Estimation of Counterfactual Distributions with a Continuous Endogenous Treatment

Santiago Pereda-Fernández*

Universidad de Cantabria

December 11, 2023

Abstract

In this paper I propose a method to estimate the counterfactual distribution of an outcome variable when the treatment is endogenous, continuous, and its effect is heterogeneous. The types of counterfactuals considered are those in which the change in treatment intensity can be correlated with the individual effects or when some of the structural functions are changed by some other group's counterparts. I characterize the outcome and the treatment with a triangular system of equations in which the unobservables are related by a copula that captures the endogeneity of the treatment, which is nonparametrically identified by inverting the quantile processes that determine the outcome and the treatment. Both processes are estimated using existing quantile regression methods, and I propose a parametric and a nonparametric estimator of the copula. To illustrate these methods, I estimate several counterfactual distributions of the birth weight of children, had their mothers smoked differently during pregnancy.

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

JEL classification: C31, C36

*Departamento de Economía, Universidad de Cantabria, Avenida de los Castros, s/n, 39005 Santander, Spain. I would like to thank the editor, Esfandiar Maasoumi, the associate editor, two anonymous referees, Manuel Arellano, Stéphane Bonhomme, Domenico Depalo, Bryan Graham, Giuseppe Ilardi, Michael Jansson, James Powell, Demian Pouzo, Enrique Sentana, and seminar participants at CEMFI and University of California, Berkeley for their helpful comments and discussion. This work is part of the I+D+i project Ref. TED2021-131763A-I00 financed by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR. I gratefully acknowledge financial support from the Spanish Ministry of Universities and the European Union-NextGenerationEU (RMZ-18). All remaining errors are my own. I can be reached via email at santiagopereda@gmail.com

1 Introduction

Consider the problem of assessing of the distributional effects of a policy intervention, or the decomposition of the difference of some variable between two or more populations into several effects. These are two examples that can be addressed by estimating counterfactual distributions. Typically, these have focused on cases in which the treatment was randomly assigned or, if it was endogenously determined, the endogeneity only affects the estimation of the structural parameters, but not the counterfactuals. For instance, if the counterfactual treatment is either randomly assigned or set at a fixed value for every individual.

In this paper I study the estimation of counterfactual distributions when the treatment is continuous, endogenous, its effect is heterogeneous, and the population can be split into two or more groups. I consider two types of counterfactuals: one in which some of the structural functions of one group are swapped by the counterparts of another group; another in which the treatment intensity for each individual is changed, such that the counterfactual value is correlated with the unobservables. Whereas the first may be more useful for decomposition purposes, the second one may be more relevant when the policy maker has some limited ability to enforce the treatment.

In this paper I make the following contributions. First, I propose two estimators of the counterfactual distribution, one based on a parametric estimator of the distribution of the unobservables, and the other on a nonparametric estimator. I study their asymptotic properties, how to conduct inference, and compare them to alternative methods. Second, I show that the distribution of the unobservables is identified, and describe how it determines the outcome distribution. This is necessary to evaluate certain counterfactuals and to compute the Marginal Treatment Effect (MTE, Heckman and Vytlacil, 2007; Florens et al., 2008). Third, I illustrate these methods with an application of how reducing smoking intensity in pregnant women would affect the weight distribution of newborns.

One aspect that is often overlooked in applied analysis is the relation between the different unobservables and how they can affect the distribution of both the treatment and the outcome variables. I model the distribution of the unobservables with a bivariate copula. Thus, one can intuitively interpret these unobservables as the conditional ranks of unobserved ability. Moreover, it is also convenient for the estimation, as it allows to represent the equations that determine the treatment and the outcome as quantile processes.

The copula is a determinant of both the treatment intensity and the outcome. As such, it is a crucial element to analyze the possible effects of a policy intervention. Additionally, the MTE can be expressed as an integral of the derivative of the outcome equation weighted by the copula. Therefore, marginal variations of the treatment intensity would have a heterogeneous effect on the outcome across individuals, partly because of the different value of the unobservables. As a consequence, the copula also has a first order effect in the estimation of counterfactuals that change the treatment distribution unevenly.

The estimation of the treatment and outcome equations is done using existing quantile regression methods. In particular, I estimate them using Quantile Regression (QR, Koenker and Bassett, 1978) and Instrumental Variables Quantile Regression (IVQR, Chernozhukov and Hansen, 2005), respectively. These estimates are used as inputs for the estimation of the copula, which can be done either parametrically, or non-parametrically. Both estimators require the inversion of the quantile processes that conform the triangular model. Then, these estimators are combined to obtain either an estimate of the actual distribution, or of one of the counterfactual distributions. Moreover, they could also be used to estimate other functionals of interest, such as the MTE or the unconditional quantile treatment effect.

I consider two types of counterfactuals. The first one involves swapping the structural functions of one group by the counterparts of another group. This type is useful for decomposition purposes. In an exogenous setting, the Oaxaca-Blinder decomposition splits differences between the two groups into differences in covariates and differences in the returns to these covariates (the slope parameters). The endogeneity of the model allows to consider two additional sources for the differences in outcomes: the first stage equation that determines treatment intensity, and the copula that captures the amount of endogeneity. The second type changes the distribution of the treatment intensity in a way that depends on the unobservables. This may be more relevant when the policy maker has the ability to partially affect the distribution of treatment intensity. For example, if it can set a minimum level of treatment. In this case, the variation in treatment intensity is correlated with the unobservables.

This paper belongs to the literature of the estimation of unconditional counterfactual distributions. Machado and Mata (2005) and Melly (2006) proposed estimators of such counterfactuals based on QR when the treatment is exogenous, which Chernozhukov et al. (2013) generalized by proposing a method to estimate any functional of interest, given an initial estimator of the conditional quantile curve or the conditional distribution function.[1] Another closely related paper was Martinez-Sanchis et al. (2012), who adapted Melly (2006) to an endogenous setting using a control function approach. The methods proposed in this paper are different. In particular, they extend Chernozhukov et al. (2013) methods to the presence of endogeneity based on an instrumental variables approach, similarly to Pereda-Fernández (2010). Moreover, I consider different counterfactuals that cannot be consistently estimated by Martinez-Sanchis et al. (2012).

Several empirical works could fit into the framework presented in this paper, such as the impact of education on earnings (Card, 2001), and on adult mortality (Lleras-Muney, 2005), the effect of family income on scholastic achievement (Dahl and Lochner, 2012), the impact of class size on scholastic achievement (Angrist and Lavy, 1999), or on long-term outcomes (Fredriksson et al., 2013), the quality of institutions on income (Acemoglu et al., 2012), or the effect of smoking during pregnancy on child's birthweight (Evans and Ringel, 1999). These studies could benefit from studying the distributional effects of an intervention that results in a different assignment of the treatment intensity for the whole population.

