Estimating Causal Effects of Discrete and Continuous Treatments with Binary Instruments

Victor Chernozhukov Iván Fernández-Val

Sukjin Han Kaspar Wüthrich

March 20, 2024

Abstract

We propose an instrumental variable framework for identifying and estimating average and quantile effects of discrete and continuous treatments with binary instruments. The basis of our approach is a local copula representation of the joint distribution of the potential outcomes and unobservables determining treatment assignment. This representation allows us to introduce an identifying assumption, so-called *copula invariance*, that restricts the local dependence of the copula with respect to the treatment propensity. We show that copula invariance identifies treatment effects for the entire population and other subpopulations such as the treated. The identification results are constructive and lead to straightforward semiparametric estimation procedures based on distribution regression. An application to the effect of sleep on well-being uncovers interesting patterns of heterogeneity.

JEL Numbers: C14, C21, C31.

Keywords: Quantile treatment effects, endogeneity, binary instruments, copula.

*The authors respectively represent MIT, BU, U of Bristol, and UCSD. We are grateful to Dalia Ghanem and Desire Kedagni for comments.

1 Introduction

Endogeneity and heterogeneity are key challenges in causal inference. Endogeneity arises because most treatments and policies of interest are the result of decisions made by economic agents. Heterogeneity also arises naturally as many of the agents' characteristics are unobserved to the researcher. Accounting for endogeneity and heterogeneity in treatment effects is crucial to answer policy questions, such as how to allocate social resources and combating inequalities. This paper contributes to the literature by proposing a flexible instrumental variable (IV) modeling framework for identifying heterogeneous treatment effects under endogeneity, which yields straightforward semiparametric estimation and inference procedures.

Without additional assumptions, IV strategies cannot point-identify meaningful treatment effects. The literature has proposed different solutions to deal with this challenge that exhibit trade-offs between adding structure to the treatment assignment mechanism and potential outcomes. One line of research has restricted the structure and heterogeneity of the potential outcomes while allowing for flexible treatment assignment mechanisms (e.g., Chernozhukov and Hansen (2005) with binary treatment and Newey and Powell (2003) with continuous treatment). Another line of research has shown the usefulness of restricting the treatment assignment while being flexible regarding how the potential outcomes are formed (e.g., Imbens and Angrist (1994) with binary treatment and Imbens and Newey (2009) with continuous treatment). There are also partial identification solutions that impose less structure on the treatment assignment and potential outcomes (e.g., Manski (1990) and Balke and Pearl (1997) for earlier references, and Chesher and Rosen (2020) for a more recent survey).

We explore an intermediate route that imposes structure on the relationship between the treatment assignment and potential outcomes to achieve point identification of meaningful heterogeneous treatment effects. The basis of this approach is a local Gaussian representation of the copula of the potential outcomes and unobservable determinants of treatment assignment. We emphasize that this representation is fully *nonparametric*, that is, it does not require that potential outcomes and treatment unobservables are jointly or marginally

Gaussian (see Chernozhukov et al., 2020a). Indeed, the bivariate Gaussian structure always holds locally by treating the correlation parameter as an implicit function that equates the bivariate Gaussian distribution with the copula. We use this representation to introduce an assumption that has not been previously considered for identification of treatment effects. This assumption, so-called *copula invariance* (CI), restricts the local dependence of the copula with respect to treatment propensity. We show that, even with a binary IV, copula invariance identifies quantile and average treatment effects (QTE and ATE) of binary and ordered treatments and quantile and average structural functions (QSF and ASF) of continuous treatments for the entire population and other subpopulations such as the treated. The same approach applies without modification to continuous, discrete and mixed continuous-discrete outcomes. When covariates are available, we impose CI conditional on these covariates, allowing for an additional source of heterogeneity in our model.

Our framework is also useful to compare the assumptions of different strategies to identify treatment effects with endogeneity. For example, our approach imposes more restrictions on the dependence structure (i.e., the form of endogeneity), while allowing for richer patterns of effect heterogeneity, compared to Chernozhukov and Hansen (2005) and Newey and Powell (2003), or more heterogeneity of the treatment assignment, compared to Imbens and Angrist (1994) and Imbens and Newey (2009).[1] In this sense, we expand the directions of modeling trade-offs. Another attractive feature of our identification strategy is that it is constructive and leads to straightforward semiparametric estimation procedures based on distribution regression for both discrete and continuous treatments.

We apply the proposed method to estimating the distributional effects of sleep on well-being. In this case, sleep time is treated as a continuous treatment. We use the data from the experimental analysis of Bessone et al. (2021b), who studied the effects of randomized interventions to increase sleep time of low-income adults in India. A simple two-stage least squares analysis suggests that sleep has moderate or insignificant average effects on well-being. Using our method, we document interesting patterns of heterogeneity across the

[1] We provide a more detailed comparison with the existing strategies below.

distributions of sleep time and well-being that standard analyses focusing on average effects miss.

1.1 Related Literature

The literature on identification of heterogeneous treatment effects with endogeneity is vast. We focus the review on approaches that do not impose distributional assumptions to achieve point identification.

A first strand of literature focused on imposing assumptions in the generation of the potential outcomes, such as rank invariance and rank similarity. Rank invariance imposes that the outcome equation is strictly monotonic in a scalar unobservable such that there is a one-to-one mapping between the potential outcome and unobservable, and the unobservable is the same for all the potential outcomes. This assumption is very convenient for the identification analysis. It allows for identification of QTE and ATE with discrete treatments (Chernozhukov and Hansen (2005)) and identification of the ASF with continuous treatments (Newey and Powell (2003), Blundell et al. (2007)). Chernozhukov and Hansen (2005)'s rank similarity is slightly weaker than rank similarity as it does not necessarily restrict the unobservable to be the same for all the potential outcomes. However, both assumptions produce the same testable restriction (Chernozhukov and Hansen (2013)). In subsequent work, Vuong and Xu (2017) showed that rank invariance and strict monotonicity are powerful enough to identify individual treatment effects (under suitable regularity conditions) in addition to the QTE and ATE. This literature remains flexible about the treatment selection process. Our CI assumption and rank similarity or invariance are non-nested. The former concerns the dependence between potential outcomes and selection unobservables, whereas the latter concerns the dependence between potential outcomes. As shown below, CI allows for more general patterns of treatment effect heterogeneity than rank invariance and rank similarity. Also, rank invariance and similarity rely on strict monotonicity on the outcome equation to achieve point identification. This assumption can only hold for continuous outcomes.

Moreover, these approaches rely on completeness conditions on the relationship between the treatment and instrument that rule out, for example, an ordered and continuous treatment when the instrument is binary. CI does not rely on monotonicity nor completeness and therefore can accommodate discrete and mixed discrete-continuous outcomes and ordered and continuous treatments with a binary instrument.

A second strand of literature focuses on assumptions imposed on the treatment assignment. Imbens and Angrist (1994) and Heckman and Vytlacil (2005) assumed that treatment assignment is determined by a scalar unobservable and combined this assumption with monotonicity of potential treatments with respect to a binary instrument to show identification of local and marginal effects of discrete treatments. Abadie et al. (2002) and Carneiro and Lee (2009) extended this approach to the corresponding quantile effects. Newey et al. (1999) and Imbens and Newey (2009) used strict monotonicity of the treatment selection equation with respect to a scalar unobservable and large support of the instrument (ruling out discrete instruments) to identify global treatment effects using a control variable approach. Compared to this strand of the literature, CI restricts the relationship between potential outcomes and treatment assignment, but allows for identifying global treatment effects of discrete and continuous treatments with discrete instruments without relying on monotonicity assumptions on the treatment assignment mechanism.

There are also approaches that combine or modify the assumptions of the previous strands. Chesher (2003) showed identification of the quantile effect of continuous treatments on continuous outcomes with continuous instruments assuming strict monotonicity of the outcome and treatment selection equations with respect to scalar unobservables. Under similar assumptions, D'Haultfoeuille and Février (2015) and Torgovitsky (2015) found that quantile effects can be identified with discrete instruments. Newey and Stouli (2021) avoided the large support requirement on the instrument of Imbens and Newey (2009) by assuming a parametric structure on the expectation of the treatment conditional on the instrument that enables extrapolation outside the instrument support. Again, these restrictions are different and not nested with CI. In Appendix A, we provide a more detailed comparison of CI with

these and other approaches.

The copula is a powerful tool that has been previously employed in econometrics for identification and estimation. For example, Chen et al. (2006) used a parametric copula to achieve efficient estimation in a class of multivariate distributions. In a semiparametric triangular model with binary dependent variables, Han and Vytlacil (2017) introduced a class of single-parameter copulas to model the dependence structure between the unobservables and established a condition on the copula under which the parameters are identified. They showed that many well-known copulas including the Gaussian copula satisfy the condition. When we restrict our attention to a binary treatment and binary outcome, the current paper's framework is relevant to Han and Vytlacil (2017). However, while they assumed a parametric copula for the dependence structure, we assume CI. Assuming a Gaussian copula in Han and Vytlacil (2017) can be viewed as an extreme special case of CI. Han and Lee (2019) developed sieve estimation and inference methods based on Han and Vytlacil (2017), and Han and Lee (2023) extended them to semiparametric models for dynamic treatment effects. Mourifié and Wan (2021) use copula as a channel to impose assumptions in characterizing identified sets in the framework of marginal treatment effects.

Arellano and Bonhomme (2017) and Chernozhukov et al. (2020a) studied IV identification of selection models using assumptions on the copula between the latent outcome and selection unobservable. Arellano and Bonhomme (2017) assumed real analytical copula and continuous instrument. Chernozhukov et al. (2020a) used the local Gaussian representation and copula exclusion, which is a special case of CI, with a binary instrument. Therefore, our framework with binary treatment is closely related to their setup. However, even in the case of binary treatment, the current setting differs from Chernozhukov et al. (2020a) in several dimensions. First, our setting requires two-way sample selection due to the switching of treatment status. Second, we introduce a general selection model that does not follow the typical threshold-crossing structure, which is important to allow for rich selection patterns. Third, because of these features, the identification analysis involves local representation and copula invariance that are specific to treatment status and the value of the IV. More importantly, the use of

local representation and CI for ordered and continuous treatments are completely new to this paper. The identification strategies in these two cases are distinct from the binary case although they rely on the same CI assumption. Finally, in the difference-in-differences setup, Athey and Imbens (2006) showed that average and quantile treatment effects on the treated (and the untreated) can be identified when the unobservable determinant of the untreated potential outcome is independent of time within groups. Ghanem et al. (2023) provide general identification results under a time invariance assumption on the copula between the potential outcomes and the group indicator and show that their assumption is equivalent to the assumptions in Athey and Imbens (2006) with continuous outcomes.

1.2 Organization of the Paper

Section 2.1 introduce key variables and parameters of interest, Section 2.2 introduces the local Gaussian representation, and Section 2.3 posits the main identifying assumptions that will be used throughout the analyses. We devote Sections 3.1–3.3 to the identification analyses with binary, ordered, and continuous treatments, respectively. Section 4 discusses copula invariance in further detail (e.g., by providing sufficient conditions) and Section 5 discusses estimation and inference. Section 6 provides the empirical application. The Appendix is organized as follows. Appendix A compares our identification approach with those in the previous literature. Appendix B shows how copulas other than Gaussian can also be used for local representation. Appendices C and D extend the identification analysis of the main text to understand the role of the support of IVs and of the presence of covariates. Appendix E presents alternative identification strategies with variants of copula invariance. Appendix G contains proofs.

1.3 Notation

For scalar random variables X and Y and possibly multivariate random variable Z, $F_{X,Y|Z}$ denotes the joint distribution of X and Y conditional on Z, $F_{X|Z}$ denotes the (marginal)

distribution of X conditional on Z, and F_Z denotes the marginal (joint) distribution of Z. We use calligraphic letters to denote support sets of random variables. For example, $\mathcal{Z}$ denotes the support of Z. The symbol $\perp\!\!\!\perp$ denotes (stochastic) independence; for example, $X \perp\!\!\!\perp Y$ means that X is independent of Y. The interior of the set $\mathcal{D}$ is denoted as $int(\mathcal{D})$.

2 Setup and Assumptions

We consider three classes of models depending on the type of treatment variable: binary, ordered or continuous. Before investigating identification, we introduce the setup, parameters of interest and identifying assumptions that are common to all classes of models.

2.1 Preliminaries

Let $Y \in \mathcal{Y} \subseteq \mathbb{R}$ denote the scalar outcome and $D \in \mathcal{D} \subseteq \mathbb{R}$ denote the scalar treatment. We consider binary, ordered and continuous treatments with $\mathcal{D} = \{0, 1\}$, $\mathcal{D} = \{1, ..., K\}$ and $\mathcal{D}$ equal to an uncountable set, respectively. The outcome is not restricted, it can be continuous, discrete or mixed continuous-discrete. Let $Z \in \{0, 1\}$ be the binary IV. We focus on a binary instrument as the most challenging case; the analysis readily extends to discrete or continuous Z. Let Y_d denote the potential outcome given $d \in \mathcal{D}$ and D_z the potential treatment given $z \in \{0, 1\}$. They are related to the observed outcome and treatment through $Y = Y_D$ and $D = D_Z$. All the identification analysis is conditional on a vector of covariates $X \in \mathcal{X} \subseteq \mathbb{R}^{d_x}$ for some positive integer d_x. We explicitly incorporate X when we discuss estimation and inference in Section 5.

We consider a general treatment assignment equation:

$$D_z = h(z, V_z), \tag{2.1}$$

where we normalize $V_z \sim U[0, 1]$. We provide examples of the function h for each type of treatment below. By allowing for a different unobservable V_z at each value of z, we essentially

permit D to be a function of the *vector* of unobservables (V_0, V_1). Even this general version of a treatment assignment model is not necessary for our analyses but simplifies the exposition; see Appendix F.

We are interested in identifying the distribution of Y_d, F_{Y_d}, for $d \in \mathcal{D}$, and functionals of F_{Y_d}, such as quantile and average structural functions. Thus, by using appropriate operators:

$$QSF_\tau(d) \equiv Q_{Y_d}(\tau) = \mathcal{Q}_\tau(F_{Y_d}),$$
$$ASF(d) \equiv E[Y_d] = \mathcal{E}(F_{Y_d}),$$

where $\mathcal{Q}_\tau(F) \equiv \inf\{y \in \mathcal{Y} : F(y) \geq \tau\}$ and $\mathcal{E}(F) \equiv \int_{\mathcal{Y}}[1 - F(y)]dy$. Quantile and average effects can be expressed as $QSF_\tau(d) - QSF_\tau(d')$ and $ASF(d) - ASF(d')$ for a binary or ordered treatment and $\partial QSF_\tau(d)/\partial d$ and $\partial ASF_\tau(d)/\partial d$ for a continuous treatment. When the treatment is binary, we may also be interested in the distribution of Y_d in subpopulations such as the treated, $F_{Y_d|D}(\cdot \mid 1)$, and untreated, $F_{Y_d|D}(\cdot \mid 0)$, for $d \in \{0, 1\}$, and functionals of these distributions.

2.2 Local Gaussian Representation

Treatment endogeneity can be captured by the joint distribution of the potential outcome and unobservable of the treatment assignment equation (2.1). We use a conditional version of the local Gaussian representation (LGR) to represent such a joint distribution. This representation is the basis of our identification and estimation strategies. Throughout the paper, let $C(u_1, u_2; \rho)$ denote the Gaussian copula with correlation coefficient ρ, that is

$$C(u_1, u_2; \rho) = \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_1); \rho),$$

where $\Phi_2(\cdot, \cdot; \rho)$ is the standard bivariate Gaussian distribution with parameter ρ and Φ is the standard univariate Gaussian distribution.

The following lemma shows that the conditional copula of any bivariate random variable

has a local Gaussian representation (Chernozhukov et al., 2020a).

Lemma 2.1 (LGR)**.** *For any random variables Y, V and Z, the joint distribution of Y and V conditional on Z admits the representation:*

$$F_{Y,V|Z}(y,v \mid z) = C(F_{Y|Z}(y \mid z), F_{V|Z}(v \mid z); \rho_{Y,V;Z}(y,v;z)), \text{ for all } (y,v,z),$$

where $\rho_{Y,V;Z}(y,v;z)$ is the unique solution in ρ to

$$F_{Y,V|Z}(y,v \mid z) = C(F_{Y|Z}(y \mid z), F_{V|Z}(v \mid z); \rho).$$

In the lemma, note that the solution $\rho_{Y,V;Z}(y,v;z)$ depends on both the dependence structure and the marginals. This distinction becomes important later. Lemma 2.1 can be equivalently stated as the LGR of a copula instead of a distribution; see Section 4.3. Note that Gaussianity is not essential for the local representation. In Appendix B, we provide other copulas that can be used for the representation. Gaussianity, however, is convenient to introduce identifying assumptions, interpret these assumptions using joint normality as a benchmark of comparison, and to develop semiparamatric estimators.

2.3 Assumptions

We maintain the following assumptions:

Assumption EX (Independence)**.** *For $d \in \mathcal{D}$ and $z \in \{0,1\}$, $Z \perp\!\!\!\perp Y_d$ and $Z \perp\!\!\!\perp V_z$.*

Assumption REL (Relevance)**.** *(i) $Z \in \{0,1\}$; (ii) $0 < \Pr(Z=1) < 1$; and (iii) for $\mathcal{D} = \{0,1\}$, $\Pr(D=1 \mid Z=1) \neq \Pr(D=1 \mid Z=0)$ and $0 < \Pr(D=1 \mid Z=z) < 1$, $z \in \{0,1\}$; for $\mathcal{D} = \{1,...,K\}$, $\Pr(D=d \mid Z=z) > 0$, $(z,d) \in \{0,1\} \times \mathcal{D}$, $F_{D|Z}(1 \mid 1) \neq F_{D|Z}(1 \mid 0)$, and $F_{D|Z}(d \mid 1) \neq F_{D|Z}(d \mid 0)$ or $F_{D|Z}(d-1 \mid 1) \neq F_{D|Z}(d-1 \mid 0)$, $d \in \{2,...,K\}$; and for uncountable $\mathcal{D}$, $F_{D|Z}(d \mid 1) \neq F_{D|Z}(d \mid 0)$ and $0 < F_{D|Z}(d \mid z) < 1$ for $(z,d) \in \{0,1\} \times int(\mathcal{D})$.*

EX is standard in IV strategies. It is weaker than $Z \perp\!\!\!\perp (\{Y_d\}_{d\in\mathcal{D}}, V_0, V_1)$ or $Z \perp\!\!\!\perp (Y_d, V_z)$ for $(d, z) \in \mathcal{D} \times \{0, 1\}$. Also, in EX, a standard exclusion restriction is implicit in the notation: $Y_d = Y_{d,z}$ almost surely, where $Y_{d,z}$ is the potential outcome given (d, z). REL(ii)–(iii) are the usual relevance condition for the IV and the boundary condition. REL(iii) for $\mathcal{D} = \{0, 1\}$ can be formulated a special case of REL(iii) for $\mathcal{D} = \{1, ..., K\}$ with $K = 2$, but we state it separately for clarity. When D is ordered, REL(iii) imposes that $F_{D|Z}(K - 1 \mid 1) \neq F_{D|Z}(K - 1 \mid 0)$ because $F_{D|Z}(K \mid 1) = F_{D|Z}(K \mid 0) = 1$, but allows for $F_{D|Z}(d \mid 1) = F_{D|Z}(d \mid 0)$ for some treatment levels d when $K \geq 4$. For example, when $K = 3$ this condition requires $F_{D|Z}(d \mid 1) \neq F_{D|Z}(d \mid 0)$ for $d \in \{1, 2\}$, but when $K = 4$ this condition allows for $F_{D|Z}(2 \mid 1) = F_{D|Z}(2 \mid 0)$. A sufficient relevance condition for ordered D is $F_{D|Z}(d \mid 1) \neq F_{D|Z}(d \mid 0)$ for $d \in \mathcal{D} \setminus \{K\}$.

