BANCA D'ITALIA

EUROSISTEMA

Temi di discussione

(Working Papers)

Identification and estimation of triangular models with a binary treatment

by Santiago Pereda Fernández

March 2019

Number 1210

BANCA D'ITALIA
EUROSISTEMA

Temi di discussione

(Working Papers)

Identification and estimation of triangular models with a binary treatment

by Santiago Pereda Fernández

Number 1210 - March 2019

The papers published in the Temi di discussione *series describe preliminary results and are made available to the public to encourage discussion and elicit comments.*

The views expressed in the articles are those of the authors and do not involve the responsibility of the Bank.

ISSN 1594-7939 (print)
ISSN 2281-3950 (online)

Printed by the Printing and Publishing Division of the Bank of Italy

IDENTIFICATION AND ESTIMATION OF TRIANGULAR MODELS WITH A BINARY TREATMENT

by Santiago Pereda Fernández*

Abstract

I study the identification and estimation of a nonseparable triangular model with an endogenous binary treatment. Unlike other studies, I do not impose rank invariance or rank similarity on the unobservable of the outcome equation. Instead, I achieve identification 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.

JEL Classification: C31, C36.
Keywords: copula, endogeneity, policy analysis, quantile regression, unconditional distributional effects.

Contents

* Bank of Italy, Structural Economic Analysis Directorate.

1 Introduction*

One of the most relevant settings in empirical works is a triangular model with a binary treatment. Allowing for heterogeneous effects in this framework is important for several reasons. First, policies are often targeted at some specific subpopulations for which the effects could be different than for those already treated. For instance, they could vary with respect to individual characteristics such as gender or age, so understanding who benefits most from the treatment is crucial for effective policy design. Moreover, many effects of interest require knowledge of the whole distribution of potential outcomes, such as when one analyzes inequality. In those cases it is often important to allow the treatment to have a different impact on individuals with the same observable characteristics. The credibility of these counterfactual distributions of potential outcomes crucially depends on modeling heterogeneous effects appropriately (Heckman et al., 1997; Bisbee et al., 2017).

Hence, the workhorse linear model has been superseded by nonparametric, nonseparable models that display different kinds of heterogeneous effects. However, identification often requires assumptions that restrict the degree of heterogeneity. Notably, many results rely on either rank invariance or rank similarity. These two are assumptions on the relation between the treatment and the disturbance term in the outcome equation. Rank invariance implies that an individual's rank in the distribution of potential outcomes is the same under both treatment status. Rank similarity is weaker, as it allows this disturbance term to differ depending on the treatment status, but the correlation between the treatment and each rank is the same. As such, if those more likely to be treated rank relatively high when treated, they also rank relatively high when they are not. On the other hand, under rank dissimilarity the correlation between the treatment and each disturbance term is allowed to differ. In other words, the amount of self selection, or endogeneity, can vary for the treated

*Banca d'Italia, Via Nazionale 91, 00184 Roma, Italy. This paper has greatly benefited from discussions with Stéphane Bonhomme. I would also like to thank Manuel Arellano, Domenico Depalo, Bryan Graham, Christian Hansen, Giovanni Mellace, Magne Mogstad, Quang Vuong, Paolo Zacchia, and seminar participants at CEMFI, the 8th ICEEE Conference and the 5th Conference of the IAAE for their helpful comments and discussion. 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.

and the untreated.

Moreover, even when the identified model can display a large degree of heterogeneity in the effects, existing estimators often rely on substantially stronger assumptions to be tractable. On top of rank similarity, estimators often require additive separability of the unobservables or that the treatment variable be the same for people with different observed characteristics. The first one places restrictions on the Marginal Treatment Effect (MTE, Björklund and Moffitt, 1987) and the quantile curves: their shape does not vary with covariates, so changes in the covariates result in parallel shifts. The other could result in the effect being overestimated for some subpopulations and underestimated for others.

In this paper, I make the following contributions. First, I show that it is possible to fully identify a nonseparable model without imposing rank similarity if the instrument is continuously distributed and the distribution of the unobservables is sufficiently smooth. Second, I analyze how rank similarity affects the mean outcome difference between the treated and the untreated and the shape of the MTE, proposing a decomposition for each. Third, I propose an estimator based on Rotated Quantile Regression (RQR, Arellano and Bonhomme, 2017) that can capture a rich amount of heterogeneous effects.

I model self selection into treatment with copulas, one for each treatment group. Hence, the amount of self selection into treatment can differ for each of them. Furthermore, the unobservables are interpretable in terms of the conditional quantiles, or ranks, of the latent distribution of potential outcomes. This makes the distributional analysis intuitive, and it points to quantile regression methods for the estimation.

The identification of the model is based on a combination of continuous variation of the instrument and a smoothness assumption on the copula. In particular, I assume that the copula is real analytic, which allows to extrapolate the identification region from the support of the propensity score to the whole unit interval. Most parametric copulas are real analytic, including Bernstein copulas, a flexible family of copulas that can arbitrarily approximate any continuous copula when its order is high enough. This assumption nests several shape restrictions on the treatment effects that have already been considered in the literature, and

it is possible to do partial identification analysis when this assumption is dropped.

The amount of self selection has traditionally received little attention. However, both the distribution of observed outcomes and the MTE can be expressed in terms of the copula. It is possible to show that under rank dissimilarity the difference in the mean outcome between the treated and the untreated can be decomposed into three terms, which reflect differences in the covariates, differences in the returns to the treatment, and differences in the amount of self selection. The first two correspond to the endowments and coefficients components in the Oaxaca-Blinder decomposition. The third term reflects the fact that ability is not unidimensional, and some individuals perform relatively better when treated and relatively worse when not. Similarly, the MTE can be decomposed into a term that equals the MTE under rank similarity, and another that reflects differences in the amount of self selection.

Building on the identification result, I propose a multi-step estimator based on RQR, similar to the one proposed in Arellano and Bonhomme (2017) for sample selection models. It consists on the estimation of the propensity score, the copula of the unobservables, and the structural quantile process of the outcome equation. 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 known function of the copula, so 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).

The estimator 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. On the other hand, the estimation of the copula is computationally expensive, and the baseline estimator hinges on parametric assumptions.[1] These issues are explicitly addressed in extensions that relax the parametric assumptions and discuss their implementation.

[1]Parametric copulas have been used to model latent variables in a variety of setups: quantile selection models (Arellano and Bonhomme, 2017), bivariate probit models with dummy endogenous regressors (Han and Vytlacil, 2017), triangular models with continuous endogenous variables (Pereda Fernández, 2016), and both linear and non-linear panel data Prokhorov and Schmidt (2009); Pereda-Fernández (2017).

There is a vast literature on the identification and estimation of triangular models with a binary treatment.[2] Das (2005) showed the identification of an additively separable model. Subsequently, Instrumental Variables Quantile Regression (IVQR, Chernozhukov and Hansen, 2005, 2006) set an important milestone in the literature, defining a quantile treatment effect framework based on a nonseparable model. Other recent works include Feng et al. (2016) and Vuong and Xu (2017). In contrast with the approach in this paper, these papers required either rank invariance or similarity for identification.

A method that does not require rank similarity is Local Instrumental Variables (LIV, Heckman and Vytlacil, 1999, 2005). Carneiro and Lee (2009) extended earlier works by showing identification of the quantile treatment effects, although it required a large amount of variation in the instrument: it should take some values such that the individuals are always treated, and others such that they are never treated. In this paper I show that it is possible to identify the entire distributions of potential outcomes with a smaller amount of variation of the instrument.

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 (LATE, Imbens and Angrist, 1994), and later on the Local Quantile Treatment Effect (LQTE, Abadie et al., 2002) for the linear model. Frandsen et al. (2012) considered distributional treatment effects in a regression discontinuity design, and Frölich and Melly (2013) focused on the estimation of unconditional quantile treatment effects under minimal assumptions. More recently, a series of papers have considered the extrapolation of these 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. (2017). Some of these use the MTE for their extrapolation, imposing shape assumptions on the distribution of the unobservables that implicitly require the copula to be real analytic.

The estimation method presented in this paper is applied to the estimation of the effect

[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).

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 test the rank similarity assumption, finding strong evidence against it. I also estimate a positive effect for most of the distribution of future earnings for both types of placements. 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 some extensions are considered in Section 4. The methods presented in this paper are illustrated with an empirical application in Section 5. Finally, Section 6 concludes. All proofs are shown in Appendix A.

2 The Model

Consider the following triangular system of equations:[3]

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

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

[3]Throughout the paper I use upper case letters to denote random variables and lower case to denote their realizations.

where Y is the continuous outcome, D is the binary treatment, X is the vector of covariates, and Z is the vector with the variables used in the selection equation, that typically include an instrument (Z_1) and the covariates, *i.e.* $Z \equiv (Z_1, X')'$.[4] 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. The latter is uniformly distributed over the unit interval, and I refer to it as the rank of the SQF.[5] Equation 2 is the selection equation. Following Heckman and Vytlacil (2005), the treatment is determined by the propensity score, $\pi(Z)$, and a uniformly distributed unobserved random variable V. As shown by Vytlacil (2002), this is equivalent to the monotonicity condition in Imbens and Angrist (1994).

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.

Under endogeneity, the disturbance terms of the SQF and the selection equation are correlated. A convenient way to model it is using copulas.[6] The joint distribution of the unobservables is therefore given by $U_0, U_1, V|X \sim C_X(U_0, U_1, V|X)$.[7] However, it is possible to observe only one of the two treatment status for each individual, so the structural relation between U_0 and U_1 is not identified.[8]

Consequently, the focus lies on the bivariate copulas between U_d and V, conditional on

[4] An extension to multivalued treatment is considered in Section 4.4.

[5] This is known as the Skorohod representation.

[6] A copula is a multivariate cdf whose arguments are the ranks of the individual effects. Formally, given a vector of random variables $W_1, ..., W_N$, with marginal distributions $F_1(w_1)$,...,$F_N(w_N)$, 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 whose arguments are the ranks of the individual components.

[7] 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).

[8] 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).

X, which are denoted by $C_{d,X}(U_d, V)$, for $d = 0, 1$. 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 is more general, as it 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. In contrast, I allow for *rank dissimilarity* ($C_{0,X} \neq C_{1,X}$). Under this assumption, those who perform relatively well when they are treated are not necessarily those who perform relatively well when they are not.

The following example clarifies the different 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 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. Finally, under rank dissimilarity, those with a high propensity to have a college degree are on average more intelligent, but they do not necessarily have a high level of physical prowess. In fact, it is possible that those less likely to have a college degree have higher level of physical prowess and 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 extending the treatment.

2.1 Identification of the Structural Functions

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

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

These two distributions are conditional on the treatment status, so the focus lies on the conditional copula. Formally, $H_X(\tau, \pi(z)) \equiv \frac{C_{1,X}(\tau,\pi(z))}{\pi(z)} = \mathbb{P}(U_1 \leq \tau | D = 1, z)$ for the treated, and $G_X(\tau, \pi(z)) \equiv \frac{\tau - C_{0,X}(\tau,\pi(z))}{1-\pi(z)} = \mathbb{P}(U_0 \leq \tau | D = 0, z)$ for the untreated. The distribution of the outcome, conditional on being treated is therefore given by

$$F_{Y|D=1,Z}(y|z) = \int_0^1 \mathbf{1}(g_1(x, u_1) \leq y) \, dH_X(u_1, \pi(z)) \tag{3}$$

where $\mathbf{1}(\cdot)$ denotes the indicator function. Evaluating Equation 3 at $y = g_1(\tau, x)$ yields $F_{Y|D=1,Z}(g_1(x, \tau)|z) = H_X(\tau, \pi(z))$. Similarly, the distribution for the untreated equals

$$F_{Y|D=0,Z}(y|z) = \int_0^1 \mathbf{1}(g_0(x, u_0) \leq y) \, dG_X(u_0, \pi(z)) \tag{4}$$

and evaluating Equation 4 at $y = g_0(\tau, x)$ yields $F_{Y|D=0,Z}(g_0(x, \tau)|z) = G_X(\tau, \pi(z))$.