I illustrate these methods using data on birth weight of children whose mothers smoked

[1] Other related works include Firpo et al. (2009), who proposed an estimator based on reweighting of the influence function to estimate distributional effects under exogeneity, or Frölich and Melly (2013), who proposed a nonparametric estimator of the unconditional quantile treatment effect for the subpopulation of compliers with an endogenous, binary treatment.

during pregnancy, focusing on differences for white and black women. Following Evans and Ringel (1999), I instrument smoke consumption during pregnancy with the state tax as a percentage of the final price. First, I estimate the distributional effect of smoking an extra daily cigarette, as well as the MTE. Then I carry out three counterfactuals in which I respectively swap the copula for each group, I reduce the amount of cigarettes smoked to one half of what was actually reported, and I limit the maximum daily amount of smoked cigarettes to ten. The results show that such reductions in smoke consumption during pregnancy by white mothers would increase the average birth weight. However, this effect would be heterogeneous, and in particular it would substantially reduce the number of newborns with low birth weight, *i.e.*, those who weigh less than 2,500 grams at birth.

The rest of the paper is organized as follows. In Section 2 I describe the framework and discuss the identification of the functionals of interest. In Section 3 I propose two estimation methods based on different assumptions of the copula. In Section 4 I apply the methodology presented in this paper to the estimation of the effect of smoking during pregnancy on birth weight. Finally, Section 5 concludes. All proofs are shown in Appendix A.

2 Framework

Consider a triangular system of equations conformed by the following observable variables: Y denotes the outcome variable of interest, $X \equiv (X_1\ X_2')'$ be the vector composed of the continuous treatment, X_1, and a set of exogenous covariates, X_2, and $Z \equiv (Z_1\ X_2')'$ be the vector composed of the instrumental variable, Z_1 and the exogenous covariates. Moreover the population can be split into a finite number of groups. Denote the group to which each individual belongs by D, such that $D \in \mathcal{D} = \left\{1, ..., \overline{D}\right\}$, where $\overline{D} < \infty$ is the number of groups. The leading case are two groups, *i.e.*, $\mathcal{D} = \{1, 2\}$. This system is completed by the unobserved random variables U_d and V, to which I refer as the *conditional ranks*, whose conditional distributions given X_2 are uniform.[2] Both the outcome and treatment equations

[2]Because these variables are unidimensional, the amount of heterogeneity of the model is restricted. In particular, it rules out models that are not monotonic on the unobservables, such as random coefficients

are nonseparable functions. Formally, for $D = d$,[3]

$$Y = g_d\left(X_1, X_2, U_d\right) \tag{1}$$

$$X_1 = k_d\left(Z_1, X_2, V\right) \tag{2}$$

where $g_d\left(\cdot,\cdot,\cdot\right)$ and $k_d\left(\cdot,\cdot,\cdot\right)$ are nonseparable and strictly increasing in their last argument. k_d represents the conditional quantile function (CQF) of the treatment X_1, which satisfies $\mathbb{P}\left(X_1 \leq k_d\left(Z_1, X_2, \tau\right)|Z_1, X_2, D = d\right) = \tau$, and g_d is the structural quantile function (SQF) of Y, which satisfies $\mathbb{P}\left(Y \leq g_d\left(X_1, X_2, \tau\right)|Z_1, X_2, D = d\right) = \tau$. The latter does not coincide with the CQF because of the endogeneity of the treatment (Chernozhukov and Hansen, 2013).

The endogeneity of this model stems from the relation between the two unobservables. Their joint distribution is established by the following assumption:

Assumption 1. $(U_d, V) \perp Z_1|X_2, \forall d \in \mathcal{D}$. *Moreover, the joint distribution of the unobservables, conditional on the covariates, is given by* $U_d, V|X_2 \sim C^d_{UV|X_2} \forall d \in \mathcal{D}$.

$C^d_{UV|X_2}$ is the conditional copula of (U_d, V) for group d.[4] This may depend on X_2, allowing the amount endogeneity to vary across individuals with different characteristics. For example, the endogeneity between income and schooling could vary with parental income, since those students whose parents have low income may not be able to afford high levels of schooling. Assumption 1 implies that, conditional on the covariates, the ranks are independent of the instrument. This is the exclusion restriction of the model.

models. Unfortunately, these models are not nonparametrically identified (Hahn and Ridder, 2011; Kasy, 2011; Hoderlein et al., 2017; Masten, 2018) and they would further complicate the present analysis. Deriving set-identification results with multidimensional unobserved heterogeneity is beyond the scope of this paper.

[3]This system uses the Skorohod representation, which states that a random variable φ_i can be written in terms of its quantile function: $\varphi_i = Q\left(U_i\right)$, where $U_i \sim U\left(0,1\right)$.

[4]By definition, a copula is the multivariate distribution of $(U_1, ..., U_m)$ such that their marginal distributions are uniformly distributed on the unit interval. Sklar (1959) showed that any multivariate distribution of the continuously distributed variables $X_1, ..., X_m$, with respective marginal cdfs $F_1\left(x_1\right), ..., F_m\left(x_m\right)$, there exists a unique cdf C, such that $\mathbb{P}\left(X_1 \leq x_1, ..., X_m \leq x_m\right) = C\left(F_1\left(x_1\right), ..., F_m\left(x_m\right)\right)$. The conditional copula is defined as $C\left(F_1\left(x_1\right), ..., F_{\overline{m}-1}\left(x_{\overline{m}-1}\right), F_{\overline{m}+1}\left(x_{\overline{m}+1}\right), ..., F_m\left(x_m\right)|F_{\overline{m}}\left(x_{\overline{m}}\right)\right) = \frac{\partial}{\partial F_{\overline{m}}\left(x_{\overline{m}}\right)} C\left(F_1\left(x_1\right), ..., F_m\left(x_m\right)\right)$. Lastly, the copula density is defined as $c\left(F_1\left(x_1\right)\right), ..., F_m\left(F_m\left(x_m\right)\right) = \frac{\partial^m}{\partial F_1\left(x_1\right)\cdot ... \cdot \partial F_m\left(x_m\right)} C\left(F_1\left(x_1\right), ..., F_m\left(x_m\right)\right)$.

The copula of the conditional ranks, despite not being the objective function of the policy maker, is nevertheless informative, as it can help identifying who would be more affected by a change in treatment intensity. To see this, consider the MTE, given by[5]

$$\mathbb{E}\left[\frac{\partial g_1\left(X_1, X_2, U_d\right)}{\partial X_1}|X_1 = x_1, X_2 = x_2, V = v, D = d\right] = \int_0^1 \nabla_1 g_d\left(x_1, x_2, u\right) c^d_{U|VX_2}\left(u|v, x_2\right) du \tag{3}$$

where ∇_1 denotes the derivative with respect to its first argument, and $C^d_{U|VX_2}$ denotes the copula conditional on V, X_2, and $D = d$.