We make the following assumption about the local dependence parameter of the LGR of (Y_d, V_z) conditional on Z:

Assumption CI (Copula Invariance)**.** *For $d \in \mathcal{D}$, $\rho_{Y_d,V_z;Z}(y, v; z)$ is a constant function of (v, z), that is*

$$\rho_{Y_d,V_z;Z}(y, v; z) = \rho_{Y_d}(y), \quad (y, v, z) \in \mathcal{Y} \times \mathcal{V} \times \{0, 1\}$$

and $\rho_{Y_d}(y) \in (-1, 1)$.

CI is a high-level condition that (together with the other assumptions we maintain) is sufficient for identification in all the cases that we consider, but it is not necessary. We provide weaker conditions for each case in the following section. Section 4 provides more interpretable conditions for CI and compares CI with alternative identifying assumptions that have been used in the literature such as rank invariance and rank similarity. The condition $\rho_{Y_d}(y) \in (-1, 1)$ rules out boundary cases, which can be dealt with as in Chernozhukov et al. (2020a).

3 Identification Analysis

3.1 Binary Treatment

We start by considering the identification of the causal effects of a binary treatment $D \in \mathcal{D} = \{0, 1\}$. To reflect this, we consider a treatment selection equation

$$D_z = h(z, V_z) = 1\{V_z \leq \pi(z)\}, \tag{3.1}$$

with propensity score

$$\Pr[D = 1 \mid Z = z] = \Pr[D_z = 1 \mid Z = z] = \Pr[V_z \leq \pi(z)] = \pi(z),$$

by EX and the normalization $V_z \sim U[0, 1]$. Note that V_1 and V_0 are two distinct unobservables that do not restrict the behavior of D_0 and D_1. The LATE monotonicity assumption of Imbens and Angrist (1994), imposes either $D_1 \geq D_0$ or $D_0 \geq D_1$ almost surely, which corresponds to $V_1 = V_0$ almost surely (Vytlacil, 2002).

For the identification analysis, consider

$$\begin{aligned}\Pr[Y \leq y, D = 1 \mid Z = z] &= \Pr[Y_1 \leq y, D_z = 1 \mid Z = z] \\ &= C(F_{Y_1|Z}(y|z), \pi(z); \rho_{Y_1,V_z;Z}(y, \pi(z); z)) \\ &= C(F_{Y_1}(y), \pi(z); \rho_{Y_1,V_z;Z}(y, \pi(z); z)), \quad (y, z) \in \mathcal{Y} \times \{0, 1\}, \end{aligned} \tag{3.2}$$

where the second equality uses equation (3.1) and Lemma 2.1, and the last equality follows from EX. For each $y \in \mathcal{Y}$, this is a system of two equations in three unknowns, $F_{Y_1}(y)$, $\rho_{Y_1,V_z;Z}(y, \pi(0); 0)$ and $\rho_{Y_1,V_z;Z}(y, \pi(1); 1)$.

The following theorem establishes that the condition

$$\rho_{Y_1,V_1;Z}(y, \pi(1); 1) = \rho_{Y_1,V_0;Z}(y, \pi(0); 0) \equiv \rho_{Y_1}(y), \quad y \in \mathcal{Y}, \tag{3.3}$$

which is implied by CI, identifies the distribution of Y_1 and the local dependence parameter of the LGR of Y_1 and V_z. A similar argument shows that the distribution of Y_0 and the local dependence parameter of the LGR of Y_0 and V_z are identified.

Theorem 3.1 (Identification for Binary Treatment)**.** *Suppose $D_z \in \{0,1\}$ satisfies (3.1) for $z \in \{0,1\}$. Under EX, REL, and CI, the functions $y \mapsto F_{Y_d}(y)$ and $y \mapsto \rho_{Y_d}(y)$ are identified on $y \in \mathcal{Y}$, for $d \in \{0,1\}$.*

The proof of Theorem 3.1 in Appendix G shows that the nonlinear system of equations (3.2) has a unique solution. This result follows from a global univalence theorem of Gale and Nikaido (1965), because the Jacobian of the system of equations is a P-matrix.

Remark 3.1 (Identification of $F_{Y_1|D}$ and $F_{Y_0|D}$)**.** *We show identification for the treated, $D = 1$, identification for the untreated follows by a similar argument. The distribution of Y_1 is trivially identified from $F_{Y_1|D}(y \mid 1) = F_{Y|D}(y \mid 1)$. Identification of the distribution of Y_0 follows from*

$$F_{Y_0|D}(y \mid 1) = \frac{F_{Y_0}(y) - (1-\pi)F_{Y|D}(y \mid 0)}{\pi},$$

where $\pi \equiv \Pr[D = 1]$ and $F_{Y_0}(y)$ is identified by Theorem 3.1.

Remark 3.2 (Random Intention to Treat)**.** *Under random intention to treat or one-sided compliance, $D = 0$ whenever $Z = 0$ (i.e. $\Pr(D = 0 \mid Z = 0) = 1$), so that REL is violated. In this case $F_{Y_1}(y)$ is no longer identified because one of the equations in (3.2) becomes uninformative as $\pi(0) = 0$. We can still identify the distributions of the potential outcomes for the treated using the same analysis of Remark 3.1 because $F_{Y_0}(y) = F_{Y|Z}(y \mid 0)$. When we impose a stronger version of CI that $\rho_{Y_1}(y) = \rho_{Y_0}(y)$ (i.e., rank similarity), then we can also identify treatment effects for the entire population; see Appendix A.1 for details.*

3.2 Ordered Treatment

We consider identification of the causal effect of a multivalued ordered treatment $D \in \mathcal{D} = \{1, \ldots, K\}$ using a binary instrument $Z \in \{0,1\}$. There are now K potential outcomes

denoted as $(Y_1, \ldots, Y_K)$, which are related to the observed outcome as $Y = \sum_{d \in \mathcal{D}} Y_d 1\{D = d\}$. As before, we denote the potential treatments as (D_0, D_1).

We assume a threshold-crossing model for the treatment selection equation, which can be viewed as a natural extension of model (3.1) from two to multiple treatment levels,

$$D_z = h(z, V_z) = \begin{cases} 1, & \pi_0(z) < V_z \leq \pi_1(z) \\ 2, & \pi_1(z) < V_z \leq \pi_2(z) \\ \vdots & \vdots \\ K, & \pi_{K-1}(z) < V_z \leq \pi_K(z) \end{cases}, \tag{3.4}$$

where $\pi_0(z) = 0$ and $\pi_K(z) = 1$. Equation (3.4) generalizes the model in Section 7.2 of Heckman and Vytlacil (2007) by allowing for a different impact of the instrument on the different cutoffs; see Remark 3.3.

Under the normalization $V_z \sim U[0,1]$ and EX, the threshold functions $\pi_d(z)$ are identified by the distribution of the observed treatment conditional on the instrument as $\pi_d(z) = F_{D|Z}(d \mid z)$ for $d \in \mathcal{D}$. For the identification analysis, consider

$$\begin{aligned} \Pr[Y \leq y, D = d \mid Z = z] &= \Pr[Y_d \leq y, \pi_{d-1}(z) < V_z \leq \pi_d(z) \mid Z = z] \\ &= C(F_{Y_d}(y), \pi_d(z); \rho_{Y_d, V_z; Z}(y, \pi_d(z); z)) \\ &\quad - C(F_{Y_d}(y), \pi_{d-1}(z); \rho_{Y_k, V_z; Z}(y, \pi_{d-1}(z); z)), \ (y, d, z) \in \mathcal{Y} \times \mathcal{D} \times \{0, 1\}, \end{aligned} \tag{3.5}$$

where the first equality follows from (3.4) and the second equality from EX and Lemma 2.1. For each $d \in \mathcal{D}$ and $y \in \mathcal{Y}$, (3.5) is a system of two equations on five unknowns: $F_{Y_d}(y)$, $\rho_{Y_d, V_0; Z}(y, \pi_{d-1}(0); 0)$, $\rho_{Y_d, V_0; Z}(y, \pi_d(0); 0)$, $\rho_{Y_d, V_1; Z}(y, \pi_{d-1}(1); 1)$, and $\rho_{Y_d, V_1; Z}(y, \pi_d(1); 1)$. REL(iii) guarantees that the two equations of the system are not redundant.

For $d \in \{1, K\}$, one of the terms in the right hand side drops out because either $\pi_{d-1}(z) = 0$ or $\pi_d(z) = 1$, yielding a system of two equations on three unknowns. The distribution of

the potential outcome and local dependence parameter can be identified using the condition (3.3) from the binary treatment case and by REL(iii).

For $d \in \mathcal{D}\backslash\{1, K\}$, condition (3.3) reduces the number of unknowns to three but is not sufficient to identify the unknowns. We impose additionally copula invariance between consecutive treatment levels

$$\rho_{Y_d,V_z;Z}(y, \pi_d(z); z) = \rho_{Y_d,V_z;Z}(y, \pi_{d-1}(z); z) \equiv \rho_{Y_d}(y), \quad (y, z) \in \mathcal{Y} \times \{0, 1\}. \tag{3.6}$$

This condition is also implied by CI and reduces the number of unknowns to two: $F_{Y_d}(y)$ and $\rho_{Y_d}(y)$. The Jacobian of the resulting system of equations, however, does not satisfy the conditions to apply the global univalence results of Gale and Nikaido (1965) even under REL(iii). We show existence and uniqueness of solution using an alternative global univalence result of Ambrosetti and Prodi (1995).[2] To apply this result we make the following sufficient condition on the distribution of the treatment conditional on the instrument:

Assumption U$_{OC}$ (Uniformity in Ordered Choice)**.** *Either* $F_{D|Z}(d \mid 0) < F_{D|Z}(d \mid 1)$ *for all* $d \in \mathcal{D}\backslash\{K\}$ *or* $F_{D|Z}(d \mid 0) > F_{D|Z}(d \mid 1)$ *for all* $d \in \mathcal{D}\backslash\{K\}$.

U$_{OC}$ does not necessarily follow from REL(iii) and imposes the same ordering between $F_{D|Z}(d \mid 0)$ and $F_{D|Z}(d \mid 1)$ for all $d \in \mathcal{D}\backslash\{K\}$. Like REL(iii), U$_{OC}$ can be directly tested from the data. Remark 3.3 shows that Heckman and Vytlacil (2007)'s ordered choice model satisfies U$_{OC}$. It is interesting to see what type of compliance behavior with respect to D_0 and D_1 is ruled out by this sufficient condition. To explore this, define the compliers and defiers of order $j \in \mathcal{D}\backslash\{K\}$ as

$$C_j \equiv \bigcup_{d=1}^{K-j} \{D_0 = d, D_1 = d + j\},$$
$$B_j \equiv \bigcup_{d=1}^{K-j} \{D_1 = d, D_0 = d + j\}.$$

[2]This results was previously used to show identification by De Paula et al. (2019) in a different setting.

Assumption EG (Exchangeability). *V_0 and V_1 are exchangeable, i.e., $C(v_0, v_1) = C(v_1, v_0)$.*

Assumption EG states that the distribution for (V_0, V_1) is symmetric; most known copulas are symmetric. Under this assumption, we can interpret Assumption U_{OC} in terms of compliance behavior:

Lemma 3.1 (Compliance Shares). *Under Assumption EG, $\pi_d(1) > \pi_d(0)$ (resp. $<$) for all $d \in \mathcal{D}\backslash\{K\}$ implies that the share of all complier groups is smaller (resp. larger) than the share of all defier groups, that is, $\Pr[\bigcup_{j=1}^{K-1} C_j] < \Pr[\bigcup_{j=1}^{K-1} B_j]$ (resp. $>$).*

The condition about the share of compliers and defiers is reminiscent of a similar assumption used in De Chaisemartin (2017) in the case of binary treatment. Another simple interpretation of Lemma 3.1 can be made under the restriction $V_0 = V_1$. In this special case, we can easily see that there is no defiers (i.e., $\Pr[D_1 < D_0] = 0$) if and only if $\pi_d(1) < \pi_d(0)$ for all $k \in \mathcal{D}\backslash\{K\}$. In general, when V_z is not restricted, Assumption EG alone does not eliminate compliers or defiers. We summarize the identification result:

Theorem 3.2 (Identification for Ordered Treatment). *Suppose D_z, $z \in \{0, 1\}$, satisfies (3.4). Under EX, REL, CI, and U_{OC}, the functions $y \mapsto F_{Y_d}(y)$ and $y \mapsto \rho_{Y_d}(y)$ are identified on $y \in \mathcal{Y}$, for $d \in \mathcal{D}$.*

The proof of Theorem 3.2 in Appendix G does not follow from the same argument as the proof of Theorem 3.1. As mentioned above, we cannot apply the global univalence result of Gale and Nikaido (1965) because the Jacobian of the system (3.5) is not a P-matrix. We show that (3.5) has a unique solution using the global univalence result of Ambrosetti and Prodi (1995) by showing that the system has a unique solution when $\rho_{Y_d}(y) = 0$ (locally no endogeneity), the function that defines the system is proper, and the Jacobian is full-rank. U_{OC} is sufficient to establish the full-rank condition.

Remark 3.3 (Comparison with Heckman and Vytlacil (2007)). *Heckman and Vytlacil (2007, Section 7.2) consider an ordered choice model, where the instrument is restricted to shift all*

cutoffs by the same amount. Suppose that

$$D_z = \begin{cases} 1, & -\infty < \mu(z) + V \leq \pi_1 \\ 2, & \pi_1 < \mu(z) + V \leq \pi_2 \\ \vdots & \vdots \\ K, & \pi_{K-1} < \mu(z) + V < \infty \end{cases} . \tag{3.7}$$

where $V \mid Z \sim N(0,1)$. *This model is a special case of the model we consider in this section if we normalize* $V_z \sim N(0,1)$.

3.3 Continuous Treatment

Suppose $D \in \mathcal{D} \subseteq \mathbb{R}$ is an uncountable set and $d \mapsto F_{D|Z}(d \mid z)$ is strictly increasing on $\mathcal{D}$, for $z \in \{0,1\}$. Assume the treatment selection equation,

$$D_z = h(z, V_z) = F^{-1}_{D|Z}(V_z \mid z), \tag{3.8}$$

where $V_z \sim U(0,1)$. For the identification analysis, consider

$$F_{Y|D,Z}(y \mid d, z) = F_{Y_d|D_z,Z}(y \mid d, z) = F_{Y_d|V_z,Z}(y \mid F_{D|Z}(d \mid z), z), \tag{3.9}$$

where the first equality follows from EX and the second from equation (3.8) and a change of variable. Let $\mu_{d,y} \equiv \Phi^{-1}(F_{Y_d}(y))$ and $\eta_v \equiv \Phi^{-1}(v)$. By the properties of the conditional distribution, Lemma 2.1, EX and the properties of the Gaussian copula,

$$F_{Y_d|V_z,Z}(y \mid v, z) = \frac{(\partial/\partial v) F_{Y_d,V_z|Z}(y, v \mid z)}{(\partial/\partial v) F_{V_z|Z}(v \mid z)} = \Phi\left(\frac{\mu_{d,y} - \rho_{Y_d,V_z;Z}(y,v;z)\eta_v}{\sqrt{1 - \rho_{Y_d,V_z;Z}(y,v;z)^2}}\right)$$
$$+ \phi_2(\mu_{d,y}, \eta_v; \rho_{Y_d,V_z;Z}(y,v;z))(\partial/\partial v)\rho_{Y_d,V_z;Z}(y,v;z). \tag{3.10}$$

Assume that

$$\rho_{Y_d,V_z;Z}(y, F_{D|Z}(d \mid 1); 1) = \rho_{Y_d,V_z;Z}(y, F_{D|Z}(d \mid 0); 0) \equiv \rho_{Y_d}(y), \quad y \in \mathcal{Y},$$

and

$$(\partial/\partial v)\rho_{Y_d,V_z;Z}(y, F_{D|Z}(d \mid z); z) = 0, \quad z \in \{0, 1\},$$

where the differentiability of $v \mapsto \rho_{Y_d,V_z;Z}(y, v; z)$ follows by continuity of (Y_d, V_z). Note that the previous conditions are implied by CI. Then, combining (3.9) and (3.10) yields

$$\Phi^{-1}\left(F_{Y|D,Z}(y \mid d, z)\right) = a_{d,y} + b_{d,y}\Phi^{-1}(F_{D|Z}(d \mid z)), \quad z \in \{0, 1\}, \tag{3.11}$$

where $a_{d,y} = \mu_{d,y}/\sqrt{1 - \rho_{Y_d}(y)^2}$ and $b_{d,y} = -\rho_{Y_d}(y)/\sqrt{1 - \rho_{Y_d}(y)^2}$. Equation (3.11) a linear system of two equations on two unknowns: $a_{d,y}$ and $b_{d,y}$, which has solution

$$\begin{aligned} a_{d,y} &= \frac{\Phi^{-1}(F_{Y|D,Z}(y \mid d, 0))\Phi^{-1}(F_{D|Z}(d \mid 1)) - \Phi^{-1}(F_{Y|D,Z}(y \mid d, 1))\Phi^{-1}(F_{D|Z}(d \mid 0))}{\Phi^{-1}(F_{D|Z}(d \mid 1)) - \Phi^{-1}(F_{D|Z}(d \mid 0))}, \\ b_{d,y} &= \frac{\Phi^{-1}(F_{Y|D,Z}(y \mid d, 1)) - \Phi^{-1}(F_{Y|D,Z}(y \mid d, 0))}{\Phi^{-1}(F_{D|Z}(d \mid 1)) - \Phi^{-1}(F_{D|Z}(d \mid 0))}, \end{aligned} \tag{3.12}$$

under REL.

The following theorem shows that the distribution of the potential outcomes and the local dependence $\rho_{Y_d}(y)$ are identified.

Theorem 3.3 (Identification for Continuous Treatment)**.** *Suppose D_z, $z \in \{0, 1\}$, satisfies (3.8). Under EX, REL, and CI, the functions $y \mapsto F_{Y_d}(y)$ and $y \mapsto \rho_{Y_d}(y)$ are identified on $y \in \mathcal{Y}$, for $d \in \mathcal{D}$ by*

$$F_{Y_d}(y) = \Phi\left(\frac{a_{d,y}}{\sqrt{1 + b_{d,y}^2}}\right), \quad \rho_{Y_d}(y) = \frac{-b_{d,y}}{\sqrt{1 + b_{d,y}^2}},$$

where $a_{d,y}$ and $b_{y,d}$ are defined in (3.12).