Therefore, $F_{Y|Z}$ depends on three components: the SQF of Y, the propensity score, 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, 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 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. 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 the system (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. Assumption 4 is a support assumption on the instrument, which is required 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 1. *Let $x \in \mathcal{X}$. Then, under assumptions 1 to 4:*

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

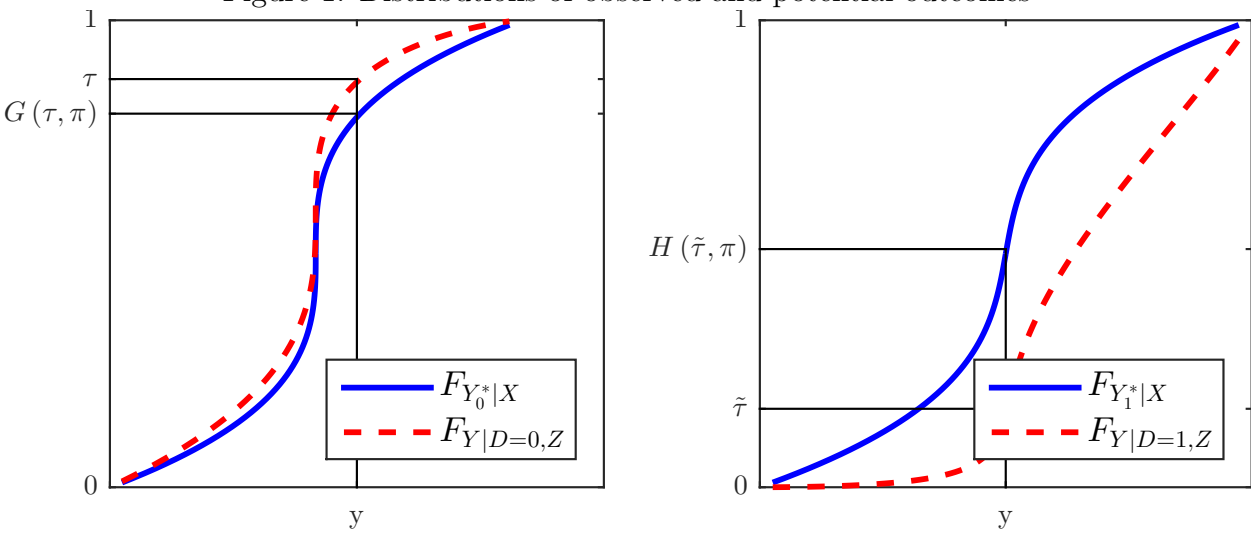
$$F_{Y|D=0,Z}\left(F^{-1}_{Y|D=0,Z}\left(\tau|z'\right)|z\right) = G_X\left(G_X^{-1}\left(\tau, \pi\left(z'\right)\right), \pi\left(z\right)\right) \forall \left(z, z'\right) \in \mathcal{Z}_x \times \mathcal{Z}_x \tag{6}$$

Moreover, for any H_X and G_X satisfying Equations 5 and 6, one can find distribution functions $F_{Y_1^*|X}(y|x)$ and $F_{Y_0^*|X}(y|x)$ such that $H_X\left(F_{Y_1^*|X}(y|x), \pi(z)\right) = F_{Y|D=1,Z}(y|z)$ and $G_X\left(F_{Y_0^*|X}(y|x), \pi(z)\right) = F_{Y|D=0,Z}(y|z)$ for all (z,y) in the support of (Z,Y) given $X = x$, where Y_1^* and Y_0^* are the potential outcomes of individuals when they are respectively untreated or treated.

Equations 5-6 require the instrument to come from a non-degenerate distribution for them to be informative about the copulas. Intuitively, variations in the propensity score due to the instrument do not affect the SQF. As a result, the change induced in the distribution of observed outcomes operates entirely through differences in self selection.

The second part of the lemma indicates the existence of potential outcomes that would be observed if the treatment was randomly allocated. These potential outcomes are related to the observed ones through a bijection determined by the copula. The two distributions coincide when the copula is independent, *i.e.* when the treatment is exogenous. Because the support of both distributions is the same, one can map the τ-th quantile of the distribution of potential outcomes to a specific quantile of the distribution of observed outcomes, as shown in Figure 1. This constitutes the basis for using RQR in the estimation.

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.

Equations 5-6 hold true for all values of τ. However, they are only well defined for values of the propensity score in $\mathcal{P}_x$. As a result, when the instrument displays so much variation that the support of the propensity score (conditional on $X = x$) equals the unit interval, it is immediate to show that the copula and the SQF are fully identified. This case is known as identification at infinity (Heckman and Vytlacil, 2007), but in the vast majority of datasets the variation of the instrument is smaller. A way to achieve identification is by making a

smoothness assumption on the copula:[9]

Assumption 5. *$\forall \tau \in (0,1)$, the functions $\pi \to C_{0,X}(\tau,\pi)$ and $\pi \to C_{1,X}(\tau,\pi)$ are real analytic on the unit interval.*[10]

Assumption 5 is a shape restriction, and it implies that both copulas, as well as all their derivatives are continuous. Most parametric copulas are based on analytic functions, such as polynomials, power functions or exponentials, and therefore real analytic.[11] A particularly relevant one is the Bernstein copula, which depends on Bernstein polynomials.[12] 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 it 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.[13] Hence, Assumption 5 can be seen as a parametric assumption for a large, flexible family of copulas.

Combining this assumption with Lemma 1 yields the main identification result:

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

Although Proposition 1 establishes the identification even if the support of the propensity score is small, the performance of an estimator depends largely on the its size. Because Assumption 5 extrapolates the identification from $\mathcal{P}_x$ to the unit interval, the larger the

[9] Other alternatives considered in the literature include parametric assumptions or using variation in the covariates.

[10] 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 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 $\mathcal{X}$.

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

[12] 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.

[13] Note that the Bernstein copulas are not the most appropriate to model extreme tail behavior. As shown in Sancetta and Satchell (2004), the Bernstein copula and its approximand converge to an arbitrary limit at a different speeds. In any case, it can capture increasing dependence as one moves to the tails.

support, the smaller the extrapolation.[14] This problem is akin to what happens to other methods that achieve identification using parametric or shape restrictions: the extrapolation is more plausible in a neighborhood of the observed support of the propensity score, but it becomes an increasingly stronger assumption as the distance increases.

Some copulas do not satisfy Assumption 5, such as the Fréchet-Hoeffding bounds, which correspond to perfect positive and negative correlation. They are not real analytic because they are piecewise defined, so their first derivative are not continuous and they do not have a well defined density. In such cases, it is still possible to study partial identification of the SQF, as shown in Section 4.3.[15] Alternatively, one could study which sets are identified using two popular approaches in the literature: LIV and IVQR. See Appendix D for further details.

2.2 Difference in Means Decomposition

Consider the mean outcome value for each treatment group. This quantity depends on three components: the distribution of the covariates, the copula, and the SQF. Define the following counterfactual mean outcomes when one or two of these components are exchanged with the other groups' counterparts:

$$\mathbb{E}\left[Y\left(j,H,k\right)\right] \equiv \int_{\mathcal{Z}} \int_0^1 g_j\left(x,u\right) dH_X\left(u,\pi\left(z\right)\right) dF_Z^{(k)}\left(z\right)$$
$$\mathbb{E}\left[Y\left(j,G,k\right)\right] \equiv \int_{\mathcal{Z}} \int_0^1 g_j\left(x,u\right) dG_X\left(u,\pi\left(z\right)\right) dF_Z^{(k)}\left(z\right)$$

where $j = 0, 1$ refers to the treatment group of the SQF and $F_Z^{(k)}$ is the distribution of the observables of treatment group $k = 0, 1$. Note that the mean outcome for those that are treated and untreated are respectively given by $\mathbb{E}\left[Y\left(1,H,1\right)\right]$ and $\mathbb{E}\left[Y\left(0,G,0\right)\right]$. The difference between these two can be attributed to differences in each of the three individual

[14] The simulations in Appendix F support this claim.

[15] Partial identification based of the triangular model under weak assumptions has also been studied by Chesher (2005) and Jun et al. (2011).

components. Hence, the following decomposition follows

$$\mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0] = \mathbb{E}[Y(1,H,1)] - \mathbb{E}[Y(1,H,0)] \tag{7}$$

$$+ \mathbb{E}[Y(1,H,0)] - \mathbb{E}[Y(1,G,0)] \tag{8}$$

$$+ \mathbb{E}[Y(1,G,0)] - \mathbb{E}[Y(0,G,0)] \tag{9}$$

Hence, Equations 7-9 respectively reflect differences in the distribution of the observables, differences in the amount of self selection, and differences in the distribution of potential outcomes. The first and third terms are the equivalent to the endowments and coefficients effects in the Oaxaca-Blinder decomposition. The endowments effect reflects the fact that individuals with some characteristics are more likely to be treated. On the other hand, the coefficients effect captures differences in the distribution of potential outcomes for the whole population, which in a linear framework correspond to the slope coefficients.

This analysis extends the Oaxaca-Blinder decomposition by adding the excess selection term, represented by differences in the copula between the treated and the untreated. When the copula is independent, *i.e.* under exogeneity, the average value of the rank equals 0.5 for both treatment groups, and Equation 8 equals 0. If there is (positive) sample selection 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 Equation 8 is also equal to 0. In contrast, with rank dissimilarity, the average rank is different for the treated and the untreated, and this translates into a non-zero excess selection effect.

2.3 Marginal Treatment Effect

A similar result holds true for the MTE which. For comparability with other studies, we focus on the conditional MTE, which can be expressed as

$$\Delta^{MTE}(x,v) = \int_0^1 g_1(x,u)\, dC_{1,X}(u|v) - \int_0^1 g_0(x,u)\, dC_{0,X}(u|v) \tag{10}$$

Under rank dissimilarity, the following decomposition holds:[16]

$$\begin{aligned}\Delta^{MTE}(x,v) &= \int_0^1 \left[g_1(x,u) - g_0(x,u)\right] dC_{1,X}(u|v) + \int_0^1 g_0(x,u_0)\, d\left[C_{1,X}(u|v) - C_{0,X}(u|v)\right] \\ &\equiv \Delta^{RIMTE}(x,v) + \Delta^{ESME}(x,v) \end{aligned} \tag{11}$$

The first term is denoted as the rank invariant marginal treatment effect (RIMTE), *i.e.* the expected gain for the marginal individual, when his unobservables are equally correlated with each treatment status. This effect depends on the difference between the two SQF weighted by the copula. On the other hand, the second term is denoted as the excess selection marginal effect (ESME), and it reflects rank dissimilarity. This effect captures the difference in the amount of selection between the two treatment status.

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. A property of the ESME is that the average across all values of v is equal to zero, *i.e.* some individuals are positively selected, and some are negatively selected. This is shown in the following lemma:

Lemma 2. *Let the copula $C_{D,X}(u,v)$ have a well-defined density for $D = 0, 1$. Then, $\int_0^1 \Delta^{ESME}(x,v)\, dv = 0$ and $\int_0^1 \Delta^{MTE}(x,v)\, dv = \int_0^1 \Delta^{RIMTE}(x,v)\, dv$.*

The MTE is the building block for several treatment effects of interest, such as the average treatment effect (ATE), the average treatment effect on the treated (TT), or on the untreated (TUT). Hence, they can also be expressed in terms of the SQF and the copula, as shown in Appendix E. Finally, note that the quantile treatment effect (QTE) only requires knowledge of the SQF. Thus, for a given quantile τ, it can be expressed as the difference between the SQF of each treatment group but not on the copula.

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

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. $g_D(x, \tau) = x'\beta_D(\tau)$ *for* $D = 0, 1$, *where* β_D *is continuous and such that* $g_D(x, \tau)$ *is increasing in its last argument.*

Assumption 8. *Let* $\beta(\tau) \equiv \left(\beta_1(\tau)', \beta_0(\tau)'\right)'$ *and* $\theta \equiv (\theta_1', \theta_0')'$. *For all* τ, $\left(\beta(\tau)', \theta', \gamma'\right)' \in$ *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* 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 12. *Let* $C_{D|X}(u, v) \equiv C_{D|X}(u, v; \theta_D)$, *with* $\dim(\theta_D) < \infty$ *for* $D = 0, 1$. $C_{D|X}(u, v; \theta_D)$ *is uniformly continuous and differentiable with respect to its arguments a.e.. Its density,* $c_{D|X}(u, v; \theta_D)$, *is well-defined and finite.*

Assumption 6 describes the sampling process of the data. The linear quantile model imposed by Assumption 7 is standard in the literature and convenient from a computational point of view.[17] Note however, that it would be possible to relax this assumption, allowing for nonlinear quantile functions as long as the resulting SQF is continuous and increasing in τ. Assumption 8 is a regularity condition. Assumption 9 restricts the analysis to dependent

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

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.

