Identification and Estimation of Triangular Models with a Binary Treatment

Santiago Pereda-Fernández*

Banca d'Italia

November 15, 2021

Abstract

I study the identification and estimation of a nonseparable triangular model with an endogenous binary treatment. I impose neither rank invariance nor rank similarity on the unobservable term of the outcome equation. Identification is achieved by using continuous variation of the instrument and a shape restriction on the distribution of the unobservables, which is modeled with a copula. The latter captures the endogeneity of the model and is one of the components of the marginal treatment effect, making it informative about the effects of extending the treatment to untreated individuals. The estimation is a multi-step procedure based on rotated quantile regression. Finally, I use the estimator to revisit the effects of Work First Job Placements on future earnings.

Keywords: Copula, endogeneity, policy analysis, quantile regression, unconditional distributional effects

JEL classification: C31, C36, J22, J31

*Banca d'Italia, Via Nazionale 91, 00184 Roma, Italy. All remaining errors are my own. The views presented in this paper do not necessarily reflect those of the Banca d'Italia. I can be reached via email at santiago.pereda@bancaditalia.it. This paper has greatly benefited from discussions with Stéphane Bonhomme. I would also like to thank the editor, the associate editor, two anonymous referees, Manuel Arellano, Dmitry Arkhangelsky, Domenico Depalo, Bryan Graham, Christian Hansen, Giovanni Mellace, Magne Mogstad, Quang Vuong, Paolo Zacchia, and seminar participants at CEMFI, Banca d'Italia, the 5th Conference of the IAAE and the 8th Conference of the ICEEE for their helpful comments and discussion.

1 Introduction

Consider the following triangular system of equations:

$$Y = g\left(D, X, U_D\right) \tag{1}$$

$$D = \mathbf{1}\left(\pi\left(Z\right) - V > 0\right) \tag{2}$$

where the observed variables are the continuous outcome Y, the binary treatment D, the vector of covariates X, and the vector Z, that typically includes an instrument (Z_1) and the covariates, *i.e.*, $Z \equiv (Z_1, X')'$; U_D and V are univariate unobservable terms. Equations 1-2 are referred to as the outcome and selection equations, respectively. Both of them are assumed to be monotonic with respect to their respective unobservable term. This paper focuses on the identification and estimation of the structural function $g\left(\cdot,\cdot,\cdot\right)$, as well as functionals of it. Notably, I consider the Marginal Treatment Effect (MTE, Björklund and Moffitt, 1987) and the unconditional Quantile Treatment Effect (QTE).

The only source of exogenous variation is the instrument Z_1, which conditional on the covariates is independent of the unobservables, *i.e.*, $(U_1, U_0, V) \perp Z_1 | X$. When it displays enough variation such that the support of the propensity score (conditional on the covariates) equals the unit interval, the model is fully nonparametrically identified, a case known as identification at infinity (Heckman, 1990; Heckman and Vytlacil, 2007b). However, instruments do not display that much variability in practice. Therefore, identification under these conditions often hinges on assumptions that restrict the degree of heterogeneity, which could hinder the credibility of the estimates of the effects of counterfactual policy interventions (Heckman et al., 1997; Bisbee et al., 2017).

A common approach is to impose shape restrictions on either the outcome equation or the MTE, which usually imply an additively linear specification, *e.g.*, Das (2005).[1] This assumption restricts the shape of the MTE, making them invariant with respect to the

[1] A related setting is one with a weakly separable but not additively separable model. This was considered by Vytlacil and Yildiz (2007), who allowed for a broader set of dependent variables, including discrete choice models.

covariates. Thus, changes in the covariates result in parallel shifts. Moreover, even when the identified model is nonseparable, existing estimators require separability to be tractable.

Another approach to achieve identification is based on assumptions on the distribution of the unobservables that constrain the type of self-selection. A prominent example is rank invariance (Vuong and Xu, 2017; Feng, 2019; Abrevaya and Xu, 2021), which implies that an individual's rank in the distribution of potential outcomes is the same under both treatment status. A less restrictive assumption is rank similarity (Chernozhukov and Hansen, 2005, 2006; Wüthrich, 2019a), which allows this disturbance term to differ depending on the treatment status. However, the relation between the treatment and each disturbance is the same across treatment status. This rules out the possibility that the treated individuals are relatively better off when treated, and relatively worse off when not. In contrast, one could be agnostic about the distribution of the unobserved term conditional on each treatment status, allowing them to differ, a case to which I refer as rank dissimilarity.

In this paper I make the following contributions. First, I show that it is possible to fully identify a nonseparable model if the instrument is continuously distributed and the joint distribution (copula) of the unobservables satisfies a shape restriction. The identification requires neither rank similarity nor a large support assumption on the instrument. Second, I propose a quantile regression estimator that can be used to estimate functionals for policy analysis and how to conduct uniformly valid inference. Third, I revisit the estimation of the effects of the Work First Job Placements program on earnings. I estimate the unconditional QTE and the MTE, analyzing the role of self-selection in shaping the results using some novel decompositions, and assessing the restrictiveness of the rank similarity assumption.

The identification strategy draws from Arellano and Bonhomme (2017a), and it is based on the combination of two assumptions: continuous variation of the instrument, and a shape restriction of the copulas of the unobservables. In particular, I assume that the copula belongs to a parametric family as in Han and Vytlacil (2017). Such an assumption allows to extrapolate the identification region from the support of the propensity score to the whole unit interval. This assumption could be further relaxed to real analyticity of the copula,

similarly to Arellano and Bonhomme (2017a) for quantile selection models. This is a shape restriction satisfied by many copulas, including the Bernstein copula. The latter is a flexible copula that can arbitrarily approximate any continuous copula when its order is high enough. Moreover, even if the copula is unrestricted, the structural functions are partially identified.

I model self-selection into treatment using a different copula for each treatment status. Hence, the amount of self-selection, understood as the distribution of U_D, conditional on the treatment status $D = d$, may be different for $d = 1$ and $d = 0$. The unobservables can then be interpreted in terms of the conditional quantiles, or ranks, of the latent distribution of potential outcomes. This makes makes the distributional analysis intuitive and naturally translates into quantile regression methods for the estimation.

Building on the identification result, I propose a multi-step estimator based on Rotated Quantile Regression (RQR, Arellano and Bonhomme, 2017a,b), that has several desirable properties: (i) it imposes neither rank similarity nor additive separability; (ii) it includes all interactions between the treatment and the covariates by default; (iii) its asymptotic distribution is Gaussian and converges at the $\sqrt{n}$ rate; (iv) it can be easily extended to multivalued treatments. Relative to Arellano and Bonhomme (2017a), I show the uniform asymptotic distribution of the estimator and how to conduct inference. Moreover, I propose an estimation algorithm when one uses Bernstein copulas.

There is a vast literature on the identification and estimation of triangular models with a binary treatment.[2] Chernozhukov and Hansen (2005, 2006) set an important milestone in the literature, defining a quantile treatment effect framework based on a nonseparable model. Moreover, they proposed the Instrumental Variables Quantile Regression (IVQR) estimator, which is not additively separable, but requires rank similarity. In contrast, Local Instrumental Variables (LIV, Heckman and Vytlacil, 1999, 2005) is an approach that does not require rank similarity. Carneiro and Lee (2009) extended earlier works by showing the identification of the quantile treatment effects. However, the identification relies on the large

[2]For triangular systems of equations with a continuous treatment see *e.g.*, Chesher (2003), Newey and Powell (2003), Horowitz and Lee (2007), Lee (2007), Imbens and Newey (2009), Jun (2009), D'Haultfœuille and Février (2015), or Torgovitsky (2015).

support assumption for the instrument.

When the instrument is binary, Local Average Treatment Effect methods (LATE, Imbens and Angrist, 1994; Abadie, 2003) identify the average treatment effect on the population of compliers. These methods have several advantages, such as their simplicity, their intuitive interpretation or their internal validity. Moreover, they have been used in other frameworks, such as the regression discontinuity design (Frandsen et al., 2012), or the estimation of conditional and unconditional quantile treatment effects for the compliers (Abadie et al., 2002; Frölich and Melly, 2013). However, one shortcoming is that they do not identify the effect on the population unaffected by the instrument, which may be more relevant for the evaluation of the effect of counterfactual policies. In contrast, the MTE approach allows to consider a wider range of policy effects that may affect individuals who were not induced to treatment by the instrument. Thus, the approach in this paper can be used to analyze such changes and is related to other approaches that consider the extrapolation of the local effects to the rest of the population, such as Angrist and Fernandez-Val (2013), Kowalski (2016), Brinch et al. (2017) or Mogstad et al. (2018).

The methods presented in this paper are applied to the estimation of the effect of Work First Job Placements on earnings. This public employment program focused on quickly finding a job for unemployed low-skilled workers. Autor and Houseman (2010) and Autor et al. (2017) found that temporary-help jobs had a negative effect on earnings at high quantiles, and null for the rest of the distribution, whereas direct-hire placements led to an increase in earnings for more than half of the distribution.

I extend their results by looking at the effects on the unconditional distribution of earnings, and using a model that allows for a larger degree of heterogeneity. I estimate a positive effect for most of the distribution of future earnings for both types of placements. I also find strong evidence against the rank similarity assumption, and around 40% of the difference between the treated and the untreated can be explained by differences in the amount of self-selection. However, the MTE takes negative values for a proportion of the population. These findings can be reconciled with the rank dissimilarity assumption, as the

excess selection in the treatment groups relative to the control group is responsible for the majority of the heterogeneity captured by the MTE. Consequently, extending the treatment to all individuals would not improve the distribution of earnings at all quantiles.

The rest of the paper is organized as follows: Section 2 introduces the model and presents the identification result. Section 3 describes the estimation method and how to conduct inference. The methods presented in this paper are illustrated with the empirical application in Section 4. Finally, Section 5 concludes. All proofs are shown in Appendix A.

2 The Model

To better understand the model, it is useful to think in terms of the potential outcomes framework. If the treatment $D = d$ could be randomly assigned to the population, then the distribution of U_d would be independent of D. However, when individuals self-select into treatment, this is no longer true, making the identification more difficult. Equations 1-2 can be derived from a generalized Roy model with imperfect information in which individuals are uncertain about the exact value of the outcome under each treatment status. However, they can form themselves an expectation based on the information they have available. Whenever the expected net surplus of being treated is positive, they choose to receive it. This model is presented in Appendix C.

Following Heckman and Vytlacil (2005), it is possible to normalize V to be uniformly distributed, which allows to interpret $\pi(Z)$ as the propensity score. As shown by Vytlacil (2002), this is equivalent to the monotonicity condition in Imbens and Angrist (1994). Similarly, under the assumption that the outcome depends monotonically on U_D, it is possible to normalize it to be uniformly distributed, such that $g(\cdot,\cdot,\cdot)$ can be interpreted as the Structural Quantile Function (SQF), and U_D is interpreted as the rank of the SQF.

To justify the normalizations, let the data generating process be determined by the functions $\tilde{g}$, $\tilde{F}_{\tilde{U}_1,\tilde{U}_0,\tilde{V}|X}$ and $\tilde{\pi}$, such that the marginal distributions of the unobservables are not necessarily uniformly distributed. The following Lemma establishes the observational

equivalence between this model and the normalized one:

Lemma 1. *Let* $Y = \tilde{g}\left(D, X, \tilde{U}_D\right)$ *and* $D = \mathbf{1}\left(\tilde{\pi}(Z) - \tilde{V} > 0\right)$*, where the distribution of the unobservables is given by* $\tilde{F}_{\tilde{U}_1,\tilde{U}_0,\tilde{V}|X}$*, with marginal distributions* $\tilde{F}_{\tilde{U}_D|X}$ *and* $\tilde{F}_{\tilde{V}|X}$*, that may depend non-trivially on* X*. Then, there exists* g *such that the model given by Equations 1-2, where* $U_D|X \sim U[0,1]$ *and* $V|X \sim U[0,1]$*, generates the same distribution of* (Y, D, Z)*.*

This normalization is particularly convenient to understand endogeneity, as the joint distribution of the unobservables is a copula.[3] Formally, the unobservables have the following conditional distribution: $U_0, U_1, V|X \sim C_X(U_0, U_1, V|X)$.[4] However, only one of the two treatment status is observed for each individual, so the joint distribution between U_0 and U_1 is not identified.[5] Hence, the focus lies on the bivariate copulas between U_d and V, conditional on $X = x$, denoted by $C_{d,x}(U_d, V)$, for $d = 0, 1$. Note that the normalization in Lemma 1 implies that the marginals are independent of X, but the copula is generally not.

Much of the literature has focused on the rank invariance ($U_0 = U_1$) or rank similarity cases ($C_{0,x} = C_{1,x}$). The first assumption implies that unobserved ability is unidimensional, so more able individuals would perform relatively well under either treatment status. Rank similarity allows ability to be bidimensional, but it is still the case that those who perform well when untreated also tend to perform well when treated. Therefore, these assumptions rule out the possibility that those who perform relatively well when they are treated are not necessarily those who perform relatively well when they are not. I allow for this possibility ($C_{0,x} \neq C_{1,x}$), and I refer to it as *rank dissimilarity*.

The following example clarifies the implications of each assumption. Denote earnings by Y, the possession of a college degree by D, and U_1 and U_0 be measures of intelligence and

[3] Given random variables $W_1, ..., W_N$, with marginals $F_1(w_1)$,...,$F_N(w_N)$, the copula is defined as $C(F_1(w_1), ..., F_N(w_N)) \equiv \mathbb{P}(W_1 \leq w_1, ..., W_N \leq w_N)$. Sklar (1959) showed that any continuous multivariate distribution can be written in terms of a copula that has the ranks of the individual components as arguments.

[4] The dimensionality of the unobservables places some restrictions on the amount of heterogeneity, *e.g.*, it rules out non-monotonic models such as random coefficients. A richer model would consider unobservables of higher dimension, although this type of models are generally not point-identified. See, *e.g.*, Hahn and Ridder (2011); Kasy (2011); Hoderlein et al. (2017); Masten (2018) for further details.

[5] This is akin to the identification of the distribution of the treatment effect, which is not point identified, but can be bounded (Fan and Park, 2010; Firpo and Ridder, 2019).

physical prowess, respectively. Moreover, assume that the productivity at work depends on intelligence when one has a college degree, and on physical prowess otherwise. Under rank invariance, both unobserved characteristics are perfectly correlated, so each individual's rank is the same in the distribution of potential earnings with and without a college degree. Rank similarity allows for differences in the level of intelligence and physical prowess. However, the correlation between holding a college degree and intelligence is the same as the correlation between holding a college degree and physical prowess. Hence, those who are likely to be top earners with a college degree, are also likely to be top earners without it. In contrast, rank dissimilarity allows those with a high propensity to have a college degree to be on average more intelligent and have a lower level of physical prowess. Moreover, those less likely to have a college degree could have a higher level of physical prowess and even have higher earnings without the college degree.

This example highlights the usefulness of the copula for policy making: it is informative about the potential effects of extending the treatment to the untreated by acknowledging how they are selected. For example, consider two types of individuals: one with a high propensity score, and another with a low one. If none of them were treated, we would expect a larger value of the unobserved variable V for the first individual. If the copula displayed a negative degree of correlation, then the first individual would be expected to rank lower than the second individual in the distribution of treated individuals. On the other hand, if the copula displayed no correlation, then both individuals would be expected to rank similarly. In other words, it is important to account for differences in self-selection to appropriately assess the impact of extending the treatment.

2.1 Identification of the Structural Functions

Define the conditional copulas by $G_{1,x}(\tau, \pi(z)) \equiv \frac{C_{1,x}(\tau,\pi(z))}{\pi(z)} = \mathbb{P}(U_1 \leq \tau | D = 1, z)$ for the treated, and $G_{0,x}(\tau, \pi(z)) \equiv \frac{\tau - C_{0,x}(\tau,\pi(z))}{1-\pi(z)} = \mathbb{P}(U_0 \leq \tau | D = 0, z)$ for the untreated. The

distribution of the outcome for the treated is given by

$$F_{Y|D=1,Z}(y|z) = \int_0^1 \mathbf{1}(g(1,x,u_1) \leq y)\, dG_{1,x}(u_1, \pi(z)) \tag{3}$$

where $\mathbf{1}(\cdot)$ denotes the indicator function. Evaluating Equation 3 at $y = g(1,x,\tau)$ yields

$$\begin{aligned} F_{Y|D=1,Z}(g(1,x,\tau)|z) &= \int_0^1 \mathbf{1}(g(1,x,u_1) \leq g(1,x,\tau))\, dG_{1,x}(u_1, \pi(z)) \\ &= \int_0^1 \mathbf{1}(u_1 \leq \tau)\, dG_{1,x}(u_1, \pi(z)) = G_{1,x}(\tau, \pi(z)) \end{aligned}$$

where the second equality follows by the monotonicity of $g(\cdot,\cdot,\cdot)$ with respect to its last argument. Similarly, the distribution for the untreated equals

$$F_{Y|D=0,Z}(y|z) = \int_0^1 \mathbf{1}(g(0,x,u_0) \leq y)\, dG_{0,x}(u_0, \pi(z)) \tag{4}$$

and evaluating Equation 4 at $y = g(0,x,\tau)$ yields $F_{Y|D=0,Z}(g(0,x,\tau)|z) = G_{0,x}(\tau, \pi(z))$.

Therefore, both distributions depend on three components: the SQF of Y, the propensity score $\pi(z)$, and the copulas $C_{0,x}$ and $C_{1,x}$. Consider the following assumptions:

Assumption 1. *(U_0, U_1, V) are jointly statistically independent of Z_1 given $X = x$.*

Assumption 2. *The bivariate distributions (U_0, V) and (U_1, V), conditional on $X = x$, are absolutely continuous with respect to the Lebesgue measure. Moreover, the marginal distributions of U_0, U_1, V are uniform conditional on all x.*

Assumption 3. *$F_{Y|D=0,Z}(y|z)$, $F_{Y|D=1,Z}(y|z)$, and their inverses in y are strictly increasing.*

Assumption 4. *Denote the support of $\pi(Z)$ conditional on $X = x$ by $\mathcal{P}_x$. $\forall x \in \mathcal{X}$, $\mathcal{P}_x \subset [0,1]$ is an open interval.*

Assumption 1 is the exclusion restriction, which imposes the independence of the ranks of the selection equation and the SQF. In terms of the copula, it can vary with X, but not with Z_1. Moreover, even if the copulas do not depend on X, Assumption 1 does not imply full

independence between (U_1, U_0, V) and Z. Assumptions 2 and 3 imply that the SQF and the propensity score display continuous variation with respect to the unobservables, ruling out jumps. Moreover, they allow Equations 1-2 to represent the conditional quantile function of the potential outcomes Y_d^*: by normalizing the marginal distributions of the ranks to be uniform, their joint distribution is a well-defined copula. The support Assumption 4 requires the instrument to display some continuous variation that maps into the propensity score.

Denote the support of X by $\mathcal{X}$ and the support of Z_1 given $X = x$ by $\mathcal{Z}_x$. Then, the following two restrictions on the copula hold:

Lemma 2. *Let $x \in \mathcal{X}$, $z \equiv (z_1, x)$ and $z' \equiv (z_1', x)$. Then, under Assumptions 1 to 4:*

$$F_{Y|D=1,Z}\left(F^{-1}_{Y|D=1,Z}\left(\tau|z'\right)|z\right) = G_{1,x}\left(G^{-1}_{1,x}\left(\tau, \pi\left(z'\right)\right), \pi\left(z\right)\right) \forall \left(z_1, z_1'\right) \in \mathcal{Z}_x \times \mathcal{Z}_x \tag{5}$$

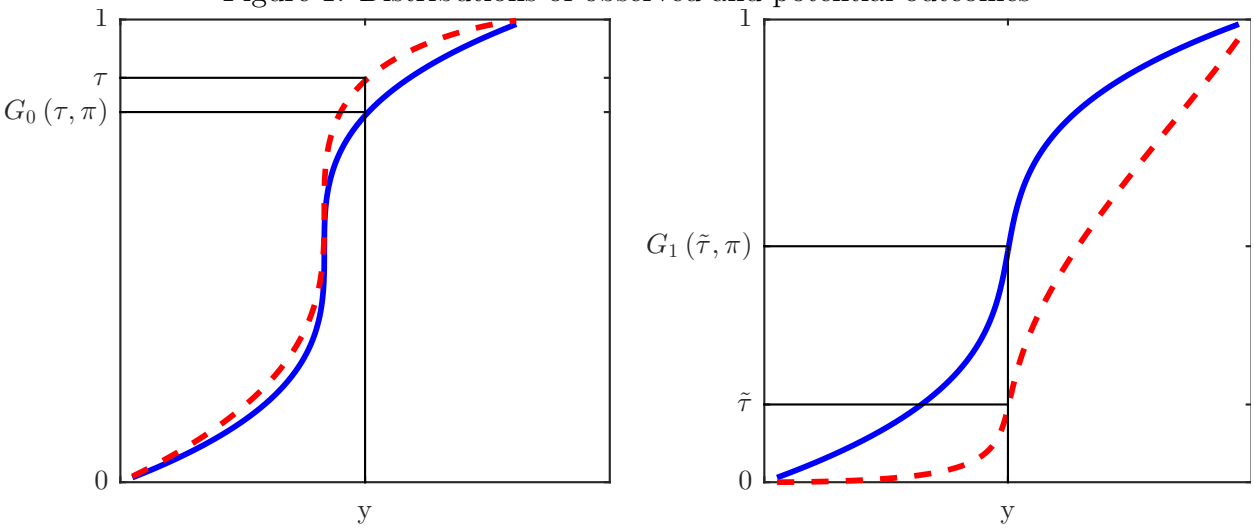
$$F_{Y|D=0,Z}\left(F^{-1}_{Y|D=0,Z}\left(\tau|z'\right)|z\right) = G_{0,x}\left(G^{-1}_{0,x}\left(\tau, \pi\left(z'\right)\right), \pi\left(z\right)\right) \forall \left(z_1, z_1'\right) \in \mathcal{Z}_x \times \mathcal{Z}_x \tag{6}$$

Equations 5-6 show that it is possible to place some restrictions on the copulas using the observed distributions $F_{Y|D=d,Z}$, $d = 0, 1$, without knowing the SQF. For them to be informative about the copulas, the instrument must come from a non-degenerate distribution. Therefore, variations in the propensity score due to the instrument do not affect the SQF, and, the change induced in the distribution of observed outcomes operates entirely through differences in self-selection.

Intuitively, the conditional copula $G_{d,x}$ can therefore be interpreted as a mapping between the rank in the latent distribution of potential outcomes, τ, and the quantile of the outcome equation conditionally on $D = d$. This is shown in Figure 1, which relates the distribution of the conditional outcome as given by Equation 1, to the distribution that would arise if the treatment was randomly allocated.

Equations 5-6 hold true for all values of τ and values of the propensity score in $\mathcal{P}_x$. Unless one can apply the identification at infinity argument, these two restrictions are not enough to fully identify the SQF and the copula. A way to achieve it is by making a parametric

Figure 1: Distributions of observed and potential outcomes



Notes: The solid line on the left and right panels denote the distribution of potential outcomes conditional on $D = 0, 1$, respectively; the dashed line on the left and right panels respectively denote $F_{Y|D=0,Z}$ and $F_{Y|D=1,Z}$; $G_0(\tau,\pi)$ and $G_1(\tilde{\tau},\pi)$ are shorthands for $G_{0,x}(\tau,\pi(z))$ and $G_{1,x}(\tilde{\tau},\pi(z))$, respectively.

assumption on the copula:

Assumption 5. *The copula $C_{d,x}(\tau,\pi)$ is known up to a scalar parameter $\theta_{d,x} \in \Theta_{d,x}$, for $d = 0, 1$ and $\forall x \in \mathcal{X}$. $C_{d,x} : (0,1)^2 \to (0,1)$ is uniformly continuous and twice continuously differentiable in its arguments and in $\theta_{d,x}$ a.e. Its density, $c_{d,x}(u,v;\theta_d)$, is well-defined and finite. Moreover, for any $\theta_1 < \theta_2$, $C_{d,x}(\tau,\pi;\theta_2)$ is strictly more stochastically increasing in joint distribution than $C_{d,x}(\tau,\pi;\theta_1)$.*

Assumption 5 restricts the shape of the copulas, conditional on $X = x$, to depend on a scalar parameter.[6] Moreover, this dependence needs to satisfy a specific ordering as the parameter value varies. We introduce the following ordering concepts (Han and Vytlacil, 2017):

Definition 1. *For random variables W_1, W_2, the conditional distribution $F_{1|2}$ is stochastically increasing if $\mathbb{P}(W_1 > w_1|W_2 = w_2) = 1 - F_{1|2}(w_1|w_2)$ is increasing in w_2 for all w_1.*

[6]Parametric copulas have been used to model latent variables in a variety of setups: quantile selection models (Arellano and Bonhomme, 2017a), bivariate probit models with dummy endogenous regressors (Han and Vytlacil, 2017), triangular models with continuous endogenous variables (Pereda-Fernández, 2016), both linear and non-linear panel data (Prokhorov and Schmidt, 2009; Pereda-Fernández, 2021), and in time series to model nonstationary and nonlinear dynamics (Chen et al., 2021) or GARCH models (Chen et al., 2021).

Definition 2. *Suppose that* $F(w_1, w_2)$ *and* $\tilde{F}(w_1, w_2)$ *are continuous in* (w_1, w_2). $\tilde{F}$ *is strictly more stochastically increasing in joint distribution than* F *if* $\tilde{F}^{-1}(w_1, F(w_1, w_2))$ *and* $\tilde{D}^{-1}(w_1, D(w_1, w_2))$ *are strictly increasing in* w_2, *where* $D(w_1, w_2) \equiv F(w_1) - F(w_1, w_2)$, $\tilde{D}(w_1, w_2) \equiv \tilde{F}(w_1) - \tilde{F}(w_1, w_2)$, *and* $F(w_1)$ *and* $\tilde{F}(w_1)$ *are the marginals of* $F(w_1, w_2)$ *and* $\tilde{F}(w_1, w_2)$, *respectively.*

The property introduced in Definition 1 captures a positive dependence between the two variables. Moreover, it is related to first order stochastic dominance, as for any $w_2' < w_2$, $F_{1|2}$ is stochastically increasing if $F_{1|2}(w_1|w_2)$ first order stochastically dominates $F_{1|2}(w_1|w_2')$. On the other hand, the second concept allows to order different distributions according to their degree of first order stochastic dominance. This is particularly useful when considering the copulas generated by the different possible values of their parameters. Several of the most common choices of copulas satisfy Assumption 5, including the Gaussian, the Clayton, and the the Frank.[7]

This assumption is based on Assumptions 3 and 6 in Han and Vytlacil (2017), which were used to identify a bivariate probit model with dummy endogenous regressors. Relative to their setting, the moments used for the identification of the copula do not depend on the remaining parameters, which simplifies the identification expressions. Despite this, the identification conditions are similar, as established by the following proposition:

Proposition 1. *Let Assumptions 1 to 5 hold, and* $x \in \mathcal{X}$. *Then, for* $d = 0, 1$

- *the correlation parameter* $\theta_{d,x}$ *is globally identified.*
- *the SQF* $\tau \to g(d, x, \tau)$ *is globally identified.*

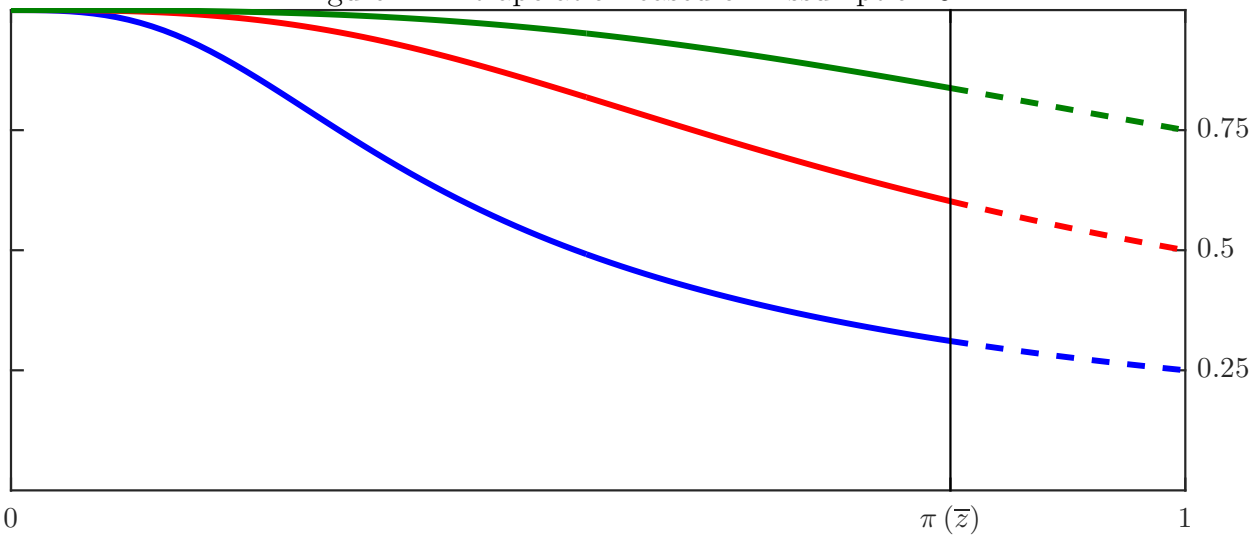
Proposition 1 is the main identification result. The identification argument is as follows: Equations 5-6 are well defined for any $\pi(z), \pi(z') \in \mathcal{P}_x$. Under the assumptions listed in the proposition, they are enough to pin down the copula correlation parameters. As a consequence, by Assumption 5 it is possible to extrapolate the identification region from $\mathcal{P}_x$

[7] See (Han and Vytlacil, 2017) for further examples.

to the unit interval, and thus the conditional copulas $G_{d,x}(\tau, \pi)$ are identified. The latter, in turn, allow to identify the SQF for all values of $\tau \in [0,1]$ and $d = 0, 1$.

