 1]$,

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

where $\Phi(\cdot, \cdot, \rho)$ is the bivariate standard normal distribution function and $\Phi(\cdot)$ is the marginal standard normal distribution function.

Example 3.2. The Plackett family: For $\rho \in [0, \infty) \setminus \{1\}$,

$$C(u_1, u_2; \rho) = \frac{1}{2\eta} \left\{ 1 + \eta(u_1 + u_2) - [(1 + \eta(u_1 + u_2))^2 - 4\rho\eta u_1 u_2]^{1/2} \right\},$$

where $\eta = \rho - 1$.

Example 3.3. The Frank family: For $\rho \in (-\infty, \infty) \setminus \{0\}$,

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

Example 3.4. The Kimeldorf and Sampson family (or the Clayton family): For $\rho \in [0, \infty)$,

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

Example 3.1 provides likely the most interesting case. With the normal copula and, additionally, marginal standard normal distributions $F_\varepsilon(\cdot) = F_\nu(\cdot) = \Phi(\cdot)$, the model of Eq. (2.1) becomes a *bivariate probit model*.

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., $Z = Z_1$ is scalar), so that

$$\begin{aligned} Y &= \mathbf{1}[\beta_0 + \delta_1 D - \varepsilon \geq 0], \\ D &= \mathbf{1}[\alpha_0 + \gamma_1 Z_1 - \nu \geq 0]. \end{aligned} \tag{4.1}$$

Let Ψ be the parameter space of $\psi \equiv (\alpha_0, \gamma_1, \beta_0, \delta_1, \rho)'$. For this simple stylized model, we further assume that Z_1 is a binary variable, namely, $Z_1 \in \text{supp}(Z_1) = \{0, 1\}$, where $\text{supp}(\cdot)$ denotes the support of its argument. We show that ψ is locally and globally identified with this minimal variation. In the following sections, we show how the results for this simple stylized model are readily generalized to the full model of Eq. (2.1) with possibly vector valued X and Z and without requiring that any element of Z be binary (Section 5), and to a model with nonparametric marginal distributions (Section 6). Before proceeding, recall that the reduced-form parameters (α_0, γ_1) are (globally) identified, since $(1, Z_1)$ does not lie in a proper linear subspace of \mathbb{R} a.s. by trivially assuming that Z_1 is non-degenerate.

4.1. Local identification

Local identification is necessary for global identification, and thus can be seen as a first step towards global identification. Particularly in our analysis, local identification results guide us to build a framework for global identification; see Section 4.2. In general, local identification requires a set of weaker assumptions

⁸ The copulas in **Examples 3.4, A.1 and A.2** in **Appendix** only allow 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 some of the copulas listed in this paper. Also, see **Nelsen (1999, p. 68, pp. 96–97)**.

than global identification. If one has additional prior knowledge to select one local solution from all others (by restricting the parameter space or by an explicit decision rule), local identification analysis itself can be useful for estimation.

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 Eq. (4.1). For instance, $\Pr[Y = 1, D = 1 | Z_1 = 0]$ can be expressed as

$$\begin{aligned} \Pr[Y = 1, D = 1 | Z_1 = 0] &= \Pr[\varepsilon \leq \beta_0 + \delta_1, \nu \leq \alpha_0; \rho] \\ &= \Pr[U_1 \leq F_\varepsilon(\beta_0 + \delta_1), U_2 \leq F_\nu(\alpha_0); \rho] \\ &= C(F_\varepsilon(\beta_0 + \delta_1), F_\nu(\alpha_0); \rho), \end{aligned}$$

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

$$(\alpha_0, \gamma_1, \beta_0, \delta_1)' \mapsto (a_0, a_1, b_0, b_1)' \tag{4.2}$$

denote a mapping such that

$$\begin{aligned} a_0 &\equiv F_\nu(\alpha_0), \\ a_1 &\equiv F_\nu(\alpha_0 + \gamma_1), \\ b_0 &\equiv F_\varepsilon(\beta_0), \\ b_1 &\equiv F_\varepsilon(\beta_0 + \delta_1), \end{aligned}$$

and note that the mapping is one-to-one since F_ν and F_ε are strictly increasing by **Assumption 2**. Let $p_{y,d,z} \equiv \Pr[Y = y, D = d | Z_1 = z]$ for $(y, d, z) \in \{0, 1\}^3$. Now, the six fitted probabilities can be written as follows:

$$\begin{aligned} p_{11,0} &= C(b_1, a_0; \rho), \\ p_{11,1} &= C(b_1, a_1; \rho), \\ p_{10,0} &= b_0 - C(b_0, a_0; \rho), \\ p_{10,1} &= b_0 - C(b_0, a_1; \rho), \\ p_{01,0} &= a_0 - C(b_1, a_0; \rho), \\ p_{01,1} &= a_1 - C(b_1, a_1; \rho). \end{aligned} \tag{4.3}$$

Eq. (4.3) contains the maximal set of probabilities that are not superfluous, since these probabilities imply the values of $p_{00,1}$ and $p_{00,0}$.

Among (4.3), $p_{01,0}$ and $p_{01,1}$ are superfluous, since a_0 and a_1 are already identified by using $p_{11,0} + p_{01,0}$ and $p_{11,1} + p_{01,1}$. Let $\theta \equiv (b_0, b_1, \rho)'$ denote the structural parameter vector in a parameter space $\Theta \subseteq (0, 1)^2 \times \Omega$ and $\tilde{\pi} \equiv (p_{11,0}, p_{11,1}, p_{10,0}, p_{10,1})'$ be a reduced-form parameter vector in a parameter space $\tilde{\Pi} \subseteq (0, 1)^4$, which is trivially identified as $p_{y,d,z}$'s are the distributions of the data. Therefore, our (local) identification problem is a question of whether we can *uniquely* recover the true structural parameter $\theta^0 \equiv (b_0^0, b_1^0, \rho^0)'$ given true reduced form parameter $\tilde{\pi}^0$.

Define $G : \Theta \subseteq (0, 1)^2 \times \Omega \rightarrow \tilde{\Pi} \subseteq (0, 1)^4$ as

$$G(\theta) \equiv G(\theta; a_0, a_1) \equiv \begin{bmatrix} C(b_1, a_0; \rho) \\ C(b_1, a_1; \rho) \\ b_0 - C(b_0, a_0; \rho) \\ b_0 - C(b_0, a_1; \rho) \end{bmatrix}, \tag{4.4}$$

and write

$$\tilde{\pi}^0 = G(\theta^0). \tag{4.5}$$

Then θ^0 is (locally) identifiable if and only if, from Eq. (4.5), $\tilde{\pi}^0$ uniquely determines θ^0 in the neighborhood of θ^0 . Let

$$J_G(\theta) \equiv \frac{\partial G(\theta)}{\partial \theta'} \tag{4.6}$$

be the Jacobian matrix of $G(\theta)$.⁹ Then by the standard implicit function theorem, the full rank of J_G ensures the identifiability (e.g., [Rothenberg \(1971, Theorem 6\)](#)):

Proposition 4.1. *Assume that there exists an open neighborhood of θ^0 in which $J_G(\theta)$ has constant rank. Then θ^0 is locally identifiable if and only if $J_G(\theta^0)$ has rank equal to $\dim(\theta)$.*

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 ρ , respectively. By conducting elementary row and column operations on the Jacobian matrix $J_G(\theta)$ for a given value of θ (see [Appendix A.4](#) in [Appendix](#)) which preserves the rank, it is easy to see that the matrix has full column rank if and only if either

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

$$\text{or } \frac{C_\rho(b_0, a_1; \rho)}{1 - C_1(b_0, a_1; \rho)} - \frac{C_\rho(b_0, a_0; \rho)}{1 - C_1(b_0, a_0; \rho)} \neq 0. \tag{4.7}$$

The main result of this section is to show that, under [Assumption 6](#), the condition (4.7) is true for θ^0 if and only if $a_0^0 \neq a_1^0$ (that is, γ_1^0 , the coefficient on Z , is nonzero).