Hence, those individuals with a copula that has higher density at values of U_d for which the marginal gain from increasing the treatment is largest, would benefit more from the treatment intensity increase. *E.g.*, assume that $\nabla_1 g_d\left(x_1, x_2, u\right) = u$ and $c^d_{UV|X_2}\left(u, v|x_2\right) = 2 + 4uv - 2u - 2v$.[6] Under exogeneity, the MTE would be equal to $1/2$ for everyone, but under endogeneity Equation 3 would equal $(1+v)/3$. Hence, individuals with high values of v would have a larger than average MTE and would benefit more from a treatment increase. As a result, the copula allows to identify which individuals benefit the most, so that the intervention can be targeted to maximize its impact.

2.1 Counterfactuals of interest

The first type of counterfactuals consists in swapping of any of the four structural functions that determine the outcome by the counterparts of another group.[7] Namely, the SQF of the outcome, the CQF of the treatment, the copula, and the distribution of the observables, $F^d_Z\left(z\right)$. These counterfactuals may focus both on the mean value and on the unconditional distribution of the outcome variable. Mathematically, the counterfactual distribution when

[5] The relation of the copula and the MTE can be immediately generalized to the case of multidimensional unobserved heterogeneity. For example, assume that it depends on $\{U_m\}^M_{m=1}$. Define the multidimensional copula by $C_{UV|X_2}\left(U_1, ..., U_M, V|x_2\right)$. Then, the MTE would be calculated as in Equation 3 by integrating the derivative of g_d with respect to the multidimensional copula, holding V constant.

[6] This is a particular case of the Bernstein copula.

[7] Note that in many circumstances the ability of the policy maker to implement such counterfactuals may be limited, making them informative of such potential effects theoretically.

one combines the SQF of group p, the CQF of group q, the copula of group r, and the distribution of the observables of group s is given by

$$F_Y^m(y) = \int_{\mathcal{Z}} \int_{[0,1]^2} \mathbf{1}\left(g_p\left(k_q\left(z_1, x_2, v\right), x_2, u\right) \leq y\right) dC^r_{UV|X_2}\left(u, v|x_2\right) dF^s_Z\left(z\right) \tag{4}$$

where $\mathbf{1}(\cdot)$ denotes the indicator function, $\mathcal{Z}$ denotes the support of Z, and $m \equiv (p, q, r, s)$ is used for brevity. Note that this type of counterfactual nests the type considered by Chernozhukov et al. (2013) in an exogenous setting, which involve changing the SQF or the distribution of the unobservables.

Alternatively, one could consider a counterfactual that involves a structural change in the determination of the treatment. In particular, let the counterfactual treatment equal $k_d^{cf}(z, v) \equiv \psi\left(k_d(z, v)\right)$, rather than by Equation 2. These counterfactuals may be more pertinent when the policy maker has the ability to partially enforce the distribution of the treatment in certain ways. To illustrate them, consider the following examples of counterfactuals:

Example 1. *Imagine a policy maker interested in increasing worker's income, Y, through an increase in the minimum level of compulsory schooling, $\underline{X}_1$. In this counterfactual scenario, those who would attain a level of education above that minimum without the intervention ($X_1 > \underline{X}_1$) would have the same level of education, but there would be an increase for those below it. Since students with higher levels of unobserved ability, U_d, tend to study more years, the increase in income for students below the threshold would be smaller than if those above it increased their education level.*

Example 2. *Consider a policy maker who wants to estimate the effect that halving the number of daily smoked cigarettes during pregnancy, X_1, would have on the birth weight of newborns, Y. Since smoking could be correlated with other unobserved bad habits, U_d, such a policy would have a different effect on women who smoked a different amount of cigarettes, which could increase the birthweight of those who would have had a low birth weight ($Y \leq 2,500$) by less than the rest of the children.*

In these examples, the respective counterfactual treatments would be given by $k_d^{cf}(z,v) = \max(k_d(z,v), \underline{x})$, for some new minimum schooling age $\underline{x}$, and $k_d^{cf}(z,v) = \frac{1}{2}k_d(z,v)$.

Remark 1. *Note that the latest type of counterfactuals does not need the population to be split into groups. However, because the group structure is used in decompositions, I maintain it for both counterfactuals.*

2.2 Identification

To discuss identification, note that the actual conditional distribution of the outcome can be expressed as a function of the SQF of Y, the CQF of X_1, and the copula of (U_d, V):

$$
\begin{aligned}
F^d_{Y|Z}(y|z) &= \int_{[0,1]^2} \mathbf{1}\left(g_d\left(k_d\left(z_1, x_2, v\right), x_2, u\right) \leq y\right) dC^d_{UV|X_2}(u, v|x_2) \\
&= \int_{[0,1]^2} \mathbf{1}\left(u \leq g_d^{-1}\left(k_d\left(z_1, x_2, v\right), x_2, y\right)\right) dC^d_{U|VX_2}(u|v, x_2)\, dv \\
&= \int_0^1 C^d_{U|VX_2}\left(g_d^{-1}\left(k_d\left(z_1, x_2, v\right), x_2, y\right)|v, x_2\right) dv \qquad (5)
\end{aligned}
$$

for $d \in \mathcal{D}$. Therefore, once these three functions are identified, the actual conditional distribution is also identified. To do so, let the following assumption hold:

Assumption 2. *Let* $\mathcal{X}_d^{\dagger}\left(z^a, z^b\right) \equiv \left\{x_1 \in \mathcal{X}^d_{z^a} \cap \mathcal{X}^d_{z^b} : F^d_{X_1|Z}(x_1|z^a) = F^d_{X_1|Z}\left(x_1|z^b\right)\right\}$, *where* $\mathcal{X}^d_{z^a} \equiv supp\left(X_1|Z = z, D = d\right)$, *and* $z^c = (z_1^c, x_2)$ *for* $c = \{a, b\}$. $\mathcal{X}_d^{\dagger}\left(z^a, z^b\right)$ *is nonempty and finite* $\forall d \in \mathcal{D}$.

This assumption is needed for the identification of the SQF of Y (Torgovitsky, 2015), and it holds even if the support of the instrument is binary. For the specific case considered in this paper in the estimation, *i.e.*, a linear quantile regression, it suffices to have some quantile v such that the coefficient at that quantile equals zero.