To get some insight, consider Figure 2. The solid line represents the value of $F_{Y|D=1,Z}(y|z)$ over the support of the propensity score, given by $\mathcal{P}_x = [0, \pi(\overline{z})]$. If $\pi(\overline{z}) < 1$, then the identification requires some kind of extrapolation. Under Assumption 5, small changes in the value of the propensity score would translate into small changes in the conditional distribution of the outcome, which operate exclusively through the selection equation. Thus, $F_{Y|D=1,Z}(y|z)$ can be extrapolated from $\mathcal{P}_x$ to the unit interval, represented with the dashed lines. It is important to note that even if the identification holds when the support of the propensity score is small, the extrapolation is more plausible in a neighborhood of the observed support of the propensity score. Therefore, Assumption 5 is increasingly stronger as the distance increases.[8]

Figure 2: Extrapolation based on Assumption 5



Notes: The solid and dashed lines represent the value of $F_{Y|D=1,Z}(g(1,x,\tau)|\overline{z}) = \tau$ for $\tau = \{0.25, 0.5, 0.75\}$ when $\pi(z) \in \mathcal{P}_x$ and when $\pi(z) \notin \mathcal{P}_x$, respectively.

There exist alternatives to obtain identification without Assumption 5. The working paper version assumed that the copulas are real analytic.[9] Such an assumption covers most

[8] This point matters particularly for the precision of the estimates. See Appendix H for further details.

[9] A function $f(x)$ is real analytic at x_0 if $\forall x$ in a neighborhood around x_0 one can write $f(x) = \sum_{j=0}^{\infty} a_j (x - x_0)^j$, where a_j, $j = 0, ...$ are the polynomial coefficients. In words, the function f can be

parametric copulas, as well as many others. Identification under this assumption, along with a comparison with alternative methods are explored in Appendix D.

2.2 Identification of Functionals

Knowledge of the structural functions allows to identify several parameters of interest, which are frequently the goal of the researchers. A remarkably important one is the MTE, which can be interpreted as the expected treatment gain for those individuals with characteristics $X = x$ and $V = v$, *i.e.*, the expected gain for those who would be indifferent between being treated and not when $\pi(z) = v$. Many parameters of interest can be obtained by averaging the MTE with appropriate known weights, such as the Average Treatment Effect or the Average Treatment Effect on the Treated (Heckman and Vytlacil, 2005). In the current setting, the MTE can be easily expressed in terms of the structural functions:

$$\Delta_{MTE}(x, v) = \int_0^1 g(1, x, u)\, dC_{1,x}(u|v) - \int_0^1 g(0, x, u)\, dC_{0,x}(u|v) \tag{7}$$

Note that the identification of the MTE may not require Assumption 5 if the support of the propensity score equals the unit interval (Heckman and Vytlacil, 2001).

Another function of interest is the Quantile Treatment Effect, *i.e.*, the difference between the treated and the untreated unconditional quantile functions of potential outcomes. To obtain the latter, integrate the conditional distributions of potential outcomes with respect to the distribution of the covariates, and then invert the resulting unconditional distributions (Machado and Mata, 2005; Chernozhukov et al., 2013). They are given by

$$\Delta_{QTE}(\tau) = Q_{Y|D=1}(\tau) - Q_{Y|D=0}(\tau) \tag{8}$$

where $Q_{Y|D=d}(\tau) = \min\left\{y : F_{Y|D=d}(y) \geq \tau\right\}$, $F_{Y|D=d}(y) = \int_{\mathcal{X}} \int_0^1 \mathbf{1}(g(d, x, u) \leq y)\, du dF_X(x)$.

expressed as a power convergence series. If $f(x)$ is real analytic at all $x \in \mathcal{X}$, where $\mathcal{X}$ is an open interval, then the function is real analytic on $\mathcal{X}$.

3 Estimation

3.1 Estimation of the Structural Parameters

For the estimation I consider the following set of assumptions:

Assumption 6. $(Y_i, D_i, Z_i')'$ *are iid for* $i = 1, ..., n$, *defined on the probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and take values in a compact set.*

Assumption 7. $g(d, x, \tau) = x'\beta_d(\tau)$ *for* $d = 0, 1$, *where* β_d *is continuous and such that* $g(d, x, \tau)$ *is increasing in its last argument.*

Assumption 8. $\pi(Z) \equiv \pi(Z; \gamma)$, *with* $\dim(\gamma) < \infty$. $\pi(Z; \gamma)$ *is continuously differentiable with respect to* γ. *Moreover, there exists an asymptotically linear estimator* $\hat{\gamma}$ *that admits the following representation:* $\hat{\gamma} - \gamma = -B^{-1}\frac{1}{n}\sum_{i=1}^{n} s(d_i, z_i; \gamma) + o_P\left(\frac{1}{\sqrt{n}}\right)$.

Assumption 9. *Let* $\beta(\tau) \equiv \left(\beta_1(\tau)', \beta_0(\tau)'\right)'$ *and* $\theta \equiv (\theta_1', \theta_0')'$. $\forall \tau \in \mathcal{T}$, $\left(\beta(\tau)', \theta', \gamma'\right)' \in int\mathcal{B} \times \Theta \times \Gamma$, *where* $\mathcal{B} \times \Theta \times \Gamma$ *is compact and convex, and* $\mathcal{T} = [\varepsilon, 1 - \varepsilon]$, *for some constant* ε *that is used to avoid the estimation of extreme quantiles.*[10]

Assumption 10. Y *has conditional density that is bounded from above and away from zero, a.s. on a compact set* $\mathcal{Y}$. *The density is given by* $f_{Y|D,Z}(y)$ *for* $D = 0, 1$.

Assumption 11. *Matrices of derivatives of the moments* $J_{\beta_d}(\tau)$, $\tilde{J}_{\beta_d}(\tau)$, $J_{\gamma_d}(\tau)$, $\tilde{J}_{\gamma_d}(\tau)$, $J_{\theta_d}(\tau)$, $\tilde{J}_{\theta_d}(\tau)$ *for* $d = 0, 1$, *as defined in Appendix A, are continuous and have full rank, uniformly over* $\mathcal{B} \times \Theta \times \Gamma \times \mathcal{T}$ *and* $d = 0, 1$.

Assumption 6 describes the sampling process of the data. The linear quantile model imposed by Assumption 7 is standard in the literature and it makes the extrapolation to values of X not observed in the data easier.[11] Note however, that it would be possible to relax this assumption, allowing for nonlinear quantile functions as long as the resulting SQF

[10] Functionals based on the conditional distribution will also be sensible to the trimming constant. See *e.g.*, Chernozhukov and Hansen (2005) for further details.

[11] See, *e.g.*, Koenker and Bassett (1978), Chernozhukov and Hansen (2005), or Angrist et al. (2006).

is continuous and increasing in τ. For example, one could specify a partially linear model (Lee, 2003) or a generalized linear model with a known link function (Horowitz et al., 2004). However, these approaches could be subject to the curse of dimensionality, making them infeasible in practice.

Assumption 8 is made for simplicity, and it is satisfied by several estimation methods, including maximum likelihood. Assumption 9 is a regularity condition. Assumption 10 restricts the analysis to dependent variables that have a well-defined and finite conditional density, and Assumption 11 requires the existence of moments and their full rank to derive the asymptotic variance of the estimator.

Because of the endogeneity of the treatment, standard quantile regression would yield inconsistent estimates of the $\beta(\cdot)$ parameters. However, note that using the mapping provided by the conditional copula and Assumption 7, it is possible to write

$$F^{-1}_{Y|D=d,Z}(\tau|z) = x'\beta_d\left(G^{-1}_{d,x}(\tau, \pi(z))\right) \tag{9}$$

where we have used the strict monotonicity of $G_{d,x}$. As highlighted by Equation 9, the quantile of the distribution of observed outcomes does not coincide with the quantile index of the SQF. Hence, by appropriately rescaling the quantile index for each observation, it is possible to consistently estimate the SQF. This constitutes the basis for using RQR, *i.e.*, using the rotated check function $\rho_{G_{d,x}(\tau,\pi;\theta)}$ rather than the standard check function ρ_τ. Specifically, the estimation consists of the following steps:

1. Estimate the propensity score by $\hat{\pi}(z_i) \equiv \pi(z_i, \hat{\gamma})$ (*e.g.*, by MLE).

2. Fix a value of $t \in \Theta$. For $d = 0, 1$ and $\tau \in \mathcal{T}$, compute $\hat{\beta}_d(\tau; t)$ as

$$\hat{\beta}_d(\tau; t) \equiv \arg\min_{b\in\mathcal{B}} \sum_{i=1}^{n} \mathbf{1}(d_i = d)\, \rho_{G_{d,x}(\tau,\hat{\pi}(z_i);t)}(y_i - x_i'b) \tag{10}$$

   where $\rho_u(x) \equiv xu\mathbf{1}(x \geq 0) - (1-u)x\mathbf{1}(x < 0)$ denotes the check function.

3. Estimate the copula parameters for $d = 0, 1$ by minimizing over $t \in \Theta$:

$$\hat{\theta}_d \equiv \arg\min_{t\in\Theta} \left\| \sum_{i=1}^{n} \int_{\varepsilon}^{1-\varepsilon} \mathbf{1}\left(d_i = d\right) \varphi\left(\tau, z_i\right) \left[\mathbf{1}\left(y_i \leq x_i' \hat{\beta}_d\left(\tau; t\right)\right) - G_{d,x}\left(\tau, \hat{\pi}\left(z_i\right); t\right) \right] d\tau \right\| \tag{11}$$

where $\varphi\left(\tau, z_i\right)$ is an instrument function.[12]

4. The slope parameters are obtained by $\hat{\beta}_d\left(\tau\right) \equiv \hat{\beta}_d\left(\tau; \hat{\theta}_d\right)$ for $d = 0, 1$.

In practice, the estimation is done for a grid of values of τ, *e.g.*, $\tau = \{0.01, ..., 0.99\}$. Equation 10 can be solved with standard quantile regression techniques by rotating the loss function. On the other hand, Equation 11 involves non-convex optimization and it constitutes the slowest step of the estimation. For most parametric copulas, the number of parameters is small, so it is possible to discretize Θ to a grid of values. In that case, step 2 would be performed for each value of t in the grid, and then in step 3 one would pick the minimizer of the criterion function in step 3 among those values. This approach suffers from the curse of dimensionality when the number of parameters is large, such as Bernstein copulas of a high enough order. An alternative approach that combines random search increasing the order of the copula sequentially is described in Appendix E.

The asymptotic distribution of this estimator is established by the following theorem:

Theorem 1. *Let* $\hat{\beta}\left(\tau\right) \equiv \left(\hat{\beta}_1\left(\tau\right)', \hat{\beta}_0\left(\tau\right)'\right)'$, $\hat{\phi} \equiv \left(\hat{\theta}_1', \hat{\theta}_0', \hat{\gamma}'\right)'$ *and* $\hat{\vartheta}\left(\tau\right) \equiv \left(\hat{\beta}\left(\tau\right)', \hat{\phi}'\right)'$. *Under Assumptions 1-11, the following hold:*

- $\sqrt{n}\left(\hat{\beta}\left(\tau\right) - \beta\left(\tau\right)\right) \Rightarrow \mathbb{Z}_\beta\left(\tau\right)$
- $\sqrt{n}\left(\hat{\phi} - \phi\right) \Rightarrow \mathbb{Z}_\phi$

$\mathbb{Z}_\beta\left(\tau\right)$ *is a zero-mean Gaussian process with covariance function* $\Sigma_\beta\left(\tau, \tau'\right)$, *and* $\mathbb{Z}_\phi$ *is a*

[12]For example, a polynomial of the propensity score. See Arellano and Bonhomme (2017a).

zero-mean normal distribution with covariance function Σ_ϕ, where

$$\Sigma_\vartheta(\tau,\tau') \equiv \begin{pmatrix} \Sigma_\beta(\tau,\tau') & \Sigma_{\beta\phi}(\tau')' \\ \Sigma_{\beta\phi}(\tau) & \Sigma_\phi \end{pmatrix} = H(\tau)\,\Sigma_R(\tau,\tau')\,H(\tau')'$$

$$\Sigma_R(\tau,\tau') \equiv \begin{pmatrix} \Sigma_{R_1}^{11} & 0 & \Sigma_{R_1}^{12}(\tau')' & 0 & 0 \\ 0 & \Sigma_{R_0}^{11} & 0 & \Sigma_{R_0}^{12}(\tau')' & 0 \\ \Sigma_{R_1}^{12}(\tau) & 0 & \Sigma_{R_1}^{22} & 0 & 0 \\ 0 & \Sigma_{R_0}^{12}(\tau) & 0 & \Sigma_{R_0}^{22} & 0 \\ 0 & 0 & 0 & 0 & \Sigma_R^{33} \end{pmatrix}$$

$$\Sigma_{R_d}^{11}(\tau,\tau') = \mathbb{E}\left[\mathbf{1}(D=d)\left(G_{d,X,\tau}\wedge G_{d',X,\tau'} - G_{d,X,\tau}G_{d',X,\tau'}\right)XX'\right]$$

$$\Sigma_{R_d}^{12}(\tau) = \mathbb{E}\left[\mathbf{1}(D=d)\int_0^1 X\varphi(u,Z)'\left[G_{d,X,\tau}\wedge G_{d',X,u} - G_{d,X,\tau}G_{d',X,u}\right]du\right]$$

$$\Sigma_{R_d}^{22}(\tau) = \mathbb{E}\left[\mathbf{1}(D=d)\int_\varepsilon^{1-\varepsilon}\int_\varepsilon^{1-\varepsilon}\varphi(u,Z)\,\varphi(v,Z)'\left[G_{d,X,u}\wedge G_{d',X,v} - G_{d,X,u}G_{d',X,v}\right]dvdu\right]$$

$$\Sigma_R^{33} = \mathbb{E}\left[s(D,Z;\gamma)\,s(D,Z;\gamma)'\right]$$

$$G_{d,X,\tau} \equiv G_{d,X}(\tau,\pi(Z);\theta_d)$$

$$H(\tau) = F^I(\tau)\left[C(\tau) + \left(I - \int_\varepsilon^{1-\varepsilon} D(u)\,F^I(u)\,du\right)\int_\varepsilon^{1-\varepsilon} D(u)\,F^I(u)\,C(u)\,du\right]$$

$$F^I(\tau) \equiv (I - F(\tau))^{-1} = I + F(\tau)$$

$$F(\tau) = \begin{bmatrix} 0 & 0 & J_{\beta_1}(\tau)^{-1}J_{\theta_1}(\tau) & 0 & J_{\beta_1}(\tau)^{-1}J_{\gamma_1}(\tau) \\ 0 & 0 & 0 & J_{\beta_0}(\tau)^{-1}J_{\theta_0}(\tau) & J_{\beta_0}(\tau)^{-1}J_{\gamma_0}(\tau) \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$D(\tau) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \left[\int_{\varepsilon}^{1-\varepsilon} \tilde{J}_{\theta_1}(u)\,du\right]^{-1} \tilde{J}_{\beta_1}(\tau) & 0 & 0 & 0 & \left[\int_{\varepsilon}^{1-\varepsilon} \tilde{J}_{\theta_1}(u)\,du\right]^{-1} \tilde{J}_{\gamma_1}(\tau) \\ 0 & \left[\int_{\varepsilon}^{1-\varepsilon} \tilde{J}_{\theta_0}(u)\,du\right]^{-1} \tilde{J}_{\beta_0}(\tau) & 0 & 0 & \left[\int_{\varepsilon}^{1-\varepsilon} \tilde{J}_{\theta_0}(u)\,du\right]^{-1} \tilde{J}_{\gamma_0}(\tau) \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$C(\tau) = \begin{bmatrix} -J_{\beta_1}(\tau)^{-1} & 0 & 0 & 0 & 0 \\ 0 & -J_{\beta_0}(\tau)^{-1} & 0 & 0 & 0 \\ 0 & 0 & \left[\int_{\varepsilon}^{1-\varepsilon} \tilde{J}_{\theta_1}(u)\,du\right]^{-1} & 0 & 0 \\ 0 & 0 & 0 & \left[\int_{\varepsilon}^{1-\varepsilon} \tilde{J}_{\theta_0}(u)\,du\right]^{-1} & 0 \\ 0 & 0 & 0 & 0 & -B^{-1} \end{bmatrix}$$

and $\wedge$ *denotes the minimum between two variables.*

The proof of Theorem 1 requires accounting for the estimation of three sets of parameters: γ, θ and β. The estimator of the propensity score parameters does not depend on the remaining ones, and by Assumption 8 its asymptotic distribution is well-behaved. However, the estimators of the copula and the slope parameters depend on each other. Consequently, the expansions of the moments used for the estimation depend on the process of all the parameters. This is captured by the matrix $H(\tau)$, which depends on the matrices defined in Assumptions 8 and 11. The latter reflect the impact of the estimator of the propensity score on the limiting variance of the remaining estimators, as well as the impact of the copula and slope parameters on the variance of each other.

An additional challenge is posed by the copula parameters, which depend on the entire quantile regression process. Hence, when taking the expansion of the moments for the copula and slope parameters, the resulting system of equations combines some functions indexed by τ at a specific value with integrals over all $\tau \in \mathcal{T}$. These equations conform a Fredholm integral equation of the second kind. Moreover, for the asymptotic variance to be finite, it requires that the matrices of derivatives of the moments have full rank uniformly over $\mathcal{T}$, so that they can be inverted, as stated in Assumption 11. Note however, that the estimation

of the SQF and the copula is done independently for each value of d, so the asymptotic distributions of the parameters specific to each group are only related because they both depend on the common parameters of the propensity score.

Remark 1. *The estimator has several desirable features: it imposes neither rank similarity nor additive separability of the unobservables, and it achieves the $\sqrt{n}$ convergence rate.*

Remark 2. *Although the SQF is linear in quantiles, the resulting MTE is not necessarily linear. Thus, the estimator displays a rich amount of heterogeneity across both the observed covariates and the propensity score, and it is tractable. From a policy perspective, this can allow to better identify which groups of individuals benefit the most from the treatment. On the other hand, the asymptotic distribution and the proof crucially relies on the linearity of the SQF. Hence, a more flexible specification would be associated with a different asymptotic distribution and proof strategy.*

Remark 3. *Both the estimator and the identification results can be extended to a multivalued treatment when the latter is ordered and the selection equation can be written as a function of a latent variable. See Appendix G for further details.*

Remark 4. *Estimated quantile coefficients may result in non-monotonic quantile curves. Whenever this is the case, one can follow Chernozhukov et al. (2010) to circumvent this problem. Regardless, the criterion function given in Equation 11 is valid even if the quantile curves are not monotonic.*

Remark 5. *The algorithm takes the parametric form of the propensity score and the copula as given. There exist tests for the propensity score (Sant'Anna and Song, 2019, 2020), which can be used to assess if it is correctly specified. To the best of my knowledge, the only specification tests for copulas of latent variables is the one proposed in Pereda-Fernández (2021) in a likelihood framework. A similar test in this framework could be based on the values of the criterion function of Equation 11. The properties of such a test are beyond the scope of this paper.*

3.2 Estimation of Functionals

Several functions of interest can be expressed in terms of the estimated structural parameters, including the QTE and the MTE.[13] Specifically, they are estimated by

$$\hat{\Delta}_{QTE}(\tau) = \hat{Q}_{Y|D=1}(\tau) - \hat{Q}_{Y|D=0}(\tau) \tag{12}$$

$$\hat{\Delta}_{MTE}(v) = \int_{\mathcal{X}} \left(\int_{\varepsilon}^{1-\varepsilon} x'\hat{\beta}_1(\tau)\, d\hat{C}_{1,x}(\tau|v) - \int_{\varepsilon}^{1-\varepsilon} x'\hat{\beta}_0(\tau)\, d\hat{C}_{0,x}(\tau|v) \right) d\hat{F}_X(x) \tag{13}$$

where $\hat{Q}_{Y|D}(\tau) = \inf\left\{y : \hat{F}_{Y|D}(y) \geq \tau\right\}$, $\hat{F}_{Y|D}(y) = \int_{\mathcal{X}} F_{Y|D,X}(y|x)\, dF_X(x)$, $\hat{F}_{Y|D,X}(y|x) = \varepsilon + \int_{\varepsilon}^{1-\varepsilon} \mathbf{1}\left(x'\hat{\beta}_D(\tau) \leq y\right) d\tau$, and $\hat{F}_X(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \leq x)$. Their asymptotic distribution are established in the following theorems:

Theorem 2. *Under Assumptions 1-11, the asymptotic distribution of $\hat{\Delta}_{QTE}(\tau)$ is given by $\sqrt{n}\left(\hat{\Delta}_{QTE}(\tau) - \Delta_{QTE}(\tau)\right) \Rightarrow \mathbb{Z}_{QTE}(\tau)$. $\mathbb{Z}_{QTE}(\tau)$ is a zero-mean Gaussian process defined in the proof.*

Theorem 3. *Under Assumptions 1-11, the asymptotic distribution of $\hat{\Delta}_{MTE}(v)$ is given by $\sqrt{n}\left(\hat{\Delta}_{MTE}(v) - \Delta^{\varepsilon}_{MTE}(v)\right) \Rightarrow \mathbb{Z}_{MTE}(v)$. $\mathbb{Z}_{MTE}(v)$ is a zero-mean Gaussian process defined in the proof.*

Remark 6. *Note that, because of the trimming constant ε, the estimated MTE converges to the following truncated version of the MTE: $\Delta^{\varepsilon}_{MTE}(x,v) = \int_{\varepsilon}^{1-\varepsilon} g(1,x,u)\, dC_{1,x}(u|v) - \int_{\varepsilon}^{1-\varepsilon} g(0,x,u)\, dC_{0,x}(u|v)$.*

3.3 Inference

The asymptotic variance of the RQR estimator depends on several density functions, making it cumbersome and impractical to estimate directly. In contrast, resampling methods are easy

[13] The estimation of other functions of interest, such as the ATE, the TT, and the TUT, is shown in Appendix F.

to implement. Therefore, I show the validity of the weighted bootstrap (Ma and Kosorok, 2005).

Assumption 12. *Let W_i be an iid sample of positive weights, such that $\mathbb{E}(W_i) = 1$, $Var(W_i) = \omega_0 > 0$ and is independent of $(Y_i, D_i, Z_i')'$ for $i = 1, ..., n$.*

The bootstrap is implemented by using these weights for the estimation of the propensity score, given by $\hat{\pi}(z_i) \equiv \pi(z_i, \hat{\gamma}^*)$, and of the slope and copula parameters by substituting Equations 10-11, given by

$$\hat{\beta}_d^*(\tau; t) \equiv \arg\min_{b \in \mathcal{B}} \sum_{i=1}^{n} w_i \mathbf{1}(d_i = d) \rho_{G_{d,x}(\tau, \hat{\pi}(z_i); t)}(y_i - x_i' b) \tag{14}$$

$$\hat{\theta}_d^* \equiv \arg\min_{t \in \Theta} \left\| \sum_{i=1}^{n} w_i \int_{\varepsilon}^{1-\varepsilon} (d_i = d) \varphi(\tau, z_i) \left[\mathbf{1}\left(y_i \leq x_i' \hat{\beta}_d(\tau; t)\right) - G_{d,x}(\tau, \hat{\pi}(z_i); t) \right] d\tau \right\| \tag{15}$$

Note that the weight for each individual is the same in every step. Combine all the bootstrap parameter estimators into $\hat{\vartheta}_d^*(\tau)$, and denote the bootstrap functional estimators by $\hat{\Delta}_{QTE}^*(\tau)$, and $\hat{\Delta}_{MTE}^*(v)$. The following theorem establishes their asymptotic validity:

Theorem 4. *Under Assumptions 1-12, the weighted bootstrap estimators $\hat{\vartheta}^*(\tau)$, $\hat{\Delta}_{QTE}^*(\tau)$, and $\hat{\Delta}_{MTE}^*(v)$ consistently estimate the limiting laws of $\hat{\vartheta}(\tau)$, $\hat{\Delta}_{QTE}(\tau)$, and $\hat{\Delta}_{MTE}(v)$. Moreover,*

$$\sqrt{\frac{n}{\omega_0}} \left(\hat{\vartheta}^*(\tau) - \hat{\vartheta}(\tau) \right) \Rightarrow \mathbb{Z}_{\vartheta}(\tau)$$
$$\sqrt{\frac{n}{\omega_0}} \left(\hat{\Delta}_{QTE}^*(\tau) - \hat{\Delta}_{QTE}(\tau) \right) \Rightarrow \mathbb{Z}_{QTE}(\tau)$$
$$\sqrt{\frac{n}{\omega_0}} \left(\hat{\Delta}_{MTE}^*(v) - \hat{\Delta}_{MTE}(v) \right) \Rightarrow \mathbb{Z}_{MTE}(v)$$

It is possible to carry out uniform inference by applying the Kolmogorov-Smirnov test statistic to the QTE process. Specifically, let the null hypothesis be given by $H_0 : \Delta_{QTE}(\tau) =$

$\tilde{\Delta}_{QTE}(\tau)$, with alternative $H_1 : \Delta_{QTE}(\tau) \neq \tilde{\Delta}_{QTE}(\tau)$. The test statistic would be given by

$$KS_n \equiv \sup_{\tau \in \mathcal{T}} \sqrt{n} \hat{\Sigma}_{QTE}(\tau)^{-1/2} \left| \hat{\Delta}_{QTE}(\tau) - \tilde{\Delta}_{QTE}(\tau) \right| \tag{16}$$

Equation 16 requires an estimate of the asymptotic variance of the QTE estimator, $\hat{\Sigma}_{QTE}(\tau)$. This, along with the critical value can be obtained with the following algorithm:[14]

1. For each repetition $t = 1, ..., T$, compute $\hat{b}^*_{QTE,t}(\tau) \equiv \sqrt{n}\left(\hat{\Delta}^*_{QTE}(\tau) - \hat{\Delta}_{QTE}(\tau)\right)$, for $\tau \in \mathcal{T}$.

2. Estimate the covariance of the QTE process by $\hat{\Sigma}_{QTE}(\tau) = \frac{q_{0.75}(\tau) - q_{0.25}(\tau)}{z_{0.75} - z_{0.25}}$, where z_p is the p-th quantile of the standard normal distribution, and $q_p(\tau)$ is the p-th quantile of the distribution of $\hat{b}^*_{QTE,t}(\tau)$, for $t = 1, ..., T$.

3. Compute $KS_{n,t} \equiv \sup_{\tau \in \mathcal{T}} \hat{\Sigma}_{QTE}(\tau)^{-1/2} \left| \hat{b}^*_{QTE,t} \right|$, for $t = 1, ..., T$.

4. The critical value is obtained as the 1-α-th quantile of $KS_{n,t}$ across $t = 1, ..., T$, *i.e.*, $c_{1-\alpha} = \inf\left\{c : \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}(KS_{n,t} \leq c) \geq 1 - \alpha\right\}$.

