

LAYERED POLICY ANALYSIS IN PROGRAM EVALUATION USING THE MARGINAL TREATMENT EFFECT

ISMAEL MOURIFIÉ[†] AND YUANYUAN WAN[‡]

ABSTRACT. In this paper, we propose a unified approach to derive sharp bounds on all conventional policy parameters when the instrumental variables (IVs) are potentially invalid. Using a *Vine Copula* approach, we propose a novel characterization of the identified sets for the marginal treatment effect (MTE) and the policy-relevant treatment effect (PRTE) parameters. Our method explicitly demonstrates how imposing different IV-related assumptions with different levels of credibility affects the identification of the MTE and PRTE parameters. It provides a tractable way to inform policy choices in the presence of uncertainty of the validity of identifying assumptions. Our approach enlarges the scope of the MTE framework by showing how it can inform policy decisions even when valid instruments are not available.

Keywords: Desegregated MTE, Vine Copula, Identified Set, Policy-relevant treatment effect, Specification Test.

JEL Classification: C12, C14, C21 and C26

This version: Tuesday 20th October, 2020

We are grateful to Peter Hull, Désiré Kédagni, Vitor Possebom, Pedro Sant’Anna, Thomas Russell, Alex Torgovitsky for valuable comments and/or discussions. We have greatly benefited from insightful comments from Jim Heckman. The research was conducted in part when Ismael Mourifié was visiting the Becker Friedman Institute (BFI) at the University of Chicago. Mourifié thanks his hosts for its hospitality and support. Wan thanks the support from the SSHRC Insight Grant #435190500.

[†] Corresponding author. Department of Economics, University of Toronto, 150 St. George Street, Toronto ON M5S 3G7, Canada. ismael.mourifie@utoronto.ca

[‡] Department of Economics, University of Toronto. yuanyuan.wan@utoronto.ca.

1. INTRODUCTION

Evaluating the impact of an intervention is fundamental for policymakers. It generates knowledge about a program's effectiveness and determines whether it should be scaled up, down, or discontinued. However, the program (treatment) effects may vary widely across economic agents, and expectations about individual treatment effects may trigger strategic participation. In such an environment, uncovering aggregate treatment effect parameters and using them as baseline information to evaluate new policies is challenging. Heckman and Vytlacil (2005, HV05 hereafter) propose a key causal parameter: the *marginal treatment effect* (MTE). The identification of the MTE allows researchers to recover conventional causal parameters of interest, such as average treatment effect (ATE), Local ATE (LATE), and the ATE on treated/untreated (ATT/ATUT). It also allows researchers to evaluate new policies through the policy-relevant treatment effect parameter (PRTE). Since its introduction, various approaches have been proposed to identify the MTE and then the PRTE. HV05 requires the treatment selection to be defined by a single threshold crossing model—which imposed a monotonicity restriction, see (see Vytlacil, 2002)—and a continuous instrument. Recently, Brinch, Mogstad, and Wiswall (2017), Mogstad, Santos, and Torgovitsky (2018) shows that the MTE can be recovered even in the presence of discrete instruments but at the cost of imposing some parametric or shape restrictions. Lee and Salanié (2018) relaxes the single threshold selection rule by a type of unordered monotonicity discussed in Heckman and Pinto (2018) and shows the identification of the MTE in the presence of multiple thresholds.

However, all existing MTE identification strategies strongly rely on the availability of valid instruments. The valid instruments assumption often creates a great deal of controversy amongst economists, see discussions in Deaton (2009) and Deaton, Heckman, and Imbens (2010). Manski (2011) questioned the “credibility” of policy predictions based on parameters obtained under doubtful, contestable, or non-testable restrictions, and asserted that it would be harmful to policy choice. Then, there is a clear tension between the strength of the assumptions used to recover the MTE and the “credibility” of any policy recommendations based on it. One way to resolve this tension, as advocated by Manski (2011), is what he called *layered policy analysis*. *Layered policy analysis*

demands researchers to visit various assumptions at different levels of credibility and analyze how this affects policy predictions.¹

This paper’s first main contribution is to show how one can use a modified version of the MTE to perform informative and credible policy analyses on conventional policy parameters, specifically the PRTE. To accommodate possibly invalid instruments, we introduce a modified MTE parameter, namely the *disaggregated marginal treatment effect* (DMTE). To fix the idea, consider the model $Y = Y_1D + Y_0(1 - D)$ and $D = 1\{P(Z) \geq V\}$, where Y_1 and Y_0 are potential outcomes, D is the treatment, Y is the observed outcome, Z are (possibly invalid) instrument variables, and V is independent with $P(Z)$ and is normalized to have $U[0, 1]$ distribution. We define the DMTE as the expectation of treatment effect conditioning on V and $P(Z)$ —the propensity score, that is, $\text{DMTE}(v, p) \equiv \mathbb{E}[Y_1 - Y_0 | V = v, P(Z) = p]$.² We show that all the aforementioned conventional policy parameters, including the MTE, can be expressed as a weighted average of the DMTE under the threshold-crossing treatment selection rule only, making the DMTE a more primitive parameter than the MTE. Unlike the MTE, the mapping between the DMTE and other policy parameters does not require any element of Z to be a valid instrument. Furthermore, the weights are directly identifiable from the data. Therefore, we can partially identify any of the conventional policy parameters as long as the identified set for DMTE is available.

Secondly, we propose a *Vine Copula* approach to partially identify the DMTE. We show that under the assumption that V is independent with $P(Z) \equiv P$, the dependence structure among variables (Y_d, P, V) is fully captured by two copula functions. The first one is the conditional copula of Y_d and V given P , i.e. $C_{Y_d, V | P}(\cdot, \cdot)$. This copula characterizes the endogenous selection in the model. If $C_{Y_d, V | P}(\cdot, \cdot)$ takes a product form, then we have “selection on observables”. There is no issue of endogenous selection once P is controlled; otherwise, “selection on unobservables” exists. The other copula function $C_{Y_d, P}(\cdot, \cdot)$ captures the dependence between the potential outcomes and the propensity score. This copula measures the “quality of instruments” and is a key function that we investigate in this paper. For example, if $P(Z)$ is a valid instrument, as assumed in existing literature, then it must be the case that $C_{Y_d, P}(x_1, x_2) = x_1x_2$. Therefore, we can view the IV

¹Regarding the *layered analysis*, Manski (2011, F289) said: “A researcher who performs an instructive layered policy analysis and exposit work clearly may see himself as having accomplished the objective of informing choice.”

²Like the LATE, the DMTE itself is an instrument-dependent parameter. It is not used here for causal interpretation, but rather as an intermediary quantity to recover conventional policy parameters.

independence assumption as a shape restriction on the unknown function $C_{Y_d,P}(x_1, x_2)$. Under the copula formulation, we show that calculating the identified set of the DMTE boils down to finding the set of conditional bivariate copulas that respect a set of *equality constraints* and any practical constraints that researchers would like to impose on $C_{Y_d,P}(x_1, x_2)$.

In our *Vine Copula* characterization, the identified set of DMTE depends on restrictions that we impose on $C_{Y_d,P}(x_1, x_2)$ in an explicit way. At one extreme, when we impose the IV-independence assumption, the set of equality constraints pins down a unique copula $C_{Y_d,V}(\cdot, \cdot)$, which therefore allows the point identification of the MTE and then other policy parameters. At the other extreme, where we impose no restrictions on $C_{Y_d,P}(\cdot, \cdot)$, our characterization recovers the sharp bounds on the DMTE under only the single threshold crossing assumption. One can also impose restrictions that are weaker than IV-independence. For instance, we show that imposing the Monotone IV assumption—see [Manski and Pepper \(2000\)](#)—is equivalent to considering only the set of copulas $C_{Y_d,P}(x_1, x_2)$ that are concave in x_2 . In this case, we recover the sharp identified set under monotone IV. As we demonstrate in more details in the main text, our approach, in general, provides empirical researchers a very flexible way to derive the identified set on the DMTE under any dependence restrictions she is willing to impose between the IV and the potential outcomes. From this perspective, our method shares the same spirit as the *layered policy analysis* discussed in [Manski \(2011\)](#).

To make our approach more operational, we also consider a *semi-parametric* version of our characterization by imposing parametric assumptions on the copulas but leaving the marginals entirely nonparametric. In this case, we show that the identification analysis boils down to constructing identified sets for only the copula function’s finite-dimensional parameters. As discussed in [Chen, Fan, and Tsyrennikov \(2006\)](#), using this type of semi-parametric approach to study multivariate distributions has gained popularity in diverse fields for its flexibility and ability to circumvent the curse of dimensionality.

It is worth noting that the assumptions — IV-independence and monotonicity assumptions—imposed by HV05 to identify the MTE have testable implications (see HV05 Appendix A). As a by-product contribution, we provide a more tractable characterization MTE assumptions’ sharp testable implications. We show that the proposed characterization is sufficient to screen all possible observable violations of the MTE assumptions, and one can test it using existing inferential methods,

e.g., [Hsu, Liu, and Shi \(2019\)](#). The rejection of the testable implication demands one to relax some of the MTE assumptions, and in such cases, our proposed layered analysis provides a possible solution.

Finally, we show that our identification strategy for MTE and PRTE also applies to models in which multiple thresholds on multiple unobserved heterogeneities define the selection equation. Therefore, this current paper (i) extends the [Lee and Salanié \(2018\)](#) identification approach to the case when none of the existing continuous covariates satisfies the IV-independence assumption; and (ii) applies also to the so-called "actual monotonicity" recently discussed in Mogstad, Torgovitsky, and Walters (2019).

We organize the rest of the paper as follows. In Section 2, we introduce the intermediate quantity DMTE and build its connection with other policy parameters. We characterize the identified set for DMTE under a variety of assumptions on the IV in Section 3. We extend the analysis to a two-threshold model in Section 4. Section 5 concludes the paper.

2. POLICY PARAMETERS AND DISAGGREGATED MTE

We adopt the framework of the potential outcomes model: $Y = Y_1D + Y_0(1 - D)$, where $Y \in \mathcal{Y} \subseteq \mathbb{R}$ is the observed outcome taking values from the support \mathcal{Y} , $D \in \{0, 1\}$ is the observed treatment indicator, and (Y_1, Y_0) are potential outcomes. [Heckman and Vytlacil \(1999\)](#) trace the genealogy of this model, and we refer to them for terminology and attribution. Let Z be a vector of covariates taking values from the support $\mathcal{Z} \subseteq \mathbb{R}^{d_z}$ for $d_z \geq 1$. Let us consider the following assumptions required for the point identification of the MTE:

Assumption 1 (Single Threshold-Crossing: STC). *The selection mechanism is governed by the following threshold crossing model $D = 1\{v(Z) \geq V\}$ for some measurable and non-trivial function v , where V follows a uniform distribution over the interval $[0, 1]$ and is statistically independent of the vector of covariates Z , i.e., $Z \perp V$.³*

Hereafter, we use the shorthand notation $P(Z)$ for $\mathbb{P}(D = 1|Z)$, a quantity directly recoverable from data. We can see that under Assumption 1, $v(\cdot)$ is identified over the support of Z since $P(z) = \mathbb{P}(V \leq v(z)|Z = z) = v(z)$ for all $z \in \mathcal{Z}$. When it does not confuse, we will use the shorthand notation P or p to denote $P(Z)$ or $P(z)$, respectively. Let $\mathcal{P} \subseteq [0, 1]$ denote the support

³ V follows uniform distribution is a normalization. Please see [Vytlacil \(2002\)](#) for primitive conditions under which this normalization is without loss of generality.

of $P(Z)$ and $P(z)$ a generic element of \mathcal{P} . For the main text, we restrict our attention to the cases where the set of limit points of \mathcal{P} , denoted by $L(\mathcal{P})$, is non-empty. This excludes the case that P is discrete.

Assumption 2 (IV Independence). *Conditioning on the first stage unobserved variable, the propensity score is statistically independent with the potential outcomes, i.e. $P(Z) \perp Y_d | V$ for $d = 0, 1$.*

Assumptions 1 and 2 and the whole remaining analysis, distributional assumptions and theoretical results, are understood to be conditional on a set of observed covariates X , which will be omitted from the notation for sake of simplicity. In the following, we will first review why MTE, or the marginal treatment response (MTR), is not point identified without Assumption 2. Then we will examine the restrictions that can be used for partial identification. Let $g : \mathcal{Y} \rightarrow \mathbb{R}$ be a real integrable function such that $\mathbb{E}[|g(Y_d)|] < \infty$. Taking $d = 1$ as illustration and following the identification strategy of HV05, for all $p \in L(\mathcal{P})$:

$$\begin{aligned} \mathbb{E}[g(Y)D|P = p] &= \mathbb{E}[g(Y_1)|D = 1, P = p]\mathbb{P}(D = 1|P = p) \\ &= \mathbb{E}[g(Y_1)|V \leq p, P = p]\mathbb{P}(V \leq p|P = p) = \mathbb{E}[g(Y_1)|V \leq p, P = p]p \\ &= \int_0^p \mathbb{E}[g(Y_1)|V = v, P = p]dF_{V|P=p} = \int_0^p \mathbb{E}[g(Y_1)|V = v, P = p]dv, \end{aligned}$$

where all equalities holds only under Assumption 1. The key equation is then:

$$\mathbb{E}[g(Y)D|P = p] = \int_0^p \mathbb{E}[g(Y_1)|V = v, P = p]dv. \quad (1)$$

By taking the derivative of the previous equation respect with p we obtain:

$$\frac{\partial}{\partial p} \mathbb{E}[g(Y)D|P = p] = \mathbb{E}[g(Y_1)|V = p, P = p] + \int_0^p \frac{\partial}{\partial p} \mathbb{E}[g(Y_1)|V = v, P = p]dv. \quad (2)$$

It can be seen that the left hand side of the equation, also known as the the Local IV (LIV) estimand proposed by HV05, can no longer identify the MTR (and MTE) because (i) $\frac{\partial}{\partial p} \mathbb{E}[g(Y_1)|V = v, P = p] \neq 0$, and (ii) $\mathbb{E}[g(Y_1)|V = p, P = p]$ is in general different from $\mathbb{E}[g(Y_1)|V = p]$ when P and Y_1 are not independent conditioning on V . Nevertheless, Equation (1) still contains useful information about the quantity

$$\theta_g^d(v, p) \equiv \mathbb{E}[g(Y_d)|V = v, P = p], \quad d = 0, 1, (v, p) \in [0, 1] \times \mathcal{P},$$

which we refer in the rest of paper as the *disaggregated marginal treatment responses* with respect to the function g and abbreviate it as DMTR_g . Analogous to the relationship between MTR and MTE, we define another intermediate quantity *disaggregated marginal treatment effect* (DMTE) as:

$$\text{DMTE}(v, p) \equiv \mathbb{E}(Y_1 - Y_0 | V = v, P = p), \quad \forall (v, p) \in [0, 1] \times \mathcal{P}. \quad (3)$$

DMTR implies DMTE since $\text{DMTE}(v, p) = \theta_g^1(v, p) - \theta_g^0(v, p)$ with $g(\cdot)$ being chosen as the identity function.^{4 5} It is apparent from Equation (3) that the identification of MTR (hence MTE) is readily available once DMTR is recovered since

$$\mathbb{E}[g(Y_d) | V = v] = \int_0^1 \mathbb{E}[g(Y_d) | V = v, P = t] f_{P|V}(t|v) dt = \int_{\mathcal{P}} \theta_g^d(v, t) f_P(t) dt,$$

and

$$\text{MTE}(v) \equiv \mathbb{E}(Y_1 - Y_0 | V = v) = \int_{\mathcal{P}} \text{DMTE}(v, t) f_P(t) dt,$$

where $f_{P|V} = f_P$ by Assumption 1 and the density f_P of P is directly identified from data. Note that when $P(Z) \perp Y_d | V$ as in HV05, the DMTE is exactly equal to the MTE and we have

$$\text{DMTE}(v, p) = \text{DMTE}(v, p') = \text{MTE}(v), \quad (4)$$

for all $(p, p') \in \mathcal{P} \times \mathcal{P}$ and $v \in [0, 1]$. Therefore, although the DMTR and DMTE are not necessarily parameters of direct interest—for being instrument-dependent—they do serve as useful intermediate quantities to identify the MTR, MTE, and other useful policy parameters such as the ATE, ATT, ATUT. Specifically, we will show in Theorem 1 below that all these mentioned parameters can be expressed as a weighted average of DMTE under Assumption 1 only. Another useful parameter that often draws interest is the policy relevant treatment effect (PRTE_g):

$$\text{PRTE}_g \equiv \frac{\mathbb{E}[g(Y) | \mathbf{a}'] - \mathbb{E}[g(Y) | \mathbf{a}]}{\mathbb{E}[D | \mathbf{a}'] - \mathbb{E}[D | \mathbf{a}]},$$

⁴One can also recover its distributional version by choosing $g(Y_d) = \mathbf{1}[Y_d \leq y]$, i.e. $\mathbb{P}(Y_1 \leq y | V = v, P = p) - \mathbb{P}(Y_0 \leq y | V = v, P = p)$, $\forall (v, p) \in [0, 1] \times \mathcal{P}$.

⁵It is worth-noting that our DMTE shares a superficial resemblance with the Redefined MTE ($\widetilde{\text{MTE}}$) introduced in Zhou and Xie (2019). In presence of a vector of exogenous covariates X , $\widetilde{\text{MTE}}(v, p) = \mathbb{E}[Y_1 - Y_0 | V = v, P(Z, X) = p]$ while $\text{DMTE}(v, p, x) = \mathbb{E}[Y_1 - Y_0 | V = v, P(Z, X) = p, X = x]$. Unlike the DMTE when $X | P(X, Z)$ has a degenerate distribution $\widetilde{\text{MTE}}$ is equal to the MTE, please see Zhou and Xie (2019, Page 3076, eq 8).

where a' and a denote the alternative policy under consideration and the baseline policy, respectively. Please refer to Heckman and Vytlacil (2001b), HV05, and Carneiro, Heckman and Vylacil (2010) for a detailed discussion about the PRTE.

Hereafter, Let Y_d^a denote the potential outcome when the agent is externally set to treatment d under policy a . In HV05, the identification of PRTE rely on both Assumptions 1 and 2 and the following policy invariant assumption:

Assumption 3 (HV05, Policy Invariance). $(Y_d^{a'}, V^{a'}) \sim (Y_d^a, V^a)$ for $a \neq a'$.

Under Assumptions 1 to 3, HV05 shows that

$$\text{PRTE}_g = \int_0^1 \frac{F_{P^a}(v) - F_{P^{a'}}(v)}{\underbrace{\mathbb{E}_{F_{P^{a'}}}[P] - \mathbb{E}_{F_{P^a}}[P]}_{w^{\text{PRTE}}(v)}} \text{MTE}_g(v) dv.$$

The PRTE is designed by HV05 to evaluate a new policy that induces a change in P but keeping the full joint distribution of latent variable unchanged from the baseline policy to the targeting alternative policy, when Assumptions 1 to 3 hold. In the following, we will propose an alternative policy invariance assumption and show that the PRTE_g can be recovered even when Assumption 2 fails to hold.

Assumption 4 (Conditional Policy Invariance). $Y_d^{a'} | V^{a'}, P^{a'} \sim Y_d^a | V^a, P^a$ with $V^{a'} \sim U[0, 1]$, $V^a \sim U[0, 1]$ and $Y_d^{a'} \sim Y_d^a$ for $a \neq a'$.

In Assumption 4, we maintain the same normalization for the distribution of V in both baseline policy and alternative policy environments. Notice that under Assumption 2, Assumption 4 holds if and only if Assumption 3 holds. Therefore, if the independence assumption holds, our conditional policy-invariance becomes equivalent to the HV05's policy-invariance assumption. The difference is that without the IV-independence assumption, Assumption 4 requires the DMTE_g^a to be invariant from a policy a to an alternative policy a' , namely $\text{DMTE}_g^{a'} = \text{DMTE}_g^a$. Unlike HV05, our conditional policy-invariance assumption does not impose the invariance of the MTE_g^a . Theorem 1 below shows that the PRTE_g can be written as a weighted average of the DMTE_g under Assumptions 1 and 4. We now summarize the above discussion in the following Theorem.

Theorem 1. *Suppose that Assumption 1 is satisfied, then*

(i) $MTE_g(v) = \int_0^1 DMTE_g(v, p) f_P(p) dp$;

(ii) for any $s \in \{ATE_g, LATE_g(u, u'), ATT_g, ATUT_g\}$ ⁶ and weights $\omega^s(v, p)$ listed in Table 1 below, we have

$$s = \int_0^1 \int_0^1 \omega^s(v, p) DMTE(v, p) dv dp. \quad (5)$$

(iii) If in addition Assumption 4 holds, Equation (5) holds with $s = PRTE$.

TABLE 1. Policy Parameters and DMTE

Parameters	weights $\omega^s(v, p)$
ATE_g	$f_P(p)$
ATT_g	$\frac{f_P(p)1\{v < p\}}{\mathbb{E}[P]}$
$ATUT_g$	$\frac{f_P(p)1\{v > p\}}{\mathbb{E}[1-P]}$
$LATE_g(u, u')$	$\frac{f_P(p)1_{\{u < v \leq u'\}}}{u' - u}$
$PRTE_g$	$\frac{[f_{pa'}(p) - f_{pa}(p)]1\{v \leq p\}}{\mathbb{E}_{F_{pa'}}[P] - \mathbb{E}_{F_{pa}}[P]}$

Proof. See Appendix C.1. □

Notice that one can easily verify that when Assumption 2 holds, $DMTE_g(v, p) = DMTE_g(v)$ and then $\int_0^1 \omega^s(v, p) dp = w^s(v)$ for any $s \in \{ATE_g, LATE_g(u, u'), ATT_g, ATUT_g, PRTE_g\}$, with $w^s(v)$ being exactly the weights derived in HV05. Although $DMTE_g$ itself may or may not be the main parameter of interest, Theorem 1 shows that it plays an important role in the identification of many common parameters of interest. Note that the weights are known and can be estimated for each value $(v, p) \in [0, 1]^2$. Thus, we can readily recover the identified sets for any of the conventional policy parameters once we have the identified set for the $DMTR_g$ (hence $DMTE_g$). Therefore, our main goal will be to provide a tractable characterization of the identified set for the DMTRs.

⁶Here $LATE_g(u, u')$ represents the average treatment effect for the group of compliers when P is externally changed from u to u' .

Definition 1. Suppose that Assumption 1 is satisfied. For any integrable real function $g(\cdot)$, the identified set $\Theta_{I,g}$ for $DMTR_g$ is defined as follows:

$$\Theta_{I,g} = \left\{ (\theta_g^0, \theta_g^1) : [0, 1] \times [0, 1] \rightarrow \mathbb{R}^2 \text{ such that} \right. \\ \left. \mathbb{E}[g(Y)1\{D = d\}|P = p] = \int_{p1\{d=0\}}^{p+(1-p)1\{d=0\}} \theta_g^d(v, p) dv \text{ for } d \in \{0, 1\} \right\}.$$

In this paper, we focus on the half-interval class $\mathcal{G} \equiv \{g(\cdot) = \mathbf{1}[\cdot \leq y], y \in \mathcal{Y}\}$ when identifying DMTRs. Under the half interval class, the $DMTR_g^d$ can then be expressed as

$$DMTR_g^d(v, p) = \mathbb{P}[Y_d \leq y|V = v, P = p] \equiv F_{Y_d|V,P}(y|v, p) \quad d \in \{0, 1\}.$$

Here, the $DMTR_g^d$ is just the conditional distribution function of Y_d given P and V . Therefore, the identification under the interval class recovers DMTRs for other classes of g functions.

3. IDENTIFICATION

In the previous section, we show that there exist intermediate quantities $DMTE_g$ or $DMTR_g$, which facilitate the partial identification of the MTE_g and the MTR_g without imposing the IV-independence Assumption 2. In this section, we will take the STC structure (Assumption 1) as given and characterize identified sets for the MTE_g under a sequence of assumptions on the copula of (P, Y_d) . In this sequence of assumptions, we have “no restrictions”⁷ and the “IV-independence assumption” representing two extreme cases.