Lemma 4.1. *Under [Assumption 3](#), the copula $C(u_1, u_2; \rho)$ satisfies [Assumption 6](#) if and only if*

$$\frac{C_\rho(u_1, u_2; \rho)}{C_1(u_1, u_2; \rho)} \text{ is strictly decreasing in } u_2, \tag{4.8}$$

and

$$\frac{C_\rho(u_1, u_2; \rho)}{1 - C_1(u_1, u_2; \rho)} \text{ is strictly increasing in } u_2, \tag{4.9}$$

for any $(u_1, u_2) \in (0, 1)^2$ and $\rho \in \Omega$.

To give some intuition behind the conditions in the lemma, with the normal copula of [Example 3.1](#), $\frac{C_\rho(u_1, u_2; \rho)}{C_1(u_1, u_2; \rho)}$ and $\frac{C_\rho(u_1, u_2; \rho)}{1 - C_1(u_1, u_2; \rho)}$ become (rescaled) inverse Mill's ratios, and thus (4.8) and (4.9) immediately hold; see [Appendix A.3.3](#) in [Appendix](#). Given the result of [Lemma 4.1](#), the desired result follows since the strict monotonicity in (4.8) and (4.9) implies that $a_0^0 = a_1^0$ if and only if $\frac{C_\rho(b_1, a_0^0; \rho)}{C_1(b_1, a_0^0; \rho)} = \frac{C_\rho(b_1, a_1^0; \rho)}{C_1(b_1, a_1^0; \rho)}$ and $\frac{C_\rho(b_0, a_1^0; \rho)}{1 - C_1(b_0, a_1^0; \rho)} = \frac{C_\rho(b_0, a_0^0; \rho)}{1 - C_1(b_0, a_0^0; \rho)}$. 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 [6](#) 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 $\gamma_1^0 \neq 0$ and Z_1 is non-degenerate.*

The identification condition is the exclusion restriction that the coefficient on the instrument Z_1 is nonzero. This condition implies that the excluded instrument plays a key role in identifying the parameters of the stylized model. This can be readily seen from the fact that, when $\gamma_1 = 0$ and hence $a_0 = a_1 \equiv a$, the fitted probabilities (4.3) reduce down to three equations, which are not enough to identify four unknowns, (a, b_0, b_1, ρ) :

$$p_{11,0} = p_{11,1} = C(b_1, a; \rho),$$

$$p_{10,0} = p_{10,1} = b_0 - C(b_0, a; \rho),$$

$$p_{01,0} = p_{01,1} = a - C(b_1, a; \rho).$$

[Assumption 6](#) characterizes an interpretable structure of copula that is minimally required for identification. It is minimal because [Assumption 6](#) is necessary and sufficient for (4.8) and (4.9) as shown in [Lemma 4.1](#). By [Lemma 3.1](#), this assumption is implied by

[Assumption 5](#), which is further well-understood in the literature. The gap between [Assumptions 5](#) and [6](#) essentially comes from the difference between the orderings defined in terms of conditional copulas and the ordering defined in terms of copulas.

4.2. Global identification

Based on the result of the local identification, we now establish global identification. In essence, the task is to show the uniqueness of a solution that satisfies a system of nonlinear equations, where the number of equations can be larger than the dimension of the solution. [Rothenberg \(1971\)](#) also derives a global identification result based on the [Gale and Nikaido \(1965\)](#) type of a global univalence result by imposing conditions on the square sub-matrix of J_G . The conditions, however, are restrictive and difficult to verify in our setting.¹⁰ Instead, we propose an identification analysis that makes use of the Hadamard's global inverse function theorem in a sub-system of equations; see, e.g., [Chernozhukov and Hansen \(2005\)](#) for a related approach.¹¹

Lemma 4.2. *For $n \leq m$, let A and B be nonempty subsets of \mathbb{R}^n and \mathbb{R}^m , respectively, and $g : A \rightarrow B$ be a continuously differentiable map. Let g_s be the s th $n \times 1$ sub-block of g for some arbitrary ordering $s = 1, \dots, \binom{m}{n}$. If there exists s such that (i) g_s is proper, (ii) the Jacobian of g_s vanishes nowhere, and (iii) $g_s(A)$ is simply connected, then a solution of $b = g(a)$ is unique.*

A mapping $g_s : A \rightarrow g_s(A)$ is proper if whenever $K \subset g_s(A)$ is compact then $g_s^{-1}(K) \subset A$ is compact. A topological space is simply connected if it is path-connected and any simple closed curve can be shrunk to a point.¹² Note that, for example, any convex subset of \mathbb{R}^n and its half spaces are simply connected.¹³ The proof of this lemma is as follows: Suppose a^\dagger is a solution of the system $b = g(a)$. By the global inverse function theorem ([Hadamard, 1906a, b](#)), the conditions (i), (ii), and (iii) guarantee that g_s is a homeomorphism and hence one-to-one and onto. Therefore, a^\dagger is the unique solution of the sub-system $b_s = g_s(a)$, where b_s is the corresponding subvector of b . Since a^\dagger must satisfy the remaining equations as well, we can conclude that a^\dagger is the unique solution of the system $b = g(a)$.

For our global identification, we apply the result of this lemma to the map (4.5) introduced in the previous section. In this case, establishing the result with any of the possible 3×1 sub-blocks of G will serve our purpose. For concreteness, we consider the following specific sub-block $G^* : \Theta \subseteq (0, 1)^2 \times \Omega \rightarrow \Pi \subseteq (0, 1)^3$

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

and a sub-system $\pi = G^*(\theta)$, where $\pi \equiv (p_{11,0}, p_{11,1}, p_{10,0})'$ in its parameter space Π . Under [Assumption 6](#),^{*} one can show that the square Jacobian matrix $J_{G^*} \equiv J_{G^*}(\theta) \equiv \frac{\partial G^*(\theta)}{\partial \theta'}$ is positive semi-definite for $a_0 > a_1$ and negative semi-definite for $a_0 < a_1$, and has full rank for all $\theta \in \Theta$ such that $a_0 \neq a_1$; see [Appendix A.4](#)

¹⁰ Using the notations of our paper, the conditions are that there exists a square matrix $\bar{J}(\theta)$ of $J_G(\theta)$ such that the determinant of $\bar{J}(\theta)$ is positive and $\bar{J}(\theta) + \bar{J}(\theta)$ is positive semidefinite throughout Θ . The latter condition appears not to be feasible to verify in all of our settings, including the stylized model, the full model, and a semiparametric model below.

¹¹ For nonparametric identification, [Chernozhukov and Hansen \(2005\)](#) apply a variant of the global inverse function theorem to their sub-system (Theorem 3, p. 258). Unlike their approach, we do not restrict our parameter space to be compact nor convex.

¹² Refer to [Rudin \(1986, p. 222\)](#) for a technical definition of the simple connectedness.

¹³ A half-space is either of the two parts into which a hyperplane divides a space.

⁹ See [Appendix](#) for the actual expression of $J(\theta)$.

in Appendix for the proof. Based on these results, we now show that G^* on restricted parameter spaces satisfies (i), (ii), and (iii) of Lemma 4.2. We then extend the map over the entire parameter space to draw our conclusion.

Define $\Theta_c \subseteq (0, 1)^2 \times \Omega$ to be a 3-dimensional bounded 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.