Assumption 11 is made for simplicity, and it is satisfied by several estimation methods, including maximum likelihood. This assumption could be relaxed to allow the propensity score to be nonparametrically estimated. This extension is considered in Section 4.1.

Similarly, Assumption 12 is imposed for convenience, and it is satisfied by most common choices of copulas, including the Gaussian or the Clayton.[18] Alternatively, the dependence on a finite number of parameters could be relaxed. Proposition 1 is satisfied when the copula is real analytic, making the Bernstein copula the natural choice to consider. This extension is explored in Section 4.2.

The estimation is done in three steps. The first one 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 the copula parameters: θ_1 and θ_0. Given some $t \in \Theta$, define $\hat{\beta}_1(\tau; t)$ as

$$\hat{\beta}_1(\tau; t) \equiv \arg\min_{b \in \mathcal{B}} \sum_{i=1}^{N} d_i \rho_{\hat{H}_{X,i,\tau}}(y_i - x_i' b) \tag{12}$$

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

$$\hat{\theta}_1 \equiv \arg\min_{t \in \Theta} \left\| \sum_{i=1}^{N} \int_0^1 d_i \varphi(\tau, z_i) \left[\mathbf{1}\left(y_i \leq x_i' \hat{\beta}_1(\tau; t)\right) - H_X(\tau, \hat{\pi}(z_i); t)\right] d\tau \right\| \tag{13}$$

where $\varphi(\tau, z_i)$ is an instrument function.[19] Similarly, $\beta_0(\tau; t)$ and θ_0 are estimated by

$$\hat{\beta}_0(\tau; t) \equiv \arg\min_{b \in \mathcal{B}} \sum_{i=1}^{N} (1 - d_i) \rho_{\hat{G}_{X,i,\tau}}(y_i - x_i' b) \tag{14}$$

[18] It is possible to allow the copula to depend on the covariates. For example, if the covariates are discrete, one could specify a different copula for each value of the covariates. Alternatively, one could use a more parsimonious approach with continuous covariates. E.g., one could let $\rho(x) = \exp\left(x'\gamma\right) - 1/\exp\left(x'\gamma\right) + 1$ for the Gaussian copula. Variations of this can be used to accommodate copulas whose correlation parameter has a different support.

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

$$\hat{\theta}_0 \equiv \arg\min_{t\in\Theta} \left\| \sum_{i=1}^{N} \int_0^1 (1-d_i)\,\varphi(\tau, z_i) \left[\mathbf{1}\left(y_i \leq x_i'\hat{\beta}_0(\tau;t)\right) - G_X\left(\tau, \hat{\pi}(z_i);t\right)\right] d\tau \right\| \tag{15}$$

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

Such an estimator is asymptotically Gaussian and converges at the parametric rate, as shown by the following theorem:

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* $\hat{\beta}_d(\tau)$ *and* $\hat{\theta}_d$ *for* $d = 0, 1$ *be the estimators defined above. Under Assumptions 1-12, their joint asymptotic distribution is given by* $\sqrt{n}\left(\hat{\vartheta}(\cdot) - \vartheta(\cdot)\right) \Rightarrow S(\cdot)$, *where* $S(\cdot)$ *is a zero-mean Gaussian process with covariance function* $\Sigma_S(\tau, \tau')$, *which is defined in the proof.*

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. *Even though the SQF is assumed to be linear in quantiles, the resulting MTE is not linear in general. 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.*

Remark 3. *From an implementation standpoint, equations 12 and 14 can be solved with standard quantile regression techniques by rotating the loss function. On the other hand, equations 13 and 15 involve non-convex optimization. When the number of parameters is small, an appealing method is grid search.*

Estimation of the MTE is straightforward using the sample analog and substituting the propensity score, the copula parameters, and the SQF by those presented above:

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

where $\hat{C}_{d,x_i,\tau,v} \equiv C_{d,x_i}\left(\tau, v; \hat{\theta}_d\right)$ for $d = 0, 1$.

4 Extensions

4.1 Nonparametric First Stage

Under Assumption 11, the propensity score depends on a finite number of parameters, ruling out nonparametric estimators of the propensity score, such as Klein and Spady (1993) or the Nadaraya-Watson estimator. In principle, one could plug-in such estimators into the estimating Equations 12 and 14. Under certain conditions on the propensity score estimator, it is possible to show that the slope parameters admit an asymptotically linear representation (Newey, 1994; Ichimura and Newey, 2017). Recently, Chernozhukov et al. (2016) have proposed the construction of locally robust moment functions. Using these moments, a nonparametric estimator of the propensity score would not affect the influence function of the slope parameters, which would allow them to retain the asymptotic normality at the $\sqrt{n}$ convergence rate. See Chernozhukov et al. (2016) for further details.

4.2 Bernstein Copula

A way to relax assumption 12 would be to use Bernstein copulas. The cumulative distribution of this copula is given by

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

where M is the order of the copula, and $\overline{P}_{m,M}(u) = \binom{M}{m} u^m (1-u)^{M-m}$. The density of this copula has a similar form, making it is very convenient to implement.[20] 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 $j = 0, 1$. The RQR estimator is the same as the one presented in Section 3, substituting θ_j by A_j.[21]

[20] 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 $c(u,v) = \sum_{m_u=0}^{M-1} \sum_{m_v=0}^{M-1} \eta\left(\frac{m_u}{M}, \frac{m_v}{M}\right) \overline{P}_{m_u,M}(u)\, \overline{P}_{m_v,M}(v)\, M^2$.

[21] 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 implementation of the estimator is more complicated than in the parametric case: the number of parameters equals $(M-1)^2$, 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 express any Bernstein copula of order M_1 as a Bernstein copula of order $M_2 > M_1$.[22] 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^*.[23]

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 programme 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

[22] 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, $A_2 = \overline{P}_2 C_1^{-1} \overline{P}_2'$.

[23] 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.

unobservables is high in absolute value, starting with a copula of a relatively high order may be advisable.[24] Finally, one should bear in mind that the random search algorithm does not guarantee that the estimator is the minimizer of the objective function.[25]

4.3 Fréchet-Hoeffding Bounds

Whenever the researcher is not willing to impose Assumption 5, it is possible to use the Fréchet-Hoeffding bounds on a bivariate copula, $\max\{u+v-1,0\} \leq C(u,v) \leq \min\{u,v\}$, to attain partial identification on the SQF defined in Equation 1. To do so, first notice that the H_X and G_X functions are bounded:

$$\max\left\{\frac{\tau+\pi(z)-1}{\pi(z)},0\right\} \leq H_X(\tau,\pi(z)) \leq \min\left\{\frac{\tau}{\pi(z)},1\right\} \tag{16}$$

$$\max\left\{\frac{\tau-\pi(z)}{1-\pi(z)},0\right\} \leq G_X(\tau,\pi(z)) \leq \min\left\{\frac{\tau}{1-\pi(z)},1\right\} \tag{17}$$

Combining Equations 16 and 17 with Equations 3 and 4, respectively, yields

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

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

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. Moreover, notice that when $\pi(z)=1$, the upper and lower bound of $g_1(x,\tau)$ coincide, and similarly for $g_0(x,\tau)$ when $\pi(z)=0$. This is the identification at infinity case, which indicates that

[24] 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.

[25] 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.

the potential bias of assuming real analyticity is smaller as the variation induced by the instrument increases.

4.4 Multivalued Treatment

Suppose that the treatment variable D can take on values $\{0, 1, ..., J\}$. The model is now given by:

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

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 G_X and $H_{j,X}$, $j = 1, ..., J$, are given by

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

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

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_X and $H_{j,X}$.

5 Empirical Application

The estimation method presented in Section 3 is applied to the estimation of the effect of Work First Job Placements on the distribution of future earnings. This is a welfare-to-work

program 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 found no job 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 categories amounted to 11583, 2762, and 16177, respectively.

Since both treatments were endogenously determined, estimation of the effects of interest requires the exogenous variation coming from an instrument. 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. The authors 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. For a detailed description on the dataset and how the instrument is constructed, see Autor and Houseman (2010).

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, either type of placement had a small and not significant effect on the lower tail of the earnings distribution. On the other hand, the effect on the upper tail was substantially large and positive for DHP, while it was negative and significant for THP. However, Autor et al. (2017) stated that they could not test the rank similarity assumption. Moreover, they highlighted the difficulty of translating the estimates of the effects of the job placements on the conditional distribution of earnings into the unconditional distribution of earnings.

Note that Autor et al. (2017) focused on the estimation of the distribution of conditional

(on X) effects, not on the effect on the unconditional distribution. This is important, as the latter may be more relevant from a policy perspective, and consequently I focus the attention on it. Hence, rather than reporting the coefficients on the treatment status from the IVQR estimates, I report the estimates of the unconditional quantile function for the whole population under each treatment status following Chernozhukov et al. (2013).[26] An additional difference with respect to Autor et al. (2017) is the number of parameters of the model: I include all the interactions between the treatment status and the covariates. In other words, I compute the RQR estimator for each group separately.

I compute the RQR for each group using different copulas. In particular, I consider the Gaussian copula and the Bernstein copula of orders 2 through 6. Among the latter, I present the estimates from the model selected using 5-fold cross validation. The propensity score is estimated with ordered multinomial logit. Finally, to assess the sensibility of the results to the rank similarity assumption I consider an additional specification with a Gaussian copula constrained to be the same for all three groups.

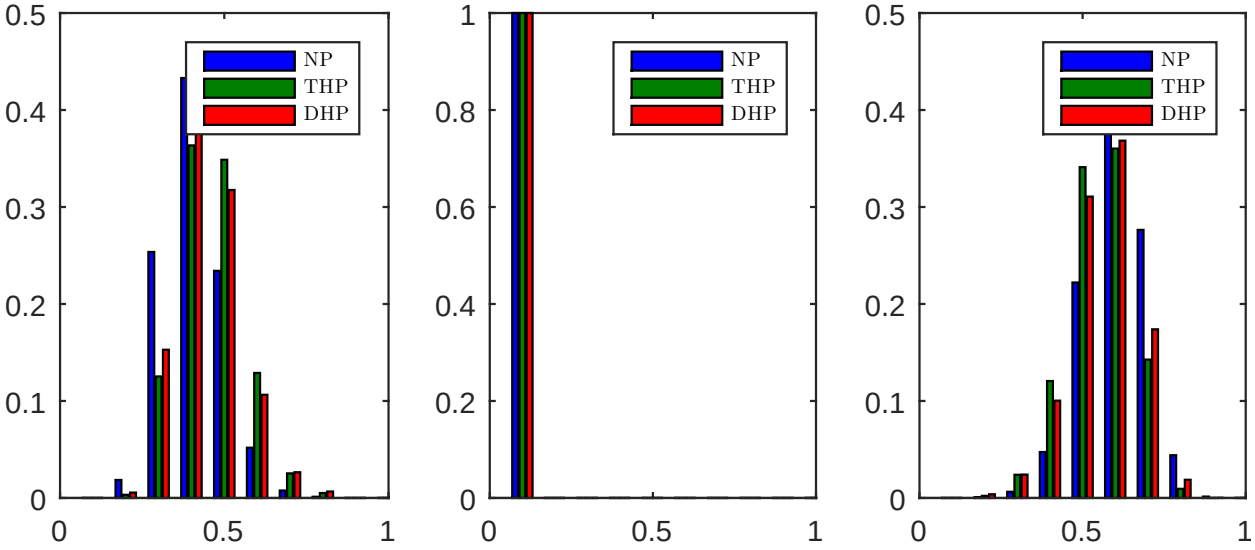
Figure 2 reports the distribution of the propensity score to be in each group for the people in each treatment group. The three histograms reveal a substantial overlap for the people in the three groups. However, the central histogram indicates that there is a little amount of variability for the propensity 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 3 compares the baseline estimates of the quantile function of future earnings with the observed empirical distribution.[27] There are two relevant findings in this figure. First, the 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 results suggest that the rank similarity assumption is unlikely to hold. More formally, I test the null hypothesis

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

[27] For the Bernstein copula, the selected orders were 2, 4, and 2 for the DHP, THP, and NP groups, respectively.