3.1. Identification under a single threshold-crossing selection rule. Once we focus on the half-interval class of g functions, the primitive parameter of interest, as defined by Definition 1 is the conditional distribution of Y_d given V and P . While we do not know precisely the full joint distribution, the STC structure (Assumption 1) does provide some restrictions. For instance, we know the distribution of two of three marginal distributions, i.e., V and P , and we know that they are independent. This feature of the model suggests using copulas decomposition to study our joint distribution of interest. Copula theory is useful to separate marginal properties from properties related to the dependence structure. Here we cite Sklar (1959)’s result:

⁷We do, however, impose certain regularity conditions, as will be made clear later.

Lemma 1 (Sklar (1959)'s Theorem). *There exist a copula $C : [0, 1]^3 \rightarrow [0, 1]$ such that*

$$\mathbb{P}(Y_d \leq y, V \leq v, P \leq p) = C_{Y_d, V, P}(F_{Y_1}(y), F_V(v), F_P(p)), \quad \text{for } y, v, p \in [-\infty, \infty].$$

Moreover, if the margins are continuous, then C is unique; otherwise it is uniquely determined on $\text{Ran}F_{Y_d} \times \text{Ran}F_V \times \text{Ran}F_P$ where $\text{Ran}F_X = F_X([-\infty, \infty])$ is the range of F_X .

Using the Sklar's result, we can decompose the joint trivariate distribution into three univariate distributions and one trivariate copula $C_{Y_d, V, P}(F_{Y_1}(y), F_V(v), F_P(p))$. However, working directly with the trivariate copula is not very convenient. Unlike the bivariate copula, the dependence of trivariate copula can be less intuitive to interpret. Also, the number of multivariate (> 2) copula families with flexible dependence is limited. Furthermore, the STC assumption (Assumption 1) already provides some restrictions, such as the independence of P and V , and their known marginals. To fully take advantage of that information, we consider the *Vine Copula* approach, which was introduced by Joe (1994) to break down the dependence structure of a multivariate copula into a sequence of bivariate copulas and conditional bivariate copulas. The Vine copula approach has proven to be useful in various existing problems such as (constrained) sampling of correlation matrices, building non-parametric continuous Bayesian networks, and various applications in finance. Here, we will make use of the Vine copula in our treatment effect context. We consider the following regularity assumption:

Assumption 5. *The joint distribution of (Y_d, V, P) is absolutely continuous respect to the Lebesgue measure.*

V is continuous by definition, and P is continuous if Z contains a continuous element. It is worth noting that we do not require the variable Z to be a valid instrument in the sense of being independent of $Y_d, d \in \{0, 1\}$. Then the presence of any continuous exogenous covariates in the selection equation ensures P is continuous. Therefore, the main restriction of Assumption 5 is to focus the analysis on applications with continuous outcomes.

Let $C_{Y_d, V|P=p}(F_{Y_d|P}(y|p), F_{V|P}(v|p); p) = F_{Y_d, V|P}(y, v|p)$ be the conditional copula of (Y_d, V) given $P = p$. Note that given our assumption that $V|P = p \sim U[0, 1]$ for all p , the second term in the parenthesis can be simplified to v since $F_{V|P}(v|p) = v$, that is, $C_{Y_d, V|P=p}(F_{Y_d|P}(y|p), v; p)$.

Under Assumption 5, the following Lemma expresses all the conditional distributions in terms of copulas.

Lemma 2 (Vine Copula). *Under Assumptions 1 and 5, for $d \in \{0, 1\}$ we have for each $y \in \mathcal{Y}$ and $p \in \mathcal{P}$,*

$$F_{Y_d|P}(y|p) = \frac{\partial}{\partial x_2} C_{Y_d, P}(x_1, x_2) \Big|_{x_1=F_{Y_d}(y), x_2=F_P(p)} \equiv c_{d, F_P(p)}(F_{Y_d}(y)), \quad (6)$$

$$F_{Y_d|V, P}(y|v, p) = \frac{\partial}{\partial x_2} C_{Y_d, V|P=p}(x_1, x_2) \Big|_{x_1=F_{Y_d|P}(y|p), x_2=v} \quad (7)$$

and there exists monotone mapping $\Psi_{1,p}$ and $\Psi_{0,p}$ such that for each $y \in \mathcal{Y}$ and $p \in \mathcal{P}$,

$$\mathbb{P}[Y \leq y, D = 1|P = p] = \Psi_{1,p}(F_{Y_1}(y)) \equiv C_{Y_1, V|P=p}(c_{1, F_P(p)}(F_{Y_1}(y)), p; p). \quad (8)$$

$$\begin{aligned} \mathbb{P}[Y \leq y, D = 0|P = p] &= \Psi_{0,p}(F_{Y_0}(y)) \\ &\equiv c_{0, F_P(p)}(F_{Y_0}(y)) - C_{Y_0, V|P}(c_{0, F_P(p)}(F_{Y_0}(y)), p; p). \end{aligned} \quad (9)$$

That is, the observed probability $\mathbb{P}[Y \leq y, D = d|P = p]$ depends on y only through $F_{Y_d}(y)$.

Proof. See Appendix C.2. □

We have some remarks on the usefulness of the Lemma 2. First, by inserting Equation (6) to Equation (7), we observe that the parameter of interest DMTR (which equals $F_{Y_d|V, P}(y|v, p)$ under the half-interval class) can be considered as a composite mapping from $F_{Y_d}(y)$, where the mapping depends only on the (partial derivatives of) two bivariate copula functions: $C_{Y_d, V|P}$ and $C_{Y_d, P}$. Instead of working with $F_{Y_d|P, V}$, which involves with dependence structure of three variables, we can now focus on two bivariate copulas. To be more specific, let \mathcal{F}^c be the set of continuous CDFs, \mathcal{C}_d^c be the set of conditional copulas for $C_{Y_d, V|P}$, and \mathcal{C}_d be the set of copulas for $C_{Y_d, P}$, the identification of DMTR is equivalent to the identification of $\tilde{\theta} \equiv (F_{Y_1}, F_{Y_0}, C_{Y_1, V|P}, C_{Y_0, V|P}, C_{Y_1, P}, C_{Y_0, P}) \in \tilde{\Theta} \equiv \mathcal{F}^c \times \mathcal{F}^c \times \mathcal{C}_1^c \times \mathcal{C}_0^c \times \mathcal{C}_1 \times \mathcal{C}_0$ under only the STC structure (Assumption 1) and the regularity condition (Assumption 5). This is important not only because bivariate copula are easier to model, but also because $C_{Y_d, P}$ and $C_{Y_d, V|P}$ have appropriate economic interpretations, as we will see in the next sessions.

Secondly, Equations (8) and (9) in the second part of Lemma 2 provide a link between $\tilde{\theta}$ and the observed data distribution. Interestingly, given $\Psi_{d,p}$ is invertible (see Appendix C.3), one can “solve” $F_{Y_d}(y) = \Psi_{d,p}^{-1}(\mathbb{P}[Y \leq y, D = d | P = p])$ from Equations (8) and (9). So once the two bivariate copula are fixed, F_{Y_d} is uniquely determined. Meanwhile, since $F_{Y_d}(y)$ does not depends on p , it should be the case that $\forall p \neq p'$ and $d = 0, 1$,

$$\Psi_{d,p}^{-1}(\mathbb{P}[Y \leq y, D = d | P = p]) = \Psi_{d,p'}^{-1}(\mathbb{P}[Y \leq y, D = d | P = p']). \quad (10)$$

Indeed, one can show any pair of copulas $C_{Y_d, V|P}$ and $C_{Y_d, P}$ that satisfy Equations (8) and (9) can be rationalized by the data and the model structure. This indeed characterizes the identified set for $\tilde{\theta}$, as we summarize in the following theorem.

Theorem 2. *Under Assumptions 1 and 5, the identified set $\Theta_{I,g}$ in Definition 1 can be equivalently expressed by the identified set Θ_I of $\tilde{\theta} \equiv (F_{Y_1}, F_{Y_0}, C_{Y_1, V|P}, C_{Y_0, V|P}, C_{Y_1, P}, C_{Y_0, P})$, which is characterized as follows:*

$$\Theta_I = \left\{ \tilde{\theta} \in \tilde{\Theta} : \text{For } d \in \{0, 1\}, (C_{Y_d, V|P}, C_{Y_d, P}) \in \mathcal{C}_d^c \times \mathcal{C}_d \text{ satisfies Equation (10)} \right. \\ \left. \text{and } \forall y \in \mathcal{Y}, F_{Y_d}(y) = \Psi_{d,p}^{-1}(\mathbb{P}[Y \leq y, D = d | P = p]) \right\}.$$

Proof. See Appendix C.3. □

Theorem 2 characterizes the identified set for $\tilde{\theta}$ under the STC restriction imposed on the treatment selection alone. It says that any pair of copulas $(C_{Y_d, V|P}, C_{Y_d, P})$ such that the mapping $\Psi_{d,p}^{-1}$ produces a flat function in p , can be rationalized by the observed data and the STC model. Meanwhile, the Theorem also provides a convenient characterization of the identified set for subvectors of the parameters. For instance, the projection of the identified set for copulas are determined by Equation (10), and once $(C_{Y_d, V|P}, C_{Y_d, P})$ are fixed, F_{Y_d} is point identified. In particular, Equation (10) essentially uses the fact that the marginal distribution of potential outcome is invariant to the propensity score. This identification approach has some similarity with the identification restriction of [Arellano and Bonhomme \(2017, Lemma 1\)](#) in their study of the sample selection model. Since the marginal distributions of potential outcomes are uniquely determined given $(C_{Y_d, V|P}, C_{Y_d, P})$, one

can expect that the assumptions that one imposes on the dependence structure among these variables largely determine the identification power of the model. Not surprisingly, the first restriction in Theorem 2 reduces to the identification equation in HV05 when Assumption 2 holds, that is, when P is independent with Y_d given V . The following corollary summarizes this observation.

Corollary 1. *Suppose Assumptions 1, 2 and 5 hold, then the identification equation postulated in Theorem 2 coincides with the identification result of HV05, that is,*

$$\frac{\partial \mathbb{P}[Y \leq y, D = 1 | P = p]}{\partial p} = \mathbb{P}[Y_1 \leq y | V = p].$$

and

$$-\frac{\partial \mathbb{P}[Y \leq y, D = 0 | P = p]}{\partial p} = \mathbb{P}[Y_0 \leq y | V = p].$$

Proof. See Appendix C.4.

Remark 1. *(Selection on observables) In our framework, the selection on observables assumption boils down to imposing that $C_{Y_d, V | P}(x_1, x_2) = x_1 x_2$, in such a case, we can easily see through Lemma 2 that we recover the well known identification result under selection on observables, i.e. $F_{Y_d | P}(y | p) = P(Y \leq y | D = d, P = p)$.*

3.2. Identification with Imperfect IVs. We previously characterized the identified set for the DMTE by only imposing the STC assumption. We also showed that if we additionally impose the independence assumption, we recover the HV05 point identification results for the MTE and subsequent parameters. However, the validity of the independence assumption is very often subject to significant controversy, see, for instance, Deaton, Heckman, and Imbens (2010). Therefore, Manski and Pepper (2000) proposed to relax the IV assumption in order to have more credible and trustworthy results.

In this subsection, we will further explore in this direction and study the identification of DMTE under imperfect IVs. Here, we refer to “imperfect IVs” as any covariates in the selection equation that could be dependent on the potential outcomes, with the type of dependence being restricted by the economic theory or the empirical context under study. We will show that these restrictions can be easily implemented in our approach to derive sharp bounds on the DMTE and, therefore, on all conventional policy parameters. Indeed, Manski and Pepper (2000) provide sharp bounds on the ATE

under the monotone IV (MIV) assumption. However, their approach is not immediately transferrable in the sense that if an applied researcher is interested in other parameters or other deviations of the IV assumptions, she has to derive the specific sharp bounds for this parameter of interest. Using DMTE as a bridge, our unified approach allows the researcher to recover sharp bounds on a variety of parameters of interest under a various IV dependence assumptions, a nice feature inherited from the classical MTE framework. This will free applied researchers from case-by-case constructions.

Before visiting various imperfect IV restrictions, let us discuss an empirical context where misspecification may generate a violation of the IV independence assumption. One recent and growing empirical application using the MTE identification strategy is the Judge leniency IV designs.⁸ We will use such empirical context to illustrate the importance of our method.

3.2.1. *Misspecification in presence of multiple treatments.* Consider a model where two simultaneous treatments determine the outcome while researchers focus only on one treatment and overlook the second one. This is a paramount concern that appears in the Judge leniency IV design literature. In this literature, researchers are interested in the causal effect of incarceration decisions on future outcomes such as recidivism, making abstraction of other potential treatments. However, trial decisions are multidimensional, with judges deciding on incarceration, fines, community service, etc.⁹ Let us consider the following model:

$$Y = \underbrace{[Y_{11}D_2 + Y_{10}(1 - D_2)]}_{Y_1} D_1 + \underbrace{[Y_{01}D_2 + Y_{00}(1 - D_2)]}_{Y_0} (1 - D_1),$$

$$D_1 = 1\{P_1 > V_1\}, \quad D_2 = 1\{P_2 > V_2\}.$$

where D_1 denotes the incarceration decision and D_2 is a second binary treatment that denotes if the agent receives a fine or not. $Y_{d_1d_2}$ denotes the potential outcome when the two treatments are externally set to $D_1 = d_1$ and $D_2 = d_2$. P_1 and P_2 are measures of the judge's stringency level for a different punishment. Assuming that the judge's assignment to cases is entirely random, we might expect the following IV-independence assumption $(V_1, V_2, Y_{d_1d_2}) \perp (P_1, P_2)$ to hold. When D_2 is

⁸See for instance Kling (2006); Aizer and Doyle Jr (2015); Di Tella and Schargrodsky (2013); Mueller-Smith (2015); Dobbie, Goldin, and Yang (2018), and Bhuller, Dahl, Loken, and Mogstad (2019) among others.

⁹Please see Bhuller, Dahl, Loken and Mogstad (2019, section 5.5) for a detailed discussion.

neglected, researchers essentially adopt the following model:

$$Y = Y_1 D_1 + Y_0 (1 - D_1), \quad (11)$$

$$D_1 = 1\{P_1 > V_1\}. \quad (12)$$

where $Y_d \equiv [Y_{d1} D_2 + Y_{d0} (1 - D_2)]$ for $d \in \{0, 1\}$. When considering this simplified model, we must be cautious about the causal interpretation of $\mathbb{E}[Y_1 - Y_0 | V_1 = v_1]$. Indeed, the latter quantity does not capture the direct causal effect of incarceration (D_1), but rather the total causal effect of the incarceration—for an individual at the margin—that is mediated by the fines effect on the recidivism. Please see [Pearl \(2013\)](#) for a detailed discussion. Now, let's presume that we are interested in the identification of the total effect. Under the IV independence $(V_1, V_2, Y_{d_1 d_2}) \perp (P_1, P_2)$, where $d_1 \in \{0, 1\}, d_2 \in \{0, 1\}$, we can show that for $p_1 \leq p'_1$ (without loss of generality) and an arbitrary $y \in \mathcal{Y}$,

$$\mathbb{P}(Y_1 \leq y | P_1 = p_1) - \mathbb{P}(Y_1 \leq y | P_1 = p'_1) = \mathbb{E}_{P_2 | P_1 = p_1} [H_y(P_2)] - \mathbb{E}_{P_2 | P_1 = p'_1} [H_y(P_2)],$$

where $H_y(p_2) \equiv \int_0^{p_2} \left\{ \mathbb{P}(Y_{11} \leq y | V_2 = v_2) - \mathbb{P}(Y_{10} \leq y | V_2 = v_2) \right\} dv_2$. This leads to the following result:

Lemma 3. *Considering the model (11, 12), where $Y_d \equiv [Y_{d1} D_2 + Y_{d0} (1 - D_2)]$ for $d \in \{0, 1\}$, $D_2 = 1\{P > V_2\}$ and $(V_1, V_2, Y_{d_1 d_2}) \perp (P_1, P_2)$. We have,*

(i) *If $\mathbb{P}(Y_{d1} \leq y | V_2 = v_2) = \mathbb{P}(Y_{d0} \leq y | V_2 = v_2)$ for all v_2 and y , then $Y_d \perp P_1$.*

(ii) *If $P_1 \perp P_2$, then $Y_d \perp P_1$.*

(iii) *If for each $v_2 \in [0, 1]$ the conditional distribution of Y_{d1} given $V_2 = v_2$ first order stochastically dominates the distribution of Y_{d0} given $V_2 = v_2$, i.e. $Y_{d0} | V_2 = v_2 \succeq_{FSD} Y_{d1} | V_2 = v_2$, and if for any pair $(p'_1, p_1) \in \mathcal{P}_1^2$ such that $p'_1 \geq p_1$, we have $P_2 | P_1 = p'_1 \succeq_{FSD} P_2 | P_1 = p_1$, then $Y_d | P_1 = p_1 \succeq_{FSD} Y_d | P_1 = p'_1$.*

Proof. See Appendix C.5

Lemma 3 (i) and (ii) provide two sufficient conditions under which ignoring D_2 does not cause the failure of IV-independence assumption (with respect to P_1). In the misspecified model, Y_1 is essentially a mixture of Y_{11} and Y_{10} —two random variables that are independent with P_1 . Condition (i) says that these two random variables have the same distribution conditioning on V_2 ; hence any

mixing between them does not change the distribution. The condition in (ii) says the mixing weights are independent of P_1 , so the mixture of Y_{11} and Y_{10} is independent of P_1 as well. However, in applications, it is hard to justify Y_{11} and Y_{10} have the same distribution conditioning on V_2 , and we also observe that P_1 and P_2 tend to be positively correlated. Therefore, if receiving a fine or not has a direct causal impact on recidivism and if a judge's stringency indexes for both treatments are correlated, then the IV independence assumption in the misspecified model is violated. [Bhuller, Dahl, Loken, and Mogstad \(2019\)](#) propose some suggestive ideas to screen for such violations; however, it is worth noting their suggestive tests are not very informative to screen violations of the MTE assumptions. In [Appendix A](#), we propose a nonparametric specification test of the MTE assumptions. We show that the test is sharp in the sense that it is the most informative one to screen all detectable violations of the MTE assumptions.

[Lemma 3](#) (iii) shows that while the IV independence assumption is violated, we can invoke a monotone IV assumption under some reasonable restrictions. More precisely, $Y_{10}|V_2 = v_2 \succeq_{FSD} Y_{11}|V_2 = v_2$ means that conditionally on V_2 , being externally assigned to both punishments (incarceration and fines), make someone less likely to reoffend than someone who is externally assigned to incarceration but with no fines. The second condition $P_2|P_1 = p'_1 \succeq_{FSD} P_2|P = p_1$ for $p'_1 > p_1$ suggests that the level of Judge's stringency is positively dependent for two types of punishments. Under those conditions, [Lemma 3](#) (iii) suggests that P_1 can be considered as monotone IV; that is, being more stringent on incarceration leads to less recidivism. Below, we will show how we could bound the MTE in such an empirical context using the monotone IV assumption.

3.2.2. *Imperfect IVs.* We will consider the following list of imperfect IVs:

Assumption 6 (Imperfect IVs).

- (a) **Affiliated IV:** We say the propensity score P is an Affiliated IV if the joint density f_{P,Y_d} of P and Y_d satisfies $f_{P,Y_d}(p, y)f_{P,Y_d}(p', y') \geq f_{P,Y_d}(p, y')f_{P,Y_d}(p', y)$ for any $p > p'$ and $y > y'$, where (p, y) and (p', y') belong to the joint support of (P, Y_d) ¹⁰.
- (b) **IHRD IV:** We say the propensity score P is an Inverse Hazard Rate Decreasing IV if $\frac{F_{Y_d|P}(y|p)}{f_{Y_d|P}(y|p)}$ is non-increasing in p for all y .

¹⁰It is also referred as $f_{P,Y_d}(y, p)$ being TP2 (Totally Positive of Order 2)

- (c) **MIV**: We say P is a monotone IV if for any pair $(p', p) \in \mathcal{P} \times \mathcal{P}$ such that $p' \geq p$, $P = p$, i.e. $Y_d|P = p' \succeq_{FSD} Y_d|P = p$. In other terms, $\mathbb{P}(Y_d > y|P = p)$ is non-decreasing in p for all $y \in \mathcal{Y}$.¹¹
- (d) **RTI-IV**: We say the propensity score P is a right tail increasing IV, if Y_d is right tail increasing in P , i.e. $RTI(Y_d|P)$ meaning that $\mathbb{P}(Y_d > y|P > p)$ is non-decreasing in p for all $y \in \mathcal{Y}$.
- (e) **LTD-IV**: We say the propensity score P is a left tail decreasing IV, if Y_d is left tail decreasing in P , i.e. $LTD(Y_d|P)$ meaning that $\mathbb{P}(Y_d \leq y|P \leq p)$ is non-increasing in p for all $y \in \mathcal{Y}$.
- (f) **PQD-IV**: We say the propensity score P is a positive quadrant dependent IV, if $\mathbb{P}(Y_d > y, P > p) \geq \mathbb{P}(Y_d > y)\mathbb{P}(P > p)$ for all $(y, p) \in \mathcal{Y} \times \mathcal{P}$.

We define the imperfect IV assumptions relative to the propensity score P . However, we can also define it with respect to the vector of instruments.¹² Affiliated IV means that it is more likely that the pair of realizations of Y_d and P simultaneously take high values or low values than for Y_d to take a high (resp. low) realization while P take a low (resp. high) realization. The MIV is a positive dependence restriction, which means that Y_d is more likely to take on larger values when P increases. The RTI-IV captures the fact that Y_d is more likely to take on larger values when P takes high values as well, while the LTD captures that Y_d is more likely to take lower values when P takes low values as well. Finally, the PQD-IV suggests that Y_d and P are more likely to take large values together or to be smaller together compare to the case if they were independent. The PQD type of assumption has been discussed in [Bhattacharya, Shaikh, and Vytlacil \(2012\)](#). The relation between these positive dependence concepts could be summarized as follows:¹³

$$\text{Affiliated IV} \Rightarrow \text{IHRD-IV} \Rightarrow \text{MIV} \Rightarrow \text{LTD-IV} \Rightarrow \text{PQD-IV},$$

$$\text{Affiliated IV} \Rightarrow \text{IHRD-IV} \Rightarrow \text{MIV} \Rightarrow \text{RTI-IV} \Rightarrow \text{PQD-IV}.$$

¹¹This property is also referred as Y_d being positively regression dependent on P .

¹²For instance, let " \succeq " denotes the component-wise partial order when comparing vectors, then we can state: for any pair $z' \succeq z$ in the support of a vector of observable variables Z , the conditional distribution of $Y_d, d \in \{0, 1\}$ given $Z = z'$ first order stochastically dominates the distribution of Y_d given $Z = z$, i.e. $Y_d|Z = z' \succeq_{FSD} Y_d|Z = z$. Then the partial ordering on \mathcal{Z} induces an ordering on \mathcal{P} which is what we considered in Assumption 6-(c).

¹³The proof of those implications can be found in [Joe \(1997, Theorem 2.3\)](#).

The next result shows that all these imperfect IV restrictions can be equivalently written in terms of restrictions only on the copula $C_{Y_d, P}(\cdot, \cdot)$.

Lemma 4. *Let Y_d and P be two continuous variables satisfying Assumption 5, then*

(a) *P is an affiliated IV if and only if*

$$c_{Y_d, P}(x_1, x_2)c_{Y_d, P}(x'_1, x'_2) \geq c_{Y_d, P}(x'_1, x_2)c_{Y_d, P}(x_1, x'_2)$$

for all $(x_1, x_2) \in [0, 1]$ and $(x'_1, x'_2) \in [0, 1]$ such that $x_1 \geq x'_1$ and $x_2 \geq x'_2$, where $c_{Y_d, P}(x_1, x_2) = \frac{\partial^2 C_{Y_d, P}(x_1, x_2)}{\partial x_1 \partial x_2}$ is the copula density.