The following Corollary establishes the validity of the test:

Corollary 1. *Under* H_0, $KS_n \xrightarrow{d} KS \equiv \sup_{\tau \in \mathcal{T}} \sqrt{n}\Sigma_{QTE}(\tau)^{-1/2} \left|\Delta_{QTE}(\tau) - \Delta^*_{QTE}(\tau)\right|$. *If* $\nu(\cdot) \equiv \Sigma_{QTE}(\cdot)^{-1/2} \left|\Delta_{QTE}(\cdot) - \Delta^*_{QTE}(\cdot)\right|$ *has a nondegenerate covariance kernel, then for any* $\alpha \leq 1/2$, $\mathbb{P}(KS_n > c_{1-\alpha}) \xrightarrow{d} \mathbb{P}(KS > c_{1-\alpha}) = \alpha$. *Moreover, under* H_1 $KS_n \xrightarrow{d} \infty$ *and* $\mathbb{P}(KS_n > c_{1-\alpha}) \to 1$.

4 Empirical Application

I apply the methodology presented in Section 3 to the estimation of the effect of Work First Job Placements on the distribution of future earnings. This was a welfare-to-work program

[14] This algorithm is based on Algorithm 3 in Chernozhukov et al. (2013). The resulting estimator of the covariance matrix of the QTE process does not require additional conditions for its validity, unlike the bootstrap estimates of the covariance matrix. See Kato (2011) for further details.

in Detroit that consisted in quickly finding an employment for low-skilled workers, with the aim of improving their future earnings. Following a week-long orientation period, workers were randomly assigned to a contractor, whose role was to help them find a job during the following weeks. Successful workers found either a direct-hire placement (DHP) or a temporary-help placement (THP), each of which could have a potentially different effect on future earnings. On the other hand, some workers obtained no placement (NP) at all. The latter constitute the control group, whereas the former are the two treatment groups. Overall, the number of individuals in the DHP, THP, and NP groups amounted to 11,583, 2,762 and 16,177, respectively.

This dataset was originally studied by Autor and Houseman (2010), who proposed to use contractor assignments as an instrument: since placement practices vary by contractor, the assignment of each contractor would lead to a different probability of obtaining a DHP or a THP. In their paper, they explain how they construct a variable that uses variation across contractors within periods and districts, which they use to estimate the effects of each type of placement on future earnings.

Autor and Houseman (2010) found a positive and significant mean effect of DHP on earnings during the following 7 quarters, whereas the mean effect of THP was negative, though not significant. Subsequently, Autor et al. (2017) studied the distributional effects using IVQR, finding a substantial amount of heterogeneity of the effects. In particular, the effect on the upper tail was substantially large and positive for DHP, while it was negative and significant for THP, and small and not significant on the lower tail of both earnings distributions.

Autor et al. (2017) highlighted the difficulty of translating the estimates of the conditional (on X) distribution of earnings into the unconditional distribution. From a policy perspective, the latter may be more relevant, so I report the estimates of the unconditional quantile function for the whole population under each treatment status following Chernozhukov et al. (2013). Also, I present the estimates of the MTE and, to assess the role of rank similarity in shaping the results, I propose a decomposition for the mean difference between two treatment

status, and another for the MTE. Additional tables and results are reported in Appendix I.[15]

Regarding the parametric forms of the different estimators, I estimate the propensity score with ordered multinomial logit and the copulas used for the RQR estimators are the Gaussian and the Bernstein of orders 2 through 6.[16] Among the latter, I present the estimates from the model selected using 5-fold cross validation.[17] To assess the sensitivity of the results to the rank similarity assumption, I estimate a constrained RQR model in which the copula is the same for all three groups. Additionally, I include all interactions between the treatment status and the covariates, unlike Autor et al. (2017). All uniform confidence intervals were computed using the weighted bootstrap (Ma and Kosorok, 2005) with 500 repetitions.

4.1 Propensity Score and Unconditional Earnings Distributions

The ordered logit specification can be tested, as proposed in Sant'Anna and Song (2020). Because of the large sample size, I could not implement their main test, but the second variant of the test they proposed in Sant'Anna and Song (2019). The value of the test statistic equals 0.0104, whereas the critical value for the 5% size test is 0.0437. Hence, the test fails to reject the null hypothesis of validity of the ordered logit specification.

Figure 3 reports the distribution of the propensity score to be into each treatment status for program participants in each treatment group. The three histograms reveal a substantial overlap among individuals in the three groups. However, the central histograms indicate that there is little amount of variability in the propensity to receive a THP. Consequently, the estimates for this group require a large degree of extrapolation with any estimation method, making them less reliable than the estimates for the other two groups.

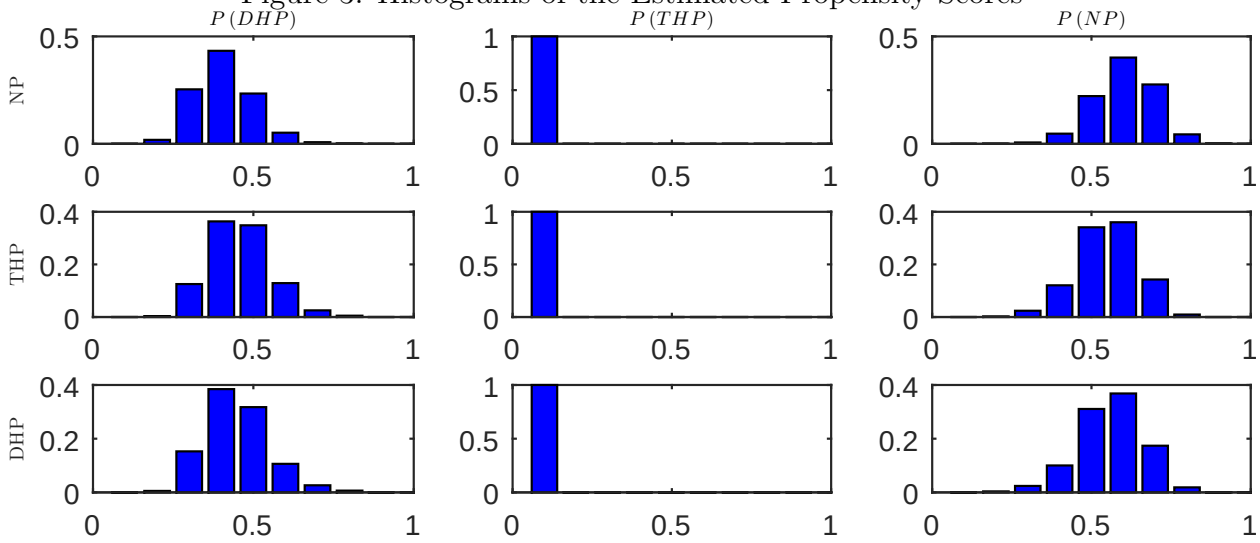
Figure 4 compares the baseline estimates of the quantile function of future earnings with the observed empirical distribution. There are two relevant findings in this figure. First, the

[15] The results of the coefficients not reported in the paper are available upon request.

[16] As stated in the text, the Gaussian copula satisfies Assumption 5. The same cannot be said of Bernstein copulas of order equal or higher than 3, as they depend on several parameters. This copula would be valid, however, under the alternative Assumption 13 considered in Appendix D.

[17] The orders of the Bernstein copulas selected through cross validation were 3, 2 and 5 for the DHP, THP and NP groups, respectively. See Table 10 in Appendix I for the cross-validated objective function for each estimator, including the Gaussian.

Figure 3: Histograms of the Estimated Propensity Scores



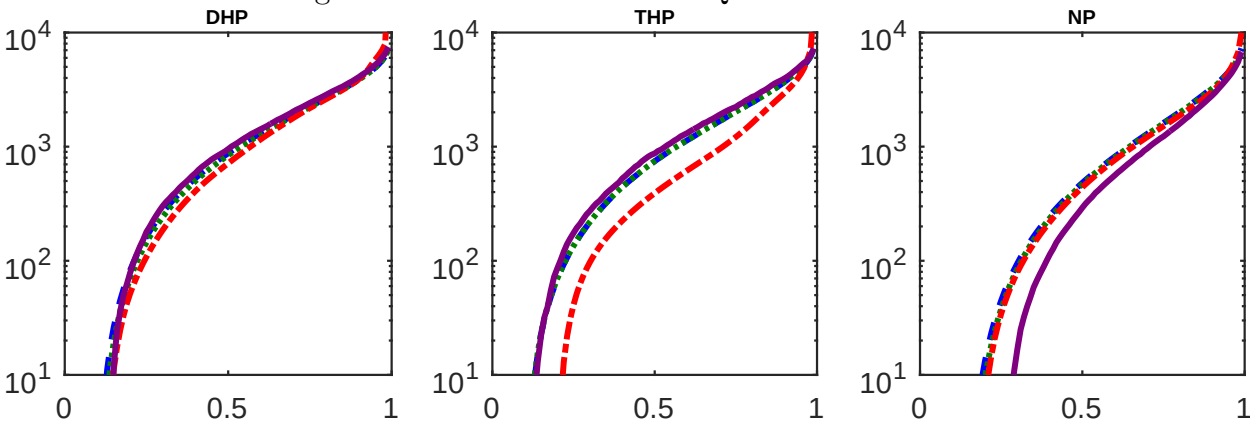
Notes: histograms with the distribution of the propensity to be in the DHP group (left panel), in the THP group (central panel), and in the NP group (right panel) for individuals in the NP group (top row), in the THP group (central row), and in the DHP group (bottom row); interval width=0.1.

observed distribution and the one estimated with RQR largely coincide for the two treatment groups. Second, the observed distribution for the NP group lies above the estimated potential distribution, regardless of the copula. These two findings suggest that the rank similarity assumption is unlikely to hold. In fact, the value of the test statistic proposed by Frandsen and Lefgren (2018) using IVQR equals 213.3, whereas the critical value for a test size of 5% is 67.5.[18] Hence, the test strongly rejects the hypothesis of rank similarity.

Even though the empirical cdf is a step function by construction, there are no noticeable jumps, as required by Assumption 3. Moreover, it makes Assumption 5 more likely to be satisfied, although it is not possible to verify. Moreover, the definition of the instrument implies that it takes values over a large range of points, making Assumption 4 plausible. On the other hand, rather than working with Assumption 1, I also take the variation in X as exogenous, so that the copulas used in the estimation do not depend on the covariates, only

[18]The IVQR estimates in this paper were obtained using Smoothed Estimating Equations (Kaplan and Sun, 2017) rather than the more common Inverse Quantile Regression (IQR; Chernozhukov and Hansen, 2006). The former is convenient from a computational standpoint, particularly to obtain standard errors of functionals based on the IVQR estimator using the bootstrap. The test statistic using the IQR implementation equals 218.8. Additionally, the instrument used for the estimation is the same used in Autor et al. (2017). The estimates using IQR are available upon request.

Figure 4: Estimated Potential Quantile Functions



Notes: in each panel, the dashed blue line represents the quantile function of the RQR estimator with the Gaussian copula, the dotted green line represents the quantile function of the RQR estimator with the Bernstein copula, the dashed-dotted red line represents the quantile function of the IVQR estimator, and the solid purple line represents the empirical distribution. The scale of the Y axis is logarithmic.

on the treatment status.

4.2 Copula Estimates

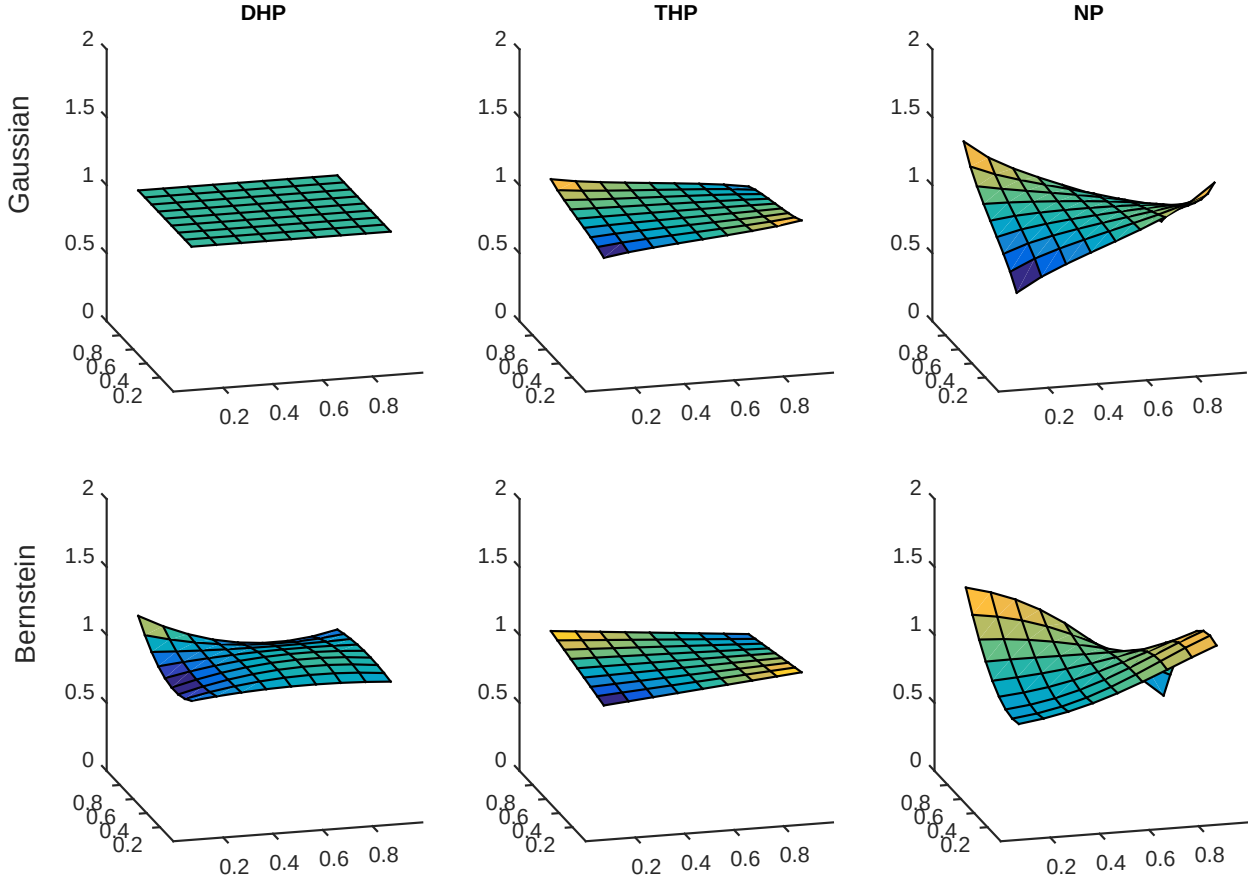
Table 1 reports the Kendall's τ correlation coefficient of the estimated copulas to give a comparable measure of the degree of correlation. These numbers provide additional evidence against the rank similarity assumption. In particular, they reflect a tiny amount of correlation between the unobservables of the selection equation and the rank of the SQF for treated individuals, and a moderate degree of correlation for those in the NP group. The latter is negative, as it can be seen in Figure 5. Hence, those more likely to be treated (low values of v) would have ranked relatively low in the distribution of potential outcomes if nobody had been treated (low values of u_0). Analogously, those less likely to be treated would have ranked higher in the distribution of potential outcomes of the NP. On the other hand, if everybody had been treated, their earnings would have been largely uncorrelated with their propensity to be treated. The results for the Bernstein copula are similar to those found for the Gaussian copula, although there is a slight increase in the amount of correlation for the two treatment groups as the order increases.

Table 1: Kendall's τ Statistic of the Estimated Copulas

Copula	Gau	Con	Ber(2)	Ber(3)	Ber(4)	Ber(5)	Ber(6)
DHP	0.00	-0.03	0.00	-0.02	-0.04	-0.03	-0.03
THP	-0.03	-0.03	-0.02	-0.04	-0.01	-0.03	-0.03
NP	-0.13	-0.03	-0.12	-0.11	-0.11	-0.12	-0.12

Notes: Gau, Con, and Ber(X) respectively stand for Gaussian copula, Gaussian copula constrained to be the same for all three groups, and Bernstein copula of order X.

Figure 5: Estimated Copula Densities

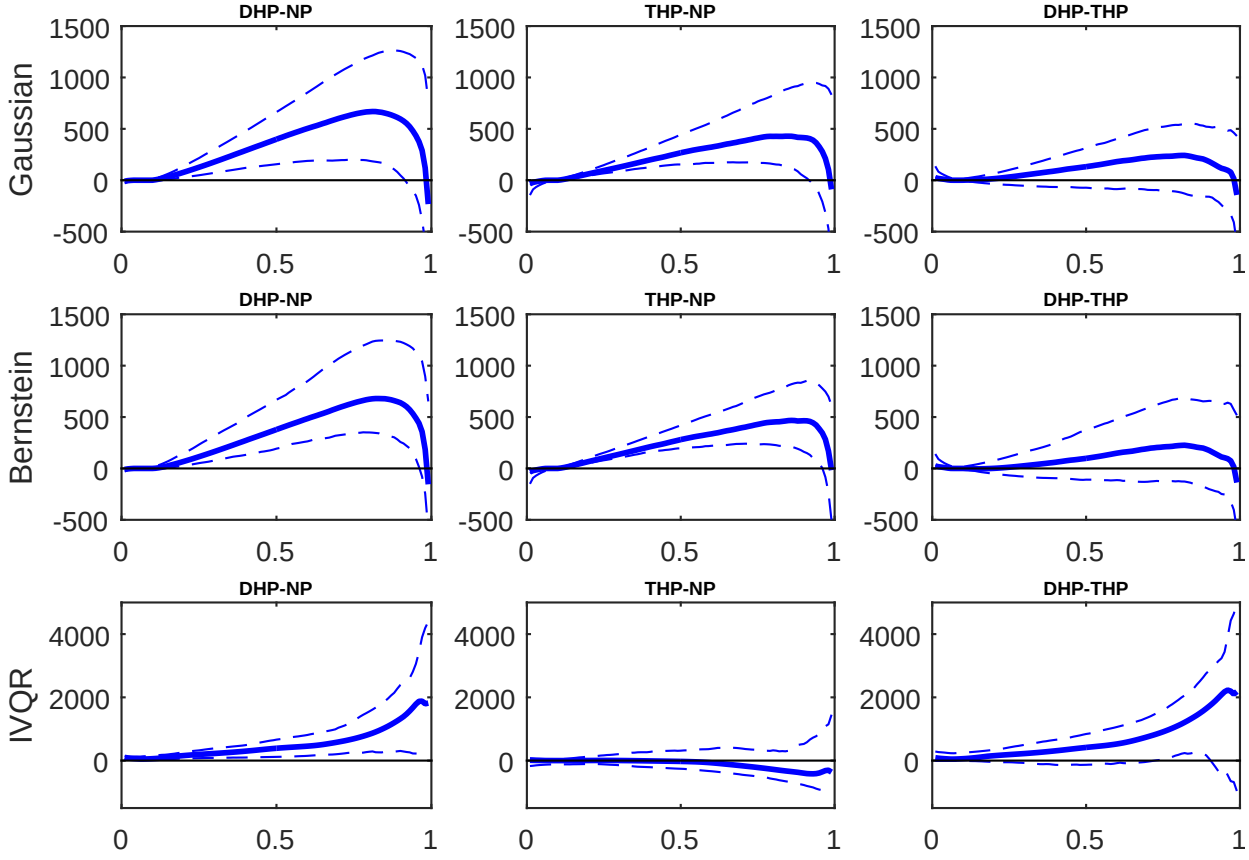


Notes: the first row shows the density of the Gaussian copula between U_D and V for $D = \{DHP, THP, NP\}$; the second row shows the density of the selected Bernstein copula.

4.3 Unconditional Quantile Treatment Effect Estimates

The estimates of the unconditional QTE based on RQR (Figure 6) indicate that receiving any kind of treatment versus not being treated increases future earnings at most quantiles of the distribution, with the only exception of those close to the extremes, for which the effect is negligible and not significant. Moreover, both QTE have an increasing profile for most of the distribution, peaking around the 80th percentile and rapidly decreasing thereafter. The largest gain comes for the DHP, whereas the gain for THP is substantially smaller at all quantiles, especially at the top of the distribution. On average, these gains amount to about \$339 and \$224 with the Gaussian copula, and \$338 and \$243 with the Bernstein copula.

Figure 6: Quantile Treatment Effect Estimates



Notes: in each panel, the solid line represents the MTE estimate, and the dashed lines represent the bootstrapped 95% uniform confidence intervals.

Relative to the findings in Autor et al. (2017), the estimated unconditional QTE for DHP is larger for the lower and central parts of the distribution, and smaller for the upper part.

Indeed, the estimates based on IVQR (Figure 6) predict that the gains for the right tail would be large and increasing, whereas the estimates based on RQR indicate that the gain would be small and decreasing. However, because of the lower level of accuracy at the upper quantiles, the 95% confidence intervals of the different estimators overlap. Hence, we cannot conclude that they are different from each other.

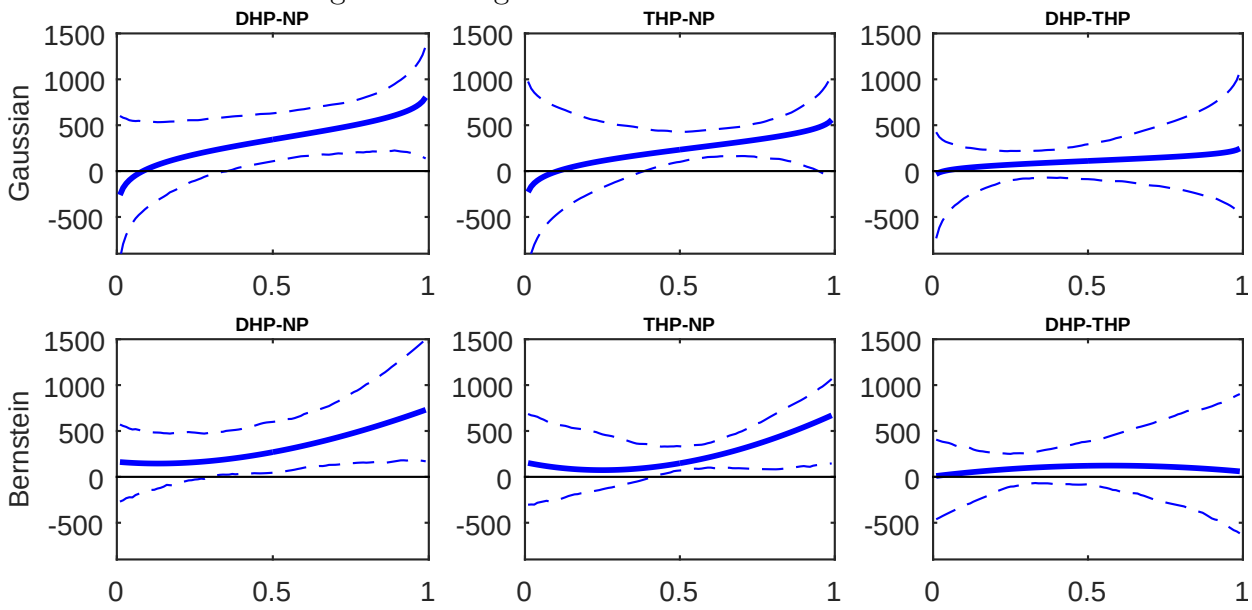
Similarly to the DHP, the estimates for THP are positive at almost every quantile, although smaller in magnitude. In contrast, the estimates based on IVQR are negative at most quantiles. Consequently, the difference between DHP and THP shows a gain for DHP with both estimators. However, while the estimator based on RQR yields an average difference close to \$100, for the estimator based on IVQR the average difference is over \$600.

4.4 Marginal Treatment Effect Estimates

The MTE estimates also display a large amount of heterogeneity (Figure 7). They have an increasing shape, meaning that those more likely to be treated (those with a small value of v) have the lowest expected gain from the treatment. This follows from the more negative correlation of the copula of (U_0, V) and the gain from the treatment. The MTE estimates with the Gaussian copula are negative for roughly $v \leq 0.1$, although they are not significantly different from zero. Thus, if all had been treated, there would have been a minority of workers with smaller future earnings, even though the distributions of potential outcomes of both DHP and THP dominate the distribution of NP.[19] However, because the density of the propensity scores around the first decile is so small, this result could be largely driven by the functional form of the copula.

[19]Interpreting this result through the lenses of the generalized Roy model with imperfect information leads to the conclusion that the expected cost of being treated is increasing in V: for Equation 2 to hold, the net surplus needs to be decreasing in V, and because the MTE is decreasing in V, the opposite must hold for the expected cost of being treated.

Figure 7: Marginal Treatment Effect Estimates



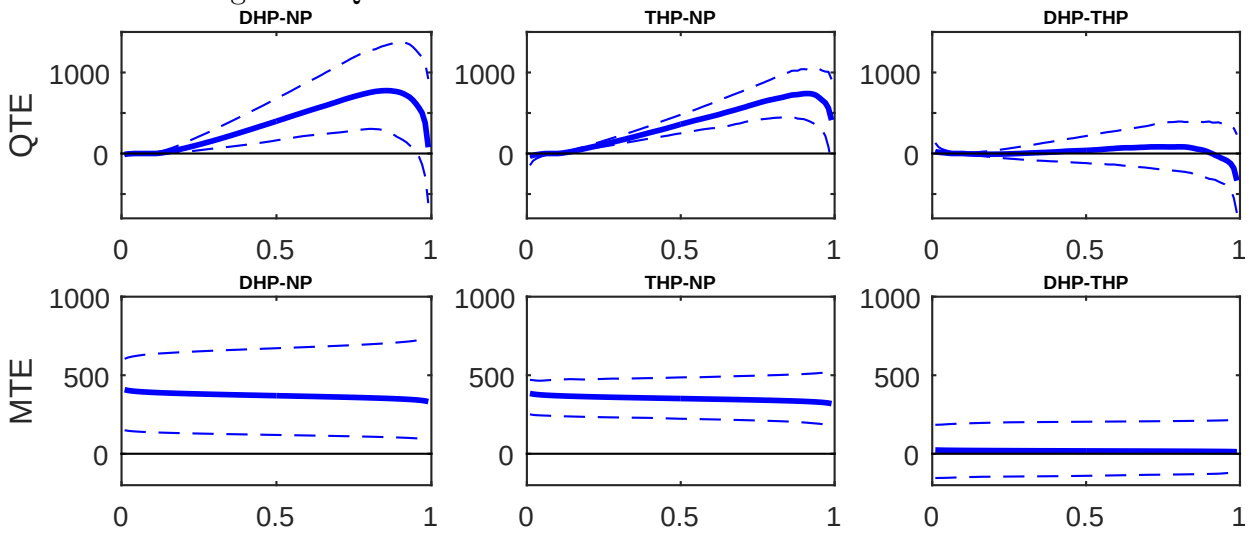
Notes: in each panel, the solid line represents the MTE estimate, and the dashed lines represent the bootstrapped 95% uniform confidence intervals.

4.5 Estimates under Rank Similarity

The results change substantially if the estimator is forced to satisfy the rank similarity hypothesis. First, the estimated amount of correlation lies in between those of the DHP and THP groups on the one hand, and that of the NP group (Table 1). Second, the estimated unconditional QTE for the DHP group relative to the NP group is slightly larger, but for the THP group it is much larger (Figure 8). Consequently, the QTE for the DHP group relative to the THP group becomes a small fraction of the unrestricted estimate. Because these results go in the opposite direction of those found with IVQR, they suggest that the difference between the RQR and IVQR estimates could also be attributed to the inclusion of interaction terms between the treatment and the covariates for the RQR estimators.

On the other hand, the constrained estimates of the MTE are radically different: they are downward sloping for both treatment status. Hence, those more likely to be treated would, on average, gain more from the treatment. However, the slope is almost flat, so these differences are actually small. Finally, the average (across v) MTE would be larger for both DHP and THP in the constrained model.

Figure 8: QTE and MTE Estimates for the Restricted Model



Notes: in each panel, the solid line represents the QTE/MTE of the restricted RQR estimator, the dashed lines represent the bootstrapped 95% uniform confidence intervals.

