# Identification and Estimation of Triangular Models with a Binary Treatment

Santiago Pereda-Fernández*

Banca d'Italia

January 5, 2018

**Abstract**

I study the identification and estimation of triangular models with an endogenous binary treatment. The system consists of a selection equation for the treatment and a structural quantile function that expresses the outcome variable as a function of the treatment. The unobservables of both equations are related by a copula, which captures the endogeneity of the model. Unlike other studies, I do not impose rank invariance or rank similarity on the unobservable of the outcome equation, and the outcome equation is non-separable. The estimation is a multi-step procedure that consists in the estimation of the propensity score, the copula of the unobservables for each treated and untreated, and the structural quantile function using rotated quantile regression. Several statistics of interest, including the marginal treatment effect, can be expressed as a function of the structural functions of the model, allowing them to be estimated using the estimates of the structural functions.

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

**JEL classification:** C31, C36

*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

# 1 Introduction

One of the most relevant settings in empirical works is a triangular system with a binary endogenous treatment. This system consists of a continuous outcome that depends on the treatment, a set of covariates, and an unobserved random variable. On the other hand, the treatment depends on an instrument, the exogenous covariates, and another unobservable random variable, which is correlated with the unobservable of the outcome equation, making the treatment endogenous. This model has been widely studied in the literature, where it has been emphasized the importance of allowing for nonlinear and heterogeneous effects.

Consequently, a variety of settings have been considered, including nonparametric models that are additively separable on the unobservables, and non-separable models. However, a commonly maintained assumption is that of rank invariance or rank inequality, *i.e.* that the disturbance term of the outcome equation is either the same for each individual regardless of the treatment, or that at least it has the same distribution. This assumption may be too strong in some applications, thus making it necessary to properly account for the difference in the distribution of the unobservables for the treated and the untreated.

The approach in this paper is to specify the outcome as a structural quantile function, and model the correlation of the unobservables of the selection and the outcome equations with a copula. Importantly, I allow each individual to have a different unobservable when they are treated and when they are not, and the copula between each of these and the unobservable of the selection equation may also be different. Using variation of the instrument, it is possible to point identify both the copula and the structural quantile function that determines the outcome, even when the support of the instrument is small.

On top of the identification result, I propose a multi-step estimator that converges at the parametric rate. It consists on the estimation of the three components of the system: the propensity score, the copula of the unobservables, and the structural quantile process of the outcome equation. This procedure draws from the one proposed in Arellano and Bonhomme (2017) for quantile regression in sample selection models, and in particular it uses rotated quantile regression (RQR). Intuitively, because of the endogeneity, the conditional quantile

of an individual with a given treatment status does not coincide with the quantile of the distribution of potential outcomes for the whole population. However, the mapping between these two variables is a function of the copula, and therefore it is possible to estimate the conditional quantile function by appropriately rotating the check function that is used in standard quantile regression (Koenker and Bassett, 1978).

Despite not being the main object of interest for the policy maker, the copula plays an important role: the Marginal Treatment Effect (MTE) and the distribution of the effects can be expressed in terms of the copula. Hence, the copula can be policy-informative. For example, if it displays a high level of correlation, then treating individuals with high (low) propensity would result in large (small) treatment effects. Conversely, if it displays little correlation, then the treatment effect could be both small and large for individuals with either high or low propensity to be treated.

Frequently, the main objects of interest are different types of treatment effects, such as the Average Treatment Effect (ATE), the Average Treatment Effect on the Treated (TT), the Average Treatment effect on the Untreated (TUT), on top of the aforementioned MTE. I show that they can be characterized as a function of the copula and the two equations of the triangular system. Hence, by appropriately integrating the estimator of the outcome equation with respect to the copula, it is possible to estimate all these treatment effects.

This paper belongs to the literature of identification and estimation of triangular systems of equations when the treatment is binary.[1] Chesher (2005) provided partial identification of the structural function under weak assumptions, which was strengthened by Jun et al. (2011). A strand of the literature focused on the identification when the model is additively separable in the unobservables, such as Das (2005) or Carneiro and Lee (2009). The latter used continuous variation of the instrument to show nonparametric identification of the distribution of potential outcomes and proposed a semiparametric estimator.

Another strand of the literature focused on models that were non-separable in the

[1]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).

unobservables. Chernozhukov and Hansen (2005, 2006) constitute an important milestone in the literature, defining a quantile treatment effect framework, which works both with continuous and discrete treatments. In their paper, they provided a linear quantile estimator, known as Instrumental Variables Quantile Regression (IVQR) estimator. More recently Jun et al. (2016) studied the identification using a control function approach, discussing the trade-offs between continuity, monotonicity, and differentiability of the propensity score, proposing a semiparametric estimator.

Finally, a series of papers have studied the identification of the effects on compliers when the instrument is binary. The early focus was on the Local Average Treatment Effect (Imbens and Angrist, 1994), but the interest has shifted to the estimation of distributional effects. For example, Abadie et al. (2002) proposed an estimator of a linear quantile model, which was compared to Chernozhukov and Hansen (2005) by Wüthrich et al. (2015). More recent works include Vuong and Xu (2017), who studied nonparametric identification using the monotonicity of the latent variable, and Feng et al. (2016), who proposed an estimator based on Vuong and Xu (2017) that achieves $\sqrt{n}$ convergence rate when the covariates are discrete. On the other hand, Frölich and Melly (2013) proposed an estimator of the unconditional treatment effect, and Frandsen et al. (2012) estimated local quantile treatment effects in the regression discontinuity design.

The identification results of most of these papers hinge on either rank continuity or rank invariance, thus limiting the amount of heterogeneity that the model can display. Works based on Local Instrumental Variables (LIV, Heckman and Vytlacil, 1999, 2005) are a prominent exception, and they explicitly acknowledge the existence of two latent variables that determine the outcome: one when treated, and one when untreated. (Carneiro and Lee, 2009) proposed a semiparametric estimator that achieves the parametric rate of convergence under additive separability of the unobservables. The results in this paper strengthen their identification argument through the use of copulas. Moreover, the estimation method presented in this paper does not require additive separability of the unobservables.

The rest of the paper is organized as follows: section 2 introduces the model and discusses

the identification. Section 3 describes the estimation method. Section 4 presents some extensions to the baseline estimator, and section 5 concludes.

# 2 The Model

Consider the following triangular system of equations:

$$Y = g_D(X, U_D) \tag{1}$$

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

where $Y$ is the continuous outcome, $D$ is the binary treatment, $X$ is the vector of covariates, and $Z \equiv (Z_1, X')'$ includes the instrument $Z_1$. The treatment is determined by equation 2, known as the selection equation. It is modeled following Heckman and Vytlacil (2005), *i.e.* it depends on the propensity score, $\pi(Z)$, and a uniformly distributed unobserved random variable V. On the other hand, equation 1 is the Structural Quantile Function (SQF), which models the outcome as a function of the treatment, the covariates, and a univariate unobservable $U_D$, which is uniformly distributed over the unit interval.[2] The system does not impose separability between the observables and the unobservables.

The joint distribution of the disturbance terms of the system is modeled with a copula that captures the endogeneity of the system. In particular, I assume the existence of three unobservable variables $(U_0, U_1, V)$ which are distributed as $U_0, U_1, V|X \sim C_X(U_0, U_1, V|X)$.[3] Throughout the paper I respectively refer to $V$ and $U_D$ as the ranks of the selection equation and the SQF.

Denote the copula between $U_D$ and $V$, conditional on $(D, X')'$, by $C_{D,X}(U_D, V)$. Since only one of the two treatment status is observed for each individual, any structural function of

[2]This is known as the Skorohod representation.

[3]Even though this setting allows for heterogeneous effects, even for individuals with the same treatment and covariates, 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 in general not point-identified (Hahn and Ridder, 2011; Kasy, 2011; Hoderlein et al., 2017; Masten, 2017).

the data can be expressed in terms of $C_{D,X}$, and it is not possible to identify the structural relation between $U_0$ and $U_1$.[4] Much of the literature has focused on the rank invariance ($U_0 = U_1$) or rank similarity cases ($C_{0,X} = C_{1,X}$). In contrast, I assume *rank dissimilarity*, which allows the distribution of $U_D$ to vary with the treatment status, *i.e.* $C_{0,X} \neq C_{1,X}$. In practice, this means that an individual with a high probability of having a high level of outcome when untreated (*i.e.* high $U_0$), does not necessarily has a high probability of having a high level of outcome when treated.

To better understand this concept, consider the following example. Denote earnings by $Y$, the possession of a college degree by $D$, and let $U_D$ be a measure of ability at work with education status $D$. Under rank invariance, the ability of an individual would be the same regardless of whether he is in possession of a college degree or not, and it would rank the same in the distribution of potential earnings with and without college degree. Under rank similarity, an individual with a particular level of ability when he has college education, can have a different level of ability when he has no college education. However, conditional on the probability to have a college degree, the distribution of ability would be the same for those with and without a college degree. Finally, under rank dissimilarity, these two distributions can be different. Hence, those with a high propensity to have a college degree may be very likely to have high ability when they have college education, but the reverse could be true when they have no college degree.

The copula has traditionally received little attention in the treatment effects literature. However, its knowledge can increase the external validity of the estimates, since it is closely related to the MTE. To see how it can be informative on the potential effects of the treatment on the untreated, notice that if the copula of the unobservables displays a high correlation, individuals with a high propensity to be treated are those who have a higher treatment effect, and consequently extending the treatment to untreated individuals will result in smaller effects for the newly treated. In contrast, if the correlation is close to zero, the effect of extending the treatment to those with a small propensity will be of a similar size as the

[4]This is akin to the identification of the distribution of the treatment effect, which is not point identified, but can be bounded (Firpo and Ridder, 2008; Fan and Park, 2010).

effect on those with a high propensity.

## 2.1 Identification of the Structural Functions

The distribution of the outcome variable conditional on $Z$ can be decomposed into the weighted sum of two distributions: one for the treated and another for the untreated:

$$F_{Y|Z}(y|z) = F_{Y|D=0,Z}(y|z)(1-\pi(z)) + F_{Y|D=1,Z}(y|z)\pi(z)$$

The copula, appropriately scaled by the propensity score, is an argument of each these distributions. For the untreated, it is given by

$$F_{Y|D=0,Z}(y|z) = \int_0^1 \mathbf{1}(g_0(x,u_0) \leq y)\, dG_X(u_0, \pi(z)) \tag{3}$$

where $G_X(\tau, \pi(z)) \equiv \frac{\tau - C_{0,X}(\tau,\pi(z))}{1-\pi(z)} = \mathbb{P}(U_0 \leq \tau | D=0, z)$, and $\mathbf{1}(\cdot)$ is the indicator function. Evaluating equation 3 at $y = g_0(\tau, x)$ yields $F_{Y|D=0,Z}(g_0(x,\tau)|z) = G_X(\tau, \pi(z))$. Similarly, the distribution for the treated is given by

$$F_{Y|D=1,Z}(y|z) = \int_0^1 \mathbf{1}(g_1(x,u_1) \leq y)\, dH_X(u_1, \pi(z)) \tag{4}$$

where $H_X(\tau, \pi(z)) \equiv \frac{C_{1,X}(\tau,\pi(z))}{\pi(z)} = \mathbb{P}(U_1 \leq \tau | D=1, z)$, and evaluating equation 4 at $y = g_1(\tau, x)$ yields $F_{Y|D=1,Z}(g_1(x,\tau)|z) = H_X(\tau, \pi(z))$.