Remark 3.4 (Comparison with Imbens and Newey (2009) and Torgovitsky (2010))**.** *Unlike Imbens and Newey (2009), this approach does not require an instrument with large support nor rank invariance in the treatment selection equation. i.e.,* $V_1 = V_0 = V$*, but imposes CI. Unlike Torgovitsky (2010), CI does not impose any structural models for* Y *and* D *nor restrictions on the dimension of the structural unobservables. Also note that Imbens and Newey (2009) motivate their control function assumption from the joint independence between* Z *and* (Y_d, V)*, whereas we only need "marginal" independence between* Z *and* Y_d *and between* Z *and* V_z*. See Appendix A for a more detailed comparison.*

4 Discussion on Copula Invariance

To further understand CI, we provide sets of simple sufficient conditions (Sections 4.1), interpret the condition under normality (Section 4.2), and invoke the implicit function theorem (Section 4.3). For ease of illustration, we focus on binary D in this section.

4.1 Sufficient Conditions

4.1.1 Joint Independence

First, we provide sufficient conditions for EX and CI.

Assumption EX′ (Joint Independence)**.** *For* $d, z \in \{0, 1\}$*,* $Z \perp\!\!\!\perp (Y_d, V_z)$*.*

Assumption CI′ (Unconditional CI)**.** *For* $d \in \{0, 1\}$*,*

$$\rho_{Y_d,V_1}(y, \pi(1)) = \rho_{Y_d,V_0}(y, \pi(0)) \equiv \rho_{Y_d}(y).$$

Proposition 4.1. *EX′ and CI′ imply EX and CI.*

Note that $\rho_{Y_d,V_z;Z}(\cdot,\cdot;z) = \rho_{Y_d,V_z}(\cdot,\cdot)$ by EX′[3] and $\rho_{Y_d,V_z}(y, \pi(z)) = \rho_{Y_d}(y)$ by CI′. Therefore, EX′ and CI′ are sufficient for CI. A sufficient condition for EX′ is $(Y_0, Y_1, V_z) \perp\!\!\!\perp Z$,

[3]In fact, not only EX′ implies $\rho_{Y_d,V_z;Z}(\cdot,\cdot;z) = \rho_{Y_d,V_z}(\cdot,\cdot)$, but the converse is also true. See Remark 4.3 below.

which is imposed in Imbens and Angrist (1994) and Vytlacil (2002) with $V_0 = V_1 = V$ almost surely, although it is sufficient for the LATE result to have $(Y_d, V) \perp\!\!\!\perp Z$ for $d \in \{0, 1\}$.

Remark 4.1 (CI′)**.** *We might wonder if CI′ imposes any restriction on the dependence between V_0 and V_1. The following example shows that if Y_d, V_0, and V_1 are jointly normal with the same marginal distributions, the restriction $F_{Y_d,V_0} = F_{Y_d,V_1}$, which is equivalent to CI′, does not restrict F_{V_0,V_1}. Let*

$$\begin{pmatrix} Y_d \\ V_0 \\ V_1 \end{pmatrix} \sim \mathcal{N}_3 \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{Y_d,V_0} & \rho_{Y_d,V_1} \\ \rho_{Y_d,V_0} & 1 & \rho_{V_0,V_1} \\ \rho_{Y_d,V_1} & \rho_{V_0,V_1} & 1 \end{bmatrix} \right).$$

Under CI′, $\rho_{Y_d,V_0} = \rho_{Y_d,V_1}$, so that (Y_d, V_0) and (Y_d, V_1) have the same distribution. Moreover, the matrix

$$\begin{bmatrix} 1 & \rho_{Y_d,V_0} & \rho_{Y_d,V_0} \\ \rho_{Y_d,V_0} & 1 & \rho_{V_0,V_1} \\ \rho_{Y_d,V_0} & \rho_{V_0,V_1} & 1 \end{bmatrix}$$

is positive definite for any $-1 < \rho_{V_0,V_1} < 1$. In other words, the condition $\rho_{Y_d,V_0} = \rho_{Y_d,V_1}$ does not restrict the distribution of (V_0, V_1).

4.1.2 Monotonicity

Next, we investigate the interaction between CI and LATE monotonicity (Imbens and Angrist, 1994). The specification of D_z in (2.1) is weaker than LATE monotonicity due to the vector of unobservables (V_1, V_0). Specifically, we can generate any compliance patterns from

the joint distribution of (V_1, V_0):

$$\Pr[D_1 = 1, D_0 = 1] = \Pr[V_1 \leq \pi(1), V_0 \leq \pi(0)]$$
$$\Pr[D_1 = 0, D_0 = 1] = \Pr[V_1 > \pi(1), V_0 \leq \pi(0)]$$
$$\Pr[D_1 = 1, D_0 = 0] = \Pr[V_1 \leq \pi(1), V_0 > \pi(0)]$$
$$\Pr[D_1 = 0, D_0 = 0] = \Pr[V_1 > \pi(1), V_0 > \pi(0)]$$

Suppose we maintain EX′ for ease of discussion. Then, the following conditions are sufficient for CI.

Assumption RI$_S$ (Rank Invariance in Selection)**.** *For $d \in \{0, 1\}$, $V_1 = V_0 = V$ almost surely.*

Assumption RI$_J$ (Joint Rank Invariance in Selection)**.** *For $d \in \{0, 1\}$, (Y_d, V_1) and (Y_d, V_0) are identically distributed such that $\rho_{Y_d,V_0}(y, v) = \rho_{Y_d,V_1}(y, v) \equiv \rho_{Y_d,V}(y, v)$.*

Assumption CI″ (CI in Treatment Propensity)**.** $\rho_{Y_d,V}(y, v) = \rho_{Y_d,V}(y)$.

RI$_S$ is equivalent to LATE montonicity, and it implies RI$_J$. RI$_J$ and CI″ (or RI$_S$ and CI″) being sufficient for CI shows how CI may interfere with treatment compliance patterns.

Proposition 4.2. *Under EX′, RI$_J$ and CI″ imply CI. RI$_S$ implies RI$_J$.*

4.2 Normality

Here we want to further understand CI by imposing normality on the joint distribution of unobservables. Normality is useful to compare CI with the rank similarity (RS) and rank invariance (RI) assumptions of Chernozhukov and Hansen (2005). In general, CI restricts the relationship between the parameters across conditional distributions whereas RI and RS directly restrict the relationship between the potential outcomes. In this sense, CI allows for more effect heterogeneity.

For ease of notation, we consider a slightly different version of the treatment selection equation (3.1):

$$D_z = 1\{\tilde{V}_z \leq q(z)\},$$

where $\tilde{V}_z \mid Z \sim N(0,1)$ is an alternative normalization such that $q(z) \equiv \Phi^{-1}(\pi(z))$.[4] We maintain EX and REL.

Assumption NM (Normality)**.** $(Y_0, Y_1, \tilde{V}_0, \tilde{V}_1)$ *are jointly normal conditional on* Z*,*

$$\begin{pmatrix} Y_0 \\ Y_1 \\ \tilde{V}_0 \\ \tilde{V}_1 \end{pmatrix} \mid Z = z \sim \mathcal{N}_4 \left(\begin{pmatrix} \mu_0 \\ \mu_1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_{01}(z) & \sigma_{0V_0}(z) & \sigma_{0V_1}(z) \\ \sigma_{01}(z) & \sigma_1^2 & \sigma_{1V_0}(z) & \sigma_{1V_1}(z) \\ \sigma_{0V_0}(z) & \sigma_{1V_1}(z) & 1 & \rho_{V_0V_1}(z) \\ \sigma_{0V_1}(z) & \sigma_{1V_1}(z) & \rho_{V_0V_1}(z) & 1 \end{pmatrix} \right).$$

In this distribution, some of the parameters do not depend on z because of EX. To compare CI, RI and RS, it is convenient to work with the standardized potential outcomes:

$$\tilde{Y}_0 = \frac{Y_0 - \mu_0}{\sigma_0}, \quad \tilde{Y}_1 = \frac{Y_1 - \mu_1}{\sigma_1}.$$

By NM,

$$\begin{pmatrix} \tilde{Y}_0 \\ \tilde{Y}_1 \\ \tilde{V}_0 \\ \tilde{V}_1 \end{pmatrix} \mid Z = z \sim \mathcal{N}_4 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{01}(z) & \rho_{0V_0}(z) & \rho_{0V_1}(z) \\ \rho_{01}(z) & 1 & \rho_{1V_0}(z) & \rho_{1V_1}(z) \\ \rho_{0V_0}(z) & \rho_{1V_0}(z) & 1 & \rho_{V_0V_1}(z) \\ \rho_{0V_1}(z) & \rho_{1V_1}(z) & \rho_{V_0V_1}(z) & 1 \end{pmatrix} \right),$$

where

$$\rho_{dd'}(z) = \frac{\sigma_{dd'}(z)}{\sigma_d \sigma_{d'}}, \quad \rho_{dV_z}(z) = \frac{\sigma_{dV_z}(z)}{\sigma_d}$$

[4]This is because of the normalization $V_z \mid Z \sim U[0,1]$ and thus $\tilde{V}_z = \Phi^{-1}(V_z) \mid Z \sim N(0,1)$.

for $d, d', z \in \{0,1\}$. Then, CI imposes that

$$\rho_{dV_1}(1) = \rho_{dV_0}(0), \quad d \in \{0,1\},$$

that is, the correlation between the potential outcomes and the latent treatment assignment does not depend on the value of the instrument. Note that CI still imposes substantial invariance restrictions even under normality.[5] However, CI does not impose any restriction on the correlation between the potential outcomes, $\rho_{01}(z)$.

In terms of the standardized variables, RS can be stated as follows:

Assumption RS (Rank Similarity)**.** *Conditional on* $\tilde{V}_z = v$ *and* $Z = z$, $\tilde{Y}_0 \sim \tilde{Y}_1$.

By NM, $\tilde{Y}_d \mid \tilde{V}_z = v, Z = z \sim \mathcal{N}\left(\rho_{dV_z}(z)v, 1 - \rho_{dV_z}(z)^2\right)$ for $d \in \{0,1\}$. RS therefore imposes that $\rho_{0V_z}(z) = \rho_{1V_z}(z)$ for $z \in \{0,1\}$. Compared to CI, RS imposes that the correlation of the two potential outcomes with the latent treatment assignment is the same for all the values of the instrument, but does not require that the correlation is independent of the value of the instrument. Unlike CI, RS implicitly imposes restrictions on $\rho_{01}(z)$; see Appendix A.1. Note that RS also restricts selection on gains because

$$\mathrm{Cov}(Y_1 - Y_0, \tilde{V}_z \mid Z = z) = \mathrm{Cov}(\sigma_1\tilde{Y}_1 - \sigma_0\tilde{Y}_0, \tilde{V}_z \mid Z = z) = \sigma_1\rho_{1V_z}(z) - \sigma_0\rho_{0V_z}(z),$$

which, for example, vanishes if $\sigma_1 = \sigma_0$ under RS. CI, instead, imposes that the selection on gains is independent of the instrument, that is

$$\mathrm{Cov}(Y_1 - Y_0, \tilde{V}_1 \mid Z = 1) = \mathrm{Cov}(Y_1 - Y_0, \tilde{V}_0 \mid Z = 0).$$

In general, CI and RS are not nested.

In terms of the standardized variables, RI can be stated as follows:

[5] CI does *not* trivially holds under joint normality. This is because even under the independence assumptions $(Y_0, V_z) \perp\!\!\!\perp Z$ and $(Y_1, V_z) \perp\!\!\!\perp Z$ (i.e. EX′), we only achieve $\rho_{dV_z}(z) = \rho_{dV_z}$. After imposing independence and RI for the selection, normality becomes sufficient for CI.

Assumption RI (Rank Invariance)**.** $\tilde{Y}_0 = \tilde{Y}_1$, *almost surely.*

Under NM, RI imposes that $\rho_{01}(z) = 1$ and $\rho_{0V_z}(z) = \rho_{1V_z}(z)$. RI is therefore more restrictive than RS as it imposes the same restrictions plus perfect correlation between the standardized potential outcomes. Appendix A.1 contains further discussions on RS and RI in comparison to CI without assuming normality.

4.3 Local Dependence as Implicit Function

The LGR in Lemma 2.1 can also be expressed in terms of the copula function. In particular,

$$\tilde{C}(u_1, u_2 \mid z) = C(u_1, u_2; \rho(u_1, u_2; z)). \tag{4.1}$$

Here $\tilde{C}(u_1, u_2 \mid z)$ is the conditional copula of (Y, V_z) given $Z = z$, that is the joint distribution of $U_1 = F_{Y|Z}(Y \mid Z)$ and $U_2 = F_{V_z|Z}(V_z \mid Z)$ conditional on $Z = z$, and C is the Gaussian copula. Note that $U_2 = V_z$ under EX and the normalization $V_z \sim U(0, 1)$.

The parameter $\rho(u_1, u_2; z)$ can be viewed as an implicit function in (4.1). For any $z \neq z'$, consider

$$\begin{aligned}\tilde{C}(u_1, u_2 \mid z) - \tilde{C}(u_1, u_2 \mid z') &= C_\rho(u_1, u_2; \tilde{\rho}) \left\{\rho(u_1, u_2; z) - \rho(u_1, u_2; z')\right\} \\ &= \phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \tilde{\rho}) \left\{\rho(u_1, u_2; z) - \rho(u_1, u_2; z')\right\},\end{aligned}$$

where $\tilde{\rho}$ lies between $\rho(u_1, u_2; z)$ and $\rho(u_1, u_2; z')$ and $\phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho) \neq 0$ for $(u_1, u_2) \in (0, 1)^2$.

To focus on CI, we maintain EX′ and RI_J, so that CI boils down to CI″, $\rho_{Y_d,V}(y, v) = \rho_{Y_d,V}(y)$. To understand CI″, consider the LGR of the (unconditional) copula of (Y_d, V):

$$\tilde{C}(u_1, u_2) = C(u_1, u_2; \rho(u_1, u_2)).$$

CI″ is equivalent to $\rho(u_1, u_2) = \rho(u_1)$. Since $\tilde{C}$ and C are differentiable in almost all (u_1, u_2)

(by the definition of copula), so is ρ by the implicit function theorem. Then, for $\tilde{C}(u_1 \mid u_2)$ and $C(u_1 \mid u_2)$ being conditional copulas,

$$\begin{aligned}\tilde{C}(u_1 \mid u_2) &= C(u_1 \mid u_2; \rho(u_1, u_2)) + C_\rho(u_1, u_2; \rho(u_1, u_2))\frac{\partial \rho(u_1, u_2)}{\partial u_2} \\ &= C(u_1 \mid u_2; \rho(u_1, u_2)) + \phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho(u_1, u_2))\frac{\partial \rho(u_1, u_2)}{\partial u_2}.\end{aligned}$$

We can interpret $\phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho(u_1, u_2))\frac{\partial \rho(u_1, u_2)}{\partial u_2}$ as the adjustment term that equates the two conditional copulas.[6] In general, the LGR for the joint distribution does *not* imply the same representation for the conditional distribution. Rewrite the equation to have

$$\frac{\partial \rho(u_1, u_2)}{\partial u_2} = \frac{\tilde{C}(u_1 \mid u_2) - C(u_1 \mid u_2; \rho(u_1, u_2))}{\phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho(u_1, u_2))}$$

for $(u_1, u_2) \in (0, 1)^2$, which captures a (normalized) deviation from local Gaussianity. Given this result, CI″ w.r.t. u_2 is equivalent to $\tilde{C}(u_1 \mid u_2) = C(u_1 \mid u_2; \rho(u_1))$. We thus have the following result:

Proposition 4.3. *Under EX′ and RI_J, CI holds if $\tilde{C}(u_1 \mid u_2) = C(u_1 \mid u_2; \rho(u_1))$.*

Remark 4.2 (Stochastic Monotonicity)**.** *$\tilde{C}(u_1 \mid u_2) = C(u_1 \mid u_2; \rho(u_1))$ implies that $u_2 \mapsto \tilde{C}(u_1 \mid u_2)$ is monotonic for each u_1. This restricts the dependence between U_1 and U_2. For example, if U_1 and U_2 are continuous, then the effect of U_2 on the τ-quantile of U_1 cannot change sign with respect to the value of U_2, but can change sign with τ. Note that if U_1 and U_2 are jointly normal, then this τ-quantile effect cannot change sign with τ. More generally, the stochastic monotonicity condition holds if, for example, the conditional distribution has a monotone likelihood ratio.*

Remark 4.3. *Under the LGR (4.1), we can show that the equivalence between statistical*

[6] Note that $\phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho(u_1, u_2)) \to 0$ as $u_1 \to 1$ or 0, which is consistent with $\tilde{C}(u_1 \mid u_2)$ and $C(u_1 \mid u_2)$ being CDFs (and similarly in the previous case).

independence and a restriction on the dependence parameter as an implicit function:

$$(U_1, U_2) \perp\!\!\!\perp Z \Leftrightarrow \tilde{C}(u_1, u_2 \mid z) - \tilde{C}(u_1, u_2 \mid z') \neq 0 \quad \textit{for any } z \neq z' \textit{ and } (u_1, u_2) \in (0,1)^2$$
$$\Leftrightarrow \rho(u_1, u_2; Z) = \rho(u_1, u_2) \quad \textit{almost surely, for any } (u_1, u_2) \in (0,1)^2.$$

5 Estimation and Inference

We estimate flexible semiparametric models for the distributions of the potential outcomes for the three types of treatments based on distribution regression (DR). We also show how to construct estimators of treatment effect parameters such as the conditional and unconditional QTE and the ATE using the plug-in rule. In this section, we make the role of the covariates X explicit. For estimation, we assume we have access to a random sample of size n from (Y, D, Z, X), $\{(Y_i, D_i, Z_i, X_i)\}_{i=1}^n$.

Let $B(X_i)$, $B(X_i, Z_i)$, and $B(D_i, X_i, Z_i)$ denote vectors of transformations of X_i, (X_i, Z_i), and (D_i, X_i, Z_i), respectively. Define the indicators $I_i(y) \equiv 1\{Y_i \leq y\}$ and $J_i(d) \equiv 1\{D_i \leq d\}$. Let $\bar{\mathcal{D}}$ and $\bar{\mathcal{Y}}$ be two finite grids covering $\mathcal{D}$ and $\mathcal{Y}$.[7]

5.1 Binary Treatments

We consider a DR model for the conditional potential outcome distributions,

$$F_{Y_d|X}(y|x) = \Phi(B(x)'\beta_d(y)), \quad d \in \{0,1\}, \tag{5.1}$$

and a Probit model for the propensity score,

$$\pi(z,x) = \Pr[D = 1 \mid Z = z, X = x] = \Phi(B(z,x)'\pi). \tag{5.2}$$

[7] We can set $\bar{\mathcal{Y}} = \mathcal{Y}$ when $\mathcal{Y}$ is finite. We only use $\bar{\mathcal{D}}$ when D is continuous.

We model the local dependence parameter as

$$\rho_{Y_d;X}(y;x) = \rho(B(x)'\gamma_d(y)), \quad d \in \{0,1\}, \tag{5.3}$$

where $\rho(u) = \tanh(u) \in [-1,1]$, the Fisher transformation.[8] Define $\theta_d(y) \equiv (\beta_d(y), \gamma_d(y))$.

Together, (5.1), (5.2), and (5.3) imply the bivariate DR model

$$\begin{aligned}
\Pr(Y \leq y, D = 1 \mid Z = z) &= \Phi_2(B(x)'\beta_1(y), B(z,x)'\pi; \rho(B(x)'\gamma_1(y))), \\
\Pr(Y \leq y, D = 0 \mid Z = z) &= \Phi_2(B(x)'\beta_0(y), -B(z,x)'\pi; -\rho(B(x)'\gamma_0(y))),
\end{aligned}$$

where we have used the symmetry properties of the bivariate Gaussian distribution.