Figure 2: Histograms of the Estimated Propensity Scores



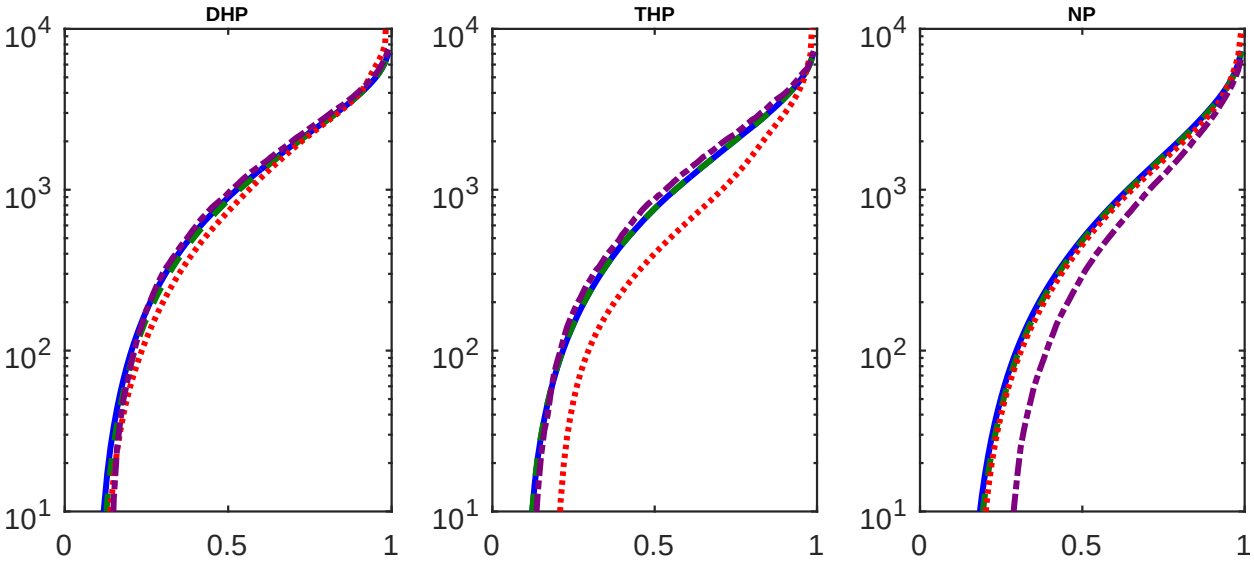
Notes: histograms with the distribution of the propensity to be in the DHP group (left panel), in the THP group (central panel), and in the NP group (right panel), split by the actual treatment received.

of rank similarity with the test proposed in Frandsen and Lefgren (2018) using the IVQR estimates. The test statistic equals 213.3, whereas the critical value for a test size of 5% is 67.5.[28] Hence, this test strongly rejects the hypothesis of rank similarity.

The shape of the estimated copulas is shown in Figure 4. I report the Kendall's τ correlation coefficient in Table 1 to give a comparable measure of correlation intensity. 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. Note that the negative correlation between U_0 and V implies that those individuals less likely to be treated (*i.e.* those with high values of V) would rank relatively low in the distribution of potential outcomes when $D = 0$ (*i.e.* low values of U_0). The results for the Bernstein copula are similar to those found for the Gaussian copula,

[28]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 estimates using IQR, which are available upon request, were similar with the exception of the tails, where they displayed an erratic behavior. The test statistic using the IQR implementation equals 218.8.

Figure 3: Estimated Potential Quantile Functions



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

although there is a slight increase in the amount of correlation for the two treatment groups as the order increases.

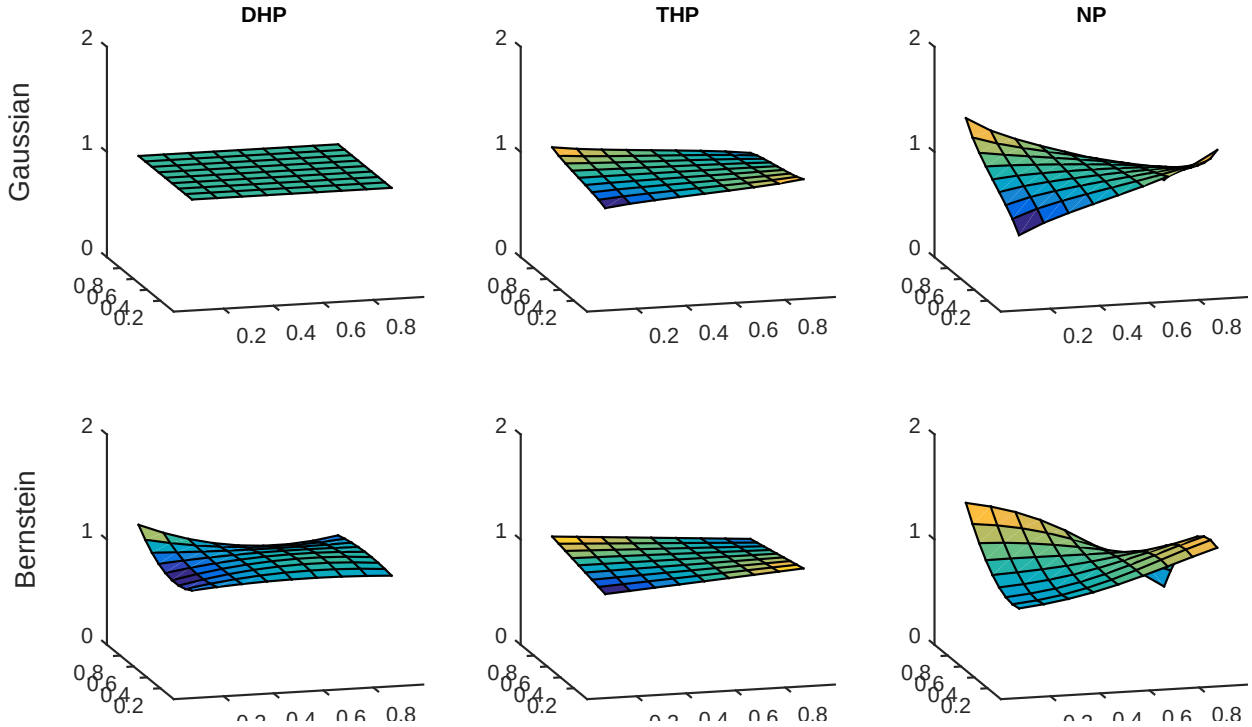
Table 1: Kendall's τ Statistic of the Estimated Copulas

Copula	Gau	Con	Ber(2)	Ber(3)	Ber(4)	Ber(5)	Ber(6)
DHP	0.00	0.06	-0.01	0.02	0.06	0.06	0.07
THP	0.04	0.06	0.02	0.02	0.04	0.05	0.05
NP	0.13	0.06	0.16	0.14	0.15	0.14	0.14

Notes: DHP, THP, NP, Gau, Con, and Ber(X) respectively stand for direct-hire placement, temporary-help placement, no placement, Gaussian copula, Gaussian copula constrained to be the same for all three groups, and Bernstein copula of order X.

The estimates of the unconditional QTE based on RQR (Table 2) 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 of the distribution, for which the effect is negligible and not significantly different from zero. Moreover, both QTE have an increasing profile for most of the distribution, peaking around the 80th percentile and rapidly decreasing thereafter, as shown in Figure 5. The largest gain comes from the

Figure 4: 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; DHP, THP, and NP respectively stand for direct-hire placement, temporary-help placement, and no placement.

DHP, whereas the gain from THP is substantially smaller for all quantiles, specially at the top of the distribution. On average, these gains are estimated to equal \$340 and \$225 with the Gaussian copula, and \$340 and \$245 with the Bernstein copula.[29]

Table 2: Quantile Treatment Effect Estimates

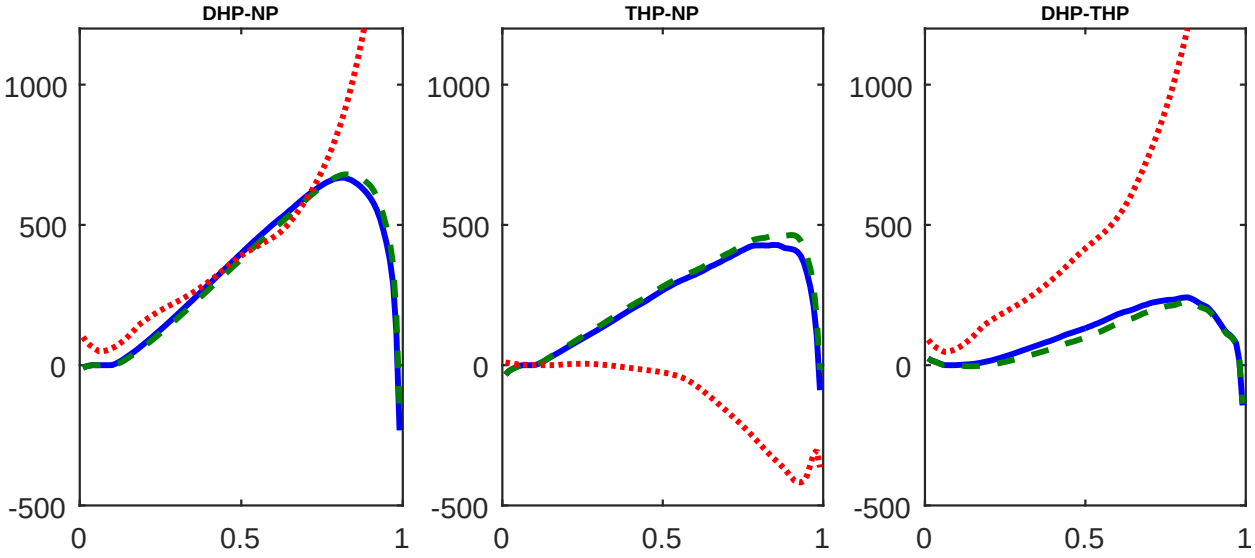
	u					
	0.15	0.25	0.5	0.75	0.85	Mean
$RQR(DHP, NP; Gau)$	33.0	126.3	398.9	641.9	653.2	340.2
	(15.0)	(42.4)	(124.4)	(225.7)	(274.4)	(138.4)
$RQR(THP, NP; Gau)$	28.4	94.7	267.4	410.3	428.5	227.4
	(7.3)	(21.2)	(79.7)	(168.5)	(224.0)	(102.9)
$RQR(DHP, THP; Gau)$	4.6	31.6	131.5	231.6	224.6	112.8
	(16.5)	(41.4)	(107.0)	(169.0)	(185.1)	(101.1)
$RQR(DHP, NP; Ber)$	26.0	114.4	378.6	637.5	678.3	338.8
	(15.7)	(43.6)	(118.9)	(199.4)	(224.2)	(116.2)
$RQR(THP, NP; Ber)$	29.6	102.7	280.6	428.4	465.7	245.1
	(7.4)	(17.9)	(55.6)	(109.6)	(141.0)	(62.7)
$RQR(DHP, THP; Ber)$	-3.6	11.8	98.0	209.1	212.6	93.7
	(17.5)	(45.4)	(119.6)	(192.6)	(208.9)	(111.4)
$RQR(DHP, NP; Con)$	20.7	110.1	398.7	702.8	775.8	377.3
	(11.8)	(41.0)	(126.6)	(223.3)	(260.1)	(381.9)
$RQR(THP, NP; Con)$	28.5	114.4	360.9	620.5	711.7	356.0
	(6.6)	(18.2)	(58.5)	(112.0)	(139.6)	(358.9)
$RQR(DHP, THP; Con)$	-7.8	-4.3	37.8	82.3	64.2	21.2
	(11.0)	(31.1)	(84.5)	(142.2)	(161.9)	(23.0)
$IVQR(DHP, NP)$	102.7	195.2	390.4	691.7	1038.1	483.6
	(38.6)	(63.9)	(139.0)	(272.3)	(411.4)	(177.3)
$IVQR(THP, NP)$	0.1	5.1	-25.7	-217.7	-337.0	-115.0
	(45.0)	(71.4)	(153.9)	(223.3)	(260.6)	(148.1)
$IVQR(DHP, THP)$	102.6	190.1	416.1	909.4	1375.1	598.6
	(75.0)	(117.9)	(250.6)	(402.1)	(534.0)	(273.6)

Notes: DHP, THP, NP, Gau, Con, and Ber respectively stand for direct-hire placement, temporary-help placement, no placement, Gaussian copula, Gaussian copula constrained to be the same for all three groups, and the selected Bernstein copula; u denotes the quantile; mean denotes the average across all quantiles in the estimation grid; boostrapped standard errors in parenthesis.

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.

[29]The orders of the Bernstein copulas selected through cross validation were 3, 2 and 5 for the DHP, THP and NP groups, respectively.