(b) *P is an IHRD IV if and only if for any $x_1 \in [0, 1]$,*

$$\frac{\partial^2 \log \frac{\partial C_{Y_d, P}(x_1, x_2)}{\partial x_2}}{\partial x_1 \partial x_2} \geq 0.$$

(c) *P is an MIV if and only if $C_{Y_d, P}(x_1, x_2)$ is concave in x_2 for all $x_1 \in [0, 1]$,*

$$\frac{\partial^2 C_{Y_d, P}(x_1, x_2)}{\partial x_2^2} \leq 0. \quad (13)$$

(d) *P is an RTI-IV if and only if for any $x_2 \in [0, 1]$ and almost all x_1 ,*

$$\frac{\partial C_{Y_d, P}(x_1, x_2)}{\partial x_1} \geq \frac{[x_2 - C_{Y_d, P}(x_1, x_2)]}{1 - x_1}. \quad (14)$$

(e) *P is an LTD-IV if and only if for any $x_2 \in [0, 1]$ and almost all x_1 ,*

$$\frac{\partial C_{Y_d, P}(x_1, x_2)}{\partial x_1} \leq \frac{C_{Y_d, P}(x_1, x_2)}{x_1}. \quad (15)$$

(f) *P is an PQD-IV if and only if for all $(x_1, x_2) \in [0, 1]^2$,*

$$C_{Y_d, P}(x_1, x_2) \geq x_1 x_2. \quad (16)$$

Proof. See Appendix C.6.

This proposition has a significant practical advantage since it allows us to see how the identified set for the DMTRs under the STC assumption shrinks with each of the above assumptions. We do not need to recompute the identified set for the DMTRs for each of these assumptions; instead, we only have to intersect the set of copulas that rationalize the model under the STC with the set of

copulas that respect the IV restrictions that researchers would like to maintain. Notice that in this layered analysis, only the copula $C_{Y_d, P}(\cdot, \cdot)$ is involved.

Let us denote by $\mathcal{C}_d^{\mathbf{r}}$ the set of copulas $C_{Y_d, P}(\cdot, \cdot)$ that respect a given restriction \mathbf{r} . More precisely, $\mathcal{C}_d^{\mathbf{r}} \equiv \{C_{Y_d, P}(\cdot, \cdot) \in \mathcal{C}_d \text{ such that the restriction } \mathbf{r} \text{ is satisfied}\}$.

Theorem 3. *Suppose that Assumptions 1 and 5 are satisfied.*

$$\Theta_I^{\mathbf{r}} = \left\{ \tilde{\theta} \in \tilde{\Theta} : \text{For } d \in \{0, 1\}, (C_{Y_d, V|P}, C_{Y_d, P}) \in \mathcal{C}_d^c \times \mathcal{C}_d^{\mathbf{r}} \text{ satisfies Equation (10)} \right. \\ \left. \text{and } \forall y \in \mathcal{Y}, F_{Y_d}(y) = \Psi_{d, p}^{-1}(\mathbb{P}[Y \leq y, D = d | P = p]) \right\}$$

for $\mathbf{r} \in \{\text{Affiliated IV, MIV, RTI-IV, LTD-IV, PQD-IV}\}$

Remark 2. *While we restrict the statement of Theorem 3 to the restrictions*

$\mathbf{r} \in \{\text{Affiliated IV, MIV, RTI-IV, LTD-IV, PQD-IV}\}$ *for the copula $C_{Y_d, P}(\cdot, \cdot)$, it applies to any type of restrictions the researcher would like to impose on either $C_{Y_d, V|P}(\cdot, \cdot)$ or $C_{Y_d, P}(\cdot, \cdot)$. For instance, the HV05 identification assumptions, i.e. Assumptions 1 and 2 impose in our context that $C_{Y_d, V|P}(\cdot, \cdot) = C_{Y_d, V}(\cdot, \cdot)$ and $C_{Y_d, P}(x_1, x_2) = x_1 x_2$.*

Notice that if we have a sequence of restrictions $(\mathbf{r}_1, \dots, \mathbf{r}_j)$ on the copula $C_{Y_d, P}(\cdot, \cdot)$ such that \mathbf{r}_j is more restrictive than \mathbf{r}_l for $l < j$, we have the following: $\Theta_I^{\mathbf{r}_j} \subseteq \dots \subseteq \Theta_I^{\mathbf{r}_1}$. For instance, we have $\Theta_I^{\text{Affiliated IV}} \subseteq \Theta_I^{\text{MIV}} \subseteq \Theta_I^{\text{RTI-IV}} \subseteq \Theta_I^{\text{PQD-IV}}$. An interesting feature of our approach is that the set of equality restrictions that characterize the identified set does not change with \mathcal{C}_d^c nor $\mathcal{C}_d^{\mathbf{r}}$. This feature will ease the computation of the identified set of the DMTRs under different layers of assumptions. As previously discussed, the identified set for $(F_{Y_1}, F_{Y_0}, C_{Y_1, V|P}, C_{Y_0, V|P}, C_{Y_1, P}, C_{Y_0, P})$ has a particular structure that once the copulas are fixed, the marginal distributions of potential outcomes are uniquely determined. The “size” or “volume” of the projected identified set for (F_{Y_1}, F_{Y_0}) is then determined by how many or what kind of restrictions one would like to impose on the copulas. For example, if we assume $Y_d \perp P|V$ as in HV05, (F_{Y_1}, F_{Y_0}) become point-identified, see Corollary 1. If we do not make any assumptions on the dependence between Y_d and P , either conditioning on V or not, then we obtain the identified set as shown in Theorem 2. If we are willing to take a middle ground on the “perfectness” of the instrument P or have prior information on the type of selection into treatment,

we can use an analogous version of Theorem 3 that applies to the context and then obtain directly the identified set that corresponds to it.

3.3. Semi-parametric Identification. The identification results in Theorems 2 and 3 are fully non-parametric. In this subsection, we will consider an alternative approach by parametrizing the copulas with a finite-dimensional parameter θ . However, we will leave the marginals fully nonparametric. From this perspective, we consider semi-parametric identification in this subsection. As discussed in Chen, Fan, and Tsyrennikov (2006), such a semi-parametric approach has gained popularity in studying some features of multivariate distributions in diverse fields. It is flexible and circumvents the curse of dimensionality. It is worth-noting that the copula-based (partial) identification approach we propose below significantly differs from the one proposed in Chen, Fan, and Tsyrennikov (2006) and subsequent papers. In their models, all the marginals are nonparametrically (point) identified, while in our case, the marginal potential outcomes $F_{Y_d}(y), d \in \{0, 1\}$ are not (point) identified. We make the following assumption.

Assumption 7 (Parametric Copula). *There exists $\theta = (\alpha^0, \alpha^1, \delta^1, \delta^0) \in \Theta_0^c \times \Theta_1^c \times \Theta_0 \times \Theta_1 \equiv \Theta \subseteq \mathbb{R}^T$ with $T < \infty$ such that $C_{Y_d, P}(x_1, x_2) = C_{Y_d, P}(x_1, x_2; \alpha^d)$ and $C_{Y_d, V|P=p}(x_1, x_2) = C_{Y_d, V}(x_1, x_2; \sigma_d(p))$ where $\sigma_d(p)$ is known up to a finite number of parameters $\delta^d, d \in \{0, 1\}$.*

With this copula parametrization, our key unknown parameters of interest are θ and $F_{Y_d}(y), d \in \{0, 1\}$. Let \mathcal{F}^c define the set of continuous CDFs. The mapping $\Psi_{d,p}^{-1}(\mathbb{P}[Y \leq y, D = d|P = p]; \theta)$ is now known up to the finite dimensional parameter θ . The following theorem characterizes the identified set for $\tilde{\theta} \equiv (\theta, F_{Y_1}(\cdot), F_{Y_0}(\cdot)) \in \Theta \times \mathcal{F}^c \times \mathcal{F}^c$ under the semi-parametric specification.

Theorem 4. *Under Assumptions 1, 5 and 7, the semi-parametric identified set Θ_I^{SP} of $\tilde{\theta} \equiv (\theta, F_{Y_1}(\cdot), F_{Y_0}(\cdot))$ is characterized as follows:*

$$\Theta_I^{SP} = \left\{ \begin{array}{l} \text{For } d \in \{0, 1\}, \theta \in \Theta \text{ satisfying Equation (10),} \\ \text{and for all } \forall y \in \mathcal{Y}, F_{Y_d}(y) = \Psi_{d,p}^{-1}(\mathbb{P}[Y \leq y, D = d|P = p]; \theta) \end{array} \right\}.$$

If in addition, a restriction \mathbf{r} is imposed on the copula $C_{Y_d, P}(\cdot, \cdot)$, we have then:

$$\Theta_1^{SP, \mathbf{r}} = \left\{ \text{For } d \in \{0, 1\}, \theta \in \Theta_0^c \times \Theta_1^c \times \Theta_0^{\mathbf{r}} \times \Theta_1^{\mathbf{r}} \text{ satisfying Equation (10)}, \right. \\ \left. \text{for all } \forall y \in \mathcal{Y}, F_{Y_d}(y) = \Psi_{d,p}^{-1}(\mathbb{P}[Y \leq y, D = d | P = p]; \theta) \right\},$$

for $\mathbf{r} \in \{\text{Affiliated IV, MIV, RTI-IV, LTD-IV, PQD-IV}\}$.

Remark 3. As Theorem 4 shows, the identification of the infinite-dimensional parameter F_{Y_d} boils down to the identification of a finite-dimensional parameter θ . The sharp identification region of θ is characterized by a set of equality constraints, which are easy to work with because they only contain finite-dimensional parameters and known quantities.

Remark 4 (Specification tests). Recently, there have been an increasing number of papers that develop specifications tests for the assumptions usually maintained to identify causal effects, see for instance [Kitagawa \(2015\)](#), [Huber and Mellace \(2015\)](#), [Mourifié and Wan \(2017\)](#), and [Kédagni and Mourifié \(2017\)](#), etc. Our approach provides a unified way to do specification tests for the assumptions the researcher is willing to maintain. Indeed, each of the identified sets proposed in Theorems 2 to 4 can be empty if we cannot find copulas that respects the equality constraints. The “largest” identified set that imposes the minimum structure so far is the one derived in Theorem 3. If empty, this means imposing the STC specification for treatment selection is too stringent for the data. In Section 4 below, we relax this assumption to allow for multi-dimensional unobserved heterogeneity in the treatment selection.

Choice of Copulas: Frank, and Bernstein copulas. In many applications, it is unknown ex-ante if there is positive or negative selection into the treatment. In such contexts, it is essential to consider a comprehensive copula family for $C_{Y_d, V|P}$. Comprehensive parametric copulas are copulas that (i) approach the countermonotonicity copula (resp. comonotonicity copula), i.e., Fréchet Lower Bound (resp. Fréchet Upper Bound copula) for certain values of their parameters in their permissible range, (ii) and cover the entire domain between the Fréchet lower and upper copula bounds including the product copula as special case. Using these copulas, we may test the absence of selection by checking if the confidence region of $\sigma_d(p)$ excludes the value that corresponds to the product copula, which corresponds to the independence case. Comprehensive copulas, for instance Gaussian and Frank

copulas, parameterize the full range of dependence. On the other hand, non-comprehensive copulas such as Farlie-Gumbel-Morgenstern (FGM), Clayton, Gumbel, and Joe copulas, are only able to capture dependence in a limited manner. In practice, it will be useful to use a different family of copula to analyze how sensitive the results are depending on the copula parametrization.

3.3.1. *Frank Copula.* In Corollary 2 below, we show how the characterization of the identified set simplifies when considering the Frank Copula.¹⁴

Assumption 8 (Frank Copula). *There exists $\theta = (\alpha_0, \alpha_1, \delta_1, \delta_0) \in \Theta \subseteq \mathbb{R}^T$ with $T < \infty$ such that $C_{Y_d, P}(x_1, x_2) = -\frac{1}{\alpha_d} \ln \left[1 + \frac{(e^{-\alpha_d x_1} - 1)(e^{-\alpha_d x_2} - 1)}{(e^{-\alpha_d} - 1)} \right]$ for $\alpha_d \in (0, \infty)$ and $C_{Y_d, V|P=p}(x_1, x_2) = -\frac{1}{\sigma_d(p)} \ln \left[1 + \frac{(e^{-\sigma_d(p)x_1} - 1)(e^{-\sigma_d(p)x_2} - 1)}{(e^{-\sigma_d(p)} - 1)} \right]$ for $\sigma_d(p) \in (-\infty, 0) \cup (0, +\infty)$, $d \in \{0, 1\}$, where $\sigma_d(p)$ is known up to a finite number of parameters δ_d .*

Corollary 2. *Under Assumptions 1, 5, 6 and 8, the identified set Θ_I^{SP} of $\tilde{\theta} \equiv (\theta, (F_{Y_d}(y) : d \in \{0, 1\}, y \in \mathbb{R}))$ is characterized as follows:*

$$\Theta_I^{SP} = \left\{ \text{For } d \in \{0, 1\}, F_{Y_d}(y; \theta) = -\frac{1}{\alpha_d} \ln \left[1 + \frac{H_d(y, p, \sigma_d(p))(e^{-\alpha_d} - 1)}{e^{-\alpha_d F_P(p)} - H_d(y, p, \sigma_d(p))(e^{-\alpha_d F_P(p)} - 1)} \right] \right. \\ \left. \text{and } \theta = (\alpha_d, \delta_d) \text{ satisfies } \alpha_d f_P(p)(1 - H_d)H_d + \frac{\partial H_d}{\partial p} = 0 \text{ for all } p \text{ and } y. \right\}$$

where

$$H_1(y, p, \sigma_1(p)) = -\frac{1}{\sigma_1(p)} \ln \left[1 + \frac{(e^{-\sigma_1(p)F_{Y, D|P}(y, 1|p)} - 1)(e^{-\sigma_1(p)} - 1)}{(e^{-\sigma_1(p)p} - 1)} \right]$$

and

$$H_0(y, p, \sigma_0(p)) = \frac{1}{\sigma_0(p)} \ln \left[1 + \frac{(e^{\sigma_0(p)F_{Y, D|P}(y, 0|p)} - 1)(e^{-\sigma_0(p)} - 1)}{e^{-\sigma_0(p)} - e^{-\sigma_0(p)p}} \right]$$

and $\sigma_d(p)$ is parameterized by δ_d : $\sigma_d(p) \equiv \sigma_d(p; \delta_d)$.

Proof. See Appendix B.3.

Remark 5. *When the function σ_d (or its finite dimensional parameter δ_d) is given, α_d is uniquely determined. To see this, fixing an arbitrary y , integrating both side of $\alpha_d f_P(p)(1 - H_d)H_d + \frac{\partial H_d}{\partial p} =$*

¹⁴In Appendix B we derive a similar characterization for the FGM copula.

0 from 0 to 1 with respect to p gives

$$\begin{aligned} \alpha_d \mathbb{E}[H_d(y, P, \sigma_d(P; \delta_d))(1 - H_d(y, P, \sigma_d(P; \delta_d)))] + \int_0^1 \frac{\partial H_d}{\partial p} dp &= 0 \\ \Rightarrow \alpha_d &= -\frac{H_d(y, 1, \sigma_d(1; \delta_d)) - H_d(y, 0, \sigma_d(0; \delta_d))}{\mathbb{E}[H_d(y, P, \sigma_d(P; \delta_d))(1 - H_d(y, P, \sigma_d(P; \delta_d)))]} \end{aligned}$$

where (with abusing of notation) $\frac{\partial H_d(P)}{\partial p}$ denotes the derivative of $H_d(y, p, \sigma_d(p; \delta_d))$ with respect to p . Then δ_d can be estimated using a minimum distance estimator by inserting pre-estimated \hat{p} and \hat{f}_P . When inserting the true $\sigma_d(\cdot)$ into the right hand side of the above equation, we have $H_d(y, P, \sigma_d(P; \delta_d)) = F_{Y_d|P}(y|P)$ and hence

$$\alpha_d = -\frac{F_{Y_d|P}(y|1) - F_{Y_d|P}(y|0)}{\mathbb{E}[F_{Y_d|P}(y|P)(1 - F_{Y_d|P}(y|P))]}.$$

This states that α_d is positive iff the distribution of $Y_d|P = 1$ first order stochastically dominates $Y_d|P = 0$.

3.3.2. *Bernstein Copula.* We now turn our attention to the Bernstein copula. More precisely, let's introduce the following assumption:

Assumption 9 (Bernstein Copulas).

$$(i) C_{Y_d, V|P}(x_1, x_2) = C_{Y_d, V|P}(x_1, x_2; \alpha^d) = K_d L_d \sum_{k=1}^{K_d} \sum_{l=1}^{L_d} \alpha_{kl}^d B_{k-1, K_d-1}(x_1) B_{l-1, L_d-1}(x_2)$$

$$(ii) C_{Y_d, P}(x_1, x_2) = C_{Y_d, P}(x_1, x_2; \beta^d) = R_d S_d \sum_{r=1}^{R_d} \sum_{s=1}^{S_d} \beta_{rs}^d B_{r-1, R_d-1}(x_1) B_{s-1, S_d-1}(x_2)$$

where $B_{i, I}(u) = \int_0^u b_{i, I}(t) dt$ for $u \in [0, 1]$, $b_{i, I}(u) = \binom{I}{i} u^i (1-u)^{I-i}$, and $\alpha_{kl}^d \geq 0$ and $\beta_{rs}^d \geq 0$ are unknown parameters that satisfy $K_d \sum_{l=1}^{L_d} \alpha_{kl}^d = 1$, $L_d \sum_{k=1}^{K_d} \alpha_{kl}^d = 1$, $S_d \sum_{r=1}^{R_d} \beta_{rs}^d = 1$, $R_d \sum_{s=1}^{S_d} \beta_{rs}^d = 1$.¹⁵

The Bernstein copula is a very useful and an important copula since because of the Weirstrass-approximation theorem, we know that any copula can be approximated uniformly over $[0, 1]^2$ for K_d and L_d sufficiently large, see Kingsley (1951). Note that in Assumption 9-(i), the parameters α_{kl} do not depend on P , so it implicitly assumes that the joint distribution of Y_d and V depends on P through

¹⁵Please see Dou et al (2016) for this formulation of the Bernstein copula.

the marginal distribution $F_{Y_d|P}$. One special case of Assumption 9-(i) is “selection on observable”, which happens when $K_d = L_d = 1$, so that $F_{Y_d,V|P} = C_{Y_d,V|P}(F_{Y_d|P}, F_{V|P}) = F_{Y_d|P} \times F_{V|P}$. On the other hand, if we take $R_d = S_d = 1$, then Assumption 9-(ii) implies $F_{Y_d,P} = C_{Y_d,P}(F_{Y_d}, F_P) = F_{Y_d} \times F_P$. In this case, we have a valid instrument.

In Theorem 7 relegated in Appendix B.2, we characterize the identified set of the DMTR_g for arbitrary fixed values of $R_d, S_d, K_d, L_d, d \in \{0, 1\}$. For sake of simplicity, we present below the special case where $R_d = S_d = K_d = L_d = 2$, for $d \in \{0, 1\}$ and $g(Y) = Y$.

Let's consider the following notations:

$$\begin{aligned}\psi_{10}^d(p) &= \left(4\beta_{11}^d + 2(1 - 4\beta_{11}^d)F_P(p)\right)4\alpha_{11}^d, \\ \psi_{11}^d(p) &= (4\beta_{11}^d - 1)(4F_P(p) - 2)4\alpha_{11}^d + (4\beta_{11}^d + 2F_P(p)(1 - 4\beta_{11}^d))^2(1 - 4\alpha_{11}^d), \\ \psi_{12}^d(p) &= (12\beta_{11}^d + 6(1 - 4\beta_{11}^d)F_P(p))(1 - 4\alpha_{11}^d)(4\beta_{11}^d - 1)(2F_P(p) - 1), \\ \psi_{13}^d(p) &= 2(4\beta_{11}^d - 1)^2(2F_P(p) - 1)^2(1 - 4\alpha_{11}^d),\end{aligned}$$

with $\boldsymbol{\psi}_1^d(p) = (\psi_{10}^d(p), \psi_{11}^d(p), \psi_{12}^d(p), \psi_{13}^d(p))'$, and

$$\begin{aligned}\psi_{20}^d(p) &= -2\left(4\beta_{11}^d + 2(1 - 4\beta_{11}^d)F_P(p)\right), \\ \psi_{21}^d(p) &= 4(4\beta_{11}^d + 2F_P(p)(1 - 4\beta_{11}^d))^2 - 2(4\beta_{11}^d - 1)(4F_P(p) - 2), \\ \psi_{22}^d(p) &= (48\beta_{11}^d + 24(1 - 4\beta_{11}^d)F_P(p))(4\beta_{11}^d - 1)(2F_P(p) - 1), \\ \psi_{23}^d(p) &= 8(4\beta_{11}^d - 1)^2(2F_P(p) - 1)^2,\end{aligned}$$

with $\boldsymbol{\psi}_2^d(p) = (\psi_{20}^d(p), \psi_{21}^d(p), \psi_{22}^d(p), \psi_{23}^d(p))'$. In addition, let's define

$\boldsymbol{\gamma}^d \equiv (\gamma_0^d, \gamma_1^d, \gamma_2^d, \gamma_3^d)$ and then define the following sets:

$$\begin{aligned}\Theta^{BC} &\equiv \left\{(\alpha_{11}^d, \beta_{11}^d, \alpha_{11}^0, \beta_{11}^0) : 0 \leq \alpha_{11}^d \leq \frac{1}{2}, 0 \leq \beta_{11}^d \leq \frac{1}{2}, \text{ for } d \in \{0, 1\}\right\} \\ \Gamma &\equiv \left\{(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^0) : \liminf \mathcal{Y} \leq \gamma_j^d \leq \limsup \mathcal{Y}, \text{ for } j \in \{0, 1, 2, 3\}, \text{ for } d \in \{0, 1\}\right\}.\end{aligned}$$

In this special case the identified set of the DMTRs can be characterized as follows:

$$\Theta_I = \left\{ \left(\mathbb{E}[Y_1|V = v, P = p; \bar{\theta}], \mathbb{E}[Y_0|V = v, P = p; \bar{\theta}] \right) \text{ where} \right.$$

$$\mathbb{E}[Y_d|V = v, P = p; \bar{\theta}] = \boldsymbol{\psi}_1^d(p) \cdot \boldsymbol{\gamma}^d + (\boldsymbol{\psi}_2^d(p) \cdot \boldsymbol{\gamma}^d)v,$$

$$\forall \bar{\theta} \equiv (\alpha_{11}^1, \beta_{11}^1, \alpha_{11}^0, \beta_{11}^0, \boldsymbol{\gamma}^1, \boldsymbol{\gamma}^0) \in \Theta^{BC} \times \Gamma \text{ such that the following equations are satisfied}$$

$$\mathbb{E}[YD|P = p] = (\boldsymbol{\psi}_1^1(p) \cdot \boldsymbol{\gamma}^1)p + (\boldsymbol{\psi}_2^1(p) \cdot \boldsymbol{\gamma}^1)\frac{p^2}{2}$$

$$\mathbb{E}[Y(1 - D)|P = p] = (\boldsymbol{\psi}_1^0(p) \cdot \boldsymbol{\gamma}^0)(1 - p) + (\boldsymbol{\psi}_2^0(p) \cdot \boldsymbol{\gamma}^0)\left(\frac{1}{2} - \frac{p^2}{2}\right) \left. \right\}.$$

Notice that in this special case, imposing a MIV restriction is equivalent to simply restrict the range of β_{11}^d , from $0 \leq \beta_{11}^d \leq \frac{1}{2}$ to $\frac{1}{4} \leq \beta_{11}^d \leq \frac{1}{2}$.

Remark 6. *Our semi-parametric characterization of the DMTR_g identified set using Bernstein copula — Theorem 7 relegated in Appendix B.2, shows that when $R_d = S_d = 1$, meaning the IV is valid, we have $\mathbb{E}[g(Y_d)|V = v] = \sum_{l=1}^{L_d} \theta_{dl}^g b_{l-1, L_d-1}(v)$, then we recover the parametric form Mogstad, Santos, and Torgovitsky (2018) imposed on the MTR. Mogstad, Santos, and Torgovitsky (2018) approach imposes $\mathbb{E}[g(Y_d)|V = v] = \sum_{l=1}^{L_d} \theta_{dl}^g b_{l-1, L_d-1}(v)$ as a primitive, while in contrast we show that under a valid IV assumption —Assumption 2, imposing such a structure on the MTRs is equivalent to parametrize the “selection on unobservables” dependence — $C_{Y_d, V|P}(x_1, x_2; \alpha^d)$ — using a Bernstein copula of order L_d .*

4. EXTENSION: MULTIPLE THRESHOLD-CROSSING MODELS

There are many empirical applications where a model imposing only STC model cannot adequately model the selection to the treatment. Various examples are given in Heckman and Pinto (2018), Lee and Salanié (2018). In presence of multiple potential instruments, one way to relax the “strong” monotonicity assumption is to consider the “AM monotonicity” —in the language of Mogstad, Torgovitsky and Walters (2019), which can be modelled by considering the multiple hurdle model entertained in Lee and Salanié (2018). Our approach can be applied to the case where selection into treatment is defined by a finite number of thresholds. However, for the sake of simplicity, we will consider the case with two thresholds.