Note that $G^*|_{\Theta_{c1}}(\cdot)$ 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 Θ_{c1} and Θ_{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) Θ_{c1} and Θ_{c2} are simply connected, (ii) $G^*|_{\Theta_{c1}}(\theta)$ and $G^*|_{\Theta_{c2}}(\theta)$ are continuous on Θ_{c1} and Θ_{c2} , respectively, and (iii) J_{G^*} is positive and negative semi-definite on Θ_{c1} and Θ_{c2} , respectively, it follows that Π_{c1} , and Π_{c2} are also simply connected.¹⁴

Lastly, J_{G^*} has full rank over Θ_{c1} and Θ_{c2} . Therefore, $G^*|_{\Theta_{c1}} : \Theta_{c1} \rightarrow \Pi_{c1}$ and $G^*|_{\Theta_{c2}} : \Theta_{c2} \rightarrow \Pi_{c2}$ satisfy all the conditions in Lemma 4.2, which means that $\tilde{\pi} = G(\theta)$ has a unique solution on Θ_{c1} and Θ_{c2} , respectively. Since there exist $G^*|_{\Theta_{c1}}^{-1}(\cdot)$ and $G^*|_{\Theta_{c2}}^{-1}(\cdot)$, such a solution can be expressed as $\theta = G^*|_{\Theta_{c1}}^{-1}(\pi) \in \Theta_{c1}$ for $\pi \in \Pi_{c1}$ and $\theta = G^*|_{\Theta_{c2}}^{-1}(\pi) \in \Theta_{c2}$ for $\pi \in \Pi_{c2}$. This proves that the parameter θ is globally identified in Θ_{c1} and in Θ_{c2} .

Now we use these results to derive global identification over $\Theta_1 \equiv \{\theta \in \Theta : a_0 > a_1\}$ and $\Theta_2 \equiv \{\theta \in \Theta : a_0 < a_1\}$, which are not necessarily bounded. Above results suggest that θ is globally identified in any given subset of Θ_1 or Θ_2 that is a bounded simply connected set. Assume that the original parameter space Ψ 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 one-to-one map defined in (4.2), the transformed parameter space Θ is open and Θ_1 and Θ_2 are open, simply connected. Then Θ_1 and Θ_2 can be represented by a countable union of bounded 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 bounded open simply connected sets in Θ_1 such that $\Theta_{11} \subset \Theta_{12} \subset \dots \subset \Theta_1$. Also, let $G^*(\Theta_{1i}) \equiv \Pi_{1i}$ for $i = 1, 2, \dots$, so that $\Pi_1 = G^*(\Theta_1) = G^*(\bigcup_{i=1}^{\infty} \Theta_{1i}) = \bigcup_{i=1}^{\infty} G^*(\Theta_{1i}) = \bigcup_{i=1}^{\infty} \Pi_{1i}$ and $\Pi_{11} \subset \Pi_{12} \subset \dots \subset \Pi_1$. Then, for any given $\pi \in \Pi_1$, we have that $\pi \in \Pi_{1i}$ for all $i \geq q$ (for some q), then $G^*|_{\Theta_{1i}}^{-1}(\pi) \in \Theta_{1i}$ for all $i \geq q$ from the previous result, and hence $G^{*-1}(\pi) = G^*|_{\bigcup_{i=q}^{\infty} \Theta_{1i}}^{-1}(\pi) \in \bigcup_{i=q}^{\infty} \Theta_{1i} = \Theta_1$. As $G^{*-1}(\pi)$ is the unique solution of the sub-system $\pi = G^*(\theta)$ on Θ_1 , it is the unique solution of the full system $\tilde{\pi} = G(\theta)$ on Θ_1 by similar reasoning as in the proof of Lemma 4.2. Therefore, θ is globally identified in Θ_1 . Then, adding the reduced-form parameters, we can conclude that ψ is globally identified in Ψ_1 . By similar arguments, ψ is globally identified in Ψ_2 . Since γ_1 is already identified, it is known whether ψ lies in Ψ_1 or Ψ_2 and consequently, ψ is globally identified in Ψ if $\gamma_1 \neq 0$.

The following theorem summarizes the results:

¹⁴ This is because simple connectedness is preserved under a monotone map; see, e.g., Arnold (2009, p. 33). For an arbitrary function $\tilde{G} : \Theta \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$, $\tilde{G}(\cdot)$ is monotone on Θ if for all $\theta_1, \theta_2 \in \Theta$, $(\theta_1 - \theta_2)'(\tilde{G}(\theta_1) - \tilde{G}(\theta_2))$ is non-negative or non-positive. By the mean value theorem,

$$\tilde{G}(\theta_1) - \tilde{G}(\theta_2) = \frac{\partial \tilde{G}(\theta^*)}{\partial \theta'}(\theta_1 - \theta_2)$$

where the intermediate value θ^* may differ across the rows of $\partial \tilde{G}(\theta^*)/\partial \theta'$. Then,

$$(\theta_1 - \theta_2)'(\tilde{G}(\theta_1) - \tilde{G}(\theta_2)) = (\theta_1 - \theta_2)' \frac{\partial \tilde{G}(\theta^*)}{\partial \theta'}(\theta_1 - \theta_2).$$

Therefore, as long as $\partial \tilde{G}(\theta^*)/\partial \theta'$ is positive (negative) semi-definite for all θ^* then $\tilde{G}(\cdot)$ is monotone.

Theorem 4.2. In model (4.1), let Assumptions 1–3 and 6 hold. Then $(\alpha_0, \gamma_1, \beta_0, \delta_1, \rho) \in \Psi$ is globally identified if (i) $\gamma_1 \neq 0$ and Z_1 is non-degenerate; (ii) Ψ is open and Ψ_1 and Ψ_2 are simply connected.

Again, Assumption 5 is sufficient for Assumption 6. To satisfy (ii), one can simply have $\Psi = \mathbb{R}^4 \times \Omega$ where Ω is open. In fact, any open convex Ψ is sufficient, although it is not necessary. Note that we do not assume the compactness of the parameter space either.

5. Identification in full model

In this section, we conduct identification analysis of the full model of Eq. (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 , and we allow for the possibility of instruments Z being vector valued without requiring any element of Z to be binary. We present results for global identification of $\tilde{\psi} = (\alpha', \beta', \delta_1, \gamma', \rho)'$ in $\tilde{\Psi}$. Local identification results can be obtained by a similar argument as in the previous section; see discussions at the end of this section.

Recall that (α, γ) are identified. Suppose that γ is a nonzero vector, i.e., there exists at least one variable in Z with one 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 so 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 a one-to-one map as

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

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

$$\begin{aligned} p_{11,xz} &= C(r_1, s_0; \rho), \\ p_{11,x\tilde{z}} &= C(r_1, s_1; \rho), \\ p_{10,xz} &= r_0 - C(r_0, s_0; \rho), \\ p_{10,x\tilde{z}} &= r_0 - C(r_0, s_1; \rho), \\ p_{01,xz} &= s_0 - C(r_1, s_0; \rho), \\ p_{01,x\tilde{z}} &= s_1 - C(r_1, s_1; \rho). \end{aligned} \tag{5.1}$$