We propose computationally tractable two-step maximum likelihood estimators, building on Chernozhukov et al. (2020a).

Algorithm 5.1 (Estimation of Binary Treatment Model)**.** *We compute the estimator in 2 stages:*

1. *Treatment equation: estimate π using a Probit regression:*

$$\widehat{\pi} = \arg\max_{c} \sum_{i=1}^{n} \left[D_i \log \Phi(B(X_i, Z_i)'c) + (1 - D_i) \log(1 - \Phi(B(X_i, Z_i)'c)) \right]$$

2. *Outcome equation: for $y \in \bar{\mathcal{Y}}$ and $d \in \{0,1\}$, $\widehat{F}_{Y_d|X}(y|x) = \Phi(B(x)'\widehat{\beta}_d(y))$, where*

$$\begin{aligned}
\widehat{\theta}_1(y) &= \arg\max_{t=(b,g)} \sum_{i=1}^{n} D_i \big[I_i(y) \log \Phi_2(X_i'b, B(X_i, Z_i)'\widehat{\pi}, \rho(X_i'g)) \\
&\quad + (1 - I_i(y)) \log \Phi_2(-B(X_i)'b, B(X_i, Z_i)'\widehat{\pi}, \rho(B(X_i)'g)) \big], \\
\widehat{\theta}_0(y) &= \arg\max_{t=(b,g)} \sum_{i=1}^{n} (1 - D_i) \big[I_i(y) \log \Phi_2(B(X_i)'b, -B(X_i, Z_i)'\widehat{\pi}, -\rho(B(X_i)'g)) \\
&\quad + (1 - I_i(y)) \log \Phi_2(-B(X_i)'b, -B(X_i, Z_i)'\widehat{\pi}, -\rho(B(X_i)'g)) \big].
\end{aligned}$$

[8] To simplify the notation, we use the same vector of transformations in models (5.1) and (5.3). This is not essential, and one can use different specifications in both models.

Rearrange the estimates $y \mapsto \widehat{F}_{Y_d|X}(y \mid x)$ *on* $\bar{\mathcal{Y}}$ *if needed.*

Remark 5.1 (Computation)**.** *The first stage of Algorithm 5.1 is a conventional Probit regression, and estimation can proceed using existing software. The second stage is computationally more expensive since it involves a nonlinear smooth optimization problem. This optimization problem can be solved using standard algorithms such as Newton-Raphson.*

5.2 Ordered Treatments

As for binary treatments, we model all the components using flexible generalized linear and DR models:

$$\begin{aligned} F_{Y_d|X}(y \mid x) &= \Phi(B(x)'\beta_d(y)), \\ \rho_{Y_d;X}(y \mid x) &= \rho(B(x)'\gamma_d(y)), \\ \pi_d(x,z) &= F_{D|Z,X}(d \mid z,x) = \Phi(B(z,x)'\pi(d)), \end{aligned}$$

where $\rho(u) = \tanh(u)$.

Algorithm 5.2 (Estimation of Ordered Treatment Model)**.** *We compute the estimator in 2 stages:*

1. *Treatment equation: set* $\widehat{\pi}_0(z,x) = 0$ *and* $\widehat{\pi}_K(z,x) = 1$ *for all* (z,x)*. For* $d \in \{1,\ldots,K-1\}$*,* $\widehat{\pi}_d(z,x) = \Phi(B(z,x)'\widehat{\pi}(d))$*, where*

$$\widehat{\pi}(d) \in \arg\max_p \sum_{i=1}^n \left[J_i(d)\log\Phi(B(Z_i,X_i)'p) + (1-J_i(d))\log\Phi(-B(Z_i,X_i)'p)\right].$$

Rearrange the estimates $d \mapsto \widehat{\pi}_d(z,x)$ *on* $\mathcal{D}$ *if needed. This rearrangement is important to avoid having logarithms of negative numbers in the second stage.*[9]

[9] In the second stage, $g_{d,i}(b,g) > 0$ and $\bar{g}_{d,i}(b,g) > 0$ a.s. if $\widehat{\pi}_d(Z_i,X_i) > \widehat{\pi}_{d-1}(Z_i,X_i)$ a.s.

2. *Outcome equation: for* $y \in \bar{\mathcal{Y}}$ *and* $d \in \bar{\mathcal{D}}$, $\widehat{F}_{Y_d|X}(y \mid x) = \Phi(B(x)'\widehat{\beta}_d(y))$, *where*

$$\widehat{\theta}_d(y) = (\widehat{\beta}_d(y), \widehat{\gamma}_d(y)) \in \arg\max_{b,g} \sum_{i=1}^{n} 1\{D_i = d\} \left[I_i(y) \log g_{d,i}(b,g) + (1 - I_i(y)) \log \bar{g}_{d,i}(b,g) \right],$$

where

$$g_{d,i}(b,g) \equiv \Phi_2(B(X_i)'b, \widehat{\pi}_d(Z_i, X_i), \rho(B(X_i)'g)) - \Phi_2(B(X_i)'b, \widehat{\pi}_{d-1}(Z_i, X_i), \rho(B(X_i)'g)),$$

and

$$\bar{g}_{d,i}(b,g) \equiv \widehat{\pi}_d(Z_i, X_i) - \widehat{\pi}_{d-1}(Z_i, X_i) - g_{d,i}(b,g).$$

Rearrange the estimates $y \mapsto \widehat{F}_{Y_d|X}(y \mid x)$ *on* $\bar{\mathcal{Y}}$ *if needed.*

Remark 5.2 (Computation)**.** *The first stage is a sequence of Probit regressions that can be solved using standard software, as in Algorithm 5.1. The second stage is a nonlinear smooth optimization problem that can be solved using standard algorithms such as Newton-Raphson.*

5.3 Continuous Treatments

We construct plug-in estimators based on the closed-form solutions in Section 3.3. We consider DR models for $F_{Y|D,Z,X}$ and $F_{D|Z,X}$,

$$F_{Y|D,Z,X}(y \mid d, z, x) = \Phi(B(d, z, x)'\beta(y)), \tag{5.4}$$

$$F_{D|Z,X}(d \mid z, x) = \Phi(B(z, x)'\pi(d)). \tag{5.5}$$

Algorithm 5.3 (Estimation of Continuous Treatment Model)**.** *We compute the estimator in two stages:*

1. *Observable conditional distributions: for* $y \in \bar{\mathcal{Y}}$ *and* $d \in \bar{\mathcal{D}}$, $\widehat{F}_{Y|D,Z,X}(y|d,z,x) =$

$\Phi(B(d,z,x)'\widehat{\beta}(y))$ *and* $\widehat{F}_{D|Z,X}(d|z,x) = \Phi(B(z,x)'\widehat{\beta}(d))$*, where*

$$\begin{aligned}
\widehat{\beta}(y) &= \arg\max_b \sum_{i=1}^n [I_i(y)\log\Phi(B(D_i,Z_i,X_i)'b) + (1-I_i(y))\log(1-\Phi(B(D_i,Z_i,X_i)'b))] \\
\widehat{\pi}(d) &= \arg\max_p \sum_{i=1}^n [J_i(d)\log\Phi(B(Z_i,X_i)'p) + (1-J_i(d))\log(1-\Phi(B(Z_i,X_i)'p))]
\end{aligned}$$

2. *Potential outcome distributions: for* $y \in \bar{\mathcal{Y}}$ *and* $d \in \bar{\mathcal{D}}$, $\widehat{F}_{Y_d|X}(y|x) = \Phi(\widehat{\mu}_{d,y;x})$*, where* $\widehat{\mu}_{d,y;x} = \widehat{a}_{d,y;x}/\sqrt{1+\widehat{b}^2_{d,y;x}}$ *and*

$$\begin{aligned}
\widehat{a}_{d,y;x} &= \frac{(B(d,0,x)'\widehat{\beta}(y))(B(1,x)'\widehat{\pi}(d)) - (B(d,1,x)'\widehat{\beta}(y))(B(0,x)'\widehat{\pi}(d))}{B(1,x)'\widehat{\pi}(d) - B(0,x)'\widehat{\pi}(d)}, \\
\widehat{b}_{d,y;x} &= \frac{B(d,1,x)'\widehat{\beta}(y) - B(d,0,x)'\widehat{\beta}(y)}{B(1,x)'\widehat{\pi}(d) - B(0,x)'\widehat{\pi}(d)}.
\end{aligned}$$

Rearrange the estimates $y \mapsto \widehat{F}_{Y_d|X}(y \mid x)$ *on* $\bar{\mathcal{Y}}$ *if needed.*

5.4 Estimation of QSF and ASF

Finally, we can estimate the marginal distribution of the potential outcomes, quantile structural function (QSF) and average structural function (ASF) by plugging in the estimators obtained above. To give a unified treatment to all the treatment cases, we set $\bar{\mathcal{D}} = \mathcal{D}$ when D is binary or ordered.

Algorithm 5.4 (Estimation of F_{Y_d}, QSF and ASF)**.** *Estimation proceeds in two steps.*

1. *Unconditional distribution: for* $y \in \bar{\mathcal{Y}}$ *and* $d \in \bar{\mathcal{D}}$*,*

$$\widehat{F}_{Y_d}(y) = \frac{1}{n}\sum_{i=1}^n \widehat{F}_{Y_d|X}(y \mid X_i)$$

For $y \in \mathcal{Y} \setminus \bar{\mathcal{Y}}$ *and* $d \in \bar{\mathcal{D}}$*,*

$$\widehat{F}_{Y_d}(y) = \max\{\widehat{F}_{Y_d}(\bar{y}) : \bar{y} < y, \bar{y} \in \bar{\mathcal{Y}}\}.$$[10]

[10] In practice, one can also use linear extrapolation when D is continuous.

2. Quantile and average structural function:

$$\widehat{QSF}_\tau(d) = \mathcal{Q}_\tau(\widehat{F}_{Y_d}),$$
$$\widehat{ASF}(d) = \mathcal{E}(\widehat{F}_{Y_d}).$$

The quantile and average effects can be be obtained accordingly.

5.5 Inference

The target parameters in Section 5.4 are function-valued. Inference on these parameters can be performed using resampling methods. To provide a unified framework, we denote the functional parameters by $u \mapsto \delta_u, u \in \mathcal{U}$, where $\mathcal{U} \subset \mathcal{Y} \times \mathcal{D} \times \mathcal{T}$, where $\mathcal{T} \subset (c, 1-c)$ for $c > 0$. For example, if we are interested in $\tau \mapsto QSF_\tau(d)$ on $[.05, .95]$, then $u = \tau$, $\delta_u = QSF_u(d)$ and $\mathcal{U} = [.05, .95]$. In practice, we approximate $\mathcal{U}$ using a fine grid $\bar{\mathcal{U}}$. We denote the estimator of δ_u obtained from Algorithms 5.1, 5.2, 5.3, and 5.4 as $\widehat{\delta}_u$.

We focus on constructing uniform confidence bands, $CB_{(1-\alpha)}(\delta_u)$, satisfying

$$\lim_{n\to\infty} \Pr[\delta_u \in CB_{(1-\alpha)}(\delta_u), \text{ for all } u \in \mathcal{U}] = 1 - \alpha.$$

Uniform confidence bands can be used to test a variety of hypotheses of interest, such as the hypotheses of no effect or constant effects when applied to treatment effect parameters, or stochastic dominance.

Under standard regularity conditions, $\sqrt{n}(\widehat{\delta}_u - \delta_u)$ converges in distribution to a mean-zero Gaussian process $G_\delta(u)$ in $\ell^\infty(\mathcal{U})$ and the bootstrap is valid. This follows from standard arguments for DR (e.g., Chernozhukov et al., 2013, 2020a) and the functional delta method.

The following algorithm provides a generic bootstrap construction of $CB_{(1-\alpha)}(\delta_u)$. When δ_u is equal to the QSF or quantile effect functions, the algorithm only applies when the outcomes are continuous. When the outcomes are discrete or mixed discrete-continuous, the estimators of the QSF or quantile effect functions are not be asymptotically Gaussian in

general. In this case, uniform confidence bands for the QSF and quantile effect functions can be constructed as described in Chernozhukov et al. (2020b).

Algorithm 5.5 (Uniform Confidence Bands for Functional Parameters[11])**.**

1. *For $u \in \bar{\mathcal{U}}$, obtain B bootstrap draws of the estimator $\widehat{\delta}_u$, $\{\widehat{\delta}_u^{(b)} : 1 \leq b \leq B\}$.*

2. *For $u \in \bar{\mathcal{U}}$, compute the standard error,*

$$SE(\widehat{\delta}_u) = (\widehat{Q}_\delta(0.75, u) - (\widehat{Q}_\delta(0.25, u))/(\Phi^{-1}(0.75) - (\Phi^{-1}(0.25)),$$

   where $\widehat{Q}_\delta(\tau, u)$ is the τ-quantile of $\{\widehat{\delta}_u^{(b)} : 1 \leq b \leq B\}$.

3. *Compute the critical value as*

$$cv(1-\alpha) = (1-\alpha)\text{-quantile of } \left\{ \max_{u \in \bar{\mathcal{U}}} \frac{|\widehat{\delta}_u^{(b)} - \widehat{\delta}_u|}{SE(\widehat{\delta}_u)} : 1 \leq b \leq B \right\}.$$

4. *Compute the $(1-\alpha)$ uniform confidence band as*

$$CB_{(1-\alpha)}(\delta_u) = [\widehat{\delta}_u \pm cv(1-\alpha) SE(\widehat{\delta}_u)], \quad u \in \bar{\mathcal{U}}.$$

The estimation algorithms for binary and ordered treatments (Algorithms 5.1 and 5.2) involve nonlinear optimization problems. Therefore, following Chernozhukov et al. (2020a), we recommend using the multiplier bootstrap in Step 1 of Algorithm 5.5. The multiplier bootstrap is a computationally efficient resampling procedure based on the influence function of the estimator that avoids re-estimating the parameters in Algorithms 5.1 and 5.2 in each of the B bootstrap iterations.

The estimation approach for continuous treatments in Algorithm 5.3 does not involve solving a nonlinear optimization problem in the second step and is computationally less expensive than Algorithms 5.1 and 5.2. Therefore, the standard empirical bootstrap is a

[11] See, for example, Chernozhukov et al. (2013, 2020a,b) for similar algorithms.

natural alternative to the multiplier bootstrap in Step 1 when the sample size is small or moderate, as for example in Section 6.

6 Empirical Application

We illustrate our method by estimating the distributional effects of sleep on well-being. We use the data from the experimental analysis of Bessone et al. (2021b), who analyzed the effects of randomized interventions to increase sleep of low-income adults in India.[12] Specifically, they considered two main treatments (see their Section III for details): (i) *devices + encouragement* (information on the benefits of and tips to improve sleep, encouragements and a sleep tracker, and devices to improve sleep environment) and (ii) *devices + incentives* (same as (i) and payments for each minute of sleep increase). In addition, they cross-randomized a *nap treatment* (the opportunity to nap during the day).

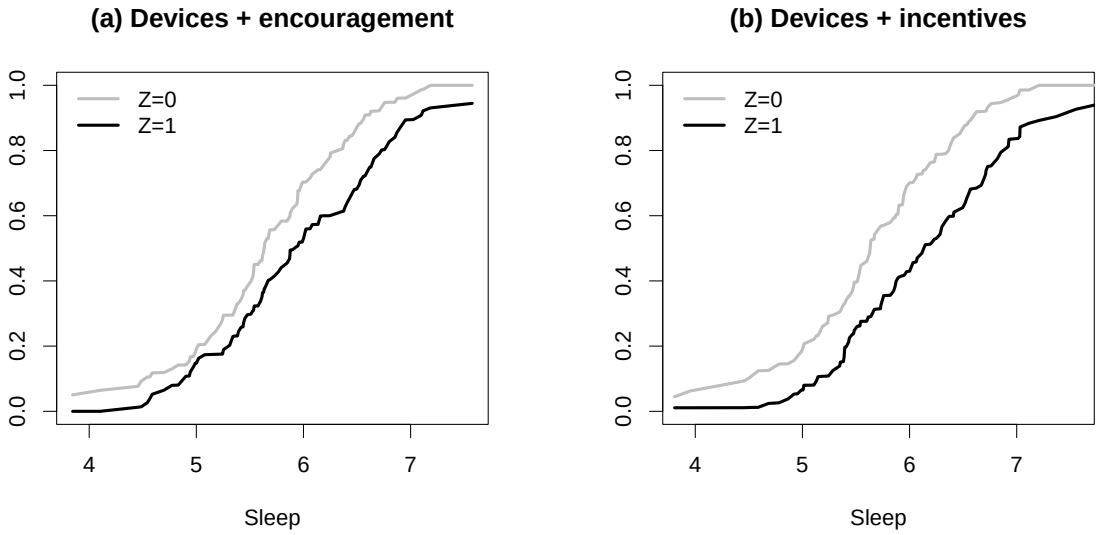
The outcome of interest (Y) is an overall index of individual well-being. The treatment (D) is the number of hours of sleep per night. Following Bessone et al. (2021b, Table A.XVII) and the recent reanalysis by Dong and Lee (2023), we use the randomly assigned experimental treatments as instruments for sleep. We focus on the two main experimental treatments, devices + encouragement (Z_1) and devices + incentives (Z_2), and restrict the sample to individuals who did not receive the nap treatment, as in Dong and Lee (2023). Both instruments are equal to one if the individuals received the experimental treatment and zero otherwise. The vector of covariates (X) includes controls for gender, three age indicators, and the baseline well-being index, as in Bessone et al. (2021b, Table A.XVII). The treatment takes on many values and is treated as continuous. We therefore use the estimators for continuous treatments in Section 5.3 in this application.

We start by analyzing the (distributional) first-stage relationship between D and Z_1 and D and Z_2 to shed light on the plausibility of REL. Figure 1 plots $\widehat{F}_{D|Z}(\cdot \mid 1) = n^{-1}\sum_{i=1}^{n} \widehat{F}_{D|Z,X}(\cdot \mid 1, X_i)$ and $\widehat{F}_{D|Z}(\cdot \mid 0) = n^{-1}\sum_{i=1}^{n} \widehat{F}_{D|Z,X}(\cdot \mid 0, X_i)$ for both instruments.

[12] We downloaded the data from the Harvard Dataverse replication package (Bessone et al., 2021a).

It shows that both instruments induce a shift in the distribution of D. The shift is more pronounced for Z_2, which is not surprising given the additional financial incentives relative to Z_1.

Figure 1: Distributional First Stage



Notes: The sample sizes are $n = 152$ in Figure 1(a) and $n = 151$ in Figure 1(b). All specifications control for gender, three age indicators, and the baseline well-being index.