Therefore, $F_{Y|Z}$ depends on three components: the SQF of $Y$, the propensity score, and the copulas of $C_{0,X}$ and $C_{1,X}$. To identify them, I work with the following assumptions:

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

**Assumption 2.** *The bivariate distributions $(U_0, V)$ and $(U_1, V)$, conditional on $X$, are absolutely continuous with respect to the Lebesgue measure. Moreover, $U_0$, $U_1$, and $V$ are uniformly distributed on the unit interval.*

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

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

**Assumption 5.** $\forall \tau \in (0,1)$, *the functions* $\pi \to C_{1,X}(\tau, \pi)$ *and* $\pi \to C_{0,X}(\tau, \pi)$ *are real analytic on the unit 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 $Z_1$. Assumption 2 implies that the SQF is continuous with respect to $U_D$, and similarly for the propensity score. It also normalizes the marginal distributions of the ranks to be uniform, making their joint distribution a well-defined copula, and together with assumption 3, it allows the system (1)-(2) to represent the conditional quantile function of the potential outcomes $Y_D^*$. Assumption 4 is a support assumption on the instrument, which is required to display some continuous variation that maps into the propensity score. Finally, assumption 5 assumes that the copulas $C_{1,X}$ and $C_{0,X}$ are real analytic, and so are the functions $G_X$ and $H_X$. This assumption is required to achieve identification when $\mathcal{P}_X$ does not cover the unit interval.[5] This assumption is discussed after proposition 1

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 for the copula hold:

**Lemma 1.** *Let* $x \in \mathcal{X}$. *Then, under assumptions 1 to 4:*

$$F_{Y|D=0,Z}\left(F^{-1}_{Y|D=0,Z}\left(\tau|z'\right)|z\right) = G_X\left(G_X^{-1}\left(\tau, \pi\left(z'\right)\right), \pi\left(z\right)\right) \forall \left(z, z'\right) \in \mathcal{Z}_x \times \mathcal{Z}_x \tag{5}$$

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

*Moreover, for any* $G_X$ *and* $H_X$ *satisfying equations 5 and 6, one can find distribution functions* $F_{Y_0^*|X}(y|x)$ *and* $F_{Y_1^*|X}(y|x)$ *such that* $G_X\left(F_{Y_0^*|X}(y|x), \pi(z)\right) = F_{Y|D=0,Z}(y|z)$ *and*

[5] When there exist some $z$ such that the propensity equals either 0 or 1, it is possible to invoke the identification at infinity argument (Heckman and Vytlacil, 2007) to achieve the identification of the structural functions for the untreated and the treated, respectively.

$H_X\left(F_{Y_1^*|X}\left(y|x\right),\pi\left(z\right)\right) = F_{Y|D=1,Z}\left(y|z\right)$ *for all* $(z,y)$ *in the support of* $(Z,Y)$ *given* $X = x$*, where* $Y_0^*$ *and* $Y_1^*$ *are the potential outcomes of individuals when they are respectively untreated or treated.*

Equations 5 and 6 require the instrument $Z$ to come from a non-degenerate distribution for them to be informative about the copulas. Also, the second part of the lemma already indicates the existence of potential outcomes, denoted by $Y_0^*$ and $Y_1^*$, which would be observed if the treatment was randomly assigned, and how they relate to the observed outcomes when the treatment is endogenous. Specifically, there is a bijection between the two of them that depends on the copula, and their distributions would coincide if the copula was independent, *i.e.* in the absence of endogeneity. The support of the potential outcomes is the same as the support of the realized outcomes for both treated and untreated individuals, and therefore one can link the $\tau$-th quantile of the distribution of potential outcomes to a particular quantile of the distribution of realized outcomes. This constitutes the basis for using RQR (Arellano and Bonhomme, 2017) in the estimation.

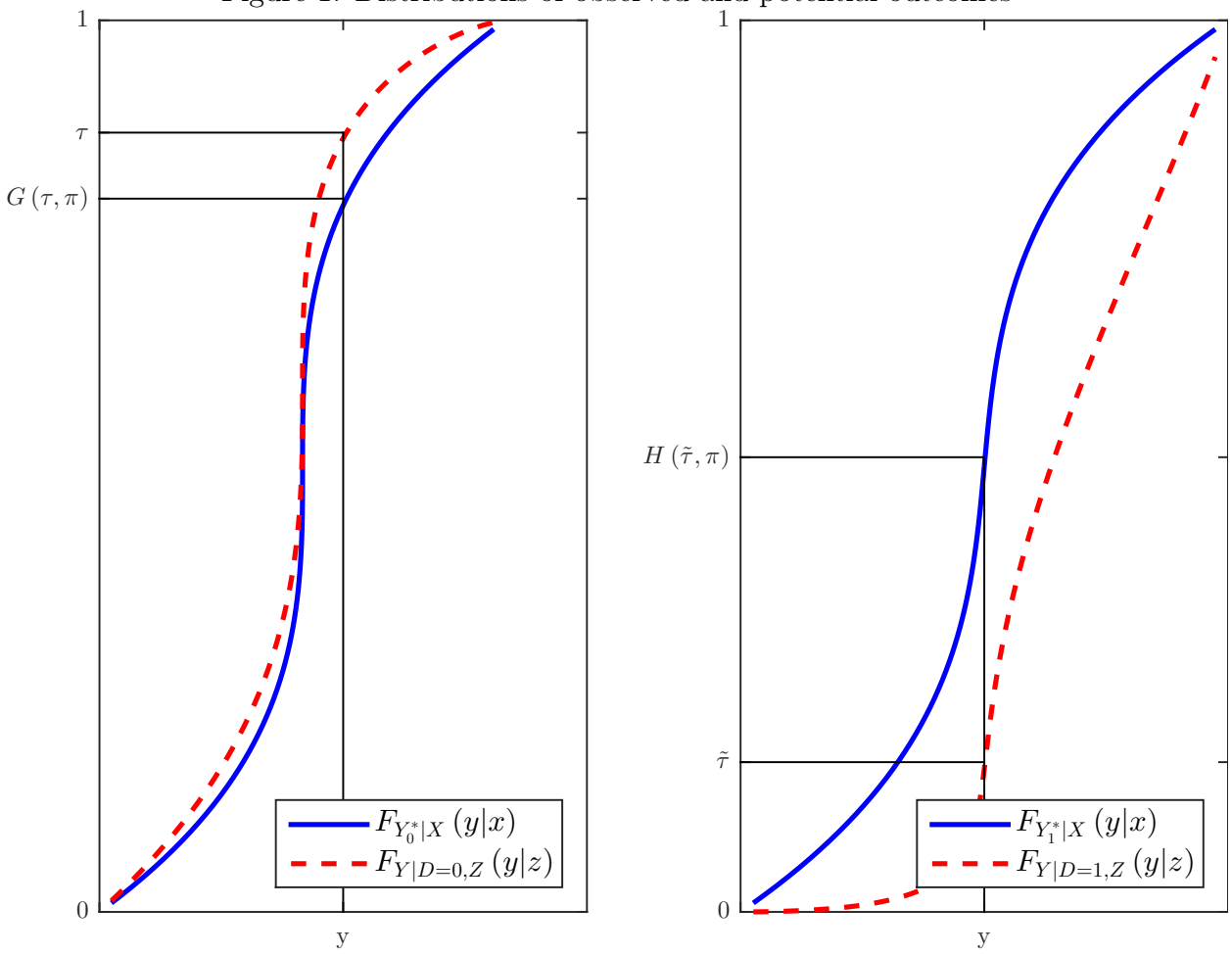
Graphically, the distribution of potential outcomes is a nonlinear translation of the observed distribution, which is mapped by the copula, as shown in figure 1. Because the unobservables for the treated and the untreated have a different distribution, the translation is different for each group. In the example shown in figure 1, there is positive correlation, so the distribution of observed outcomes first order stochastically dominates the distribution of potential outcomes for the treated, and the reverse is true for the distribution of the untreated.

Denote the support of $\pi\left(Z\right)$ conditional on $X = x$ by $\mathcal{P}_x$. The following proposition establishes the identification of $G_X$ and $H_X$:

**Proposition 1.** *Let assumptions 1 to 5 hold, and* $x \in \mathcal{X}$*. Then, the functions* $(\tau,\pi) \to G_X\left(\tau,\pi\right)$*,* $(\tau,\pi) \to H_X\left(\tau,\pi\right)$*, and* $\tau \to g_D\left(x,\tau\right)$ *for* $D = 0,1$ *are nonparametrically identified.*

The identification rests on the assumption that the functions $G_X$ and $H_X$ are real analytic, which allows the extrapolation from $\mathcal{P}_x$ to the whole real line. To assess the

Figure 1: Distributions of observed and potential outcomes




Notes: $G(\tau, \pi)$ and $H(\tilde{\tau}, \pi)$ are shorthands for $G_X(\tau, \pi(z))$ and $H_X(\tilde{\tau}, \pi(z))$, respectively.

plausibility of assumption 5, consider the Bernstein copula, which is constructed using Bernstein polynomials, and thus real analytic by construction. Bernstein (1912) showed that these polynomials can arbitrarily approximate any bounded continuous function on the unit interval, a result also known as Stone-Weierstrass approximation theorem. This was extended by Sancetta and Satchell (2004), who showed that the set of Bernstein polynomials is dense in the space of bounded continuous functions in the $k$-dimensional hypercube $[0,1]^k$ (lemma 1). Consequently, Bernstein copulas can approximate any arbitrary copula that has a well-defined density.[6]

Thus, the identification is achieved by finding the real analytic copulas that satisfy equations 5-6 $\forall \tau \in [0,1]$ and $\forall \pi \in \mathcal{P}_X$. Assumption 5 extends the identification from $\mathcal{P}_X$ to $[0,1]$, which separately identifies the copulas from the SQF. The identification is shown graphically in figure 2. If one changes the propensity score by changing the instrument from $z$ to $\tilde{z}$, then the distribution of observed outcomes also changes, both for the treated and the untreated. Using equations 5 and 6, it is possible to link each pair of distributions using the copula.

## 2.2 Comparison with Identification based on LIV

The identification conditions of LIV 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

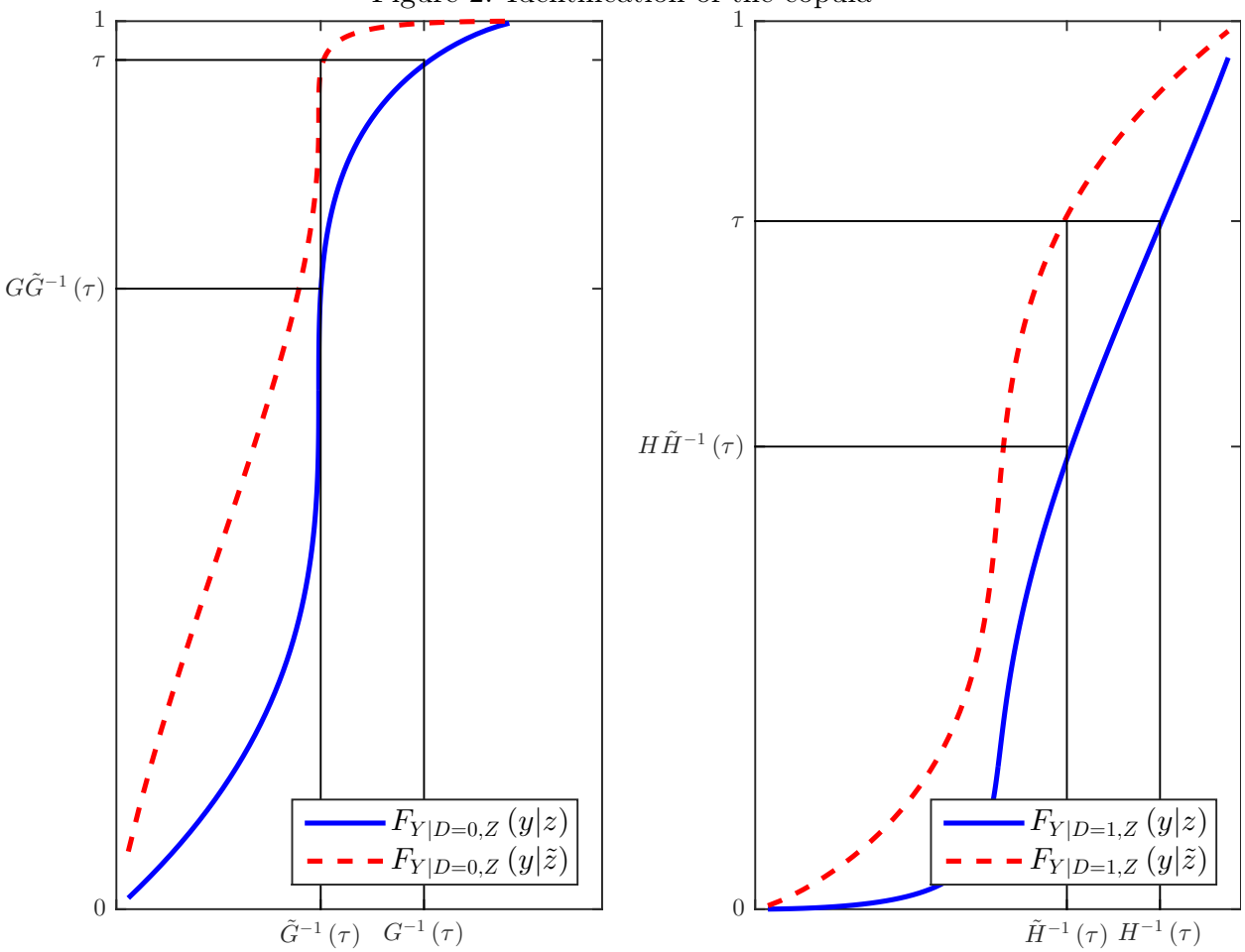
$$\frac{\partial}{\partial p} C_{0,X}\left(F_{Y_0^*}(y|x), p\right)\Big|_{p=\pi(z)} = F_{Y|D=0,Z}(y|z) - (1-\pi(z))\frac{\partial}{\partial \pi(z)} F_{Y|D=0,Z}(y|z) \quad (7)$$