The set of equations has the same form as (4.3) in the previous section. By pursuing a similar argument as in the previous section, identification of $\theta_x \equiv (r_0, r_1, \rho)'$ in its parameter space Θ_x is equivalent to being able to show the uniqueness of the solution for

$$\tilde{\pi}_x = G(\theta_x) \equiv G(\theta_x; s_0, s_1), \tag{5.2}$$

where G is defined in (4.4) and $\tilde{\pi}_x \equiv (p_{11,xz}, p_{11,x\tilde{z}}, p_{10,xz}, p_{10,x\tilde{z}})'$ in its parameter space $\tilde{\Pi}_x$. The subscript x emphasizes the objects' dependence on x . Now we proceed similar to the proof of Theorem 4.2: Under Assumption 6, $J_{G^*}(\theta_x)$ is either positive or negative semi-definite and has full rank for any θ_x and x , since $z'\gamma \neq \tilde{z}'\gamma$ implies $s_0 \neq s_1$. By Lemma 4.2, θ_x is identified in simply connected half spaces of a bounded open set. Assume that the original parameter space $\tilde{\Psi}$ is open and convex. Then, for any x , its linear map $\{(x'\beta, x'\beta + \delta_1, \rho) : \tilde{\psi} \in \tilde{\Psi}\}$ is also open and convex, and hence simply connected. Since Θ_x is a continuous and one-to-one map of this set, Θ_x is open and simply connected. This implies that $\Theta_{1,x} \equiv \{\theta_x \in \Theta_x : s_0 > s_1\}$ and $\Theta_{2,x} \equiv \{\theta_x \in \Theta_x : s_0 < s_1\}$ are also open and simply connected, and therefore can be approximated by sequences of bounded, open, and simply

connected sets. Eventually, it follows that, for any given $\pi_x \in G^*(\Theta_{1,x})$,

$$\theta_x = G^{*-1}(\pi_x) \in \Theta_{1,x}$$

is the unique solution of $\tilde{\pi}_x = G(\theta_x)$ and hence θ_x is globally identified in $\Theta_{1,x}$. Similarly, θ_x is globally identified in $\Theta_{2,x}$. Since s_0 and s_1 are known, we can conclude that θ_x is globally identified in Θ_x . Identification of δ_1 follows from

$$\delta_1 = F_\varepsilon^{-1}(r_1) - F_\varepsilon^{-1}(r_0).$$

Let

$$\mathcal{X} \equiv \bigcup_{\substack{z', \gamma \neq z', \gamma \\ z, \tilde{z} \in \text{supp}(Z)}} \text{supp}(X|Z = z) \cap \text{supp}(X|Z = \tilde{z}).$$

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

$$x'\beta = F_\varepsilon^{-1}(r_0),$$

assuming that \mathcal{X} does not lie in a proper linear subspace of \mathbb{R}^k a.s. The following theorem summarizes the identification result.

Theorem 5.1. *In model (2.1), let Assumptions 1–4 and 6 hold. Then $(\alpha', \beta', \delta_1, \gamma, \rho) \in \tilde{\Psi}$ are globally identified if (i) γ is a nonzero vector; (ii) \mathcal{X} is not empty and does not lie in a proper linear subspace of \mathbb{R}^k a.s.; (iii) $\tilde{\Psi}$ is open and convex.*

Condition (i) requires an exclusion restriction. A sufficient condition for Condition (ii) is that $\text{supp}(X, Z) = \text{supp}(X) \times \text{supp}(Z)$ and Assumption 4 holds, since $\mathcal{X} = \text{supp}(X)$ in this case. Note that Condition (ii) implies that there exist z and \tilde{z} in $\text{supp}(Z)$ such that $\text{supp}(X|Z = z) \cap \text{supp}(X|Z = \tilde{z})$ is nonempty. Note that local identification is achieved maintaining Assumptions 1–4 and 6 and (i) and (ii) of Theorem 5.1. Compared to Theorem 4.1, the rank conditions relevant to the full model (Assumption 4 and (iii)) are added.

6. Identification with unknown marginals

As is mentioned earlier, the assumption that the marginal distributions of the error terms (ε, ν) are known (Assumption 2) is not essential in the identification analyses of this paper. In this section, we extend our identification results of the previous section by relaxing Assumption 2. Here, we identify the structural and reduced-form parameters as well as the unknown marginal distributions. In order to identify the marginal distributions, it is necessary to have sufficient exogenous variation in each equation, which can be provided by the common exogenous covariate X present in each equation. We illustrate the proof using the full model (2.1).

Assumption 7. (i) F_ε and F_ν are (unknown) marginal distributions of ε and ν , respectively, that are strictly increasing and are absolutely continuous with respect to Lebesgue measure. (ii) The index structure in each equation of (2.1) has no intercept and the first coefficient is 1.

Assumption 7(i) relaxes the assumption of known marginal distributions in Assumption 2. For convenience, instead of imposing $E[\varepsilon] = E[\nu] = 0$ and $\text{Var}(\varepsilon) = \text{Var}(\nu) = 1$ as location and scale normalizations as in Assumption 2, Assumption 7(ii) imposes an alternative location and scale normalizations to facilitate the analysis of this section.¹⁵ The next assumption is the additional support condition.

¹⁵ The model (2.1) with one type of normalizations can always be rewritten with another type of normalization. For example, under Assumption 5(ii), let $\mu \equiv E[\varepsilon]$ and $\sigma^2 \equiv \text{Var}(\varepsilon)$, and also let $\tilde{X} \equiv (X_1, \dots, X_k)'$ and $\tilde{\beta} \equiv (\beta_1, \dots, \beta_k)'$. Then $Y = \mathbf{1}[\tilde{X}'\tilde{\beta} + \delta_1 D \geq \varepsilon] = \mathbf{1}[-\mu/\sigma + \tilde{X}'\tilde{\beta}/\sigma + (\delta_1/\sigma)D \geq (\varepsilon - \mu)/\sigma]$ so that $-\mu/\sigma$ becomes an intercept and $(\varepsilon - \mu)/\sigma$ becomes a new error term with mean zero and variance one. A similar argument applies for the D equation.

Assumption 8. (i) The distributions of X_i and Z_j are absolutely continuous with respect to Lebesgue measure for $1 \leq i \leq k$ and $1 \leq j \leq l$. (ii) There exists at least one element X_i in X such that its support conditional on $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_k)$ is \mathbb{R} and $\alpha_i \neq 0$ and $\beta_i \neq 0$. Without loss of generality, let $i = 1$.

Assumption 8(i) guarantees differentiability with respect to $X_i = x_i$ and $Z_j = z_j$. Assumption 8(ii) is a “large support” type of assumption. We require Assumption 8(i) but not the large support Assumption 8(ii) for identification of $(\alpha', \beta', \delta_1, \gamma, \rho) \in \tilde{\Psi}$. We additionally require the large support Assumption 8(ii) only for identification of the marginal distributions $F_\varepsilon(\cdot)$ and $F_\nu(\cdot)$.

For any x , we obtain global identification of $\theta_x \equiv (r_0, r_1, \rho)$ from (5.2) and the proof of Theorem 5.1, and global identification of (s_0, s_1) from $p_{11,xz} + p_{01,xz}$ and $p_{11,x\tilde{z}} + p_{01,x\tilde{z}}$ in (5.1). Recall that

$$s_0 \equiv F_\nu(x'\alpha + z'\gamma). \tag{6.1}$$

First, given identification of s_0 , we will now use Eq. (6.1) to identify α, γ , and $F_\nu(\cdot)$. The statistical independence assumption (Assumption 1) implies quantile independence as well as index sufficiency. Under this assumption and under assumptions similar to Assumptions 7–8, Manski (1988) provides identification results that follow his proof under quantile independence. Here we follow his proof strategy under index sufficiency.¹⁶ Under Assumptions 7 and 8(i), differentiating Eq. (6.1) yields

$$\frac{\partial s_0}{\partial x_1} = f_\nu(x'\alpha + z'\gamma)$$

and, for $2 \leq i \leq k$,

$$\frac{\partial s_0}{\partial x_i} = f_\nu(x'\alpha + z'\gamma)\alpha_i,$$

and for $1 \leq j \leq l$,

$$\frac{\partial s_0}{\partial z_j} = f_\nu(x'\alpha + z'\gamma)\gamma_j,$$

where $f_\nu(\cdot)$ is the density of ν . Then α_i and γ_j are identified for all i and j by

$$\alpha_i = \frac{\partial s_0 / \partial x_i}{\partial s_0 / \partial x_1}$$

and

$$\gamma_j = \frac{\partial s_0 / \partial z_j}{\partial s_0 / \partial x_1}.$$

Using that $(X, Z) \perp \nu$, we have $\Pr[D = 1|X = x, Z = z] = \Pr[D = 1|X'\alpha + Z'\gamma = t]$ for $t = x'\alpha + z'\gamma$ (index sufficiency). Also, $\text{Supp}(X'\alpha + Z'\gamma) = \mathbb{R}$ by Assumption 8(ii).¹⁷ Therefore, $f_\nu(\cdot)$ is identified on \mathbb{R} by

$$\begin{aligned} \frac{\partial s_0}{\partial x_1} &= \frac{\partial \Pr[D = 1|X = x, Z = z]}{\partial x_1} \\ &= \frac{\partial \Pr[D = 1|X'\alpha + Z'\gamma = t]}{\partial t} = f_\nu(t) \end{aligned}$$

for $t = x'\alpha + z'\gamma$. Since the density is identified, the distribution function $F_\nu(\cdot)$ is identified. Now we identify other components of the model by a similar fashion. Since we identify $r_0 = F_\varepsilon(x'\beta)$ for all $x \in \mathcal{X}$, similar as above,

$$\frac{\partial r_0}{\partial x_1} = f_\varepsilon(x'\beta)$$

¹⁶ Note that the normalization and assumptions (and hence the proof) in this paper are slightly different from Manski’s (1988) results for index sufficiency.