4.6 Decompositions

These estimates suggest that the differences in self-selection into each treatment status can explain a substantial amount of the difference between the treatment groups and the control group. To assess this possibility, define the following counterfactual mean outcome:

$$\mathbb{E}\left[Y^{jhk}\right] \equiv \int_{\mathcal{Z}} \int_{0}^{1} g_j\left(x,u\right) dG_{h,x}\left(u,\pi\left(z\right)\right) dF_Z^{(k)}\left(z\right)$$

where $j=0,1$ refers to the treatment group of the SQF, $h=0,1$ to the treatment group of the copula, and $k=0,1$ to the treatment group of the distribution of the observables $F_Z^{(k)}$. The difference between these two can be decomposed as follows:

$$\mathbb{E}\left[Y|D=1\right]-\mathbb{E}\left[Y|D=0\right]=\underbrace{\mathbb{E}\left[Y^{111}\right]-\mathbb{E}\left[Y^{110}\right]}_{\text{endowments effect}}+\underbrace{\mathbb{E}\left[Y^{110}\right]-\mathbb{E}\left[Y^{100}\right]}_{\text{self-selection effect}}+\underbrace{\mathbb{E}\left[Y^{100}\right]-\mathbb{E}\left[Y^{000}\right]}_{\text{coefficients effect}}$$

The first and third components are those present in the Oaxaca-Blinder decomposition. The second one, which I refer to as the self-selection effect, captures differences in the copula between the treated and the untreated. Under exogeneity, the average value of the rank

equals 0.5 for both treatment groups, so the self-selection term vanishes. If there is (positive) selection into treatment but the copulas are the same (rank invariance or similarity), then the average rank is higher than 0.5, but it is still the same for both groups, so the self-selection term is also equal to 0. In contrast, with rank dissimilarity, the average rank is different for the treated and the untreated, resulting in a non-zero self-selection effect.

Table 2 reports the size of each term of the decomposition, confirming the importance of self-selection: it explains roughly 40% of the difference between the mean earnings of those in the DHP group and the NP group, and slightly more when one compares the earnings of the THP and NP groups. In contrast, this difference vanishes when one looks at the mean difference between the DHP and THP groups, which is almost entirely explained by the coefficients effects. The endowments effect is negligible in all cases because the sample is very homogeneous with respect to the covariates.

Table 2: Means decomposition

	Gaussian copula			Bernstein copula		
	DHP, NP	THP, NP	DHP, THP	DHP, NP	THP, NP	DHP, THP
Total	486.1	383.6	102.5	491.4	388.0	103.4
effect	(20.9)	(30.4)	(32.2)	(21.5)	(29.5)	(32.4)
Endowments	0.0	-7.9	0.0	1.6	-7.1	-0.5
effect	(5.9)	(18.0)	(1.4)	(6.8)	(18.6)	(1.7)
Self-selection	186.6	188.2	-6.8	192.9	175.7	13.2
effect	(102.8)	(68.7)	(80.3)	(103.4)	(62.1)	(93.0)
Coefficients	299.5	203.3	109.3	297.0	219.4	90.7
effect	(112.7)	(67.0)	(89.7)	(109.3)	(61.3)	(98.3)

Notes: bootstrapped standard errors in parenthesis.

A similar decomposition can be applied to the MTE under rank dissimilarity:

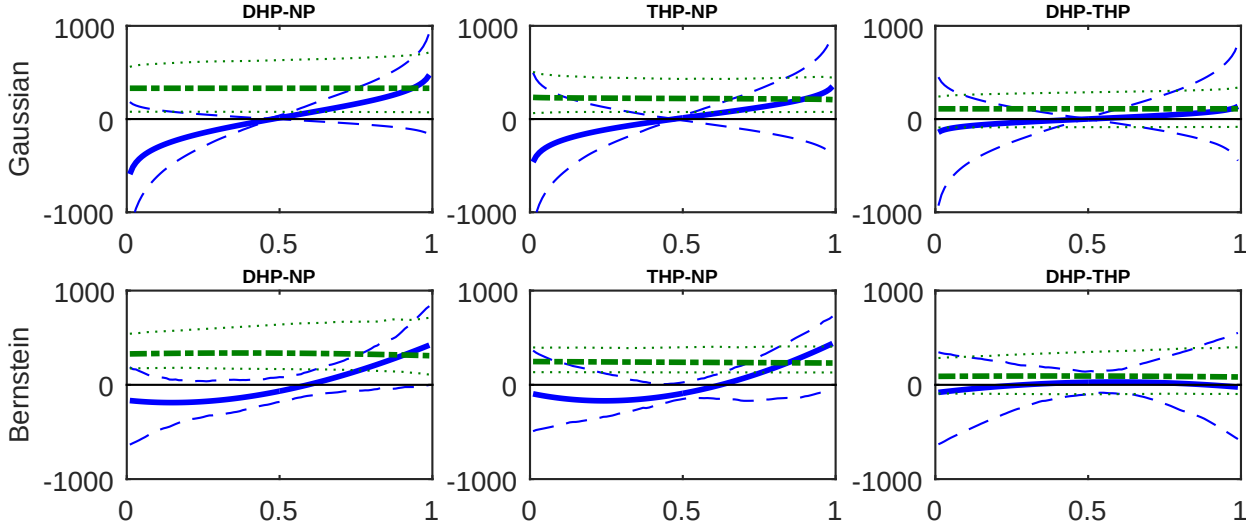
$$\Delta_{MTE}(x,v) = \underbrace{\int_0^1 \left[g_1(x,u) - g_0(x,u)\right] dC_{1,x}(u|v)}_{\equiv \Delta_{RIMTE}(x,v)} + \underbrace{\int_0^1 g_0(x,u)\, d\left[C_{1,x}(u|v) - C_{0,x}(u|v)\right]}_{\equiv \Delta_{ESME}(x,v)}$$

Denote the first term as the rank invariant marginal treatment effect (RIMTE), *i.e.*, the expected gain for the marginal individual, if the amount of self-selection was the same under each treatment status. This effect depends on the difference between the two SQF

weighted by the copula. The second term, denoted as the excess selection marginal effect (ESME), reflects rank dissimilarity, *i.e.*, the difference in the amount of selection between the two treatment status.[20] Hence, even if the SQF was the same for treated and untreated individuals, the MTE would be positive because the marginal individual would, on average, have a higher value of the rank of the SQF when treated, *i.e.*, they are more positively selected. Conversely, the ESME vanishes under either rank invariance or similarity.

The estimates shown in Figure 9 show that the shape of the MTE is determined almost entirely by the ESME, which has a positive and steep slope. In contrast, the RIMTE has the usual decreasing shape (in v), although it is almost flat. This reconciles the estimates of the constrained copula with the unconstrained ones. Moreover, the difference of the RIMTE between the estimators based on the Gaussian and Bernstein copulas is almost indistinguishable. This reinforces the claim that the difference between the MTE of both estimators is due to differences in the extrapolation of the functional form of the copula.

Figure 9: Decomposition of Marginal Treatment Effect Estimates



Notes: in each panel, the solid blue line represents the estimated ESME, the dashed blue line represent the bootstrapped 95% uniform confidence intervals, the dashed-dotted green line represents the estimated RIMTE, and the dotted green lines represent the bootstrapped 95% uniform confidence intervals.

[20]Note that there is an alternative decomposition of the MTE: $\int_0^1 \left[g_1\left(x,u\right)-g_0\left(x,u\right)\right] dC_{0,x}\left(u|v\right) + \int_0^1 g_1\left(x,u_0\right) d\left[C_{1,x}\left(u|v\right)-C_{0,x}\left(u|v\right)\right]$.

5 Conclusion

In this paper I study the identification of a nonseparable triangular model with a binary endogenous treatment. Nonparametric identification is achieved by using local variation of the instrument combined with a shape restriction on the distribution of the unobservables. The latter is modeled with copulas, explicitly allowing for rank dissimilarity. I show how it can capture differences in the mean outcome between the treated and the untreated, and how the shape of the MTE is influenced by it.

The proposed estimator is a multi-step quantile regression estimator. It estimates the SQF, the copula of the unobservables and the propensity score. I show its asymptotic distribution and how to conduct uniform inference with the exchangeable bootstrap.

Finally, the estimation methods presented are applied to the Work First Job Placements data. In contrast with previous findings, the estimates reveal that both types of placements had a positive effect on the distribution of earnings, particularly on the upper half of the distribution. Moreover, I find evidence that the rank similarity assumption was not satisfied in the data. The difference in the amount of self-selection for each treatment status was responsible for a substantial amount of the difference in outcomes between the treated and the untreated. It also affected the shape of the MTE, and it identified a share of the population whose earnings would have been higher if they had not received the treatment.

References

Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *Journal of econometrics 113*(2), 231–263.

Abadie, A., J. Angrist, and G. Imbens (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica 70*(1), 91–117.

Abrevaya, J. and H. Xu (2021). Estimation of treatment effects under endogeneous heteroskedasticity. *Journal of Econometrics Forthcoming*, 1–28.

Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006). Quantile regression under misspecification, with an application to the us wage structure. *Econometrica 74*(2), 539–563.

Angrist, J. and I. Fernandez-Val (2013). Extrapolate-ing: External validity and. In *Advances in Economics and Econometrics: Volume 3, Econometrics: Tenth World Congress*, Volume 51, pp. 401. Cambridge University Press.

Arellano, M. and S. Bonhomme (2017a). Quantile selection models. *Econometrica 85*(1), 1–28.

Arellano, M. and S. Bonhomme (2017b). Sample selection in quantile regression: a survey. In *Handbook of Quantile Regression*, pp. 209–224. Chapman and Hall/CRC.

Autor, D. H. and S. N. Houseman (2010). Do temporary-help jobs improve labor market outcomes for low-skilled workers? evidence from" work first". *American Economic Journal: Applied Economics 2*(3), 96–128.

Autor, D. H., S. N. Houseman, and S. P. Kerr (2017). The effect of work first job placements on the distribution of earnings: An instrumental variable quantile regression approach. *Journal of Labor Economics 35*(1), 149–190.

Bernstein, S. (1912). Démonstration du théoreme de weierstrass fondée sur le calcul des probabilités. *Comm. Soc. Math. Kharkov 13*, 1–2.

Bisbee, J., R. Dehejia, C. Pop-Eleches, and C. Samii (2017). Local instruments, global extrapolation: External validity of the labor supply–fertility local average treatment effect. *Journal of Labor Economics 35*(S1), S99–S147.

Björklund, A. and R. Moffitt (1987). The estimation of wage gains and welfare gains in self-selection models. *The Review of Economics and Statistics 69*(1), 42–49.

Brinch, C. N., M. Mogstad, and M. Wiswall (2017). Beyond late with a discrete instrument. *Journal of Political Economy 125*(4), 985–1039.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011). Estimating marginal returns to education. *American Economic Review 101*(6), 2754–5781.

Carneiro, P. and S. Lee (2009). Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality. *Journal of Econometrics 149*(2), 191–208.

Chen, X., Z. Huang, and Y. Yi (2021). Efficient estimation of multivariate semi-nonparametric garch filtered copula models. *Journal of Econometrics 222*(1), 484–501.

Chen, X., Z. Xiao, and B. Wang (2021). Copula-based time series with filtered nonstationarity. *Journal of Econometrics*.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010). Quantile and probability curves without crossing. *Econometrica 78*(3), 1093–1125.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2013). Inference on counterfactual distributions. *Econometrica 81*(6), 2205–2268.

Chernozhukov, V. and C. Hansen (2005). An iv model of quantile treatment effects. *Econometrica 73*(1), 245–261.

Chernozhukov, V. and C. Hansen (2006). Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics 132*(2), 491–525.

Chesher, A. (2003). Identification in nonseparable models. *Econometrica 71*(5), 1405–1441.

Chesher, A. (2005). Nonparametric identification under discrete variation. *Econometrica 73*(5), 1525–1550.

Cornelissen, T., C. Dustmann, U. Schönberg, and A. Raute (2017). Who benefits from universal child care?: Estimating marginal returns to early child care attendance. *Journal of Political Economy 41*, 47–60.

Das, M. (2005). Instrumental variables estimators of nonparametric models with discrete endogenous regressors. *Journal of Econometrics 124*(2), 335–361.

D'Haultfœuille, X. and P. Février (2015). Identification of nonseparable triangular models with discrete instruments. *Econometrica 83*(3), 1199–1210.

Eisenhauer, P., J. J. Heckman, and E. Vytlacil (2015). The generalized roy model and the cost-benefit analysis of social programs. *Journal of Political Economy 123*(2), 413–443.

Fan, Y. and S. S. Park (2010). Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory 26*(3), 931–951.

Feng, J. (2019). Identification and estimation of nonseparable models with multivalued endogeneity and a binary instrument. Technical report, arXiv preprint arXiv:1904.01159.

Firpo, S. and G. Ridder (2019). Partial identification of the treatment effect distribution and its functionals. *Journal of Econometrics 213*(1), 210–234.

Fox, J. T., K. il Kim, S. P. Ryan, and P. Bajari (2012). The random coefficients logit model is identified. *Journal of Econometrics 166*(2), 204–212.

Frandsen, B. R., M. Frölich, and B. Melly (2012). Quantile treatment effects in the regression discontinuity design. *Journal of Econometrics 168*(2), 382–395.

Frandsen, B. R. and L. J. Lefgren (2018). Testing rank similarity. *Review of Economics and Statistics 100*(1), 86–91.

Frölich, M. and B. Melly (2013). Unconditional quantile treatment effects under endogeneity. *Journal of Business & Economic Statistics 31*(3), 346–357.

Hahn, J. and G. Ridder (2011). Conditional moment restrictions and triangular simultaneous equations. *Review of Economics and Statistics 93*(2), 683–689.

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

Heckman, J. (1990). Varieties of selection bias. *The American Economic Review 80*(2), 313.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement 5*(4), 475–492.

Heckman, J. J. and R. Pinto (2018). Unordered monotonicity. *Econometrica 86*(1), 1–35.

Heckman, J. J., J. Smith, and N. Clements (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies 64*(4), 487–535.

Heckman, J. J. and E. 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.

Heckman, J. J. and E. Vytlacil (2001). Policy-relevant treatment effects. *American Economic Review 91*(2), 107–111.

Heckman, J. J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica 73*(3), 669–738.

Heckman, J. J. and E. J. Vytlacil (2007a). Econometric evaluation of social programs, part i: Causal models, structural models and econometric policy evaluation. *Handbook of econometrics 6*, 4779–4874.

Heckman, J. J. and E. J. Vytlacil (2007b). Econometric evaluation of social programs, part ii: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. *Handbook of econometrics 6*, 4875–5143.

Hoderlein, S., H. Holzmann, and A. Meister (2017). The triangular model with random coefficients. *Journal of Econometrics 201*(1), 144 – 169.

Horowitz, J. L. and S. Lee (2007). Nonparametric instrumental variables estimation of a quantile regression model. *Econometrica 75*(4), 1191–1208.

Horowitz, J. L., E. Mammen, et al. (2004). Nonparametric estimation of an additive model with a link function. *The Annals of Statistics 32*(6), 2412–2443.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–75.

Imbens, G. W. and W. K. Newey (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica 77*(5), 1481–1512.

Jun, S. J. (2009). Local structural quantile effects in a model with a nonseparable control variable. *Journal of Econometrics 151*(1), 82–97.

Jun, S. J., J. Pinkse, and H. Xu (2011). Tighter bounds in triangular systems. *Journal of Econometrics 161*(2), 122–128.

Jun, S. J., J. Pinkse, and H. Xu (2016). Estimating a nonparametric triangular model with binary endogenous regressors. *The Econometrics Journal 19*(2), 113–149.

Kaplan, D. M. and Y. Sun (2017). Smoothed estimating equations for instrumental variables quantile regression. *Econometric Theory 33*(1), 105–157.

Kasy, M. (2011). Identification in triangular systems using control functions. *Econometric Theory 27*(3), 663–671.

Kato, K. (2011). A note on moment convergence of bootstrap m-estimators. *Statistics & Risk Modeling 28*(1), 51–61.

Kitagawa, T. (2021). The identification region of the potential outcome distributions under instrument independence. *Journal of Econometrics Forthcoming*, 1–23.

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica: journal of the Econometric Society 46*, 33–50.

Kowalski, A. E. (2016). Doing more when you're running late: Applying marginal treatment effect methods to examine treatment effect heterogeneity in experiments. Technical report, National Bureau of Economic Research.

Krantz, S. G. and H. R. Parks (2002). *A primer of real analytic functions*. Springer Science & Business Media.

Lechner, M. (2002). Program heterogeneity and propensity score matching: An application to the evaluation of active labor market policies. *Review of Economics and Statistics 84*(2), 205–220.

Lee, S. (2003). Efficient semiparametric estimation of a partially linear quantile regression model. *Econometric theory 19*(1), 1–31.

Lee, S. (2007). Endogeneity in quantile regression models: A control function approach. *Journal of Econometrics 141*(2), 1131–1158.

Lee, S. and B. Salanié (2018). Identifying effects of multivalued treatments. *Econometrica 86*(6), 1939–1963.

Ma, S. and M. R. Kosorok (2005). Robust semiparametric m-estimation and the weighted bootstrap. *Journal of Multivariate Analysis 96*(1), 190–217.

Machado, J. A. and J. Mata (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of applied Econometrics 20*(4), 445–465.

Masten, M. (2018). Random coefficients on endogenous variables in simultaneous equations models. *The Review of Economic Studies 85*(2), 1193–1250.

Mogstad, M., A. Santos, and A. Torgovitsky (2018). Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica 86*(5), 1589–1619.

Nadarajah, S., E. Afuecheta, and S. Chan (2017). A compendium of copulas. *Statistica 77*(4), 279.

Newey, W. K. and J. L. Powell (2003). Instrumental variable estimation of nonparametric models. *Econometrica 71*(5), 1565–1578.

Pereda-Fernández, S. (2016). Estimation of counterfactual distributions with a continuous endogenous treatment. Technical report, Bank of Italy Temi di Discussione (Working Paper) No 1053.

Pereda-Fernández, S. (2021). Copula-based random effects models for clustered data. *Journal of Business & Economic Statistics 39*(2), 575–588.

Prokhorov, A. and P. Schmidt (2009). Likelihood-based estimation in a panel setting: robustness, redundancy and validity of copulas. *Journal of Econometrics 153*(1), 93–104.

Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica: Journal of the Econometric Society 39*(3), 577–591.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers 3*(2), 135–146.

Sancetta, A. and S. Satchell (2004). The bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric theory 20*(03), 535–562.

Sant'Anna, P. H. and X. Song (2019). Specification tests for the propensity score. *Journal of Econometrics 210*(2), 379–404.

Sant'Anna, P. H. and X. Song (2020). Specification tests for generalized propensity scores using double projections. Technical report, arXiv preprint arXiv:2003.13803.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Istitut de Statistique de l'Universitè de Paris 8*, 229–231.

Torgovitsky, A. (2015). Identification of nonseparable models using instruments with small support. *Econometrica 83*(3), 1185–1197.

van der Vaart, A. W. (2000). *Asymptotic statistics*, Volume 3. Cambridge university press.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes With Applications to Statistics*. Springer.

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

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

Vytlacil, E. and N. Yildiz (2007). Dummy endogenous variables in weakly separable models. *Econometrica 75*(3), 757–779.

Wüthrich, K. (2019a). A closed-form estimator for quantile treatment effects with endogeneity. *Journal of econometrics 210*(2), 219–235.

Wüthrich, K. (2019b). A comparison of two quantile models with endogeneity. *Journal of Business & Economic Statistics 38*(2), 443–456.

A Mathematical proofs

Let $W \equiv (Y, D, Z)$. The following notation is used throughout the Appendix:[21]

$$r(W, \beta, \theta, \gamma, \tau) \equiv \begin{bmatrix} \mathbf{1}(D=1) X \zeta_{G_{1,x}(\tau, \pi(Z;\gamma), \theta_1)}(Y - X'\beta_1) \\ \mathbf{1}(D=0) X \zeta_{G_{0,x}(\tau, \pi(Z;\gamma), \theta_0)}(Y - X'\beta_0) \\ \int_0^1 \mathbf{1}(D=1) \varphi(u, Z) \zeta_{G_{1,x}(\tau, \pi(Z;\gamma), \theta_1)}(Y - X'\beta_1) du \\ \int_0^1 \mathbf{1}(D=0) \varphi(u, Z) \zeta_{G_{0,x}(\tau, \pi(Z;\gamma), \theta_0)}(Y - X'\beta_0) du \\ s(D, Z; \gamma) \end{bmatrix}$$

$$q(W, \beta, \theta, \gamma, \tau) \equiv \begin{bmatrix} \mathbf{1}(D=1) X \rho_{G_{1,x}(\tau, \pi(Z;\gamma), \theta_1)}(Y - X'\beta_1) \\ \mathbf{1}(D=0) X \rho_{G_{0,x}(\tau, \pi(Z;\gamma), \theta_0)}(Y - X'\beta_0) \\ \mathbf{1}(D=1) \int_0^1 \varphi(u, Z) \rho_{G_{1,x}(\tau, \pi(Z;\gamma), \theta_1)}(Y - X'\beta_1) du \\ \mathbf{1}(D=0) \int_0^1 \varphi(u, Z) \rho_{G_{0,x}(\tau, \pi(Z;\gamma), \theta_0)}(Y - X'\beta_0) du \\ s(D, Z; \gamma) \end{bmatrix}$$

$f \mapsto \mathbb{E}_n[f(W)] \equiv \frac{1}{n}\sum_{i=1}^n f(W)$, $f \mapsto \mathbb{G}_n[f(W)] \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n f(W) - \mathbb{E}(f(W))$, $Q_n(\beta, \theta, \gamma, \tau) \equiv \mathbb{E}_n[q(W, \beta, \theta, \gamma, \tau)]$, and $Q(\beta, \theta, \gamma, \tau) \equiv \mathbb{E}[q(W, \beta, \theta, \gamma, \tau)]$, where $\rho_\tau(u) \equiv (\tau - \mathbf{1}(u < 0))u$, $\zeta_\tau(u) \equiv (\mathbf{1}(u < 0) - \tau)$, $\epsilon_d(\tau) \equiv Y - X'\beta_d(\tau)$, and $\hat{\epsilon}_d(\tau) \equiv Y - X'\hat{\beta}_d(\tau)$.

A.1 Proof of Lemma 1

Let $V = \tilde{F}_{V|X}\left(\tilde{V}|X\right)$. By definition, $V \sim U(0,1)$. Moreover, $\tilde{V} < \tilde{\pi}(Z) \Leftrightarrow \tilde{F}_{V|X}\left(\tilde{V}|X\right) < \tilde{F}_{V|X}(\tilde{\pi}(Z)|X) \equiv \pi(Z)$. Hence, $\tilde{F}_{D|Z} = F_{D|Z}$.

Similarly, let $U_D = \tilde{F}_{U|X}\left(\tilde{U}_D|X\right)$, which is also uniformly distributed. It follows that $Y = \tilde{g}\left(D, X, \tilde{U}_D\right) = \tilde{g}\left(D, X, \tilde{F}^{-1}_{U|X}(U_D|X)\right) \equiv g(D, X, U_D)$. The joint distribution of $\left(\tilde{U}_1, \tilde{V}\right)$ can

[21] Some of this notation is standard in the literature of empirical processes. See, *e.g.*, van der Vaart (2000).

be written as

$$\begin{aligned}\mathbb{P}\left(\tilde{U}_1 \leq \tau, \tilde{V} \leq \tilde{\pi}(z) | Z = z\right) &= \mathbb{P}\left(\tilde{F}^{-1}_{U|1,x}(U_1|x) \leq \tau, \tilde{F}^{-1}_{V|X}(V|x) \leq \tilde{\pi}(z) | Z = z\right) \\ &= \mathbb{P}\left(U_1 \leq \tilde{F}_{U|1,x}(\tau|x), V \leq \pi(z) | Z = z\right) \\ &= C_{1,x}\left(\tilde{F}_{U|1,x}(\tau|x), \pi(z)\right)\end{aligned}$$

where the first equality follows by the invertibility of $\tilde{U}_1$ and $\tilde{V}$, the second one by the first result of the Lemma, and the third one by definition of the copula. Similarly, for $d = 0$ it can be shown that

$$\mathbb{P}\left(\tilde{U}_0 \leq \tau, \tilde{V} > \tilde{\pi}(z) | Z = z\right) = \tilde{F}_{U|0,x}(\tau|x) - C_{0,x}\left(\tilde{F}_{U|0,x}(\tau|x), \pi(z)\right)$$

Define $\tilde{G}_{d,x}(\tau, \pi(z)) \equiv \mathbb{P}\left(\tilde{U}_d \leq \tau | D = d, Z = d\right)$. For $d = 1$, it can be expressed as

$$\begin{aligned}\tilde{G}_{1,x}(\tau, \tilde{\pi}(z)) &\equiv \frac{\mathbb{P}\left(\tilde{U}_1 \leq \tau, \tilde{V} \leq \tilde{\pi}(z) | Z = z\right)}{\mathbb{P}\left(\tilde{V} \leq \tilde{\pi}(z) | Z = z\right)} \\ &= \frac{\mathbb{P}\left(\tilde{F}^{-1}_{U|1,x}(U_1|x) \leq \tau, \tilde{F}^{-1}_{V|x}(V|x) \leq \tilde{\pi}(z) | Z = z\right)}{\mathbb{P}\left(\tilde{F}^{-1}_{V|x}(V|x) \leq \tilde{\pi}(z) \middle| Z = z\right)} \\ &= \frac{C_{1,x}\left(\tilde{F}_{U|1,x}(\tau|x), \pi(z)\right)}{\pi(z)} = G_{1,x}\left(\tilde{F}_{U|1,x}(\tau|x), \pi(z)\right) \qquad (17)\end{aligned}$$

and for $d = 0$ as

$$\begin{aligned}\tilde{G}_{0,x}(\tau, \tilde{\pi}(z)) &\equiv \frac{\mathbb{P}\left(\tilde{U}_0 \leq \tau, \tilde{V} > \tilde{\pi}(z) | Z = z\right)}{\mathbb{P}\left(\tilde{V} > \tilde{\pi}(z) | Z = z\right)} \\ &= \frac{\tilde{F}_{U|0,x}(\tau|x) - C_{0,x}\left(\tilde{F}_{U|0,x}(\tau|x), \pi(z)\right)}{1 - \pi(z)} = G_{0,x}\left(\tilde{F}_{U|1,x}(\tau|x), \pi(z)\right) \qquad (18)\end{aligned}$$

Then, the distribution of Y, conditional on $D = d$ and $Z = z$ equals

$$\begin{aligned}
\mathbb{P}\left(Y \leq y | D = d, Z = z\right) &= \int \mathbf{1}\left(\tilde{g}\left(d, x, \tilde{u}_1\right) \leq y\right) d\tilde{G}_{d,x}\left(\tilde{u}_d, \tilde{\pi}\left(z\right)\right) \\
&= \int \mathbf{1}\left(\tilde{g}\left(d, x, \tilde{u}_d\right) \leq y\right) dG_{d,x}\left(\tilde{F}_{U|d,x}\left(\tilde{u}_d | x\right), \pi\left(z\right)\right) \\
&= \int \mathbf{1}\left(\tilde{g}\left(d, x, \tilde{F}^{-1}_{U|d,x}\left(u_d\right)\right) \leq y\right) dG_{d,x}\left(u_d, \pi\left(z\right)\right) \\
&= \int \mathbf{1}\left(g\left(d, x, u_d\right) \leq y\right) dG_{d,x}\left(u_d, \pi\left(z\right)\right)
\end{aligned}$$

where the second equality follows by Equation 18, the third one by the invertibility of $\tilde{U}_d$, and the fourth one by the definition of $\tilde{g}$, completing the proof.

A.2 Proof of Lemma 2

By Assumption 3 and Equations 3-4, the result follows immediately.

A.3 Proof of Proposition 1

The proof is split in parts. First, I show the local identification of $\theta_{d,x}$, then I show its global identification, and finally I show the identification of the SQF.

Define the functions $M_{d,x}\left(\tau, \theta_{d,x}\right) \equiv G_{d,x}\left(G^{-1}_{d,x}\left(\tau, \pi\left(z'\right); \theta_{d,x}\right), \pi\left(z\right); \theta_{d,x}\right)$ and $\phi_{d,x}\left(\tau\right) \equiv F_{Y|D=d,Z}\left(F^{-1}_{Y|D=d,Z}\left(\tau | z'\right) | z\right)$. By Equations 5-6, $M_{d,x}\left(\tau, \theta_{d,x}\right) = \phi_{d,x}\left(\tau\right)$, $\forall x \in \mathcal{X}, d = 0, 1$. Taking the derivative with respect to the copula parameter for a generic value of θ, and dropping the (d, x) subscript from the functions M and ϕ for notational simplicity, yields

$$\begin{aligned}
\nabla_\theta M\left(\tau, \theta\right) = &\nabla_\theta G\left(G^{-1}\left(\tau, \pi\left(z'\right); \theta\right), \pi\left(z\right); \theta\right) \\
&- \nabla_u G\left(G^{-1}\left(\tau, \pi\left(z'\right); \theta\right), \pi\left(z\right); \theta\right) \frac{\nabla_\theta G\left(G^{-1}\left(\tau, \pi\left(z'\right); \theta\right), \pi\left(z'\right); \theta\right)}{\nabla_u G\left(G^{-1}\left(\tau, \pi\left(z'\right); \theta\right), \pi\left(z'\right); \theta\right)}
\end{aligned} \tag{19}$$

Because $M\left(\tau, \theta\right)$ holds for any $\tau \in (0, 1)$, there is an continuum of moments that pin down the parameter θ. Instead, consider a finite number of values of τ, given by $\{\tau_1, ..., \tau_T\}$.