$$\frac{\partial}{\partial p} C_{1,X}\left(F_{Y_1^*}(y|x), p\right)\Big|_{p=\pi(z)} = F_{Y|D=1,Z}(y|z) + \pi(z)\frac{\partial}{\partial \pi(z)} F_{Y|D=1,Z}(y|z) \quad (8)$$

If there exist values of $z$ such that the propensity score can be equal to 0 and 1 for every value of $x$ (identification at infinity), then it is possible to separately identify the copula and the distribution of potential outcomes. Otherwise, it is only possible to identify

[6]Note that this rules out the perfect correlation copulas, since they have no density.

Figure 2: Identification of the copula




Notes: $G^{-1}(\tau)$, $\tilde{G}^{-1}(\tau)$, $H^{-1}(\tau)$, $\tilde{H}^{-1}(\tau)$, $G\tilde{G}^{-1}(\tau)$, and $H\tilde{H}^{-1}(\tau)$ are shorthands for $G_X^{-1}(\tau, \pi(z))$, $G_X^{-1}(\tau, \pi(\tilde{z}))$, $H_X^{-1}(\tau, \pi(z))$, $H_X^{-1}(\tau, \pi(\tilde{z}))$, $G_X\left(G_X^{-1}(\tau, \pi(\tilde{z})), \pi(z)\right)$, and $H_X\left(H_X^{-1}(\tau, \pi(\tilde{z})), \pi(z)\right)$, respectively.

a composition of them. Also, notice that using LIV, the left hand side of equations 7-8 can only be identified over the support $\mathcal{P}_x$. The key difference with respect to the identification result in proposition 1 is the assumption that the copula is real analytic, which allows to extrapolate the identification from $\mathcal{P}_x$ to the whole real line.

## 2.3 Comparison with Identification based on IVQR

Chernozhukov and Hansen (2005, 2006) presented the IVQR model. It is based on different assumptions from those presented in this paper, and in general the models they estimate are not nested. Notably, it allows the treatment to be either discrete or continuous, whereas this paper focuses on the binary treatment case. The identification in the IVQR model rests on a moment that requires either rank invariance or rank similarity. 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:

$$\mathbb{P}\left(Y \leq g_D\left(X,\tau\right)|Z\right) = \tau - C_{0,X}\left(F_{Y_0^*}\left(g_0\left(X,\tau\right)\right),\pi\left(Z\right)\right) + C_{1,X}\left(F_{Y_1^*}\left(g_1\left(X,\tau\right)\right),\pi\left(Z\right)\right) \tag{9}$$

Hence, under rank dissimilarity, the moment $\mathbb{P}\left(Y \leq g_D\left(X,\tau\right)|Z\right) \neq \tau$, and therefore it does not point identify the SQF process. The cost of not requiring rank similarity is the specification of the selection equation and of the copula between the unobservable ranks (equation 2), and therefore the IVQR estimator is robust to the misspecification of this equation. Nevertheless, it is still possible to combine equation 9 with Frechét bounds to obtain set identification:

$$\begin{aligned} \tau + \min\left\{F_{Y_0^*}\left(g_0\left(X,\tau\right)\right),\pi\left(Z\right)\right\} - \max\left\{F_{Y_1^*}\left(g_1\left(X,\tau\right)\right) - \pi\left(Z\right),0\right\} \leq \\ \mathbb{P}\left(Y \leq g_D\left(X,\tau\right)|Z\right) \leq \\ \tau + \min\left\{F_{Y_1^*}\left(g_1\left(X,\tau\right)\right),\pi\left(Z\right)\right\} - \max\left\{F_{Y_0^*}\left(g_0\left(X,\tau\right)\right) - \pi\left(Z\right),0\right\} \end{aligned} \tag{10}$$

## 2.4 Relation with other Treatment Effects

The marginal treatment effect, as defined by Heckman and Vytlacil (2005), is interpreted as the expected treatment effect on the marginal individual entering treatment, and it can be written in terms of the copula:[7]

$$\Delta^{MTE}(x, z_1) = \int_0^1 g_1(x, u_1)\, dC_{1,X}(u|\pi(z)) - \int_0^1 g_0(x, u_0)\, dC_{0,X}(u|\pi(z)) \tag{11}$$

Notice that it can be decomposed into two different components:

$$\begin{aligned}\Delta^{MTE}(x, z_1) = &\int_0^1 [g_1(x, u) - g_0(x, u)]\, dC_{1,X}(u|\pi(z)) \\ &- \int_0^1 g_1(x, u_0)\, d[C_{1,X}(u|\pi(z)) - C_{0,X}(u|\pi(z))]\end{aligned}$$

The first term can be interpreted as the marginal treatment effect under rank invariance, *i.e.* the expected gain for the marginal individual, when his unobservables have the same distribution regardless of the treatment status. This effect depends on the difference between the distributions of potential outcomes. On the other hand, the second term captures the excess selection effect, which reflects rank dissimilarity. In this case, the effect reflects the difference in the amount of selection (or endogeneity) between the two treatment status.

To better understand this, imagine that the copula for the treated is a Bernstein copula of order 2 with $\alpha(0.5, 0.5) = .5$ (positive correlation), and the copula for the untreated is independent. Then, $C_{1,X}(u|v) - C_{0,X}(u|v) = u(u-1)(2v-1)$, which is positive for $v \leq 0.5$, and negative otherwise. Assuming that $g_1(x, \tau)$ is increasing in $\tau$, it follows that the second term is positive. In words, because the selection into treatment results in a distribution of the unobservables that displays a higher level of correlation when treated, the excess selection effect is positive.

An implication of this is that, even if the SQF is the same for treated and untreated individuals, the marginal treatment effect would be positive because the marginal individual

[7] With some abuse of notation, I denote the copula of $U_D$ conditional on $V$ and $X$ by $C_{D,X}(u|v)$.

would, on average, have a greater value of the unobservable when treated. Conversely, under either rank invariance or similarity, the excess selection effect vanishes.

Therefore, the average treatment effect on the treated and the untreated can be also characterized by the copula. In particular, they depend on $H_X$ and $G_X$, since they capture the distribution of the unobservable $U_D$, conditional on the observed treatment:[8]

$$\Delta^{TUT}(z) = \int_0^1 g_1(x, u_1)\, dG'_X(u_1, \pi(z)) - \int_0^1 g_0(x, u_0)\, dG_X(u_0, \pi(z)) \tag{12}$$

$$\Delta^{TT}(z) = \int_0^1 g_1(x, u_1)\, dH_X(u_1, \pi(z)) - \int_0^1 g_0(x, u_0)\, dH'_X(u_0, \pi(z)) \tag{13}$$

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

$$\Delta^{ATE}(z) = \Delta^{TUT}(z)(1 - \pi(z)) + \Delta^{TT}(z)\pi(z) = \int_0^1 (g_1(x, u) - g_0(x, u))\, du \tag{14}$$

# 3 Estimation

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.**

$$g_0(x, \tau) = x'\beta_0(\tau)$$

$$g_1(x, \tau) = x'\beta_1(\tau)$$

[8] From these expressions, obtaining the unconditional average treatment effect for the whole population, the untreated, and the treated, is straightforward by integrating over the distribution of $Z$: $ATE = \int_{\mathcal{Z}} ATE(z)\, dF_Z(z)$, $TUT = \int_{\mathcal{Z}} TUT(z)\, dF_Z(z)$, and $TT = \int_{\mathcal{Z}} TT(z)\, dF_Z(z)$.

*where $\beta_j$ is continuous and such that $g_j(x,\tau)$ is increasing in its last argument.*

**Assumption 8.** *Let $\beta(\tau) \equiv \left(\beta_1(\tau)', \beta_0(\tau)'\right)'$ and $\theta \equiv (\theta_1', \theta_0')'$. For all $\tau$, $\left(\beta(\tau)', \theta', \gamma'\right)' \in int\mathcal{B} \times \Theta \times \mathcal{G}$, where $\mathcal{B} \times \Theta \times \mathcal{G}$ is compact and convex.*

**Assumption 9.** *$Y$ has conditional density that is bounded from above and away from zero, a.s. on compact set $\mathcal{Y}$. The density is given by $f_{Y|D,Z}(y)$ for $D = 0, 1$.*

**Assumption 10.** *Matrices of derivatives of the moments $J_0(\tau)$, $\tilde{J}_0(\tau)$, $J_1(\tau)$, $\tilde{J}_1(\tau)$, $P_{01}(\tau)$, $\tilde{P}_{01}(\tau)$, $P_{02}(\tau)$, $\tilde{P}_{02}(\tau)$, $P_{11}(\tau)$, $\tilde{P}_{11}(\tau)$, $P_{12}(\tau)$, $\tilde{P}_{12}(\tau)$, as defined in appendix A, are continuous and have full rank, uniformly over $\mathcal{B} \times \Theta \times \Gamma \times \mathcal{T}$.*

**Assumption 11.** *$\pi(Z) \equiv \pi(Z;\gamma)$, with $\dim(\gamma) < \infty$. $\pi(Z;\gamma)$ is continuously differentiable with respect to $\gamma$. Moreover, there exists an asymptotically linear estimator $\hat{\gamma}$ that admits the following representation: $\hat{\gamma} - \gamma = -H^{-1}\frac{1}{n}\sum_{i=1}^{n} s(d_i, z_i; \gamma) + o_P\left(\frac{1}{\sqrt{n}}\right)$.*

**Assumption 12.** *Let $C_{D|X_2}(u,v) \equiv C_{D|X_2}(u,v;\theta_D)$, with $\dim(\theta_D) < \infty$ for $D = 0, 1$. $C_{D|X_2}(u,v;\theta_D)$ is uniformly continuous and differentiable with respect to its arguments a.e.. Its density, $c_{D|X_2}(u,v;\theta_D)$, is well-defined and finite.*

Although the model (1)-(2) allows for nonlinear quantile functions, assumption 7 is standard in the literature and convenient from a computational point of view.[9] Assumption 9 restricts the analysis to dependent variables that have a well-defined and finite conditional density. Assumption 10 requires the existence of moments and their full rank in order to derive the asymptotic variance of the estimator. Assumptions 6 and 8 are regularity conditions.