The following proposition establishes the identification of the structural functions:

Proposition 1. *Let Assumptions 1-2 hold. Then, for* $d \in \mathcal{D}$, *the structural functions* g_d, k_d, *and* $C_{UV|X_2}$ *are identified.*

The identification of g_d and k_d are based on the conditions spelled by Torgovitsky (2015) and Matzkin (2003), respectively. The latter is well established, whereas the identification of g_d has received a lot of attention in the literature.[8]

Consequently, the copula is the remaining structural function that needs to be identified. This is achieved by inverting the SQF and the CQF, which is possible by the continuity and the monotonicity of both functions in their last argument. To better understand how the instrument allows to identify the copula, notice that, conditional on (V, Z), there is a bijection between Y and U_d. This is given by $F^d_{Y|ZV}(y|z,v) = C^d_{U|VX}\left(g_d^{-1}(k_d(z_1,x_2,v),x_2,y)|v,x_2\right)$. Therefore, variations in Z induce a variation in the distribution of Y through a change in the treatment, without affecting the unobservables. To illustrate this relation, consider the exogenous case. If U_d and V were independent of each other, conditional on Z, the conditional copula would simplify to $C^d_{U|VX_2}(u|v,x_2) = u$. Moreover, by Equation 2, when Z and V are known, so is X_1. Thus, it follows that $F^d_{Y|ZV}(y|z,v) = g_d^{-1}(k_d(z_1,x_2,v),x_2,y) = g_d^{-1}(x_1,x_2,y) \equiv u$, whereas under endogeneity $F^d_{Y|ZV}(y|z,v) \neq u$.

3 Estimation

3.1 Baseline Estimator

Let the following assumptions hold:

Assumption 3.

$$g_d(X_1,X_2,U_d) = X'\beta_d(U_d)$$

$$k_d(Z_1,X_2,V) = Z'\gamma_d(V)$$

[8] Chesher (2003) and Imbens and Newey (2009), who studied the nonparametric identification of nonseparable models using a control function approach. Other papers proposed semiparametric methods, which do not suffer from the curse of dimensionality, such as Jun (2009), or Lee (2007) who assumes the model to be separable. Alternatively, Ma and Koenker (2006) proposed a parametric model of Chesher (2003). Another recent paper is D'Haultfœuille and Février (2015), which established the identification based on similar conditions to those in Torgovitsky (2015). On the other hand, (Hoderlein and Mammen, 2007) discussed the identification without monotonicity.

where $\beta_d(\cdot)$ and $\gamma_d(\cdot)$ are continuous, and $g_d(\cdot,\cdot,\cdot)$ and $k_d(\cdot,\cdot,\cdot)$ are strictly increasing in their last argument $\forall d \in \mathcal{D}$.

Assumption 4. *$(Y_i, X_{1i}, X_{2i}', Z_{1i}, D_i)'$ are iid for $i = 1, ..., n$, defined on the probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, and take values in a compact set.*

Assumption 5. *Denote the sample size for the d-th group by n_d, and the total sample size by $n = \sum_d n_d$. $n/n_d \xrightarrow{P} \lambda_d \in [1, \infty)\, \forall d \in \mathcal{D}$ as $n \to \infty$.*

Assumption 6. *Y and X_1 have conditional density that is bounded from above and away from zero, a.s. on compact sets $\mathcal{Y}$ and $\mathcal{X}_1$, respectively. The copula $C^d_{UV|X_2}(u, v|x_2)$ is uniformly continuous and differentiable with respect to its arguments a.e.*

Assumption 7. *For all $(\tau, \nu) \in \mathcal{U} \times \mathcal{U}$, $\left(\beta_d(\tau)', \gamma_d(\nu)'\right)' \in int\mathcal{B} \times \mathcal{G}$, where $\mathcal{B} \times \mathcal{G}$ is compact and convex. $\mathcal{U} = [\epsilon, 1-\epsilon]$, for some small ϵ.*[9]

Assumption 8.

$$\Pi_d(\beta, \iota, \gamma, \tau, \nu) \equiv \mathbb{E}\begin{bmatrix} \left(\tau - \mathbf{1}\left(Y < X'\beta + \Phi_d(\tau, Z)'\iota\right)\right)\Psi_D(\tau, Z)\mathbf{1}(D = d) \\ (\nu - \mathbf{1}(X_1 < Z'\gamma))Z\mathbf{1}(D = d) \end{bmatrix}$$

$$\Pi_d(\beta, \gamma, \tau, \nu) \equiv \mathbb{E}\begin{bmatrix} (\tau - \mathbf{1}(Y < X'\beta))\Psi_D(\tau, Z)\mathbf{1}(D = d) \\ (\nu - \mathbf{1}(X_1 < Z'\gamma))Z\mathbf{1}(D = d) \end{bmatrix}$$

where $\Psi_d(\tau, Z) \equiv \left[\Phi_d(\tau, Z)', X_2'\right]'$, $\Phi_d(\tau, Z)$ is a vector of transformation of instruments, Jacobian matrices $\frac{\partial}{\partial(\beta',\gamma')}\Pi_d(\beta, \gamma, \tau, \nu)$ and $\frac{\partial}{\partial\left(\beta_2',\iota',\gamma'\right)}\Pi_d(\beta, \iota, \gamma, \tau, \nu)$ are continuous and have full rank, uniformly over $\mathcal{B} \times \mathcal{I} \times \mathcal{G} \times \mathcal{U} \times \mathcal{U}$, and the image of $\mathcal{B} \times \mathcal{G}$ under the mapping $(\beta, \gamma) \mapsto \Pi_d(\beta, \gamma, \tau, \nu)$ is simply-connected.

Assumption 9. *$wp \to 1$, there exists some function $\hat{\Phi}_d(\tau, z) \in \mathcal{F}_Z$ and $\hat{\Phi}_d(\tau, z) \xrightarrow{p} \Phi_d(\tau, z)$ uniformly in (τ, z) over compact sets, where $\Phi_d(\tau, z) \in \mathcal{F}_Z$. $\mathcal{F}_Z$ is the class of uniformly smooth functions in z with the uniform smoothness order $\omega > \dim\left((d, x_1, z')'\right)/2$. Moreover, for $f(\tau, z) \in \mathcal{F}$, $\|f(\tilde{\tau}, z) - f(\tau, z)\| < A|\tau - \tilde{\tau}|^a$, $A > 0, a > 0$, for all $(z, \tau, \tilde{\tau})$.*

[9]This constant serves to avoid the estimation of extreme quantiles. See, *e.g.*, Chernozhukov and Hansen (2005).