Local identification holds when the matrix that collects the Jacobian for all values in this set is of full rank, as required by Theorem 6 in Rothenberg (1971). Because it is a scalar parameter, full rank is attained if $\nabla_\theta M(\tau,\theta) \neq 0$ for any of the values of τ, *i.e.*,

$$\frac{\nabla_\theta G\left(G^{-1}\left(\tau,\pi\left(z'\right);\theta\right),\pi\left(z\right);\theta\right)}{\nabla_u G\left(G^{-1}\left(\tau,\pi\left(z'\right);\theta\right),\pi\left(z\right);\theta\right)} - \frac{\nabla_\theta G\left(G^{-1}\left(\tau,\pi\left(z'\right);\theta\right),\pi\left(z'\right);\theta\right)}{\nabla_u G\left(G^{-1}\left(\tau,\pi\left(z'\right);\theta\right),\pi\left(z'\right);\theta\right)} \neq 0 \tag{20}$$

Let $\tau' \equiv G^{-1}\left(\tau,\pi\left(z'\right);\theta\right) \Leftrightarrow \tau = G\left(\tau',\pi\left(z'\right);\theta\right)$. Then, Equation 20 can be rewritten as

$$\frac{\nabla_\theta G\left(\tau',\pi\left(z\right);\theta\right)}{\nabla_u G\left(\tau',\pi\left(z\right);\theta\right)} - \frac{\nabla_\theta G\left(\tau',\pi\left(z'\right);\theta\right)}{\nabla_u G\left(\tau',\pi\left(z'\right);\theta\right)} \neq 0 \tag{21}$$

By the definition of the conditional copula, ${\nabla_\theta G(\tau,\pi;\theta)}/{\nabla_u G(\tau,\pi;\theta)} = {\nabla_\theta C(\tau,\pi;\theta)}/{\nabla_u C(\tau,\pi;\theta)}$ for $d=1$, and ${\nabla_\theta G(\tau,\pi;\theta)}/{\nabla_u G(\tau,\pi;\theta)} = {\nabla_\theta(\tau - C(\tau,\pi;\theta))}/{\nabla_u(\tau - C(\tau,\pi;\theta))}$ for $d=0$, so Equation 21 is equivalent to

$$\frac{\nabla_\theta C\left(\tau',\pi\left(z\right);\theta\right)}{\nabla_u C\left(\tau',\pi\left(z\right);\theta\right)} - \frac{\nabla_\theta C\left(\tau',\pi\left(z'\right);\theta\right)}{\nabla_u C\left(\tau',\pi\left(z'\right);\theta\right)} \neq 0 \tag{22}$$

for $d=1$ and

$$-\frac{\nabla_\theta C\left(\tau',\pi\left(z\right);\theta\right)}{1-\nabla_u C\left(\tau',\pi\left(z\right);\theta\right)} + \frac{\nabla_\theta C\left(\tau',\pi\left(z'\right);\theta\right)}{1-\nabla_u C\left(\tau',\pi\left(z'\right);\theta\right)} \neq 0 \tag{23}$$

for $d=0$. These are equivalent to the two equations in Condition 4.7 in Han and Vytlacil (2017). By Lemma 4.1 in Han and Vytlacil (2017), under Assumption 5, the copula $G_{d,x}(\tau,\pi)$ satisfies Assumption 6 in Han and Vytlacil (2017) if and only if Equations 22-23 are strictly decreasing in the second argument of the copula. If $\pi(z) \neq \pi(z')$, *i.e.*, if the instrument does not come from a degenerate distribution, then the copula parameter $\theta_{d,x}$ is locally identifiable by Proposition 4.1 in Han and Vytlacil (2017).

For global identification, I first show that it is possible to apply Lemma 4.2 in Han and Vytlacil (2017) on a restricted parameter space, extending it subsequently to the entire

parameter space. Note that Equation 19 can be rewritten as

$$\nabla_\theta M(\tau,\theta) = \nabla_u G(\tau', \pi(z);\theta)\left[\frac{\nabla_\theta G(\tau', \pi(z);\theta)}{\nabla_u G(\tau', \pi(z);\theta)} - \frac{\nabla_\theta G(\tau', \pi(z');\theta)}{\nabla_u G(\tau', \pi(z');\theta)}\right]$$

where $\tau' \equiv G^{-1}(\tau, \pi(z');\theta)$. By Lemma 4.1 in Han and Vytlacil (2017), the term in brackets is positive when $\pi(z) < \pi(z')$. Moreover, $\nabla_u G(\tau,\pi;\theta) = \frac{1}{\pi}\nabla_u C(\tau,\pi;\theta) > 0$ for $d = 1$, and $\nabla_u G(\tau,\pi;\theta) = \frac{1}{1-\pi}(1 - \nabla_u C(\tau,\pi;\theta)) > 0$ for $d = 0$.[22] Therefore, the Jacobian $\nabla_\theta M(\tau,\theta)$ is positive semidefinite if $\pi(z) < \pi(z')$ and negative semidefinite if $\pi(z) > \pi(z')$. Moreover, it has full rank for any θ as long as $\pi(z) \neq \pi(z')$.

Let $\Theta_c \subseteq \Theta$ be a bounded open space with half spaces $\Theta_{c_1} \equiv \{\theta \in \Theta_c : \pi(z) < \pi(z')\}$, and $\Theta_{c_2} \equiv \{\theta \in \Theta_c : \pi(z) > \pi(z')\}$, which are simply connected. Define $\phi_{c_1}(\tau) = M(\tau, \Theta_{c_1})$ and $\phi_{c_2}(\tau) = M(\tau, \Theta_{c_2})$, and let $M|_{\Theta_{c_1}} : \Theta_{c_1} \to \phi_{c_1}$ and $M|_{\Theta_{c_2}} : \Theta_{c_2} \to \phi_{c_2}$ be the function $M(\tau,\cdot)$ on its restricted domains.

Because $M|_{\Theta_{c_1}}(\tau,\cdot)$ and $M|_{\Theta_{c_2}}(\tau,\cdot)$ are continuous, the pre-image of a closed set under $M|_{\Theta_{c_1}}(\tau,\cdot)$ and $M|_{\Theta_{c_2}}(\tau,\cdot)$ is closed. Because Θ_{c_1} and Θ_{c_2} are bounded, the pre-image of a bounded set is bounded. Thus, $M|_{\Theta_{c_1}}(\tau,\cdot)$ and $M|_{\Theta_{c_2}}(\tau,\cdot)$ are proper.

Because Θ_{c_1} and Θ_{c_2} are simply connected, $M|_{\Theta_{c_1}}(\tau,\cdot)$ and $M|_{\Theta_{c_2}}(\tau,\cdot)$ are continuous on Θ_{c_1} and Θ_{c_2}, respectively, and the Jacobian $\nabla_\theta M(\tau,\cdot)$ is positive semidefinite and negative semidefinite on Θ_{c_1} and Θ_{c_2}, respectively, it follows that ϕ_{c_1} and ϕ_{c_2} are simply connected.

Also, $\nabla_\theta M(\tau,\cdot)$ has full rank over Θ_{c_1} and Θ_{c_2}. Thus, by Lemma 4.2 in Han and Vytlacil (2017), $\phi(\tau) = M(\tau,\theta)$ has a unique solution on Θ_{c_1} and Θ_{c_2}, respectively. Because there exist $M|^{-1}_{\Theta_{c_1}}(\tau,\cdot) \in \Theta_{c_1}$ for $\phi \in \phi_{c_1}$ and $M|^{-1}_{\Theta_{c_2}}(\tau,\cdot) \in \Theta_{c_2}$ for $\phi \in \phi_{c_2}$, θ is globally identified.

Now let $\Theta_1 \equiv \{\theta \in \Theta : \pi(z) < \pi(z')\}$ and $\Theta_2 \equiv \{\theta \in \Theta : \pi(z) > \pi(z')\}$ be two simply connected, possibly unbounded spaces. Θ_1 and Θ_2 can be represented as a countable union of bounded open simply connected sets. *E.g.*, $\Theta_j = \cup_{i=1}^{\infty}\Theta_{ji}$, where Θ_{ji} is a sequence of bounded open simply connected sets in Θ_j such that $\Theta_{j1} \subset \Theta_{j2} \subset ... \subset \Theta_j$ for $j = 1, 2$.

Let $\phi_{ji}(\tau) \equiv M(\tau, \Theta_{ji})$ for $i = 1, 2, ...$ and $j = 1, 2$. Then, $\phi_j(\tau) = M(\tau, \Theta_j) =$

[22] To see why the latter holds, note that $\nabla_u C(u,v) = \mathbb{P}(V \leq v | U = u)$ is itself a probability, and therefore bounded between 0 and 1.

$M(\tau, \cup_{i=1}^{\infty}\Theta_{ji}) = \cup_{i=1}^{\infty} M(\tau, \Theta_{ij}) = \cup_{i=1}^{\infty}\phi_{ij}(\tau)$, and $\phi_{j1} \subset \phi_{j2} \subset ... \subset \phi_j$. Then, for any given $\phi \in \phi_j$, $\exists q : \phi \in \phi_{ji} \forall i \geq q$, so $M|_{\Theta_{c_j}}^{-1}(\tau, \phi) \in \Theta_{ji} \forall i \geq q$, and therefore $M^{-1}(\tau, \phi) = M|_{\cup_{i=q}^{\infty}\Theta_{ji}}^{-1}(\tau, \phi) \in \cup_{i=q}^{\infty}\Theta_{ji} = \Theta_j$. Because $M^{-1}(\tau, \phi)$ is the unique solution on Θ_j, it is the unique solution of the full system with $\tau = \{\tau_1, ..., \tau_T\}$. Thus, θ is globally identified in Θ_j.

Having established the global identification of the copula parameter, it is straightforward to identify the SQF. To see this, note that by Equations 3-4, $F_{Y|D=d,z}(g(d, x, \tau)|z) = G_{d,x}(\tau, \pi(z); \theta_{d,x})$ for $d = 0, 1$. Therefore, one can solve for g and express it in terms of either observed or identified functions: $g(d, x, \tau) = F_{Y|D=d,z}^{-1}(G_{d,x}(\tau, \pi(z); \theta_{d,x}))$. This finishes the proof.

A.4 Proof of Theorem 1

First I show consistency of $\hat{\vartheta}(\tau)$. By Assumptions 5, 7, 8,and 10, $Q(\beta, \theta, \gamma, \tau)$ is continuous over $\mathcal{B} \times \Theta \times \Gamma \times \mathcal{T}$. By Lemma 6, $\sup_{(\beta,\theta,\gamma,\tau)\in\mathcal{B}\times\Theta\times\Gamma\times\mathcal{T}} \|Q_n(\beta, \theta, \gamma, \tau) - Q(\beta, \theta, \gamma, \tau)\| \xrightarrow{P} 0$, uniformly in $\mathcal{D}$. Thus, by Lemma 5, $\sup_{\tau\in\mathcal{T}} \left\|\hat{\vartheta}(\tau) - \vartheta(\tau)\right\| \xrightarrow{P} 0$, uniformly in $\mathcal{D}$.

Second, I show its asymptotic distribution. By Theorem 3 in Koenker and Bassett (1978), it is possible to show that

$$O\left(\frac{1}{\sqrt{n}}\right) = \sqrt{n}\mathbb{E}_n\left[\mathbf{1}(D = d) X \zeta_{G_{d,x}\left(\tau,\pi(Z;\hat{\gamma}),\hat{\theta}_d\right)}\left(\hat{\epsilon}_d(\tau)\right)\right]$$

By Lemma 6 and Assumption 11, the following expansion holds in $\ell^{\infty}(\mathcal{T})$:

$$\begin{aligned}
O\left(\frac{1}{\sqrt{n}}\right) &= \mathbb{G}_n\left[\mathbf{1}(D = d) X \zeta_{G_{d,x}\left(\tau,\pi(Z;\hat{\gamma});\hat{\theta}_d\right)}\left(\hat{\epsilon}_d(\tau)\right)\right] + \sqrt{n}\mathbb{E}\left[\mathbf{1}(D = d) X \zeta_{G_{d,x}\left(\tau,\pi(Z;\hat{\gamma});\hat{\theta}_d\right)}\left(\hat{\epsilon}_d(\tau)\right)\right] \\
&= \mathbb{G}_n\left[\mathbf{1}(D = d) X \zeta_{G_{d,x}(\tau,\pi(Z;\gamma);\theta_d)}\left(\epsilon_d(\tau)\right)\right] + o_P(1) \\
&+ \sqrt{n}\mathbb{E}\left[\mathbf{1}(D = d) X \zeta_{G_{d,x}\left(\tau,\pi(Z;\hat{\gamma});\hat{\theta}_d\right)}\left(\hat{\epsilon}_d(\tau)\right)\right] \\
&= \mathbb{G}_n\left[\mathbf{1}(D = d) X \zeta_{G_{d,x}(\tau,\pi(Z;\gamma);\theta_d)}\left(\epsilon_d(\tau)\right)\right] + J_{\beta_d}(\tau)\sqrt{n}\left(\hat{\beta}_d(\tau) - \beta_d(\tau)\right) \\
&- J_{\gamma_d}(\tau)\sqrt{n}(\hat{\gamma} - \gamma) - J_{\theta_d}(\tau)\sqrt{n}\left(\hat{\theta}_d - \theta_d\right) + o_P(1)
\end{aligned}$$

where

$$J_{\beta_d}(\tau) \equiv \frac{\partial \mathbb{E}\left[\mathbf{1}(D=d) X \zeta_{G_{d,x}(\tau,\pi(Z;\gamma);\theta_d)}(\epsilon_d(\tau))\right]}{\partial \beta_d}$$

$$J_{\gamma_d}(\tau) \equiv -\frac{\partial \mathbb{E}\left[\mathbf{1}(D=d) X \zeta_{G_{d,x}(\tau,\pi(Z;\gamma);\theta_d)}(\epsilon_d(\tau))\right]}{\partial \gamma}$$

$$J_{\theta_d}(\tau) \equiv -\frac{\partial \mathbb{E}\left[\mathbf{1}(D=d) X \zeta_{G_{d,x}(\tau,\pi(Z;\gamma);\theta_d)}(\epsilon_d(\tau))\right]}{\partial \theta_d}$$

Rearranging and solving for $\sqrt{n}\left(\hat{\beta}_d(\tau) - \beta_d(\tau)\right)$,

$$\begin{aligned}\sqrt{n}\left(\hat{\beta}_d(\tau) - \beta_d(\tau)\right) = -J_{\beta_d}(\tau)^{-1}\Big\{&\mathbb{G}_n\left[\mathbf{1}(D=d) X \zeta_{G_{d,x}(\tau,\pi(Z;\gamma);\theta_d)}(\epsilon_d(\tau))\right] \\ &-J_{\gamma_d}(\tau)\sqrt{n}(\hat{\gamma}-\gamma) - J_{\theta_d}(\tau)\sqrt{n}\left(\hat{\theta}_d - \theta_d\right)\Big\} + o_P(1) \qquad (24)\end{aligned}$$

in $\ell^\infty(\mathcal{T})$.

Using Theorem 3 in Koenker and Bassett (1978) again, it is possible to show that

$$O\left(\frac{1}{\sqrt{n}}\right) = \sqrt{n}\mathbb{E}_n\left[\int_\varepsilon^{1-\varepsilon} \mathbf{1}(D=d)\varphi(u,Z)\zeta_{G_{d,x}\left(u,\pi(Z;\hat{\gamma});\hat{\theta}_d\right)}(\hat{\epsilon}_d(u))\,du\right]$$

By Lemma 6 and Assumption 11, the following expansion holds:

$$
\begin{aligned}
O\left(\frac{1}{\sqrt{n}}\right) &= \mathbb{G}_n\left[\int_{\varepsilon}^{1-\varepsilon} \mathbf{1}\left(D=d\right)\varphi\left(u,Z\right)\zeta_{G_{d,x}\left(u,\pi(Z;\hat{\gamma});\hat{\theta}_d\right)}\left(\hat{\epsilon}_d\left(u\right)\right)du\right] \\
&\quad + \sqrt{n}\int_{\varepsilon}^{1-\varepsilon} \mathbb{E}\left[\mathbf{1}\left(D=d\right)\varphi\left(u,Z\right)\zeta_{G_{d,x}\left(u,\pi(Z;\hat{\gamma});\hat{\theta}_d\right)}\left(\hat{\epsilon}_d\left(u\right)\right)\right]du \\
&= \mathbb{G}_n\left[\int_{\varepsilon}^{1-\varepsilon} \mathbf{1}\left(D=d\right)\varphi\left(u,Z\right)\zeta_{G_{d,x}(u,\pi(Z;\gamma);\theta_d)}\left(\epsilon_d\left(u\right)\right)du\right] + o_P\left(1\right) \\
&\quad + \sqrt{n}\int_{\varepsilon}^{1-\varepsilon} \mathbb{E}\left[\mathbf{1}\left(D=d\right)\varphi\left(u,Z\right)\zeta_{G_{d,x}\left(u,\pi(Z;\hat{\gamma});\hat{\theta}_d\right)}\left(\hat{\epsilon}_d\left(u\right)\right)\right]du \\
&= \mathbb{G}_n\left[\int_{\varepsilon}^{1-\varepsilon} \mathbf{1}\left(D=d\right)\varphi\left(u,Z\right)\zeta_{G_{d,x}(u,\pi(Z;\gamma);\theta_d)}\left(\epsilon_d\left(u\right)\right)du\right] \\
&\quad + \sqrt{n}\int_{\varepsilon}^{1-\varepsilon} \tilde{J}_{\beta_d}\left(u\right)\left(\hat{\beta}_d\left(u\right)-\beta_d\left(u\right)\right)du \\
&\quad - \sqrt{n}\int_{\varepsilon}^{1-\varepsilon} \tilde{J}_{\theta_d}\left(u\right)du\left(\hat{\theta}_d-\theta_d\right) - \sqrt{n}\int_{\varepsilon}^{1-\varepsilon} \tilde{J}_{\gamma_d}\left(u\right)du\left(\hat{\gamma}-\gamma\right) + o_P\left(1\right)
\end{aligned}
$$

where

$$
\tilde{J}_{\beta_d}\left(\tau\right) \equiv \frac{\partial \mathbb{E}\left[\mathbf{1}\left(D=d\right)\varphi\left(\tau,Z\right)\zeta_{G_{d,x}(\tau,\pi(Z;\gamma);\theta_d)}\left(\epsilon_d\left(\tau\right)\right)\right]}{\partial \beta_d}
$$

$$
\tilde{J}_{\gamma_d}\left(\tau\right) \equiv -\frac{\partial \mathbb{E}\left[\mathbf{1}\left(D=d\right)\varphi\left(\tau,Z\right)\zeta_{G_{d,x}(\tau,\pi(Z;\gamma);\theta_d)}\left(\epsilon_d\left(\tau\right)\right)\right]}{\partial \gamma}
$$

$$
\tilde{J}_{\theta_d}\left(\tau\right) \equiv -\frac{\partial \mathbb{E}\left[\mathbf{1}\left(D=d\right)\varphi\left(\tau,Z\right)\zeta_{G_{d,x}(\tau,\pi(Z;\gamma);\theta_d)}\left(\epsilon_d\left(\tau\right)\right)\right]}{\partial \theta_d}
$$

Rearranging and solving for $\sqrt{n}\left(\hat{\theta}_d-\theta_d\right)$,

$$
\begin{aligned}
\sqrt{n}\left(\hat{\theta}_d-\theta_d\right) &= \left[\int_{\varepsilon}^{1-\varepsilon} \tilde{J}_{\theta_d}\left(u\right)du\right]^{-1}\left\{\mathbb{G}_n\left[\int_{\varepsilon}^{1-\varepsilon} \mathbf{1}\left(D=d\right)\varphi\left(u,Z\right)\zeta_{G_{d,x}(u,\pi(Z;\gamma);\theta_d)}\left(\epsilon_d\left(u\right)\right)du\right]\right. \\
&\quad \left. + \sqrt{n}\int_{\varepsilon}^{1-\varepsilon} \tilde{J}_{\beta_d}\left(u\right)\left(\hat{\beta}_d\left(u\right)-\beta_d\left(u\right)\right)du - \sqrt{n}\int_{\varepsilon}^{1-\varepsilon} \tilde{J}_{\gamma_d}\left(u\right)du\left(\hat{\gamma}-\gamma\right)\right\} + o_P\left(1\right)
\end{aligned} \tag{25}
$$

Define $A(\tau) \equiv \hat{\vartheta}(\tau) - \vartheta(\tau)$ and $\psi(\tau) \equiv r(W, \beta(\tau), \theta, \gamma, \tau)$. Combining Equations 24 and 25 yields

$$A(\tau) = F(\tau) A(\tau) + \int_{\varepsilon}^{1-\varepsilon} D(u) A(u) du + C(\tau) \frac{1}{\sqrt{n}} \mathbb{G}_n \psi(\tau) + o_P\left(\frac{1}{\sqrt{n}}\right) \tag{26}$$

in $\ell^{\infty}(\mathcal{T})$. Equation 26 is a particular case of a Fredholm integral equation of the second kind. The solution to this type of equations is a Liouville-Neumann series. By Lemma 4, the solution to this equation is given by:

$$\begin{aligned}\sqrt{n}A(\tau) = {} & F^I(\tau) \left(I - \int_{\varepsilon}^{1-\varepsilon} D(u) F^I(u) du\right)^{-1} \int_{\varepsilon}^{1-\varepsilon} D(u) F^I(u) C(u) \mathbb{G}_n \psi(u) du \\ & + F^I(\tau) C(\tau) \mathbb{G}_n \psi(\tau) + o_P(1)\end{aligned} \tag{27}$$

in $\ell^{\infty}(\mathcal{T})$. Using the Functional Delta Method and Lemmas 3 and 5, it follows that

A.5 Proof of Theorem 2

This proof is split into several steps. The first one is to show the joint asymptotic distribution of $\left(\hat{F}_{Y|d,X}(y|x), \int_{\mathcal{X}} f d\hat{F}_X\right)$. By Proposition 2 in Chernozhukov et al. (2010), the mapping $b \to \varepsilon + \int_{\varepsilon}^{1-\varepsilon} \mathbf{1}(x'b(u) \leq y) du$ is Hadamard differentiable at $b(\cdot) = \beta_d(\cdot)$ tangentially to $C(\mathcal{T})^{d_x}$, where the derivative is equal to $D_h(y|x) = -f_{Y|d,X}(y|x) x' h\left(F_{Y|d,X}(y|x)\right)$, and d_x is the dimension of $\mathcal{X}$. By Lemma E.4 in Chernozhukov et al. (2013), $\int_{\mathcal{X}} f(y,x) d\left(\hat{F}_X(x) - F_X(x)\right) \Rightarrow \mathbb{Z}_X(f)$. Using these two results, Theorem 1 and the functional delta method, it follows that

$$\sqrt{n}\begin{pmatrix} \hat{F}_{Y|d,X}(y|x) - F_{Y|d,X}(y|x) \\ \int_{\mathcal{X}} f d\left(\hat{F}_X - F_X\right) \end{pmatrix} \Rightarrow \begin{pmatrix} \mathbb{Z}_{Y_d}(y,x) \\ \mathbb{Z}_X(f) \end{pmatrix}$$

where $\mathbb{Z}_{Y_d}(y,x) = -f_{Y|d,X}(y|x) x' \mathbb{Z}_{\beta_d}\left(F_{Y|d,X}(y|x)\right)$, and $\mathbb{Z}_{\beta_d}(\tau)$ is the first component of $\mathbb{Z}_{\vartheta_d}(\tau)$.

The second step is to show the asymptotic distribution of $\hat{F}_{Y|d}(y)$. By the functional

delta method,

$$\begin{aligned}\sqrt{n}\left(\hat{F}_{Y|d}(y)-F_{Y|d}(y)\right) &= \sqrt{n}\int_{\mathcal{X}}\left(\hat{F}_{Y|d,X}(y|x)-F_{Y|d,X}(y|x)\right)dF_X(x)\\ &\quad+\sqrt{n}\int_{\mathcal{X}}F_{Y|d,X}(y|x)\,d\left(\hat{F}_X(x)-F_X(x)\right)+o_P(1)\\ &\Rightarrow \int_{\mathcal{X}}\mathbb{Z}_{Y_d}(y,x)\,dF_X(x)+\mathbb{Z}_X\left(F_{Y|d,X}(y|\cdot)\right)\equiv\mathbb{Z}_{Y_d}(y)\end{aligned}$$

jointly in $d\in\mathcal{D}$. $\mathbb{Z}_{Y_d}(y,x)$ is a.s. uniformly continuous with respect to (y,x), and $\mathbb{Z}_X(f)$ is continuous with respect to f under the metric λ. Moreover, by the uniform continuity of $F_{Y|d,X}(y|\cdot)$ with respect to y under the metric λ, $y\to\mathbb{Z}_X\left(F_{Y|d,X}(y|\cdot)\right)$ is a.s. uniformly continuous with respect to y.

The third step is to show the asymptotic distribution of $\hat{Q}_{Y|d}(\tau)$. By the functional delta method,

$$\begin{aligned}\sqrt{n}\left(\hat{Q}_{Y|d}(\tau)-Q_{Y|d}(\tau)\right) &= -\frac{\sqrt{n}\left(\hat{F}_{Y|d}\left(Q_{Y|d}(\tau)\right)-F_{Y|d}\left(Q_{Y|d}(\tau)\right)\right)}{f_{Y|d}\left(Q_{Y|d}(\tau)\right)}+o_P(1)\\ &\Rightarrow -\frac{\mathbb{Z}_{Y_d}\left(Q_{Y|d}(\tau)\right)}{f_{Y|d}\left(Q_{Y|d}(\tau)\right)}\equiv\mathbb{Z}_{Q_d}(\tau)\end{aligned}$$

jointly in $d\in\mathcal{D}$, where I have used the Hadamard differentiability of the quantile operator. $\tau\to Q_{Y|d}(\tau)$ is a.s. uniformly continuous by Assumption 10, and together with the a.s. uniform continuity of $\mathbb{Z}_{Y_d}(y)$, it follows that $\mathbb{Z}_{Q_d}(\tau)$ is a.s. uniformly continuous with respect to τ.

The last step is to show the asymptotic distribution of Equation 12. By the extended continuous mapping theorem,

$$\sqrt{n}\left(\hat{\Delta}_{QTE}(\tau)-\Delta_{QTE}(\tau)\right)\Rightarrow\mathbb{Z}_{Q_1}(\tau)-\mathbb{Z}_{Q_0}(\tau)\equiv\mathbb{Z}_{QTE}(\tau)$$

$\mathbb{Z}_{QTE}(\tau)$ is a.s. uniformly continuous with respect to τ by the Hadamard differentiability of $\tau\to\Delta_{QTE}(\tau)$, so the desired result follows.