Assumption 11 is made for simplicity, and it is satisfied by several estimation methods, including maximum likelihood. This assumption could be relaxed at the cost of assuming that the propensity score is smooth enough, but unknown. The smoothness is required to ensure that the RQR estimator achieves the $\sqrt{n}$ convergence rate, even if a nonparametric

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

estimator of the propensity score is used, as long as it is based on a bias-reduction method. This extension is considered in appendix 4.1.

Similarly, the dependence of the copula on a finite number of parameters is made for simplicity, and assumption 12 could be relaxed to allow for a more complex dependence. Since the identification result of proposition 1 is satisfied when the copula is real analytic, the Bernstein copula becomes a natural flexible choice to consider. This extension is explored in appendix 4.2. Moreover, assumption 12 is satisfied by most common choices of copulas, including the Gaussian or the Clayton. In any case, even if the copula is not well-specified, it would be possible to estimate bounds on the parameters of interest by using Fréchet bounds, as shown in appendix 4.3.

## 3.1 Estimation of the Structural Functions

The estimation is done in three steps. The first step is the most straightforward, and it consists in the estimation of the propensity score: $\hat{\pi}(z_i) \equiv \pi(z_i, \hat{\gamma})$. The second step consists in the estimation of $\theta_1$ and $\theta_0$. Conditional on $t \in \Theta$, define the estimator $\hat{\beta}_1(\tau; t)$ as

$$\hat{\beta}_1(\tau; t) \equiv \arg\min_{b \in \mathcal{B}} \sum_{i=1}^{N} d_i \rho_{\hat{H}_{X,i,\tau}}(y_i - x_i' b) \tag{15}$$

where $\rho_u(x) \equiv xu\mathbf{1}(x \geq 0) - (1-u)x\mathbf{1}(x < 0)$ is the check function, and $\hat{H}_{X,i,\tau} \equiv H_X\left(\tau, \hat{\pi}(z_i); \hat{\theta}_1\right)$. Then, $\theta_1$ is estimated by

$$\hat{\theta}_1 \equiv \arg\min_{t \in \Theta} \left\| \sum_{i=1}^{N} \int_0^1 d_i \varphi(\tau, z_i) \left[\mathbf{1}\left(y_i \leq x_i' \hat{\beta}_1(\tau; t)\right) - H_X(\tau, \hat{\pi}(z_i); t)\right] d\tau \right\| \tag{16}$$

where $\varphi(\tau, z_i)$ is an instrument function.[10] Similarly, $\beta_0(\tau; t)$ and $\theta_0$ are respectively

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

estimated by

$$\hat{\beta}_0(\tau;t) \equiv \arg\min_{b\in\mathcal{B}} \sum_{i=1}^{N} (1-d_i)\,\rho_{\hat{G}_{X,i,\tau}}(y_i - x_i'b) \tag{17}$$

$$\hat{\theta}_0 \equiv \arg\min_{t\in\Theta} \left\| \sum_{i=1}^{N} \int_0^1 (1-d_i)\,\varphi(\tau,z_i)\left[\mathbf{1}\left(y_i \le x_i'\hat{\beta}_0(\tau;t)\right) - G_X\left(\tau,\hat{\pi}(z_i);t\right)\right] d\tau \right\| \tag{18}$$

where $\hat{G}_{X,i,\tau} \equiv G_X\left(\tau,\hat{\pi}(z_i);\hat{\theta}_0\right)$. Finally, $\beta_1(\tau)$ and $\beta_0(\tau)$ are estimated in the third step as $\hat{\beta}_1(\tau) \equiv \hat{\beta}_1\left(\tau;\hat{\theta}_1\right)$ and $\hat{\beta}_0(\tau) \equiv \hat{\beta}_0\left(\tau;\hat{\theta}_0\right)$.

## 3.2 Estimation of the Treatment Effects

Estimation of equations 12-11 is straightforward using the sample analog and substituting the propensity score, the copula parameters, and the quantile estimates by those presented in section 3.1. Hence, the estimated MTE, TUT, TT, and ATE are given by

$$\hat{\Delta}^{MTE}(x_i,z_{1i}) = \int_0^1 x_i'\hat{\beta}_1(\tau)\,d\hat{C}_{1,X,i,\tau} - \int_0^1 x_i'\hat{\beta}_0(\tau)\,d\hat{C}_{0,X,i,\tau} \tag{19}$$

$$\hat{\Delta}^{TUT}(z_i) = \int_0^1 x_i'\hat{\beta}_1(\tau)\,d\hat{G}'_{X,i,\tau} - \int_0^1 x_i'\hat{\beta}_0(\tau)\,d\hat{G}_{X,i,\tau} \tag{20}$$

$$\hat{\Delta}^{TT}(z_i) = \int_0^1 x_i'\hat{\beta}_1(\tau)\,d\hat{H}_{X,i,\tau} - \int_0^1 x_i'\hat{\beta}_0(\tau)\,d\hat{H}'_{X,i,\tau} \tag{21}$$

$$\hat{\Delta}^{ATE}(z_i) = \int_0^1 x_i'\left(\hat{\beta}_1(\tau) - \hat{\beta}_0(\tau)\right) d\tau \tag{22}$$

where $\hat{G}'_{X,i,\tau} \equiv G'_X\left(\tau,\hat{\pi}(z_i);\hat{\theta}_0\right)$, $\hat{H}'_{X,i,\tau} \equiv H'_X\left(\tau,\hat{\pi}(z_i);\hat{\theta}_1\right)$, $\hat{C}_{0,X,i,\tau} \equiv C_{0,X}\left(\tau,\hat{\pi}(z_i);\hat{\theta}_0\right)$, and $\hat{C}_{1,X,i,\tau} \equiv C_{1,X}\left(\tau,\hat{\pi}(z_i);\hat{\theta}_1\right)$. Finally, the unconditional counterparts of these treatment

effects are obtained by averaging over $i = 1, ..., N$.

### 3.3 Asymptotic Distribution

**Theorem 1.** *Let* $\hat{\vartheta}(\tau) \equiv \left(\hat{\beta}_1(\tau)', \hat{\beta}_0(\tau)', \hat{\theta}_1', \hat{\theta}_0', \hat{\gamma}'\right)'$, *where* $\hat{\beta}_j(\tau)$ *and* $\hat{\theta}_j$ *for* $j = 0, 1$ *be the estimators defined in equations 15-18. Under assumptions 1-12, their joint asymptotic distribution is given by:*

$$\sqrt{n}\left(\hat{\vartheta}(\cdot) - \vartheta(\cdot)\right) \Rightarrow \mathbb{G}_2(\cdot)$$

*where* $\mathbb{G}_2(\cdot)$ *is a zero-mean Gaussian process with covariance function* $\Sigma_2(\tau, \tau')$, *which is defined in the proof.*

## 4 Extensions

### 4.1 Nonparametric First Stage

By assumption 11, the propensity score depends on a finite number of parameters. This assumption is used for simplicity, but if it were not true, it would be possible to use a more flexible estimator of the propensity score, such as Klein and Spady (1993) or the Nadaraya-Watson estimator. For the case at hand, it would require to overcome two issues:

1. The functions $G_X$ and $H_X$ respectively depend on $1-\pi(z)$ and $\pi(z)$, so if the estimator equals 0 or 1 for some values of $Z$, then the estimator would not be properly defined. This can be solved by adding a trimming term $\epsilon$ that considers only observations that lie on $[\epsilon, 1-\epsilon]$, where $\epsilon$ vanishes to zero as the sample size increases.

2. If the number of continuous covariates is large enough, it is possible that the RQR estimator does not admit an asymptotically linear representation. This issue has been studied by Newey (1991), Newey (1994), Andrews (1994), and Ichimura and Newey (2017), who establish conditions under which semiparametric estimators are

asymptotically linear. One way to get around this is to impose smoothness assumptions on the propensity score, and use bias-reduction methods that insure that the uniform convergence rate of the asymptotic bias is fast enough. For example, Heckman et al. (1998) use a kernel whose moments up to a high enough order equal zero. The same strategy could be used for the propensity score in this paper, and the RQR estimator would retain the $\sqrt{n}$ convergence rate.

## 4.2 Bernstein Copula

Because the parametric assumption on the copula can be strong, one possibility is to use Bernstein copulas in the estimation. They are a flexible family of copulas that can arbitrarily approximate any other copula (Sancetta and Satchell, 2004). Its cumulative distribution is given by

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

where $M$ is the order of the copula, and $P_{m,M}(u) = \binom{M}{m} u^m (1-u)^{M-m}$. The density of this copula has a similar form, making it is very convenient to implement.[11] Because the $P_{m,M}$ terms are known, the estimation of the copula amounts to the estimation of the $\alpha$ terms. Let $A_j$ denote the matrix that stacks the $\alpha\left(\frac{m_u}{M}, \frac{m_v}{M}\right)$ parameters for $j = 0, 1$. It is possible to estimate $A_j$, together with $\beta_j(\tau)$ as in equations 16 and 18.[12]

## 4.3 Fréchet Bounds

As shown by Fréchet, a bivariate copula can be bounded by $\max\{u+v-1, 0\} \leq C(u,v) \leq \min\{u, v\}$. Hence, one can use them to estimate bounds on equation 1. Specifically, applying

[11] For completeness, define $\beta\left(\frac{m_u}{M}, \frac{m_v}{M}\right) = \alpha\left(\frac{m_u+1}{M}, \frac{m_v+1}{M}\right) - \alpha\left(\frac{m_u+1}{M}, \frac{m_v}{M}\right) - \alpha\left(\frac{m_u}{M}, \frac{m_v+1}{M}\right) + \alpha\left(\frac{m_u}{M}, \frac{m_v}{M}\right)$. The density is given by $c(u,v) = \sum_{m_u=0}^{M} \sum_{m_v=0}^{M} \beta\left(\frac{m_u}{M}, \frac{m_v}{M}\right) P_{m_u,M}(u) P_{m_v,M}(v)$.

[12] Sancetta and Satchell (2004) propose a way to estimate the copula using the realizations of the copula. However, these are not observed in the data, and even then, $V$ can only be bounded using the propensity score: either $V \leq \pi(Z)$ if $D = 1$, or $V \geq \pi(Z)$ if $D = 0$.

the Fréchet bounds to the $G_X$ and $H_X$ functions yields

$$\min\left\{1, \frac{\tau}{1-\pi(z)}\right\} \leq G_X(\tau, \pi(z)) \leq \max\left\{0, \frac{\tau-\pi(z)}{1-\pi(z)}\right\} \tag{23}$$

$$\max\left\{\frac{\tau+\pi(z)-1}{\pi(z)}, 0\right\} \leq H_X(\tau, \pi(z)) \leq \min\left\{\frac{\tau}{\pi(z)}, 1\right\} \tag{24}$$

Combining equations 23 and 24 with equations 3 and 4, respectively, yields

$$\inf_{z\in\mathcal{Z}_X} F^{-1}_{Y|D=0,Z}\left(\min\left\{1, \frac{\tau}{1-\pi(z)}\right\}|z\right) \leq g_0(x,\tau) \leq \sup_{z\in\mathcal{Z}_X} F^{-1}_{Y|D=0,Z}\left(\max\left\{0, \frac{\tau-\pi(z)}{1-\pi(z)}\right\}|z\right)$$

$$\sup_{z\in\mathcal{Z}_X} F^{-1}_{Y|D=1,Z}\left(\max\left\{\frac{\tau+\pi(z)-1}{\pi(z)}, 0\right\}|z\right) \leq g_1(x,\tau) \leq \inf_{z\in\mathcal{Z}_X} F^{-1}_{Y|D=1,Z}\left(\min\left\{\frac{\tau}{\pi(z)}, 1\right\}|z\right)$$

Thus, even if the copula is not analytic, or if the instrument does not have continuous variation, it is possible to set-identify $g_D$ for $D = 0, 1$ and estimate those bounds.

# 5 Conclusion

In this paper I study the identification of a triangular non-separable model with a binary endogenous treatment. Nonparametric identification of the structural quantile functions is achieved by using local variation of the instrument combined with the properties of real analytic copulas, even without imposing rank invariance or rank similarity.