¹⁷ Note that for the D equation, the large support assumption can alternatively be imposed on Z .

and, for $2 \leq i \leq k$,

$$\frac{\partial r_0}{\partial x_i} = f_\varepsilon(x' \beta) \beta_i,$$

which identify β_i and $f_\varepsilon(\cdot)$. Finally, δ_1 can be identified by

$$\delta_1 = F_\varepsilon^{-1}(r_1) - F_\varepsilon^{-1}(r_0).$$

Theorem 6.1. *In model (2.1), suppose Assumptions 1, 3, 4, 6, 7 and 8(i) hold. Then $(\alpha', \beta', \delta_1, \gamma, \rho) \in \tilde{\Psi}$ are globally identified if (i) γ is a nonzero vector; (ii) \mathcal{X} is not empty and does not lie in a proper linear subspace of \mathbb{R}^k a.s.; (iii) $\tilde{\Psi}$ is open and convex. Additionally, if Assumption 8(ii) holds, $F_\varepsilon(\cdot)$ and $F_v(\cdot)$ are identified.*

7. Conclusions

We derive conditions for local and global identification in a class of models that generalize bivariate probit models. We show that the parameters are identified in such models with instruments, i.e., with covariates that enter into the equation for the endogenous treatment variable but are excluded from the equation for the outcome variable. We show that such models are identified with or without common exogenous regressors that enter into both equations. 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 the joint normality assumption by introducing a broad class of copulas for the joint distribution of the latent error terms while allowing their marginal distributions to be arbitrary but known. We show that our identification results extend to the case where the marginal distributions are unknown, with an additional large support for the identification of the distribution.

Based on the identification results of this paper, one can proceed to estimate the parameter $\tilde{\psi}$ and conduct inference on it. When the model is parametric (i.e., the triangular threshold crossing model (2.1) with Assumptions 2 and 3), one can employ standard maximum likelihood (ML) or generalized method of moment (GMM) procedures. When the model is semiparametric (i.e., the model (2.1) with Assumptions 3 and 7), one can apply similar semiparametric estimation methods, such as the plug-in sieve ML method (Chen et al., 2006) or the semiparametric GMM method (Chen et al., 2003). Han and Lee (2017) establish the asymptotic theory for sieve ML estimators in this semiparametric model, where the sieve is introduced to approximate the nonparametric marginal distributions. For a smooth functional of the sieve estimators, such as those for the parametric components (e.g., δ_1 and ρ in our notation), they establish asymptotic normality and derive the variance-covariance estimator which can be used for inference. Using Monte Carlo simulation, they also document the finite sample performance of the ML estimates based on a parametric model and the same parametric parts of the sieve ML estimates. Their simulation evidence suggests that: (i) in a correctly specified parametric model, the performance of the ML estimates in terms of MSE's is what one can expect from standard ML estimation, i.e., negligible bias and small variance; (ii) when the model is misspecified, either from a misspecified copula or misspecified marginal distributions, both bias and variance of the ML estimates substantially deteriorate; (iii) with the same data generating process as in (ii), the performance of the same parametric parts of the sieve ML estimates is significantly improved over the ML estimates of the misspecified parametric model. See Han and Lee (2017) for details and other related results.

In the parametric model, the performance of the ML estimates is also studied in Freedman and Sekhon (2010). One of their simulation findings is that the performance of the ML estimates deteriorates as the exogenous variation shrinks to zero (Fig. 2 in

Freedman and Sekhon, 2010). It is left unanswered in their paper, however, whether this finding is due to the failure of their large support assumption or of the requirement of any variation at all. The present paper suggests that it is in fact the latter, by showing that the parameters can be identified even with minimal variation in the excluded instrument (i.e., with a binary instrument).

The deterioration of the performance (such as larger bias) from shrinking exogenous variation is related to the fact that the finite sample distribution of the estimators becomes non-normal in this situation. This non-normality implies that standard inference methods based on the normal distribution show poor performances, such as size distortion. This opens up an interesting question on how to conduct inference that is robust to weak instruments in bivariate probit models and the more general class of models considered in this paper. While there is an extensive literature on weak instruments in linear models (see Andrews and Stock, 2007 for a complete survey), there is relatively little literature on weak instruments in nonlinear models (see, e.g., Stock and Wright, 2000; Kleibergen, 2005; Andrews and Mikusheva, 2016b, a; Andrews and Guggenberger, 2015), and no previous literature on inference under weak identification that nests the class of models considered in this paper. In current work, Han and McCloskey (2017) develop inference that is robust to non- and weak-identification in a broad class of models where the implied Jacobian has general deficient rank when identification fails, and where the source of such identification failure is known. As one example of their more general analysis, they develop an inference procedure for generalized bivariate probit models (with known marginals) that is robust to weak instruments. They exploit the identification results of the present paper in order to understand when the Jacobian will be nearly singular, and to introduce a transformation method to separately treat the weakly and strongly identified parameters in deriving nonstandard asymptotic theory. Based on their results, one can conduct a hypothesis test, say, for the average treatment effect ($F_\varepsilon(x' \beta + \delta_1) - F_\varepsilon(x' \beta)$ using our notation) that has correct asymptotic size regardless of identification strength and good power properties.

Appendix

A.1. Proof of Lemma 3.1

We provide the proof of Lemma 3.1 (which is restated here), naturally followed by the proof of Lemma 4.1 in Appendix A.2. Let $C : (0, 1)^2 \rightarrow (0, 1)$ and $\tilde{C} : (0, 1)^2 \rightarrow (0, 1)$ be two distinct copulas, succinctly denoted as $C(u_1, u_2) \equiv C(u_1, u_2; \rho_1)$ and $\tilde{C}(u_1, u_2) \equiv C(u_1, u_2; \rho_2)$, respectively, where $\rho_1 < \rho_2$. Define $D(u_1, u_2) \equiv u_1 - C(u_1, u_2)$ and $\tilde{D}(u_1, u_2) \equiv u_1 - \tilde{C}(u_1, u_2)$.

Lemma A.1. *Suppose $C(u_1|u_2) <_S \tilde{C}(u_1|u_2)$, 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 $C(u_1, u_2) <_{Sf} \tilde{C}(u_1, u_2)$, i.e., $u_1^*(u_1, u_2) = \tilde{C}^{-1}(C(u_1, u_2), u_2)$ and $u_1^{**}(u_1, u_2) = \tilde{D}^{-1}(D(u_1, u_2), u_2)$ are strictly increasing in u_2 .*

Proof of Lemma A.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 $\tilde{C}(u_1^*, u_2) = C(u_1, u_2)$ and $u_1^{**} = u_1^{**}(u_1, u_2)$ is strictly increasing in u_2 with u_1^{**} being the root of $u_1^{**} - \tilde{C}(u_1^{**}, u_2) = u_1 - C(u_1, u_2)$.