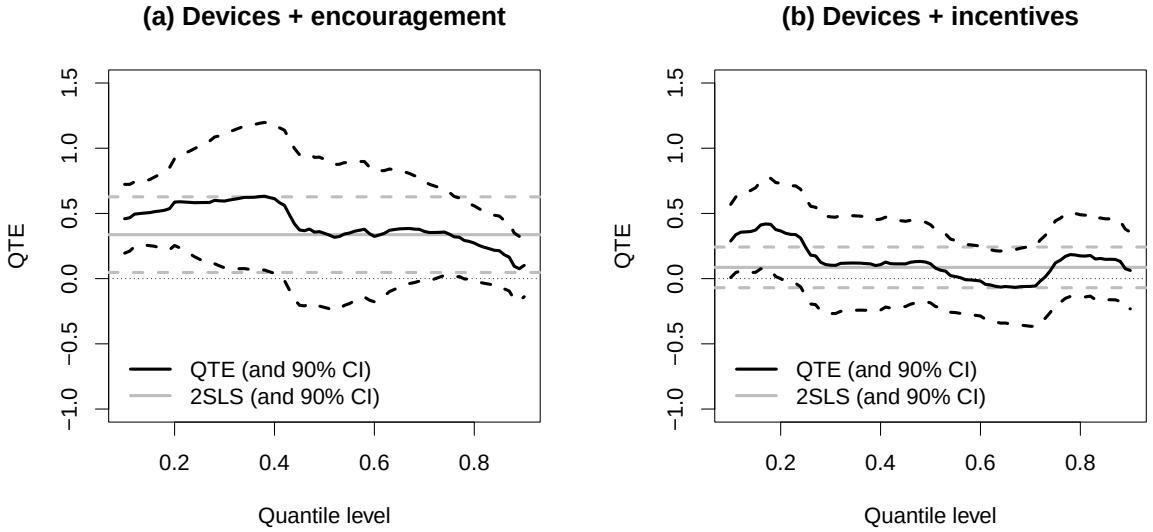
In addition to REL, our method relies on EX and CI. The random assignment of the instruments renders the independence assumptions in EX plausible.[13] CI allows the local dependence between potential well-being and the unobservable determinants of sleep to depend on the level of well-being, but not on the level of the unobservable determinants of sleep and the instruments.

Figure 2 plots estimates of the normalized QTE, $\tau \mapsto (\mathcal{Q}_\tau(\widehat{F}_{Y_{d''}}) - \mathcal{Q}_\tau(\widehat{F}_{Y_{d'}}))/(d'' - d')$, including 90% pointwise confidence intervals (CIs) computed using empirical bootstrap. We set d'' and d' to be equal to the 75% and 25% quantile of the empirical distribution of sleep, respectively. We report the results separately for each instrument (setting the other instrument to zero). For comparison, we also report two-stage least squares (2SLS) estimates using the same set of covariates with 90% confidence intervals.

Figure 2 reveals interesting effect heterogeneity across the distribution of well-being. For

[13]Note that random assignment does not automatically imply the (implicit) exclusion restriction, which requires that the instruments have no direct effect on the well-being.

Figure 2: Quantile Treatment Effects



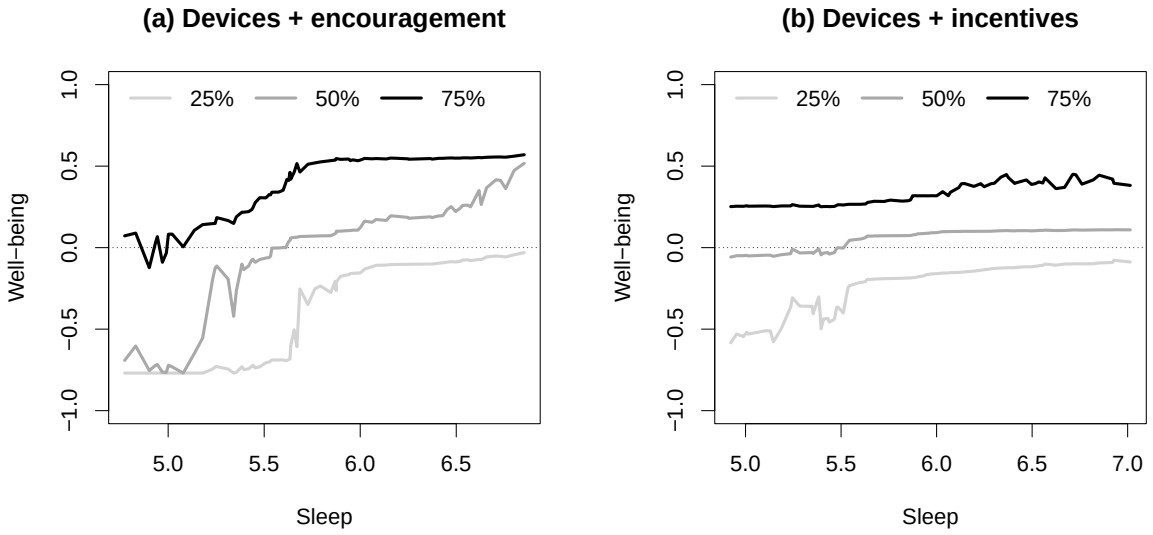
Notes: The sample sizes are $n = 152$ in Figure 2(a) and $n = 151$ in Figure 2(b). Pointwise CIs for the normalized QTE are computed using the empirical bootstrap with 10,000 repetitions. All specifications control for gender, three age indicators, and the baseline well-being index.

both instruments, the QTEs are the largest and significant at the 90% level over a range of quantile levels below the median and smaller and insignificant at the upper tail. The 2SLS estimates based on Z_1 are larger than those based on Z_2. The estimates based on Z_1 are significant at the 90% level, while those based on Z_2 are not.

We treat sleep as a continuous random variable. It is therefore also interesting to investigate the effect heterogeneity across the distribution of sleep. Figure 3 displays the estimated QSF, $d \mapsto \widehat{QSF}_\tau(d)$, for $\tau \in \{0.25, 0.5, 0.75\}$. The estimated QSFs display an increasing overall pattern for both instruments and all three quantile levels. However, there are some notable nonmonotonicities and interesting patterns of heterogeneity, especially for values of sleep at the lower tail of the sleep distribution.

The analysis in this section showcases the value-added that our method can bring to standard empirical analyses focusing on average effects. While a simple two-stage least squares analysis suggests that sleep has moderate or insignificant average effects on well-being, the proposed approach uncovers interesting patterns of heterogeneity in the effect of sleep on well-being.

Figure 3: Quantile Structural Function



Notes: The sample sizes are $n = 152$ in Figure 2(a) and $n = 151$ in Figure 2(b). All specifications control for gender, three age indicators, and the baseline well-being index.

References

Abadie, A. (2002): "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 97, 284–292. A.2

Abadie, A., J. Angrist, and G. Imbens (2002): "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings," *Econometrica*, 70, 91–117. 1.1

Ambrosetti, A. and G. Prodi (1995): *A primer of nonlinear analysis*, 34, Cambridge University Press. 3.2, 3.2, G.2

Angrist, J. D. and I. Fernández-Val (2013): *ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework*, Cambridge University Press, vol. 3 of *Econometric Society Monographs*, 401–434. A.2

Arellano, M. and S. Bonhomme (2017): "Quantile selection models with an application to understanding changes in wage inequality," *Econometrica*, 85, 1–28. 1.1

Athey, S. and G. W. Imbens (2006): "Identification and Inference in Nonlinear Difference-in-Differences Models," *Econometrica*, 74, 431–497. 1.1

Balke, A. and J. Pearl (1997): "Bounds on treatment effects from studies with imperfect compliance," *Journal of the American Statistical Association*, 92, 1171–1176. 1

Bessone, P., G. Rao, F. Schilbach, H. Schofield, and M. Toma (2021a): "Replication Data for: 'The Economic Consequences of Increasing Sleep among the Urban Poor'," . 12

——— (2021b): "The Economic Consequences of Increasing Sleep Among the Urban Poor*," *The Quarterly Journal of Economics*, 136, 1887–1941. 1, 6

Blundell, R., X. Chen, and D. Kristensen (2007): "Semi-nonparametric IV estimation of shape-invariant Engel curves," *Econometrica*, 75, 1613–1669. 1.1

Carneiro, P. and S. Lee (2009): "Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality," *Journal of Econometrics*, 149, 191–208. 1.1

Chen, X., Y. Fan, and V. Tsyrennikov (2006): "Efficient estimation of semiparametric multivariate copula models," *Journal of the American Statistical Association*, 101, 1228–1240. 1.1

Chernozhukov, V., I. Fernández-Val, and S. Luo (2020a): "Distribution regression with sample selection, with an application to wage decompositions in the UK," *arXiv preprint arXiv:1811.11603*. 1, 1.1, 2.2, 2.3, 5.1, 5.5, 5.5, 11, F

Chernozhukov, V., I. Fernández-Val, and B. Melly (2013): "Inference on Counterfactual Distributions," *Econometrica*, 81, 2205–2268. 5.5, 11

Chernozhukov, V., I. Fernández-Val, B. Melly, and K. Wüthrich (2020b): "Generic Inference on Quantile and Quantile Effect Functions for Discrete Outcomes," *Journal of the American Statistical Association*, 115, 123–137. 5.5, 11

Chernozhukov, V. and C. Hansen (2005): "An IV model of quantile treatment effects," *Econometrica*, 73, 245–261. 1, 1.1, 4.2, 6, A.1, A.1, E.2

——— (2013): "Quantile models with endogeneity," *Annu. Rev. Econ.*, 5, 57–81. 1.1

Chesher, A. (2003): "Identification in nonseparable models," *Econometrica*, 71, 1405–1441. 1.1

Chesher, A. and A. M. Rosen (2020): "Generalized instrumental variable models, methods, and applications," in *Handbook of Econometrics*, Elsevier, vol. 7, 1–110. 1

De Chaisemartin, C. (2017): "Tolerating defiance? Local average treatment effects without monotonicity," *Quantitative Economics*, 8, 367–396. 3.2

De Paula, Á., I. Rasul, and P. Souza (2019): "Identifying network ties from panel data: Theory and an application to tax competition," *arXiv preprint arXiv:1910.07452*. 2

D'Haultfoeuille, X. and P. Février (2015): "Identification of Nonseparable Triangular Models With Discrete Instruments," *Econometrica*, 83, 1199–1210. 1.1

Dong, Y. and Y.-Y. Lee (2023): "Nonparametric Doubly Robust Identification of Causal Effects of a Continuous Treatment using Discrete Instruments," . 6

Gale, D. and H. Nikaido (1965): "The Jacobian matrix and global univalence of mappings," *Mathematische Annalen*, 159, 81–93. 3.1, 3.2, 3.2, G.1

Ghanem, D., D. Kédagni, and I. Mourifié (2023): "Evaluating the Impact of Regulatory Policies on Social Welfare in Difference-in-Difference Settings," . 1.1

Hadamard, J. (1906): "Sur les transformations ponctuelles," *Bull. Soc. Math. France*, 34, 71–84. 16

Han, S. and S. Lee (2019): "Estimation in a generalization of bivariate probit models with dummy endogenous regressors," *Journal of Applied Econometrics*, 34, 994–1015. 1.1

——— (2023): "Semiparametric Models for Dynamic Treatment Effects and Mediation Analyses with Observational Data," *University of Bristol and Sogang University*. 1.1

Han, S. and E. J. Vytlacil (2017): "Identification in a generalization of bivariate probit models with dummy endogenous regressors," *Journal of Econometrics*, 199, 63–73. 1.1, B, C, G.1

Han, S. and H. Xu (2023): "A note on model restrictions and identification power of the monotonicity condition," *U of Bristol, UT Austin*. A.1

Han, S. and S. Yang (2023): "A Computational Approach to Identification of Treatment Effects for Policy Evaluation," *arXiv preprint arXiv:2009.13861*. A.2

Heckman, J., J. L. Tobias, and E. Vytlacil (2003): "Simple Estimators for Treatment Parameters in a Latent-Variable Framework," *The Review of Economics and Statistics*, 85, 748–755. A.2

Heckman, J. J. and E. Vytlacil (2005): "Structural equations, treatment effects, and econometric policy evaluation1," *Econometrica*, 73, 669–738. 1.1

Heckman, J. J. and E. J. Vytlacil (2007): "Chapter 71 Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments," Elsevier, vol. 6 of *Handbook of Econometrics*, 4875–5143. 3.2, 3.2, 3.3, A.2, A.2

Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475. 1, 1.1, 3.1, 4.1.1, 4.1.2, 6, A.1, A.2

Imbens, G. W. and W. K. Newey (2009): "Identification and estimation of triangular simultaneous equations models without additivity," *Econometrica*, 77, 1481–1512. 1, 1.1, 3.4, 6, A.3

Manski, C. F. (1990): "Nonparametric bounds on treatment effects," *The American Economic Review*, 80, 319–323. 1

Mogstad, M., A. Santos, and A. Torgovitsky (2018): "Using instrumental variables for inference about policy relevant treatment parameters," *Econometrica*, 86, 1589–1619. A.2

Mourifié, I. and Y. Wan (2021): "Layered policy analysis program evaluation using the marginal treatment effect," Tech. rep., cemmap working paper. 1.1

Newey, W. and S. Stouli (2021): "Control variables, discrete instruments, and identification of structural functions," *Journal of Econometrics*, 222, 73–88. 1.1, A.3

Newey, W. K. and J. L. Powell (2003): "Instrumental variable estimation of nonparametric models," *Econometrica*, 71, 1565–1578. 1, 1.1

Newey, W. K., J. L. Powell, and F. Vella (1999): "Nonparametric estimation of triangular simultaneous equations models," *Econometrica*, 67, 565–603. 1.1

Torgovitsky, A. (2010): "Identification and Estimation of Nonparametric Quantile Regressions with Endogeneity," . 3.4, 6, A.4, 14, A.4

——— (2015): "Identification of Nonseparable Models Using Instruments With Small Support," *Econometrica*, 83, 1185–1197. 1.1, 14

Vuong, Q. and H. Xu (2017): "Counterfactual mapping and individual treatment effects in nonseparable models with binary endogeneity," *Quantitative Economics*, 8, 589–610. 1.1, 6, A.1, A.1, E.2, E.2

Vytlacil, E. (2002): "Independence, monotonicity, and latent index models: An equivalence result," *Econometrica*, 70, 331–341. 3.1, 4.1.1, A.1

——— (2006): "Ordered discrete-choice selection models and local average treatment effect assumptions: Equivalence, nonequivalence, and representation results," *The Review of Economics and Statistics*, 88, 578–581. A.2

Wüthrich, K. (2020): "A Comparison of Two Quantile Models With Endogeneity," *Journal of Business & Economic Statistics*, 38, 443–456. A.2

Appendix to

"Estimating Causal Effects of Discrete and Continuous Treatments with Binary Instruments"

A Comparisons to Previous Studies

Here we compare the proposed LGR-based identification approach to the literature and show that it can unify and complement existing methods. To simplify the exposition, we will abstract from covariates.

A.1 Rank Similarity and Rank Invariance in Chernozhukov and Hansen (2005) and Vuong and Xu (2017)

The instrumental variables quantile regression (IVQR) model of Chernozhukov and Hansen (2005) provides an alternative set of conditions under which the QSF_τ is point identified, provided that the outcome is continuous. The IVQR model is based on the Skorohod representation

$$Y_d = Q_{Y_d}(U_d), \quad U_d \sim U[0,1], \quad d \in \{0,1\}.$$

Chernozhukov and Hansen (2005) consider a general selection mechanism, $D = \delta(Z, V)$, where V can be vector-valued and the instrument is assumed to satisfy $U_d \perp\!\!\!\perp Z$ (which is implied by Assumption EX). The key assumption of the IVQR model is RS, $U_1 \overset{d}{=} U_0 \mid Z, V$. RS weakens the classical RI assumption, which requires $U_1 = U_0$ almost surely. The IVQR model yields the following conditional moment restriction (Chernozhukov and Hansen, 2005, Theorem 1),

$$\tau = \Pr[Y_1 \leq Q_{Y_1}(\tau), D_z = 1 | Z = z] + \Pr[Y_0 \leq Q_{Y_0}(\tau), D_z = 0 | Z = z], \quad z \in \{0,1\}. \quad \text{(A.1)}$$

Under the general selection model (2.1) and the independence assumption, equation (A.1) can be rewritten as

$$\begin{aligned}\tau &= \Pr[Y_1 \leq Q_{Y_1}(\tau), V_z \leq \pi(z) | Z = z] + \Pr[Y_0 \leq Q_{Y_0}(\tau), V_z > \pi(z) | Z = z] \\ &= \Pr[Y_1 \leq Q_{Y_1}(\tau), V_z \leq \pi(z) | Z = z] + \tau - \Pr[Y_0 \leq Q_{Y_0}(\tau), V_z \leq \pi(z) | Z = z].\end{aligned}$$

Thus, (A.1) can alternatively be written as

$$\Pr[Y_1 \leq Q_{Y_1}(\tau), V_z \leq \pi(z)|Z = z] = \Pr[Y_0 \leq Q_{Y_0}(\tau), V_z \leq \pi(z)|Z = z], \quad z \in \{0, 1\}.$$

Using Lemma 2.1, we can further rewrite it as

$$C(\tau, \pi(z), \rho_{Y_1,V_z;Z}(Q_{Y_1}(\tau), \pi(z); z)) = C(\tau, \pi(z), \rho_{Y_0,V_z;Z}(Q_{Y_0}(\tau), \pi(z); z)), \quad z \in \{0, 1\}.$$

This shows that the IVQR model also relies on a copula invariance assumption,

$$\rho_{Y_1,V_z;Z}(Q_{Y_1}(\tau), \pi(z); z) = \rho_{Y_0,V_z;Z}(Q_{Y_0}(\tau), \pi(z); z), \quad z \in \{0, 1\}. \tag{A.2}$$

The IVQR model imposes a restriction across potential outcomes, whereas CI is a restriction between the outcome and the unobservable component in the selection equation. To make this more explicit, it is instructive to rewrite equation (A.2) using the "counterfactual mapping" $\phi(y) \equiv Q_{Y_1}(F_{Y_0}(y))$ of Vuong and Xu (2017) as

$$\rho_{Y_1,V_z;Z}(\phi(y), \pi(z); z) = \rho_{Y_0,V_z;Z}(y, \pi(z); z), \quad z \in \{0, 1\}.$$

Note that the copula invariance assumption underlying the IVQR model restricts treatment effect heterogeneity, unlike CI.

The conditional moment restriction (A.1) and the implied copula invariance assumption are not sufficient for point identification. Chernozhukov and Hansen (2005) provide a set of sufficient conditions for point identification that require the Jacobian of (A.1) to be of full rank and continuous. For the case where D and Z are binary, Vuong and Xu (2017) provide weaker conditions by directly analyzing identification of the counterfactual mapping ϕ, which satisfies $Y_1 = \phi(Y_0)$ under continuity of the outcome and RI. Their key identification

condition is piecewise strict monotonicity of

$$\Delta_d(\cdot) = (-1)^d(\Pr[Y \leq \cdot, D = d|Z = 0] - \Pr[Y \leq \cdot, D = d|Z = 1]).$$

The approach of Vuong and Xu (2017) is extended to the case of ordered treatments in Han and Xu (2023).

In sum, both the IVQR model and our approach rely on copula invariance assumptions to identify causal effects for the overall population using instruments. Relative to the IVQR model, the proposed identification approach has two main advantages. First, it allows for unrestricted treatment effect heterogeneity (see Section 4.2). Second, it does not rely on continuity of the outcome to achieve point identification, and it naturally accommodates discrete and mixed discrete-continuous outcomes.