Figure 5: Quantile Treatment Effect Estimates



Notes: in each panel, the solid blue line represents the QTE of the RQR estimator with the Gaussian copula, the dashed green line represents the QTE of the RQR estimator with the Bernstein copula, and the dashed red line represents the QTE of the IVQR estimator; DHP, THP, and NP respectively stand for direct-hire placement, temporary-help placement, and no placement.

Indeed, the estimates based on IVQR 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. The estimates for THP are positive at almost every quantile, although not as large as those of the other treatment group. In contrast, the estimates based on IVQR are negative for the majority of the quantiles. Consequently, the difference between DHP and THP shows a gain for DHP with both estimators, but the magnitude is remarkably different: for the estimator based on RQR, the difference is on average between \$110 and \$90, whereas for the estimator based on IVQR this difference is about \$600 on average.

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. This is confirmed by the estimated self-selection effect of the means decomposition (Table 3): 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 group with the control group. 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. Note that because the sample is very homogeneous with respect to the covariates, the endowments effect is negligible in all cases.

Table 3: Means decomposition

	Gaussian copula			Bernstein copula		
	DHP, NP	DHP, NP	DHP, THP	DHP, NP	DHP, NP	DHP, THP
Total	486.1	383.6	102.5	491.4	388.0	103.4
effect	(25.1)	(35.1)	(36.3)	(20.6)	(33.2)	(36.1)
Endowments	0.0	-7.9	0.0	1.6	-7.1	-0.5
effect	(6.3)	(16.3)	(1.7)	(7.9)	(19.9)	(2.2)
Self-selection	186.6	188.2	-6.8	192.9	175.7	13.2
effect	(124.5)	(87.6)	(82.4)	(107.0)	(57.5)	(95.3)
Coefficients	299.5	203.3	109.3	297.0	219.4	90.7
effect	(139.1)	(102.2)	(100.9)	(112.3)	(61.2)	(111.4)

Notes: DHP, THP, and respectively stand for direct-hire placement, temporary-help placement, and no placement; boostrapped standard errors in parenthesis.

Similarly, the MTE also reflects a large amount of heterogeneity in the effects, as shown in Tables 4-5 and Figure 6. If one considers the RIMTE, it has the usual decreasing shape (in v), although the slope is almost flat. In contrast, the ESME has a positive and steep slope. This implies that those individuals more likely to be treated (*i.e.* those with a small value of v) are those with the expected lowest gain from the treatment. The reason behind this is the difference in the amount of selection, reflected on the copulas. Because the correlation is more negative for the NP group, it means that individuals with a small value of v would tend to rank higher in the distribution of potential outcomes of the NP (*i.e.* $u_0 > u_1$). Even though for a given quantile there is a gain from being treated, the excess selection reduces the expected gain.

Moreover, the estimates of the MTE with the Gaussian copula are negative for roughly $V \leq 0.1$. Hence, even if the distributions of potential outcomes for each of the two treatments dominate the distribution of no placement, there would be a minority of workers who would have less future earnings, had all of them received one of the two treatments. 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 the selection equation to be represented as in equation 2, 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.

Table 4: 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
	(231.5)	(183.8)	(131.6)	(153.1)	(185.2)	(138.8)
$RQR(THP, NP; Gau)$	42.4	112.1	233.4	343.0	396.7	221.6
	(263.7)	(196.1)	(94.4)	(101.8)	(146.3)	(103.6)
$RQR(DHP, THP; Gau)$	47.8	69.9	110.6	150.7	171.9	110.0
	(121.0)	(84.3)	(102.4)	(186.8)	(236.5)	(101.3)
$RQR(DHP, NP; Ber)$	145.6	159.2	269.6	469.9	571.0	328.9
	(133.8)	(119.3)	(126.9)	(183.8)	(229.8)	(116.2)
$RQR(THP, NP; Ber)$	86.5	73.6	148.8	358.5	475.9	237.3
	(181.7)	(143.2)	(65.4)	(122.4)	(163.0)	(63.1)
$RQR(DHP, THP; Ber)$	59.1	85.7	120.8	111.4	95.1	91.7
	(124.5)	(91.4)	(122.2)	(224.9)	(282.9)	(111.5)
$RQR(DHP, NP; Con)$	386.7	380.8	369.7	358.7	352.9	369.8
	(126.6)	(129.8)	(136.1)	(142.5)	(146.0)	(374.7)
$RQR(THP, NP; Con)$	366.0	360.9	351.5	342.1	337.0	351.5
	(60.3)	(62.4)	(66.8)	(71.7)	(74.5)	(353.5)
$RQR(DHP, THP; Con)$	20.7	19.8	18.2	16.6	15.8	18.3
	(83.9)	(84.0)	(84.2)	(84.3)	(84.3)	(21.3)

Notes: DHP, THP, NP, Gau, Con, and Ber respectively stand for direct-hire placement, temporary-help placement, no placement, 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; boostrapped standard errors in parenthesis.

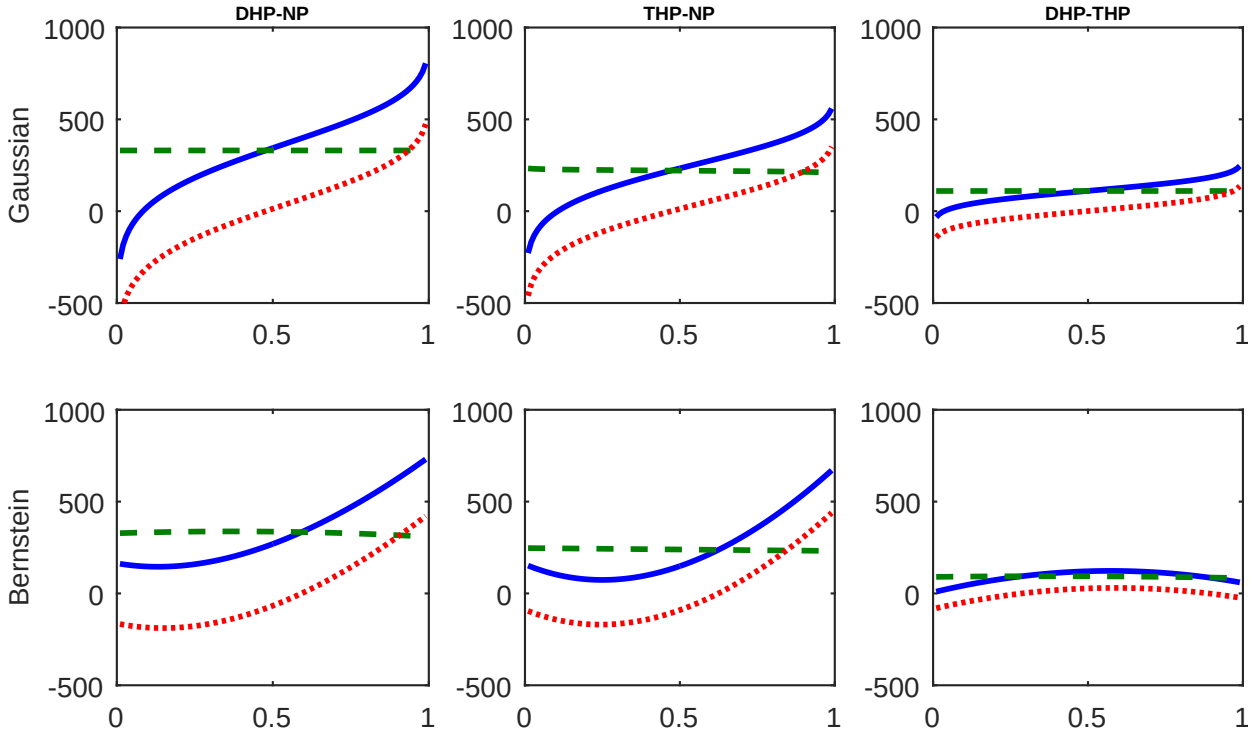
These results do not support the rank similarity hypothesis. Therefore, it is pertinent to study the estimates for the constrained model and compare them with those that are unconstrained. First, the amount of correlation lies in between those of the DHP and THP groups on the one hand, and that of the 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 (Table 2). Consequently, the QTE for the DHP relative to the THP group becomes a small fraction of the unrestricted estimate. These results do not coincide with the IVQR estimates, in fact they go on the opposite direction. This suggests

Table 5: 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
		(130.0)	(133.0)	(139.3)	(146.1)	(150.0)	(139.3)
	$RQR(DHP, NP; Gau)$	226.2	224.3	220.8	217.2	215.3	220.7
		(110.0)	(107.6)	(104.1)	(101.8)	(101.2)	(104.3)
	$RQR(THP, NP; Gau)$	109.9	109.9	109.9	109.9	109.9	109.9
		(95.2)	(97.3)	(101.3)	(105.3)	(107.5)	(101.2)
	$RQR(DHP, NP; Ber)$	334.2	336.7	336.5	327.1	320.8	330.5
		(98.1)	(103.5)	(117.3)	(130.5)	(136.6)	(139.3)
	$RQR(DHP, NP; Ber)$	244.7	243.1	239.2	235.3	233.7	239.2
		(64.6)	(63.9)	(63.4)	(64.8)	(65.9)	(104.3)
	$RQR(THP, NP; Ber)$	92.4	93.3	93.3	90.0	87.7	91.2
		(101.2)	(104.9)	(113.1)	(119.3)	(121.6)	(101.2)
ESME	$RQR(THP, NP; Gau)$	-240.4	-148.6	13.4	163.0	238.0	1.0
		(148.7)	(91.4)	(11.3)	(101.3)	(146.7)	(0.7)
	$RQR(DHP, THP; Gau)$	-183.8	-112.2	12.7	125.8	181.5	0.9
		(173.4)	(106.4)	(12.7)	(118.3)	(171.1)	(0.8)
	$RQR(DHP, THP; Gau)$	-62.1	-40.0	0.7	40.8	62.0	0.1
		(154.7)	(98.7)	(5.1)	(102.5)	(154.0)	(0.4)
	$RQR(THP, NP; Ber)$	-188.5	-177.4	-66.8	142.8	250.3	-1.5
		(119.9)	(95.8)	(66.8)	(92.6)	(128.2)	(1.3)
	$RQR(DHP, THP; Ber)$	-158.2	-169.6	-90.4	123.2	242.2	-2.0
		(146.5)	(109.9)	(42.0)	(110.4)	(144.4)	(0.8)
	$RQR(DHP, THP; Ber)$	-33.3	-7.7	27.5	21.4	7.4	0.5
		(174.0)	(129.5)	(58.3)	(124.6)	(183.5)	(1.1)

Notes: DHP, THP, NP, Gau, Con, and Ber respectively stand for direct-hire placement, temporary-help placement, no placement, 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; boostrapped standard errors in parenthesis.

Figure 6: Marginal Treatment Effect Estimates



Notes: in each panel, the solid blue line represents the MTE, the dashed green line represents the RIMTE, and the dashed red line represents the ESME. The first row shows the estimates with the Gaussian copula; the second row shows the estimates with the selected Bernstein copula; DHP, THP, and NP respectively stand for direct-hire placement, temporary-help placement, and no placement.

that the difference between the RQR and IVQR estimates is due to both the rank similarity assumption and the interaction effect between the treatments and the covariates.

In terms of the MTE (Table 4), the results are markedly different. In particular, the MTE of the constrained estimator has a slightly downward slope for both treatment status, and they are roughly the same size. Moreover, they are positive for the entire distribution, hiding the negative mean effects for those with a high enough value of V.

6 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 distribution of the unobservables is modeled with copulas, 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. This is complemented by proposing two decompositions: one for the mean difference between the treated and the untreated that extends the Oaxaca-Blinder decomposition by adding the self-selection term; another for the MTE, that is split into the sum of the MTE under rank invariance and a term that captures the excess selection of individuals into each treatment status.

The proposed estimator is a three-step quantile regression estimator. It estimates the SQF, the copula of the unobservables, and the propensity score. The baseline estimator uses a parametric copula, but in an extension I consider using Bernstein copulas, which are a flexible family that can approximate any well-defined continuous copula.

Finally, the estimation methods presented are applied to the Work First Job Placements data. In contrast with what had been found in the literature, 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. Moreover, the shape of the MTE was severely affected by it, and it identified a share of the population whose earnings would have been higher if they had not received the treatment.

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.

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 (2017). Quantile selection models. *Econometrica 85*(1), 1–28.

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. 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., J. C. Escanciano, H. Ichimura, and W. K. Newey (2016). Locally robust semiparametric estimation. *arXiv preprint arXiv:1608.00033*.