We first prove for u_1^* . 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

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

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

$$\tilde{C}^{-1}(C(u_1|u_2)|u_2) > u_1^*(u_1, u_2), \tag{A.1}$$

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

$$C(u_1|u_2) > \tilde{C}(u_1^*(u_1, u_2)|u_2). \tag{A.2}$$

Next, differentiating $\tilde{C}(u_1^*, u_2) = C(u_1, u_2)$ w.r.t. u_2 yields

$$\tilde{C}_1(u_1^*, u_2) \cdot \frac{\partial u_1^*}{\partial u_2} + \tilde{C}_2(u_1^*, u_2) = C_2(u_1, u_2), \tag{A.3}$$

or $\tilde{C}_1(u_1^*, u_2) \cdot \frac{\partial u_1^*}{\partial u_2} = C_2(u_1, u_2) - \tilde{C}_2(u_1^*, u_2) = C(u_1|u_2) - \tilde{C}(u_1^*|u_2)$.¹⁸

But since $\tilde{C}_1(u_1^*, u_2) = \tilde{C}(u_2|u_1^*) > 0$ for $u_2 \in (0, 1)$, (A.2) implies that $\frac{\partial u_1^*}{\partial u_2} > 0$.

Similarly, we prove for u_1^{**} . For any $u_2' > u_2$, we have $u_1^\dagger(u_1, u_2') > u_1^\dagger(u_1, u_2)$, and thus $\tilde{C}(u_1^\dagger(u_1, u_2')|u_2') > \tilde{C}(u_1^\dagger(u_1, u_2)|u_2)$. Then it follows that

$$\begin{aligned} u_1^\dagger(u_1, u_2) - \tilde{C}(u_1^\dagger(u_1, u_2), u_2) &= \int_{u_2}^1 \tilde{C}(u_1^\dagger(u_1, u_2)|u_2') du_2' \\ &< \int_{u_2}^1 \tilde{C}(u_1^\dagger(u_1, u_2')|u_2') du_2' \\ &= \int_{u_2}^1 C(u_1|u_2') du_2' \\ &= u_1 - C(u_1, u_2) \\ &= u_1^{**}(u_1, u_2) - \tilde{C}(u_1^{**}(u_1, u_2), u_2). \end{aligned}$$

Therefore, since $u_1 - \tilde{C}(u_1, u_2)$ is strictly increasing in u_1 , it follows that

$$u_1^\dagger(u_1, u_2) < u_1^{**}(u_1, u_2),$$

or

$$\tilde{C}^{-1}(C(u_1|u_2)|u_2) < u_1^{**}(u_1, u_2),$$

or

$$C(u_1|u_2) < \tilde{C}(u_1^{**}(u_1, u_2)|u_2). \tag{A.4}$$

Now, differentiating $u_1^{**} - \tilde{C}(u_1^{**}, u_2) = u_1 - C(u_1, u_2)$ w.r.t. u_2 yields

$$\left(1 - \tilde{C}_1(u_1^{**}, u_2)\right) \frac{\partial u_1^{**}}{\partial u_2} - \tilde{C}_2(u_1^{**}, u_2) = -C_2(u_1, u_2), \tag{A.5}$$

or $\left(1 - \tilde{C}_1(u_1^{**}, u_2)\right) \frac{\partial u_1^{**}}{\partial u_2} = -C_2(u_1, u_2) + \tilde{C}_2(u_1^{**}, u_2) = -C(u_1|u_2) + \tilde{C}(u_1^{**}|u_2)$. Since $1 - \tilde{C}_1(u_1^{**}, u_2) = 1 - \tilde{C}(u_2|u_1^{**}) > 0$ for $u_2 \in (0, 1)$, (A.4) implies that $\frac{\partial u_1^{**}}{\partial u_2} > 0$. \square

A.2. Proof of Lemma 4.1

Let $\rho_1 < \rho_2$ and follow the same notations as in Appendix A.1. Given Assumption 6, $u_1^* = u_1^*(u_1, u_2) = \tilde{C}^{-1}(C(u_1, u_2), u_2)$ is strictly increasing in u_2 with u_1^* being the root of

$$C(u_1^*, u_2; \rho_2) = C(u_1, u_2; \rho_1). \tag{A.6}$$

By (A.3) in the previous proof, $\frac{\partial u_1^*}{\partial u_2} > 0$ is equivalent to $C_2(u_1, u_2; \rho_1) - C_2(u_1^*, u_2; \rho_2) > 0$. Since $u_1^* = u_1^*(u_1, u_2, \rho_1, \rho_2) \rightarrow u_1$ as $\rho_1 \rightarrow \rho_2$ from (A.6), it is also equivalent to $\frac{\partial}{\partial \rho} C_2(u_1^*(\rho), u_2; \rho) < 0$, or

$$C_{12}(u_1^*(\rho), u_2; \rho) \cdot \frac{\partial u_1^*}{\partial \rho} + C_{\rho 2}(u_1^*(\rho), u_2; \rho) < 0. \tag{A.7}$$

Note that $\frac{\partial u_1^*}{\partial \rho} = -\frac{C_\rho}{C_1}$ by differentiating (A.6) w.r.t ρ_2 and by letting $\rho = \rho_2$. Therefore, (A.7) can be expressed as $C_{\rho 2} C_1 - C_\rho C_{12} < 0$, which is in turn equivalent to the condition (4.8).

Similarly, given Assumption 6, $u_1^{**} = u_1^{**}(u_1, u_2)$ is strictly increasing in u_2 with u_1^{**} being the root of

$$u_1^{**} - C(u_1^{**}, u_2; \rho_2) = u_1 - C(u_1, u_2; \rho_1). \tag{A.8}$$

By (A.5) in the previous proof, $\frac{\partial u_1^{**}}{\partial u_2} > 0$ is equivalent to $C_2(u_1, u_2; \rho_1) - C_2(u_1^{**}, u_2; \rho_2) < 0$ or $\frac{\partial}{\partial \rho} C_2(u_1^{**}(\rho), u_2; \rho) > 0$, or

$$C_{12}(u_1^{**}(\rho), u_2; \rho) \cdot \frac{\partial u_1^{**}}{\partial \rho} + C_{\rho 2}(u_1^{**}(\rho), u_2; \rho) > 0. \tag{A.9}$$

Note that $\frac{\partial u_1^{**}}{\partial \rho} = \frac{C_\rho}{1 - C_1}$ by differentiating (A.8) w.r.t ρ_2 and by letting $\rho = \rho_2$. Therefore, (A.9) can be expressed as $C_{\rho 2}(1 - C_1) + C_\rho C_{12} > 0$, which is in turn equivalent to the condition (4.9). \square

A.3. More copulas and verification of the assumptions

Here we list more copulas that satisfy the assumptions for identification. In each example, Ω is defined as the interior of the parameter space of ρ .

Example A.1. The Joe family: For $\rho \in [1, \infty)$,

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

Example A.2. The Gumbel family: For $\rho \in [1, \infty)$,

$$C(u_1, u_2; \rho) = \exp\{-[(-\log u_1)^\rho + (-\log u_2)^\rho]^{1/\rho}\}.$$

Example A.3. 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 A.4. The Farlie–Gumbel–Morgenstern family: For $\rho \in [-1, 1]$,

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

Examples A.1 and A.2 satisfy Assumption 5; see Joe (1997, pp. 140–142). Examples A.3 and A.4 are shown to satisfy Assumption 6. Define

$$\mu(u_1, u_2; \rho) \equiv \frac{C_\rho(u_1, u_2; \rho)}{C_1(u_1, u_2; \rho)}$$

and

$$\tilde{\mu}(u_1, u_2; \rho) \equiv \frac{C_\rho(u_1, u_2; \rho)}{1 - C_1(u_1, u_2; \rho)},$$

and their derivatives w.r.t. the second argument as $\mu_2(u_1, u_2; \rho)$ and $\tilde{\mu}_2(u_1, u_2; \rho)$.