Remark A.1 (RS Restricts Treatment Effect Heterogeneity)**.** *Consider the setting of Section 4.2 where we impose the classical LATE assumptions,* $\tilde{V}_1 = \tilde{V}_0 = \tilde{V}$ *and* $(Y_1, Y_0, \tilde{V}) \perp\!\!\!\perp Z$ *(Imbens and Angrist, 1994; Vytlacil, 2002). Under these assumptions, by NM,*

$$\begin{pmatrix} \tilde{Y}_0 \\ \tilde{Y}_1 \\ V \end{pmatrix} \mid Z = z \sim \mathcal{N}_3 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{01} & \rho_{0V} \\ \rho_{01} & 1 & \rho_{1V} \\ \rho_{0V} & \rho_{1V} & 1 \end{pmatrix} \right).$$

CI holds by construction. CI therefore does not restrict treatment effect heterogeneity since ρ_{01} *and hence the relationship between* $\tilde{Y}_1$ *and* $\tilde{Y}_0$ *is completely unrestricted. RS imposes that* $\rho_{0V} = \rho_{1V} = \rho_V$*, which restrict treatment effect heterogeneity. Thus, the requirement that*

$$\Sigma \equiv \begin{pmatrix} 1 & \rho_{01} & \rho_V \\ \rho_{01} & 1 & \rho_V \\ \rho_V & \rho_V & 1 \end{pmatrix}$$

is a variance-covariance matrix and hence positive semi-definite poses implicit restrictions

on ρ_{01}*. Specifically, all three eigenvalues of* Σ*,*

$$\lambda_1 = 1 - \rho_{01}, \quad \lambda_2 = \frac{1}{2}\left(-\sqrt{\rho_{01}^2 + 8\rho_V^2} + \rho_{01} + 2\right), \quad \textit{and } \lambda_3 = \frac{1}{2}\left(\sqrt{\rho_{01}^2 + 8\rho_V^2} + \rho_{01} + 2\right),$$

need to be positive. While λ_1 *and* λ_3 *are always positive, the requirement that* $\lambda_2 \geq 0$ *is equivalent to* $\rho_{01} \geq 2\rho_V^2 - 1$*. This shows that RS imposes restrictions on* ρ_{01} *and thus on effect heterogeneity. These restrictions crucially depend on the degree of endogeneity, measured by* ρ_V*. There are two polar cases: (i) if the treatment is exogenous so that* $\rho_V = 0$*, then RS does not imply any restrictions on* ρ_{01}*; (ii) if the endogeneity is "maximal" so that* $\rho_V = 1$*, then* $\rho_{01} = 1$ *(as under RI). In other words, the stronger the endogeneity the "closer" RS becomes to RI.*

A.2 LATE Monotonicity in Imbens and Angrist (1994)

The LATE framework of Imbens and Angrist (1994) provides conditions for identifying causal effects for the subpopulation of compliers. The compliers are the individuals who react to the instrument so that $D_1 \geq D_0$ almost surely. Compared to the assumptions in Section 3.1, the LATE framework restricts the selection model by setting $V_1 = V_0 = V$ almost surely, and relies a stronger joint independence assumption, $(Y_0, Y_1, V) \perp\!\!\!\perp Z$, but does not impose any copula invariance assumption.

Under the LATE assumptions, the marginal potential outcome distributions for the compliers are identified (e.g., Abadie, 2002, Lemma 2.1):

$$F_{Y_1|D_1 \geq D_0}(y) = \frac{E[1\{Y \leq y\}D|Z=1] - E[1\{Y \leq y\}D|Z=0]}{E[D|Z=1] - E[D|Z=0]}$$
$$F_{Y_0|D_1 \geq D_0}(y) = \frac{E[1\{Y \leq y\}(1-D)|Z=1] - E[1\{Y \leq y\}(1-D)|Z=0]}{E[1-D|Z=1] - E[1-D|Z=0]}$$

The main drawback of the LATE framework is its lack of external validity: it only identifies effects for a subpopulation that depends on the instrument. Different approaches exist for extrapolating from the LATE to externally valid global treatment effects. Examples include

approaches based on structural models (e.g., Heckman et al., 2003), covariates (e.g., Angrist and Fernández-Val, 2013), rank similarity (e.g., Wüthrich, 2020), and the smoothness of the marginal treatment effect function (e.g., Mogstad et al., 2018; Han and Yang, 2023).

The identification results in Section 3.1 apply under the LATE assumptions augmented with Assumption CI. Therefore, the proposed method provides a new approach to external validity based on copula invariance assumptions. It has important practical advantages relative to the existing methods referenced above. It avoids parametric structural assumptions, does not require covariates, accommodates discrete outcomes, and does not resort to partial identification.

For ordered treatments, Remark 3.3 discusses Heckman and Vytlacil (2007), which is a special case of our model. Vytlacil (2006) considers the ordered selection model with random thresholds:

$$D_z = \begin{cases} 1, & -\infty < v(z) \leq V_1 \\ 2, & V_1 < v(z) \leq V_2 \\ & \\ K, & V_{K-1}(z) < v(z) \leq \infty \end{cases}$$

where $V_1, \ldots, V_{K-1}$ are random thresholds such that $V_1 \leq \cdots \leq V_{K-1}$ and assumes $Z \perp (V_1, \ldots, V_{K-1}, Y_1, \ldots, Y_K)$. Under the normalization $V_z \sim N(0,1)$, our model has a similar representation:

$$D_z = \begin{cases} 1, & -\infty < v(z) \leq V_1(z) \\ 2, & V_1(z) < v(z) \leq V_2(z) \\ & \\ K, & V_{K-1}(z) < v(z) \leq \infty \end{cases}$$

upon setting $V_k(z) = -V_z + \pi_k(z) + v(z)$. This representation looks more general because

the random thresholds can vary with z. To get the same representation, we can impose in our model that $V_0 = V_1 = V$ and $\pi_k(z) = \pi_k - v(z)$, and set $V_k = -V + \pi_k$, but this version of the model is more restrictive than Vytlacil's model because it imposes restrictions on the joint distribution of $V_1, \ldots, V_{K-1}$. More generally, our identification analysis carries over with suitable adjustments if we allow for threshold-specific unobservables, that is

$$D_z = \begin{cases} 1, & -\infty \leq \tilde{V}_{1,z} \leq \pi_1(z) \\ 2, & \pi_1(z) < \tilde{V}_{2,z} \leq \pi_2(z) \\ \vdots & \vdots \\ K, & \pi_{K-1}(z) < \tilde{V}_{K,z} \leq \infty \end{cases}. \tag{A.3}$$

where $\tilde{V}_{1,z} \leq \cdots \leq \tilde{V}_{K,z}$. This model is more general than Vytlacil's model. It is equivalent when $V_{k,1} = V_{k,0}$ for all k and $\pi_k(z) = \pi_k - v(z)$, upon setting $V_k = -\tilde{V}_k + \pi_k$.

Remark A.2. *We continue Remark 3.3 and discuss Heckman and Vytlacil (2007)'s model in comparison to ours using the notation introduced in the proof of Theorem 3.2. The cutoffs of the two models are related as $\pi_\ell(z) = \pi_\ell - \mu(z)$. Under this model, we have*

$$r_\ell(z) = \frac{\pi_\ell - \mu(z) - \rho_{d,y} F_{d,y}}{\sqrt{1 - \rho_{d,y}^2}}$$

which is particularly easy-to-interpret because

$$r_d(z) - r_{d-1}(z) = \frac{\pi_d - \pi_{d-1}}{\sqrt{1 - \rho_{d,y}^2}}.$$

On the other hand, in the general model with the normalization $V_z \mid Z \sim N(0,1)$, we have

$$r_d(z) - r_{d-1}(z) = \frac{\pi_d(z) - \pi_{d-1}(z)}{\sqrt{1 - \rho_{d,y}^2}}.$$

Note that in the simplified model, the full-rank condition holds by construction. Specifically,

since $r_d(z) - r_{d-1}(z)$ *does not depend on* z*, we can write* $r_d(z) = r_{d-1}(z) + c$ *for* $c > 0$*. Therefore, we can rewrite* $\lambda(z)$ *as*

$$\frac{\phi(r_{d-1}(z) + c) - \phi(r_{d-1}(z))}{\Phi(r_{d-1}(z) + c) - \Phi(r_{d-1}(z))}.$$

This function is monotonically decreasing in $r_{d-1}(z)$*. Thus, as long as* $\mu(1) \neq \mu(0)$ *so that* $r_{d-1}(1) \neq r_{d-1}(0)$*, the full rank condition holds.*

A.3 Control Function Approach in Imbens and Newey (2009)

For a continuous treatment, Imbens and Newey (2009) consider identification based on a control function approach. A simple version of their model consists of a structural outcome equation, $Y = g(D, \varepsilon)$, and a reduced form treatment assignment equation, $D = h(Z, V)$, where V is scalar and $v \mapsto h(\cdot, v)$ is strictly monotone. The main idea is to use $V = F_{D|Z}(D \mid Z)$ as a control function that satisfies $D \perp\!\!\!\perp \varepsilon \mid V$. Under the assumption that $(\varepsilon, V) \perp\!\!\!\perp Z$, the latter holds. The key to their identification approach is a common support assumption that the support of V conditional on D equals the support of V, which potentially requires a large support of Z. This assumption and the control function assumption yield

$$E[Y \mid D = d, V = v] = \int g(d, e) dF_{\varepsilon|D,V}(e \mid d, v) = \int g(d, e) dF_{\varepsilon|V}(e \mid v).$$

In terms of our framework, while they require $V_1 = V_0 = V$ almost surely in (2.1) and strict monotonicity with respect to V. We also do not require Z to have a large support and allow for binary Z. On the other hand, we assume CI as a trade-off.

To avoid the large support assumption of Imbens and Newey (2009), Newey and Stouli (2021) impose parametric structure for extrapolation. Their main assumptions are $E[Y \mid D, V] = \beta' p(D) \otimes q(V)$ where $p(\cdot)$ and $q(\cdot)$ are known and $E[(p(D) \otimes q(V))(p(D) \otimes q(V))']$ is positive definite. Then binary IV is enough to identify the average structural function. They also generalize the result for the case where either $p(\cdot)$ or $q(\cdot)$ is known. While their

approach imposes parametric structure on the conditional mean, we impose copula invariance, restricting the dependence structure of the unobservables.

A.4 Conditional Copula Invariance in Torgovitsky (2010)

For a continuous treatment, Torgovitsky (2010) considers identification based on a conditional copula invariance assumption.[14] He assumes that Y and D are continuous and considers the model, $Y = m(D, U)$, where $u \mapsto m(d, u)$ is strictly increasing for every d, which implies RI. The (possibly binary) instrument Z is assumed to be marginally independent of U, $U \perp\!\!\!\perp Z$, which is implied by EX. Moreover, he imposes a weak local dependence assumption between D and Z.

The key condition of Torgovitsky (2010) is the conditional copula invariance assumption. Let $R \equiv F_{D|Z}(D \mid Z)$ and consider $\Pr[U \leq u, D \leq Q_{D|Z}(r \mid z) \mid Z = z]$. Then the assumption requires that the copula of $(U, D) \mid Z = 1$ is equal to the copula of $(U, D) \mid Z = 0$ (focusing on binary Z). Under $U \perp\!\!\!\perp Z$, this can be written as

$$\tilde{C}(F_U(u), r; 1) = \tilde{C}(F_U(u), r; 0). \tag{A.4}$$

Using Lemma 2.1, these two copulas have the following LGR: for $z \in \{0, 1\}$,

$$\tilde{C}(F_U(u), r; z) = C(F_U(u), r; \rho_{U,D;Z}(F_U(u), r; z)).$$

Hence, the conditional copula invariance assumption (A.4) can be written as

$$\rho_{U,D;Z}(F_U(u), r; 1) = \rho_{U,D;Z}(F_U(u), r; 0). \tag{A.5}$$

Comparing equation (A.5) to CI, we can see that both copula invariance assumptions restrict the dependence of the joint distribution of (Y_d, D) on Z by requiring the correlation parameter

[14] Portions of Torgovitsky (2010) were published in Torgovitsky (2015). Since the role of the conditional copula invariance assumption is only discussed in Torgovitsky (2010), we focus on comparing our method and assumptions to this paper.

not to depend on $Z = z$. Since Torgovitsky (2010) maintains the RI assumption, restricting the copula of (U, D) is sufficient. Our identification strategy does not depend on RI such that we need to impose copula invariance restrictions for both potential outcomes. As a trade-off of not assuming RI, we require copula invariance that the correlation parameter is not a function of the reduced-form parameter.

Overall, our identification results complement Torgovitsky (2010) by showing that copula invariance assumptions also are useful with binary treatments. While the underlying copula invariance assumptions are related, our identification strategy fundamentally differs from Torgovitsky (2010). It accommodates binary and ordered treatments and does not rely in RI assumptions.

B Local Representation with Other Copulas

It may be the case that a joint distribution can be locally represented using a copula that is not necessarily Gaussian. We ask what other single-parameter copulas can be used for local representation. Consider the case of binary D. In showing the full rank of Jacobian in the identification proof that employs Hadamard's global inverse function theorem, the following quantity typically arises once the Jacobian is transformed using elementary operations:

$$\frac{C_\rho(F_{d,y}, \pi(z); \rho)}{C_1(F_{d,y}, \pi(z); \rho)} - \frac{C_\rho(F_{d,y}, \pi(z'); \rho)}{C_1(F_{d,y}, \pi(z'); \rho)},$$

where C_ρ and C_1 are the derivatives w.r.t. ρ and the first argument, respectively. Therefore, any copula that satisfies

$$\frac{C_\rho(F_{d,y}, \pi(z); \rho)}{C_1(F_{d,y}, \pi(z); \rho)} \neq \frac{C_\rho(F_{d,y}, \pi(z'); \rho)}{C_1(F_{d,y}, \pi(z'); \rho)} \tag{B.1}$$

for $\pi(z) \neq \pi(z')$ will yield the full rank Jacobian. Han and Vytlacil (2017) show that any single-parameter copula that follows the ordering of stochastic increasingness w.r.t. ρ satisfies (B.1). Therefore, among them, a comprehensive copula can be a candidate for the

local representation. The following Archimedean copulas are such copulas:

Example 1 (Clayton copula)**.**

$$C(u_1, u_2; \rho) = \max\{u_1^{-\rho} + u_2^{-\rho} - 1, 0\}^{-1/\rho}, \quad \rho \in [-1, \infty)\backslash\{0\}$$

and $C \to C_L$ *when* $\rho \to -1$, $C \to C_U$ *when* $\rho \to \infty$, *and* $C \to C_I$ *when* $\rho \to 0$.

Example 2 (Frank copula)**.**

$$C(u_1, u_2; \rho) = -\frac{1}{\rho} \ln\left(1 + \frac{(e^{-\rho u_1} - 1)(e^{-\rho u_2} - 1)}{e^{-\rho} - 1}\right), \quad \rho \in (-\infty, \infty)\backslash\{0\}$$

and $C \to C_L$ *when* $\rho \to -\infty$, $C \to C_U$ *when* $\rho \to \infty$, *and* $C \to C_I$ *when* $\rho \to 0$.

C Identification Power of IV Support

When the instrument takes more than two values, we may be able to relax Assumption CI. We illustrate this focusing on the case of binary D. Let $\boldsymbol{Z} \equiv (Z_1, ..., Z_K)$ be the vector of binary IVs, that is, $Z_k \in \{0, 1\}$ for $k = 1, ..., K$. This vector might arise from having multiple binary instruments or constructing indicators from a multivalued instrument.

We focus on the binary treatment case. Define a selection equation

$$D_{\boldsymbol{z}} = 1\{V_{\boldsymbol{z}} \leq \pi(\boldsymbol{z})\}, \tag{C.1}$$

where $\pi(\boldsymbol{z}) \equiv \Pr[D = 1 \mid \boldsymbol{Z} = \boldsymbol{z}]$. We make the following assumptions.

Assumption EX2. *For* $d, z_k \in \{0, 1\}$ *for all* k, $\boldsymbol{Z} \perp\!\!\!\perp Y_d$ *and* $\boldsymbol{Z} \perp\!\!\!\perp V_{\boldsymbol{z}}$.

Assumption CI2 (Partial Copula Invariance)**.** *For* $d \in \{0, 1\}$, $\rho_{d,y}(0, ..., 0, 1) = \rho_{d,y}(0, ..., 0, 0) \equiv \rho^0_{d,y}$, *where* $\rho_{d,y}(\boldsymbol{z}) \equiv \rho_{Y_d, V_{\boldsymbol{z}}}(y, \pi(\boldsymbol{z}))$.

Assumption REL2. *(i)* $\boldsymbol{Z} \in \{0,1\}^K$*; (ii)* $0 < \Pr(\boldsymbol{Z} = \boldsymbol{z}) < 1$ *and* $0 < \Pr(D = d \mid \boldsymbol{Z} = \boldsymbol{z}) < 1$*, for* $d \in \mathcal{D}$ *and* $\boldsymbol{z} \in \{(0,...,0,0),(0,...,0,1)\}$*; and (iii)* $\Pr(D = d \mid \boldsymbol{Z} = (0,...,0,0)) \neq \Pr(D = d \mid \boldsymbol{Z} = (0,...,0,1))$ *for* $d \in \mathcal{D}$.

EX2 is the analog of EX in the multiple instrument case. CI2 can be justified if, conditional on $Z_k = 0$ for $k = 1,...,K-1$ (i.e., the status quo), Z_K does not shift the joint distribution of $(Y_d, V_{\boldsymbol{z}})$. In general, with the vector of IVs, there will always be only one more parameter than the number of identifying equations, which is 2^K. CI2 reduces this additional parameter. For illustration, let $K = 2$, that is, consider two binary IVs, Z_1 and Z_2, in $\{0,1\}$. Then, the resulting equations for $D = 1$ are

$$
\begin{aligned}
F_{Y|D,\boldsymbol{Z}}(y \mid 1,(1,1))\pi(1,1) &= C(F_{Y_1}(y), \pi(1,1); \rho_{1,y}(1,1)),\\
F_{Y|D,\boldsymbol{Z}}(y \mid 1,(1,0))\pi(1,0) &= C(F_{Y_1}(y), \pi(1,0); \rho_{1,y}(1,0)),\\
F_{Y|D,\boldsymbol{Z}}(y \mid 1,(0,1))\pi(0,1) &= C(F_{Y_1}(y), \pi(0,1); \rho^0_{1,y}),\\
F_{Y|D,\boldsymbol{Z}}(y \mid 1,(0,0))\pi(0,0) &= C(F_{Y_1}(y), \pi(0,0); \rho^0_{1,y}),
\end{aligned}
$$

where there are four unknowns: $(F_{Y_1}(y), \rho_{1,y}(1,1), \rho_{1,y}(1,0), \rho^0_{1,y})$. Then we can show that the corresponding Jacobian has full rank as long as

$$
\frac{C_\rho(F_{Y_1}(y), \pi(0,1); \rho^0_{1,y})}{C_1(F_{Y_1}(y), \pi(0,1); \rho^0_{1,y})} \neq \frac{C_\rho(F_{Y_1}(y), \pi(0,0); \rho^0_{1,y})}{C_1(F_{Y_1}(y), \pi(0,0); \rho^0_{1,y})},
$$

which is guaranteed since $\pi(0,1) \neq \pi(0,0)$ by REL2, and Lemma 4.1 in Han and Vytlacil (2017).[15] For general K, we can prove that the Jacobian of the corresponding system of

[15]Note that in proving full rank of the Jacobian, only $\pi(0,1) \neq \pi(0,0)$ is used. However, $\pi(z_1,z_2) \neq \pi(z_1',z_2')$ for any $(z_1,z_2) \neq (z_1',z_2')$ are fully utilized to generate the four equations above, because otherwise we have less equations than unknown.[IFV: I do not think this is needed. $F_{Y_1}(y)$ and $\rho^0_{1,y}$ are identified from the last 2 equations using the same argument as in the binary instrument case. $\rho_{1,y}(1,1)$ and $\rho_{1,y}(1,0)$ are then identified from the first 2 equations without additional restrictions]

equations for $D = 1$ has full rank as long as

$$\frac{C_\rho(F_{Y_1}(y), \pi(0, ..., 0, 1); \rho^0_{1,y})}{C_1(F_{Y_1}(y), \pi(0, ..., 0, 1); \rho^0_{1,y})} \neq \frac{C_\rho(F_{Y_1}(y), \pi(0, ..., 0, 0); \rho^0_{1,y})}{C_1(F_{Y_1}(y), \pi(0, ..., 0, 0); \rho^0_{1,y})},$$

which is guaranteed by $\pi(0, ..., 0, 1) \neq \pi(0, ..., 0, 0)$. Then we identify $2^K \times 1$ vector $(F_{Y_1}(y), \{\rho_{1,y}(\boldsymbol{z}) : \boldsymbol{z}_{-K} \neq \mathbf{0}\}, \rho^0_{1,y})$. A desirable aspect is that no matter how large the system is (i.e., how large 2^K is), the proof of full rank always amounts to checking the ratio of copula derivatives between the two groups defined by the last instrument Z_K given $\boldsymbol{Z}_{-K} = (0, ..., 0)$, the status quo.