A.6 Proof of Theorem 3

This proof is split into two steps. In the first step, I show the joint asymptotic distribution of $\left(\hat{g}_d(x,\tau), \int_\varepsilon^{1-\varepsilon} f_1 d\hat{C}_{d,x}(\tau|v), \int_{\mathcal{X}} f_2 d\hat{F}_X\right)$. Note that, by Assumption 5,

$$\begin{aligned}
\sqrt{n}\int_\varepsilon^{1-\varepsilon} f_1 d\left(\hat{C}_{d,x}(\tau,v) - C_{d,x}(\tau,v)\right) &= \sqrt{n}\int_\varepsilon^{1-\varepsilon} f_1\left(\hat{c}_{d,x}(\tau,v) - c_{d,x}(\tau,v)\right)d\tau \\
&= \sqrt{n}\int_\varepsilon^{1-\varepsilon} f_1 \nabla_{\theta_d} c_{d,x}(\tau,v;\theta_d)\left(\hat{\theta}_d - \theta_d\right)d\tau + o_P(1) \\
&\Rightarrow \int_\varepsilon^{1-\varepsilon} f_1 \nabla_{\theta_d} c_{d,x}(\tau,v;\theta_d)\, d\tau \mathbb{Z}_{\theta_d}(\cdot) \equiv \mathbb{Z}_{C_d}(f_1)
\end{aligned}$$

where the second equality follows by the functional delta method. Moreover, the mapping $b \to x'b(u)$ is linear and therefore Hadamard differentiable at $b(\cdot) = \beta_d(\cdot)$ tangentially to $C(\mathcal{T})^{d_x}$, where the derivative equals $D_h(\tau,x) = x'h(\tau)$. By Lemma E.4 in Chernozhukov et al. (2013), $\int_{\mathcal{X}} f d\left(\hat{F}_X - F_X\right) \Rightarrow \mathbb{Z}_X(f)$. Using these results, together with Theorem 1 and the functional delta method, it follows that

$$\sqrt{n}\begin{pmatrix} \hat{g}(d,x,\tau) - g(d,x,\tau) \\ \int_\varepsilon^{1-\varepsilon} f_1 d\left(\hat{C}_{d,x}(\tau,v) - C_{d,x}(\tau,v)\right) \\ \int_{\mathcal{X}} f_2 d\left(\hat{F}_X(x) - F_X(x)\right) \end{pmatrix} \Rightarrow \begin{pmatrix} \mathbb{Z}_{g_d}(\tau,x) \\ \mathbb{Z}_{C_d}(f_1,v) \\ \mathbb{Z}_X(f_2) \end{pmatrix}$$

The second step is to show the asymptotic distribution of Equation 13. By the functional

delta method,

$$\begin{aligned}
\sqrt{n}\left(\hat{\Delta}_{MTE}(v)-\Delta_{MTE}(v)^{\varepsilon}\right) &= \sqrt{n}\int_{\mathcal{X}}\int_{\varepsilon}^{1-\varepsilon}\left(\hat{g}(1,x,\tau)-g(1,x,\tau)\right)dC_{1,x}(\tau|v)\,dF_X(x)\\
&+\sqrt{n}\int_{\mathcal{X}}\int_{\varepsilon}^{1-\varepsilon}g(1,x,\tau)\,d\left(\hat{C}_{1,x}(\tau|v)-C_{1,x}(\tau|v)\right)dF_X(x)\\
&+\sqrt{n}\int_{\mathcal{X}}\int_{\varepsilon}^{1-\varepsilon}g(1,x,\tau)\,dC_{1,x}(\tau|v)\,d\left(\hat{F}_X(x)-F_X(x)\right)\\
&-\sqrt{n}\int_{\mathcal{X}}\int_{\varepsilon}^{1-\varepsilon}\left(\hat{g}(0,x,\tau)-g(0,x,\tau)\right)dC_{0,x}(\tau|v)\,dF_X(x)\\
&-\sqrt{n}\int_{\mathcal{X}}\int_{\varepsilon}^{1-\varepsilon}g(0,x,\tau)\,d\left(\hat{C}_{0,x}(\tau|v)-C_{0,x}(\tau|v)\right)dF_X(x)\\
&-\sqrt{n}\int_{\mathcal{X}}\int_{\varepsilon}^{1-\varepsilon}g(0,x,\tau)\,dC_{0,x}(\tau|v)\,d\left(\hat{F}_X(x)-F_X(x)\right)+o_P(1)\\
&\Rightarrow\int_{\mathcal{X}}\int_{\varepsilon}^{1-\varepsilon}\mathbb{Z}_{g_1}(\tau,x)\,dC_{1,x}(\tau|v)\,dF_X(x)\\
&+\sqrt{n}\int_{\mathcal{X}}\mathbb{Z}_{C_1}\left(g(1,x,\cdot),v\right)dF_X(x)\\
&+\sqrt{n}\mathbb{Z}_X\left(\int_{\varepsilon}^{1-\varepsilon}g(1,\cdot,\tau)\,dC_{1,x}(\tau|v)\right)\\
&-\sqrt{n}\int_{\mathcal{X}}\int_{\varepsilon}^{1-\varepsilon}\mathbb{Z}_{g_0}(\tau,x)\,dC_{0,x}(\tau|v)\,dF_X(x)\\
&-\sqrt{n}\int_{\mathcal{X}}\mathbb{Z}_{C_0}\left(g(0,x,\cdot),v\right)dF_X(x)\\
&-\sqrt{n}\mathbb{Z}_X\left(\int_{\varepsilon}^{1-\varepsilon}g(0,\cdot,\tau)\,dC_{0,x}(\tau|v)\right)\equiv\mathbb{Z}_{MTE}(v)
\end{aligned}$$

$\mathbb{Z}_{g_d}(\tau,x)$ is a.s. uniformly continuous with respect to (τ,x), jointly in $d\in\mathcal{D}$, $\mathbb{Z}_{C_d}(f_1)$ is continuous with respect to f_1 under the metric λ, jointly in $d\in\mathcal{D}$, and $\mathbb{Z}_X(f_2)$ is continuous with respect to f_2 under the metric λ. By the uniform continuity of $g(d,x,\tau)$ with respect to x under the metric λ, $x\to\mathbb{Z}_{C_d}\left(g(d,x,\cdot),v\right)$ is a.s. uniformly continuous with respect to x. By the uniform continuity of $\int_{\varepsilon}^{1-\varepsilon}g(d,\cdot,\tau)\,dC_{d,\cdot}(\tau,v)$ with respect to (x,v) under the metric λ, $(x,v)\to\mathbb{Z}_X\left(\int_{\varepsilon}^{1-\varepsilon}g(d,\cdot,\tau)\,dC_{d,\cdot}(\tau|v)\right)$ is a.s. uniformly continuous with respect to (x,v). Hence, the desired result follows.

A.7 Proof of Theorem 4

First, I show the distribution of $\vartheta^*(\tau)$. Using the same arguments used in Theorem 1 and Assumption 12, it follows that

$$\sqrt{n}\left(\hat{\vartheta}^*(\tau)-\vartheta(\tau)\right)=F^I(\tau)\left(I-\int_\varepsilon^{1-\varepsilon}D(u)F^I(u)\,du\right)^{-1}\int_\varepsilon^{1-\varepsilon}D(u)F^I(u)C(u)\,\mathbb{G}_n^*\psi(u)\,du$$
$$+F^I(\tau)C(\tau)\,\mathbb{G}_n^*\psi(\tau)+o_P(1) \tag{28}$$

where $f\mapsto\mathbb{G}_n^*[f(W)]\equiv\frac{1}{\sqrt{n}}\sum_{i=1}^n w_i f(W)-\mathbb{E}(f(W))$. Therefore, $\sqrt{n}\left(\hat{\vartheta}^*-\vartheta\right)\Rightarrow\mathbb{Z}_\vartheta^*(\tau)\equiv\sqrt{\omega_0}\mathbb{Z}_\vartheta(\tau)$, a zero-mean Gaussian process with covariance $\omega_0\Sigma_\vartheta(\tau,\tau')$.

Now subtract Equation 27 from Equation 28 to get

$$\sqrt{n}\left(\hat{\vartheta}^*(\tau)-\hat{\vartheta}(\tau)\right)$$
$$=F^I(\tau)\left(I-\int_\varepsilon^{1-\varepsilon}D(u)F^I(u)\,du\right)^{-1}\int_\varepsilon^{1-\varepsilon}D(u)F^I(u)C(u)\frac{1}{\sqrt{n}}\sum_i^n(w_i-1)\psi(u)\,du$$
$$+F^I(\tau)C(\tau)\frac{1}{\sqrt{n}}\sum_i^n(w_i-1)\psi(\tau)+o_P(1) \tag{29}$$

By Assumption 12, it follows that $\sqrt{\frac{n}{\omega_0}}\left(\hat{\vartheta}^*(\tau)-\hat{\vartheta}(\tau)\right)\Rightarrow\mathbb{Z}_\vartheta(\tau)$. By the functional delta method, Theorems 2 and 3, and the previous result, it is straightforward to show that $\sqrt{\frac{n}{\omega_0}}\left(\hat{\Delta}_{QTE}^*(\tau)-\Delta_{QTE}(\tau)\right)\Rightarrow\mathbb{Z}_{QTE}(\tau)$, $\sqrt{\frac{n}{\omega_0}}\left(\hat{\Delta}_{QTE}^*(\tau)-\hat{\Delta}_{QTE}(\tau)\right)\Rightarrow\mathbb{Z}_{QTE}(\tau)$, $\sqrt{\frac{n}{\omega_0}}\left(\hat{\Delta}_{MTE}^*(v)-\Delta_{MTE}^\varepsilon(v)\right)\Rightarrow\mathbb{Z}_{MTE}(v)$ and $\sqrt{\frac{n}{\omega_0}}\left(\hat{\Delta}_{MTE}^*(v)-\hat{\Delta}_{MTE}(v)\right)\Rightarrow\mathbb{Z}_{MTE}(v)$.

A.8 Proof of Corollary 1

This is a particular case of Theorem 4 in Chernozhukov and Hansen (2006), so the result follows immediately.

B Auxiliary Lemmas

B.1 Jacobian Matrix of $M(\tau, \theta_{d,x})$

Let $\tau' \equiv G^{-1}(\tau, \pi(z'); \theta)$. Then, Equation 19 can be written as

$$\begin{aligned}
\nabla_\theta M(\tau,\theta) &= \nabla_\theta G(\tau', \pi(z); \theta) + \nabla_u G(\tau', \pi(z); \theta) \frac{\nabla_\theta G(\tau', \pi(z'); \theta)}{\nabla_u G(\tau', \pi(z'); \theta)} \\
&= \nabla_\theta G(\tau', \pi(z); \theta) \frac{\nabla_u G(\tau', \pi(z); \theta)}{\nabla_u G(\tau', \pi(z); \theta)} + \nabla_u G(\tau', \pi(z); \theta) \frac{\nabla_\theta G(\tau', \pi(z'); \theta)}{\nabla_u G(\tau', \pi(z'); \theta)} \\
&= \nabla_u G(\tau', \pi(z); \theta) \left[\frac{\nabla_\theta G(\tau', \pi(z); \theta)}{\nabla_u G(\tau', \pi(z); \theta)} - \frac{\nabla_\theta G(\tau', \pi(z'); \theta)}{\nabla_u G(\tau', \pi(z'); \theta)} \right] \\
&= \frac{\nabla_u C(\tau', \pi(z); \theta)}{\pi(z)} \left[\frac{\nabla_\theta C(\tau', \pi(z); \theta)}{\nabla_u C(\tau', \pi(z); \theta)} - \frac{\nabla_\theta C(\tau', \pi(z'); \theta)}{\nabla_u C(\tau', \pi(z'); \theta)} \right]
\end{aligned}$$

By the properties of the copula and the propensity score, the term outside the brackets is positive. Moreover, by Lemma 4.1 in Han and Vytlacil (2017) the term in brackets is positive if $\pi(z) < \pi(z')$. Therefore, $\nabla_\theta M(\tau, \theta)$ is positive semidefinite if $\pi(z) < \pi(z')$, and negative semidefinite if $\pi(z) > \pi(z')$.

B.2 Hadamard Derivative

Lemma 3. *Let the operator $\kappa : \ell^\infty(\mathcal{T}) \to \mathbb{R}$ defined by $\kappa(\nu(\cdot)) = \int_0^1 \lambda(\cdot)\nu(\cdot)\, d\cdot$. Define $\kappa(u; h_t) \equiv \int_0^1 \lambda(u)(\nu(u) + t h_t(u))\, du$. As $t \to 0$,*

$$D_{h_t}(t) = \frac{\int_0^1 \lambda(u)(\nu(u) + t h_t(u))\, du - \int_0^1 \lambda(u)\nu(u)\, du}{t} \to D_h$$

where $D_h \equiv \int_0^1 \lambda(u) h(u)\, du$. The convergence holds uniformly in any compact subset of $\mathcal{T}$ for any $h_t : \|h_t - h\|_\infty \to 0$, where $h_t \in \ell^\infty(\mathcal{T})$ and $h \in C(\mathcal{T})$.

Proof.

$$\begin{aligned} D_{h_t}(h_t) &= \frac{\int_0^1 \lambda(u)(\nu(u) + th_t(u))\,du - \int_0^1 \lambda(u)\nu(u)\,du}{t} \\ &= \frac{1}{t}\int_0^1 \lambda(u)\,th_t(u)\,du \to D_h \end{aligned}$$

□

B.3 Solution to the Fredholm Integral Equation

Lemma 4. *Let* $L(\tau) = M_1(\tau)L(\tau) + M_2(\tau) + \int_0^1 M_3(u)L(u)\,du$ *be a Fredholm integral equation of the second kind. Moreover, define* $\tilde{M}_2(\tau) \equiv (I - M_1(\tau))^{-1} M_2(\tau)$ *and* $\tilde{M}_3(\tau) \equiv M_3(\tau)(I - M_1(\tau))^{-1}$. *Let*

(i) $I - M_1(\tau)$ *is invertible* $\forall \tau \in [0,1]$

(ii) $\lim_{n\to\infty}\left[\int_0^1 \tilde{M}_3(u)\,du\right]^n = 0$

Under (i)-(ii), the solution to this equation is given by

$$L(\tau) = \tilde{M}_2(\tau) + (I - M_1(\tau))^{-1}\left(I - \int_0^1 \tilde{M}_3(u)\,du\right)^{-1}\int_0^1 \tilde{M}_3(u)M_2(u)\,du$$

Proof.

$$\begin{aligned} L(\tau) &= M_1(\tau)L(\tau) + M_2(\tau) + \int_0^1 M_3(u)L(u)\,du \\ &= \tilde{M}_2(\tau) + (I - M_1(\tau))^{-1}\int_0^1 M_3(u)L(u)\,du \\ &= \tilde{M}_2(\tau) + (I - M_1(\tau))^{-1}\sum_{n=0}^{\infty}\left[\int_0^1 \tilde{M}_3(u)\,du\right]^n\int_0^1 \tilde{M}_3(u)M_2(u)\,du \\ &+ \lim_{n\to\infty}(I - M_1(\tau))^{-1}\left[\int_0^1 \tilde{M}_3(u)\,du\right]^n\int_0^1 M_3(u)L(u)\,du \\ &= \tilde{M}_2(\tau) + (I - M_1(\tau))^{-1}\left(I - \int_0^1 \tilde{M}_3(u)\,du\right)^{-1}\int_0^1 \tilde{M}_3(u)M_2(u)\,du \end{aligned}$$

where the second equality follows by (i), the third one by iteratively substituting $L(u)$ inside the integral, and the fourth one by (ii) and the following result: define $S \equiv \sum_{n=0}^{\infty} C^n$, and A, B and C be square matrices. Then, $ASB - ACSB = A(I-C)SB = AB$. If $I - C$ is invertible, then $S = (I-C)^{-1}$. Premultiply both sides of the equation by A and postmultiply them by B to obtain the desired result. □

B.4 Argmax Process

Lemma 5. *(Chernozhukov and Hansen, 2006) Suppose that uniformly in π in a compact set Π and for a compact set K (i) $Z_n(\pi)$ is s.t. $Q_n(Z_n(\pi)|\pi) \geq \sup_{z\in K} Q_n(z|\pi) - \epsilon_n$, $\epsilon \searrow 0$; $Z_n(\pi) \in K$ wp $\to 1$, (ii) $Z_\infty(\pi) \equiv \arg\sup_{z\in K} Q_\infty(z|\pi)$ is a uniquely defined continuous process in $\ell^\infty(\Pi)$, (iii) $Q_n(\tau|\tau) \xrightarrow{p} Q_\infty(\tau|\tau)$ in $\ell^\infty(K\times\Pi)$, where $Q_\infty(\tau|\tau)$ is continuous. Then $Z_n(\tau) = Z_\infty(\tau) + o_P(1)$ in $\ell^\infty(\Pi)$*

Proof. See Chernozhukov and Hansen (2006). □

B.5 Stochastic Expansion

Lemma 6. *Under Assumptions 5-11, the following statements hold uniformly over $d \in \mathcal{D}$:*

1. $\sup_{(\beta,\theta,\gamma,\tau)\in\mathcal{B}\times\Theta\times\Gamma\times\mathcal{T}} |\mathbb{E}_n[q(W,\beta,\theta,\gamma,\tau)] - \mathbb{E}[q(W,\beta,\theta,\gamma,\tau)]| = o_P(1)$
2. $\mathbb{G}_n r(W,\beta(\tau),\theta,\gamma,\tau) \Rightarrow \mathbb{Z}_R(\tau)$ *in* $\ell^\infty(\mathcal{T})$, *where* $\mathbb{Z}_R(\tau)$ *is a zero-mean Gaussian process with covariance* $\Sigma_R(\tau,\tau')$ *defined below in the proof. Moreover, for any* $\hat{\vartheta}(\tau)$ *such that* $\sup_{\tau\in\mathcal{T}} \left\|\hat{\vartheta}(\tau) - \vartheta(\tau)\right\| = o_P(1)$, *the following holds:*

$$\sup_{\tau\in\mathcal{T}} \left\|\mathbb{G}_n r\left(W,\hat{\beta}(\tau),\hat{\theta},\hat{\gamma},\tau\right) - \mathbb{G}_n r(W,\beta(\tau),\theta,\gamma,\tau)\right\| = o_P(1)$$

Proof. Let $\mathcal{F}$ be the class of uniformly smooth functions in z with the uniform smoothness order $\omega > \frac{\dim(d,z)}{2}$ and $\|f(\tau',z) - f(\tau,z)\| < \overline{K}(\tau-\tau')^a$ for $\overline{K} > 0$, $a > 0$, $\forall(z,\tau,\tau')\forall f \in \mathcal{F}$. The bracketing number of $\mathcal{F}$, by Corollary 2.7.4 in van der Vaart and Wellner (1996)

satisfies

$$\log N_{[\cdot]}\left(\epsilon, \mathcal{F}, L_2\left(P\right)\right) = O\left(\epsilon^{-\frac{\dim(z)}{\omega}}\right) = O\left(\epsilon^{-2-\delta}\right)$$

for some $\delta < 0$. Therefore, $\mathcal{F}$ is Donsker with a constant envelope. By Corollary 2.7.4, the bracketing number of

$$\mathcal{D} \equiv \{\beta \mapsto X'\beta, \beta \in \mathcal{B}\}$$

satisfies

$$\log N_{[\cdot]}\left(\epsilon, \mathcal{D}, L_2\left(P\right)\right) = O\left(\epsilon^{-\frac{\dim(d,x)}{\omega}}\right) = O\left(\epsilon^{-2-\delta'}\right)$$

for some $\delta' < 0$. Since the indicator function is bounded and monotone, and the density functions $f_{Y|d,Z}\left(y|z\right)$ are bounded by Assumption 10 for $d = 0, 1$, the bracketing number of

$$\mathcal{E} \equiv \{\beta \mapsto \mathbf{1}\left(Y < X'\beta\right), \beta \in \mathcal{B}\}$$

satisfies

$$\log N_{[\cdot]}\left(\epsilon, \mathcal{E}, L_2\left(P\right)\right) = O\left(\epsilon^{-2-\delta'}\right)$$

Since $\mathcal{E}_d$ has a constant envelope, it is Donsker. Now consider the function $G_{d,x}$. By Assumptions 4 and 5, the mean value theorem can be applied to show

$$\left\|G_{d,x}\left(\tau, \pi\left(z,\gamma\right);\theta_d\right) - G_{d,x}\left(\tau', \pi\left(z,\gamma\right);\theta_d\right)\right\| = \left\|\tau - \tau'\right\| \left\| \frac{\partial}{\partial \tau} G_{d,x}\left(\tau'', \pi\left(z,\gamma\right);\theta_d\right)\right\|$$

for some τ'' between τ and τ'. By Assumptions 4 and 5, the second term is bounded $\forall z, \tau'', d$,

so it follows that $G_{d,x} \in \mathcal{F}$.[23] Let $\mathcal{T} \equiv \{\tau \mapsto \tau\}$ and define

$$\mathcal{H} \equiv \{h = (\beta, \theta, \gamma, \tau) \mapsto r(W, \beta, \theta, \gamma, \tau), (\beta, \theta, \gamma) \in \mathcal{B} \times \Theta \times \Gamma\}$$

The first subvector of $\mathcal{H}$ is $\mathcal{E} \times \mathcal{F} - \mathcal{T} \times \mathcal{F}$, the second subvector is $\mathcal{E} \times \mathcal{F} - \mathcal{T} \times \mathcal{F}$, and the third subvector is $\mathcal{F}$. Since $\mathcal{H}$ is Lipschitz over $(\mathcal{T}, \mathcal{F}, \mathcal{E})$, it follows that it is Donsker by Theorem 2.10.6 in van der Vaart and Wellner (1996). Define

$$h \equiv (\beta, \theta, \gamma, \tau) \mapsto \mathbb{G}_n r(W, \beta, \theta, \gamma, \tau)$$

h is Donsker in $\ell^\infty(\mathcal{H})$. Consider the process

$$\tau \mapsto \mathbb{G}_n r(W, \beta, \theta, \gamma, \tau)$$

By the uniform Hölder continuity of $\tau \mapsto (\tau, \beta(\tau))$ in τ with respect to the supremum norm, it is also Donsker in $\ell^\infty(\mathcal{T})$. Hence,

$$\mathbb{G}_n r(W, \beta(\tau), \theta, \gamma, \tau) \Rightarrow \mathbb{Z}_R(\tau)$$

with covariance function $\Sigma_R(\tau, \tau')$. Define ξ as the $L_2(P)$ pseudometric on $\mathcal{H}_d$:

$$\xi\left(\tilde{h}, h\right) \equiv \sqrt{\mathbb{E}\left\|r\left(W, \tilde{\beta}, \tilde{\theta}, \tilde{\gamma}, \tilde{\tau}\right) - r(W, \beta, \theta, \gamma, \tau)\right\|^2}$$

Define $\delta_n \equiv \sup_{\tau \in \mathcal{T}} \xi\left(\tilde{h}(\tau), h(\tau)\right)\Big|_{\tilde{h}(\tau) = \hat{h}(\tau)}$. Since $\hat{\vartheta}(\tau) \xrightarrow{p} \vartheta(\tau)$ uniformly in τ, by

[23] To see this, notice that both $\frac{\partial}{\partial \tau} C_{d,x}(\tau, \pi) \in [0,1]$ and $\pi(\tau) \in [0,1]$. Hence, it suffices to show that $\lim_{\pi \to 1} \frac{\partial}{\partial \tau} G_{d,x}(\tau, \pi) = \lim_{\pi \to 1} C_{d,x}(\tau, \pi) < \infty$, where I have used L'Hôpital rule. Since the derivative is bounded by Assumption 5, the result follows.

Assumption 10, $\delta_n \xrightarrow{p} 0$. Therefore, as $\delta_n \xrightarrow{p} 0$,

$$\begin{aligned}
&\sup_{\tau\in\mathcal{T}} \left\| \mathbb{G}_n r\left(W, \hat{\beta}, \hat{\theta}, \hat{\gamma}, \tau\right) - \mathbb{G}_n r\left(W, \beta, \theta, \gamma, \tau\right) \right\| \\
&\leq \sup_{\substack{\xi\left(\tilde{h},h\right)\leq\delta_n \\ \tilde{h},h\in\mathcal{H}}} \left\| \mathbb{G}_n r\left(W, \hat{\beta}, \hat{\theta}, \hat{\gamma}, \tau\right) - \mathbb{G}_n r\left(W, \beta, \theta, \gamma, \tau\right) \right\| = o_P\left(1\right)
\end{aligned}$$

by stochastic equicontinuity of $h \mapsto \mathbb{G}_n r\left(W, \beta, \theta, \gamma, \tau\right)$, which proves claim 2. To prove claim 1, define

$$\mathcal{A} \equiv \left\{ (\beta, \theta, \gamma, \tau) \mapsto q\left(W, \beta, \theta, \gamma, \tau\right) \right\}$$

By Assumption 6, $\mathcal{A}$ is bounded, and it is also uniformly Lipschitz over $\mathcal{B}\times\Theta\times\Gamma\times\mathcal{T}$, so by Theorem 2.10.6 in van der Vaart and Wellner (1996), $\mathcal{A}$ is Donsker. Hence, the following ULLN holds:

$$\sup_{h\in\mathcal{H}} \left| \mathbb{E}_n q\left(W, \beta, \theta, \gamma, \tau\right) - \mathbb{E} q\left(W, \beta, \theta, \gamma, \tau\right) \right| \xrightarrow{p} 0$$

which gives

$$\sup_{(\beta,\theta,\gamma,\tau)\in\mathcal{B}\times\Theta\times\Gamma\times\mathcal{T}} \left| \mathbb{E}_n q\left(W, \beta, \theta, \gamma, \tau\right) - \mathbb{E} q\left(W, \beta, \theta, \gamma, \tau\right) \right| \xrightarrow{p} 0$$

which implies claim 1. □

C A Generalized Roy Model with Imperfect Information

The Roy model (Roy, 1951) is a useful tool to better understand the economic foundation of triangular models. Several models have enriched the original model by adding a cost function in the individual decision, not imposing parametric assumptions, or dealing with uncertainty at the time the individual's decision to be treated (see, *e.g.* Heckman and Vytlacil, 2005, 2007a; Carneiro et al., 2011; Eisenhauer et al., 2015). However, the models considered by

these authors either assume additive separability of the unobserved heterogeneity, or no uncertainty. The following generalized model with imperfect information accommodates both features.

Let the outcome of an individual be determined by the switching model $Y = (Y_1 - Y_0) D + Y_0$, where Y_d is the potential outcome under treatment status d as defined by equation 1. Individuals also face a cost for being treated, equal to $K \equiv k(Z) + U_K$, where Z is the vector of instruments that includes the covariates X. The cost function is linearly separable in the error term. When individuals know these variables with certainty, they choose to be treated if their net surplus is positive, *i.e.*, if $S \equiv Y_1 - Y_0 - K \geq 0$.

Instead, assume that individuals do not know exactly the value of the outcome under each treatment nor its cost. Their information set is composed of the vector of instruments Z and a variable V that is correlated with all the other unobservable variables (U_1, U_0, U_K). V is not observed by the econometrician and is normalized to be uniformly distributed on the unit interval. In this setting, individuals would consider the *expected* net surplus to decide whether or not to receive the treatment:

$$\mathbb{E}[S|Z,V] = \mathbb{E}[g(1,X,U_1) - g(0,X,U_0)|Z,V] - k(Z) - \mathbb{E}[U_K|Z,V]$$

If the net surplus is positive, then the individual would choose to be treated. Defining $\mu_d(X,V) \equiv \mathbb{E}[g(d,X,U_d)|Z,V]$ for $d = 0, 1$, the selection equation can be written as

$$D = \mathbf{1}(\mu_1(X,V) - \mu_0(X,V) - k(Z) - \mathbb{E}[U_k|Z,V] \geq 0)$$

In general, this selection rule cannot be written in terms of the propensity score. In that case, the method of instrumental variables does not identify all the relevant effects (Heckman and Vytlacil, 2005, 2007b). However, if the expected net surplus is monotone in V, then it is possible to rewrite the selection equation in terms of the propensity score as

$$D = \mathbf{1}(V \leq \pi(Z))$$

where $\pi(Z) \equiv \{p : \mu_1(X,p) - \mu_0(X,p) - k(Z) - \mathbb{E}[U_k|Z,p] = 0\}$. Note that the net surplus in this case depends on two terms, the MTE, $\mu_1(X,V) - \mu_0(X,V)$, and the expected cost, $k(Z) + \mathbb{E}[U_k|Z,V]$. Thus, the decision to be treated depends on which of these two terms is the largest.

To get some insight, consider the case in which the net surplus is decreasing in V. An individual with a small value of V would predict that the expected net surplus from being treated is large, and would choose to be treated. As V decreases, one would eventually attain the value that makes the net surplus zero, *i.e.*, when V equals the propensity score. An individual with such value of V would be indifferent between being treated or not, and if it were smaller than the propensity score, the expected net surplus would be negative, and the individual would choose not to be treated.

This covers several interesting cases. For example, when the treatment has an expected positive effect for all values of X and there is either rank invariance or rank similarity. In other words, the distribution of potential outcomes for the treated dominates the distribution of potential outcomes for the untreated, conditional on any value of the covariates, and the expected value of the unobservables U_1 and U_0 conditional on V is the same. Alternatively, even if $g(1,X,u) - g(0,X,u) = 0$ for all possible values of u, it is possible to obtain a net surplus from the treatment if the difference between the expected value of U_1 and U_0 given V, is large enough.