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.

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.

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. 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., 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 (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.

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.

Kitagawa, T. (2009). Identification region of the potential outcome distributions under instrument independence. Technical report, cemmap working paper.

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.

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.

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.

Mogstad, M., A. Santos, and A. Torgovitsky (2017). Using instrumental variables for inference about policy relevant treatment effects. Technical report, National Bureau of Economic Research.

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

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.

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. (2017). Copula-based random effects models for clustered data. Technical report, Bank of Italy Temi di Discussione (Working Paper) No 1092.

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.

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.

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.

Wüthrich, K. (2016). A comparison of two qantile models with endogeneity. Technical report, Discussion Papers, Universität Bern, Department of Economics.

A Mathematical proofs

Let $W \equiv (Y, D, Z)$. The following notation is used throughout the Appendix:[30]

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

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

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

A.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 6. Pick a $z_x \in \mathcal{Z}_x$, and define $F_{Y_0^*|X}(y|x) \equiv$

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

$G_X^{-1}\left(F_{Y|D=0,Z}\left(y|z_x\right),\pi\left(z_x\right)\right)$. For all (z,y) in the support of (Z,Y) given $X=x$, we have:

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

By a parallel argument, for H_X satisfying Equation 5, one can get that for all (z,y) in the support of (Z,Y) given $X=x$, we have $H_X\left(F_{Y_1^*|X}\left(y|x\right),\pi\left(z\right)\right)=F_{Y|D=1,Z}\left(y|z\right)$.

A.2 Proof of Lemma 2

$$\begin{aligned} \int_0^1 \Delta^{ESME}\left(x,v\right)dv &= \int_0^1\int_0^1 g_1\left(x,u\right)d\left[C_{1,X}\left(u|v\right)-C_{0,X}\left(u|v\right)\right]dv \\ &= \int_0^1 g_1\left(x,u\right)\int_0^1\left[c_{1,X}\left(u,v\right)-c_{0,X}\left(u,v\right)\right]dvdu \\ &= \int_0^1 g_1\left(x,u\right)\left[C_{1,X}\left(v|u\right)-C_{0,X}\left(v|u\right)\right]_0^1 du \\ &= \int_0^1 g_1\left(x,u\right)\left[u-u\right]_0^1 du = 0 \end{aligned}$$

The second part of the lemma follows trivially.

A.3 Proof of Proposition 1

Let G_X and $\tilde{G}_X$ satisfy Equation 4, and $\pi_1,\pi_2\in\mathcal{P}_x$. Then,

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

Hence, $\forall\tau\in(0,1)$, $(\pi_1,\pi_2)\to G_X\left(G_X^{-1}\left(\tau,\pi_2\right),\pi_1\right)-\tilde{G}_X\left(\tilde{G}_X^{-1}\left(\tau,\pi_2\right),\pi_1\right)$. $C_{0,X}$ is real analytic by Assumption 5, so G_X is also real analytic, and therefore the composition is real analytic. Hence, because it is zero on a product of two open neighborhoods, it is zero

everywhere on $(0,1)\times(0,1)$. 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 3 and taking limits at $\pi_2 = 1$, H_X, $C_{1,X}$, and $g_1(x,u)$ are identified.

A.4 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 6, $\sup_{(\beta,\theta,\gamma)\in\mathcal{B}\times\Theta\times\Gamma} \|Q_n(\beta,\theta,\gamma,\tau) - Q(\beta,\theta,\gamma,\tau)\| \xrightarrow{P} 0$. Thus, by Lemma 5, $\sup_{\tau\in\mathcal{T}} \left\|\hat{\vartheta}(\tau) - \vartheta(\tau)\right\| \xrightarrow{P} 0$.

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 6 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

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

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

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

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

$$\begin{aligned}\sqrt{n}\left(\hat{\beta}_1(\tau) - \beta_1(\tau)\right) = -J_1(\tau)^{-1}\Big\{&\mathbb{G}_n\left[DX\zeta_{H_X(\tau,\pi(Z;\gamma);\theta_1)}(\varepsilon_1(\tau))\right] \\ &-P_{11}(\tau)\sqrt{n}(\hat{\gamma}-\gamma) - P_{12}(\tau)\sqrt{n}\left(\hat{\theta}_1 - \theta_1\right)\Big\} + o_P(1) \qquad (18)\end{aligned}$$

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

$$\begin{aligned}\sqrt{n}\left(\hat{\beta}_0(\tau) - \beta_0(\tau)\right) = -J_0(\tau)^{-1}\Big\{&\mathbb{G}_n\left[(1-D)X\zeta_{G_X(\tau,\pi(Z;\gamma);\theta_0)}(\varepsilon_0(\tau))\right] \\ &-P_{01}(\tau)\sqrt{n}(\hat{\gamma}-\gamma) - P_{02}(\tau)\sqrt{n}\left(\hat{\theta}_0 - \theta_0\right)\Big\} + o_P(1) \qquad (19)\end{aligned}$$

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 6 and Assumption 10, the following expansion holds:

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

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

$$
\begin{aligned}
\sqrt{n}\left(\hat{\theta}_1-\theta_1\right) &= \left[\int_0^1 \tilde{P}_{12}\left(u\right)du\right]^{-1}\left\{\mathbb{G}_n\left[\int_0^1 D\varphi\left(u,Z\right)\zeta_{H_X\left(u,\pi\left(Z;\gamma\right);\theta_1\right)}\left(\varepsilon_1\left(u\right)\right)du\right]\right. \\
&\quad\left.+\sqrt{n}\int_0^1 \tilde{J}_1\left(u\right)\left(\hat{\beta}_1\left(u\right)-\beta_1\left(u\right)\right)du-\sqrt{n}\int_0^1 \tilde{P}_{11}\left(u\right)du\left(\hat{\gamma}-\gamma\right)\right\}+o_P\left(1\right)
\end{aligned}
\tag{20}
$$

By a parallel argument, it can be shown that

$$\sqrt{n}\left(\hat{\theta}_0 - \theta_0\right) = \left[\int_0^1 \tilde{P}_{02}(u)\,du\right]^{-1}\left\{\mathbb{G}_n\left[\int_0^1 (1-D)\,\varphi(u,Z)\,\zeta_{G_X(u,\pi(Z;\gamma);\theta_0)}(\varepsilon_0(u))\,du\right] + \sqrt{n}\int_0^1 \tilde{J}_0(u)\left(\hat{\beta}_0(u) - \beta_0(u)\right)du - \sqrt{n}\int_0^1 \tilde{P}_{01}(u)\,du\,(\hat{\gamma}-\gamma)\right\} + o_P(1) \tag{21}$$

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

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

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

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

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

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

Combining Equations 18, 19, 20, and 21 yields

$$A(\tau) = F(\tau) A(\tau) + \int_0^1 D(u) A(u)\,du + C(\tau) \frac{1}{\sqrt{n}} \mathbb{G}_n \psi(\tau) + o_P\left(\frac{1}{\sqrt{n}}\right) \tag{22}$$

in $\ell^\infty(\mathcal{T})$. Equation 22 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:

$$\begin{aligned} \sqrt{n} A(\tau) = {} & F^I(\tau) \left(I - \int_0^1 D(u) F^I(u)\,du\right)^{-1} \int_0^1 D(u) F^I(u) C(u) \mathbb{G}_n \psi(u)\,du \\ & + F^I(\tau) C(\tau) \mathbb{G}_n \psi(\tau) + o_P(1) \end{aligned} \tag{23}$$

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

and Lemmas 3 and 5, it follows that $\sqrt{n}\left(\hat{\vartheta}(\tau)-\vartheta(\tau)\right)\Rightarrow S(\tau)$, where

$$S(\tau)\equiv F^{I}(\tau)\left[C(\tau)+\left(I-\int_{0}^{1}D(u)F^{I}(u)\,du\right)^{-1}\int_{0}^{1}D(u)F^{I}(u)C(u)\,du\right]R(\tau)$$

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

$$\Sigma_S(\tau,\tau')=F^{I}(\tau)\left[C(\tau)+\left(I-\int_{0}^{1}D(u)F^{I}(u)\,du\right)^{-1}\int_{0}^{1}D(u)F^{I}(u)C(u)\,du\right]\Sigma_R(\tau,\tau')$$
$$\left\{F^{I}(\tau')\left[C(\tau')+\left(I-\int_{0}^{1}D(u)F^{I}(u)\,du\right)^{-1}\int_{0}^{1}D(u)F^{I}(u)C(u)\,du\right]\right\}'$$

and $\Sigma_R(\tau,\tau')$ is defined in Lemma 6.

B Auxiliary Lemmas

B.1 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 $\lambda(h_t)\equiv\int_0^1\lambda(u)(\nu(u)+th_t(u))\,du$. As $t\to0$,*

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

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 $h_t:\|h_t-h\|_{\infty}\to0$, 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.2 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.3 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.4 Stochastic Expansion

Lemma 6. *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. $\mathbb{G}_n r(W,\beta(\tau),\theta,\gamma,\tau) \Rightarrow R(\tau)$ *in* $\ell^\infty(\mathcal{T})$, *where* $R(\tau)$ *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

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

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 τ'' between τ and τ'. By Assumptions 4 and 12, the second term is bounded $\forall z, \tau''$, so it follows that $G_X \in \mathcal{F}$.[31] Using a parallel argument, it can be shown that $H_X \in \mathcal{F}$. Let $\mathcal{T} \equiv \{\tau \mapsto \tau\}$ and define

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

[31] 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 bounded by Assumption 12, the result follows.

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 $\mathcal{F}$.[32] 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

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

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

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

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,

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

with covariate function

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

where

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

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

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

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

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

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

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

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

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 ξ as the $L_2(P)$ pseudometric on $\mathcal{H}$:

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

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 Assumption 9, $\delta_n \xrightarrow{p} 0$. Therefore, as $\delta_n \xrightarrow{p} 0$,

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

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

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

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:

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

which gives

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

which implies claim 1. □

C A Generalized Roy Model with Imperfect Information

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 $S \equiv Y_1 - Y_0 - K \geq 0$. This model is known as the generalized Roy model with perfect information.

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:

$$\mathbb{E}\left[S|Z,V\right] = \mathbb{E}\left[g_1\left(X,U_1\right) - g_0\left(X,U_0\right)|Z,V\right] - k\left(Z\right) - \mathbb{E}\left[U_K|Z,V\right]$$

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

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

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, 2007). 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

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

where $\pi\left(Z\right) \equiv \{p : \mu_1\left(X,p\right) - \mu_0\left(X,p\right) - k\left(Z\right) - \mathbb{E}\left[U_k|Z,p\right] = 0\}$. Note that the net surplus in this case depends on two terms, the MTE, $\mu_1\left(X,V\right) - \mu_0\left(X,V\right)$, and the expected cost, $k\left(Z\right) + \mathbb{E}\left[U_k|Z,V\right]$. 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 Comparison with Alternative Identification Conditions

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

$$\frac{\partial}{\partial p} C_{0,X}\left(F_{Y_0^*}\left(y|x\right), p\right)\Big|_{p=\pi(z)} = F_{Y|D=0,Z}\left(y|z\right) - \left(1-\pi\left(z\right)\right)\frac{\partial}{\partial \pi\left(z\right)} F_{Y|D=0,Z}\left(y|z\right) \tag{24}$$

$$\frac{\partial}{\partial p} C_{1,X}\left(F_{Y_1^*}\left(y|x\right), p\right)\Big|_{p=\pi(z)} = F_{Y|D=1,Z}\left(y|z\right) + \pi\left(z\right)\frac{\partial}{\partial \pi\left(z\right)} F_{Y|D=1,Z}\left(y|z\right) \tag{25}$$

Without extra assumptions, LIV identifies the left had side of Equations 24-25 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 5, 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 5. 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.[33] Another possibility is to relax Assumption 1 to allow for full independence between the unobservables and both the instrument and the covariates. Then, one could use variation in X as a source of identification. This assumption, however, imposes severe restrictions on the amount of heterogeneity that can be displayed by the model. In particular, 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 with respect to the covariates.[34]

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. (2017) 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,

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

[34] This assumption is strong enough to achieve identification of the MTE even when the instrument is binary. See Kitagawa (2009) for further discussion on the identified sets when both the instrument and the treatment are binary.