Let $\mathcal{V}_{\boldsymbol{z}}$ denote the support of $\pi(\boldsymbol{z})$. The following theorem gathers the identification result:

Theorem C.1 (Identification Binary Treatment with Multiple Instruments)**.** *Suppose* $D_{\boldsymbol{z}} \in \{0, 1\}$ *satisfies* (C.1) *for* $\boldsymbol{z} \in \{0, 1\}^K$*. Under EX2, REL2, and CI2, the functions* $y \mapsto F_{Y_d}(y)$ *and* $(y, v) \mapsto \rho_{Y_d, V_{\boldsymbol{z}}}(y, v)$ *are identified on* $y \in \mathcal{Y}$ *and* $(y, v) \in \mathcal{Y} \times \mathcal{V}_{\boldsymbol{z}}$*, respectively, for* $d \in \{0, 1\}$*.*

CI2 may become innocuous with large K, because within a finer cell (defined by $\boldsymbol{Z}$), individuals tend to be homogeneous and thus share the same joint distribution of $(Y_1, V_{\boldsymbol{z}})$, justifying the copula invariance. The trade-off is that in this case instruments may be weak (i.e., $\pi(0, ..., 0, 1) \approx \pi(0, ..., 0, 0)$) for the same reason. Therefore, a large K may not necessarily be preferred. Finally, note that Z_k being binary is not essential and we can have discrete or continuous Z_k with different supports across instruments.

D Identification with Covariates

In this section, we extend our main identification analyses to the case where covariates X present. We focus on binary and continuous D. The case with ordered D is analogous to that with binary D. Consider (potentially endogenous) covariate $X \in \mathcal{X}$.

Assumption EX3 (Conditional Independence)**.** *For $d \in \mathcal{D}$ and $z \in \{0,1\}$, $Z \perp\!\!\!\perp Y_d \mid X$ and $Z \perp\!\!\!\perp V_z \mid X$.*

Assumption REL3 (Relevance)**.** *(i) $Z \in \{0,1\}$; (ii) $0 < \Pr(Z = 1 \mid X) < 1$, almost surely; and (iii) for $\mathcal{D} = \{0,1\}$, $\Pr(D = 1 \mid Z = 1, X) \neq \Pr(D = 1 \mid Z = 0, X)$ and $0 < \Pr(D = 1 \mid Z = z, X) < 1$ almost surely, for $z \in \{0,1\}$; and, for uncountable $\mathcal{D}$, $F_{D|Z,X}(d \mid 1, X) \neq F_{D|Z}(d \mid 0, X)$ and $0 < F_{D|Z,X}(d \mid z, X) < 1$ almost surely, for $(z, d) \in \{0,1\} \times int(\mathcal{D})$.*

Assumption CI3 (Conditional Copula Invariance)**.** *For $d \in \mathcal{D}$, $\rho_{Y_d,V_z;Z,X}(y, v; z, x)$ is a constant function of (v, z), that is*

$$\rho_{Y_d,V_z;Z,X}(y, v; z, X) = \rho_{Y_d;X}(y; X), \text{ almost surely, } \quad (y, v, z) \in \mathcal{Y} \times \mathcal{V} \times \{0,1\}.$$

Note that copula invariance is allowed to hold conditional on covariates. Therefore, we allow for observed heterogeneity in the dependence structure.

Remark D.1. *CI3 and CI2 are complementary. Which one to impose depends on the plausibility in given applications. On the one hand, CI3 imposes invariance for every subgroup defined by $X = x$, whereas CI2 imposes invariance for a single subgroup defined by $\boldsymbol{Z}_{-K} = (0, ..., 0)$. On the other hand, CI2 imposes stronger exclusion restrictions.*

We show the identifiability of $F_{d,y}(x) \equiv F_{Y_d|X}(y \mid x)$, from which we can construct conditional parameters:

$$QSF_\tau(d; x) \equiv Q_{Y_d|X}(\tau|x) = \mathcal{Q}_\tau(F_{Y_d|X}(\cdot|x)),$$
$$ASF(d; x) \equiv E[Y_d|X = x] = \mathcal{E}(F_{Y_d|X}(\cdot|x)).$$

Marginal QSF_τ and ASF are also identified from

$$F_{Y_d}(y) = \int F_{Y_d|X}(y \mid x) \mathrm{d}F_X(x),$$

where F_X is the distribution of X.

D.1 Binary Treatment

Define a selection equation

$$D_z = 1\{V_z \leq \pi(z, X)\}, \tag{D.1}$$

where $\pi(z, x) \equiv \Pr[D = 1 \mid Z = z, X = x]$. Consider

$$\begin{aligned}\Pr[Y \leq y, D = 1 \mid Z = z, X = x] &= \Pr[Y_1 \leq y, V_z \leq \pi(z, x) \mid Z = z, X = x] \\ &= C(F_{Y_1|X}(y \mid x), \pi(z, x); \rho_{Y_1;X}(y; x)), \quad z \in \{0, 1\},\end{aligned}$$

where the last equation is by EX3 and CI3. Now, let $F_{d,y}(x) \equiv F_{Y_d|X}(y \mid x)$, and $\rho_{d,y}(x) \equiv \rho_{Y_d}(y; x)$. Then, we have the system of two equations

$$\begin{aligned}\Pr[Y \leq y, D = 1 \mid Z = 1, X = x] &= C(F_{1,y}(x), \pi(1, x); \rho_{1,y}(x)), \\ \Pr[Y \leq y, D = 1 \mid Z = 0, X = x] &= C(F_{1,y}(x), \pi(0, x); \rho_{1,y}(x)),\end{aligned}$$

with two unknowns for every $x \in \mathcal{X}$: $(F_{1,y}(x), \rho_{1,y}(x))$. This system has full rank if

$$\frac{C_\rho(F_{1,y}(x), \pi(1, x); \rho_{1,y}(x))}{C_1(F_{1,y}(x), \pi(1, x); \rho_{1,y}(x))} \neq \frac{C_\rho(F_{1,y}(x), \pi(0, x); \rho_{1,y}(x))}{C_1(F_{1,y}(x), \pi(0, x); \rho_{1,y}(x))}$$

for all x, which is guaranteed by REL3.

The following theorem gathers the identification result:

Theorem D.1 (Identification Binary Treatment with Covariates)**.** *Suppose $D_z \in \{0, 1\}$ satisfies (D.1) for $z \in \{0, 1\}$. Under EX3, REL3, and CI3, the functions $(y, x) \mapsto F_{Y_d|X}(y \mid x)$ and $y \mapsto \rho_{Y_d;X}(y; x)$ are identified on $(y, x) \in \mathcal{Y} \times \mathcal{X}$, for $d \in \{0, 1\}$.*

The proof of Theorem D.1 is omitted because it is similar to the proof of Theorem 3.1.

D.2 Continuous Treatment

Let $F_{d,y}(x) \equiv F_{Y_d|X}(y \mid x)$ and $\pi_{z,d}(x) \equiv F_{D_z|X}(d \mid x)$. Assume that $d \mapsto F_{D_z|X}(d \mid X)$ is strictly increasing almost surely.

For the generalized selection, let

$$D_z = h(z, X, V_z), \tag{D.2}$$

where $V_z \equiv F_{D_z|X}(D_z \mid X)$ and $h(z, X, \cdot) \equiv F^{-1}_{D_z|X}(\cdot \mid X)$, so that $\pi_{z,\cdot}(x) = h^{-1}(z, x, \cdot)$. Also note that $V_z \perp\!\!\!\perp Z$ by Assumption EX3 and $v \mapsto h(z, x, v)$ is strictly increasing. Then, by EX3,

$$F_{Y_d,D_z|Z,X}(y, d \mid z, x) = C(F_{d,y}(x), \pi_{z,d}(x); \rho_{Y_d,V_z;Z,X}(y, \pi_{z,d}(x); z, x)).$$

By Assumption CI3, $\rho_{Y_d,V_z;Z,X}(y, \pi_{z,d}(x); z, x) \equiv \rho_{Y_d}(y; x) \equiv \rho_{d,y}(x)$. Then,

$$\begin{aligned} F_{Y|D,Z,X}(y \mid d, z, x) = F_{Y_d|D_z,Z,X}(y \mid d, z, x) &= \frac{(\partial/\partial d) F_{Y_d,D_z|Z,X}(y, d \mid z, x)}{(\partial/\partial d) F_{D_z|Z,X}(d \mid z, x)} \\ &\equiv \Phi\left(a_{d,y}(x) + b_{d,y}(x)\Phi^{-1}(\pi_{z,d}(x))\right) \end{aligned}$$

where

$$a_{d,y}(x) \equiv \Phi^{-1}(F_{d,y}(x))/\sqrt{1 - \rho^2_{d,y}(x)}, \text{ and } b_{d,y}(x) \equiv -\rho_{d,y}(x)/\sqrt{1 - \rho^2_{d,y}(x)}. \tag{D.3}$$

The argument from here is the same as in the case without covariates.

The following theorem gathers the identification result:

Theorem D.2 (Identification Continuous Treatment with Covariates)**.** *Suppose D_z, $z \in \{0, 1\}$, satisfies (D.2). Under EX3, REL3, and CI3, the functions $(y, x) \mapsto F_{Y_d|X}(y \mid x)$ and*

$(y,x) \mapsto \rho_{Y_d;X}(y;x)$ *are identified on* $(y,x) \in \mathcal{Y} \times \mathcal{X}$*, for* $d \in \mathcal{D}$ *by*

$$F_{Y_d|X}(y \mid x) = \Phi\left(\frac{a_{d,y}(x)}{\sqrt{1+b_{d,y}(x)^2}}\right), \quad \rho_{Y_d;X}(y;x) = \frac{-b_{d,y}(x)}{\sqrt{1+b_{d,y}(x)^2}},$$

where $a_{d,y}(x)$ *and* $b_{y,d}(x)$ *are defined in* (D.3).

E Alternative Identification Strategies

For the case of binary D, we show there can be alternative identification strategies using a version of copula invariance. The analysis can be extended to the ordered treatment case. Here we assume EX and REL and the treatment selection equation $D_z = 1[V_z \leq p(z)]$ where $V_z|Z = z \sim U[0,1]$. We consider strategies that use a subpopulation defined by each treatment level separately and strategies that combine the two subpopulations.

E.1 Restrictions Within Treatment Levels

We focus here on the level of $D = 1$. A similar analysis follows for $D = 0$. For $y \in \mathcal{Y}$, consider the LGR of the observed probabilities, that is

$$\Pr(Y_1 \leq y, D = 1 \mid Z = z) = C(F_{Y_1}(y), \pi(z); \rho_1(y, \pi(z); z)), \quad z \in \{0,1\},$$

where $\rho_d(y, \pi(z); z) \equiv \rho_{Y_d,V_z;Z}(y, \pi(z); z)$. The identification problem is that we have two probabilities to identify three parameters: $F_{Y_1}(y)$, $\rho_1(y, \pi(0); 0)$ and $\rho_1(y, \pi(1); 1)$. So far, we have reduced the number of parameters by imposing the condition:

$$\rho_1(y, \pi(0); 0) = \rho_1(y, \pi(1); 1).$$

This restriction is imposed separately for each value of y. However, it is also possible to impose restrictions across values of y. Assume that there exists $y' \in \mathcal{Y}$ be such that $F_{Y_1}(y) \neq F_{Y_1}(y')$

and

$$\rho_1(y, \pi(z); z) = \rho_1(y', \pi(z); z), \quad z \in \{0, 1\}. \tag{E.1}$$

This condition leads to the following system of four equations with four unknowns:

$$\begin{aligned}
\Pr(Y_1 \leq y, D = 1 \mid Z = z) &= C(F_{Y_1}(y), \pi(z); \rho_1(y, \pi(z); z)), \quad z \in \{0, 1\}, \\
\Pr(Y_1 \leq y', D = 1 \mid Z = z) &= C(F_{Y_1}(y'), \pi(z); \rho_1(y, \pi(z); z)), \quad z \in \{0, 1\}.
\end{aligned}$$

Then, it is possible to find conditions under which the solution to this system exists and is unique. This condition is appealing in that it does not imposes restrictions across levels of z.

Let C_j denote the partial derivative of the Gaussian copula C with respect to the jth argument, $j \in \{1, 2, 3\}$. The Jacobian of the system of equations is

$$J(y, y') = \begin{pmatrix}
C_1(y, 1) & 0 & C_3(y, 1) & 0 \\
C_1(y, 0) & 0 & 0 & C_3(y, 0) \\
0 & C_1(y', 1) & C_3(y', 1) & 0 \\
0 & C_1(y', 0) & 0 & C_3(y', 0)
\end{pmatrix},$$

where $C_j(k, z) := C_j(F_{Y_1}(k), \pi(z); \rho_1(k, \pi(z); z)) > 0$ for $j \in \{1, 3\}$, $k \in \{y, y'\}$ and $z \in \{0, 1\}$. By the Laplace expansion, the Jacobian determinant is

$$\det(J(y, y')) = C_1(y, 0)C_1(y', 1)C_3(y, 1)C_3(y', 0) - C_1(y, 1)C_1(y', 0)C_3(y', 1)C_3(y, 0),$$

which does not vanish if

$$\frac{C_3(y, 1)}{C_1(y, 1)} \frac{C_3(y', 0)}{C_1(y', 0)} \neq \frac{C_3(y, 0)}{C_1(y, 0)} \frac{C_3(y', 1)}{C_1(y', 1)}.$$

Let

$$\lambda(k, z) = \frac{\phi\left(u(k, z)\right)}{\Phi\left(u(k, z)\right)}, \quad u(k, z) := \frac{\Phi^{-1}(\pi(z)) - \rho_1(k, \pi(z); z)\Phi^{-1}(F_{Y_1}(k))}{\sqrt{1 - \rho_1(k, \pi(z); z)^2}},$$

where ϕ and Φ are the standard normal PDF and CDF, respectively. Then, using that $C_3(k,z) = \lambda(k,z)C_1(k,z)$, the previous condition can be expressed as

$$\frac{\lambda(y,1)}{\lambda(y,0)} \neq \frac{\lambda(y',1)}{\lambda(y',0)} \quad \text{or} \quad \frac{\lambda(y,1)}{\lambda(y',1)} \neq \frac{\lambda(y,0)}{\lambda(y',0)}, \tag{E.2}$$

that is, the change in the conditional inverse Mills ratio from $z = 0$ to $z = 1$ is different at y and y', or the change in the conditional inverse Mills ration from y to y' is different at $z = 0$ and $z = 1$. For example, if the identification condition holds locally for $y' = y + \mathrm{d}y$, then the condition becomes

$$\frac{\partial \log \lambda(y,1)}{\partial y} \neq \frac{\partial \log \lambda(y,0)}{\partial y}.$$

Theorem E.1. *Suppose $D \in \{0,1\}$ satisfies (3.1). Suppose Assumptions EX and REL holds. Given $y \in \mathcal{Y}$, suppose that there exists $y' \in \mathcal{Y}$ such that (E.1) and (E.2) hold. Then, $F_{Y_d}(y)$ and $\rho_{Y_d}(y)$ are identified for $d \in \{0,1\}$.*

In general, let d_y be the number of values of Y that we use to construct the system of equations. Then, we have $2d_y$ equations and $3d_y$ unknowns. Therefore, we need to reduce d_y parameters by whichever combinations of copula invariance (E.1) and the alternative assumptions.

E.2 Restrictions Between Treatment Levels

Alternative to the previous subsection, we can impose restrictions involving parameters for different treatment levels. This strategy is based on the system of equations

$$\begin{aligned}
\Pr(Y_1 \leq y, D = 1 \mid Z = z) &= C(F_{Y_1}(y), \pi(z); \rho_1(y, \pi(z); z)), \quad z \in \{0,1\}, \\
\Pr(Y_0 \leq y, D = 0 \mid Z = z) &= F_{Y_0}(y) - C(F_{Y_0}(y), \pi(z); \rho_0(y, \pi(z); z)), \quad z \in \{0,1\},
\end{aligned}$$

where $\rho_d(y, \pi(z); z)) \equiv \rho_d(F_{Y_d}(y), \pi(z); z))$ for clarification. Assume RS analogous to that in Chernozhukov and Hansen (2005):

$$\rho_0(y, \pi(z); z) = \rho_1(\phi(y), \pi(z); z), \quad z \in \{0, 1\}, \tag{E.3}$$

where $\phi(y)$ is such that $F_{Y_0}(y) = F_{Y_1}(\phi(y))$, leading to a system of four equations with four unknowns

$$\begin{aligned} \Pr(Y_1 \leq \phi(y), D = 1 \mid Z = z) &= C(F_{Y_0}(y), \pi(z); \rho_0(y, \pi(z); z)), \quad z \in \{0, 1\}, \\ \Pr(Y_0 \leq y, D = 0 \mid Z = z) &= F_{Y_0}(y) - C(F_{Y_0}(y), \pi(z); \rho_0(y, \pi(z); z)), \quad z \in \{0, 1\}. \end{aligned}$$

Adding the previous equations yields

$$\Pr(Y_1 \leq \phi(y), D = 1 \mid Z = z) + \Pr(Y_0 \leq y, D = 0 \mid Z = z) = F_{Y_0}(y), \quad z \in \{0, 1\},$$

so that $\phi(y)$ can be identified from

$$\begin{aligned} &\Pr(Y_1 \leq \phi(y), D = 1 \mid Z = 1) + \Pr(Y_0 \leq y, D = 0 \mid Z = 1) \\ &\qquad = \Pr(Y_1 \leq \phi(y), D = 1 \mid Z = 0) + \Pr(Y_0 \leq y, D = 0 \mid Z = 0), \end{aligned} \tag{E.4}$$

provided that this equation has unique solution; see Section A.1 for related discussions. This is the same condition used in Vuong and Xu (2017), but we do not rely on absolute continuity to arrive to this equation. Another feature of our approach is that we identify the local dependence parameter. If Y_0 and Y_1 are discrete, however, this equation might not have solution. If $\phi(y)$ is identified, then $F_{Y_0}(y)$ and $\rho_0(y, p(0); 0)$ are identified from

$$\begin{aligned} \Pr(Y_1 \leq \phi(y), D = 1 \mid Z = 0) &= C(F_{Y_0}(y), \pi(0); \rho_0(y, \pi(0); 0)), \\ \Pr(Y_0 \leq y, D = 0 \mid Z = 0) &= F_{Y_0}(y) - C(F_{Y_0}(y), \pi(0); \rho_0(y, \pi(0); 0)), \end{aligned}$$

because the Jacobian of this system

$$J(y) = \begin{pmatrix} C_1(y,0) & C_3(y,0) \\ 1 - C_1(y,0) & -C_3(y,0) \end{pmatrix}$$

has negative determinant everywhere:

$$\det(J(y)) = -C_3(y,0) < 0.$$

Finally, $\rho_0(y, \pi(1); 1)$ is identified from

$$\Pr(Y_1 \leq \phi(y), D = 1 \mid Z = 1) = C(F_{Y_0}(y), \pi(1); \rho_0(y, \pi(1); 1)),$$

because the right hand side is monotonically increasing in $\rho_0(y, \pi(1); 1)$.