I propose a three-step quantile regression estimator to estimate the SQF, the copula of the unobservables, and the propensity score. The baseline estimator is based on a parametric copula, and in an extension I consider using Bernstein copulas, which are a flexible family that can approximate any well-defined copula.

# References


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.

Andrews, D. W. (1994). Empirical process methods in econometrics. *Handbook of econometrics 4*, 2247–2294.

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.

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

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

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.

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.

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.

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, Q., Q. Vuong, and H. Xu (2016). Estimation of heterogeneous individual treatment effects with endogenous treatments. Technical report, Mimeo.

Firpo, S. and G. Ridder (2008). Bounds on functionals of the distribution of treatment effects. Technical report, Mimeo.

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

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.

Heckman, J. J., H. Ichimura, and P. Todd (1998). Matching as an econometric evaluation estimator. *The review of economic studies 65*(2), 261–294.

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

Heckman, J. J. and E. J. Vytlacil (2007). 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.

Ichimura, H. and W. K. Newey (2017). The influence function of semiparametric estimators. Technical report, Mimeo.

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.

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

Klein, R. W. and R. H. Spady (1993). An efficient semiparametric estimator for binary response models. *Econometrica: Journal of the Econometric Society 61*(2), 387–421.

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

Lee, L.-f. (2007). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics 140*(2), 333–374.

Masten, M. (2017). Random coefficients on endogenous variables in simultaneous equations models. *The Review of Economic Studies 0*, 1 – 58.

Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica: Journal of the Econometric Society 59*(4), 1161–1167.

Newey, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica 62*(6), 1349–1382.

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

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.

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.

Wüthrich, K. et al. (2015). Semiparametric estimation of quantile treatment effects with endogeneity. Technical report, Universitaet Bern, Departement Volkswirtschaft.

Let $W \equiv (Y, D, Z)$. The following notation is used throughout the appendix:[13]

$$f \mapsto \mathbb{E}_n [f(W)] \equiv \frac{1}{n} \sum_{i=1}^{n} f(W)$$

$$f \mapsto \mathbb{G}_n [f(W)] \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f(W) - \mathbb{E}(f(W))$$

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

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

$$Q_n(\beta, \theta, \gamma, \tau) \equiv \mathbb{E}_n [q(W, \beta, \theta, \gamma, \tau)]$$

$$Q(\beta, \theta, \gamma, \tau) \equiv \mathbb{E}[q(W, \beta, \theta, \gamma, \tau)]$$

where I have used $\rho_\tau(u) \equiv (\tau - \mathbf{1}(u < 0))u$, $\zeta_\tau(u) \equiv (\mathbf{1}(u < 0) - \tau)$, $\varepsilon_D(\tau) \equiv Y -$

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

$X'\beta_D(\tau)$, and $\hat{\varepsilon}_D(\tau) \equiv Y - X'\hat{\beta}_D(\tau)$, $\vartheta(\tau) \equiv \left(\beta(\tau)', \theta', \gamma'\right)'$.

# A Mathematical proofs

## A.1 Proof of Lemma 1

By assumption 3 and equations 3 and 4, the first part of the lemma follows immediately.

Let $x \in \mathcal{X}$, and $G_X$ satisfy equation 5. Pick a $z_x \in \mathcal{Z}_x$, and define $F_{Y_0^*|X}(y|x) \equiv G_X^{-1}\left(F_{Y|D=0,Z}(y|z_x), \pi(z_x)\right)$. For all $(z, y)$ in the support of $(Z, Y)$ given $X = x$, we have:

$$\begin{aligned} G_X\left(F_{Y_0^*|X}(y|x), \pi(z)\right) &= G_X\left(G_X^{-1}\left(F_{Y|D=0,Z}(y|z_x), \pi(z_x)\right), \pi(z)\right) \\ &= F_{Y|D=0,Z}\left(F_{Y|D=0,Z}^{-1}\left(F_{Y|D=0,Z}(y|z_x)|z_x\right)|z\right) \\ &= F_{Y|D=0,Z}(y|z) \end{aligned}$$

By a parallel argument, for $H_X$ satisfying equation 6, one can get that for all $(z, y)$ in the support of $(Z, Y)$ given $X = x$, we have $H_X\left(F_{Y_1^*|X}(y|x), \pi(z)\right) = F_{Y|D=1,Z}(y|z)$.

## A.2 Proof of Proposition 1

Let $G_X$ and $\tilde{G}_X$ satisfy equation 3, and $\pi_1, \pi_2 \in \mathcal{O}_x$. Then,

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

Hence, $\forall \tau \in (0,1)$, $(\pi_1, \pi_2) \to G_X\left(G_X^{-1}(\tau, \pi_2), \pi_1\right) - \tilde{G}_X\left(\tilde{G}_X^{-1}(\tau, \pi_2), \pi_1\right)$, which is real analytic by assumption 5, is zero on a product of two open neighborhoods. Hence, it is zero everywhere on $(0,1) \times (0,1)$, and taking limits at $\pi_2 = 0$ yields

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

Hence, $G_X(\tau, \pi_1)$ and $\tilde{G}_X(\tau, \pi_1)$ coincide on $(0,1) \times (0,1)$. Consequently, $G_X$ is identified,

and so are $C_{0,X}$ and $g_0(x,u)$. By a parallel argument, using equation 4 and taking limits at $\pi_2 = 1$, $H_X$, $C_{1,X}$, and $g_1(x,u)$ are identified.

## A.3 Proof of Theorem 1

First I show consistency of the estimator. By assumptions 7, 9, 11, and 12, $Q(\beta,\theta,\gamma,\tau)$ is continuous over $\mathcal{B}\times\Theta\times\Gamma\times\mathcal{T}$. By lemma 5, $\sup_{(\beta,\theta,\gamma)\in\mathcal{B}\times\Theta\times\Gamma}\|Q_n(\beta,\theta,\gamma,\tau) - Q(\beta,\theta,\gamma,\tau)\| \xrightarrow{P} 0$. Thus, by lemma 4, $\sup_{\tau\in\mathcal{T}}\left\|\hat{\vartheta}(\tau) - \vartheta(\tau)\right\| \xrightarrow{P} 0$.

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

$$O\left(\frac{1}{\sqrt{n}}\right) = \sqrt{n}\mathbb{E}_n\left[DX\zeta_{H_X\left(\tau,\pi(Z;\hat{\gamma}),\hat{\theta}_1\right)}\left(\hat{\varepsilon}_1(\tau)\right)\right]$$

By lemma 5 and assumption 10, the following expansion holds in $\ell^\infty(\mathcal{T})$:

$$\begin{aligned}
O\left(\frac{1}{\sqrt{n}}\right) &= \mathbb{G}_n\left[DX\zeta_{H_X\left(\tau,\pi(Z;\hat{\gamma});\hat{\theta}_1\right)}\left(\hat{\varepsilon}_1(\tau)\right)\right] + \sqrt{n}\mathbb{E}\left[DX\zeta_{H_X\left(\tau,\pi(Z;\hat{\gamma});\hat{\theta}_1\right)}\left(\hat{\varepsilon}_1(\tau)\right)\right] \\
&= \mathbb{G}_n\left[DX\zeta_{H_X(\tau,\pi(Z;\gamma);\theta_1)}\left(\varepsilon_1(\tau)\right)\right] + o_P(1) + \sqrt{n}\mathbb{E}\left[DX\zeta_{H_X\left(\tau,\pi(Z;\hat{\gamma});\hat{\theta}_1\right)}\left(\hat{\varepsilon}_1(\tau)\right)\right] \\
&= \mathbb{G}_n\left[DX\zeta_{H_X(\tau,\pi(Z;\gamma);\theta_1)}\left(\varepsilon_1(\tau)\right)\right] + J_1(\tau)\sqrt{n}\left(\hat{\beta}_1(\tau) - \beta_1(\tau)\right) \\
&- P_{11}(\tau)\sqrt{n}(\hat{\gamma} - \gamma) - P_{12}(\tau)\sqrt{n}\left(\hat{\theta}_1 - \theta_1\right) + o_P(1)
\end{aligned}$$

where

$$J_1(\tau) \equiv \frac{\partial\mathbb{E}\left[DX\zeta_{H_X(\tau,\pi(Z;\gamma);\theta_1)}\left(\varepsilon_1(\tau)\right)\right]}{\partial\beta_1}$$

$$P_{11}(\tau) \equiv -\frac{\partial\mathbb{E}\left[DX\zeta_{H_X(\tau,\pi(Z;\gamma);\theta_1)}\left(\varepsilon_1(\tau)\right)\right]}{\partial\gamma}$$

$$P_{12}(\tau) \equiv -\frac{\partial \mathbb{E}\left[DX\zeta_{H_X(\tau,\pi(Z;\gamma);\theta_1)}(\varepsilon_1(\tau))\right]}{\partial \theta_1}$$

Rearranging and solving for $\sqrt{n}\left(\hat{\beta}_1(\tau) - \beta_1(\tau)\right)$,

$$\begin{aligned}\sqrt{n}\left(\hat{\beta}_1(\tau) - \beta_1(\tau)\right) = -J_1(\tau)^{-1}\Big\{&\mathbb{G}_n\left[DX\zeta_{H_X(\tau,\pi(Z;\gamma);\theta_1)}(\varepsilon_1(\tau))\right] \\ &-P_{11}(\tau)\sqrt{n}(\hat{\gamma}-\gamma) - P_{12}(\tau)\sqrt{n}\left(\hat{\theta}_1 - \theta_1\right)\Big\} + o_P(1) \qquad (25)\end{aligned}$$

in $\ell^\infty(\mathcal{T})$. By a parallel argument, it can be shown that

$$\begin{aligned}\sqrt{n}\left(\hat{\beta}_0(\tau) - \beta_0(\tau)\right) = -J_0(\tau)^{-1}\Big\{&\mathbb{G}_n\left[(1-D)X\zeta_{G_X(\tau,\pi(Z;\gamma);\theta_0)}(\varepsilon_0(\tau))\right] \\ &-P_{01}(\tau)\sqrt{n}(\hat{\gamma}-\gamma) - P_{02}(\tau)\sqrt{n}\left(\hat{\theta}_0 - \theta_0\right)\Big\} + o_P(1) \qquad (26)\end{aligned}$$

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

$$J_0(\tau) \equiv \frac{\partial \mathbb{E}\left[(1-D)X\zeta_{G_X(\tau,\pi(Z;\gamma);\theta_0)}(\varepsilon_0(\tau))\right]}{\partial \beta_0}$$

$$P_{01}(\tau) \equiv -\frac{\partial \mathbb{E}\left[(1-D)X\zeta_{G_X(\tau,\pi(Z;\gamma);\theta_0)}(\varepsilon_0(\tau))\right]}{\partial \gamma}$$

$$P_{02}(\tau) \equiv -\frac{\partial \mathbb{E}\left[(1-D)X\zeta_{G_X(\tau,\pi(Z;\gamma);\theta_0)}(\varepsilon_0(\tau))\right]}{\partial \theta_0}$$

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

$$O\left(\frac{1}{\sqrt{n}}\right) = \sqrt{n}\mathbb{E}_n\left[\int_0^1 D\varphi(u,Z)\,\zeta_{H_X\left(u,\pi(Z;\hat{\gamma});\hat{\theta}_1\right)}(\varepsilon_1(u))\,du\right]$$