Assumption 10. $C^d_{UV|X_2}(U_d, V|x_2) = C^d_{UV|X_2}(U_d, V|x_2; \theta_d)$ *(i) is known up to the vector of parameters* $\theta_d \in int(\Theta)$*, where* Θ *is bounded, and its dimension,* d_{θ_d} *is finite; (ii) its pdf,* $c^d_{UV|X_2}(U_d, V|x_2; \theta_d)$*, is three times continuously differentiable with respect to* θ *and continuously differentiable with respect to* (u, v)*; (iii)* $\forall \tilde{\theta}_d \neq \theta_d$*,* $c^d_{UV|X_2}\left(U_d, V|x_2; \tilde{\theta}_d\right) \neq c^d_{UV|X_2}(U_d, V|x_2; \theta_d)$ *a.s.*

Assumption 3 imposes linearity on the two quantile processes of the triangular system, which greatly simplifies the computation of estimator. Specifically, $\beta_d(\cdot)$ can be estimated by IVQR and $\gamma_d(\cdot)$ by QR. This assumption is more restrictive than other alternatives, such as a partially linear model (Lee, 2003) or a generalized linear model with a known link function (Horowitz et al., 2004), although such flexible approaches are subject to the curse of dimensionality.[10] Moreover, imposing linearity on both equations allows the estimator of the counterfactual distribution to be asymptotically linear, thus attaining the $\sqrt{n}$ convergence rate. Regardless, note that IVQR does not require linearity of the first stage equation (Chernozhukov and Hansen, 2005).[11]

Assumptions 4 and 5 describe the sampling process, whereas Assumptions 6 to 9 are regularity conditions needed for the asymptotic Gaussianity of the IVQR and QR estimators. In Chernozhukov and Hansen (2006), $\hat{\Phi}_d(\tau, z)$ is a linear projection of X_1 on Z. Additionally, Assumption 6 rules out perfect correlation, as the copula does not have a well-defined density. Lastly, Assumption 10 is a parametric assumption of the copula that ensures its identification, and allows us to obtain the joint asymptotic distribution of the SQF and the copula estimators.

Note that these assumptions differ from those considered by other authors in a similar setting, such as Lee (2007) or Martinez-Sanchis et al. (2012). A comparison with the estimators proposed in those papers is considered in Appendix D.

Denote the IVQR and QR estimators by $\hat{\beta}_d(\cdot)$ and $\hat{\gamma}_d(\cdot)$, respectively. These are used to

[10] Additionally, the linear specification requires regressors to take either positive or negative values, but not both, as that would make the process non-monotonic. See Koenker (2005) for further details.

[11] I discuss the properties of the estimator when this assumption is relaxed in Appendix C.

obtain the fitted values of individual pairs of unobservables, given by:

$$\hat{U}_{d,i} = \epsilon + \int_{\epsilon}^{1-\epsilon} \mathbf{1}\left(X_i'\hat{\beta}_d(u) \leq Y_i\right) du$$
$$\hat{V}_i = \epsilon + \int_{\epsilon}^{1-\epsilon} \mathbf{1}\left(Z_i'\hat{\gamma}_d(v) \leq X_{1i}\right) dv$$

Define $\ell_{d,i}(u, v, \theta) \equiv \log\left(c^d_{UV|X_2}(u, v|X_{2i}; \theta)\right)\mathbf{1}(D_i = d)$. The estimator of the copula parameters, $\hat{\theta}_d$, is given by

$$\begin{aligned}\hat{\theta}_d &= \arg\max_{\theta} \frac{1}{n_d}\sum_{i=1}^{n} \ell_{d,i}\left(\hat{U}_{d,i}, \hat{V}_i, \theta\right) \\ &= \arg\max_{\theta} \frac{1}{n_d}\sum_{i=1}^{n}\left(\ell_{d,i}(U_{d,i}, V_i, \theta) + \log\left(\frac{c^d_{UV|X_2}\left(\hat{U}_{d,i}, \hat{V}_i|X_{2i}; \theta\right)}{c^d_{UV|X_2}(U_{d,i}, V_i|X_{2i}; \theta)}\right)\mathbf{1}(D_i = d)\right) \qquad (6)\end{aligned}$$

The first term in Equation 6 is the log likelihood function. However, because the actual values of the copula are not observed, the objective function depends on a second term that accounts for the estimated conditional ranks. The estimator of the copula is obtained by replacing θ_d with $\hat{\theta}_d$: $\hat{C}^d_{UV|X_2}(u, v|x_2) = C_{UV|X_2}\left(u, v|x_2; \hat{\theta}_d\right)$. Additionally, the estimator of the counterfactual SQF, conditional on Z and V, is given by

$$\hat{g}_p\left(\hat{k}_q(z_1, x_2, v), x_2, u\right) = \left(z'\hat{\gamma}_q(v), x_2\right)'\hat{\beta}_p(u) \qquad (7)$$

3.2 Counterfactual Estimators

The first kind of counterfactual involves combining several of the structural functions from different groups. The counterfactual conditional (on $Z = z$) distribution of Y is given by

$$\begin{aligned}\hat{F}^m_{Y|Z}(y|z) &= \epsilon + \int_{\mathcal{U}^2} \mathbf{1}\left(\left(z'\hat{\gamma}_q(v), x_2\right)'\hat{\beta}_p(u) \leq y\right) dC^r_{UV|X_2}\left(u, v|x_2; \hat{\theta}_r\right) \\ &= \epsilon + \int_{\epsilon}^{1-\epsilon} C^r_{U|VX_2}\left(\hat{g}_p^{-1}\left(z'\hat{\gamma}_q(v), x_2, y\right)|v, x_2; \hat{\theta}_r\right) dv \qquad (8)\end{aligned}$$

where $\hat{g}_p^{-1}(x_1, x_2, y)$ denotes the inverse with respect to its third argument.

Define $\mathcal{F}$ to be the class of measurable functions that includes $\left\{F_{Y|Z}(y|z) : y \in \mathcal{Y}, z \in \mathcal{Z}\right\}$ as well as the indicators of all the rectangles in $\overline{\mathbb{R}}^{d_z}$, such that $\mathcal{F}$ is totally bounded under the metric $\zeta\left(f, \tilde{f}\right) = \left[\int \left(f - \tilde{f}\right)^2 dF_Z\right]^{1/2}$; denote the space of real-valued bounded functions defined on the index set by the supremum norm by ℓ^∞; let $\mathcal{M} \equiv \{(p, q, r, s) \in \mathcal{D}^4\}$. The joint distribution of this estimator and of the vector Z, $F_Z^m(z)$, is established in the following proposition:[12]

Proposition 2. *Let Assumptions 3 to 10 hold. The joint asymptotic distribution of* $\hat{F}^m_{Y|Z}(y|z)$ *and* $\hat{F}^s_Z(z)$ *is given by*

$$\sqrt{n}\begin{pmatrix} \hat{F}^m_{Y|Z}(y|z) - F^{m,\epsilon}_{Y|Z}(y|z) \\ \int_{\mathcal{Z}} f d\left(\hat{F}^s_Z(z) - F^s_Z(z)\right)\end{pmatrix} \rightsquigarrow \begin{pmatrix} \mathbb{Z}_{F_m|Z}(y, z) \\ \sqrt{\lambda_s}\mathbb{Z}_{Z_s}(f)\end{pmatrix}$$

a stochastic process in metric space $\ell^\infty(\mathcal{YZFM})$*, where* $\mathbb{Z}_{F_m|Z}(y, z)$ *and* $\mathbb{Z}_{Z_s}(f)$ *are zero-mean tight Gaussian processes, defined in Appendix A, with a.s. uniformly continuous paths in* $\mathcal{YZFM}$.