Assumption 10 (Double Hurdle (DH) model). *The selection mechanism is governed by $D = 1\{Q_1(Z) > V_1, Q_2(Z) > V_2\}$ for some measurable and non-trivial function (Q_1, Q_2) , where (V_1, V_2) has a joint continuous distribution over interval $[0, 1]^2$ with marginal uniform distributions and are statistically independent of the vector of $(Q_1(Z), Q_2(Z))$, i.e. $(Q_1(Z), Q_2(Z)) \perp (V_1, V_2)$.*

Unlike in the STC model, $Q_1(Z)$ and $Q_2(Z)$ are not readily identified from the choice probability $\mathbb{P}(D = 1|Z)$. Theorem 4.2 in [Lee and Salanié \(2018\)](#) provides conditions under which $Q_1(Z)$, $Q_2(Z)$ and the joint distribution $F_{V_1, V_2}(v_1, v_2)$ are non-parametrically identified from the propensity score $\mathbb{P}(D = 1|Z)$. Their non-parametric identification approach requires two continuous “exogenous” covariates that generate all possible values of the thresholds. In our current approach, the exogeneity refers only to the selection equation, i.e., $(Z \perp (V_1, V_2))$; Z could be correlated with the potential outcomes. Without loss of generality, we use Z_1 and Z_2 to denote those exogenous covariates such that $Q_1(Z) \equiv Q_1(Z_1)$ does not depend Z_2 and $Q_2(Z) \equiv Q_2(Z_2)$ does not depend on Z_1 .¹⁶ Hereafter, we will assume that the regularity conditions of [Lee and Salanié \(2018, Theorem 4.2\)](#) are valid and that $F_{V_1, V_2}(v_1, v_2)$, $Q_1(Z_1)$, and $Q_2(Z_2)$ are identifiable from the data. To ease the notation, we will write $\mathbf{V} = (V_1, V_2)$ and $\mathbf{Q} = (Q_1, Q_2)$. In the presence of multiple unobserved heterogeneity in the selection mechanism, we define the DMTR as follows:

$$\text{DMTR}_g^d(\mathbf{v}, \mathbf{q}) = \mathbb{P}[Y_d \leq y | \mathbf{V} = \mathbf{v}, \mathbf{Q} = \mathbf{q}] \equiv F_{Y_d | \mathbf{V}, \mathbf{Q}}(y | \mathbf{v}, \mathbf{q}),$$

for $\mathbf{v} \in [0, 1]^2$, $\mathbf{q} \in \mathcal{Q}_1 \times \mathcal{Q}_2$, and $d \in \{0, 1\}$. First, we can show that all conventional policy parameters can also be written as a weighted average of the DMTR_g^d even in the presence of multiple thresholds. Before doing so, let’s introduce the following assumption:

Assumption 11 (Conditional Policy Invariance). $Y_d^{a'} | \mathbf{V}^{a'}, \mathbf{Q}^{a'} \sim Y_d^a | \mathbf{V}^a, \mathbf{Q}^a$ with $\mathbf{V}^{a'} \sim \mathbf{V}^a$ and $Y_d^{a'} \sim Y_d^a$ for $a \neq a'$.

We have the following results for the double hurdle model.

Theorem 5. *Suppose that Assumption 10 is satisfied, then*

$$(i) \text{MTE}(\mathbf{v}) = \int_{\mathbf{q}} f_{\mathbf{Q}}(\mathbf{q}) \text{DMTE}(\mathbf{v}, \mathbf{q}) d\mathbf{q};$$

¹⁶For the entire list of requirements please see Assumption 4.2 in [Lee and Salanié \(2018\)](#). They also discussed identification under weaker conditions.

(ii) For any $s \in \{ATE, LATE(\mathbf{u}, \mathbf{u}'), ATT, ATUT\}$ ¹⁷ and weights $\omega^s(\mathbf{v}, \mathbf{q})$ listed in Table 2 below, we have

$$s = \int_{\mathbf{v}} \int_{\mathbf{q}} \omega^s(\mathbf{v}, \mathbf{q}) DMTE(\mathbf{v}, \mathbf{q}) d\mathbf{v} d\mathbf{q}. \quad (17)$$

(iii) If in addition Assumption 11 holds, Equation (17) holds with $s = PRTE$.

(iv) If in addition, *Lee and Salanié (2018, Assumption 4.2)* holds, then all the weights $\omega^s(\mathbf{v}, \mathbf{q})$ are point identified.

Proof. See Appendix C.7. □

TABLE 2. Policy Parameters and DMTE in the multiple thresholds case.

Parameters	weights $\omega^s(\mathbf{v}, \mathbf{q})$
ATE	$f_{\mathbf{Q}}(\mathbf{q})f_{\mathbf{V}}(\mathbf{v})$
ATT	$\frac{f_{\mathbf{Q}}(\mathbf{q})f_{\mathbf{V}}(\mathbf{v})1_{\{\mathbf{v} \in [0, q_1] \times [0, q_2]\}}}{\mathbb{E}[F_{\mathbf{V}}(\mathbf{Q})]}$
ATUT	$\frac{f_{\mathbf{Q}}(\mathbf{q})f_{\mathbf{V}}(\mathbf{v})1_{\{\mathbf{v} \notin [0, q_1] \times [0, q_2]\}}}{\mathbb{E}[1 - F_{\mathbf{V}}(\mathbf{Q})]}$
LATE(\mathbf{u}, \mathbf{u}')	$\frac{f_{\mathbf{Q}}(\mathbf{q})f_{\mathbf{V}}(\mathbf{v})1_{\{\mathbf{v} \in [u_1, u'_1] \times [u_2, u'_2]\}}}{F_{\mathbf{V}}(\mathbf{v} \in [u_1, u'_1] \times [u_2, u'_2])}$
PRTE	$\frac{[f_{\mathbf{Q}^{a'}}(\mathbf{q}) - f_{\mathbf{Q}^a}(\mathbf{q})]f_{\mathbf{V}^a}(\mathbf{v})1_{\{\mathbf{v} \in [0, q_1] \times [0, q_2]\}}}{\mathbb{E}[F_{\mathbf{V}}(\mathbf{Q}^{a'})] - \mathbb{E}[F_{\mathbf{V}}(\mathbf{Q}^a)]}$

Remark 7. In presence of valid IVs, i.e. $(Z_1, Z_2) \perp Y_d | \mathbf{V}$ our weights $\omega^s(\mathbf{v}, \mathbf{q})$ for any $s \in \{ATE, ATT, ATUT, PRTE\}$ collapse to the weights proposed by *Lee and Salanié (2018)* for the DH model. If in addition both \mathbf{V} and \mathbf{Q} are scalar-valued random variable, then the weights in Table 2 reduce to the weights in Table 1.

Notice that $LATE_g(\mathbf{u}, \mathbf{u}') \equiv \mathbb{E}[g(Y_1) - g(Y_0) | u_1 < V_1 \leq u'_1, u_2 < V_2 \leq u'_2]$ is a generalization of the LATE defined in *Imbens and Angrist (1994)* when the selection into treatment is defined by two thresholds. This type of parameter has recently received attention from empirical researchers, e.g. *Arteaga (2018)*, who suggested a way to identify this parameter under the assumption that the IV

¹⁷Here $LATE(\mathbf{u}, \mathbf{u}')$ represents the average treatment effect for the group of compliers when P is externally changed from \mathbf{u} to \mathbf{u}' : $LATE_g(\mathbf{u}, \mathbf{u}') \equiv \mathbb{E}[g(Y_1) - g(Y_0) | u_1 < V_1 \leq u'_1, u_2 < V_2 \leq u'_2]$.

is valid. Our approach extends and generalizes [Arteaga \(2018\)](#)'s identification approach to the case where the instruments are not necessarily valid.

Assumption 12. *The joint distribution of $(Y_d, \mathbf{V}, \mathbf{Q})$ is absolutely continuous respect to the Lebesgue measure.*

Lemma 5. *[Vine Copula] Under Assumptions 10 and 12, for $d \in \{0, 1\}$ we have for each $y \in \mathcal{Y}$ and $\mathbf{q} \in \mathcal{Q}_1 \times \mathcal{Q}_2$,*

$$F_{Y_d|Q_1}(y|q_1) = \frac{\partial}{\partial x_2} C_{Y_d, Q_1}(x_1, x_2) \Big|_{x_1=F_{Y_d}(y), x_2=F_{Q_1}(q_1)} \equiv c_{d, F_{Q_1}(q_1)}^I(F_{Y_d}(y)), \quad (18)$$

$$F_{Y_d|Q}(y|\mathbf{q}) = \frac{\partial}{\partial x_2} C_{Y_d, Q_1|Q_2=q_2}(x_1, x_2) \Big|_{x_1=F_{Y_d|Q_1}(y|q_1), x_2=F_{Q_1|Q_2}(q_1|q_2)} \equiv c_{d, F_{Q_1|Q_2}}^{II}(F_{Y_d|Q_1}(y|q_1)), \quad (19)$$

$$F_{Y_d|V_2, Q}(y|v_2, \mathbf{Q}) = \frac{\partial}{\partial x_2} C_{Y_d, V_2|Q=\mathbf{q}}(x_1, x_2) \Big|_{x_1=F_{Y_d|Q}(y|\mathbf{q}), x_2=v_2} \equiv c_{d, v_2}^{III}(F_{Y_d|Q}(y|\mathbf{q})), \quad (20)$$

$$F_{Y_d|V, Q}(y|\mathbf{v}, \mathbf{Q}) = \frac{\partial}{\partial x_2} C_{Y_d, V_1|Q=\mathbf{q}}(x_1, x_2) \Big|_{x_1=F_{Y_d|V_2, Q}(y|v_2, \mathbf{q}), x_2=F_{V_1|V_2}(v_1|v_2)}, \quad (21)$$

and there exists monotone mapping $\Psi_{1, \mathbf{q}}$ and $\Psi_{0, \mathbf{q}}$ such that for each $y \in \mathcal{Y}$ and $\mathbf{q} \in \mathcal{Q}_1 \times \mathcal{Q}_2$,

$$\mathbb{P}[Y \leq y, D = 1 | \mathbf{Q} = \mathbf{q}] = \Psi_{1, \mathbf{q}}(F_{Y_1}(y)), \quad (22)$$

and

$$\mathbb{P}[Y \leq y, D = 0 | \mathbf{Q} = \mathbf{q}] = \Psi_{0, \mathbf{q}}(F_{Y_0}(y)) \quad (23)$$

where the expressions for $\Psi_{d, \mathbf{q}}$ are collected in [Appendix C.8](#).

Proof. See [Appendix C.8](#). □

Given [Lemma 5](#), the identified set for DMTR or DMTE can be constructed as in [Theorems 2 to 4](#).

5. CONCLUSION

In this paper, we show how to use the MTE framework to perform a layered sensitivity analysis in the program evaluation literature under the assumption that the propensity score is continuously distributed, as assumed in HV05. There are various empirical applications where all potential IVs are discrete random variables. We may be able to extend our framework to incorporate discrete

invalid IVs and discrete outcomes using the sub-copula approach entertained in [Mourifié \(2015\)](#) and [Torgovitsky \(2019\)](#). We leave this important question for future research.

APPENDIX A. TESTING MTE ASSUMPTIONS

Theorem 6 (Sharp characterization of the MTE assumptions). *Let $Y, D, Y_1, Y_0, P(Z)$ define a potential outcome model $Y = Y_1D + Y_0(1 - D)$. (i) If Assumptions 1 and 2 hold, then for all $y, y' \in \mathcal{Y}$, $\mathbb{P}(y < Y \leq y', D = 1|P = p)$ and $-\mathbb{P}(y < Y \leq y', D = 0|P = p)$ are non-decreasing in p for all $p \in \mathcal{P}$. (ii) If for all $y, y' \in \mathcal{Y}$ $\mathbb{P}(y < Y \leq y', D = 1|P = p)$ and $-\mathbb{P}(y < Y \leq y', D = 0|P = p)$ are non-decreasing in p for all $p \in \mathcal{P}$, there exists a joint distribution of $(\tilde{Y}, \tilde{Y}_1, \tilde{Y}_0, P(Z))$ such that Assumptions 1 and 2 hold, and $(\tilde{Y}, \tilde{D}, P(Z))$ has the same distribution as $(Y, D, P(Z))$.*

Proof. See Appendix C.9.

It is worth noting that testable implications of the MTE assumptions has been previously derived in Heckman and Vytlacil (2005, Appendix A). More precisely, HV05 derived testable implications of Assumptions 1 and 2 are: for any non-negative measurable function, i.e. $G^+(\cdot) : \mathcal{Y} \rightarrow \mathbb{R}^+$ we have

$$\frac{\partial}{\partial p} \mathbb{E}[G^+(Y)D|P = p] \geq 0, \quad (24)$$

$$\frac{\partial}{\partial p} \mathbb{E}[G^+(Y)(1 - D)|P = p] \leq 0. \quad (25)$$

The main contributions of Theorem 6 are (i) it shows that the testable implication is sharp, and (ii) it shows that we do not need to visit all those non-negative measurable functions, but we can restricts our attention to a tractable sub-class which is sufficient to screen all possible observable violations of the MTE assumptions. This tractable characterization has the direct advantage to propose a formal statistical test to screen the violation of the MTE assumptions.

A.1. Testing Procedures. The sharp testable implications of the MTE assumptions can be summarized as follows:

$$\mathbb{P}(y < Y \leq y', D = 1|P = \tilde{p}) \geq \mathbb{P}(y < Y \leq y', D = 1|P = p), \quad (26)$$

$$-\mathbb{P}(y < Y \leq y', D = 0|P = \tilde{p}) \geq -\mathbb{P}(y < Y \leq y', D = 0|P = p), \quad (27)$$

for all $\tilde{p} \geq p$ with $(\tilde{p}, p) \in \mathcal{P}^2$ and $y, y' \in \mathcal{Y}$.

Recently, Hsu et al (2018) propose a uniform a valid test for null hypothesis of the type:

$$\mathbb{E}[f^{(1)}(W, \tau)|S = \tilde{s}, X = x] \geq \mathbb{E}[f^{(2)}(W, \tau)|S = s, X = x], \quad (28)$$

$$\tilde{s} \geq s \text{ with } (\tilde{s}, s) \in \mathcal{S}^2, x \in \mathcal{X}, \text{ and } \tau \in \Omega,$$

where W is a vector of observed random variables that contains S and X as elements, $f^{(1)}(W, \tau), f^{(2)}(W, \tau)$ are known real valued functions indexed by $\tau \in \Omega$ —which can be finite or infinite. By defining the following $\tau \equiv (d, k) \in \Omega \equiv \{(d, k) : d \in \{0, 1\}, k = (y, y') : -\infty \leq y \leq y' \leq \infty\}$, $f^{(1)}(W, (1, k)) = f^{(2)}(W, (1, k)) = 1\{Y \in C_k\}1\{D = 1\}$, $f^{(1)}(W, (0, k)) = f^{(2)}(W, (0, k)) = -1\{Y \in C_k\}1\{D = 0\}$ with $C_k \equiv [y, y'] \cap \mathcal{Y}$, and $P = S$; we see that the MTE sharp testable implications, i.e. eqs (26,27) could be recasted into Hsu, Liu, and Shi (2019) testing framework if P was an observed random variable. In our context, P is not observed but could be consistently estimated. To test the MTE

testable implications one can proceed in two steps: (a) the first step consist in finding a consistent estimator for P , i.e. \hat{P} , (b) second consist in using [Hsu, Liu, and Shi \(2019\)](#) testing procedure with \hat{P} . A theoretical challenge is to ensure that the pre-estimation does not affect the statistical properties of [Hsu, Liu, and Shi \(2019\)](#) testing procedure.

APPENDIX B. ADDITIONAL RESULTS ON COPULAS

B.1. Farlie-Gumbel-Morgenstern (FGM) Copula.

Assumption 13 (FGM Copula). *There exists $\theta = (\alpha_0, \alpha_1, \delta_1, \delta_0) \in \Theta \subseteq \mathbb{R}^T$ with $T < \infty$ such that $C_{Y_d, P}(x_1, x_2) = x_1 x_2 (1 + \alpha_d (1 - x_1)(1 - x_2))$ and $C_{Y_d, V|P=p}(x_1, x_2) = x_1 x_2 (1 + \sigma_d(p)(1 - x_1)(1 - x_2))$ for $\sigma_d(p) \in [-1, 1]$, $d \in \{0, 1\}$, where $\sigma_d(p)$ is known up to a finite number of parameters δ_d .*

Corollary 3. *Under Assumptions 1, 5, 6 and 13, the identified set Θ_1^{SP} of $\tilde{\theta} \equiv (\theta, (F_{Y_d}(y) : d \in \{0, 1\}, y \in \mathbb{R}))$ is characterized as follows:*

$$\Theta_1^{SP} = \left\{ \text{For } d \in \{0, 1\}, F_{Y_d}(y; \theta) = \frac{\alpha_d(1 - 2F_P(p)) + 1 - \sqrt{(\alpha_d(1 - 2F_P(p)) + 1)^2 - 4\alpha_d(1 - 2F_P(p))H_d}}{2\alpha_d(1 - 2F_P(p))} \right. \\ \left. \theta = (\alpha_d, \delta_d) \text{ satisfies } (\Delta_d - \Delta_d^{\frac{1}{2}}) \frac{\partial B_d}{\partial p} = B_d \frac{\partial \Delta_d}{\partial p} \text{ for all } p \right\},$$

where $B_d = \alpha_d(1 - 2F_P(p))$, $\Delta_d = (B_d + 1)^2 - 4B_d H_d$, and

$$H_1(y, p, \sigma_1(p)) = \frac{\sigma_1(p)p(1-p) + p - \sqrt{(\sigma_1(p)p(1-p) + p)^2 - 4\sigma_1(p)p(1-p)F_{Y, D|P}(y, 1|p)}}{2\sigma_1(p)p(1-p)}.$$

and

$$H_0(y, p, \sigma_0(p)) = \frac{\sigma_0(p)p(1-p) + p + 1 + \sqrt{(\sigma_0(p)p(1-p) + p + 1)^2 + 4\sigma_0(p)p(1-p)F_{Y, D|P}(y, 0|p)}}{2\sigma_0(p)p(1-p)}.$$

Proof. Consider $d = 1$. Note first by equation (8) and assumption 13,

$$F_{Y, D|P}(y, 1|p) = c_{1, F_P(p)}(F_{Y_1}(y))p(1 + \sigma_1(p)(1 - c_{1, F_P(p)}(F_{Y_1}(y)))(1 - p)).$$

Let $A = \sigma_1(p)p(1 - p)$. We focus on the case $A \neq 0$, that is, we do not consider $p = 0$ or $p = 1$ or values of p such that $\sigma_1(p) = 0$, because the solution is straightforward for those values. From the above equation, we have two possible solutions for $c_{1, F_P(p)}(F_{Y_1}(y))$,

$$\frac{A + p + \sqrt{(A + p)^2 - 4AF_{Y, D|P}(y, 1|p)}}{2A} \quad \text{or} \quad \frac{A + p - \sqrt{(A + p)^2 - 4AF_{Y, D|P}(y, 1|p)}}{2A}$$

The expression in the square root sign is always non-negative. This is obviously true when $A < 0$. When $A > 0$,

$$(A + p)^2 - 4AF_{Y, D|P}(y, 1|p) = (A - p)^2 + 4A(p - F_{Y, D|P}(y, 1|p)) \geq 4A(p - F_{Y, D|P}(y, 1|p)) \geq 0,$$

where the last inequality holds because $p = \mathbb{P}(D = 1|P = p) \geq \mathbb{P}(Y \leq y, D = 1|P = p) = F_{Y,D|P}(y, 1|p)$.

Although both solutions are well defined, the first solution is not valid for this context. When $\sigma_1(p) > 0$, we have $p > A > 0$ (unless $p = 0$), and,

$$\frac{A + p + \sqrt{(A + p)^2 - 4AF_{Y,D|P}(y, 1|p)}}{2A} \geq \frac{A + p}{2A} > 1.$$

When $\sigma_1(p) < 0$, we have $A < 0$ and $p + A > 0$, and,

$$\frac{A + p + \sqrt{(A + p)^2 - 4AF_{Y,D|P}(y, 1|p)}}{2A} < 0$$

because the numerator and denominator have opposite sign. Therefore, the first solution can not be valid since $c_{1,F_P(p)}(F_{Y_1}(y)) = F_{Y_d|P}(y, p)$ must takes values from $[0, 1]$.

It remains to verify the second solution is valid. First consider the case $A > 0$. Let $\Delta = (A + p)^2 - 4AF_{Y,D|P}(y, 1|p)$. Then,

$$\begin{aligned} 0 \leq F_{Y,D|P}(y, 1|p) \leq p &\Rightarrow (A - p)^2 \leq (A + p)^2 - 4AF_{Y,D|P}(y, 1|p) \leq (A + p)^2 \\ &\Rightarrow 0 \leq A + p - \sqrt{\Delta} \leq 2A \Rightarrow 0 \leq \frac{A + p - \sqrt{\Delta}}{2A} \leq 1 \end{aligned}$$

Next consider the case $A < 0$,

$$\begin{aligned} 0 \leq F_{Y,D|P}(y, 1|p) \leq p &\Rightarrow (p + A)^2 \leq (A + p)^2 - 4AF_{Y,D|P}(y, 1|p) \leq (p - A)^2 \\ &\Rightarrow 2A \leq A + p - \sqrt{\Delta} \leq 0 \Rightarrow 0 \leq \frac{A + p - \sqrt{\Delta}}{2A} \leq 1 \end{aligned}$$

To summarize, we have one valid solution

$$c_{1,F_P(p)}(F_{Y_1}(y)) = \frac{\sigma_1(p)p(1-p) + p - \sqrt{(\sigma_1(p)p(1-p) + p)^2 - 4\sigma_1(p)p(1-p)F_{Y,D|P}(y, 1|p)}}{2\sigma_1(p)p(1-p)} \equiv H_1(y, p, \sigma_1(p)).$$

The right hand side only depends on unknown finite dimensional parameters and quantities that can be identified from data.