By lemma 5 and assumption 10, the following expansion holds:

$$\begin{aligned}
O\left(\frac{1}{\sqrt{n}}\right) &= \mathbb{G}_n\left[\int_0^1 D\varphi\left(u,Z\right)\zeta_{H_X\left(u,\pi\left(Z;\hat{\gamma}\right);\hat{\theta}_1\right)}\left(\hat{\varepsilon}_1\left(u\right)\right)du\right] \\
&+ \sqrt{n}\int_0^1 \mathbb{E}\left[D\varphi\left(u,Z\right)\zeta_{H_X\left(u,\pi\left(Z;\hat{\gamma}\right);\hat{\theta}_1\right)}\left(\hat{\varepsilon}_1\left(u\right)\right)\right]du \\
&= \mathbb{G}_n\left[\int_0^1 D\varphi\left(u,Z\right)\zeta_{H_X\left(u,\pi\left(Z;\gamma\right);\theta_1\right)}\left(\varepsilon_1\left(u\right)\right)du\right] + o_P\left(1\right) \\
&+ \sqrt{n}\int_0^1 \mathbb{E}\left[D\varphi\left(u,Z\right)\zeta_{H_X\left(u,\pi\left(Z;\hat{\gamma}\right);\hat{\theta}_1\right)}\left(\hat{\varepsilon}_1\left(u\right)\right)\right]du \\
&= \mathbb{G}_n\left[\int_0^1 D\varphi\left(u,Z\right)\zeta_{H_X\left(u,\pi\left(Z;\gamma\right);\theta_1\right)}\left(\varepsilon_1\left(u\right)\right)du\right] + \sqrt{n}\int_0^1 \tilde{J}_1\left(u\right)\left(\hat{\beta}_1\left(u\right)-\beta_1\left(u\right)\right)du \\
&- \sqrt{n}\int_0^1 \tilde{P}_{12}\left(u\right)du\left(\hat{\theta}_1-\theta_1\right) - \sqrt{n}\int_0^1 \tilde{P}_{11}\left(u\right)du\left(\hat{\gamma}-\gamma\right) + o_P\left(1\right)
\end{aligned}$$

where

$$\tilde{J}_1\left(\tau\right) \equiv \frac{\partial \mathbb{E}\left[D\varphi\left(\tau,Z\right)\zeta_{H_X\left(\tau,\pi\left(Z;\gamma\right);\theta_1\right)}\left(\varepsilon_1\left(\tau\right)\right)\right]}{\partial \beta_1}$$

$$\tilde{P}_{11}\left(\tau\right) \equiv -\frac{\partial \mathbb{E}\left[D\varphi\left(\tau,Z\right)\zeta_{H_X\left(\tau,\pi\left(Z;\gamma\right);\theta_1\right)}\left(\varepsilon_1\left(\tau\right)\right)\right]}{\partial \gamma}$$

$$\tilde{P}_{12}\left(\tau\right) \equiv -\frac{\partial \mathbb{E}\left[D\varphi\left(\tau,Z\right)\zeta_{H_X\left(\tau,\pi\left(Z;\gamma\right);\theta_1\right)}\left(\varepsilon_1\left(\tau\right)\right)\right]}{\partial \theta_1}$$

Rearranging and solving for $\sqrt{n}\left(\hat{\theta}_1-\theta_1\right)$,

$$\begin{aligned}
\sqrt{n}\left(\hat{\theta}_1-\theta_1\right) &= \left[\int_0^1 \tilde{P}_{12}\left(u\right)du\right]^{-1}\left\{\mathbb{G}_n\left[\int_0^1 D\varphi\left(u,Z\right)\zeta_{H_X\left(u,\pi\left(Z;\gamma\right);\theta_1\right)}\left(\varepsilon_1\left(u\right)\right)du\right]\right. \\
&\left. + \sqrt{n}\int_0^1 \tilde{J}_1\left(u\right)\left(\hat{\beta}_1\left(u\right)-\beta_1\left(u\right)\right)du - \sqrt{n}\int_0^1 \tilde{P}_{11}\left(u\right)du\left(\hat{\gamma}-\gamma\right)\right\} + o_P\left(1\right)
\end{aligned} \tag{27}$$

By a parallel argument, it can be shown that

$$\sqrt{n}\left(\hat{\theta}_0 - \theta_0\right) = \left[\int_0^1 \tilde{P}_{02}(u)\,du\right]^{-1}\left\{\mathbb{G}_n\left[\int_0^1 (1-D)\,\varphi(u,Z)\,\zeta_{G_X(u,\pi(Z;\gamma);\theta_0)}(\varepsilon_0(u))\,du\right] + \sqrt{n}\int_0^1 \tilde{J}_0(u)\left(\hat{\beta}_0(u) - \beta_0(u)\right)du - \sqrt{n}\int_0^1 \tilde{P}_{01}(u)\,du\,(\hat{\gamma}-\gamma)\right\} + o_P(1) \tag{28}$$

where

$$\tilde{J}_0(\tau) \equiv \frac{\partial \mathbb{E}\left[(1-D)\,\varphi(\tau,Z)\,\zeta_{G_X(\tau,\pi(Z;\gamma);\theta_0)}(\varepsilon_0(\tau))\right]}{\partial \beta_0}$$

$$\tilde{P}_{01}(\tau) \equiv -\frac{\partial \mathbb{E}\left[(1-D)\,\varphi(\tau,Z)\,\zeta_{G_X(\tau,\pi(Z;\gamma);\theta_0)}(\varepsilon_0(\tau))\right]}{\partial \gamma}$$

$$\tilde{P}_{02}(\tau) \equiv -\frac{\partial \mathbb{E}\left[(1-D)\,\varphi(\tau,Z)\,\zeta_{G_X(\tau,\pi(Z;\gamma);\theta_0)}(\varepsilon_0(\tau))\right]}{\partial \theta_0}$$

Now define

$$A(\tau) \equiv \hat{\vartheta}(\tau) - \vartheta(\tau)$$

$$C(\tau) \equiv \begin{bmatrix} -J_1(\tau)^{-1} & 0 & 0 & 0 & 0 \\ 0 & -J_0(\tau)^{-1} & 0 & 0 & 0 \\ 0 & 0 & \left[\int_0^1 \tilde{P}_{12}(u)\,du\right]^{-1} & 0 & 0 \\ 0 & 0 & 0 & \left[\int_0^1 \tilde{P}_{02}(u)\,du\right]^{-1} & 0 \\ 0 & 0 & 0 & 0 & -H^{-1} \end{bmatrix}$$

$$D(\tau) \equiv \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \left[\int_0^1 \tilde{P}_{12}(u)\,du\right]^{-1} \tilde{J}_1(\tau) & 0 & 0 & 0 & -\left[\int_0^1 \tilde{P}_{12}(u)\,du\right]^{-1} \tilde{P}_{11}(\tau) \\ 0 & \left[\int_0^1 \tilde{P}_{02}(u)\,du\right]^{-1} \tilde{J}_0(\tau) & 0 & 0 & -\left[\int_0^1 \tilde{P}_{02}(u)\,du\right]^{-1} \tilde{P}_{01}(\tau) \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$F(\tau) \equiv \begin{bmatrix} 0 & 0 & J_1(\tau)^{-1} P_{12}(\tau) & 0 & J_1(\tau)^{-1} P_{11}(\tau) \\ 0 & 0 & 0 & J_0(\tau)^{-1} P_{02}(\tau) & J_0(\tau)^{-1} P_{01}(\tau) \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\psi(\tau) \equiv r(W, \beta(\tau), \theta, \gamma, \tau)$$

Combining equations 25, 26, 27, and 28 yields

$$A(\tau) = F(\tau) A(\tau) + \int_0^1 D(u) A(u)\,du + C(\tau) \frac{1}{\sqrt{n}} \mathbb{G}_n[\psi(\tau)] + o_P\left(\frac{1}{\sqrt{n}}\right) \tag{29}$$

in $\ell^\infty(\mathcal{T})$. Equation 29 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 3, the solution to this equation is given by:

$$\begin{aligned} \sqrt{n} A(\tau) = {} & F^I(\tau) \left(I - \int_0^1 D(u) F^I(u)\,du\right)^{-1} \int_0^1 D(u) F^I(u) C(u) \mathbb{G}_n[\psi(u)]\,du \\ & + F^I(\tau) C(\tau) \mathbb{G}_n[\psi(\tau)] + o_P(1) \end{aligned} \tag{30}$$

in $\ell^\infty(\mathcal{T})$, where $F^I(\tau) \equiv (I - F(\tau))^{-1} = I + F(\tau)$. Using the Functional Delta Method

and lemmas 2 and 4, it follows that

$$\sqrt{n}\left(\hat{\vartheta}(\tau)-\vartheta(\tau)\right)\Rightarrow\mathbb{G}_2(\tau) \tag{31}$$

where

$$\mathbb{G}_2(\tau)\equiv F^I(\tau)\left[C(\tau)+\left(I-\int_0^1 D(u)F^I(u)\,du\right)^{-1}\int_0^1 D(u)F^I(u)C(u)\,du\right]\mathbb{G}_1(\tau)$$

which is a zero-mean Gaussian process with covariance $\Sigma_2(\tau,\tau')$, where

$$\begin{aligned}\Sigma_2(\tau,\tau')=F^I(\tau)&\left[C(\tau)+\left(I-\int_0^1 D(u)F^I(u)\,du\right)^{-1}\int_0^1 D(u)F^I(u)C(u)\,du\right]\Sigma_1(\tau,\tau')\\&\left\{F^I(\tau')\left[C(\tau')+\left(I-\int_0^1 D(u)F^I(u)\,du\right)^{-1}\int_0^1 D(u)F^I(u)C(u)\,du\right]\right\}'\end{aligned}$$

and $\Sigma_1(\tau,\tau')$ is defined in lemma 5.

# B Auxiliary Lemmas

## B.1 Hadamard Derivative of $\int_0^1 R(u)S(u)\,du$ with respect to $S(u)$

**Lemma 2.** *Let the operator $\mathcal{R}:\ell^\infty(\mathcal{T})\to\mathbb{R}$ defined by $\mathcal{R}(S(\cdot))=\int_0^1 R(\cdot)S(\cdot)\,d\cdot$. Define $R(h_t)\equiv\int_0^1 R(u)(S(u)+th_t(u))\,du$. As $t\to 0$,*

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

*where $D_h\equiv\int_0^1 R(u)h(u)\,du$. 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.*

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

□

## B.2 Solution to the Fredholm Integral Equation

**Lemma 3.** *Let $M(\tau) = N_1(\tau)M(\tau) + N_2(\tau) + \int_0^1 N_3(u)M(u)\,du$ be a Fredholm integral equation of the second kind. Moreover, define $\tilde{N}_2(\tau) \equiv (I - N_1(\tau))^{-1} N_2(\tau)$ and $\tilde{N}_3(\tau) \equiv N_3(\tau)(I - N_1(\tau))^{-1}$. Let*

*(i) $I - N_1(\tau)$ is invertible $\forall \tau \in [0,1]$*

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

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

$$M(\tau) = \tilde{N}_2(\tau) + (I - N_1(\tau))^{-1}\left(I - \int_0^1 \tilde{N}_3(u)\,du\right)^{-1}\int_0^1 \tilde{N}_3(u)N_2(u)\,du$$

*Proof.*

$$\begin{aligned}
M(\tau) &= N_1(\tau)M(\tau) + N_2(\tau) + \int_0^1 N_3(u)M(u)\,du \\
&= \tilde{N}_2(\tau) + (I - N_1(\tau))^{-1}\int_0^1 N_3(u)M(u)\,du \\
&= \tilde{N}_2(\tau) + (I - N_1(\tau))^{-1}\sum_{n=0}^{\infty}\left[\int_0^1 \tilde{N}_3(u)\,du\right]^n\int_0^1 \tilde{N}_3(u)N_2(u)\,du \\
&+ \lim_{n\to\infty}(I - N_1(\tau))^{-1}\left[\int_0^1 \tilde{N}_3(u)\,du\right]^n\int_0^1 N_3(u)M(u)\,du \\
&= \tilde{N}_2(\tau) + (I - N_1(\tau))^{-1}\left(I - \int_0^1 \tilde{N}_3(u)\,du\right)^{-1}\int_0^1 \tilde{N}_3(u)N_2(u)\,du
\end{aligned}$$

where the second equality follows by (i), the third one by iteratively substituting $M(u)$ 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.3 Argmax Process

**Lemma 4.** *(Chernozhukov and Hansen, 2006) Suppose that uniformly in $\pi$ in a compact set $\Pi$ 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.4 Stochastic Expansion

**Lemma 5.** *Under assumptions 6-12, the following statements hold:*

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. $\mathbb{G}_n r(W,\beta(\tau),\theta,\gamma,\tau) \Rightarrow \mathbb{G}_1(\tau)$ *in* $\ell^\infty(\mathcal{T})$, *where* $\mathbb{G}_1(\tau)$ *is a zero-mean Gaussian process with covariance* $\Sigma_1(\tau,\tau')$ *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 $\|f(\tau',z) - f(\tau,z)\| < C(\tau-\tau')^a$ for $C>0$, $a>0$, $\forall(z,\tau,\tau')\,\forall f\in\mathcal{F}$.