¹⁸ In general, a copula satisfies that $C_1(u, v) = C(v|u)$ and $C_2(u, v) = C(u|v)$.

A.3.1. The Ali–Mikhail–Haq family (Example A.3)

Let $h(u_2) \equiv 1 - \rho(1 - u_1)(1 - u_2)$ for abbreviation. Then simple algebra yields

$$\mu(u_1, u_2; \rho) = \frac{u_1(1 - u_1)(1 - u_2)}{h(u_2) - \rho u_1(1 - u_2)}$$

and

$$\tilde{\mu}(u_1, u_2; \rho) = \frac{u_1 u_2(1 - u_1)(1 - u_2)}{h(u_2)^2 - u_2 \{h(u_2) - \rho u_1(1 - u_2)\}}.$$

Then after some algebra, one can show that

$$\mu_2(u_1, u_2; \rho) = -\frac{u_1(1 - u_1)}{\{h(u_2) - \rho u_1(1 - u_2)\}^2} < 0$$

and

$$\mu_2(u_1, u_2; \rho) = \frac{u_1(1 - u_1)(1 - u_2)^2 \{1 - \rho(1 - u_1^2)\}}{\{h(u_2)^2 - u_2 \{h(u_2) - \rho u_1(1 - u_2)\}\}^2} > 0$$

for $(u_1, u_2) \in (0, 1)^2$ and $\rho \in [-1, 1)$, since $1 - \rho(1 - u_1^2)$ is bounded from below by $1 - (1 - u_1^2) = u_1^2 > 0$. This verifies (4.8) and (4.9). □

A.3.2. The Farlie–Gumbel–Morgenstern family (Example A.4)

Simple algebra yields

$$\mu(u_1, u_2; \rho) = \frac{u_1(1 - u_1)(1 - u_2)}{1 + \rho(1 - 2u_1)(1 - u_2)}$$

and

$$\tilde{\mu}(u_1, u_2; \rho) = \frac{u_1(1 - u_1)(1 - u_2)}{1/u_2 - 1 - \rho(1 - 2u_1)(1 - u_2)}.$$

Then after some algebra, one can easily show that

$$\mu_2(u_1, u_2; \rho) = -\frac{u_1(1 - u_1)}{\{1 + \rho(1 - 2u_1)(1 - u_2)\}^2} < 0$$

and

$$\tilde{\mu}_2(u_1, u_2; \rho) = \frac{u_1(1 - u_1) \{(u_2 - 1)/u_2\}^2}{\{1/u_2 - 1 - \rho(1 - 2u_1)(1 - u_2)\}^2} > 0$$

for $(u_1, u_2) \in (0, 1)^2$ and $\rho \in [-1, 1]$, which verifies (4.8) and (4.9). □

A.3.3. The normal copula

With the normal copula, we show that the expressions in condition (4.8) and (4.9) have nice interpretable forms by themselves. The following proposition provides an interesting and useful result.

Proposition A.1 (Plackett, 1954). Let $f(\tilde{u}_1, \tilde{u}_2; \rho)$ be a normal density function with a correlation coefficient ρ . Then $\partial f(\tilde{u}_1, \tilde{u}_2; \rho) / \partial \rho = \partial^2 f(\tilde{u}_1, \tilde{u}_2; \rho) / \partial \tilde{u}_1 \partial \tilde{u}_2$.

Denote $\tilde{U}_1 \equiv \Phi^{-1}(U_1)$ and $\tilde{U}_2 \equiv \Phi^{-1}(U_2)$. Let $\phi(\cdot, \cdot; \rho), \phi(\cdot | \cdot; \rho)$, and $\phi(\cdot)$ be the bivariate, conditional, and marginal standard normal density functions, respectively. By Proposition A.1, it follows that

$$\begin{aligned} \mu(u_1, u_2; \rho) &= \frac{\phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho)}{\frac{1}{\phi(\Phi^{-1}(u_1))} \Phi_1(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho)} \\ &= \frac{\phi(\tilde{u}_1, \tilde{u}_2; \rho)}{\Phi(\tilde{u}_2 | \tilde{U}_1 = \tilde{u}_1; \rho)} \\ &= \phi(\tilde{u}_1) \lambda(\tilde{u}_2 | \tilde{U}_1 = \tilde{u}_1; \rho), \end{aligned}$$

and

$$\begin{aligned} \tilde{\mu}(u_1, u_2; \rho) &= \frac{\phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho)}{1 - \frac{1}{\phi(\Phi^{-1}(u_1))} \Phi_1(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho)} \\ &= \frac{\phi(\tilde{u}_1, \tilde{u}_2; \rho)}{1 - \Phi(\tilde{u}_2 | \tilde{U}_1 = \tilde{u}_1; \rho) \phi(\tilde{u}_1)} \\ &= \phi(\tilde{u}_1) \tilde{\lambda}(\tilde{u}_2 | \tilde{U}_1 = \tilde{u}_1; \rho), \end{aligned}$$

where $\lambda(\tilde{u}_2 | \tilde{U}_1 = \tilde{u}_1; \rho) = \frac{\phi(\tilde{u}_2 | \tilde{U}_1 = \tilde{u}_1; \rho)}{\Phi(\tilde{u}_2 | \tilde{U}_1 = \tilde{u}_1; \rho)}$ and $\tilde{\lambda}(\tilde{u}_2 | \tilde{U}_1 = \tilde{u}_1; \rho) = \frac{\phi(\tilde{u}_2 | \tilde{U}_1 = \tilde{u}_1; \rho)}{1 - \Phi(\tilde{u}_2 | \tilde{U}_1 = \tilde{u}_1; \rho)}$ are the standard (conditional) inverse Mill's ratios. □

It is well-known that $\lambda(\tilde{u}_2 | \tilde{U}_1 = \tilde{u}_1; \rho)$ is strictly decreasing and $\tilde{\lambda}(\tilde{u}_2 | \tilde{U}_1 = \tilde{u}_1; \rho)$ strictly increasing in \tilde{u}_2 (and hence in u_2). Therefore (4.8) and (4.9) automatically hold, which is in line with the discussion in Section 3 that the normal copula satisfies their sufficient condition, i.e., Assumption 5.

A.4. Jacobian matrix of $G(\theta)$ and $G^*(\theta)$

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 ρ , respectively. The Jacobian matrix $J_G(\theta) = \frac{\partial G(\theta)}{\partial \theta'}$ has the following expression:

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

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

$$\begin{aligned} E_1 \cdot J_G(\theta) \cdot E_2 &= \begin{bmatrix} 0 & 0 & \frac{C_\rho(b_1, a_0; \rho)}{C_1(b_1, a_0; \rho)} - \frac{C_\rho(b_1, a_1; \rho)}{C_1(b_1, a_1; \rho)} \\ 0 & 0 & \frac{C_\rho(b_0, a_1; \rho)}{1 - C_1(b_0, a_1; \rho)} - \frac{C_\rho(b_0, a_0; \rho)}{1 - C_1(b_0, a_0; \rho)} \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \\ E_1 &\equiv \begin{bmatrix} \frac{1}{C_1(b_1, a_0; \rho)} & -\frac{1}{C_1(b_1, a_1; \rho)} & 0 & 0 \\ 0 & 0 & \frac{1}{1 - C_1(b_0, a_0; \rho)} & -\frac{1}{1 - C_1(b_0, a_1; \rho)} \\ 0 & \frac{1}{C_1(b_1, a_1; \rho)} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{1 - C_1(b_0, a_1; \rho)} \end{bmatrix} \\ E_2 &\equiv \begin{bmatrix} 1 & 0 & \frac{C_\rho(b_0, a_1; \rho)}{1 - C_1(b_0, a_1; \rho)} \\ 0 & 1 & -\frac{C_\rho(b_1, a_1; \rho)}{C_1(b_1, a_1; \rho)} \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Now we prove that, given (4.8), the Jacobian matrix $J_{G^*} = J_{G^*}(\theta) = \frac{\partial G^*(\theta)}{\partial \theta'}$ is positive semi-definite and has full rank for all $\theta \in \Theta$ such that $a_0 \neq a_1$. Note that J_{G^*} equals $J_G(\theta)$ above with the last row