Next, recall that

$$\begin{aligned} H_1(y, p, \sigma_1(p)) &= c_{1,F_P(p)}(F_{Y_1}(y)) = \frac{\partial C_{Y_1,P}(x_1, x_2)}{\partial x_2} \Big|_{x_1=F_{Y_1}(y), x_2=F_P(p)} \\ &= x_1 + \alpha_1 x_1(1 - x_1)(1 - 2x_2) \Big|_{x_1=F_{Y_1}(y), x_2=F_P(p)} = F_{Y_1}(y) + \alpha_1 F_{Y_1}(y)(1 - F_{Y_1}(y))(1 - 2F_P(p)) \end{aligned}$$

Solve $F_{Y_1}(y)$ from the above equation, we get again two possible solutions.

$$\frac{B_1 + 1 + \sqrt{(B_1 + 1)^2 - 4B_1 H_1}}{2B_1}, \quad \text{or} \quad \frac{B_1 + 1 - \sqrt{(B_1 + 1)^2 - 4B_1 H_1}}{2B_1}$$

where $B_1 = \alpha_1(1 - 2F_P(p))$. Again, we restrict our attention to the case $B_1 \neq 0$, otherwise $F_{Y_1}(y)$ has a unique solution which equals to $H_1(y, p, \sigma_1(p))$. Note also that $B_1 \in [-1, 1]$, so $B_1 + 1 \geq 2B_1$ and $B_1 + 1 \geq 0$. Following similar argument as above (and use the fact that $0 \leq H_1 \leq 1$, we can show that the first solution is not valid while the second solution is. Therefore, it must be the case that

$$F_{Y_1}(y) = \frac{\alpha_1(1 - 2F_P(p)) + 1 - \sqrt{(\alpha_1(1 - 2F_P(p)) + 1)^2 - 4\alpha_1(1 - 2F_P(p))H_1(y, p, \sigma_1(p))}}{2\alpha_1(1 - 2F_P(p))},$$

where

$$H_1 = \frac{\sigma_1(p)p(1-p) + p - \sqrt{(\sigma_1(p)p(1-p) + p)^2 - 4\sigma_1(p)p(1-p)F_{Y,D|P}(y, 1|p)}}{2\sigma_1(p)p(1-p)}.$$

Now consider $d = 0$. Note first by equation (9) and assumption 13,

$$F_{Y,D|P}(y, 0|p) = c_{0,F_P(p)}(F_{Y_0}(y)) - c_{0,F_P(p)}(F_{Y_0}(y))p(1 + \sigma_0(p)(1 - c_{0,F_P(p)}(F_{Y_0}(y)))(1 - p)).$$

From the above equation, we have two possible solutions for $c_{0,F_P(p)}(F_{Y_0}(y))$,

$$\frac{A_0 + p - 1 + \sqrt{(A_0 + p - 1)^2 + 4A_0F_{Y,D|P}(y, 0|p)}}{2A_0} \quad \text{or} \quad \frac{A_0 + p - 1 - \sqrt{(A_0 + p - 1)^2 + 4A_0F_{Y,D|P}(y, 0|p)}}{2A_0}$$

where $A_0 = \sigma_0(p)p(1-p)$, so $A_0 + p - 1 = (1-p)(\sigma_0(p)p - 1) < 0$. The term inside the square root is always non-negative. It is obvious when $A_0 > 0$. When $A_0 < 0$, we have

$$\begin{aligned} (A_0 + p - 1)^2 + 4A_0F_{Y,D|P}(y, 0|p) &= (A_0 - p + 1)^2 + 4A_0(F_{Y,D|P}(y, 0|p) - (1 - p)) \\ &\geq 4A_0(F_{Y,D|P}(y, 0|p) - (1 - p)) = 4A_0(P(Y > y, D = 1|P = p) - P(Y > y|P = p)) \geq 0. \end{aligned}$$

Although both solutions are well-defined, we will argue the second solution is not valid. To see this, when $A_0 > 0$, the numerator is negative and denominator is positive, implies

$$\frac{A_0 + p - 1 - \sqrt{(A_0 + p - 1)^2 + 4A_0F_{Y,D|P}(y, 0|p)}}{2A_0} < 0.$$

When $A_0 < 0$, since $A_0 + 1 - p > 0$ so $2A_0 > A_0 + p - 1$, we have

$$\frac{A_0 + p - 1 - \sqrt{(A_0 + p - 1)^2 + 4A_0F_{Y,D|P}(y, 0|p)}}{2A_0} \geq \frac{A_0 + p - 1}{2A_0} > \frac{2A_0}{2A_0} = 1.$$

On the other hand, the first solution always falls between 0 and 1. When $\sigma_0(p) > 0$, $A_0 > 0$, we have

$$\begin{aligned} \frac{A_0 + p - 1 + \sqrt{(A_0 + p - 1)^2 + 4A_0F_{Y,D|P}(y, 0|p)}}{2A_0} &= \frac{A_0 + p - 1 + \sqrt{(A_0 - p + 1)^2 + 4A_0(F_{Y,D|P}(y, 0|p) - (1 - p))}}{2A_0} \\ &\leq \frac{A_0 + p - 1 + \sqrt{(A_0 - p + 1)^2}}{2A_0} = \frac{2A_0}{2A_0} = 1, \end{aligned}$$

and

$$\frac{A_0 + p - 1 + \sqrt{(A_0 + p - 1)^2 + 4A_0F_{Y,D|P}(y, 0|p)}}{2A_0} \geq \frac{A_0 + p - 1 + \sqrt{(A_0 + p - 1)^2}}{2A_0} \geq 0.$$

When $\sigma_0(p) < 0$, $A_0 < 0$, we have

$$\frac{A_0 + p - 1 + \sqrt{(A_0 + p - 1)^2 + 4A_0F_{Y,D|P}(y, 0|p)}}{2A_0} \geq \frac{A_0 + p - 1 + \sqrt{(A_0 + p - 1)^2}}{2A_0} = \frac{0}{2A_0} = 0,$$

$$\begin{aligned} \frac{A_0 + p - 1 + \sqrt{(A_0 + p - 1)^2 + 4A_0F_{Y,D|P}(y, 0|p)}}{2A_0} &= \frac{A_0 + p - 1 + \sqrt{(A_0 - p + 1)^2 + 4A_0(F_{Y,D|P}(y, 0|p) - (1 - p))}}{2A_0} \\ &\leq \frac{A_0 + p - 1 + \sqrt{(A_0 - p + 1)^2}}{2A_0} = \frac{2A_0}{2A_0} = 1, \end{aligned}$$

To conclude, we must have

$$c_{0,F_P(p)}(F_{Y_0}(y)) = \frac{\sigma_0 p(1-p) + p + 1 + \sqrt{(\sigma_0 p(1-p) + p + 1)^2 + 4\sigma_0 p(1-p)F_{Y,D|P}(y, 0|p)}}{2\sigma_0 p(1-p)} \equiv H_0(y, p, \sigma_0(p)).$$

Finally, repeating what we did for $F_{Y_1}(y)$ and $c_{1,F_P(p)}(F_{Y_1}(y))$, we can do the same and obtain

$$F_{Y_0}(y) = \frac{\alpha_0(1 - 2F_P(p)) + 1 - \sqrt{(\alpha_0(1 - 2F_P(p)) + 1)^2 - 4\alpha_0(1 - 2F_P(p))H_0(y, p, \sigma_0(p))}}{2\alpha_0(1 - 2F_P(p))}.$$

□

B.2. Bernstein Copulas. The density of the Bernstein copula is given by:

$$c_{Y_d, V|P}(x_1, x_2; \alpha^d) = K_d L_d \sum_{k=1}^{K_d} \sum_{l=1}^{L_d} \alpha_{kl}^d b_{k-1, K_d-1}(x_1) b_{l-1, L_d-1}(x_2)$$

Note that $b_{i,I}(u)$ has an alternative representation:

$$b_{i,I}(u) = \sum_{j=i}^I (-1)^{j-i} \binom{I}{j} \binom{j}{i} u^j \quad (29)$$

First let assume that Assumptions 1 and 5 hold, then under Assumption 9(i), we have

$$\begin{aligned} \mathbb{E}[g(Y_d)|V = v, P = p] &= \int_{\mathcal{Y}} g(y) K_d L_d \sum_{k=1}^{K_d} \sum_{l=1}^{L_d} \alpha_{kl}^d b_{k-1, K_d-1}(F_{Y_d|P}(y|p)) b_{l-1, L_d-1}(v) f_{Y_d|P}(y|p) dy \\ &= K_d L_d \sum_{k=1}^{K_d} \sum_{l=1}^{L_d} \alpha_{kl}^d b_{l-1, L_d-1}(v) \int_{\mathcal{Y}} g(y) b_{k-1, K_d-1}(F_{Y_d|P}(y|p)) f_{Y_d|P}(y|p) dy. \end{aligned}$$

In addition, under Assumption 9(ii) we can derive the following:

$$F_{Y_d|P}(y|p) = R_d S_d \sum_{r=1}^{R_d} \sum_{s=1}^{S_d} \underbrace{\beta_{rs}^d B_{r-1, R_d-1}(F_{Y_d}(y)) b_{s-1, S_d-1}(F_P(p))}_{\chi_{rs}^d},$$

$$f_{Y_d|P}(y|p) = f_{Y_d}(y) R_d S_d \sum_{r=1}^{R_d} \sum_{s=1}^{S_d} \underbrace{\beta_{rs}^d b_{r-1, R_d-1}(F_{Y_d}(y)) b_{s-1, S_d-1}(F_P(p))}_{\zeta_{rs}^d}.$$

To ease the notation, when there is no confusion we will make the following abuse of notation $F_P \equiv F_P(p)$ and $F_{Y_d} \equiv F_{Y_d}(y)$.

$$\begin{aligned} b_{k-1, K_d-1}(F_{Y_d|P}(y|p)) &= \sum_{j=k-1}^{K_d-1} (-1)^{j-k+1} \binom{K_d-1}{j} \binom{j}{k-1} (F_{Y_d|P}(y|p))^j \\ &= \sum_{j=k-1}^{K_d-1} (-1)^{j-k+1} \binom{K_d-1}{j} \binom{j}{k-1} (R_d S_d)^j \left(\sum_{r=1}^{R_d} \sum_{s=1}^{S_d} \chi_{rs}^d \right)^j \\ &= \sum_{j=k-1}^{K_d-1} (-1)^{j-k+1} \binom{K_d-1}{j} \binom{j}{k-1} (R_d S_d)^j \sum_{n_{11}+n_{12}+\dots+n_{R_d S_d}=j} \binom{j}{n_{11}, n_{12}, \dots, n_{R_d S_d}} (\chi_{11}^d)^{n_{11}} (\chi_{12}^d)^{n_{12}} \dots (\chi_{R_d S_d}^d)^{n_{R_d S_d}} \end{aligned}$$

and,

$$\begin{aligned} b_{k-1, K_d-1}(F_{Y_d|P}(y|p)) f_{Y_d|P}(y|p) &= \sum_{r=1}^{R_d} \sum_{s=1}^{S_d} \left\{ \sum_{j=k-1}^{K_d-1} (-1)^{j-k+1} \binom{K_d-1}{j} \binom{j}{k-1} (R_d S_d)^{j+1} \right. \\ &\quad \left. \sum_{n_{11}+n_{12}+\dots+n_{R_d S_d}=j} \binom{j}{n_{11}, n_{12}, \dots, n_{R_d S_d}} (\chi_{11}^d)^{n_{11}} (\chi_{12}^d)^{n_{12}} \dots (\chi_{R_d S_d}^d)^{n_{R_d S_d}} \right\} \zeta_{rs}^d f_{Y_d}(y) \end{aligned}$$

We have the following factorization:

$$\begin{aligned} &(\chi_{11}^d)^{n_{11}} (\chi_{12}^d)^{n_{12}} \dots (\chi_{R_d S_d}^d)^{n_{R_d S_d}} \zeta_{rs}^d \\ &= \beta_{rs}^d b_{s-1, S_d-1}(F_P) \prod_{e=1, f=1}^{R_d, S_d} (\beta_{ef}^d)^{n_{ef}} \prod_{e=1}^{R_d} (B_{e-1, R_d-1}(F_{Y_d}))^{n_e} \prod_{f=1}^{S_d} (b_{f-1, S_d-1}(F_P))^{n_f} b_{r-1, R_d-1}(F_{Y_d}) \end{aligned}$$

where $n_e = \sum_{f=1}^{S_d} n_{ef}$ and $n_f = \sum_{e=1}^{R_d} n_{ef}$. Then, we have

$$\begin{aligned} &\int_{\mathcal{Y}} g(y) (\chi_{11}^d)^{n_{11}} (\chi_{12}^d)^{n_{12}} \dots (\chi_{R_d S_d}^d)^{n_{R_d S_d}} \zeta_{rs}^d f_{Y_d}(y) dy \\ &= \beta_{rs}^d b_{s-1, S_d-1}(F_P) \prod_{f=1}^{S_d} (b_{f-1, S_d-1}(F_P))^{n_f} \prod_{e=1, f=1}^{R_d, S_d} (\beta_{ef}^d)^{n_{ef}} \underbrace{\mathbb{E} \left[g(Y_d) \prod_{e=1}^{R_d} (B_{e-1, R_d-1}(F_{Y_d}))^{n_e} b_{r-1, R_d-1}(F_{Y_d}) \right]}_{\gamma_{r-1, g}^{d, n_e}} \end{aligned}$$

Therefore, we can write:

$$\int_{\mathcal{Y}} g(y) b_{k-1, K_d-1}(F_{Y_d|P}(y|p)) f_{Y_d|P}(y|p) dy = \sum_{r=1}^{R_d} \sum_{s=1}^{S_d} \beta_{rs}^d b_{s-1, S_d-1}(F_P) \left\{ \sum_{j=k-1}^{K_d-1} (-1)^{j-k+1} \binom{K_d-1}{j} \binom{j}{k-1} (R_d S_d)^{j+1} \times \right. \\ \left. \sum_{n_{11}+n_{12}+\dots+n_{R_d S_d}=j} \binom{j}{n_{11}, n_{12}, \dots, n_{R_d S_d}} \prod_{e=1, f=1}^{R_d, S_d} (\beta_{ef}^d)^{n_{ef}} \prod_{f=1}^{S_d} (b_{f-1, S_d-1}(F_P))^{n_{\cdot f}} \gamma_{r-1, g}^{d, n_{\cdot e}} \right\}$$

Finally, we have

$$\mathbb{E}[g(Y_d)|V = v, P = p] = K_d L_d \sum_{k=1}^{K_d} \sum_{l=1}^{L_d} \alpha_{kl}^d b_{l-1, L_d-1}(v) \sum_{r=1}^{R_d} \sum_{s=1}^{S_d} \beta_{rs}^d b_{s-1, S_d-1}(F_P) \left\{ \sum_{j=k-1}^{K_d-1} (-1)^{j-k+1} \binom{K_d-1}{j} \binom{j}{k-1} (R_d S_d)^{j+1} \times \right. \\ \left. \sum_{n_{11}+n_{12}+\dots+n_{R_d S_d}=j} \binom{j}{n_{11}, n_{12}, \dots, n_{R_d S_d}} \prod_{e=1, f=1}^{R_d, S_d} (\beta_{ef}^d)^{n_{ef}} \prod_{f=1}^{S_d} (b_{f-1, S_d-1}(F_P))^{n_{\cdot f}} \gamma_{r-1, g}^{d, n_{\cdot e}} \right\} \quad (30)$$

Remark that when $R_d = S_d = 1$, $C_{Y_d, P}(x_1, x_2; \beta^d) = x_1 x_2$ which is equivalent to $Y_d \perp P$, in such a case Equation (30) simplifies to

$$\mathbb{E}[g(Y_d)|V = v, P = p] = K_d L_d \sum_{k=1}^{K_d} \sum_{l=1}^{L_d} \alpha_{kl}^d b_{l-1, L_d-1}(v) \underbrace{\int_{\mathcal{Y}} g(y) b_{k-1, K_d-1}(F_{Y_d}(y)) f_{Y_d}(y) dy}_{\tau_{gk}^d} \\ = \sum_{l=1}^{L_d} \underbrace{\left(K_d L_d \sum_{k=1}^{K_d} \alpha_{kl}^d \tau_{gk}^d \right)}_{\theta_{dl}^g} b_{l-1, L_d-1}(v) = \sum_{l=1}^{L_d} \theta_{dl}^g b_{l-1, L_d-1}(v) = \mathbb{E}[g(Y_d)|V = v]. \quad (31)$$

As can be seen we recover the parametric form [Mogstad, Santos, and Torgovitsky \(2018\)](#) imposed on the MTR. [Mogstad, Santos, and Torgovitsky \(2018\)](#) approach imposes $\mathbb{E}[g(Y_d)|V = v] = \sum_{l=1}^{L_d} \theta_{dl}^g b_{l-1, L_d-1}(v)$ as a primitive, while in contrast we show that under a valid *IV* assumption —Assumption 2, imposing such a structure on the MTRs is equivalent to parametrize the “selection on unobservables” dependence — $C_{Y_d, V|P}(x_1, x_2; \alpha^d)$ — using a Bernstein Copula of order L_d .

Going back to the general context, and by integrating the DMTEs we obtain the following model restriction:

$$\mathbb{E}[g(Y)1\{D = d\}|P = p] = \int_{p1\{d=0\}}^{p+(1-p)1\{d=0\}} \mathbb{E}[g(Y_d)|V = v, P = p] dv \\ = K_d L_d \sum_{k=1}^{K_d} \sum_{l=1}^{L_d} \alpha_{kl}^d \int_{p1\{d=0\}}^{p+(1-p)1\{d=0\}} b_{l-1, L_d-1}(v) dv \sum_{r=1}^{R_d} \sum_{s=1}^{S_d} \beta_{rs}^d b_{s-1, S_d-1}(F_P) \left\{ \sum_{j=k-1}^{K_d-1} (-1)^{j-k+1} \binom{K_d-1}{j} \binom{j}{k-1} (R_d S_d)^{j+1} \times \right. \\ \left. \sum_{n_{11}+n_{12}+\dots+n_{R_d S_d}=j} \binom{j}{n_{11}, n_{12}, \dots, n_{R_d S_d}} \prod_{e=1, f=1}^{R_d, S_d} (\beta_{ef}^d)^{n_{ef}} \prod_{f=1}^{S_d} (b_{f-1, S_d-1}(F_P))^{n_{\cdot f}} \gamma_{r-1, g}^{d, n_{\cdot e}} \right\}$$

Remark, we can show that $B_{l-1,L_d-1}(1) \equiv \int_0^1 b_{l-1,L_d-1}(v)dv = 1/L_d = \underbrace{\int_0^p b_{l-1,L_d-1}(v)dv + \int_p^1 b_{l-1,L_d-1}(v)dv}_{B_{l-1,L_d-1}(p)}$.

Therefore, we have:

$$\mathbb{E}[g(Y)D|P = p] = K_1 L_1 \sum_{k=1}^{K_1} \sum_{l=1}^{L_1} \alpha_{kl}^1 B_{l-1,L_1-1}(p) \sum_{r=1}^{R_1} \sum_{s=1}^{S_1} \beta_{rs}^1 b_{s-1,S_1-1}(Fp) \left\{ \sum_{j=k-1}^{K_1-1} (-1)^{j-k+1} \binom{K_1-1}{j} \binom{j}{k-1} (R_1 S_1)^{j+1} \times \right. \\ \left. \sum_{n_{11}+n_{12}+\dots+n_{R_1 S_1}=j} \binom{j}{n_{11}, n_{12}, \dots, n_{R_1 S_1}} \prod_{e=1, f=1}^{R_1, S_1} (\beta_{ef}^1)^{n_{ef}} \prod_{f=1}^{S_1} (b_{f-1, S_d-1}(Fp))^{n_{ef}} \gamma_{r-1, g}^{1, n_e} \right\} \quad (32)$$

and

$$\mathbb{E}[g(Y)(1-D)|P = p] = K_0 L_0 \sum_{k=1}^{K_0} \sum_{l=1}^{L_0} \alpha_{kl}^0 (1/L_0 - B_{l-1,L_0-1}(p)) \sum_{r=1}^{R_0} \sum_{s=1}^{S_0} \beta_{rs}^0 b_{s-1,S_0-1}(Fp) \left\{ \sum_{j=k-1}^{K_0-1} (-1)^{j-k+1} \binom{K_0-1}{j} \binom{j}{k-1} (R_0 S_0)^{j+1} \times \right. \\ \left. \sum_{n_{11}+n_{12}+\dots+n_{R_0 S_0}=j} \binom{j}{n_{11}, n_{12}, \dots, n_{R_0 S_0}} \prod_{e=1, f=1}^{R_0, S_0} (\beta_{ef}^0)^{n_{ef}} \prod_{f=1}^{S_0} (b_{f-1, S_0-1}(Fp))^{n_{ef}} \gamma_{r-1, g}^{0, n_e} \right\}. \quad (33)$$

Restrictions on $\gamma_{r-1, g}^{d, j_e}$. Recall $\gamma_{r-1, g}^{d, j_e} \equiv \mathbb{E}\left[g(Y_d) \prod_{e=1}^{R_d} (B_{e-1, R_d-1}(F_{Y_d}))^{j_e} b_{r-1, R_d-1}(F_{Y_d})\right]$ for $j_e \in \{1, \dots, L_d\}$. $\gamma_{r-1, g}^{d, j_e}$ is an unknown parameter to estimate, but the set of potential values it can take is restricted by the model.

In fact, each choice of $g(\cdot)$ imposes a restriction on $\gamma_{r-1, g}^{d, j_e}$, for instance for $g(\cdot) = 1\{\cdot \leq y\}$, $g(\cdot) = 1\{y < \cdot \leq y'\}$, $g^+ : \mathcal{Y} \rightarrow \mathbb{R}^+$, or $g^- : \mathcal{Y} \rightarrow \mathbb{R}^-$ we have respectively

$$\gamma_{r-1, y}^{d, j_e} = \mathbb{E}\left[1\{Y \leq y\} \prod_{e=1}^{R_d} (B_{e-1, R_d-1}(F_{Y_d}))^{j_e} b_{r-1, R_d-1}(F_{Y_d})\right] \geq 0, \quad (34)$$

$$\gamma_{r-1, y'}^{d, j_e} - \gamma_{r-1, y}^{d, j_e} = \mathbb{E}\left[1\{y < Y \leq y'\} \prod_{e=1}^{R_d} (B_{e-1, R_d-1}(F_{Y_d}))^{j_e} b_{r-1, R_d-1}(F_{Y_d})\right] \geq 0 \quad \forall y' > y, \quad (35)$$

$$\gamma_{r-1, g^+}^{d, j_e} = \mathbb{E}\left[g^+(Y_d) \prod_{e=1}^{R_d} (B_{e-1, R_d-1}(F_{Y_d}))^{j_e} b_{r-1, R_d-1}(F_{Y_d})\right] \geq 0, \quad (36)$$

$$\gamma_{r-1, g^-}^{d, j_e} = \mathbb{E}\left[g^-(Y_d) \prod_{e=1}^{R_d} (B_{e-1, R_d-1}(F_{Y_d}))^{j_e} b_{r-1, R_d-1}(F_{Y_d})\right] \leq 0. \quad (37)$$

Remark that, while the dimensionality of $\gamma_{r-1, g}^{d, j_e}$ depends on the complexity of the $g(\cdot)$, the set of unknown parameters $\theta = (\alpha^0, \alpha^1, \beta^1, \beta^0) \in \Theta^{BC}$ where

$$\Theta^{BC} \equiv \left\{ \alpha_{kl}^d \geq 0, \beta_{rs}^d \geq 0, 1 \leq l \leq L_d, 1 \leq k \leq K_d, 1 \leq r \leq R_d, 1 \leq s \leq S_d, \text{ such that} \right.$$

$$\left. K_d \sum_{l=1}^{L_d} \alpha_{kl}^d = 1, L_d \sum_{k=1}^{K_d} \alpha_{kl}^d = 1, S_d \sum_{r=1}^{R_d} \beta_{rs}^d = 1, R_d \sum_{s=1}^{S_d} \beta_{rs}^d = 1, \text{ for } d \in \{0, 1\} \right\}$$

is invariant to the choice of $g(\cdot)$. So choosing a more informative class of $g(\cdot)$, will provide a tighter identified set of the copula parameters Θ^{BC} . To do so, we will consider the half-interval class $\mathcal{G} \equiv \{g(\cdot) = \mathbf{1}[\cdot \leq y], y \in \mathcal{Y}\}$ which allow us

to recover the distributional DMTR $F_{Y_d|V,P}(y|v,p)$. We then consider $(\gamma_{r-1,y}^{j_e})_{y \in \mathcal{Y}} \equiv (\gamma_{r-1,y'}^{0j}, \gamma_{r-1,y}^{1j})_{y \in \mathcal{Y}} \in \Gamma^{BC}$ where

$$\Gamma^{BC} \equiv \left\{ \gamma_{r-1,y}^{d,j_e} \geq 0 \quad \forall y \in \mathcal{Y} \text{ such that } \gamma_{r-1,y'}^{d,j_e} - \gamma_{r-1,y}^{d,j_e} \geq 0 \quad \forall \infty \geq y' > y \geq -\infty, \text{ for } d \in \{0,1\} \right\}.$$