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.[35] 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:

$$\mathbb{P}\left(Y \leq g_D\left(X,\tau\right)|Z\right) = \tau - C_{0,X}\left(F_{Y_0^*}\left(g_0\left(X,\tau\right)\right),\pi\left(Z\right)\right) + C_{1,X}\left(F_{Y_1^*}\left(g_1\left(X,\tau\right)\right),\pi\left(Z\right)\right) \tag{26}$$

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 (Equation 2) and the copula. Nevertheless, it is still possible to combine Equation 26 with Frechét-Hoeffding bounds to obtain set identification:

$$\begin{aligned} \tau + \min\left\{F_{Y_0^*}\left(g_0\left(X,\tau\right)\right),\pi\left(Z\right)\right\} - \max\left\{F_{Y_1^*}\left(g_1\left(X,\tau\right)\right) - \pi\left(Z\right),0\right\} \leq \\ \mathbb{P}\left(Y \leq g_D\left(X,\tau\right)|Z\right) \leq \\ \tau + \min\left\{F_{Y_1^*}\left(g_1\left(X,\tau\right)\right),\pi\left(Z\right)\right\} - \max\left\{F_{Y_0^*}\left(g_0\left(X,\tau\right)\right) - \pi\left(Z\right),0\right\} \end{aligned} \tag{27}$$

[35] When the instrument is binary, the IVQR estimator is closely connected with the LQTE estimator. See Wüthrich (2016) for further details.

E Additional Treatment Effects

Using Equation 10, it is possible to express the TUT and the TT in terms of the SQF and the copula:

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

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

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 ATE:

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

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 their estimation, one just needs to replace the SQF, the copula, and the propensity score by their estimated counterparts:

$$\hat{\Delta}^{TUT}(z_i) = \int_0^1 x_i' \hat{\beta}_1(\tau)\, d\hat{G}'_{X,i,\tau} - \int_0^1 x_i' \hat{\beta}_0(\tau)\, d\hat{G}_{X,i,\tau} \tag{28}$$

$$\hat{\Delta}^{TT}(z_i) = \int_0^1 x_i' \hat{\beta}_1(\tau)\, d\hat{H}_{X,i,\tau} - \int_0^1 x_i' \hat{\beta}_0(\tau)\, d\hat{H}'_{X,i,\tau} \tag{29}$$

$$\hat{\Delta}^{ATE}(z_i) = \int_0^1 x_i' \left(\hat{\beta}_1(\tau) - \hat{\beta}_0(\tau)\right) d\tau \tag{30}$$

where $\hat{G}'_{X,i,\tau} \equiv G'_X\left(\tau, \hat{\pi}(z_i); \hat{\theta}_0\right)$ and $\hat{H}'_{X,i,\tau} \equiv H'_X\left(\tau, \hat{\pi}(z_i); \hat{\theta}_1\right)$. Finally, the unconditional treatment effects can be obtained by taking the average over $i = 1, ..., N$.

F Monte Carlo

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

$$y_i = \beta_{d_i,1}\left(\tau_{d_i,i}\right) + x_i \beta_{d_i,2}\left(u_{d_i,i}\right) \tag{31}$$

$$d_i = \mathbf{1}\left(\gamma_1 + x_i\gamma_2 + z_i\gamma_3 + \Lambda^{-1}\left(v_i\right) > 0\right) \tag{32}$$

$$u_{0,i}, u_{1,i}, v_i | z_i \sim Gaussian\left(\Sigma\right) \tag{33}$$

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 6 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 7, 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 6: 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 7: 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 8): 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.[36]

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 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.

[36] 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 8: 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.

F.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.[37] 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 9), and the distance between the estimated copula and its true value is smaller (Table 10). 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.

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

As shown in Table 11, 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,

[37] 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 9: 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 10: 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 $D = 0, 1$.

as shown in Table 12.

Table 11: 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 12: Quantile Regression Coefficients, Non-Analytical Copula

		$\tau^{1,1}$						$\tau^{0,1}$					
		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

		$\tau^{1.2}$						$\tau^{0.2}$					
		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.

RECENTLY PUBLISHED "TEMI" (*)

N. 1188 – *Raising aspirations and higher education: evidence from the UK's Widening Participation Policy*, by Lucia Rizzica (September 2018).

N. 1189 – *Nearly exact Bayesian estimation of non-linear no-arbitrage term structure models*, by Marcello Pericoli and Marco Taboga (September 2018).

N. 1190 – *Granular Sources of the Italian business cycle*, by Nicolò Gnocato and Concetta Rondinelli (September 2018).

N. 1191 – *Debt restructuring with multiple bank relationships*, by Angelo Baglioni, Luca Colombo and Paola Rossi (September 2018).

N. 1192 – *Exchange rate pass-through into euro area inflation. An estimated structural model*, by Lorenzo Burlon, Alessandro Notarpietro and Massimiliano Pisani (September 2018).

N. 1193 – *The effect of grants on university drop-out rates: evidence on the Italian case*, by Francesca Modena, Enrico Rettore and Giulia Martina Tanzi (September 2018).

N. 1194 – *Potential output and microeconomic heterogeneity*, by Davide Fantino (November 2018).

N. 1195 – *Immigrants, labor market dynamics and adjustment to shocks in the Euro Area*, by Gaetano Basso, Francesco D'Amuri and Giovanni Peri (November 2018).

N. 1196 – *Sovereign debt maturity structure and its costs*, by Flavia Corneli (November 2018).

N. 1197 – *Fiscal policy in the US: a new measure of uncertainty and its recent development*, by Alessio Anzuini and Luca Rossi (November 2018).

N. 1198 – *Macroeconomics determinants of the correlation between stocks and bonds*, by Marcello Pericoli (November 2018).

N. 1199 – *Bank capital constraints, lending supply and economic activity*, by Antonio M. Conti, Andrea Nobili and Federico M. Signoretti (November 2018).

N. 1200 – *The effectiveness of capital controls*, by Valerio Nispi Landi and Alessandro Schiavone (November 2018).

N. 1201 – *Contagion in the CoCos market? A case study of two stress events*, by Pierluigi Bologna, Arianna Miglietta and Anatoli Segura (November 2018).

N. 1202 – *Is ECB monetary policy more powerful during expansions?*, by Martina Cecioni (December 2018).

N. 1203 – *Firms' inflation expectations and investment plans*, by Adriana Grasso and Tiziano Ropele (December 2018).

N. 1204 – *Recent trends in economic activity and TFP in Italy with a focus on embodied technical progress*, by Alessandro Mistretta and Francesco Zollino (December 2018).

N. 1205 – *Benefits of Gradualism or Costs of Inaction? Monetary Policy in Times of Uncertainty*, by Giuseppe Ferrero, Mario Pietrunti and Andrea Tiseno (February 2019).

N. 1206 – *Machine learning in the service of policy targeting: the case of public credit guarantees*, by Monica Andini, Michela Boldrini, Emanuele Ciani, Guido de Blasio, Alessio D'Ignazio and Andrea Paladini (February 2019).

N. 1207 – *Do the ECB's monetary policies benefit Emerging Market Economies? A GVAR analysis on the crisis and post-crisis period*, by Andrea Colabella (February 2019).

N. 1208 – *The Economic Effects of Big Events: Evidence from the Great Jubilee 2000 in Rome*, by Raffaello Bronzini, Sauro Mocetti and Matteo Mongardini (February 2019).

N. 1209 – *The added value of more accurate predictions for school rankings*, by Fritz Schiltz, Paolo Sestito, Tommaso Agasisti and Kristof De Witte (February 2019).

(*) Requests for copies should be sent to:
Banca d'Italia – Servizio Studi di struttura economica e finanziaria – Divisione Biblioteca e Archivio storico – Via Nazionale, 91 – 00184 Rome – (fax 0039 06 47922059). They are available on the Internet www.bancaditalia.it.

2017

AABERGE, R., F. BOURGUIGNON, A. BRANDOLINI, F. FERREIRA, J. GORNICK, J. HILLS, M. JÄNTTI, S. JENKINS, J. MICKLEWRIGHT, E. MARLIER, B. NOLAN, T. PIKETTY, W. RADERMACHER, T. SMEEDING, N. STERN, J. STIGLITZ, H. SUTHERLAND, *Tony Atkinson and his legacy*, Review of Income and Wealth, v. 63, 3, pp. 411-444, **WP 1138 (September 2017).**

ACCETTURO A., M. BUGAMELLI and A. LAMORGESE, *Law enforcement and political participation: Italy 1861-65*, Journal of Economic Behavior & Organization, v. 140, pp. 224-245, **WP 1124 (July 2017).**

ADAMOPOULOU A. and G.M. TANZI, *Academic dropout and the great recession*, Journal of Human Capital, V. 11, 1, pp. 35–71, **WP 970 (October 2014).**

ALBERTAZZI U., M. BOTTERO and G. SENE, *Information externalities in the credit market and the spell of credit rationing*, Journal of Financial Intermediation, v. 30, pp. 61–70, **WP 980 (November 2014).**

ALESSANDRI P. and H. MUMTAZ, *Financial indicators and density forecasts for US output and inflation*, Review of Economic Dynamics, v. 24, pp. 66-78, **WP 977 (November 2014).**

BARBIERI G., C. ROSSETTI and P. SESTITO, *Teacher motivation and student learning*, Politica economica/Journal of Economic Policy, v. 33, 1, pp.59-72, **WP 761 (June 2010).**

BENTIVOGLI C. and M. LITTERIO, *Foreign ownership and performance: evidence from a panel of Italian firms*, International Journal of the Economics of Business, v. 24, 3, pp. 251-273, **WP 1085 (October 2016).**

BRONZINI R. and A. D'IGNAZIO, *Bank internationalisation and firm exports: evidence from matched firm-bank data*, Review of International Economics, v. 25, 3, pp. 476-499 **WP 1055 (March 2016).**

BRUCHE M. and A. SEGURA, *Debt maturity and the liquidity of secondary debt markets*, Journal of Financial Economics, v. 124, 3, pp. 599-613, **WP 1049 (January 2016).**

BURLON L., *Public expenditure distribution, voting, and growth*, Journal of Public Economic Theory,, v. 19, 4, pp. 789–810, **WP 961 (April 2014).**

BURLON L., A. GERALI, A. NOTARPIETRO and M. PISANI, *Macroeconomic effectiveness of non-standard monetary policy and early exit. a model-based evaluation*, International Finance, v. 20, 2, pp.155-173, **WP 1074 (July 2016).**

BUSETTI F., *Quantile aggregation of density forecasts*, Oxford Bulletin of Economics and Statistics, v. 79, 4, pp. 495-512, **WP 979 (November 2014).**

CESARONI T. and S. IEZZI, *The predictive content of business survey indicators: evidence from SIGE*, Journal of Business Cycle Research, v.13, 1, pp 75–104, **WP 1031 (October 2015).**

CONTI P., D. MARELLA and A. NERI, *Statistical matching and uncertainty analysis in combining household income and expenditure data*, Statistical Methods & Applications, v. 26, 3, pp 485–505, **WP 1018 (July 2015).**

D'AMURI F., *Monitoring and disincentives in containing paid sick leave*, Labour Economics, v. 49, pp. 74-83, **WP 787 (January 2011).**

D'AMURI F. and J. MARCUCCI, *The predictive power of google searches in forecasting unemployment*, International Journal of Forecasting, v. 33, 4, pp. 801-816, **WP 891 (November 2012).**

DE BLASIO G. and S. POY, *The impact of local minimum wages on employment: evidence from Italy in the 1950s*, Journal of Regional Science, v. 57, 1, pp. 48-74, **WP 953 (March 2014).**

DEL GIOVANE P., A. NOBILI and F. M. SIGNORETTI, *Assessing the sources of credit supply tightening: was the sovereign debt crisis different from Lehman?*, International Journal of Central Banking, v. 13, 2, pp. 197-234, **WP 942 (November 2013).**

DEL PRETE S., M. PAGNINI, P. ROSSI and V. VACCA, *Lending organization and credit supply during the 2008–2009 crisis*, Economic Notes, v. 46, 2, pp. 207–236, **WP 1108 (April 2017).**

DELLE MONACHE D. and I. PETRELLA, *Adaptive models and heavy tails with an application to inflation forecasting*, International Journal of Forecasting, v. 33, 2, pp. 482-501, **WP 1052 (March 2016).**

FEDERICO S. and E. TOSTI, *Exporters and importers of services: firm-level evidence on Italy*, The World Economy, v. 40, 10, pp. 2078-2096, **WP 877 (September 2012).**