Remark 2. *Note that the estimator* $\hat{F}^m_{Y|Z}(y|z)$ *converges in distribution to* $F^{m,\epsilon}_{Y|Z}(y|z) \equiv \epsilon + \int_{\mathcal{U}^2} \mathbf{1}(g_p(k_q(z_1, x_2, v), x_2, u) \leq y)\, dC^r_{UV|X_2}(u, v|x_2)$*, the counterpart of the conditional distribution with trimmed extreme quantiles. Hence,* $\hat{F}^m_{Y|Z}(y|z)$ *is a slightly biased estimator of* $F^m_{Y|Z}(y|z)$.

The estimator of $F^m_Y(y)$ is the sample analog of Equation 4:

$$\begin{aligned}\hat{F}^m_Y(y) &= \epsilon + \frac{1}{n_s}\sum_{i=1}^{n}\int_{\mathcal{U}^2} \mathbf{1}\left((Z_i'\hat{\gamma}_q(v), X_{2i})'\hat{\beta}_p(u) \leq y\right) dC^r_{UV|X_2}\left(u, v|X_{2i}; \hat{\theta}_r\right)\mathbf{1}(D_i = s) \\ &= \epsilon + \frac{1}{n_s}\sum_{i=1}^{n}\int_{\epsilon}^{1-\epsilon} C^r_{U|VX_2}\left(\hat{g}_p^{-1}(Z_i'\hat{\gamma}_q(v), X_{2i}, y)\,|v, X_{2i}; \hat{\theta}_r\right) dv\mathbf{1}(D_i = s) \qquad (9)\end{aligned}$$

The estimator of the unconditional quantile function is then calculated as:

$$\hat{Q}^m_Y(\tau) = \inf\left\{y : \tau \leq \hat{F}^m_Y(y)\right\} \qquad (10)$$

[12] With some slight abuse of notation, the metric spaces used for the convergence omit the spaces of group indicators. They should be implicitly included throughout the paper.

The following theorem characterizes their asymptotic distribution:[13]

Theorem 1. *Let Assumptions 3 to 10 hold. The asymptotic distribution of* $\hat{Q}_Y^m(\tau)$ *is given by*

$$\sqrt{n}\left(\hat{Q}_Y^m(\tau) - Q_Y^{m,\epsilon}(\tau)\right) \rightsquigarrow \mathbb{Z}_{Q_m}(\tau)$$

a stochastic process in metric space $\ell^\infty(\mathcal{TM})$, $\mathbb{Z}_{Q_m}(\tau) \equiv -\left(f_Y^m\left(Q_Y^m(\tau)\right)\right)^{-1}\mathbb{Z}_{F_m}\left(Q_Y^m(\tau)\right)$ *is a zero-mean tight Gaussian process with a.s. uniformly continuous paths in* $\mathcal{TM}$, *where* $\mathcal{T}$ *is a closed interval such that* $\mathcal{T} \subset (0,1)$*. Moreover, the asymptotic distribution of* $\hat{F}_Y^m(y)$*is given by*

$$\sqrt{n}\left(\hat{F}_Y^m(y) - F_Y^{m,\epsilon}(y)\right) \rightsquigarrow \mathbb{Z}_{F_m}(y)$$

a stochastic process in metric space $\ell^\infty(\mathcal{YM})$*, and where* $\mathbb{Z}_{F_m}(y) \equiv \int_{\mathcal{Z}} \mathbb{Z}_{F_m|Z}(y,z)\, dF_Z^s(z) + \sqrt{\lambda_s}\mathbb{Z}_{Z_s}\left(F_{Y|Z}^m(y|z)\right)$ *is a zero-mean Gaussian tight process, defined in Appendix A, with a.s. uniformly continuous paths in* $\mathcal{YM}$.

Remark 3. *As it was the case for Proposition 2, the estimators of the unconditional quantile and the unconditional cdf converge in distribution to their trimmed counterparts, defined by* $F_Y^{m,\epsilon}(y) \equiv \int_{\mathcal{Z}} F_{Y|Z}^{m,\epsilon}(y|z)\, dF_Z(z)$*, and* $Q_Y^{m,\epsilon}(\tau) \equiv \inf\left\{y : \tau \leq F_Y^{m,\epsilon}(y)\right\}$*. Consequently, they are slightly biased for the non-trimmed functions.*

Note that the estimator of the distribution can be used to estimate any other function that depends on it by plugging it in as in Chernozhukov et al. (2013).

The previous result would hold true if one used some alternative estimators of the SQF and the conditional copula. As long as the resulting estimator of the conditional cdf satisfies a Gaussian law as in Proposition 2, Theorem 1 would apply. For example, one could use methods not based on quantile regression to estimate either Equation 1 or 2.

[13]The finite sample performance is shown in a Monte Carlo exercise in Appendix E.

The second type of counterfactuals directly changes the structural determination of the treatment. The counterfactual conditional distribution for individuals in group d is given by

$$\hat{F}^{d,cf}_{Y|Z}(y|z) = \epsilon + \int_{\epsilon}^{1-\epsilon} C^{d}_{U|VX_2}\left(\hat{g}_d^{-1}\left(\hat{k}^{cf}_d(v), x_2, y\right)|v, x_2; \hat{\theta}_d\right) dv \tag{11}$$

where $\hat{k}^{cf}_d(v) \equiv \psi(z'\hat{\gamma}_d(v))$. Denote the set of uniformly continuous functions mapping a set $\mathcal{W}$ to the real line by $\mathcal{C}(\mathcal{W})$. The counterfactual unconditional cdf, $\hat{F}^{d,cf}_{Y}(y)$, and quantile functions, $\hat{Q}^{d,cf}_{Y}(\tau)$, are computed as in Equations 9-10, but using $\hat{F}^{d,cf}_{Y|Z}(y|z)$ instead of $\hat{F}^{m}_{Y|Z}(y|z)$. To derive the asymptotic properties of this type of counterfactual estimator, I impose the following regularity assumption:

Assumption 11. *The map $\psi(k_d(z,v))$ is Hadamard directionally differentiable at $k_d(z,v) \in \mathcal{C}(\mathcal{ZU})$ tangentially to a set $\mathcal{C}(\mathcal{ZU})$ with Hadamard directional derivative $\psi'_{k_d}: \mathcal{C}(\mathcal{ZU}) \mapsto \mathcal{C}(\mathcal{ZU})$.*

This assumption ensures that it is possible to apply the functional delta method to obtain the asymptotic distribution of this counterfactual estimator. It is satisfied by a large class of counterfactuals, including the two examples presented in Section 2, for which the Hadamard derivative equals $\psi'_{k_d}(h) = \max(h,0)\,\mathbf{1}(k_d(z,v) = \underline{x}) + h\mathbf{1}(k_d(z,v) > \underline{x})$ and $\psi'_{k_d}(h) = \frac{1}{2}h$, respectively. This assumption is weaker than Hadamard differentiability, which is not satisfied by the counterfactual of the first example.[14]

The asymptotic distribution of this second type of counterfactuals is a nonlinear function of a Gaussian process. It is established by the following proposition, which is based on the results in Fang and Santos (2019):[15]

Proposition 3. *Let Assumptions 3 to 11 hold. The joint asymptotic distribution of $\hat{F}^{d,cf}_{Y|Z}(y|z)$*

[14] See Fang and Santos (2019) for further details.

[15] The asymptotic distribution of the counterfactual unconditional distribution and quantile function is obtained analogously to that of the first kind of counterfactual, shown in Theorem 1. Its proof is omitted for the sake of brevity.

and $\hat{F}_Z^d(z)$, *and those of* $\hat{F}_Y^{d,cf}(y)$ $\hat{Q}_Y^{d,cf}(\tau)$ *are given by*

$$\sqrt{n}\begin{pmatrix}\hat{F}_{Y|Z}^{d,cf}(y|z) - F_{Y|Z}^{d,cf,\epsilon}(y|z) \\ \int_{\mathcal{Z}} fd\left(\hat{F}_Z^d(z) - F_Z^d(z)\right)\end{pmatrix} \rightsquigarrow \begin{pmatrix}\mathbb{Z}_{F_d^{cf}|Z}(y,z) \\ \sqrt{\lambda_d}\mathbb{Z}_{Z_d}(f)\end{pmatrix}$$

$$\sqrt{n}\left(\hat{F}_Y^{d,cf}(y) - F_Y^{d,cf,\epsilon}(y)\right) \rightsquigarrow \mathbb{Z}_{F_d^{cf}}(y)$$

$$\sqrt{n}\left(\hat{Q}_Y^{d,cf}(\tau) - Q_Y^{d,cf,\epsilon}(\tau)\right) \rightsquigarrow \mathbb{Z}_{Q_d^{cf}}(y)$$

which are stochastic processes in metric spaces $\ell^\infty(\mathcal{YZFD})$, $\ell^\infty(\mathcal{YD})$, *and* $\ell^\infty(\mathcal{TD})$, *respectively, where* $\mathbb{Z}_{F_d^{cf}|Z}(y,z)$, $\mathbb{Z}_{Z_d}(f)$, $\mathbb{Z}_{F_d^{cf}}(y)$, *and* $\mathbb{Z}_{Q_d^{cf}}(\tau)$ *are zero-mean tight processes, defined in Appendix A, with a.s. uniformly continuous paths in* $\mathcal{YZFD}$, $\mathcal{YD}$, *and* $\mathcal{TD}$, *respectively.*

Remark 4. *Analogously to the first type of counterfactuals, these are centered around a slightly biased version of the conditional cdf, unconditional cdf, and unconditional quantile function, which are respectively given by* $F_{Y|Z}^{d,cf,\epsilon}(y|z) \equiv \epsilon + \int_{\mathcal{U}^2} \mathbf{1}\left(g_p\left(k_d^f(z,v), x_2, u\right) \leq y\right) \cdot dC_{UV|X_2}^d(u,v|x_2)$, $F_Y^{d,cf,\epsilon}(y) \equiv \int_{\mathcal{Z}} F_{Y|Z}^{d,cf,\epsilon}(y|z)\, dF_Z(z)$, *and* $Q_Y^{d,cf,\epsilon}(\tau) \equiv \inf\left\{y : \tau \leq F_Y^{d,cf,\epsilon}(y)\right\}$.

3.3 Nonparametric Estimator of the Copula

Assumption 10 is convenient because it reduces the estimation of the copula to a finite set of parameters. However, because the copula is nonparametrically identified, one can consider a nonparametric estimator of the copula by relaxing the parametric assumption:

Assumption 12. *The copula of the unobservables, conditional on* X_2, *does not depend on* X_2 $\forall d \in \mathcal{D}$, *i.e.,* $C_{UV|X_2}^d(u,v|x_2) = C_{UV}^d(u,v)$.

In words, the copula is left unspecified, but it is the same for all values of X_2. This could be further relaxed, but the estimator of the copula would then converge at a rate slower than $\sqrt{n}$. The nonparametric estimator of the copula equals the empirical copula based on the estimated values of the conditional ranks:

$$\check{C}_{UV}^d(u,v) \equiv \frac{1}{n_d}\Sigma_{j=1}^n \mathbf{1}\left(\hat{U}_{d,j} \leq u\right)\mathbf{1}\left(\hat{V}_j \leq v\right)\mathbf{1}(D_j = d) \tag{12}$$

The estimator of the counterfactual distribution of Y, conditional on Z, is obtained by plugging-in the nonparametric estimator of the copula:

$$\begin{aligned}\check{F}^{m}_{Y|Z}(y|z) &\equiv \epsilon + \int_{\mathcal{U}^2} \mathbf{1}\left((z'\hat{\gamma}_q(v), x_2)' \hat{\beta}_p(u) \leq y\right) d\check{C}^{r}_{UV}(u,v) \\ &= \epsilon + \frac{1}{n_r}\Sigma_{j=1}^{n} \mathbf{1}\left(\left(Z_j'\hat{\gamma}_q\left(\hat{V}_j\right), X_{2j}\right)' \hat{\beta}_p\left(\hat{U}_{r,j}\right) \leq y\right) \mathbf{1}(D_j = r) \end{aligned} \quad (13)$$

As it was the case for the counterfactuals with a parametric copula, the counterfactual unconditional cdf, $\check{F}^{d}_{Y}(y)$, and quantile functions, $\check{Q}^{d}_{Y}(\tau)$, with the nonparametric estimator of the copula are computed like in Equations 9-10, but using $\check{F}^{m}_{Y|Z}(y|z)$ instead of $\hat{F}^{m}_{Y|Z}(y|z)$.