Theorem 7. Under Assumptions 1, 5 and 9, the identified set of the Bernstein copulas parameters $\bar{\Theta}_I^{BC}$ of $(\theta, (\gamma_{r-1,y}^{j_e})_{y \in \mathcal{Y}}) \in \Theta^{BC} \times \Gamma^{BC}$ is characterized as follows:

$$\bar{\Theta}_I^{BC} = \left\{ \left(\theta, (\gamma_{r-1,y}^{j_e})_{y \in \mathcal{Y}} \right) \in \Theta^{BC} \times \Gamma^{BC} \text{ that satisfies Equations (32) and (33), for all } g(\cdot) \in \mathcal{G} \right\},$$

and for any integrable real function $g(\cdot)$, the identified set $\Theta_{I,g}$ for DMTR $_g$ is defined as follows:

$$\begin{aligned} \Theta_{I,g} = & \left\{ (\mathbb{E}[g(Y_1)|V=v, P=p; \bar{\theta}], \mathbb{E}[g(Y_0)|V=v, P=p; \bar{\theta}]) \text{ such that} \right. \\ \mathbb{E}[g(Y_d)|V=v, P=p; \bar{\theta}] = & K_d L_d \sum_{k=1}^{K_d} \sum_{l=1}^{L_d} \alpha_{kl}^d b_{l-1, L_d-1}(v) \sum_{r=1}^{R_d} \sum_{s=1}^{S_d} \beta_{rs}^d b_{s-1, S_d-1}(F_P) \left\{ \sum_{j=k-1}^{K_d-1} (-1)^{j-k+1} \binom{K_d-1}{j} \binom{j}{k-1} (R_d S_d)^{j+1} \times \right. \\ & \left. \sum_{n_{11}+n_{12}+\dots+n_{R_d S_d}=j} \binom{j}{n_{11}, n_{12}, \dots, n_{R_d S_d}} \prod_{e=1, f=1}^{R_d, S_d} (\beta_{ef}^d)^{n_{ef}} \prod_{f=1}^{S_d} (b_{f-1, S_d-1}(F_P))^{n_f} \gamma_{r-1,g}^{d, n_e} \right\} \\ & \left. \forall \bar{\theta} \equiv \left(\theta, (\gamma_{r-1,y}^{j_e})_{y \in \mathcal{Y}} \right) \in \bar{\Theta}_I^{BC} \right\}. \end{aligned}$$

Example 1. Now we consider an example with $K_d = L_d = R_d = S_d = 2$. In this cases

$$\begin{aligned} b_{0,1}(u) &= \binom{1}{0} u^0 (1-u)^{1-0} = 1-u; & b_{1,1}(u) &= \binom{1}{1} u^1 (1-u)^{1-1} = u. \\ B_{0,1}(u) &= u - \frac{u^2}{2}; & B_{1,1}(u) &= \frac{u^2}{2}. \end{aligned}$$

Therefore, we have

$$C_{Y_d, V|P}(x_1, x_2; \alpha^d) = 4 \{ \alpha_{11}^d B_{0,1}(x_1) B_{0,1}(x_2) + \alpha_{12}^d B_{0,1}(x_1) B_{1,1}(x_2) + \alpha_{21}^d B_{1,1}(x_1) B_{0,1}(x_2) + \alpha_{22}^d B_{1,1}(x_1) B_{1,1}(x_2) \},$$

where $\alpha_{kl}^d \geq 0$ and $\beta_{rs}^d \geq 0$ satisfy $2(\alpha_{11}^d + \alpha_{12}^d) = 1$, $2(\alpha_{21}^d + \alpha_{22}^d) = 1$, $2(\alpha_{11}^d + \alpha_{21}^d) = 1$, $2(\alpha_{12}^d + \alpha_{22}^d) = 1$. In this case, we can express other α s in terms of α_{11}^d , that is, $\alpha_{22}^d = \alpha_{11}^d$, $\alpha_{21}^d = \alpha_{12}^d = \frac{1}{2} - \alpha_{11}^d$. To ensures all the parameters are greater or equal to zero, we need to have $0 \leq \alpha_{11}^d \leq \frac{1}{2}$. Therefore, we can write

$$\begin{aligned} C_{Y_d, V|P}(x_1, x_2; \alpha^d) &= 4 \left\{ \alpha_{11}^d \left(x_1 - \frac{x_1^2}{2} \right) \left(x_2 - \frac{x_2^2}{2} \right) + \left(\frac{1}{2} - \alpha_{11}^d \right) \left(x_1 - \frac{x_1^2}{2} \right) \frac{x_2^2}{2} + \left(\frac{1}{2} - \alpha_{11}^d \right) \left(x_2 - \frac{x_2^2}{2} \right) \frac{x_1^2}{2} + \alpha_{11}^d \frac{x_1^2}{2} \frac{x_2^2}{2} \right\} \\ &= 4\alpha_{11}^d x_1 x_2 + (1 - 4\alpha_{11}^d)(x_1 x_2^2 + x_2 x_1^2) + (4\alpha_{11}^d - 1)x_1^2 x_2^2. \end{aligned}$$

$$c_{Y_d, V|P}(x_1, x_2; \alpha^d) = 4\alpha_{11}^d + (1 - 4\alpha_{11}^d)(2x_1 + 2x_2) + 4(4\alpha_{11}^d - 1)x_1 x_2.$$

Note that if we define $4\alpha_{11}^d - 1 \equiv \delta$, then

$$C_{Y_d, V|P}(x_1, x_2; \alpha^d) = x_1 x_2 (1 + \delta(1 - x_1)(1 - x_2))$$

which is the FGM copula with dependence parameter δ . If we impose $\alpha_{11} = \frac{1}{4}$, then we are imposing the selection-on-observable assumption.

Likewise, when $R_d = S_d = 2$, we have

$$C_{Y_d, P}(x_1, x_2; \beta^d) = 4\beta_{11}^d x_1 x_2 + (1 - 4\beta_{11}^d)(x_1 x_2^2 + x_2 x_1^2) + (4\beta_{11}^d - 1)x_1^2 x_2^2.$$

and

$$c_{Y_d, P}(x_1, x_2; \beta^d) = 4\beta_{11}^d + (1 - 4\beta_{11}^d)(2x_1 + 2x_2) + 4(4\beta_{11}^d - 1)x_1 x_2.$$

We need $0 \leq \beta_{11}^d \leq \frac{1}{2}$. When we impose $\beta_{11}^d = \frac{1}{4}$, it follows that $C_{Y_d, P}(x_1, x_2; \beta^d) = x_1 x_2$, that is, the IV independence assumption is satisfied.

Recall that $f_{Y_d|V, P}(y|v, p) = c_{Y_d, V|P}(F_{Y_d|P}(y|p), F_{V|P}(v|p); \alpha^d) f_{Y_d|P}(y|p) f_{V|P}(v|p) = c_{Y_d, V|P}(F_{Y_d|P}(y|p), F_{V|P}(v|p); \alpha^d) f_{Y_d|P}(y|p)$, we have

$$\begin{aligned} \mathbb{E}[g(Y_d)|V = v, P = p] &= \int_{\mathcal{Y}} g(y) f_{Y_d|V, P}(y|v, p) dy \\ &= \int_{\mathcal{Y}} g(y) \left\{ 4\alpha_{11}^d + (1 - 4\alpha_{11}^d)(2F_{Y_d|P}(y|p) + 2v) + 4(4\alpha_{11}^d - 1)F_{Y_d|P}(y|p)v \right\} f_{Y_d|P}(y|p) dy \\ &= \underbrace{\int_{\mathcal{Y}} g(y) f_{Y_d|P}(y|p) \left(4\alpha_{11}^d + (1 - 4\alpha_{11}^d)F_{Y_d|P}(y|p) \right) dy}_{RHS1^d(p)} + v \underbrace{\int_{\mathcal{Y}} g(y) f_{Y_d|P}(y|p) \left(4\alpha_{11}^d - 1 \right) \left(4F_{Y_d|P}(y|p) - 2 \right) dy}_{RHS2^d(p)} \quad (38) \end{aligned}$$

So the RHS of the Equation (38) is linear in v . To proceed, recall that

$$F_{Y_d|P}(y|p) = \frac{C_{Y_d, P}(x_1, x_2)}{\partial x_2} \Big|_{x_1 = F_{Y_d}(y), x_2 = F_P(p)} = 4\beta_{11}^d F_{Y_d}(y) + (1 - 4\beta_{11}^d)(2F_{Y_d} F_P + F_{Y_d}^2) + 2(4\beta_{11}^d - 1)F_{Y_d}^2 F_P$$

and

$$f_{Y_d|P}(y|p) = f_{Y_d} \left(4\beta_{11}^d + (1 - 4\beta_{11}^d)(2F_{Y_d} + 2F_P) + 4(4\beta_{11}^d - 1)F_{Y_d} F_P \right)$$

Insert the above two equations into the first RHS term of Equation (38) and re-arrange, we have

$$RHS1^d(p) = \psi_{10}^d \underbrace{\mathbb{E}[g(Y_d)]}_{\gamma_0^d(g)} + \psi_{11}^d \underbrace{\mathbb{E}[g(Y_d)F_{Y_d}(Y_d)]}_{\gamma_1^d(g)} + \psi_{12}^d \underbrace{\mathbb{E}[g(Y_d)F_{Y_d}^2(Y_d)]}_{\gamma_2^d(g)} + \psi_{13}^d \underbrace{\mathbb{E}[g(Y_d)F_{Y_d}^3(Y_d)]}_{\gamma_3^d(g)} \equiv \boldsymbol{\psi}_1^d \cdot \boldsymbol{\gamma}^d(g),$$

where $\boldsymbol{\psi}_1^d(p) = (\psi_{10}^d(p), \psi_{11}^d(p), \psi_{12}^d(p), \psi_{13}^d(p))'$,

$$\psi_{10}^d(p) = \left(4\beta_{11}^d + 2(1 - 4\beta_{11}^d)F_P(p) \right) 4\alpha_{11}^d$$

$$\psi_{11}^d(p) = (4\beta_{11}^d - 1)(4F_P(p) - 2)4\alpha_{11}^d + (4\beta_{11}^d + 2F_P(p)(1 - 4\beta_{11}^d))^2(1 - 4\alpha_{11}^d)$$

$$\psi_{12}^d(p) = (12\beta_{11}^d + 6(1 - 4\beta_{11}^d)F_P(p))(1 - 4\alpha_{11}^d)(4\beta_{11}^d - 1)(2F_P(p) - 1)$$

$$\psi_{13}^d(p) = 2(4\beta_{11}^d - 1)^2(2F_P(p) - 1)^2(1 - 4\alpha_{11}^d)$$

The γ parameters are indexed by function g , and $\boldsymbol{\gamma}^d(g) = (\gamma_0^d(g), \gamma_1^d(g), \gamma_2^d(g), \gamma_3^d(g))'$.

For the RHS2 term, we have

$$RHS2^d(p) = \psi_{20}^d \gamma_0^d(g) + \psi_{21}^d \gamma_1^d(g) + \psi_{22}^d \gamma_2^d(g) + \psi_{23}^d \gamma_3^d(g) \equiv \boldsymbol{\psi}_2^d \cdot \boldsymbol{\gamma}^d(g)$$

where $\boldsymbol{\psi}_2^d(p) = (\psi_{20}^d(p), \psi_{21}^d(p), \psi_{22}^d(p), \psi_{23}^d(p))'$,

$$\psi_{20}^d(p) = -2(4\beta_{11}^d + 2(1 - 4\beta_{11}^d)F_P(p))$$

$$\psi_{21}^d(p) = 4(4\beta_{11}^d + 2F_P(p)(1 - 4\beta_{11}^d))^2 - 2(4\beta_{11}^d - 1)(4F_P(p) - 2)$$

$$\psi_{22}^d(p) = (48\beta_{11}^d + 24(1 - 4\beta_{11}^d)F_P(p))(4\beta_{11}^d - 1)(2F_P(p) - 1)$$

$$\psi_{23}^d(p) = 8(4\beta_{11}^d - 1)^2(2F_P(p) - 1)^2$$

Finally, consider

$$\begin{aligned} \mathbb{E}[g(Y)1\{D = d\}|P = p] &= \int_{p1\{d=0\}}^{p+(1-p)1\{d=0\}} \mathbb{E}[g(Y_d)|V = v, P = p]dv \\ &= RHS1^d(p) \int_{p1\{d=0\}}^{p+(1-p)1\{d=0\}} dv + RHS2^d(p) \int_{p1\{d=0\}}^{p+(1-p)1\{d=0\}} vdv \end{aligned}$$

Setting $d = 1$ we have

$$\mathbb{E}[g(Y)1\{D = 1\}|P = p] = (\boldsymbol{\psi}_1^1(p) \cdot \boldsymbol{\gamma}^1(g))p + (\boldsymbol{\psi}_2^1(p) \cdot \boldsymbol{\gamma}^1(g))\frac{p^2}{2} \quad (39)$$

$$\mathbb{E}[g(Y)1\{D = 0\}|P = p] = (\boldsymbol{\psi}_1^0(p) \cdot \boldsymbol{\gamma}^0(g))(1 - p) + (\boldsymbol{\psi}_2^0(p) \cdot \boldsymbol{\gamma}^0(g))\left(\frac{1}{2} - \frac{p^2}{2}\right) \quad (40)$$

In this case, the set

$$\Theta^{BC} \equiv \left\{ (\alpha_{11}^d, \beta_{11}^d, \alpha_{11}^0, \beta_{11}^0) : 0 \leq \alpha_{11}^d \leq \frac{1}{2}, 0 \leq \beta_{11}^d \leq \frac{1}{2}, \text{ for } d \in \{0, 1\} \right\}$$

If we consider the class of g functions to be $\mathcal{G} \equiv \{g(\cdot) = \mathbf{I}[\cdot \leq y], y \in \mathcal{Y}\}$, then the set

$$\Gamma^{BC} \equiv \left\{ (\boldsymbol{\gamma}^1(y), \boldsymbol{\gamma}^0(y)) : \gamma_j^d(y) \geq 0, \gamma_j^d(y) \geq \gamma_j^d(y'), \text{ for all } y \geq y', \text{ for } j \in \{0, 1, 2, 3\}, \text{ for } d \in \{0, 1\} \right\}.$$

The identified set of the Bernstein copulas parameters $\overline{\Theta}_I^{BC}$ of $(\alpha_{11}^d, \beta_{11}^d, \gamma_j^d(y))_{y \in \mathcal{Y}} \in \Theta^{BC} \times \Gamma^{BC}$ is characterized as follows:

$$\overline{\Theta}_I^{BC} = \left\{ (\alpha_{11}^d, \beta_{11}^d, \gamma_j^d(y))_{y \in \mathcal{Y}} \in \Theta^{BC} \times \Gamma^{BC} \text{ that satisfies Equations (39) and (40)} \right\},$$

and for any integrable real function $g(\cdot)$, the identified set $\Theta_{1,g}$ for DMTR $_g$ is defined as follows:

$$\Theta_{1,g} = \left\{ \left(\mathbb{E}[g(Y_1)|V = v, P = p; \bar{\theta}], \mathbb{E}[g(Y_0)|V = v, P = p; \bar{\theta}] \right) \text{ such that} \right. \\ \left. \mathbb{E}[g(Y_d)|V = v, P = p; \bar{\theta}] = \boldsymbol{\psi}_1^d(p) \cdot \boldsymbol{\gamma}^d(g) + (\boldsymbol{\psi}_2^d(p) \cdot \boldsymbol{\gamma}^d(g))v, \quad \forall \bar{\theta} \equiv \left(\alpha_{11}^1, \beta_{11}^1, \alpha_{11}^0, \beta_{11}^0, \boldsymbol{\gamma}^1(y), \boldsymbol{\gamma}^0(y) \right) \in \bar{\Theta}_I^{BC} \right\}.$$

Remark 8. It is possible to additionally impose some conditions listed in Lemma 4. In Example 1, further assuming MIV is amount to assume the second derivative of $C_{Y_d,P}$ with respect to x_2 is non-positive, that is

$$\frac{\partial^2 C_{Y_d,P}(x_1, x_2)}{\partial x_2^2} = 2(4\beta_{11}^d - 1)(x_1^2 - x_1) \leq 0, \forall x_1 \in [0, 1]$$

This is satisfied if we assume $\frac{1}{2} \geq \beta_{11}^d \geq \frac{1}{4}$.

B.3. Frank Copula: Proof of Corollary 2. Consider $d = 1$. Note first,

$$F_{Y,D|P}(y, 1|p) = -\frac{1}{\sigma_1(p)} \ln \left[1 + \frac{(e^{-\sigma_1(p)c_{1,F_P(p)}(F_{Y_1}(y))} - 1)(e^{-\sigma_1(p)p} - 1)}{(e^{-\sigma_1(p)} - 1)} \right]$$

Solve $c_{1,F_P(p)}(F_{Y_1}(y))$ from the above equation we have

$$c_{1,F_P(p)}(F_{Y_1}(y)) = -\frac{1}{\sigma_1(p)} \ln \left[1 + \frac{(e^{-\sigma_1(p)F_{Y,D|P}(y,1|p)} - 1)(e^{-\sigma_1(p)} - 1)}{(e^{-\sigma_1(p)p} - 1)} \right] \equiv H_1(y, p, \sigma_1(p))$$

Here, the $H_1(y, p, \sigma_1(p))$ only depends on quantities that are directly identifiable from the data and the finite dimensional parameters. Then recall that

$$H_1(y, p, \sigma_1(p)) = c_{1,F_P(p)}(F_{Y_1}(y)) = \frac{\partial C_{Y_1,P}(x_1, x_2)}{\partial x_2} \Big|_{x_1=F_{Y_1}(y), x_2=F_P(p)} \\ = \frac{(e^{-\alpha_1 F_{Y_1}(y)} - 1)e^{-\alpha_1 F_P(p)}}{(e^{-\alpha_1} - 1) + (e^{-\alpha_1 F_{Y_1}(y)} - 1)(e^{-\alpha_1 F_P(p)} - 1)}$$

Again, solving $F_{Y_1}(y)$ from it yields

$$F_{Y_1}(y; \theta) = -\frac{1}{\alpha_1} \ln \left[1 + \frac{H_1(y, p, \sigma_1(p))(e^{-\alpha_1} - 1)}{e^{-\alpha_1 F_P(p)} - H_1(y, p, \sigma_1(p))(e^{-\alpha_1 F_P(p)} - 1)} \right]$$

Next consider $d = 0$. We know that

$$F_{Y,D|P}(y, 0|p) = c_{0,F_P(p)}(F_{Y_0}(y)) + \frac{1}{\sigma_0(p)} \ln \left[1 + \frac{(e^{-\sigma_0(p)c_{0,F_P(p)}(F_{Y_0}(y))} - 1)(e^{-\sigma_0(p)p} - 1)}{(e^{-\sigma_0(p)} - 1)} \right]$$

Solving $c_{0,F_P(p)}(F_{Y_0}(y))$ from the above equation we have

$$c_{0,F_P(p)}(F_{Y_0}(y)) = \frac{1}{\sigma_0(p)} \ln \left[1 + \frac{(e^{\sigma_0(p)F_{Y,D|P}(y,0|p)} - 1)(e^{-\sigma_0(p)} - 1)}{e^{-\sigma_0(p)} - e^{-\sigma_0(p)p}} \right] \equiv H_0(y, p, \sigma_0(p))$$

Again, recall that

$$H_0(y, p, \sigma_0(p)) = c_{0, F_P(p)}(F_{Y_0}(y)) = \frac{\partial C_{Y_0, P}(x_1, x_2)}{\partial x_2} \Big|_{x_1=F_{Y_0}(y), x_2=F_P(p)} = \frac{(e^{-\alpha_0 F_{Y_0}(y)} - 1)e^{-\alpha_0 F_P(p)}}{(e^{-\alpha_0} - 1) + (e^{-\alpha_0 F_{Y_0}(y)} - 1)(e^{-\alpha_0 F_P(p)} - 1)}$$

Solve $F_{Y_0}(y)$ from the above equation yields

$$F_{Y_0}(y; \theta) = -\frac{1}{\alpha_0} \ln \left[1 + \frac{H_0(y, p, \sigma_1(p))(e^{-\alpha_0} - 1)}{e^{-\alpha_0 F_P(p)} - H_0(y, p, \sigma_0(p))(e^{-\alpha_1 F_P(p)} - 1)} \right].$$

Finally, since $F_d(y; \theta)$ does not depends on p , its partial derivative with respect to p must be flat at 0 for all value of p , and so does the right hand side of the equation. Therefore, for all p , noticing $e^{-\alpha_d F_P(p)} \neq 0$ and $H_d \neq 0$,

$$\frac{\partial \left\{ \frac{e^{-\alpha_d F_P(p)} - H_d(y, p, \sigma_d(p))(e^{-\alpha_d F_P(p)} - 1)}{H_d(y, p, \sigma_d(p))} \right\}}{\partial p} = 0 \Rightarrow e^{-\alpha_d F_P(p)} \frac{\alpha_d f_P(p)(H_d - 1)H_d - \frac{\partial H_d}{\partial p}}{H_d^2} = 0$$

$$\Rightarrow \alpha_d f_P(p)(1 - H_d)H_d + \frac{\partial H_d}{\partial p} = 0.$$

To obtain the identified set for the distributional DMTR, simply note that $H_d(y, p, \sigma_d(p)) = c_{d, F_P(p)}(F_{Y_d}(y)) = F_{Y_d|P}(y|p)$, hence

$$F_{Y_d|P, V}(y|p, v) = \frac{\partial C_{Y_d, V|P}(x_1, x_2)}{\partial x_2} \Big|_{x_1=H_d, x_2=v} = \frac{(e^{-\sigma_d H_d} - 1)e^{-\sigma_d v}}{(e^{-\sigma_d} - 1) + (e^{-\sigma_d H_d} - 1)(e^{-\sigma_d v} - 1)},$$

and restrict θ taking values from Θ_I^{SP} and

$$\begin{aligned} \mathbb{E}[Y_d|P = p, V = v] &= \int y f_{Y_d|P, V}(y|p, v) dy = \int y c_{Y_d, V|P}(y, v|p) f_{Y_d|P}(y|p) dy \\ &= \int y \frac{-\sigma_d(p)(e^{-\sigma_d(p)} - 1)e^{-\sigma_d(p)(H_d+v)}}{(e^{-\sigma_d} - 1) + (e^{-\sigma_d H_d} - 1)(e^{-\sigma_d v} - 1)} \frac{\partial H_d(y, p, \sigma_d(p))}{\partial y} dy \end{aligned}$$

APPENDIX C. PROOFS OF RESULTS IN THE MAIN TEXT.

C.1. Proof of Theorem 1. It is easy to see that

$$MTE_g(v) \equiv \mathbb{E}[g(Y_1) - g(Y_0)|V = v] = \int_0^1 f_P(p) DMTE_g(v, p) dp,$$

and

$$ATE_g \equiv \mathbb{E}[g(Y_1) - g(Y_0)] = \int_0^1 \int_0^1 \underbrace{f_P(p)}_{w^{ATE}(v, p)} DMTE_g(v, p) dp dv.$$

Regarding LATE,

$$LATE_g(u, u') \equiv \mathbb{E}[g(Y_1) - g(Y_0)|u < V \leq u'] = \int_0^1 \int_0^1 \frac{f_P(p) \mathbf{1}_{\{u < v \leq u'\}}}{\underbrace{u' - u}_{w^{LATE}(u, u')(v, p)}} DMTE_g(v, p) dv dp$$

For ATT, we have

$$\begin{aligned}
ATT_g &\equiv \int_0^1 \mathbb{E}[g(Y_1) - g(Y_0)|D = 1, P = p] dF_{P|D=1}(p) \\
&= \int_0^1 \mathbb{E}[g(Y_1) - g(Y_0)|V \leq p, P = p] dF_{P|D=1}(p) = \int_0^1 \frac{1}{p} \int_0^p \mathbb{E}[g(Y_1) - g(Y_0)|V = v, P = p] dv dF_{P|D=1}(p) \\
&= \int_0^1 \frac{1}{p} \int_0^p DMTE(v, p) dv \frac{p}{\mathbb{P}(D=1)} f_P(p) dp = \int_0^1 \int_0^1 DMTE(v, p) dv \frac{f_P(p) \mathbb{1}\{v \leq p\}}{\mathbb{P}(D=1)} dp \\
&= \int_0^1 \int_0^1 \underbrace{\frac{f_P(p) \mathbb{1}\{v \leq p\}}{\mathbb{E}[P]}}_{w^{ATT}(v, p)} DMTE_g(v, p) dv dp,
\end{aligned}$$

where $dF_{P|D=1}(p) = \frac{p}{\mathbb{P}(D=1)} f_P(p) dp$ by Bayesian rule and $\mathbb{P}(D = 1) = \mathbb{E}[\mathbb{E}[D|P]] = \mathbb{E}[P]$.