Theorem E.2. *Suppose $D \in \{0,1\}$ satisfies (3.1). Suppose Assumptions EX and REL hold. Also, suppose that (E.3) holds and (E.4) has a unique solution in ϕ. Then, $F_{Y_d}(y)$ and $\rho_{Y_d}(y)$ are identified for $d \in \{0,1\}$ and $y \in \mathcal{Y}$.*

Vuong and Xu (2017) assume continuity and RI on the potential outcomes such that $Y_1 = \phi(Y_0)$ a.s. for a strictly monotone transformation $y \mapsto \phi(y)$. Using the equations above with $y' = \phi(y)$ and taking differences between $z = 1$ and $z = 0$ yields

$$\begin{aligned} &\Pr(Y_0 \leq y, D = 0 \mid Z = 0) - \Pr(Y_0 \leq y, D = 0 \mid Z = 1) \\ &\qquad = \Pr(Y_1 \leq \phi(y), D = 1 \mid Z = 1) - \Pr(Y_1 \leq \phi(y), D = 1 \mid Z = 0). \end{aligned}$$

This equation identifies the mapping $y \mapsto \phi(y)$ under monotonicity and support conditions. Another possibility is to impose LATE monotonicity assumptions on the treatment selection equation and identify ϕ from the distributions of the potential outcomes for the compliers. Note that we can also combine the restrictions within and between treatment levels.

F Alternative Selection Equation

When D is binary, alternative to (3.1), we may consider

$$D_z = 1\{D_z^* \leq 0\}, \tag{F.1}$$

where D_1^* and D_0^* are two distinct r.v.'s. Let $D^* \equiv D_Z^*$ so that $D = 1\{D^* \leq 0\}$, which is consistent with the notation in Chernozhukov et al. (2020a). Note that (3.1) is a special case of (F.1) where $V_z \equiv D_z^* + \pi(z)$ so that $D_z^* = V_z - \pi(z)$ and is normalized to $V_z \sim U[0,1]$. The alternative LGR using D_z^* still requires a similar version of CI due to the following:

$$\begin{aligned} F_{Y|D,Z}(y \mid D=1, Z=z)\pi(z) &= \Pr[Y_1 \leq y, D_z^* \leq 0 \mid Z=z] \\ &= C(F_{Y_1|Z}(y \mid z), F_{D_z^*|Z}(0 \mid z); \rho_{Y_1,D_z^*;Z}(y,0;z)) \\ &= C(F_{Y_1}(y), F_{D_z^*|Z}(0 \mid z); \rho_{Y_1}(y)), \end{aligned}$$

where the last equation is by EX and the following CI:

$$\rho_{Y_1,D_1^*;Z}(y,0;1) = \rho_{Y_1,D_0^*;Z}(y,0;0). \tag{F.2}$$

This CI looks weaker than the original CI under (3.1). This is not the case, however, which becomes clear if we use the original correlation coefficient function from Lemma 2.1:

$$\rho_{U_1,U_{21};Z}(F_{Y_1}(y), F_{D_1^*|Z}(0 \mid 1); 1) = \rho_{U_1,U_{20};Z}(F_{Y_1}(y), F_{D_0^*|Z}(0 \mid 0); 0),$$

where $U_1 \equiv F_{Y_1}(Y_1)$ and $U_{2z} \equiv F_{D_z^*|Z}(D_z^*|Z)$. Note that we require ρ being a constant function of the second argument (in addition to other invariance requirements). This is still the case even with a stronger version of independence $(Y_d, D_z^*) \perp\!\!\!\perp Z$, in which we require

$$\rho_{U_1,U_{21}}(F_{Y_1}(y), F_{D_1^*}(0)) = \rho_{U_1,U_{20}}(F_{Y_1}(y), F_{D_0^*}(0)), \tag{F.3}$$

where $U_1 \equiv F_{Y_1}(Y_1)$ and $U_{2z} \equiv F_{D_z^*}(D_z^*)$.

Similar discussions can be made in the case of ordered and continuous treatments. For example, with continuous D, we can alternatively introduce the following copula invariance without introducing V_z: (i) $\rho_{Y_d,D_z;Z}(y,d;z) = \rho_{Y_d,D_z}(y,d)$ and (ii) $\rho_{Y_d,D_z}(y,d) = \rho_{Y_d,D_z}(y)$. Note that (i) holds if and only if $(Y_d, D_z) \perp\!\!\!\perp Z$ and (ii) is a CI that resembles (F.3).

G Proofs

G.1 Proof of Theorem 3.1

Note that $\pi(z)$ is identified as a reduced-form parameter. Let $F_{d,y} \equiv F_{Y_d}(y)$ and $\rho_{d,y} \equiv \rho_{Y_d}(y)$ be the structural parameters of interest. Consider the following mapping between the structural and reduced-form parameters:

$$F_{Y|D,Z}(y|1,0)\pi(0) = C(F_{1,y}, \pi(0); \rho_{1,y}), \tag{G.1}$$

$$F_{Y|D,Z}(y|1,1)\pi(1) = C(F_{1,y}, \pi(1); \rho_{1,y}), \tag{G.2}$$

or $\pi_y = G(\theta_y)$ where $\theta_y \equiv (F_{1,y}, \rho_{1,y})'$ and $\pi_y \equiv (F_{Y|D,Z}(y|1,0)\pi(0), F_{Y|D,Z}(y|1,1)\pi(1))'$. Let C_1, C_2 and C_ρ denote the derivative of copula $C(u_1, u_2; \rho)$ with respect to u_1, u_2 and ρ, respectively. Consider the Jacobian of the system of nonlinear equations (G.1)–(G.2):

$$J = \frac{\partial G}{\partial \theta_y} = \begin{bmatrix} C_1(F_{1,y}, \pi(0); \rho_{1,y}) & C_\rho(F_{1,y}, \pi(0); \rho_{1,y}) \\ C_1(F_{1,y}, \pi(1); \rho_{1,y}) & C_\rho(F_{1,y}, \pi(1); \rho_{1,y}) \end{bmatrix}.$$

The matrix has full rank if and only if

$$\frac{C_\rho(F_{1,y}, \pi(1); \rho_{1,y})}{C_1(F_{1,y}, \pi(1); \rho_{1,y})} \neq \frac{C_\rho(F_{1,y}, \pi(0); \rho_{1,y})}{C_1(F_{1,y}, \pi(0); \rho_{1,y})}, \tag{G.3}$$

which is true by Assumption REL and Lemma 4.1 in Han and Vytlacil (2017) as Gaussian copula satisfies the stochastically increasing ordering condition (Assumption 6 in Han and

Vytlacil (2017)). Therefore, the matrix is a P-matrix (with $\rho \in (-1, 1)$), and thus one can apply Gale and Nikaido (1965)'s global univalence theorem, which identifies θ_y.[16]

Analogously, we have

$$\begin{aligned} F_{Y|D,Z}(y|D=0,Z=z)(1-\pi(z)) &= \Pr[Y_0 \le y|Z=z] - \Pr[Y_0 \le y, V_z \le \pi(z)|Z=z] \\ &= \Pr[Y_0 \le y|Z=z] - C(F_{Y_0|Z}(y|z), \pi(z); \rho_{Y_0,V_z;Z}(y, \pi(z); z)) \\ &= \Pr[Y_0 \le y] - C(F_{Y_0}(y), \pi(z); \rho_{Y_0}(y)) \end{aligned}$$

and

$$F_{y|0,0} \cdot (1-\pi(0)) = F_{Y_0}(y) - C(F_{Y_0}(y), \pi(0); \rho_{0,y}), \tag{G.4}$$

$$F_{y|0,1} \cdot (1-\pi(1)) = F_{Y_0}(y) - C(F_{Y_0}(y), \pi(1); \rho_{0,y}), \tag{G.5}$$

and the mapping has a unique solution for $\tilde{\theta}_y \equiv (F_{Y_0}(y), \rho_{0,y})'$ by a similar argument as above.

G.2 Proof of Theorem 3.2

Recall from the text that we additionally impose copula invariance between a pair of levels:

$$\rho_{Y_d,V_z;Z}(y, \pi_d(z); z) = \rho_{Y_d,V_z;Z}(y, \pi_{d-1}(z); z) \equiv \rho_{Y_d}(y) \equiv \rho_{d,y}. \tag{G.6}$$

Now, following Ambrosetti and Prodi (1995), we show that (i) the system has a unique solution when $\rho_{d,y} = 0$, (ii) the function that defines the system is continuous and proper with a range that is a connected set, and (iii) it is locally invertible. Similar to the argument in Footnote 16, (ii) can be easily shown. Note that (i) is trivially true. Therefore, we are

[16] In cases where we combine more equations, the principle minors of the resulting Jacobian may be zero. In that case, Hadamard's global inverse function theorem can be applied instead. According to Hadamard's theorem (Hadamard, 1906), the solution of $\pi_y = G(\theta_y)$ is unique if (i) G is proper, (ii) the Jacobian of G vanishes nowhere, and (iii) $G(\Theta_y)$ is simply connected. Condition (i) trivially holds with our definition of G. Since the parameter space $\Theta_y = [0, 1] \times (-1, 1)$ is simply connected and G is continuous, Condition (iii) holds if the Jacobian of G is positive or negative semi-definite on Θ_y because simple connectedness is preserved under a monotone map. We can show that the Jacobian is semidefinite and has full rank, which prove Conditions (iii) and (ii), respectively, and hence the uniqueness of the solution.

remained to prove (iii) by showing the full rank of the following Jacobian with $F_{d,y} \equiv F_{Y_d}(y)$:

$$J_d = \begin{bmatrix} C_1(F_{d,y}, \pi_d(0); \rho_{d,y}) - C_1(F_{d,y}, \pi_{d-1}(0); \rho_{d,y}) & C_\rho(F_{d,y}, \pi_d(0); \rho_{d,y}) - C_\rho(F_{d,y}, \pi_{d-1}(0); \rho_{d,y}) \\ C_1(F_{d,y}, \pi_d(1); \rho_{d,y}) - C_1(F_{d,y}, \pi_{d-1}(1); \rho_{d,y}) & C_\rho(F_{d,y}, \pi_d(1); \rho_{d,y}) - C_\rho(F_{d,y}, \pi_{d-1}(1); \rho_{d,y}) \end{bmatrix}.$$

This Jacobian has full rank if and only if

$$\frac{C_\rho(F_{d,y}, \pi_d(1); \rho_{d,y}) - C_\rho(F_{d,y}, \pi_{d-1}(1); \rho_{d,y})}{C_1(F_{d,y}, \pi_d(1); \rho_{d,y}) - C_1(F_{d,y}, \pi_{d-1}(1); \rho_{d,y})} \neq \frac{C_\rho(F_{d,y}, \pi_d(0); \rho_{d,y}) - C_\rho(F_{d,y}, \pi_{d-1}(0); \rho_{d,y})}{C_1(F_{d,y}, \pi_d(0); \rho_{d,y}) - C_1(F_{d,y}, \pi_{d-1}(0); \rho_{d,y})}. \tag{G.7}$$

Showing this is more involved than showing (G.3) with the binary treatment, because the equality can arise due to two points on the indifference curve. Nonetheless, the full-rank condition (G.7) can be expressed as $\lambda(0) \neq \lambda(1)$, where

$$\lambda(z) \equiv \frac{\phi(r_d(z)) - \phi(r_{d-1}(z))}{\Phi(r_d(z)) - \Phi(r_{d-1}(z))}, \quad r_\ell(z) \equiv \frac{\Phi^{-1}(\pi_\ell(z)) - \rho_{d,y} F_{d,y}}{\sqrt{1 - \rho_{d,y}^2}},$$

and $\Phi(\cdot)$ and $\phi(\cdot)$ are univariate Gaussian CDF and PDF, respectively. To interpret this condition, we note that it can be related to the mean of truncated Gaussian random variable (r.v.): for $A \sim N(\mu, \sigma^2)$,

$$E(A \mid l < A < u) = \mu - \sigma \frac{\phi\left(\frac{u-\mu}{\sigma}\right) - \phi\left(\frac{l-\mu}{\sigma}\right)}{\Phi\left(\frac{u-\mu}{\sigma}\right) - \Phi\left(\frac{l-\mu}{\sigma}\right)}.$$

This formula simplifies to the standard inverse Mills ratio under one-sided truncation. Therefore, the full-rank condition $\lambda(0) \neq \lambda(1)$ can be equivalently expressed as

$$E[A \mid \pi_{d-1}(0) < \Phi(A) < \pi_d(0)] \neq E[A \mid \pi_{d-1}(1) < \Phi(A) < \pi_d(1)]$$

with $\mu = \rho_{d,y} F_{d,y}$ and $\sigma^2 = 1 - \rho_{d,y}^2$. For example, this holds when threshold functions are such that $\pi_{d-1}(0) < \pi_{d-1}(1)$ and $\pi_d(0) < \pi_d(1)$. By transitivity, Assumption U_{OC} guarantees

this.

G.3 Proof of Lemma 3.1

Before presenting a formal proof, it is helpful to consider an illustrative example with $K = 4$. In this case we have three complier groups and three defier groups:

$$
\begin{aligned}
C_1 &\equiv \{D_0 = 1, D_1 = 2\} \cup \{D_0 = 2, D_1 = 3\} \cup \{D_0 = 3, D_1 = 4\}, \\
C_2 &\equiv \{D_0 = 1, D_1 = 3\} \cup \{D_0 = 2, D_1 = 4\}, \\
C_3 &\equiv \{D_0 = 1, D_1 = 4\}, \\
B_1 &\equiv \{D_1 = 1, D_0 = 2\} \cup \{D_1 = 2, D_0 = 3\} \cup \{D_1 = 3, D_0 = 4\}, \\
B_2 &\equiv \{D_1 = 1, D_0 = 3\} \cup \{D_1 = 2, D_0 = 4\}, \\
B_3 &\equiv \{D_1 = 1, D_0 = 4\}.
\end{aligned}
$$

Note that the union of

$$
\begin{aligned}
C_1 &\equiv \{D_0 = 1, D_1 = 2\} \cup \{D_0 = 2, D_1 = 3\} \cup \{D_0 = 3, D_1 = 4\}, \\
C_2 &\equiv \{D_0 = 1, D_1 = 3\} \cup \{D_0 = 2, D_1 = 4\}, \\
C_3 &\equiv \{D_0 = 1, D_1 = 4\}
\end{aligned}
$$

is identical to the double-counting union of

$$
\begin{aligned}
C_1 &\equiv \{D_0 = 1, D_1 = 2\} \cup \{D_0 = 2, D_1 = 3\} \cup \{D_0 = 3, D_1 = 4\}, \\
C_2 &\equiv \{D_0 = 1, D_1 = 3\} \cup \{D_0 = 2, D_1 = 4\}, \\
C_3 &\equiv \{D_0 = 1, D_1 = 4\}, \\
C_2 &\equiv \{D_0 = 1, D_1 = 3\} \cup \{D_0 = 2, D_1 = 4\}, \\
C_3 &\equiv \{D_0 = 1, D_1 = 4\} \cup \{D_0 = 1, D_1 = 4\}.
\end{aligned}
$$

Then, by taking the union in each column of above expression, we have

$$\begin{aligned}
\Pr\left[\bigcup_{j=1}^{3} C_j\right] &= \Pr[\{0 < V_0 \leq \pi_1(0), \pi_1(1) < V_1 \leq 1\} \\
&\qquad \cup \{0 < V_0 \leq \pi_2(0), \pi_2(1) < V_1 \leq 1\} \\
&\qquad \cup \{0 < V_0 \leq \pi_3(0), \pi_3(1) < V_1 \leq 1\}] \\
&= \Pr[\{0 < V_1 \leq \pi_1(0), \pi_1(1) < V_0 \leq 1\} \\
&\qquad \cup \{0 < V_1 \leq \pi_2(0), \pi_2(1) < V_0 \leq 1\} \\
&\qquad \cup \{0 < V_1 \leq \pi_3(0), \pi_3(1) < V_0 \leq 1\}] \\
&< \Pr[\{0 < V_1 \leq \pi_1(1), \pi_1(0) < V_0 \leq 1\} \\
&\qquad \cup \{0 < V_1 \leq \pi_2(1), \pi_2(0) < V_0 \leq 1\} \\
&\qquad \cup \{0 < V_1 \leq \pi_3(1), \pi_3(0) < V_0 \leq 1\}] \\
&= \Pr\left[\bigcup_{j=1}^{3} B_j\right],
\end{aligned}$$

where the second equality is by Assumption EG and the inequality is by $\pi_d(1) > \pi_d(0)$ for all $d = 1, 2, 3$.

Now, the following is the formal proof of the lemma. Let $\pi_0(z) = 0$ and $\pi_K(z) = 1$ for all

z. Then,

$$\begin{aligned}
\Pr\left[\bigcup_{j=1}^{K-1} C_j\right] &= \Pr\left[\bigcup_{j=1}^{K-1}\bigcup_{s=0}^{s+j+1=K} \{\pi_s(0) < V_0 \leq \pi_{s+1}(0), \pi_{s+j}(1) < V_1 \leq \pi_{s+j+1}(1)\}\right] \\
&= \Pr\left[\bigcup_{j=1}^{K-1} \{0 < V_0 \leq \pi_j(0), \pi_j(1) < V_1 \leq 1\}\right] \\
&= \Pr\left[\bigcup_{j=1}^{K-1} \{0 < V_1 \leq \pi_j(0), \pi_j(1) < V_0 \leq 1\}\right] \\
&< \Pr\left[\bigcup_{j=1}^{K-1} \{0 < V_1 \leq \pi_j(1), \pi_j(0) < V_0 \leq 1\}\right] \\
&= \Pr\left[\bigcup_{j=1}^{K-1} B_j\right],
\end{aligned}$$

where the second equality is from the derivation similar to the case of $K = 4$, the third equality is by Assumption EG, and the inequality is by $\pi_d(1) > \pi_d(0)$ for all $d \in \mathcal{D}\backslash\{K\}$. The proof of the opposite direction of inequality is symmetric.