This framework highlights the advantages of using copulas to model the treatment effect and shows how the rank invariance assumption can mask some effects of interest. The copula $C_{d,x}$ reflects the amount of information that an individual has about its potential outcome under treatment d. A negative correlation between U_d and V implies that the individual would rank higher in the distribution of potential outcomes under treatment d, the lower the value of V is. A more negative correlation of the copula under treatment status d relative to d' implies that individuals with low values of V would tend to rank higher in the distribution of potential outcomes of treatment d.

D Alternative Identification Conditions

D.1 Real Analytic Copulas

Consider the following assumption:

Assumption 13. *$\forall \tau \in (0,1)$, the functions $\pi \to C_{0,x}(\tau, \pi)$ and $\pi \to C_{1,x}(\tau, \pi)$ are real analytic on the unit interval.*

Assumption 13 is a shape restriction. It implies that both copulas, as well as all their derivatives are continuous with respect to the second argument. Moreover, the distribution of the outcome conditional on each treatment status (and on $X = x$) is real analytic with respect to the propensity score.

The class of real analytic copulas is large, so Assumption 13 could also be interpreted as a parametric assumption for a flexible family of copulas. Most parametric copulas are based on analytic functions, such as polynomials, power functions or exponentials, and are therefore real analytic.[24] This includes many copulas that satisfy Assumption 5, such as those listed in Han and Vytlacil (2017). However, it is useful to consider which copulas are ruled out by it. The most prominent type are those with kinks or piece-wisely defined, such that the density of the copula has different left and right derivatives with respect to the second element, but there may be infinitely smooth copulas that are not real analytic everywhere.[25] These include the Fréchet-Hoeffding bounds, which correspond to perfect positive and negative correlation and therefore do not have a well defined density. Unfortunately, the real analyticity condition cannot be verified, although it is possible to study partial identification of the SQF without it. One could use the Fréchet-Hoeffding Bounds, as in Arellano and Bonhomme (2017a),

[24] See, *e.g.*, Nadarajah et al. (2017).

[25] For example, suppose that for $v = 0.5$, a copula has the following functional form:

$$C\left(u, \frac{1}{2}\right) = \frac{1}{2}\left[1 + \frac{e^{-\frac{1}{3u-2}}}{e^{-\frac{1}{3u-2}} + e^{-\frac{1}{3-3u}}}\mathbf{1}\left(u \geq \frac{2}{3}\right) + \frac{e^{-\frac{1}{1-3u}}}{e^{-\frac{1}{1-3u}} + e^{-\frac{1}{3u}}}\mathbf{1}\left(u \leq \frac{1}{3}\right)\right]$$

It is easy to verify that this function is within the Fréchet bounds, and it is smooth and monotonic everywhere on the unit interval, but it is not real analytic at 1/3 and 2/3.

or follow Chesher (2005) or Jun et al. (2011) for partial identification under even weaker assumptions.

To further assess the strength of Assumption 13, consider the Bernstein copula, which depends on Bernstein polynomials.[26] Bernstein (1912) showed that these polynomials can arbitrarily approximate any bounded continuous function on the unit interval, a result known as Stone-Weierstrass approximation theorem. Lemma 1 in Sancetta and Satchell (2004) strengthened this result by showing that the set of Bernstein polynomials is dense in the space of bounded continuous functions in the k-dimensional hypercube $[0,1]^k$.

This formal argument implies that Bernstein copulas can approximate any arbitrary continuous copula that has a well-defined density. Hence, because real analytic functions can be expressed as polynomials of infinite order, an alternative identification result can be achieved using Assumption 13:

Proposition 2. *Let Assumptions 1 to 4 and 13 hold, and $x \in \mathcal{X}$. Then, the functions $(\tau,\pi) \to G_{1,x}(\tau,\pi)$, $(\tau,\pi) \to G_{0,x}(\tau,\pi)$, and $\tau \to g(d,x,\tau)$ for $d=0,1$ are nonparametrically identified.*

Proof. Let $G_{0,x}$ and $\tilde{G}_{0,x}$ satisfy Equation 4, and $\pi_1, \pi_2 \in \mathcal{P}_x$. Then,

$$G_{0,x}\left(G_{0,x}^{-1}(\tau,\pi_2),\pi_1\right) - \tilde{G}_{0,x}\left(\tilde{G}_{0,x}^{-1}(\tau,\pi_2),\pi_1\right) = 0 \forall (\pi_1,\pi_2) \in \mathcal{P}_x \times \mathcal{P}_x$$

Hence, $\forall \tau \in (0,1)$, $(\pi_1,\pi_2) \to G_{0,x}\left(G_{0,x}^{-1}(\tau,\pi_2),\pi_1\right) - \tilde{G}_{0,x}\left(\tilde{G}_{0,x}^{-1}(\tau,\pi_2),\pi_1\right)=0$. $C_{0,x}$ is real analytic by Assumption 13, so $G_{0,x}$ is also real analytic, and hence the composition is real analytic (see *e.g.*, Krantz and Parks, 2002). Because it is zero on a product of two open neighborhoods, it is zero everywhere on $(0,1)\times(0,1)$(Fox et al., 2012). To see this, note that if $h(z)$ is real analytic on an open interval $\mathcal{Z}$, and $h(z_0)=0 \forall z_0 \in \mathcal{Z}_0 \subset \mathcal{Z}$, then the l-th order derivative $h^l(z_0)=0$ for any $z_0 \in \mathcal{Z}_0$. Set $\overline{\mathcal{Z}} = \mathcal{Z} \cap \left\{z : h^l(z) = 0 \, for \, l = 0,1,2,...\right\}$. By continuity, $\overline{\mathcal{Z}}$ is closed in the relative topology of $\mathcal{Z}$, and by assumption on $\mathcal{Z}$, $\overline{\mathcal{Z}}$ is open. Hence, by the connectedness of $\mathcal{Z}$, $\overline{\mathcal{Z}} = \mathcal{Z}$, concluding the proof.

[26] A Bernstein polynomial is given by $\sum_{m=0}^{M} a_{m,M}\left(\frac{M}{m}\right) x^m (1-x)^{M-m}$, where $a_{m,M}$, $m=0,...,M$ are the polynomial coefficients.

Taking limits at $\pi_2 = 0$ yields:

$$\lim_{\pi_2 \to 0} G_{0,x}\left(G_{0,x}^{-1}(\tau, \pi_2), \pi_1\right) - \tilde{G}_{0,x}\left(\tilde{G}_{0,x}^{-1}(\tau, \pi_2), \pi_1\right) = G_{0,x}(\tau, \pi_1) - \tilde{G}_{0,x}(\tau, \pi_1) = 0 \forall \pi_1 \in (0,1)$$

Hence, $G_{0,x}(\tau, \pi_1)$ and $\tilde{G}_{0,x}(\tau, \pi_1)$ coincide on $(0,1) \times (0,1)$. Consequently, $G_{0,x}$ is identified, and so are $C_{0,x}$ and $g(0, x, u)$. By a parallel argument, using Equation 3 and taking limits at $\pi_2 = 1$, $G_{1,x}$, $C_{1,x}$, and $g(1, x, u)$ are identified.

□

The identification relies on an extrapolation, similarly to the one under Assumption 5. Moreover, it is related to those used on the shape of the MTE when the support of the instrument (or the propensity score) is a finite set of points. *E.g.*, Brinch et al. (2017) modeled the MTE as a polynomial with respect to the propensity score, extrapolating the MTE outside the support of the propensity score using the polynomial. If the number of points of support approaches infinity, the order of the polynomial could also increase to infinity. Hence, Assumption 13 could be considered a limiting case when the support approximates an open interval.

D.2 Comparison with Alternative Methods

An important benchmark in the literature of triangular models with a binary treatment is LIV. Recent works (Carneiro and Lee, 2009; Jun et al., 2016) have studied the identification of distributional effects, extending the original contributions that focused on the mean effect (Heckman and Vytlacil, 1999). The model defined by Equations 1-2 is closely related to the model in Carneiro and Lee (2009), and its identification conditions can be represented in terms of the copula and the distribution of potential outcomes. In particular, the two

equations of Theorem 1 in Carneiro and Lee (2009) can be written as

$$\frac{\partial}{\partial p} C_{0,x}\left(F_{Y_0^*}\left(y|x\right), p\right)\Big|_{p=\pi(z)} = F_{Y|D=0,Z}\left(y|z\right) - \left(1-\pi\left(z\right)\right)\frac{\partial}{\partial \pi\left(z\right)} F_{Y|D=0,Z}\left(y|z\right) \tag{30}$$

$$\frac{\partial}{\partial p} C_{1,x}\left(F_{Y_1^*}\left(y|x\right), p\right)\Big|_{p=\pi(z)} = F_{Y|D=1,Z}\left(y|z\right) + \pi\left(z\right)\frac{\partial}{\partial \pi\left(z\right)} F_{Y|D=1,Z}\left(y|z\right) \tag{31}$$

Without extra assumptions, LIV identifies the left had side of Equations 30-31 only over the support $\mathcal{P}_x$. Thus, it is not possible to separately identify the copula and the distribution of potential outcomes. To achieve that identification result, one would need to invoke the identification at infinity argument, *i.e.*, $\mathcal{P}_x = [0,1]$. The key difference with respect to the identification result in Proposition 1 is Assumption 13, which allows the extrapolation of the identification region from $\mathcal{P}_x$ to the whole unit interval.

The literature has already considered a variety of alternative assumptions that achieve this extrapolation, some of which are stronger than Assumption 13. For example, if the disturbances have a known parametric distribution, then the shape of the MTE depends on these distributions, allowing the extrapolation from $\mathcal{P}_x$ to the unit interval.[27] Another possibility is to relax Assumption 1 to allow for full independence between the unobservables and both the instrument and the covariates, *i.e.*, (U_0, U_1, V) are jointly independent of (Z_1, X). Then, one could use variation in X as a source of identification: if the support of $\pi\left(Z\right)$, denoted by $\mathcal{P}$, equals the unit interval, then one can invoke the identification at infinity argument. Moreover, if at least one of the covariates is continuously distributed, one could achieve identification with a discrete instrument by using the same extrapolation argument in Proposition 1, even if the instrument is discrete. This assumption, however, imposes severe restrictions on the amount of heterogeneity that can be displayed by the model. In particular, it requires the copulas to be invariant with respect to the covariates that are used as a source of exogenous variation, ruling out differences in selection for individuals with different covariates. Thus, if the SQF was additively separable between U_D and X, the MTE would also be additively separable, and its shape would be constant up to the intercept

[27] See *e.g.*, Cornelissen et al. (2017) for the normally distributed model (Heckman, 1976).

with respect to the covariates.[28]

More recently, shape restrictions have been directly imposed on the MTE. For example, Brinch et al. (2017) consider a separable model in which the term of the MTE that depends on the unobservables can be expressed as a linear combination of parameters. Similarly, Mogstad et al. (2018) consider a nonseparable model in which the MTE can be expressed as a linear basis. They propose two kinds of basis: one consisting of Bernstein polynomials, and another one piece-wise constant. The former model and the latter with the Bernstein polynomial basis are real analytic with respect to the propensity score. Hence, because real analyticity is maintained under the integral sign, the underlying copula is also real analytic, making them particular cases of the model considered in this paper.

A different approach is considered in the IVQR model (Chernozhukov and Hansen, 2005, 2006). Importantly, the IVQR model is general enough to allow the treatment to be either discrete or continuous.[29] However, the identification result of the IVQR requires either rank invariance or rank similarity to hold. When this assumption is dropped, and using this paper's notation, equation 2.6 from Theorem 1 in Chernozhukov and Hansen (2005) can be written as:

$$\mathbb{P}\left(Y \leq g\left(D, X, \tau\right) | Z\right) = \tau - C_{0,x}\left(F_{Y_0^*}\left(g\left(0, X, \tau\right)\right), \pi\left(Z\right)\right) + C_{1,x}\left(F_{Y_1^*}\left(g\left(1, X, \tau\right)\right), \pi\left(Z\right)\right) \tag{32}$$

Hence, under rank dissimilarity, the moment $\mathbb{P}\left(Y \leq g\left(D, X, \tau\right) | Z\right) \neq \tau$, and therefore it does not point identify the SQF process. The cost of not requiring rank similarity is the specification of the selection equation (Equation 2) and the copula. Nevertheless, it is still

[28] This assumption is strong enough to achieve identification of the MTE even when the instrument is binary. See Kitagawa (2021) for further discussion on the identified sets when both the instrument and the treatment are binary.

[29] When the instrument is binary, the IVQR estimator is closely connected with the LQTE estimator. See Wüthrich (2019b) for further details.

possible to combine Equation 32 with Frechét-Hoeffding bounds to obtain set identification:

$$\begin{aligned}\tau+\min\left\{F_{Y_0^*}\left(g\left(0,X,\tau\right)\right),\pi\left(Z\right)\right\}-\max\left\{F_{Y_1^*}\left(g_1\left(X,\tau\right)\right)-\pi\left(Z\right),0\right\}\leq\\ \mathbb{P}\left(Y\leq g_D\left(X,\tau\right)|Z\right)\leq\\ \tau+\min\left\{F_{Y_1^*}\left(g_1\left(X,\tau\right)\right),\pi\left(Z\right)\right\}-\max\left\{F_{Y_0^*}\left(g_0\left(X,\tau\right)\right)-\pi\left(Z\right),0\right\}\end{aligned}\tag{33}$$

E Bernstein Copula

Assumption 5 can be relaxed by considering Bernstein copulas which, as shown by Lemma 1 in Sancetta and Satchell (2004), they can approximate arbitrary parametric copulas.[30] The cumulative distribution of this copula is given by

$$C\left(u,v\right)=\sum_{m_u=0}^{M}\sum_{m_v=0}^{M}\alpha\left(\frac{m_u}{M},\frac{m_v}{M}\right)\overline{P}_{m_u,M}\left(u\right)\overline{P}_{m_v,M}\left(v\right)$$

where M is the order of the copula, and $\overline{P}_{m,M}\left(u\right)=\binom{M}{m}u^m\left(1-u\right)^{M-m}$. The density of this copula has a similar form, making it is very convenient to implement.[31] Because the $\overline{P}_{m,M}$ terms are known, the estimation of the copula amounts to the estimation of the α coefficients. Let A_j denote the matrix that stacks the $\alpha\left(\frac{m_u}{M},\frac{m_v}{M}\right)$ parameters for $j=0,1$. For a given order M, the based on this copula is a particular case of the one presented in Section 3, substituting θ_j by A_j.[32]

The implementation of the estimator is more complicated than in the parametric case: the number of parameters equals $\left(M-1\right)^2$, so it grows at a faster rate than the order of the copula. Hence, grid search methods are subject to the curse of dimensionality. An alternative to these is a sequential random search using a property of Bernstein copulas that allows to

[30]Note that the Bernstein copulas are not the most appropriate to model extreme tail behavior, as the copula copula and its approximand converge to an arbitrary limit at a different speeds. Regardless, it can capture increasing dependence as one moves to the tails. See Sancetta and Satchell (2004) for further details.

[31]For completeness, define $\eta\left(\frac{m_u}{M},\frac{m_v}{M}\right)=\alpha\left(\frac{m_u+1}{M},\frac{m_v+1}{M}\right)-\alpha\left(\frac{m_u+1}{M},\frac{m_v}{M}\right)-\alpha\left(\frac{m_u}{M},\frac{m_v+1}{M}\right)+\alpha\left(\frac{m_u}{M},\frac{m_v}{M}\right)$. The density is given by $c\left(u,v\right)=\sum_{m_u=0}^{M-1}\sum_{m_v=0}^{M-1}\eta\left(\frac{m_u}{M},\frac{m_v}{M}\right)\overline{P}_{m_u,M}\left(u\right)\overline{P}_{m_v,M}\left(v\right)M^2$.

[32]In principle it would be possible to allow M to grow to infinity as N increases. However, the asymptotic distribution of such estimator may not coincide with the one stated in Theorem 1.

express any Bernstein copula of order M_1 as a Bernstein copula of order $M_2 > M_1$.[33] The algorithm is as follows:

1. Given an order M, fix one value of the copula, denoted by A_M^0.

2. Compute the objective function at randomly chosen point in the neighborhood of A_M^0, A_M^*.[34]

3. If the objective function decreases, repeat step 2 replacing A_M^0 by A_M^*; otherwise, repeat step 2 until a value of A that decreases the objective function is found, or the maximum number of iterations without an improvement is reached.

4. Denote the estimated copula by $\hat{A}_M$. Then, for the copula of order $M+1$, use $A_{M+1}^0 \equiv A_{M+1} = \overline{P}_{M+1} C_M^{-1} \overline{P}'_{M+1}$ as the starting initial value of the parameter for the copula of order $M+1$.

5. Stop when the one obtains the estimates of the highest order copula considered.

This is a sequential estimator that requires solving the linear program once per iteration. This estimator has two main advantages: it can combine the fast grid search over $[0, 0.5]$ for the copula of order 2, and the initial candidate for the optimum makes increasing the order not excessively burdensome. However, the amount of correlation that the Bernstein copula can display is limited by the order. Hence, if the correlation of the unobservables is high in absolute value, starting with a copula of a relatively high order may be advisable.[35]

[33] In particular, let C_1 denote the copula of order M_1, A_2 denote the matrix with the α parameters of the copula of order M_2, and $\overline{P}_2 \equiv \left(\overline{P}_{m,M_2}\left(\frac{0}{M_2+1}\right), ..., \overline{P}m, M_2\left(\frac{M_2+1}{M_2+1}\right)\right)'$. Then, $A_2 = \overline{P}_2 C_1^{-1} \overline{P}'_2$.

[34] In particular, the point is selected with a Markov chain sampling for doubly stochastic matrices. Define B as the $(M+1) \times (M+1)$ matrix whose (i,j) element is given by $\eta\left(\frac{i}{M}, \frac{j}{M}\right)$. First, pick two columns and two rows at random and denote the matrix formed by their intersection by $\overline{B}$. Draw a random number, ϵ, uniformly from $\left(-\underline{b}, \overline{b}\right)$, where $\underline{b}$ denotes the minimum element of $\overline{B}$. Add ϵ to the diagonal elements and subtract it from the off-diagonal elements, replacing the elements originally selected from matrix B. Apply the inverse mapping from B to A, obtaining the randomly chosen neighbor of the original A matrix.

[35] An initial value of the parameter can be obtained by doing a grid search that interpolates the value of all parameters of the Bernstein copulas of a given order with the minimum and maximum possible amount of correlation for that order.

Finally, one should bear in mind that the random search algorithm does not guarantee that the estimator is the minimizer of the objective function.[36]

F Additional Treatment Effects

Similarly to the MTE, it is possible to express the TUT and the TT in terms of the SQF and the copula:

$$\Delta_{TUT}(z) = \int_0^1 g(1,x,u_1)\, dG'_{0,x}(u_1, \pi(z)) - \int_0^1 g(0,x,u_0)\, dG_{0,x}(u_0, \pi(z)) \tag{34}$$

$$\Delta_{TT}(z) = \int_0^1 g(1,x,u_1)\, dG_{1,x}(u_1, \pi(z)) - \int_0^1 g(0,x,u_0)\, dG'_{1,x}(u_0, \pi(z)) \tag{35}$$

where $G'_{0,x}(\tau, \pi(z)) \equiv \mathbb{P}(U_1 \leq \tau | D = 0, z)$, and $G'_{1,x}(\tau, \pi(z)) \equiv \mathbb{P}(U_0 \leq \tau | D = 1, z)$. These two quantities, along with the propensity score, determine the ATE:

$$\Delta_{ATE}(z) = \Delta_{TUT}(z)(1 - \pi(z)) + \Delta_{TT}(z)\pi(z) = \int_0^1 (g(1,x,u) - g(0,x,u))\, du \tag{36}$$

To obtain the unconditional counterparts of these treatment effects, simply integrate them over the distribution of Z: $ATE = \int_{\mathcal{Z}} ATE(z)\, dF_Z(z)$, $TUT = \int_{\mathcal{Z}} TUT(z)\, dF_Z(z)$, and $TT = \int_{\mathcal{Z}} TT(z)\, dF_Z(z)$.

Regarding the estimation, it can be done using the sample analog of Equations 34-36:

$$\hat{\Delta}_{TUT}(z_i) = \int_{\varepsilon}^{1-\varepsilon} x_i' \hat{\beta}_1(\tau)\, dG'_{0,x}\left(\tau, \hat{\pi}(z_i); \hat{\theta}_0\right) - \int_{\varepsilon}^{1-\varepsilon} x_i' \hat{\beta}_0(\tau)\, dG_{0,x}\left(\tau, \hat{\pi}(z_i); \hat{\theta}_0\right) \tag{37}$$

[36] As such, the properties of the estimator obtained with this algorithm may be slightly different from those presented in Section 3. Studying the properties of this estimator are beyond the scope of this paper.

$$\hat{\Delta}_{TT}(z_i) = \int_{\varepsilon}^{1-\varepsilon} x_i'\hat{\beta}_1(\tau)\, d\hat{G}_{1,x,i,\tau} \equiv G_{1,x}\left(\tau, \hat{\pi}(z_i); \hat{\theta}_1\right) - \int_{\varepsilon}^{1-\varepsilon} x_i'\hat{\beta}_0(\tau)\, dG'_{1,x}\left(\tau, \hat{\pi}(z_i); \hat{\theta}_1\right) \tag{38}$$

$$\hat{\Delta}_{ATE}(z_i) = \int_{\varepsilon}^{1-\varepsilon} x_i'\left(\hat{\beta}_1(\tau) - \hat{\beta}_0(\tau)\right) d\tau \tag{39}$$

Finally, the unconditional treatment effects can be obtained by taking the average over $i = 1, ..., N$.

G Multivalued Treatment

Consider a model in which the treatment can take J distinct values:

$$\begin{aligned} Y &= g(D, X, U_D) \\ D &= \sum_{j=1}^{J} \mathbf{1}\left(\sum_{h=j}^{J} \pi_h(Z) - V > 0\right) \end{aligned}$$

where π_j is the propensity score of treatment $j = 0, ..., J-1$ and $\pi_0 = 1 - \sum_{j=1}^{J} \pi_j$. This corresponds to an ordered choice model, and the vector of unobservables has the same dimension as the number of distinct treatment status. Let $\pi(Z) \equiv [\pi_1(Z), ..., \pi_J(Z)]'$. Then, the conditional copulas are given by

$$G_{0,x}(\tau, \pi(z)) = \frac{\tau - C_{0,x}\left(\tau, \sum_{h=1}^{J} \pi_h(z)\right)}{1 - \sum_{h=1}^{J} \pi_h(z)}$$

$$G_{j,x}(\tau, \pi(z)) = \frac{C_{j,x}\left(\tau, \sum_{h=j}^{J} \pi_h(z)\right) - C_{j,x}\left(\tau, \sum_{h=j+1}^{J} \pi_h(z)\right)}{\pi_j(z)}$$

where $j = 1, ..., J-1$. Using these equations, it is straightforward to adapt the estimation method presented in Section 3, by firstly estimating the propensity score for each treatment

status, and then applying RQR using the conditional copulas $G_{j,x}$. Note however, that a richer model that allows for a multidimensional vector V has been considered by, *e.g.*, Lee and Salanié (2018) or Heckman and Pinto (2018).

H Monte Carlo

The finite sample performance of the estimator is shown in the following Monte Carlo exercise. The data generating process is as follows:

$$y_i = \beta_{d_i,1}\left(\tau_{d_i,i}\right) + x_i\beta_{d_i,2}\left(u_{d_i,i}\right) \tag{40}$$

$$d_i = \mathbf{1}\left(\gamma_1 + x_i\gamma_2 + z_i\gamma_3 + \Lambda^{-1}\left(v_i\right) > 0\right) \tag{41}$$

$$u_{0,i}, u_{1,i}, v_i | z_i \sim Gaussian\left(\Sigma\right) \tag{42}$$

where $\beta_0\left(\tau\right) = \left[\Phi^{-1}\left(\tau\right) - 2, 1 + \frac{\exp(2\tau)}{1+\tau}\right]$, $\beta_1\left(\tau\right) = \left[\tan\left(\tau - 0.5\right) + \Phi^{-1}\left(\tau\right) + 2, \frac{\exp(2\tau)+1}{1+\tau} + 2\tau\right]$, $\gamma = \left(-2, 0.4, 2\right)'$, Σ is a symmetric correlation matrix with unit diagonal, and off diagonal $\Sigma_{12} = 0$, $\Sigma_{13} = 0.5$, and $\Sigma_{23} = 0.25$ elements, $x_i \sim U\left(1, 2\right)$, $z_i \sim U\left(0, 1\right)$, $\Phi\left(\cdot\right)$ is the cdf of the standard normal distribution, and $\Lambda\left(\cdot\right)$ is the cdf of the logistic distribution. The experiment consists of $R = 500$ repetitions, with a sample size of $N = 2000$.

I compute the estimates of the two quantile processes using the method described in this paper using a variety of copulas: the correctly specified copula (Gaussian), a misspecified copula (Clayton), Bernstein copulas of orders 2 through 6, and the true copula, *i.e.* as if the true copula was known. On top of those, I compute the estimates of the IVQR estimator.

Table 3 reports the values of the objective function for each specification of the RQR estimator. Among those that depend on one parameter, the lowest value corresponds to the correctly specified copula. For the estimator based on the Bernstein copula to achieve a smaller value of the objective function, the order needs to be increased to 6 and 3 for the treatment and control groups, respectively. Hence, the number of free parameters of the copula equals 16 and 4, respectively. For a given sample size, increasing the order of

the Bernstein copula results in overfitting of the objective function. This can be seen in Table 4, which displays the average distance across repetitions between the true copula and the estimated ones. As expected, both the mean and maximum distance is smallest for the correctly specified copula. Moreover, note that the Bernstein copula does a better job than the Clayton copula even if its order is small.

Table 3: Objective Function, Baseline

Copula	Gau	Cla	Ber(2)	Ber(3)	Ber(4)	Ber(5)	Ber(6)
Eq (16)	0.101	0.114	1.317	0.590	0.291	0.155	0.089
Eq (18)	0.110	0.121	0.188	0.049	0.013	0.004	0.002

Notes: Gau, Cla and Ber(X) stand for Gaussian, Clayton and Bernstein copula of order X.

Table 4: Estimated Copula, Baseline

Copula	Gau	Cla	Ber(2)	Ber(5)	Ber(8)	Ber(5)	Ber(6)
Mean (C_1)	0.008	0.014	0.013	0.009	0.008	0.009	0.009
Sup (C_1)	0.018	0.036	0.024	0.022	0.023	0.024	0.025
Mean (C_0)	0.010	0.014	0.008	0.009	0.009	0.009	0.009
Sup (C_0)	0.020	0.036	0.018	0.021	0.021	0.022	0.022

Notes: Gau, Cla and Ber(X) stand for Gaussian, Clayton and Bernstein copula of order X; mean (C_D) and sup (C_D) respectively denote the mean and supremum distance across quantiles between the estimated copula and the true copula, averaged across repetitions, for $D = 0, 1$.

The difference in the precision of the estimation of the copula is reflected in the estimates of β (Table 5): with the misspecified parametric copula, the RQR estimates display a small bias, and with the Bernstein copula, this bias diminishes as the order increases. Despite that, even the RQR estimates with a incorrectly specified copula perform better than IVQR, which suffers from two sources of misspecification: the rank similarity assumption, and the interaction effect between the treatment and the covariate.[37]

In terms of the dispersion of the estimates, the results are the opposite, as the IVQR estimates have the smallest interquantile range (IQR). This is explained by the the number

[37]Although the IVQR estimator allows for such interactions, the standard approach is to use the basic linear-in-parameters model, which depends on $\dim(X) + \dim(D)$ parameters. Because of the grid search algorithm employed by the estimator, this is convenient from a computational point of view. See Chernozhukov and Hansen (2006) for further details.