Following the similar derivation as ATT, we can show that

$$\begin{aligned}
ATUT_g &\equiv \int_0^1 \mathbb{E}[g(Y_1) - g(Y_0)|D = 0, P = p] dF_{P|D=0}(p) \\
&= \int_0^1 \int_0^1 \underbrace{\frac{f_P(p) \mathbb{1}\{v > p\}}{\mathbb{E}[1-P]}}_{w^{ATUT}(v, p)} DMTE_g(v, p) dv dp
\end{aligned}$$

Concerning the $PRTE_g$, under Assumption 1 only, we have:

$$\begin{aligned}
\mathbb{E}[g(Y^a)] &= \int_0^1 \mathbb{E}[g(Y^a)|P^a = p] dF_{P^a}(p) = \int_0^1 \mathbb{E}[(g(Y_1^a) - g(Y_0^a))D^a|P^a = p] dF_{P^a}(p) + \mathbb{E}[g(Y_0^a)] \\
&= \int_0^1 \int_0^1 \mathbb{1}\{v \leq p\} f_{P^a}(p) \mathbb{E}[g(Y_1^a) - g(Y_0^a)|V^a = v, P^a = p] dp dv + \mathbb{E}[g(Y_0^a)] \\
&= \int_0^1 \int_0^1 \mathbb{1}\{v \leq p\} f_{P^a}(p) DMTE_g^a(v, p) dp dv + \mathbb{E}[g(Y_0^a)]
\end{aligned}$$

Since we have $DMTE_g^{a'} = DMTE_g^a$ and $\mathbb{E}[g(Y_0^{a'})] = \mathbb{E}[g(Y_0^a)]$ under Assumption 4, then under both Assumptions 1 and 4 we have:

$$\mathbb{E}[g(Y^{a'}) - g(Y^a)] = \int_0^1 \int_0^1 [f_{P^{a'}}(p) - f_{P^a}(p)] \mathbb{1}\{v \leq p\} DMTE_g^a(v, p) dp dv$$

Therefore,

$$PRTE_g = \int_0^1 \int_0^1 \frac{[f_{P^{a'}}(p) - f_{P^a}(p)] \mathbb{1}\{v \leq p\}}{\underbrace{\mathbb{E}_{F_{P^{a'}}}[P] - \mathbb{E}_{F_{P^a}}[P]}_{w^{PRTE}(v, p)}} DMTE_g(v, p) dp dv.$$

C.2. Proof of Lemma 2. To show Equation (6), note that first that

$$f_{Y_d, P}(y, p) = \frac{\partial^2 F_{Y_d, P}(t_1, t_2)}{\partial t_1 \partial t_2} \Big|_{t_1=y, t_2=p} = \frac{\partial^2 C_{Y_d, P}(x_1, x_2)}{\partial x_1 \partial x_2} \Big|_{x_1=F_{Y_d}(y), x_2=F_P(p)} f_{Y_d}(y) f_P(p)$$

Therefore,

$$\begin{aligned}
F_{Y_d|P}(y|p) &= \int_{-\infty}^y f_{Y_d|P}(t|p) dt = \int_{-\infty}^y \frac{f_{Y_d,P}(t,p)}{f_P(p)} dt \\
&= \int_{-\infty}^y \frac{\frac{\partial^2 C_{Y_d,P}(x_1, x_2)}{\partial x_1 \partial x_2} \Big|_{x_1=F_{Y_d}(t), x_2=F_P(p)} f_{Y_d}(t) f_P(p)}{f_P(p)} dt = \frac{\partial C_{Y_d,P}(x_1, x_2)}{\partial x_2} \Big|_{x_1=F_{Y_d}(y), x_2=F_P(p)} \\
&\equiv c_{d, F_P(p)}(F_{Y_d}(y)),
\end{aligned}$$

where we write the RHS as $c_{d, F_P(p)}(F_{Y_d}(y))$ since the RHS depends on y only through $F_{Y_d}(y)$ and the mapping $c_{d, F_P(p)}$ depends on the joint distribution of Y_d and P . Note that for any given p , both $F_{Y_d|P}(y|p)$ and $F_{Y_d}(y)$ are strictly increasing in y . Therefore, the mapping $c_{d, F_P(p)}(\cdot)$ is strictly increasing and we can express

$$F_{Y_d}(y) = c_{d, F_P(p)}^{-1}(F_{Y_d|P}(y|p)). \quad (41)$$

To see Equation (7), note that

$$\begin{aligned}
f_{Y_d|V,P}(y|v,p) &= \frac{f_{Y_d,V|P}(y,v|p)}{f_{V|P}(v|p)} = \frac{\frac{\partial^2 C_{Y_d,V|P}(x_1, x_2; p)}{\partial x_1 \partial x_2} \Big|_{x_1=F_{Y_d}(y|p), x_2=F_{V|P}(v|p)} f_{Y_d|P}(y|p) f_{V|P}(v|p)}{f_{V|P}(v|p)} \\
&= \frac{\partial^2 C_{Y_d,V|P}(x_1, x_2; p)}{\partial x_1 \partial x_2} \Big|_{x_1=F_{Y_d}(y|p), x_2=F_{V|P}(v|p)} f_{Y_d|P}(y|p)
\end{aligned}$$

Therefore,

$$\begin{aligned}
F_{Y_d|V,P}(y|v,p) &= \int_{-\infty}^y f_{Y_d|V,P}(t|v,p) dt \\
&= \int_{-\infty}^y \frac{\partial^2 C_{Y_d,V|P}(x_1, x_2; p)}{\partial x_1 \partial x_2} \Big|_{x_1=F_{Y_d}(t|p), x_2=F_{V|P}(v|p)} f_{Y_d|P}(t|p) dt \\
&= \frac{\partial C_{Y_d,V|P}(x_1, x_2; p)}{\partial x_2} \Big|_{x_1=F_{Y_d}(y|p), x_2=v}
\end{aligned}$$

where again we use $F_{V|P}(v|p) = v$.

At last, we consider Equation (8). Suppose $d = 1$

$$\mathbb{P}[Y \leq y, D = 1|P = p] = \mathbb{P}[Y_1 \leq y, V \leq p|P = p] = C_{Y_1, V|P}(c_{1, F_P(p)}(F_{Y_1}(y)), p; p)$$

Inserting Equation (41) yield the result. If $d = 0$, then

$$\begin{aligned}
\mathbb{P}[Y \leq y, D = 0|P = p] &= \mathbb{P}[Y_0 \leq y, V > p|P = p] \\
&= \mathbb{P}[Y_0 \leq y|P = p] - \mathbb{P}[Y_0 \leq y, V \leq p|P = p] \\
&= c_{0, F_P(p)}(F_{Y_0}(y)) - C_{Y_0, V|P}(c_{0, F_P(p)}(F_{Y_0}(y)), p; p).
\end{aligned}$$

C.3. Proof of Theorem 2. Let $\mathbb{P}(Y \leq y, D = d|P = p)$ be the distribution of observables. It is apparent from Definition 1, $(C_{Y_d, V|P}, C_{Y_d, P}, F_d)$ satisfy Equations (8) and (9), then they can rationalize the data and model; on the other hand, if $(C_{Y_d, V|P}, C_{Y_d, P}, F_d)$ are the true model parameters, they they must connect with the implied data distribution through Equations (8) and (9). In this sense, the set defined in Definition 1 is sharp.

To verify the set defined in Theorem 2 is also sharp, it is sufficient to show that for that Equations (8) and (9) in Definition 1 and Equation (10) in Theorem 2 are equivalent. First, it is straightforward to see that Equations (8) and (9) imply Equation (10). Second, suppose Equation (10) hold, that is, $\Psi_{d,p}^{-1}(\mathbb{P}[Y \leq y, D = d|P = p])$ is flat in p and only varies as a function of y . Note that under Assumption 5, F_{Y_d} and F_P are continuous and strictly increasing, and both $C_{Y_d, V|P}$ and $\frac{\partial C_{Y_d, P}(x_1, x_2)}{\partial x_2}$ are increasing in their first arguments. Therefore, $\Psi_{d,p}^{-1}$ is strictly increasing in y by construction. Next from the definitions in Equations (8) and (9) we know that for any $p > 0$

$$p = C_{Y_1, V|P=p}(1, p; p) \quad 0 = C_{Y_1, V|P=p}(0, p; p),$$

and

$$1 = c_{d, F_P(p)}(1), \quad 0 = c_{d, F_P(p)}(0).$$

Therefore it is easy to see that $\Psi_{d,p}^{-1}(\mathbb{P}[Y \leq -\infty, D = d|P = p]) = \Psi_{d,p}^{-1}(0) = 0$ and $\Psi_{d,p}^{-1}(\mathbb{P}[Y \leq \infty, D = d|P = p]) = \Psi_{d,p}^{-1}(\mathbb{P}[D = d|P = p]) = \Psi_{d,p}^{-1}(p) = 1$. This says that $\Psi^{-1}(\mathbb{P}[Y \leq \cdot, D = d|P = p])$, as a function of y , is a valid distribution function, which we can choose as the counterfactual distribution F_{Y_d} . This completes the proof.

C.4. Proof to Corollary 1. We take $d = 1$ as an example; the case for $d = 0$ is similar. Suppose for simplicity the mapping Ψ^{-1} is differentiable with respect to p , and all inverse functions below are properly defined. Then the restriction is equivalent to

$$\frac{\partial \left\{ \Psi_{1,p}^{-1}(\mathbb{P}[Y \leq y, D = 1|P = p]) \right\}}{\partial p} = 0.$$

On the other hand, $P \perp Y_1|V$ and $P \perp V$ implies $(Y_1, V) \perp P$. By the definition in Equation (8), the mapping $\Psi_{1,p}$ then reduces to the following simple form:

$$\Psi_{1,p}(F_{Y_1}(y)) = C_{Y_1, V}(F_{Y_1}(y), p).$$

Let $C_{1, Y_1, V}^{-1}(\cdot, x_2)$ be the inverse of $C_{Y_1, V}(\cdot, x_2)$ with respect to the first argument, then,

$$\begin{aligned} 0 &= \frac{\partial \left\{ \Psi_{1,p}^{-1}(\mathbb{P}[Y \leq y, D = 1|P = p]) \right\}}{\partial p} = \frac{\partial \left\{ C_{1, Y_1, V}^{-1}(\mathbb{P}[Y \leq y, D = 1|P = p], p) \right\}}{\partial p} \\ &= \frac{\partial C_{1, Y_1, V}^{-1}(y, x_2)}{\partial y} \Big|_{y=\mathbb{P}[Y \leq y, D=1|P=p], x_2=p} \times \frac{\partial \mathbb{P}[Y \leq y, D = 1|P = p]}{\partial p} \\ &\quad + \frac{\partial C_{1, Y_1, V}^{-1}(y, x_2)}{\partial x_2} \Big|_{y=\mathbb{P}[Y \leq y, D=1|P=p], x_2=p}. \end{aligned}$$

This implies

$$\begin{aligned} \frac{\partial \mathbb{P}[Y \leq y, D = 1 | P = p]}{\partial p} &= - \frac{\frac{\partial C_{1,Y_1,V}^{-1}(y, x_2)}{\partial x_2} \Big|_{y=\mathbb{P}[Y \leq y, D=1 | P=p], x_2=p}}{\frac{\partial C_{1,Y_1,V}^{-1}(y, x_2)}{\partial y} \Big|_{y=\mathbb{P}[Y \leq y, D=1 | P=p], x_2=p}} \\ &= \frac{\partial C_{Y_1,V}(x_1, x_2)}{\partial x_2} \Big|_{x_1=F_1(y), x_2=p} = \mathbb{P}[Y_1 \leq y | V = p], \end{aligned}$$

where the first equality solves the previous displayed equation, the third equality is due to the definition of a copula and $V \sim U[0, 1]$. For the second equality, note first that $C_{1,Y_1,V}^{-1}(C_{Y_1,V}(x_1, x_2), x_2) = x_1$ and differentiate both sides with respect to x_2 yields

$$\frac{\partial C_{1,Y_1,V}^{-1}(C_{Y_1,V}(x_1, x_2), x_2)}{\partial x_2} + \frac{\partial C_{1,Y_1,V}^{-1}(C_{Y_1,V}(x_1, x_2), x_2)}{\partial y} \frac{\partial C_{Y_1,V}(x_1, x_2)}{\partial x_2} = 0,$$

then the second equality holds by noticing $C_{Y_1,V}(F_1(y), p) = \mathbb{P}[Y \leq y, D = 1 | P = p]$ under the independence Assumption 2.

C.5. Proof of Lemma 3. (i) and (ii) hold obviously, in which cases $\mathbb{P}(Y_1 \leq y | P_1 = p_1) = \mathbb{P}(Y_1 \leq y | P_1 = p'_1)$ for all y, p_1 and p'_1 . To see (iii) holds, note first $Y_{d0} | V_2 = v_2 \succeq_{FSD} Y_{d1} | V_2 = v_2$ implies $H_y(\cdot)$ is an increasing function. Since $P_2 | P_1 = p'_1 \succeq_{FSD} P_2 | P_1 = p_1$, it follows that $\mathbb{E}_{P_2 | P_1 = p_1}[H_y(P_2)] - \mathbb{E}_{P_2 | P_1 = p'_1}[H_y(P_2)] \leq 0$. Therefore, $\mathbb{P}(Y_1 \leq y | P_1 = p_1) - \mathbb{P}(Y_1 \leq y | P_1 = p'_1) \leq 0$ for any $p'_1 \geq p_1$, that is, $Y_d | P_1 = p_1 \succeq_{FSD} Y_d | P_1 = p'_1$ for any $p'_1 \geq p_1$. \square

C.6. Proof of Lemma 4. For Property (c) see [Nelsen \(2007, Theorem 5.2.10\)](#). For Properties (d) and (e) see [Nelsen \(2007, Corollary 5.2.6\)](#).

Properties (a) and (f) are obvious. For property (b), note first that $F_{Y_d|P}(y|p) = \frac{\partial C_{Y_d,P}(F_{Y_d}(y), F_P(p))}{\partial x_2}$ and $f_{Y_d|P}(y|p) = \frac{\partial^2 C_{Y_d,P}(F_{Y_d}(y), F_P(p))}{\partial x_2 \partial x_2} f_{Y_d}(y)$. Therefore,

$$\frac{f_{Y_d|P}(y|p)}{F_{Y_d|P}(y|p)} = \frac{\frac{\partial^2 C_{Y_d,P}(F_{Y_d}(y), F_P(p))}{\partial x_2 \partial x_2} f_{Y_d}(y)}{\frac{\partial C_{Y_d,P}(F_{Y_d}(y), F_P(p))}{\partial x_2}} = \frac{\partial \log \frac{\partial C_{Y_d,P}(F_{Y_d}(y), F_P(p))}{\partial x_2}}{\partial x_1} f_{Y_d}(y).$$

If P is an IHRD IV, then $\frac{f_{Y_d|P}(y|p)}{F_{Y_d|P}(y|p)}$ is non-decreasing in p for all y . Since $f_{Y_d}(y) > 0$ and $F_P(p)$ is increasing in p , it is equivalent to say $\frac{\partial \log \frac{\partial C_{Y_d,P}(x_1, x_2)}{\partial x_2}}{\partial x_1}$ is non-decreasing in x_2 .

C.7. Proof of Theorem 5. under Assumption 10 we have: $f_{\mathbf{Q}, \mathbf{V}}(\mathbf{q}, \mathbf{v}) = f_{\mathbf{Q}}(\mathbf{q}) f_{\mathbf{V}}(\mathbf{v})$, this latter equality will be directly used in all the derivations below. For (i) to (iii) we have:

$$MTE_g(\mathbf{v}) \equiv \mathbb{E}[g(Y_1) - g(Y_0) | \mathbf{V} = \mathbf{v}] = \int_{\mathbf{q}} f_{\mathbf{Q}}(\mathbf{q}) DMTE_g(\mathbf{v}, \mathbf{q}) d\mathbf{q}$$

$$ATE_g \equiv \mathbb{E}[g(Y_1) - g(Y_0)] = \int_{\mathbf{v}} \int_{\mathbf{q}} \underbrace{f_{\mathbf{Q}}(\mathbf{q})f_{\mathbf{V}}(\mathbf{v})}_{w^{ATE}(\mathbf{v}, \mathbf{q})} DMTE_g(\mathbf{v}, \mathbf{q}) d\mathbf{q} d\mathbf{v}$$

$$\begin{aligned} LATE_g(\mathbf{u}, \mathbf{u}') &\equiv \mathbb{E}[g(Y_1) - g(Y_0) | u_1 < V_1 \leq u'_1, u_2 < V_2 \leq u'_2] \\ &= \int_{\mathbf{v}} \int_{\mathbf{q}} \underbrace{\frac{f_{\mathbf{Q}}(\mathbf{q})f_{\mathbf{V}}(\mathbf{v}) \mathbf{1}_{\{\mathbf{v} \in [u_1, u'_1] \times [u_2, u'_2]\}}}{F_{\mathbf{V}}(\mathbf{v} \in [u_1, u'_1] \times [u_2, u'_2])}}_{w^{LATE}(\mathbf{u}, \mathbf{u}')(\mathbf{v}, \mathbf{q})} DMTE_g(\mathbf{v}, \mathbf{q}) d\mathbf{q} d\mathbf{v} \end{aligned}$$

For vector \mathbf{a} and \mathbf{b} , let $\mathbf{a} \leq \mathbf{b}$ denote ‘‘component-wise smaller or equal to’’. Then,

$$\begin{aligned} ATT_g &\equiv \int_{\mathbf{q}} \mathbb{E}[g(Y_1) - g(Y_0) | D = 1, \mathbf{Q} = \mathbf{q}] dF_{\mathbf{Q}|D=1}(\mathbf{q}) \\ &= \int_{\mathbf{q}} \mathbb{E}[g(Y_1) - g(Y_0) | \mathbf{V} \leq \mathbf{q}, \mathbf{Q} = \mathbf{q}] dF_{\mathbf{Q}|D=1}(\mathbf{q}) \\ &= \int_{\mathbf{q}} \int_{\mathbf{v} \leq \mathbf{q}} \frac{1}{F_{\mathbf{V}}(\mathbf{q})} \mathbb{E}[g(Y_1) - g(Y_0) | \mathbf{V} = \mathbf{v}, \mathbf{Q} = \mathbf{q}] dF_{\mathbf{Q}|D=1}(\mathbf{q}) \\ &= \int_{\mathbf{v}} \int_{\mathbf{q}} \underbrace{\frac{f_{\mathbf{Q}}(\mathbf{q})f_{\mathbf{V}}(\mathbf{v}) \mathbf{1}_{\{\mathbf{v} \in [0, q_1] \times [0, q_2]\}}}{\mathbb{E}[F_{\mathbf{V}}(\mathbf{Q})]}}_{w^{ATT}(\mathbf{v}, \mathbf{q})} DMTE_g(\mathbf{v}, \mathbf{q}) d\mathbf{q} d\mathbf{v} \end{aligned}$$

where $dF_{\mathbf{Q}|D=1}(\mathbf{q}) = \frac{f_{\mathbf{Q}}(\mathbf{q})F_{\mathbf{V}}(\mathbf{q})}{\mathbb{P}(D=1)}$ by Bayesian rule and $\mathbb{P}(D = 1) = \mathbb{E}[\mathbb{E}[D|\mathbf{Q}]] = \mathbb{E}[F_{\mathbf{V}}(\mathbf{Q})]$. Likewise, we can derive the ATUT weights as follows:

$$\begin{aligned} ATUT_g &\equiv \int_{\mathbf{q}} \mathbb{E}[g(Y_1) - g(Y_0) | D = 0, \mathbf{Q} = \mathbf{q}] dF_{\mathbf{Q}|D=0}(\mathbf{q}) \\ &= \int_{\mathbf{v}} \int_{\mathbf{q}} \underbrace{\frac{f_{\mathbf{Q}}(\mathbf{q})f_{\mathbf{V}}(\mathbf{v}) \mathbf{1}_{\{\mathbf{v} \notin [0, q_1] \times [0, q_2]\}}}{\mathbb{E}[1 - F_{\mathbf{V}}(\mathbf{Q})]}}_{w^{ATUT}(\mathbf{v}, \mathbf{q})} DMTE_g(\mathbf{v}, \mathbf{q}) d\mathbf{q} d\mathbf{v} \end{aligned}$$

Concerning the $PRTE_g$, under Assumption 10 only, we have:

$$\begin{aligned} \mathbb{E}[g(Y^a)] &= \int_{\mathbf{q}} \mathbb{E}[g(Y^a) | \mathbf{Q}^a = \mathbf{q}] dF_{\mathbf{Q}^a}(\mathbf{q}) = \int_{\mathbf{q}} \mathbb{E}[(g(Y_1^a) - g(Y_0^a))D^a | \mathbf{Q}^a = \mathbf{q}] dF_{\mathbf{Q}^a}(\mathbf{q}) + \mathbb{E}[g(Y_0^a)] \\ &= \int_{\mathbf{v}} \int_{\mathbf{q}} \mathbf{1}_{\{\mathbf{v} \in [0, q_1] \times [0, q_2]\}} f_{\mathbf{Q}^a}(\mathbf{q}) f_{\mathbf{V}^a}(\mathbf{v}) \mathbb{E}[g(Y_1^a) - g(Y_0^a) | \mathbf{V}^a = \mathbf{v}, \mathbf{Q}^a = \mathbf{q}] d\mathbf{p} d\mathbf{v} + \mathbb{E}[g(Y_0^a)] \\ &= \int_{\mathbf{v}} \int_{\mathbf{q}} \mathbf{1}_{\{\mathbf{v} \in [0, q_1] \times [0, q_2]\}} f_{\mathbf{Q}^a}(\mathbf{q}) f_{\mathbf{V}^a}(\mathbf{v}) DMTE_g^a d\mathbf{q} d\mathbf{v} + \mathbb{E}[g(Y_0^a)] \end{aligned}$$

Since we have $DMTE_g^a = DMTE_g^a$, $f_{\mathbf{V}^a} = f_{\mathbf{V}^a}$, and $\mathbb{E}[g(Y_0^a)] = \mathbb{E}[g(Y_0^a)]$ under Assumption 11, then under both Assumptions 10 and 11 we have:

$$\mathbb{E}[g(Y^{a'}) - g(Y^a)] = \int_{\mathbf{v}} \int_{\mathbf{q}} [f_{\mathbf{Q}^{a'}}(\mathbf{q}) - f_{\mathbf{Q}^a}(\mathbf{q})] f_{\mathbf{V}^a}(\mathbf{v}) \mathbf{1}_{\{\mathbf{v} \in [0, q_1] \times [0, q_2]\}} DMTE_g^a d\mathbf{q} d\mathbf{v}$$

Therefore,

$$P RTE_g = \int_{\mathbf{v}} \int_{\mathbf{q}} \underbrace{\frac{[f_{\mathbf{Q}^{a'}}(\mathbf{q}) - f_{\mathbf{Q}^a}(\mathbf{q})] f_{\mathbf{V}^a}(\mathbf{v}) \mathbf{1}_{\{\mathbf{v} \in [0, q_1] \times [0, q_2]\}}}{\mathbb{E}[F_{\mathbf{V}}(\mathbf{Q}^{a'})] - \mathbb{E}[F_{\mathbf{V}}(\mathbf{Q}^a)]}}_{w^{P RTE}(\mathbf{v}, \mathbf{q})} DMTE_g^a d\mathbf{q} d\mathbf{v}.$$

For (iv): $Q_1(Z_1)$, $Q_2(Z_2)$ and $F_{\mathbf{V}}(\mathbf{v})$ are shown to be identified in Lee and Salanié (2018, Theorem 4.2). The remaining point that we need to show is that the joint distribution $F_{\mathbf{Q}}(\mathbf{q})$ is also point identified. Indeed, we have for $\mathbf{Z} = (Z_1, Z_2)$,

$$\begin{aligned} F_{\mathbf{Q}}(\mathbf{q}) &= \mathbb{P}(Q_1(Z_1) \leq q_1, Q_2(Z_2) \leq q_2) \\ &= \mathbb{P}(Z_1 \leq Q_1^{-1}(q_1), Z_2 \leq Q_2^{-1}(q_2)) = F_{\mathbf{Z}}(Q_1^{-1}(q_1), Q_2^{-1}(q_2)) \end{aligned}$$

where the joint distribution $F_{\mathbf{Z}}(\cdot, \cdot)$ is directly observed from the data. The invertibility of Q_1 and Q_2 is ensured by Lee and Salanié (2018, Assumption 4.2).