dropped. We show that the k th leading principal minor M_k of J_{G^*} is non-negative for all $1 \leq k \leq 5$:¹⁹ We have $M_1 = M_2 = 0$ and

$$\begin{aligned} M_3 &= \{1 - C_1(b_0, a_0; \rho)\} \{C_1(b_1, a_0; \rho) C_\rho(b_1, a_1; \rho) \\ &\quad - C_1(b_1, a_1; \rho) C_\rho(b_1, a_0; \rho)\} \\ &= \{1 - C_1(b_0, a_0; \rho)\} C_1(b_1, a_0; \rho) C_1(b_1, a_1; \rho) \\ &\quad \times \left\{ \frac{C_\rho(b_1, a_1; \rho)}{C_1(b_1, a_1; \rho)} - \frac{C_\rho(b_1, a_0; \rho)}{C_1(b_1, a_0; \rho)} \right\} \end{aligned}$$

is positive for $a_0 > a_1$ and negative for $a_0 < a_1$ by (4.8) and the fact that $C_1(u_1, u_2; \rho) > 0$ and $1 - C_1(u_1, u_2; \rho) > 0$ for all $(u_1, u_2) \in (0, 1)^2$ and $\rho \in \Omega$. Moreover, since M_3 is nonzero for $a_0 \neq a_1$, J_{G^*} has full rank.

One can similarly show that all other possible choices of G^* will also yield J_{G^*} that has full rank and is either positive or negative semi-definite. We omit the proof here.

References

- Altonji, J.G., Elder, T.E., Taber, C.R., 2005. An evaluation of instrumental variable strategies for estimating the effects of catholic schooling. *J. Hum. Resour.* 40 (4), 791–821.
- Andrews, D.W.K., Guggenberger, P., 2015. Identification- and Singularity-Robust Inference for Moment Condition Models, Cowles Foundation Discussion Paper No. 1978.
- Andrews, D.W.K., Stock, J.H., 2007. Inference with weak instruments. In: *Advances in Econometrics: Proceedings of the Ninth World Congress of the Econometric Society*.
- Andrews, I., Mikusheva, A., 2016a. Conditional inference with a functional nuisance parameter. *Econometrica* 84, 1571–1612.
- Andrews, I., Mikusheva, A., 2016b. A geometric approach to nonlinear econometric models. *Econometrica* 84, 1249–1264.
- Arnold, V.I., 2009. *Collected Works: Representations of Functions, Celestial Mechanics and KAM Theory, 1957–1965*. Springer, Berlin, Heidelberg.
- Bhattacharya, J., Goldman, D., McCaffrey, D., 2006. Estimating probit models with self-selected treatments. *Stat. Med.* 25 (3), 389–413.
- Chen, X., Fan, Y., Tsyrennikov, V., 2006. Efficient estimation of semiparametric multivariate copula models. *J. Amer. Statist. Assoc.* 101 (475), 1228–1240.
- Chen, X., Linton, O., Van Keilegom, I., 2003. Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* 71 (5), 1591–1608.
- Chernozhukov, V., Hansen, C., 2005. An IV model of quantile treatment effects. *Econometrica* 73 (1), 245–261.
- Chiburis, R., 2010. Semiparametric bounds on treatment effects. *J. Econometrics* 159 (2), 267–275.
- Evans, W.N., Schwab, R.M., 1995. Finishing high school and starting college: Do Catholic schools make a difference? *Q. J. Econ.* 110 (4), 941–974.
- Fan, Y., Liu, R., 2015. Partial Identification and Inference in Censored Quantile Regression: A Sensitivity Analysis. Working Paper, University of Washington and Emory University.
- Fan, Y., Park, S.S., 2010. Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory* 26 (03), 931–951.
- Fan, Y., Wu, J., 2010. Partial identification of the distribution of treatment effects in switching regime models and its confidence sets. *Rev. Econom. Stud.* 77 (3), 1002–1041.
- Freedman, D.A., Sekhon, J.S., 2010. Endogeneity in probit response models. *Polit. Anal.* 18 (2), 138–150.
- Gale, D., Nikaido, H., 1965. The Jacobian matrix and global univalence of mappings. *Math. Ann.* 159 (2), 81–93.
- Goldman, D., Bhattacharya, J., McCaffrey, D., Duan, N., Leibowitz, A., Joyce, G., Morton, S., 2001. Effect of insurance on mortality in an HIV-positive population in care. *J. Amer. Statist. Assoc.* 96 (455).
- Hadamard, J., 1906a. Sur les transformations planes. *C. R. Seances Acad. Sci. Paris* 74, 142.
- Hadamard, J., 1906b. Sur les transformations ponctuelles. *Bull. Soc. Math. France* 34, 71–84.
- Han, S., Lee, S., 2017. Sensitivity Analysis in Triangular Systems of Equations with Binary Endogenous Variables. Working Paper, University of Texas at Austin.
- Han, S., McCloskey, A., 2017. Estimation and Inference with a (Nearly) Singular Jacobian. Working Paper, University of Texas at Austin and Brown University.
- Heckman, J., 1978. Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46, 931–959.
- Heckman, J.J., 1979. Sample selection bias as a specification error. *Econometrica* 47 (1), 153–162.
- Joe, H., 1997. *Multivariate Models and Multivariate Dependence Concepts*, In: Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis.
- Kleibergen, F., 2005. Testing parameters in GMM without assuming that they are identified. *Econometrica* 73, 1103–1123.
- Lee, L.-F., 1983. Generalized econometric models with selectivity. *Econometrica* 51 (2), 507–512.
- Manski, C.F., 1988. Identification of binary response models. *J. Amer. Statist. Assoc.* 83 (403), 729–738.
- Marra, G., Radice, R., 2011. Estimation of a semiparametric recursive bivariate probit model in the presence of endogeneity. *Canad. J. Statist.* 39 (2), 259–279.
- Meango, R., Mourifié, I., 2014. A note on the identification in two equations probit model with dummy endogenous regressor. *Econom. Lett.* 125, 360–363.
- Neal, D.A., 1997. The effects of catholic secondary schooling on educational achievement. *J. Labor Econ.* 15 (1, Part 1), 98–123.
- Nelsen, R.B., 1999. *An Introduction to Copulas*. Springer Verlag.
- Plackett, R., 1954. A reduction formula for normal multivariate integrals. *Biometrika* 41, 351–360.
- Radice, R., Marra, G., Wojtyś, M., 2015. Copula regression spline models for binary outcomes. *Stat. Comput.* 1–15.
- Rhine, S.L., Greene, W.H., Toussaint-Comeau, M., 2006. The importance of check-cashing businesses to the unbanked: Racial/ethnic differences. *Rev. Econ. Statist.* 88 (1), 146–157.
- Rothenberg, T.J., 1971. Identification in parametric models. *Econometrica* 577–591.
- Rudin, W., 1986. *Real and Complex Analysis*. third ed. McGraw-Hill, New York.
- Stock, J.H., Wright, J.H., 2000. GMM with Weak Identification. *Econometrica* 68, 1055–1096.
- Trivedi, P., Zimmer, D., 2007. *Copula Modeling: An Introduction for Practitioners*. In: *Foundations and Trends and Econometrics*, vol. 1, pp. 1–111.
- Wilde, J., 2000. Identification of multiple equation probit models with endogenous dummy regressors. *Econom. Lett.* 69 (3), 309–312.
- Winkelmann, R., 2012. Copula bivariate probit models: with an application to medical expenditures. *Health Econ.* 21 (12), 1444–1455.

¹⁹ The k th leading principal minor of J^* is the determinant of its upper-left $k \times k$ sub-matrix.