Table 5: Quantile Regression Coefficients, Baseline

		$\beta_{1,1}$						$\beta_{0,1}$					
		τ						τ					
		0.1	0.25	0.5	0.75	0.9	Mean	0.1	0.25	0.5	0.75	0.9	Mean
	Gau	-0.03	-0.05	-0.02	0.00	-0.01	0.02	0.04	0.06	0.01	0.02	-0.01	0.04
	Cla	-0.05	-0.13	-0.20	-0.32	-0.31	0.19	0.08	0.06	-0.12	-0.39	-0.55	0.24
	Ber(2)	-0.01	-0.04	-0.10	-0.23	-0.18	0.11	-0.22	-0.28	-0.35	-0.33	-0.46	0.33
	Ber(3)	-0.01	-0.05	-0.04	-0.15	-0.19	0.08	-0.13	-0.10	-0.13	-0.18	-0.34	0.16
MD	Ber(4)	-0.01	-0.05	-0.05	-0.13	-0.19	0.08	-0.12	-0.06	-0.06	-0.09	-0.25	0.11
	Ber(5)	-0.01	-0.04	-0.05	-0.14	-0.19	0.08	-0.09	-0.03	-0.04	-0.07	-0.26	0.10
	Ber(6)	-0.01	-0.04	-0.04	-0.14	-0.21	0.08	-0.08	-0.03	-0.03	-0.09	-0.26	0.09
	True	-0.01	-0.04	0.03	0.05	0.02	0.02	0.03	0.07	0.00	-0.05	-0.01	0.04
	IVQR	-0.78	-1.00	-1.53	-2.06	-2.49	1.56	0.01	1.18	1.78	2.57	4.10	1.87
	Gau	1.31	1.72	2.54	2.90	3.37	2.41	2.13	2.61	3.70	4.64	5.04	3.70
	Cla	1.36	1.72	2.30	2.54	2.83	2.23	2.09	2.64	3.56	3.99	4.40	3.40
	Ber(2)	1.22	1.54	2.21	2.68	2.96	2.19	1.83	2.35	3.18	3.97	4.64	3.24
	Ber(3)	1.26	1.62	2.39	2.94	2.92	2.29	1.94	2.40	3.45	4.08	4.80	3.39
IQR	Ber(4)	1.26	1.70	2.37	2.84	2.94	2.31	2.00	2.54	3.60	4.13	4.88	3.44
	Ber(5)	1.26	1.70	2.43	2.81	2.94	2.32	1.96	2.51	3.65	4.22	4.96	3.48
	Ber(6)	1.25	1.69	2.45	2.84	2.94	2.33	2.00	2.55	3.70	4.20	4.91	3.50
	True	1.25	1.67	2.37	2.82	3.06	2.35	1.85	2.46	3.51	4.04	4.59	3.48
	IVQR	0.82	0.90	1.09	1.31	1.47	1.18	1.83	1.87	2.65	3.60	5.16	3.02

		$\beta_{1.2}$						$\beta_{0.2}$					
		τ						τ					
		0.1	0.25	0.5	0.75	0.9	Mean	0.1	0.25	0.5	0.75	0.9	Mean
	Gau	0.00	-0.02	-0.03	-0.03	-0.01	0.02	0.01	0.03	0.01	0.03	0.02	0.03
	Cla	-0.05	-0.10	-0.10	-0.03	0.02	0.07	-0.17	-0.11	-0.02	0.00	-0.02	0.07
	Ber(2)	0.01	-0.01	-0.02	0.01	0.04	0.02	0.06	0.05	0.02	0.01	-0.03	0.04
	Ber(3)	0.01	-0.02	-0.02	-0.01	0.03	0.02	0.06	0.05	0.03	0.03	-0.02	0.04
MD	Ber(4)	0.00	-0.02	-0.03	-0.01	0.03	0.02	0.07	0.04	0.03	0.04	0.00	0.04
	Ber(5)	0.00	-0.02	-0.03	-0.01	0.03	0.02	0.07	0.04	0.03	0.03	0.00	0.04
	Ber(6)	0.00	-0.03	-0.03	-0.01	0.02	0.02	0.07	0.04	0.03	0.03	0.01	0.04
	True	-0.01	-0.03	-0.03	-0.03	-0.02	0.02	0.01	0.05	0.04	0.02	0.02	0.03
	IVQR	0.43	0.50	0.47	0.44	0.31	0.44	-0.63	-0.89	-0.86	-0.75	-0.42	0.75
	Gau	1.06	1.09	1.24	1.48	1.43	2.41	1.45	1.44	1.65	1.97	1.93	3.70
	Cla	1.15	1.09	1.27	1.50	1.46	2.23	1.81	1.43	1.68	1.96	2.02	3.40
	Ber(2)	1.02	1.07	1.25	1.49	1.45	2.19	1.39	1.42	1.66	2.08	1.97	3.24
	Ber(3)	1.01	1.06	1.23	1.50	1.44	2.29	1.38	1.43	1.62	2.08	1.99	3.39
IQR	Ber(4)	1.03	1.06	1.25	1.50	1.45	2.31	1.39	1.40	1.61	2.06	2.04	3.44
	Ber(5)	1.04	1.07	1.24	1.48	1.45	2.32	1.39	1.41	1.63	2.06	2.04	3.48
	Ber(6)	1.04	1.07	1.26	1.48	1.45	2.33	1.39	1.43	1.63	2.06	2.02	3.50
	True	1.00	0.96	1.17	1.39	1.37	2.35	1.46	1.35	1.70	2.14	1.97	3.48
	IVQR	0.83	0.91	1.12	1.34	1.48	1.20	1.45	1.52	1.65	1.83	2.04	1.76

Notes: Gau, Cla and Ber(X) stand for Gaussian, Clayton and Bernstein copula of order X; MD denotes the mean distance in absolute value between the estimated and true parameters; IQR denotes the 95% interquantile range of the estimated parameters.

of parameters, which is almost twice as large for the RQR estimator: the slope coefficients of the IVQR estimator use information from all observations, whereas the coefficient of the RQR estimator uses the information from the observations of one of the treatment status. Hence, the IQR of the RQR estimator is smaller, and its magnitude is similar regardless of the copula.

H.1 Support of the Propensity Score

Another experiment compares the performance of the estimator that uses a correctly specified Bernstein copula of order 2 ($\alpha(0.5, 0.5) = 0.375$), when the support of the propensity score changes. Since the shape assumption helps extending the identification argument from the observed interval to the unit line, it is pertinent to assess the performance of the estimator for different sizes of the observed interval. For a number of different supports, I draw the actual propensity score uniformly.[38] In other words, the first step in the implementation of the RQR estimator is not required.

Increasing the support of the propensity score improves the performance of the estimator: the RQR estimates are also more precise, as their bias tends to diminish (Table 6), and the distance between the estimated copula and its true value is smaller (Table 7). On the other hand, the IQR across repetitions of the RQR estimator is very stable and increases only slightly as the support of the propensity score diminishes. Thus, even if the model is identified with small variation of the propensity score, the performance of the estimator greatly depends on the amount of exogenous variation reflected by the propensity score.

H.2 Non-Analytical Copula

The identification result presented in this paper relies on the copula being analytic. The following simulation assesses the performance of the RQR estimator when the true copula is not analytic. In particular, it is a mixture between the lower Fréchet copula and the

[38] In particular, I consider the following sets of support: $[0.1, 0.9]$, $[0.15, 0.85]$, $[0.2, 0.8]$, $[0.25, 0.75]$, $[0.3, 0.7]$, $[0.35, 0.65]$, $[0.4, 0.5]$, and $[0.45, 0.55]$.

Table 6: Quantile Regression Coefficients, Varying Support

		$\beta_{1,1}$						$\beta_{0,1}$					
		τ						τ					
		0.1	0.25	0.5	0.75	0.9	Mean	0.1	0.25	0.5	0.75	0.9	Mean
MD	(1)	-0.03	0.00	-0.01	-0.01	-0.02	0.02	0.00	0.00	0.00	-0.02	-0.04	0.01
	(2)	0.00	-0.02	0.02	0.02	-0.02	0.02	-0.02	-0.01	0.01	0.00	-0.05	0.02
	(3)	-0.04	-0.01	0.00	-0.02	-0.03	0.02	-0.02	-0.01	0.01	-0.05	-0.01	0.01
	(4)	-0.07	-0.03	0.00	0.00	-0.03	0.03	0.01	0.01	-0.01	0.00	0.00	0.01
	(5)	-0.04	0.00	0.06	-0.02	0.00	0.02	-0.01	0.00	0.01	0.00	-0.02	0.01
	(6)	-0.05	-0.02	0.01	-0.03	-0.02	0.02	0.00	-0.02	0.00	-0.03	-0.04	0.02
	(7)	-0.01	0.02	0.06	0.04	0.02	0.02	-0.02	-0.04	-0.07	-0.11	-0.04	0.05
	(8)	0.03	0.08	0.15	0.10	0.07	0.08	-0.03	-0.04	-0.11	-0.15	-0.11	0.08
IQR	(1)	1.66	1.76	2.06	2.38	2.13	2.01	0.92	0.97	1.25	1.71	1.85	1.37
	(2)	1.65	1.89	2.18	2.21	1.98	2.02	0.88	0.92	1.33	1.65	1.96	1.34
	(3)	1.67	1.75	2.21	2.09	1.96	1.99	0.90	0.95	1.34	1.75	1.81	1.37
	(4)	1.66	1.79	2.21	2.33	2.01	2.02	0.88	0.94	1.35	1.76	1.92	1.40
	(5)	1.60	1.91	2.38	2.30	2.11	2.07	0.84	0.97	1.33	1.72	1.85	1.39
	(6)	1.78	2.05	2.22	2.26	2.03	2.09	0.89	0.96	1.39	1.83	1.97	1.41
	(7)	1.70	2.04	2.50	2.32	1.94	2.17	0.87	0.94	1.39	1.84	1.81	1.42
	(8)	1.95	2.14	2.71	2.49	2.11	2.32	0.88	0.96	1.46	1.87	1.91	1.44

		$\beta_{1.2}$						$\beta_{0.2}$					
		τ						τ					
		0.1	0.25	0.5	0.75	0.9	Mean	0.1	0.25	0.5	0.75	0.9	Mean
MD	(1)	0.04	0.01	-0.03	0.02	0.06	0.03	0.01	0.01	0.00	0.01	0.09	0.02
	(2)	0.03	0.00	-0.01	-0.01	0.08	0.03	0.02	0.00	0.00	0.01	0.07	0.02
	(3)	0.04	0.01	0.02	0.08	0.10	0.04	0.01	0.00	-0.02	0.03	0.00	0.01
	(4)	0.11	0.09	0.05	0.03	0.07	0.05	0.01	-0.02	-0.01	0.00	0.01	0.02
	(5)	0.05	0.02	0.02	0.07	0.07	0.05	0.00	-0.01	-0.06	-0.02	0.00	0.02
	(6)	0.08	0.03	0.06	0.09	0.06	0.07	-0.03	-0.05	-0.07	-0.03	0.03	0.04
	(7)	0.08	0.09	0.09	0.06	0.06	0.09	-0.03	-0.06	-0.08	-0.03	-0.05	0.06
	(8)	0.13	0.09	0.10	0.19	0.13	0.14	-0.09	-0.14	-0.15	-0.09	-0.04	0.12
IQR	(1)	2.52	2.62	3.12	3.34	3.05	2.92	1.35	1.45	1.98	2.64	2.62	2.06
	(2)	2.49	2.64	3.07	3.25	2.83	2.94	1.28	1.30	1.94	2.36	2.77	1.95
	(3)	2.54	2.58	3.02	3.00	2.88	2.89	1.32	1.47	2.00	2.54	2.46	2.00
	(4)	2.57	2.76	3.00	3.14	2.81	2.89	1.37	1.42	1.91	2.62	2.65	2.02
	(5)	2.64	2.80	3.24	3.28	2.83	2.98	1.28	1.60	2.00	2.54	2.85	2.05
	(6)	2.80	2.72	3.19	3.27	2.85	3.01	1.39	1.48	1.92	2.57	2.89	2.03
	(7)	2.70	2.85	3.23	3.13	2.85	3.03	1.43	1.56	2.01	2.44	2.71	2.05
	(8)	2.82	2.97	3.31	3.39	3.02	3.13	1.66	1.68	1.99	2.58	2.80	2.13

Notes: rows (1)-(8) denote the the different support of the propensity score used in each specification, in decreasing order; MD denotes the mean distance in absolute value between the estimated and true parameters; IQR denotes the 95% interquantile range of the estimated parameters.

Table 7: Estimated Copula, Varying Support

Propensity	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean (C_1)	0.006	0.007	0.007	0.008	0.010	0.012	0.016	0.020
Sup (C_1)	0.013	0.014	0.016	0.018	0.022	0.026	0.034	0.045
Mean (C_0)	0.006	0.006	0.007	0.008	0.010	0.013	0.017	0.022
Sup (C_0)	0.013	0.014	0.016	0.019	0.022	0.028	0.037	0.048

Notes: columns (1)-(8) denote the the different support of the propensity score used in each specification, in decreasing order; mean (C_D) and sup (C_D) respectively denote the mean and supremum distance across quantiles between the estimated copula and the true copula, averaged across repetitions, for $D = 0, 1$.

independence copula, with proportions $(0.25, 0.75)$ for the treated, and $(0.5, 0.5)$ for the untreated.

As shown in Table 8, the distance between the estimated copula and the true one is similar to the distance found when the copula was analytic. Note that the distance slightly increases as the order increases, although it is roughly stable across different orders. On the other hand, increasing order of the Bernstein copula reduces the bias of the RQR estimates, as shown in Table 9.

Table 8: Estimated Copula, Non-Analytical Copula

Copula	Ber(2)	Ber(3)	Ber(4)	Ber(5)	Ber(6)
Mean (C_1)	0.010	0.011	0.011	0.011	0.011
Sup (C_1)	0.029	0.033	0.034	0.034	0.034
Mean (C_0)	0.016	0.012	0.011	0.011	0.012
Sup (C_0)	0.063	0.048	0.045	0.043	0.043

Notes: Ber(X) stands for Bernstein copula of order X; mean (C_D) and sup (C_D) respectively denote the mean and supremum distance across quantiles between the estimated copula and the true copula, averaged across repetitions, for $D = 0, 1$.

Table 9: Quantile Regression Coefficients, Non-Analytical Copula

		$\beta_{1,1}$						$\beta_{0,1}$					
		τ						τ					
		0.1	0.25	0.5	0.75	0.9	Mean	0.1	0.25	0.5	0.75	0.9	Mean
MD	Ber(2)	0.08	0.08	0.09	-0.07	-0.11	0.07	-0.28	-0.25	0.06	0.26	0.21	0.16
	Ber(3)	0.06	0.09	0.09	-0.07	-0.11	0.07	-0.29	-0.29	0.03	0.24	0.19	0.16
	Ber(4)	0.05	0.06	0.09	-0.10	-0.11	0.07	-0.28	-0.30	0.02	0.25	0.16	0.16
	Ber(5)	0.04	0.07	0.09	-0.11	-0.10	0.07	-0.28	-0.29	0.00	0.24	0.16	0.16
	Ber(6)	0.04	0.07	0.09	-0.11	-0.10	0.07	-0.28	-0.29	-0.01	0.24	0.16	0.16
	True	-0.02	-0.01	0.02	0.02	-0.02	0.02	0.02	0.04	0.00	-0.01	0.04	0.03
IQR	Ber(2)	2.03	2.51	2.50	2.09	1.88	2.30	3.14	3.84	4.11	3.27	2.93	3.57
	Ber(3)	1.98	2.49	2.67	2.20	1.90	2.30	3.12	3.94	4.19	3.27	2.86	3.63
	Ber(4)	1.99	2.50	2.67	2.25	1.88	2.31	3.17	3.96	4.37	3.19	2.78	3.64
	Ber(5)	1.96	2.49	2.65	2.26	1.89	2.31	3.20	4.02	4.35	3.25	2.75	3.64
	Ber(6)	1.96	2.45	2.69	2.26	1.89	2.31	3.20	4.03	4.40	3.25	2.78	3.64
	True	2.10	2.44	2.59	2.21	1.92	2.29	3.17	3.61	3.97	3.20	2.91	3.44
		$\beta_{1.2}$						$\beta_{0.2}$					
		τ						τ					
		0.1	0.25	0.5	0.75	0.9	Mean	0.1	0.25	0.5	0.75	0.9	Mean
MD	Ber(2)	-0.02	-0.10	-0.23	-0.29	-0.03	0.16	0.04	0.09	0.02	-0.19	-0.37	0.11
	Ber(3)	0.00	-0.06	-0.14	-0.08	0.12	0.09	0.06	0.10	0.05	-0.20	-0.34	0.13
	Ber(4)	0.00	-0.04	-0.12	-0.03	0.17	0.08	0.07	0.10	0.03	-0.20	-0.33	0.13
	Ber(5)	0.00	-0.04	-0.09	-0.01	0.21	0.07	0.08	0.10	0.03	-0.22	-0.31	0.13
	Ber(6)	0.00	-0.04	-0.09	0.02	0.21	0.07	0.08	0.11	0.03	-0.24	-0.34	0.13
	True	0.00	-0.02	-0.01	-0.05	-0.03	0.03	-0.01	0.03	0.01	0.06	0.04	0.04
IQR	Ber(2)	0.81	0.79	0.98	1.48	2.24	2.30	1.14	1.18	1.46	2.21	3.19	3.57
	Ber(3)	0.81	0.76	1.07	1.70	2.22	2.30	1.15	1.19	1.43	2.41	3.08	3.63
	Ber(4)	0.77	0.77	1.06	1.77	2.15	2.31	1.15	1.16	1.46	2.45	3.17	3.64
	Ber(5)	0.77	0.78	1.10	1.87	2.16	2.31	1.16	1.17	1.43	2.52	3.10	3.64
	Ber(6)	0.77	0.78	1.12	1.88	2.16	2.31	1.15	1.16	1.47	2.50	3.07	3.64
	True	0.81	0.78	1.07	1.85	2.19	2.29	1.15	1.17	1.57	2.61	3.14	3.44

Notes: Ber(X) stands Bernstein copula of order X; MD denotes the mean distance in absolute value between the estimated and true parameters; IQR denotes the 95% interquantile range of the estimated parameters.

I Additional results

In this Appendix I report the Tables containing the results discussed in Section 4, as well as some robustness checks. Table 11 shows the cross-validated values of the objective function of the baseline estimators. Table 11 shows the QTE estimates based on the RQR estimator with the Gaussian copula, with the selected Bernstein copula, with the constrained Gaussian copula and on IVQR. Additionally, it shows the QTE estimates with a Gaussian copula but merging both treatments into the same category: any placement (AP). Finally, it shows the pairwise RQR estimates using the conditional propensity scores as in Lechner (2002).

Table 10: Cross-Validated Objective Function

Copula	Gau	Ber(2)	Ber(3)	Ber(4)	Ber(5)	Ber(6)
DHP	36.55	36.53	35.99	36.33	36.18	36.23
THP	22.36	22.22	22.48	22.32	22.47	22.27
NP	45.46	45.62	45.03	44.87	44.55	44.61

Notes: Gau, and Ber(X) respectively stand for Gaussian copula, and Bernstein copula of order X.

The QTE estimates of AP (any placement) relative to NP lie in between the QTE estimates of DHP relative to NP and of THP relative to NP. However, they are much closer to the latter, and each of the sets of estimates lies inside the 95% confidence interval of the other set of estimates. This result is not unexpected, since the majority of the treated belong to the DHP group.

Regarding the pairwise RQR estimates based on Lechner (2002), they are substantially different from the baseline RQR estimates. The main reason for this is that when one considers the choice between two of the three treatment status, the unobservable variable of the selection equation no longer corresponds to the unobservable equation of the same equation with the three treatment status. Hence, they are not directly comparable.

Table 12 reports the MTE estimates based on the RQR estimator with the Gaussian copula, the RQR estimator with the Gaussian copula, with the selected Bernstein copula, with the constrained Gaussian copula and the MTE estimates using Carneiro and Lee (2009) estimator. The latter can only be applied to binary data, so consequently it can only be used

Table 11: Quantile Treatment Effect Estimates

	u					
	0.15	0.25	0.5	0.75	0.85	Mean
$RQR(DHP.NP; Gau)$	26.3	118.6	393.4	640.1	654.6	339.0
	(12.1)	(33.7)	(96.9)	(171.7)	(203.7)	(104.9)
$RQR(THP.NP; Gau)$	22.8	89.9	264.0	408.9	428.8	224.6
	(7.7)	(17.4)	(58.1)	(103.9)	(134.9)	(60.9)
$RQR(DHP.THP; Gau)$	3.4	28.7	129.4	231.2	225.8	114.4
	(13.4)	(34.4)	(89.8)	(142.9)	(159.6)	(91.2)
$RQR(DHP.NP; Ber)$	19.3	107.1	372.9	634.8	678.6	337.7
	(12.7)	(38.4)	(110.5)	(196.3)	(216.4)	(112.3)
$RQR(THP.NP; Ber)$	23.2	97.3	277.0	426.6	465.1	242.5
	(7.3)	(18.1)	(55.9)	(113.1)	(143.6)	(62.0)
$RQR(DHP.THP; Ber)$	-3.9	9.7	95.9	208.2	213.5	95.3
	(14.0)	(38.2)	(108.3)	(175.6)	(189.0)	(95.3)
$RQR(DHP.NP; Con)$	19.4	120.4	451.5	805.7	899.9	438.8
	(12.0)	(38.5)	(119.2)	(197.9)	(234.8)	(125.5)
$RQR(THP.NP; Con)$	23.1	113.3	380.4	665.0	770.6	380.3
	(6.2)	(14.9)	(50.2)	(106.5)	(137.3)	(61.4)
$RQR(DHP.THP; Con)$	-3.7	7.1	71.1	140.6	129.4	58.4
	(11.5)	(29.9)	(74.6)	(121.8)	(136.9)	(72.2)
$IVQR(DHP.NP)$	93.2	190.2	386.4	685.5	1030.8	526.6
	(37.3)	(63.8)	(164.8)	(284.1)	(441.9)	(187.7)
$IVQR(THP.NP)$	0.1	5.0	-24.7	-215.2	-335.8	-109.1
	(36.6)	(51.4)	(141.4)	(236.6)	(269.6)	(156.7)
$IVQR(DHP.THP)$	93.1	185.1	411.2	900.6	1366.6	635.6
	(58.7)	(107.3)	(280.4)	(391.5)	(531.4)	(288.0)
$RQR(AP.NP; Gau)$	20.9	104.1	360.4	607.2	644.8	321.9
	(9.2)	(26.4)	(73.9)	(123.9)	(135.6)	(71.4)
$RQR(DHP.NP; Gau; PW)$	27.8	123.3	403.6	649.7	659.3	344.4
	(11.7)	(31.2)	(80.6)	(124.7)	(143.8)	(77.8)
$RQR(THP.NP; Gau; PW)$	8.1	61.6	185.0	276.1	278.7	146.7
	(16.3)	(82.2)	(281.7)	(544.5)	(632.4)	(313.7)
$RQR(DHP.THP; Gau; PW)$	-65.9	-151.7	-345.5	-530.8	-643.5	-355.4
	(53.2)	(131.1)	(341.2)	(573.9)	(660.2)	(347.6)

Notes: Gau, Con, Ber, and PW respectively stand for Gaussian copula, Gaussian copula constrained to be the same for all three groups, the selected Bernstein copula, and conditional pairwise propensity score; u denotes the quantile; mean denotes the average across all quantiles in the estimation grid; bootstrapped standard errors in parenthesis.

to compute the MTE of the combination of the DHP and THP groups relative to the NP group. The estimates of the latter are substantially different from those of the MTE of either the DHP or THP groups relative to the NP group with the RQR estimators. In particular, the estimate is above 1,000 for small values of v, decreasing towards the center of the distribution to about 300, and increasing again for high values of v. Because the identification conditions for this estimator require large support of the instrument, the estimates at extreme values of v are less reliable than those around the center. It is for these values of v that the estimates based on Carneiro and Lee (2009) are most similar to those based on RQR.

Table 12: Marginal Treatment Effect Estimates

			v			
Treatment	0.15	0.25	0.5	0.75	0.85	Mean
$RQR(DHP.NP; Gau)$	90.2	182.0	344.0	493.7	568.6	331.6
	(137.0)	(107.1)	(102.4)	(138.7)	(168.2)	(107.5)
$RQR(THP.NP; Gau)$	42.4	112.1	233.4	343.0	396.7	221.6
	(179.8)	(125.9)	(62.3)	(121.8)	(160.8)	(62.7)
$RQR(DHP.THP; Gau)$	47.8	69.9	110.6	150.7	171.9	110.0
	(144.5)	(87.1)	(90.7)	(168.4)	(214.1)	(90.0)
$RQR(DHP.NP; Ber)$	145.6	159.2	269.6	469.9	571.0	328.9
	(125.8)	(120.1)	(128.0)	(174.1)	(216.1)	(113.6)
$RQR(THP.NP; Ber)$	86.5	73.6	148.8	358.5	475.9	237.3
	(192.8)	(141.7)	(79.8)	(110.1)	(147.1)	(63.6)
$RQR(DHP.THP; Ber)$	59.1	85.7	120.8	111.4	95.1	91.7
	(115.4)	(80.8)	(100.4)	(220.2)	(265.4)	(98.1)
$RQR(DHP.NP; Con)$	446.1	442.1	434.7	427.3	423.4	434.7
	(101.7)	(109.1)	(125.4)	(141.7)	(149.5)	(125.5)
$RQR(THP.NP; Con)$	390.3	387.1	381.2	375.3	372.1	381.2
	(47.5)	(51.9)	(61.3)	(70.1)	(76.3)	(61.3)
$RQR(DHP.THP; Con)$	55.8	55.0	53.5	52.1	51.3	53.5
	(65.2)	(68.0)	(73.0)	(78.2)	(80.2)	(73.0)
$CL(DHP.NP)$	1123.5	884.8	305.4	486.6	737.3	696.3
	(442.0)	(258.1)	(83.1)	(324.4)	(558.3)	(236.2)

Notes: Gau, Con, Ber, anc CL respectively stand for Gaussian copula, Gaussian copula constrained to be the same for all three groups, the selected Bernstein copula, and Carneiro and Lee (2009) estimator; v denotes the conditioned value of the unobservable of the selection equation; mean denotes the average across all v in the estimation grid; bootstrapped standard errors in parenthesis.

Table 13: Marginal Treatment Effect Decomposition Estimates

		v					
	Treatment	0.15	0.25	0.5	0.75	0.85	Mean
RIMTE	$RQR(DHP.NP; Gau)$	330.6	330.6	330.6	330.6	330.6	330.6
		(90.8)	(96.7)	(107.8)	(117.4)	(124.5)	(107.9)
	$RQR(DHP.NP; Gau)$	226.2	224.3	220.8	217.2	215.3	220.7
		(64.1)	(62.7)	(62.5)	(67.3)	(70.4)	(63.0)
	$RQR(THP.NP; Gau)$	109.9	109.9	109.9	109.9	109.9	109.9
		(79.3)	(83.9)	(89.9)	(96.8)	(100.2)	(90.1)
	$RQR(DHP.NP; Ber)$	334.2	336.7	336.5	327.1	320.8	330.5
		(94.0)	(102.5)	(116.0)	(125.1)	(125.2)	(107.9)
	$RQR(DHP.NP; Ber)$	244.7	243.1	239.2	235.3	233.7	239.2
		(61.9)	(62.1)	(63.6)	(64.1)	(66.7)	(63.0)
	$RQR(THP.NP; Ber)$	92.4	93.3	93.3	90.0	87.7	91.2
		(88.6)	(93.6)	(100.4)	(105.9)	(104.4)	(90.1)
ESME	$RQR(THP.NP; Gau)$	-240.4	-148.6	13.4	163.0	238.0	1.0
		(92.5)	(61.0)	(2.7)	(59.7)	(92.8)	(0.2)
	$RQR(DHP.THP; Gau)$	-183.8	-112.2	12.7	125.8	181.5	0.9
		(138.4)	(86.9)	(5.4)	(92.7)	(137.6)	(0.4)
	$RQR(DHP.THP; Gau)$	-62.1	-40.0	0.7	40.8	62.0	0.1
		(163.3)	(103.7)	(5.1)	(108.5)	(162.7)	(0.4)
	$RQR(THP.NP; Ber)$	-188.5	-177.4	-66.8	142.8	250.3	-1.5
		(125.7)	(97.9)	(70.0)	(89.7)	(135.2)	(1.2)
	$RQR(DHP.THP; Ber)$	-158.2	-169.6	-90.4	123.2	242.2	-2.0
		(143.9)	(100.2)	(51.3)	(108.0)	(133.0)	(0.9)
	$RQR(DHP.THP; Ber)$	-33.3	-7.7	27.5	21.4	7.4	0.5
		(175.6)	(122.2)	(46.5)	(127.7)	(175.7)	(0.9)

Notes: Gau, Con, and Ber respectively stand for Gaussian copula, Gaussian copula constrained to be the same for all three groups, and the selected Bernstein copula; v denotes the conditioned value of the unobservable of the selection equation; mean denotes the average across all v in the estimation grid; bootstrapped standard errors in parenthesis.