C.8. Proof of Lemma 5. Equations (18) to (21) are direct applications of Joe (1996, Property 2), with some simplifications due to the fact that Assumption 10 imposes that $\mathbf{V} \perp \mathbf{Q}$.

Next we prove Equation (22). Given $d = 1$, $y \in \mathcal{Y}$, and $\mathbf{q} \in \mathcal{Q}_1 \times \mathcal{Q}_2$,

$$\begin{aligned} \mathbb{P}[Y \leq y, D = 1 | \mathbf{Q} = \mathbf{q}] &= \mathbb{P}[Y_1 \leq y, \mathbf{V} \leq \mathbf{q} | \mathbf{Q} = \mathbf{q}] = \mathbb{P}[Y_1 \leq y, V_1 \leq q_1 | V_2 \leq q_2, \mathbf{Q} = \mathbf{q}] q_2 \\ &= \int_0^{q_2} \mathbb{P}[Y_1 \leq y, V_1 \leq q_1 | V_2 = v_2, \mathbf{Q} = \mathbf{q}] dF_{V_2 | \mathbf{Q}}(v_2 | \mathbf{q}) = \int_0^{q_2} \mathbb{P}[Y_1 \leq y, V_1 \leq q_1 | V_2 = v_2, \mathbf{Q} = \mathbf{q}] dv_2 \quad (42) \end{aligned}$$

For a given v_2 , the integrand $\mathbb{P}[Y_1 \leq y, V_1 \leq q_1 | V_2 = v_2, \mathbf{Q} = \mathbf{q}]$ can be handled in a similar way as in the STC case:

$$\begin{aligned} \mathbb{P}[Y_1 \leq y, V_1 \leq q_1 | V_2 = v_2, \mathbf{Q} = \mathbf{q}] &= C_{Y_1, V_1 | V_2, \mathbf{Q}}(F_{Y_1 | V_2, \mathbf{Q}}(y | v_2, \mathbf{q}), F_{V_1 | V_2, \mathbf{Q}}(q_1 | v_2, \mathbf{q}); v_2, \mathbf{q}) \\ &= C_{Y_1, V_1 | V_2, \mathbf{Q}}(c_{1, v_2}^{III}(F_{Y_1 | \mathbf{Q}}(y | \mathbf{q})), F_{V_1 | V_2}(q_1 | v_2); v_2, \mathbf{q}) = C_{Y_1, V_1 | V_2, \mathbf{Q}}(c_{1, v_2}^{III} \circ c_{1, F_{Q_1 | Q_2}}^{II}(F_{Y_1 | Q_1}(y | q_1)), F_{V_1 | V_2}(q_1 | v_2); v_2, \mathbf{q}) \\ &= C_{Y_1, V_1 | V_2, \mathbf{Q}}(c_{1, v_2}^{III} \circ c_{1, F_{Q_1 | Q_2}}^{II} \circ c_{1, F_{Q_1}(q_1)}^I(F_{Y_1}(y)), F_{V_1 | V_2}(q_1 | v_2); v_2, \mathbf{q}) \equiv \tilde{\Psi}_{1, \mathbf{q}}(F_{Y_1}(y), v_2). \quad (43) \end{aligned}$$

where the equalities hold by Equations 18, 19, and 20 and the assumption that $\mathbf{Q} \perp \mathbf{V}$, and the “ \circ ” denotes composite functions. Since for every y , \mathbf{q} , and v_2 , the functions c^I , c^{II} and c^{III} are monotone, and $C_{Y_1, V_1 | V_2, \mathbf{Q}}$ is also monotone in its first argument, then it follows that for each given $y \in \mathcal{Y}$ and $\mathbf{q} \in \mathcal{Q}_1 \times \mathcal{Q}_2$,

$$\mathbb{P}[Y \leq y, D = 1 | \mathbf{Q} = \mathbf{q}] = \int_0^{q_2} \tilde{\Psi}_{1, \mathbf{q}}(F_{Y_1}(y), v_2) dv_2 \equiv \Psi_{1, \mathbf{q}}(F_{Y_1}(y))$$

is also a monotone function in $F_{Y_1}(y)$.

Remark 9. If there was no V_2 and Q_2 , that is, if the model is single threshold crossing model, then we would not have the additional integration in Equation (42), and we do not need to use the two layers of vine-copula operation c^{III} and c^I for V_2 and Q_2 , respectively. In this case, the expression for $\Psi_{1,\mathbf{q}}$ exactly reduces to the expression for $\Psi_{1,p}$ in Lemma 2.

For the case of $d = 0$, note that

$$\begin{aligned} \mathbb{P}[Y \leq y, D = 0 | \mathbf{Q} = \mathbf{q}] &= \mathbb{P}[Y_0 \leq y | \mathbf{Q} = \mathbf{q}] - \mathbb{P}[Y_0 \leq y, D = 1 | \mathbf{Q} = \mathbf{q}] \\ &= c_{0,F_{Q_1|Q_2}}^{II} \circ c_{0,F_{Q_1}(q_1)}^I(F_{Y_0}(y)) - \int_0^{q_2} \tilde{\Psi}_{0,\mathbf{q}}(F_{Y_0}(y), v_2) dv_2 \equiv \Psi_{0,\mathbf{q}}(F_{Y_0}(y)), \end{aligned}$$

where $\tilde{\Psi}_{0,\mathbf{q}}$ is defined in the same way as $\tilde{\Psi}_{1,\mathbf{q}}$ with index “1” being replaced by “0”.

C.9. Proof of Theorem 6. Theorem 6-(i) is a direct application of HV05 testable implications where $G(Y) = 1\{Y \in [y, y']\}$ for $y \leq y'$. We show (ii) is true. We will assume that $\frac{\partial \mathbb{P}(y < Y \leq y', D=1 | P=p)}{\partial p}$ is continuous over the set of limit points of \mathcal{P} . First, we construct \tilde{V} and \tilde{D} as follows:

$$\mathbb{P}(\tilde{V} \leq t | P = p) = t \quad \forall t \in [0, 1] \text{ and } \forall p \in \mathcal{P}. \quad (44)$$

$$\tilde{D} = 1\{P(Z) \geq \tilde{V}\}. \quad (45)$$

Note that by construction, Assumption 1 is satisfied. Let $L(\mathcal{P})$ be the set of limit points of \mathcal{P} , $L^o(\mathcal{P})$ be a set of interior point of \mathcal{P} , and $C(\mathcal{P})$ be the closure of \mathcal{P} . Furthermore, let $I(\mathcal{P}) = C(\mathcal{P})/L^o(\mathcal{P})$ be the complement of $L^o(\mathcal{P})$ in the closure of \mathcal{P} . So $I(\mathcal{P})$ also contains isolation points. Note that $L^o(\mathcal{P})$ can be written as a union of countable or finite exclusive open intervals: $L^o(\mathcal{P}) = \cup_{j=1}^J (a_j, b_j)$, where $(a_j, b_j) \subseteq \mathcal{P}$, $b_j < a_{j+1}$, and J can be infinity. Let $\Omega(\mathcal{P})$ be a collection of intervals belonging to $(0, 1]$ defined as follows:

$$\Omega(\mathcal{P}) \equiv \{(p, p') : p, p' \in I(\mathcal{P}) \cup \{0, 1\} \text{ and } \nexists \tilde{p} \in \mathcal{P}, \text{ such that } p < \tilde{p} < p'\}.$$

Then we again knows that Ω contains a generic element $(c_k, d_k]$, where $c_k, d_k \in I(\mathcal{P})$, $d_k \leq c_{k+1}$, $k = 1, 2, \dots, K$ with K possibly taking ∞ . Note that with above notation, for any $v \in (0, 1]$, v must belongs to one of the following categories: (i) an element of $L^o(\mathcal{P})$ so that $v \in (a_j, b_j)$ for some j , (ii) $v \in L(\mathcal{P})/L^o(\mathcal{P})$, and (iii) there exist an integer k such that $v \in (c_k, d_k]$. The following figure illustrates the partition of the unit interval.



FIGURE 1. An illustration: $\mathcal{P} = \{p_1, p_2, p_5\} \cup [p_3, p_4] \cup [p_6, p_7]$, $L^o(\mathcal{P}) = (p_3, p_4) \cup (p_6, p_7)$, and $\Omega(\mathcal{P}) = \{(0, p_1], (p_1, p_2], (p_4, p_5], (p_5, p_6], (p_7, 1]\}$.

Next, we propose the following distribution for $\tilde{Y}_1|\tilde{V}, P$. For any arbitrary $p \in \mathcal{P}$ and $v \in (0, 1]$, we define

$$\mathbb{P}(\tilde{Y}_1 \leq y|\tilde{V} = v, P = p) = \begin{cases} \frac{\partial}{\partial t} \mathbb{P}(Y \leq y, D = 1|P = t)|_{t=v} & \text{if } v \in L^o(\mathcal{P}) \\ \lim_{\tilde{v} \rightarrow v} \frac{\partial}{\partial t} \mathbb{P}(Y \leq y, D = 1|P = t)|_{t=\tilde{v}} & \text{if } v \in L(\mathcal{P})/L^o(\mathcal{P}) \\ \frac{\mathbb{P}(Y \leq y, D=1|P=d_k) - \mathbb{P}(Y \leq y, D=1|P=c_k)}{d_k - c_k} & \text{if } v \notin L(\mathcal{P}) \text{ but } v \in (c_k, d_k] \in \Omega(\mathcal{P}). \end{cases}$$

$$\mathbb{P}(\tilde{Y}_0 \leq y|\tilde{V} = v, P = p) = \begin{cases} -\frac{\partial}{\partial v} \mathbb{P}(Y \leq y, D = 0|P = t)|_{t=v} & \text{if } v \in L^o(\mathcal{P}) \\ -\lim_{\tilde{v} \rightarrow v} \frac{\partial}{\partial v} \mathbb{P}(Y \leq y, D = 0|P = t)|_{t=\tilde{v}} & \text{if } v \in L^o(\mathcal{P}) \\ \frac{\mathbb{P}(Y \leq y, D=0|P=c_k) - \mathbb{P}(Y \leq y, D=0|P=d_k)}{d_k - c_k} & \text{if } v \notin L^o(\mathcal{P}) \text{ but } v \in (c_k, d_k] \in \Omega(\mathcal{P}). \end{cases}$$

Note that the conditioning on $\tilde{V} = v$, the distribution of \tilde{Y}_1 does not depend on P . Hence, Assumption 2 is satisfied by construction.

We now show that the distribution function constructed above is well defined. We focus on $\mathbb{P}(\tilde{Y}_1 \leq y|\tilde{V} = v, P = p)$ and the verification for $\mathbb{P}(\tilde{Y}_0 \leq y|\tilde{V} = v, P = p)$ is analogous.

- (1) $\mathbb{P}(\tilde{Y}_1 < \underline{y} - \epsilon|\tilde{V} = v, P = p) = 0$ for all $v \in [0, 1]$ and for any arbitrarily small $\epsilon > 0$. To see this, suppose $v \notin L(\mathcal{P})$, then there exists $[c_k, d_k] \in \Omega(\mathcal{P})$ such that $v \in (c_k, d_k]$, therefore,

$$\begin{aligned} \mathbb{P}(\tilde{Y}_1 \leq \underline{y} - \epsilon|\tilde{V} = v, P = p) \\ = \frac{\mathbb{P}(Y \leq \underline{y} - \epsilon, D = 1|P = d_k) - \mathbb{P}(Y \leq \underline{y} - \epsilon, D = 1|P = c_k)}{d_k - c_k} = \frac{0 - 0}{d_k - c_k} = 0. \end{aligned}$$

On the other hand, if $v \in L^o(\mathcal{P})$, then $\mathbb{P}(Y \leq \underline{y} - \epsilon, D = 1|P = \tilde{v}) = 0$ for all \tilde{v} in a small neighborhood of v , which implies $\frac{\partial}{\partial v} \mathbb{P}(Y \leq \underline{y} - \epsilon, D = 1|P = v) = 0$. The case that $v \in L^o(\mathcal{P})$ follows straightforwardly.

- (2) $\mathbb{P}(\tilde{Y}_1 \leq \bar{y}|\tilde{V} = v, P = p) = 1$. First, if $v \in L^o(\mathcal{P})$, then $\mathbb{P}(Y \leq \bar{y}, D = 1|P = v) = \mathbb{P}(D = 1|P = v) = v \Rightarrow \frac{\partial}{\partial v} \mathbb{P}(Y \leq \bar{y}, D = 1|P = v) = 1$. On the other hand, if $v \notin L(\mathcal{P})$, then

$$\begin{aligned} \mathbb{P}(\tilde{Y}_1 \leq \bar{y}|\tilde{V} = v, P = p) \\ = \frac{\mathbb{P}(Y \leq \bar{y}, D = 1|P = d_k) - \mathbb{P}(Y \leq \bar{y}, D = 1|P = c_k)}{p' - p} = \frac{d_k - c_k}{d_k - c_k} = 1. \end{aligned}$$

- (3) $\mathbb{P}(\tilde{Y}_1 \leq y|\tilde{V} = v, P = p)$ is nondecreasing in y . For $y < y'$ we have

$$\mathbb{P}(\tilde{Y}_1 \leq y' | \tilde{V} = v, P = p) - \mathbb{P}(\tilde{Y}_1 \leq y | \tilde{V} = v, P = p) = \begin{cases} \frac{\partial}{\partial t} \mathbb{P}(y < Y \leq y', D = 1 | P = t) |_{t=v} \geq 0 & \text{if } v \in L^o(\mathcal{P}), \\ \lim_{\tilde{v} \rightarrow v} \frac{\partial}{\partial t} \mathbb{P}(y < Y \leq y', D = 1 | P = t) |_{t=\tilde{v}} \geq 0 & \text{if } v \in L(\mathcal{P})/L^o(\mathcal{P}) \\ \frac{\mathbb{P}(y < Y \leq y', D=1 | P=d_k) - \mathbb{P}(y < Y \leq y', D=1 | P=c_k)}{d_k - c_k} \geq 0 & \text{if } v \notin L^o(\mathcal{P}) \text{ but } v \in [c_k, d_k] \in \Omega(\mathcal{P}), \end{cases}$$

where the last inequalities hold whenever the testable implications hold, i.e. $\mathbb{P}(y < Y \leq y', D = 1 | P = p)$ is a non-decreasing function for all $p \in \mathcal{P}$ and all $y < y'$, and by the continuously differentiability of $\mathbb{P}(y < Y \leq y', D = 1 | P = p)$ over $L(\mathcal{P})$.

Finally, we show that $(\tilde{V}, \tilde{Y}_d, P(Z))$, $d \in \{0, 1\}$ is observationally equivalent to $(V, Y_d, P(Z))$ $d \in \{0, 1\}$. For this, we show that the joint distribution of $(\tilde{Y}, \tilde{D}, P(Z))$ is the same as the joint distribution of $(Y, D, P(Z))$. Take an arbitrary $z \in \mathcal{Z}$ and let $p = p(z) \in \mathcal{P}$.

Suppose first $p \notin L^o(\mathcal{P})$, then $(0, p]$ can be expressed as unions of exclusive intervals $(\cup_{j=1}^{J^*} (a_j, b_j)) \cup (\cup_{k=1}^{K^*} (c_k, d_k])$ for some J^* and K^* , where (a_j, b_j) s are connected subsets of \mathcal{P} . Therefore,

$$\begin{aligned} \mathbb{P}(\tilde{Y} \leq y, \tilde{D} = 1 | P = p) &= \mathbb{P}(\tilde{Y}_1 \leq y, \tilde{V} \leq p | P = p) = \int_0^p \mathbb{P}(\tilde{Y}_1 \leq y | \tilde{V} = v, P = p) dv \\ &= \sum_{j=1}^{J^*} \int_{a_j}^{b_j} \mathbb{P}(\tilde{Y}_1 \leq y | \tilde{V} = v, P = p) dv + \sum_{k=1}^{K^*} \int_{c_k}^{d_k} \mathbb{P}(\tilde{Y}_1 \leq y | \tilde{V} = v, P = p) dv \\ &= \sum_{j=1}^{J^*} \left(\mathbb{P}(Y \leq y, D = 1 | P = b_j) - \mathbb{P}(Y \leq y, D = 1 | P = a_j) \right) \\ &\quad + \sum_{k=1}^{K^*} \left(\mathbb{P}(Y \leq y, D = 1 | P = d_k) - \mathbb{P}(Y \leq y, D = 1 | P = c_k) \right) \\ &= \mathbb{P}(Y \leq y, D = 1 | P = p) - \mathbb{P}(Y \leq y, D = 1 | P = 0) = \mathbb{P}(Y \leq y, D = 1 | P = p), \end{aligned}$$

where the first equality is by definition, the second equality is by construction that \tilde{V} satisfies Assumption 1, the fourth equality is obtained by inserting the constructed counterfactual distributions, and the second last equality holds because $(0, p]$ can be expressed as unions of exclusive intervals $(\cup_{j=1}^{J^*} (a_j, b_j)) \cup (\cup_{k=1}^{K^*} (c_k, d_k])$.

Suppose that $p \in (a_{j^*}, b_{j^*}) \subseteq L^o(\mathcal{P})$ for some j^* , then the right hand side equals to

$$\begin{aligned}
\mathbb{P}(\tilde{Y} \leq y, \tilde{D} = 1 | P = p) &= \mathbb{P}(\tilde{Y}_1 \leq y, \tilde{V} \leq p | P = p) = \int_0^p \mathbb{P}(\tilde{Y}_1 \leq y | \tilde{V} = v, P = p) dv \\
&= \int_0^{a_{j^*}} \mathbb{P}(\tilde{Y}_1 \leq y | \tilde{V} = v, P = p) dv + \int_{a_{j^*}}^p \mathbb{P}(\tilde{Y}_1 \leq y | \tilde{V} = v, P = p) dv \\
&= \mathbb{P}(Y \leq y, D = 1 | P = a_{j^*}) + \int_{a_{j^*}}^p \frac{\partial}{\partial v} \mathbb{P}(Y \leq y, D = 1 | P = v) dv \\
&= \mathbb{P}(Y \leq y, D = 1 | P = a_{j^*}) + \mathbb{P}(Y \leq y, D = 1 | P = p) \\
&\quad - \mathbb{P}(Y \leq y, D = 1 | P = a_{j^*}) = \mathbb{P}(Y \leq y, D = 1 | P = p),
\end{aligned}$$

where the fourth equality holds by the above argument and the fifth equality holds by inserting the constructed counterfactual distributions. This completes the proof. \square

REFERENCES

- AIZER, A., AND J. J. DOYLE JR (2015): “Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges,” *The Quarterly Journal of Economics*, 130(2), 759–803.
- ARELLANO, M., AND S. BONHOMME (2017): “Quantile selection models with an application to understanding changes in wage inequality,” *Econometrica*, 85(1), 1–28.
- ARTEAGA, C. (2018): “The Cost of Bad Parents: Evidence from the Effects of Parental Incarceration on Children’s Education,” *Working Papers*.
- BHATTACHARYA, J., A. M. SHAIKH, AND E. VYTLACIL (2012): “Treatment effect bounds: An application to Swan–Ganz catheterization,” *Journal of Econometrics*, 168(2), 223–243.
- BHULLER, M., G. B. DAHL, K. V. LOKEN, AND M. MOGSTAD (2019): “Replication data for: Intergenerational Effects of Incarceration,” .
- BRINCH, C. N., M. MOGSTAD, AND M. WISWALL (2017): “Beyond LATE with a Discrete Instrument,” *Journal of Political Economy*, 125(4), 985–1039.
- CHEN, X., Y. FAN, AND V. TSYRENNIKOV (2006): “Efficient estimation of semiparametric multivariate copula models,” *Journal of the American Statistical Association*, 101(475), 1228–1240.
- DEATON, A. S. (2009): “Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development,” Discussion paper, National Bureau of Economic Research.
- DEATON, A. S., J. J. HECKMAN, AND G. W. IMBENS (2010): “Forum on the Estimation of Treatment Effects,” *The Journal of Economic Literature*, 48(2), 356–455.
- DI TELLA, R., AND E. SCHARGRODSKY (2013): “Criminal recidivism after prison and electronic monitoring,” *Journal of Political Economy*, 121(1), 28–73.
- DOBBIE, W., J. GOLDIN, AND C. S. YANG (2018): “The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges,” *American Economic Review*, 108(2), 201–40.
- HECKMAN, J. J., AND R. PINTO (2018): “Unordered monotonicity,” *Econometrica*, 86(1), 1–35.
- HECKMAN, J. J., AND E. VYTLACIL (2005): “Structural Equations, Treatment Effects, and Econometric Policy Evaluation1,” *Econometrica*, 73(3), 669–738.
- HECKMAN, J. J., AND E. J. VYTLACIL (1999): “Local instrumental variables and latent variable models for identifying and bounding treatment effects,” *Proceedings of the national Academy of Sciences*, 96(8), 4730–4734.
- HSU, Y.-C., C.-A. LIU, AND X. SHI (2019): “Testing generalized regression monotonicity,” *Econometric Theory*, 35(6), 1146–1200.
- HUBER, M., AND G. MELLACE (2015): “Testing instrument validity for LATE identification based on inequality moment constraints,” *Review of Economics and Statistics*, 97(2), 398–411.
- IMBENS, G. W., AND J. D. ANGRIST (1994): “Identification and Estimation of Local Average Treatment Effects,” *Econometrica*, 62(2), 467–475.

- JOE, H. (1994): “Multivariate extreme-value distributions with applications to environmental data,” *Canadian Journal of Statistics*, 22(1), 47–64.
- (1996): “Families of m -variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters,” *Lecture Notes-Monograph Series*, pp. 120–141.
- (1997): *Multivariate models and multivariate dependence concepts*. CRC Press.
- KÉDAGNI, D., AND I. MOURIFIÉ (2017): “Generalized Instrumental Inequalities: Testing IV Independence Assumption,” .
- KITAGAWA, T. (2015): “A Test for Instrument Validity,” *Econometrica*, 83(5), 2043–2063.
- KLING, J. R. (2006): “Incarceration length, employment, and earnings,” *American Economic Review*, 96(3), 863–876.
- LEE, S., AND B. SALANIÉ (2018): “Identifying effects of multivalued treatments,” *Econometrica*, 86(6), 1939–1963.
- MANSKI, C. F. (2011): “Policy analysis with incredible certitude,” *The Economic Journal*, 121(554), F261–F289.
- MANSKI, C. F., AND J. V. PEPPER (2000): “Monotone Instrumental Variables: With an Application to the Returns to Schooling,” *Econometrica*, 68(4), 997–1010.
- MOGSTAD, M., A. SANTOS, AND A. TORGOVITSKY (2018): “Using instrumental variables for inference about policy relevant treatment parameters,” *Econometrica*, 86(5), 1589–1619.
- MOURIFIÉ, I. (2015): “Sharp bounds on treatment effects in a binary triangular system,” *Journal of Econometrics*, 187(1), 74–81.
- MOURIFIÉ, I., AND Y. WAN (2017): “Testing Local Average Treatment Effect Assumptions,” *The Review of Economics and Statistics*, 99(2), 305–313.
- MUELLER-SMITH, M. (2015): “The criminal and labor market impacts of incarceration,” *Unpublished Working Paper*, 18.
- NELSEN, R. B. (2007): *An introduction to copulas*. Springer Science & Business Media.
- PEARL, J. (2013): “Direct and indirect effects,” *arXiv preprint arXiv:1301.2300*.
- TORGOVITSKY, A. (2019): “Partial identification by extending subdistributions,” *Quantitative Economics*, 10(1), 105–144.
- VYTLACIL, E. (2002): “Independence, Monotonicity, and Latent Index Models: An Equivalence Result,” *Econometrica*, 70(1), 331–341.