GIACOMELLI S. and C. MENON, *Does weak contract enforcement affect firm size? Evidence from the neighbour's court*, Journal of Economic Geography, v. 17, 6, pp. 1251-1282, **WP 898 (January 2013).**

LOBERTO M. and C. PERRICONE, *Does trend inflation make a difference?*, Economic Modelling, v. 61, pp. 351–375, **WP 1033 (October 2015).**

MANCINI A.L., C. MONFARDINI and S. PASQUA, *Is a good example the best sermon? Children's imitation of parental reading*, Review of Economics of the Household, v. 15, 3, pp 965–993, **D No. 958 (April 2014).**

MEEKS R., B. NELSON and P. ALESSANDRI, *Shadow banks and macroeconomic instability*, Journal of Money, Credit and Banking, v. 49, 7, pp. 1483–1516, **WP 939 (November 2013).**

MICUCCI G. and P. ROSSI, *Debt restructuring and the role of banks' organizational structure and lending technologies*, Journal of Financial Services Research, v. 51, 3, pp 339–361, **WP 763 (June 2010).**

MOCETTI S., M. PAGNINI and E. SETTE, *Information technology and banking organization*, Journal of Journal of Financial Services Research, v. 51, pp. 313-338, **WP 752 (March 2010).**

MOCETTI S. and E. VIVIANO, *Looking behind mortgage delinquencies*, Journal of Banking & Finance, v. 75, pp. 53-63, **WP 999 (January 2015).**

NOBILI A. and F. ZOLLINO, *A structural model for the housing and credit market in Italy*, Journal of Housing Economics, v. 36, pp. 73-87, **WP 887 (October 2012).**

PALAZZO F., *Search costs and the severity of adverse selection*, Research in Economics, v. 71, 1, pp. 171-197, **WP 1073 (July 2016).**

PATACCHINI E. and E. RAINONE, *Social ties and the demand for financial services*, Journal of Financial Services Research, v. 52, 1–2, pp 35–88, **WP 1115 (June 2017).**

PATACCHINI E., E. RAINONE and Y. ZENOU, *Heterogeneous peer effects in education*, Journal of Economic Behavior & Organization, v. 134, pp. 190–227, **WP 1048 (January 2016).**

SBRANA G., A. SILVESTRINI and F. VENDITTI, *Short-term inflation forecasting: the M.E.T.A. approach*, International Journal of Forecasting, v. 33, 4, pp. 1065-1081, **WP 1016 (June 2015).**

SEGURA A. and J. SUAREZ, *How excessive is banks' maturity transformation?*, Review of Financial Studies, v. 30, 10, pp. 3538–3580, **WP 1065 (April 2016).**

VACCA V., *An unexpected crisis? Looking at pricing effectiveness of heterogeneous banks*, Economic Notes, v. 46, 2, pp. 171–206, **WP 814 (July 2011).**

VERGARA CAFFARELI F., *One-way flow networks with decreasing returns to linking*, Dynamic Games and Applications, v. 7, 2, pp. 323-345, **WP 734 (November 2009).**

ZAGHINI A., *A Tale of fragmentation: corporate funding in the euro-area bond market*, International Review of Financial Analysis, v. 49, pp. 59-68, **WP 1104 (February 2017).**

2018

ADAMOPOULOU A. and E. KAYA, *Young adults living with their parents and the influence of peers*, Oxford Bulletin of Economics and Statistics,v. 80, pp. 689-713, **WP 1038 (November 2015).**

ANDINI M., E. CIANI, G. DE BLASIO, A. D'IGNAZIO and V. SILVESTRINI, *Targeting with machine learning: an application to a tax rebate program in Italy*, Journal of Economic Behavior & Organization, v. 156, pp. 86-102, **WP 1158 (December 2017).**

BARONE G., G. DE BLASIO and S. MOCETTI, *The real effects of credit crunch in the great recession: evidence from Italian provinces*, Regional Science and Urban Economics, v. 70, pp. 352-59, **WP 1057 (March 2016).**

BELOTTI F. and G. ILARDI *Consistent inference in fixed-effects stochastic frontier models*, Journal of Econometrics, v. 202, 2, pp. 161-177, **WP 1147 (October 2017).**

BERTON F., S. MOCETTI, A. PRESBITERO and M. RICHIARDI, *Banks, firms, and jobs*, Review of Financial Studies, v.31, 6, pp. 2113-2156, **WP 1097 (February 2017).**

BOFONDI M., L. CARPINELLI and E. SETTE, *Credit supply during a sovereign debt crisis*, Journal of the European Economic Association, v.16, 3, pp. 696-729, **WP 909 (April 2013).**

BOKAN N., A. GERALI, S. GOMES, P. JACQUINOT and M. PISANI, *EAGLE-FLI: a macroeconomic model of banking and financial interdependence in the euro area*, Economic Modelling, v. 69, C, pp. 249-280, **WP 1064 (April 2016).**

BRILLI Y. and M. TONELLO, *Does increasing compulsory education reduce or displace adolescent crime? New evidence from administrative and victimization data*, CESifo Economic Studies, v. 64, 1, pp. 15–4, **WP 1008 (April 2015).**

BUONO I. and S. FORMAI *The heterogeneous response of domestic sales and exports to bank credit shocks,* Journal of International Economics, v. 113, pp. 55-73, **WP 1066 (March 2018).**

BURLON L., A. GERALI, A. NOTARPIETRO and M. PISANI, *Non-standard monetary policy, asset prices and macroprudential policy in a monetary union,* Journal of International Money and Finance, v. 88, pp. 25-53, **WP 1089 (October 2016).**

CARTA F. and M. DE PHLIPPIS, *You've Come a long way, baby. Husbands' commuting time and family labour supply,* Regional Science and Urban Economics, v. 69, pp. 25-37, **WP 1003 (March 2015).**

CARTA F. and L. RIZZICA, *Early kindergarten, maternal labor supply and children's outcomes: evidence from Italy,* Journal of Public Economics, v. 158, pp. 79-102, **WP 1030 (October 2015).**

CASIRAGHI M., E. GAIOTTI, L. RODANO and A. SECCHI, *A "Reverse Robin Hood"? The distributional implications of non-standard monetary policy for Italian households,* Journal of International Money and Finance, v. 85, pp. 215-235, **WP 1077 (July 2016).**

CECCHETTI S., F. NATOLI and L. SIGALOTTI, *Tail co-movement in inflation expectations as an indicator of anchoring,* International Journal of Central Banking, v. 14, 1, pp. 35-71, **WP 1025 (July 2015).**

CIANI E. and C. DEIANA, *No Free lunch, buddy: housing transfers and informal care later in life,* Review of Economics of the Household, v.16, 4, pp. 971-1001, **WP 1117 (June 2017).**

CIPRIANI M., A. GUARINO, G. GUAZZAROTTI, F. TAGLIATI and S. FISHER, *Informational contagion in the laboratory,* Review of Finance, v. 22, 3, pp. 877-904, **WP 1063 (April 2016).**

DE BLASIO G, S. DE MITRI, S. D'IGNAZIO, P. FINALDI RUSSO and L. STOPPANI, *Public guarantees to SME borrowing. A RDD evaluation,* Journal of Banking & Finance, v. 96, pp. 73-86, **WP 1111 (April 2017).**

GERALI A., A. LOCARNO, A. NOTARPIETRO and M. PISANI, *The sovereign crisis and Italy's potential output,* Journal of Policy Modeling, v. 40, 2, pp. 418-433, **WP 1010 (June 2015).**

LIBERATI D., *An estimated DSGE model with search and matching frictions in the credit market,* International Journal of Monetary Economics and Finance (IJMEF), v. 11, 6, pp. 567-617, **WP 986 (November 2014).**

LINARELLO A., *Direct and indirect effects of trade liberalization: evidence from Chile,* Journal of Development Economics, v. 134, pp. 160-175, **WP 994 (December 2014).**

NUCCI F. and M. RIGGI, *Labor force participation, wage rigidities, and inflation,* Journal of Macroeconomics, v. 55, 3 pp. 274-292, **WP 1054 (March 2016).**

RIGON M. and F. ZANETTI, *Optimal monetary policy and fiscal policy interaction in a non_ricardian economy,* International Journal of Central Banking, v. 14 3, pp. 389-436, **WP 1155 (December 2017).**

SEGURA A., *Why did sponsor banks rescue their SIVs?,* Review of Finance, v. 22, 2, pp. 661-697, **WP 1100 (February 2017).**

2019

CIANI E. and P. FISHER, *Dif-in-dif estimators of multiplicative treatment effects,* Journal of Econometric Methods, v. 8. 1, pp. 1-10, **WP 985 (November 2014).**

FORTHCOMING

ACCETTURO A., W. DI GIACINTO, G. MICUCCI and M. PAGNINI, *Geography, productivity and trade: does selection explain why some locations are more productive than others?,* Journal of Regional Science, **WP 910 (April 2013).**

ALBANESE G., G. DE BLASIO and P. SESTITO, *Trust, risk and time preferences: evidence from survey data,* International Review of Economics, **WP 911 (April 2013).**

APRIGLIANO V., G. ARDIZZI and L. MONTEFORTE, *Using the payment system data to forecast the economic activity,* International Journal of Central Banking, **WP 1098 (February 2017).**

ARNAUDO D., G. MICUCCI, M. RIGON and P. ROSSI, *Should I stay or should I go? Firms' mobility across banks in the aftermath of the financial crisis,* Italian Economic Journal / Rivista italiana degli economisti, **WP 1086 (October 2016).**

BELOTTI F. and G. ILARDI, *Consistent inference in fixed-effects stochastic frontier models*, Journal of Econometrics, **WP 1147 (October 2017).**

BUSETTI F. and M. CAIVANO, *Low frequency drivers of the real interest rate: empirical evidence for advanced economies*, International Finance, **WP 1132 (September 2017).**

CHIADES P., L. GRECO, V. MENGOTTO, L. MORETTI and P. VALBONESI, *Fiscal consolidation by intergovernmental transfers cuts? Economic Modelling*, **WP 1076 (July 2016).**

CIANI E., F. DAVID and G. DE BLASIO, *Local responses to labor demand shocks: a re-assessment of the case of Italy*, IMF Economic Review, **WP 1112 (April 2017).**

COLETTA M., R. DE BONIS and S. PIERMATTEI, *Household debt in OECD countries: the role of supply-side and demand-side factors*, Social Indicators Research, **WP 989 (November 2014).**

CORSELLO F. and V. NISPI LANDI, *Labor market and financial shocks: a time-varying analysis*, Journal of Money, Credit and Banking, **WP 1179 (June 2018).**

COVA P., P. PAGANO and M. PISANI, *Domestic and international macroeconomic effects of the Eurosystem Expanded Asset Purchase Programme*, IMF Economic Review, **WP 1036 (October 2015).**

D'AMURI F., *Monitoring and disincentives in containing paid sick leave*, Labour Economics, **WP 787 (January 2011).**

D'IGNAZIO A. and C. MENON, *The causal effect of credit Guarantees for SMEs: evidence from Italy*, Scandinavian Journal of Economics, **WP 900 (February 2013).**

ERCOLANI V. and J. VALLE E AZEVEDO, *How can the government spending multiplier be small at the zero lower bound?*, Macroeconomic Dynamics, **WP 1174 (April 2018).**

FEDERICO S. and E. TOSTI, *Exporters and importers of services: firm-level evidence on Italy*, The World Economy, **WP 877 (September 2012).**

GERALI A. and S. NERI, *Natural rates across the Atlantic*, Journal of Macroeconomics, **WP 1140 (September 2017).**

GIACOMELLI S. and C. MENON, *Does weak contract enforcement affect firm size? Evidence from the neighbour's court*, Journal of Economic Geography, **WP 898 (January 2013).**

GIORDANO C., M. MARINUCCI and A. SILVESTRINI, *The macro determinants of firms' and households' investment: evidence from Italy*, Economic Modelling, **WP 1167 (March 2018).**

NATOLI F. and L. SIGALOTTI, *Tail co-movement in inflation expectations as an indicator of anchoring*, International Journal of Central Banking, **WP 1025 (July 2015).**

RIGGI M., *Capital destruction, jobless recoveries, and the discipline device role of unemployment*, Macroeconomic Dynamics, **WP 871 (July 2012).**

RIZZICA L., *Raising aspirations and higher education. evidence from the UK's widening participation policy*, Journal of Labor Economics, **WP 1188 (September 2018).**

SEGURA A., *Why did sponsor banks rescue their SIVs?*, Review of Finance, **WP 1100 (February 2017).**