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

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

satisfies

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

for some $\delta' < 0$ and $j = 0, 1$. Since the indicator function is bounded and monotone, and the density functions $f_j\left(y|x\right)$ are bounded by assumption 9, the bracketing number of

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

satisfies

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

Since $\mathcal{E}_j$ has a constant envelope, it is Donsker. Now consider the functions $G_X$ and $H_X$. By assumptions 4 and 12, the mean value theorem can be applied to show

$$\left\|G_X\left(\tau, \pi\left(z, \gamma\right); \theta_0\right) - G_X\left(\tau', \pi\left(z, \gamma\right); \theta_0\right)\right\| = \left\|\tau - \tau'\right\| \left\|\frac{\partial}{\partial \tau} G_X\left(\tau'', 1 - \pi\left(z, \gamma\right); \theta_0\right)\right\|$$

for some $\tau''$ between $\tau$ and $\tau'$. By assumptions 4 and 12, the second term is bounded $\forall z, \tau''$, so it follows that $G_X \in \mathcal{F}$.[14] Using a parallel argument, it can be shown that $H_X \in \mathcal{F}$.

[14]To see this, notice that both $\frac{\partial}{\partial \tau} C_{0,X}\left(\tau, \pi\right) \in [0, 1]$ and $\pi\left(\tau\right) \in [0, 1]$. Hence, it suffices to show that $\lim_{\pi \to 1} \frac{\partial}{\partial \tau} G\left(\tau, \pi\right) = \lim_{\pi \to 1} C_{0,X}\left(\tau, \pi\right) < \infty$, where I have used L'Hôpital rule. Since the derivative is

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 $\mathcal{E}_1 \times \mathcal{F} - \mathcal{T} \times \mathcal{F}$, the second subvector is $\mathcal{E}_0 \times \mathcal{F} - \mathcal{T} \times \mathcal{F}$, the third subvector is $\mathcal{E}_1 \times \mathcal{F} - \mathcal{T} \times \mathcal{F}$, the fourth subvector is $\mathcal{E}_0 \times \mathcal{F} - \mathcal{T} \times \mathcal{F}$, and the fifth subvector is $\mathcal{F}$.[15] Since $\mathcal{H}$ is Lipschitz over $(\mathcal{T}, \mathcal{F}, \mathcal{E}_0, \mathcal{E}_1)$, 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 $\tau$ with respect to the supremum norm, it is also Donsker in $\ell^\infty(\mathcal{T})$. Hence,

$$\mathbb{G}_n r(W, \beta(\cdot), \theta, \gamma, \cdot) \Rightarrow \mathbb{G}_1(\tau)$$

with covariate function

$$\Sigma_1(\tau, \tau') \equiv \mathbb{E}\left[\mathbb{G}_1(\tau)\mathbb{G}_1(\tau')'\right] = \begin{bmatrix} \Sigma^{11}(\tau, \tau') & 0 & \Sigma^{13}(\tau')' & 0 & 0 \\ 0 & \Sigma^{22}(\tau, \tau') & 0 & \Sigma^{24}(\tau')' & 0 \\ \Sigma^{13}(\tau) & 0 & \Sigma^{33} & 0 & 0 \\ 0 & \Sigma^{24}(\tau) & 0 & \Sigma^{44} & 0 \\ 0 & 0 & 0 & 0 & \Sigma^{55} \end{bmatrix}$$

bounded by assumption 12, the result follows.

[15] Note that it is immediate to check that $xd$ and $x(1-d) \in \mathcal{F}$.

where

$$\Sigma^{11}(\tau,\tau') = \mathbb{E}\left[d_i\left(H_{X,\tau\wedge\tau'} - H_{X,\tau}H_{X,\tau'}\right)XX'\right]$$

$$\Sigma^{22}(\tau,\tau') = \mathbb{E}\left[(1-D)\left(G_{X,\tau\wedge\tau'} - G_{X,\tau}G_{X,\tau'}\right)XX'\right]$$

$$\Sigma^{13}(\tau) = \mathbb{E}\left[D\int_0^1 X\varphi(u,Z)'\left[H_{X,\tau\wedge u} - H_{X,\tau}H_{X,u}\right]du\right]$$

$$\Sigma^{24}(\tau) = \mathbb{E}\left[(1-D)\int_0^1 X\varphi(u,Z)'\left[G_{X,\tau\wedge u} - G_{X,\tau}G_{X,u}\right]du\right]$$

$$\Sigma^{33} = \mathbb{E}\left[D\int_0^1\int_0^1 \varphi(u,Z)\varphi(v,Z)'\left[H_{X,u\wedge v} - H_{X,u}H_{X,v}\right]dvdu\right]$$

$$\Sigma^{44} = \mathbb{E}\left[(1-D)\int_0^1\int_0^1 \varphi(u,Z)\varphi(v,Z)'\left[G_{X,u\wedge v} - G_{X,u}G_{X,v}\right]dvdu\right]$$

$$\Sigma^{55} = \mathbb{E}\left[s(D,Z;\gamma)\,s(D,Z;\gamma)'\right]$$

where $\wedge$ denotes the minimum between two variables, $\hat{H}_{X,\tau} \equiv H_X(\tau,\pi(Z);\theta_1)$, and $\hat{G}_{X,\tau} \equiv G_X(\tau,\pi(Z);\theta_0)$. Define $\xi$ as the $L_2(P)$ pseudometric on $\mathcal{H}$:

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

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 $\tau$, by

assumption 9, $\delta_n \xrightarrow{p} 0$. 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 Monte Carlo

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

$$y_i = \beta_{d_i,1}\left(\tau_{d_i,i}\right) + x_i\beta_{d_i,2}\left(u_{d_i,i}\right) \tag{32}$$

$$d_i = \mathbf{1}\left(\gamma_1 + x_i\gamma_2 + z_i\gamma_3 + \Lambda^{-1}\left(v_i\right) > 0\right) \tag{33}$$

$$u_{0,i}, u_{1,i}, v_i | z_i \sim Gaussian\left(\Sigma\right) \tag{34}$$

where $\beta_0\left(\tau\right) = \left[5\Phi^{-1}\left(\tau\right) - 2, 1 + 2\frac{\exp(\tau)}{1+\tau}\right]$, $\beta_1\left(\tau\right) = \left[\tan\left(2\left(\tau - 0.5\right)\right) + 5\Phi^{-1}\left(\tau\right) + 2, 2\frac{\exp(\tau)+1}{1+\tau}\right]$, $\gamma = \left(-2, 0.4, 2\right)'$, $\Sigma$ is a symmetric correlation matrix with unit diagonal, and off diagonal $\Sigma_{12} = 0$, $\Sigma_{13} = 0.25$, and $\Sigma_{23} = 0.5$ 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), and Bernstein copulas of 2nd, 5th, and 8th order. On top of those, I compute the estimates of the IVQR estimator.

First note that in terms of the objective functions (table 1), both the parametric copulas have a similar performance, regardless of whether they are correctly (Gaussian) or incorrectly specified (Clayton). However, the difference between the estimated copula and the true one (table 2) is substantially smaller for the correctly specified copula. On the other hand, the Bernstein copula attains a smaller value of the objective function once the order is high enough, and the difference between the estimated copula and the true one is comparable to the correctly specified parametric copula.

The difference in the precision of the estimation of the copula is reflected in the estimates of $\beta$ (table 3): the RQR estimates with the Clayton copula have a small bias, that is also present with the Bernstein copula if the order is low enough. Despite that, even the RQR estimates with a incorrectly specified copula perform better than IVQR. In terms of

the dispersion of the estimates, the reverse is true: the IVQR estimates have a smaller interquantile range, than the RQR estimates. Among the latter, the performance with the Bernstein copula is similar to the one with the correctly specified copula.

Table 1: Objective function

| Copula | Gau | Clay | Bern(2) | Bern(5) | Bern(8) |
|---|---|---|---|---|---|
| Equation (16) | 0.106 | 0.119 | 1.305 | 0.034 | 0.017 |
| Equation (18) | 0.116 | 0.115 | 0.114 | 0.019 | 0.019 |

Table 2: Estimated copula

| Copula | Gau | Clay | Bern(2) | Bern(5) | Bern(8) |
|---|---|---|---|---|---|
| Mean ($C_1$) | 0.008 | 0.014 | 0.013 | 0.008 | 0.009 |
| Sup ($C_1$) | 0.018 | 0.037 | 0.024 | 0.021 | 0.024 |
| Mean ($C_0$) | 0.009 | 0.013 | 0.008 | 0.009 | 0.009 |
| Sup ($C_0$) | 0.018 | 0.034 | 0.017 | 0.020 | 0.021 |

## C.1 Non-Analytical Copula

The identification result presented in this paper relies on the copula being analytic. The following Monte Carlo simulation assesses the performance of the RQR estimator with a Bernstein copula when the true copula is not analytic. In particular, it is a mixture between the lower Fréchet copula and the independence copula, with proportions $(0.25, 0.75)$ for the treated, and $(0.5, 0.5)$ for the untreated. To isolate the performance with this copula, I simplify the data generating process by assuming no covariates, and I use the true propensity score, not the estimated one. The propensity score is assumed to be uniformly distributed over $(0.1, 0.9)$.

As shown in tables 4 and 5, increasing the order of the copula reduces both the objective function and the distance between the estimated copula and the true one. However, even though the mean difference between the estimated Bernstein copula and the true copula is of a similar size to the case in which the true copula is Gaussian, the supremum difference is not. The reason for this is that the copula is non-analytical, which is harder to approximate with the Bernstein copula. Consequently, the RQR estimates display a small bias (table 6).

Table 3: Quantile regression coefficients

| | | $\beta_{1,1}$ | | | | | $\beta_{0,1}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\tau$ | | | | | $\tau$ | | | | |
| | | 0.1 | 0.25 | 0.5 | 0.75 | 0.9 | 0.1 | 0.25 | 0.5 | 0.75 | 0.9 |
| Median | Gau | 0.02 | 0.00 | 0.10 | -0.19 | 0.00 | 0.06 | -0.09 | 0.04 | -0.06 | 0.04 |
| Difference | Clay | -0.09 | 0.17 | 0.00 | -0.31 | -0.02 | -0.04 | -0.94 | -0.05 | -0.60 | 0.00 |
| | Bern(2) | 0.19 | -1.10 | 0.23 | -1.06 | 0.14 | 0.06 | 0.18 | 0.03 | 0.05 | 0.04 |
| | Bern(5) | 0.00 | -0.12 | 0.08 | -0.08 | -0.03 | 0.12 | 0.01 | 0.07 | -0.09 | 0.00 |
| | Bern(8) | -0.01 | -0.08 | 0.03 | 0.10 | 0.00 | 0.13 | -0.03 | 0.06 | -0.10 | 0.02 |
| | IVQR | 0.06 | -0.45 | 0.01 | -0.55 | -0.01 | 0.07 | -0.78 | 0.02 | -0.61 | -0.01 |
| Interquantile | Gau | 4.25 | 7.06 | 3.24 | 5.56 | 3.85 | 3.85 | 6.08 | 3.13 | 4.71 | 2.92 |
| Range | Clay | 4.21 | 7.20 | 3.43 | 5.62 | 3.74 | 4.65 | 8.15 | 3.41 | 5.36 | 2.94 |
| | Bern(2) | 4.25 | 6.58 | 3.36 | 5.00 | 3.52 | 3.84 | 5.87 | 3.14 | 4.49 | 2.93 |
| | Bern(5) | 4.17 | 6.94 | 3.34 | 5.41 | 3.98 | 3.92 | 5.96 | 3.09 | 4.68 | 2.96 |
| | Bern(8) | 4.30 | 7.13 | 3.15 | 5.58 | 3.90 | 3.79 | 6.24 | 3.16 | 4.72 | 2.92 |
| | IVQR | 3.32 | 6.03 | 2.50 | 4.82 | 2.52 | 3.32 | 5.40 | 2.50 | 4.33 | 2.52 |
| | | $\beta_{1,2}$ | | | | | $\beta_{0,2}$ | | | | |
| Median | Gau | -0.05 | 0.02 | -0.06 | 0.06 | -0.36 | -0.09 | -0.03 | 0.02 | -0.02 | 0.05 |
| Difference | Clay | -0.77 | 0.08 | -1.50 | 0.27 | -2.22 | -0.43 | -0.01 | -0.13 | 0.01 | 0.00 |
| | Bern(2) | -1.11 | 0.08 | -1.24 | 0.19 | -1.84 | -0.06 | 0.01 | 0.03 | 0.00 | 0.11 |
| | Bern(5) | 0.29 | 0.10 | -0.07 | 0.23 | -0.81 | -0.09 | -0.06 | 0.11 | -0.02 | 0.19 |
| | Bern(8) | 0.37 | 0.18 | -0.34 | 0.33 | -1.19 | -0.08 | -0.05 | 0.15 | -0.03 | 0.22 |
| | IVQR | -0.81 | -0.02 | -1.10 | 0.01 | -1.42 | -0.37 | -0.02 | -0.23 | 0.01 | -0.20 |
| Interquantile | Gau | 7.37 | 4.77 | 8.65 | 7.53 | 13.21 | 4.40 | 3.03 | 4.56 | 3.45 | 5.25 |
| Range | Clay | 6.80 | 4.17 | 6.78 | 5.33 | 8.62 | 4.57 | 2.96 | 4.61 | 3.58 | 5.41 |
| | Bern(2) | 5.09 | 4.07 | 6.20 | 5.83 | 8.85 | 4.34 | 3.05 | 4.53 | 3.62 | 5.55 |
| | Bern(5) | 7.27 | 5.02 | 9.09 | 6.59 | 11.42 | 4.43 | 3.08 | 4.61 | 3.51 | 5.42 |
| | Bern(8) | 7.42 | 4.85 | 8.85 | 6.55 | 11.18 | 4.33 | 3.06 | 4.67 | 3.39 | 5.21 |
| | IVQR | 4.53 | 2.53 | 5.70 | 3.05 | 7.49 | 3.78 | 2.53 | 4.08 | 3.05 | 4.58 |

Table 4: Objective function

| | Bernstein copula | | | | | | | | True copula |
|---|---|---|---|---|---|---|---|---|---|
| Order | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| Equation (16) | 4.433 | 2.766 | 1.328 | 0.540 | 0.202 | 0.078 | 0.039 | 0.022 | 1.578 |
| Equation (18) | 10.920 | 9.094 | 6.971 | 2.616 | 1.047 | 0.547 | 0.403 | 0.343 | 1.479 |

Table 5: Estimated copula

| | Bernstein copula | | | | | | | | True copula |
|---|---|---|---|---|---|---|---|---|---|
| Order | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| Mean ($C_1$) | 0.016 | 0.013 | 0.011 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0 |
| Sup ($C_1$) | 0.052 | 0.044 | 0.040 | 0.039 | 0.039 | 0.039 | 0.039 | 0.039 | 0 |
| Mean ($C_0$) | 0.038 | 0.033 | 0.030 | 0.028 | 0.027 | 0.027 | 0.027 | 0.027 | 0 |
| Sup ($C_0$) | 0.114 | 0.104 | 0.098 | 0.094 | 0.094 | 0.094 | 0.094 | 0.094 | 0 |

Table 6: Quantile regression coefficients

| | | $\beta_1$ | | | | | $\beta_0$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\tau$ | | | | | $\tau$ | | | | |
| | | 0.1 | 0.25 | 0.5 | 0.75 | 0.9 | 0.1 | 0.25 | 0.5 | 0.75 | 0.9 |
| Median | 0.78 | 0.87 | 0.85 | 0.70 | 0.52 | -0.94 | -1.24 | -1.57 | -1.80 | -1.82 | |
| Difference | Bern(3) | 0.56 | 0.64 | 0.63 | 0.52 | 0.39 | -0.82 | -1.10 | -1.40 | -1.61 | -1.62 |
| | Bern(4) | 0.44 | 0.50 | 0.50 | 0.42 | 0.30 | -0.73 | -0.99 | -1.26 | -1.43 | -1.43 |
| | Bern(5) | 0.38 | 0.44 | 0.44 | 0.35 | 0.23 | -0.68 | -0.93 | -1.18 | -1.31 | -1.28 |
| | Bern(6) | 0.37 | 0.44 | 0.42 | 0.34 | 0.21 | -0.66 | -0.92 | -1.16 | -1.30 | -1.25 |
| | Bern(7) | 0.35 | 0.43 | 0.40 | 0.34 | 0.22 | -0.66 | -0.92 | -1.15 | -1.30 | -1.30 |
| | Bern(8) | 0.36 | 0.42 | 0.40 | 0.32 | 0.22 | -0.66 | -0.92 | -1.16 | -1.31 | -1.30 |
| | Bern(9) | 0.37 | 0.43 | 0.40 | 0.32 | 0.22 | -0.68 | -0.93 | -1.15 | -1.31 | -1.31 |
| | True | -0.01 | -0.01 | 0.00 | 0.01 | 0.00 | 0.00 | -0.01 | -0.01 | 0.00 | 0.01 |
| Interquantile | Bern(2) | 1.37 | 1.03 | 0.93 | 0.97 | 1.12 | 0.90 | 0.76 | 0.77 | 0.85 | 1.17 |
| Range | Bern(3) | 1.45 | 1.12 | 1.02 | 1.04 | 1.17 | 0.91 | 0.81 | 0.82 | 0.97 | 1.24 |
| | Bern(4) | 1.59 | 1.25 | 1.11 | 1.13 | 1.26 | 1.01 | 0.84 | 0.88 | 0.98 | 1.38 |
| | Bern(5) | 1.70 | 1.29 | 1.15 | 1.17 | 1.30 | 1.07 | 0.94 | 0.92 | 1.09 | 7.47 |
| | Bern(6) | 1.74 | 1.38 | 1.25 | 1.18 | 1.31 | 1.11 | 0.96 | 0.94 | 1.09 | 7.64 |
| | Bern(7) | 1.73 | 1.37 | 1.23 | 1.18 | 1.36 | 1.14 | 0.96 | 0.95 | 1.12 | 7.73 |
| | Bern(8) | 1.83 | 1.37 | 1.24 | 1.19 | 1.40 | 1.11 | 0.98 | 0.96 | 1.10 | 7.79 |
| | Bern(9) | 1.76 | 1.37 | 1.21 | 1.17 | 1.38 | 1.11 | 0.98 | 0.97 | 1.13 | 7.75 |
| | True | 1.52 | 1.06 | 0.83 | 0.95 | 1.01 | 0.85 | 0.71 | 0.70 | 0.97 | 1.60 |

## C.2 Support of the Propensity Score

The final 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. 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 $[0.45, 0.55]$.

The performance of the estimator increases as the support of the propensity score increases: the distance between the estimated copula and the true one is smaller, and the RQR estimates are centered closer to their true values and their interquantile range is smaller, too. This shows that, even if the model is identified with small variation of the propensity score, the more variation there is, the better the performance of the estimator.

Table 7: Estimated copula

| Propensity | $[0.1, 0.9]$ | $[0.15, 0.85]$ | $[0.2, 0.8]$ | $[0.25, 0.75]$ | $[0.3, 0.7]$ | $[0.35, 0.65]$ | $[0.4, 0.6]$ | $[0.45, 0.55]$ |
|---|---|---|---|---|---|---|---|---|
| Mean ($C_1$) | 0.010 | 0.010 | 0.010 | 0.011 | 0.012 | 0.013 | 0.014 | 0.014 |
| Sup ($C_1$) | 0.021 | 0.022 | 0.023 | 0.024 | 0.026 | 0.029 | 0.031 | 0.031 |
| Mean ($C_0$) | 0.010 | 0.010 | 0.010 | 0.010 | 0.012 | 0.013 | 0.014 | 0.014 |
| Sup ($C_0$) | 0.021 | 0.021 | 0.022 | 0.023 | 0.026 | 0.028 | 0.031 | 0.031 |

Table 8: Quantile regression coefficients

| | | $\beta_1$ | | | | | $\beta_0$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\tau$ | | | | | $\tau$ | | | | |
| | | 0.1 | 0.25 | 0.5 | 0.75 | 0.9 | 0.1 | 0.25 | 0.5 | 0.75 | 0.9 |
| Median | $[0.1, 0.9]$ | 0.44 | 0.49 | 0.47 | 0.44 | 0.35 | -0.31 | -0.37 | -0.42 | -0.43 | -0.41 |
| Difference | $[0.15, 0.85]$ | 0.48 | 0.55 | 0.53 | 0.46 | 0.36 | -0.33 | -0.39 | -0.44 | -0.45 | -0.42 |
| | $[0.2, 0.8]$ | 0.55 | 0.59 | 0.56 | 0.50 | 0.40 | -0.37 | -0.43 | -0.49 | -0.49 | -0.48 |
| | $[0.25, 0.75]$ | 0.58 | 0.62 | 0.61 | 0.55 | 0.44 | -0.38 | -0.45 | -0.53 | -0.53 | -0.48 |
| | $[0.3, 0.7]$ | 0.68 | 0.75 | 0.71 | 0.63 | 0.49 | -0.46 | -0.54 | -0.62 | -0.62 | -0.58 |
| | $[0.35, 0.65]$ | 0.75 | 0.83 | 0.82 | 0.73 | 0.57 | -0.52 | -0.61 | -0.68 | -0.72 | -0.66 |
| | $[0.4, 0.6]$ | 0.80 | 0.89 | 0.88 | 0.76 | 0.64 | -0.55 | -0.68 | -0.76 | -0.77 | -0.71 |
| | $[0.45, 0.55]$ | 0.82 | 0.91 | 0.89 | 0.79 | 0.64 | -0.57 | -0.69 | -0.78 | -0.81 | -0.73 |
| Interquantile | $[0.1, 0.9]$ | 1.36 | 1.07 | 0.94 | 1.04 | 1.28 | 0.98 | 0.78 | 0.83 | 0.93 | 1.18 |
| Range | $[0.15, 0.85]$ | 1.33 | 1.14 | 1.06 | 1.06 | 1.37 | 1.07 | 0.84 | 0.81 | 0.91 | 1.09 |
| | $[0.2, 0.8]$ | 1.42 | 1.17 | 1.00 | 1.12 | 1.27 | 1.06 | 0.94 | 0.97 | 0.97 | 1.26 |
| | $[0.25, 0.75]$ | 1.47 | 1.21 | 1.08 | 1.12 | 1.30 | 1.11 | 0.92 | 0.92 | 0.97 | 1.13 |
| | $[0.3, 0.7]$ | 1.52 | 1.27 | 1.14 | 1.16 | 1.26 | 1.08 | 1.01 | 1.01 | 1.03 | 1.21 |
| | $[0.35, 0.65]$ | 1.43 | 1.07 | 1.03 | 1.08 | 1.29 | 1.09 | 0.86 | 0.97 | 1.07 | 1.17 |
| | $[0.4, 0.6]$ | 1.33 | 1.02 | 1.03 | 1.02 | 1.24 | 1.03 | 0.85 | 0.87 | 0.93 | 1.17 |
| | $[0.45, 0.55]$ | 1.32 | 1.07 | 0.97 | 1.03 | 1.25 | 1.05 | 0.79 | 0.84 | 0.92 | 1.17 |