The uniform convergence of $\check{C}^{r}_{UV}(u,v)$ can be shown to be at a rate $\sqrt{n}$, so that $\check{F}^{m}_{Y|Z}(y|z)$, $\check{F}^{d}_{Y}(y)$, and $\check{Q}^{d}_{Y}(\tau)$ are indeed uniformly consistent at that rate. This is established by the following proposition:

Proposition 4. *Let Assumptions 3 to 9, and 12 hold. Then,*

$$\begin{aligned}\sup_{y,z,m} \sqrt{n}\left|\check{F}^{m}_{Y|Z}(y|z) - F^{m,\epsilon}_{Y|Z}(y|z)\right| &= O_p(1) \\ \sup_{y,m} \sqrt{n}\left|\check{F}^{m}_{Y}(y) - F^{m,\epsilon}_{Y}(y)\right| &= O_p(1) \\ \sup_{\tau,m} \sqrt{n}\left|\check{Q}^{m}_{Y}(\tau) - Q^{m,\epsilon}_{Y}(\tau)\right| &= O_p(1)\end{aligned}$$

Note, however, that it is not possible to obtain the asymptotic Gaussianity by the usual arguments: because of the nonlinearity of the indicator function, it is not possible to apply the extended continuous mapping theorem. This is a consequence of the conditional ranks not being observed. A way to overcome this problem would be to regularize the estimator by using a smooth function that converges uniformly to the indicator function.[16]

[16] Even for such estimator it would not be possible to establish the asymptotic normality as in Theorem 1, since the estimator of the conditional copula converges at a rate slower than $\sqrt{n}$. See Appendix B.7 for further details.

3.4 Inference

Because the asymptotic covariance matrix of the counterfactual estimators are complex and depend on several density functions and integrals, direct estimation is impractical. Therefore, I propose to carry out inference using the weighted bootstrap (Ma and Kosorok, 2005). The following assumption describes the properties of the weights used by this bootstrap method:

Assumption 13. *Let W_i be an iid sample of positive weights, such that $\mathbb{E}(W_i) = 1$, $Var(W_i) = \omega_0 > 0$, and is independent of $(Y_i, D_i, Z_i')'$ for $i = 1, ..., n$.*

These weights are used for the estimation of all the structural functions required for the counterfactual distributions.[17] Specifically, $\hat{\gamma}_d^*(u)$ and $\hat{\beta}_d^*(u)$ are given by the weighted QR and IVQR estimators, *i.e.*

$$\hat{\gamma}_d^*(v) = \arg\min_{\gamma \in \Gamma} \sum_{i=1}^{n} W_i \rho_v (X_{1i} - Z_i'\gamma) \mathbf{1}(D_i = d)$$
$$\hat{\beta}_d^*(u) = \left(\hat{\beta}_{d,1}^*(u), \hat{\beta}_{d,2}^*\left(\hat{\beta}_{d,1}^*(u), u\right)\right)$$

where

$$\rho_\tau(u) \equiv (\tau - \mathbf{1}(u < 0)) u$$
$$\hat{\beta}_{d,1}^*(u) = \arg\min_{\beta_1 \in \mathcal{B}_1} \|\hat{\iota}_d^*(\beta_1, u)\|_{A_d^*(u)}$$
$$\left(\hat{\beta}_{d,2}^*(\beta_1, u), \hat{\iota}^*(\beta_1, u)\right) = \arg\min_{(\beta_2, \iota) \in \mathcal{B}_2 \times \mathcal{I}} \frac{1}{n_d} \sum_{i=1}^{n} W_i \rho_u \left(Y_i - X_i'\beta - \hat{\Phi}_i^*(u)'\iota\right) \mathbf{1}(D_i = d)$$
$$A_d^*(u) = \frac{1}{n_d} \sum_{i=1}^{n} \hat{\Phi}_{d,i}^*(u) \hat{\Phi}_{d,i}^*(u)' \mathbf{1}(D_i = d)$$

and $\hat{\Phi}_{d,i}^*(u)$ is the weighted counterpart of $\hat{\Phi}(u)$. The copula parameter is given by

$$\hat{\theta}_d^* = \arg\max_{\theta} \frac{1}{n_d} \sum_{i=1}^{n} W_i \ell_{d,i}\left(\hat{U}_{d,i}^*, \hat{V}_i^*, \theta\right)$$

where $\left(\hat{U}_{d,i}^*, \hat{V}_i^*\right)$ are computed like $\left(\hat{U}_{d,i}, \hat{V}_i\right)$, but substituting $\hat{\beta}_d(\cdot)$ and $\hat{\gamma}_d(\cdot)$ by $\hat{\beta}_d^*(\cdot)$

[17] Note that the weight for each individual is the same in every step.

and $\hat{\gamma}_d^*(\cdot)$, respectively. These estimates are then combined to obtain the bootstrapped counterfactual distributions:

$$\hat{F}_{Y|Z}^{m,*}(y|z) = \epsilon + \int_{\epsilon}^{1-\epsilon} C_{U|VX_2}^{r}\left(\hat{g}_p^{*,-1}\left(z'\hat{\gamma}_q^*(v), x_2, y\right)|v, x_2; \hat{\theta}_r^*\right) dv \tag{14}$$

$$\hat{F}_{Y}^{m,*}(y) = \epsilon + \frac{1}{n_s}\sum_{i=1}^{n}\int_{\epsilon}^{1-\epsilon} W_i C_{U|VX_2}^{r}\left(\hat{g}_p^{*,-1}\left(Z_i'\hat{\gamma}_q^*(v), X_{2i}, y\right)|v, X_{2i}; \hat{\theta}_r^*\right) dv\mathbf{1}(D_i = s) \tag{15}$$

$$\hat{Q}_{Y}^{m,*}(\tau) = \inf\left\{y : \tau \leq \hat{F}_{Y}^{m,*}(y)\right\} \tag{16}$$

where $\hat{g}_d^*(x_1, x_2, u) = x'\hat{\beta}_d^*(u)$ and $\hat{g}_d^{*,-1}$ denotes its inverse with respect to the third element. The asymptotic validity of these estimates is established in the following theorem:[18]

Theorem 2. *Under Assumptions 3-10, and 13, the weighted bootstrap estimators* $\hat{F}_{Y|Z}^{m,*}(y|z)$, $\hat{F}_{Y}^{m,*}(y)$, *and* $\hat{Q}_{Y}^{m,*}(\tau)$ *consistently estimate the limiting laws of* $\hat{F}_{Y|Z}^{m}(y|z)$, $\hat{F}_{Y}^{m}(y)$, *and* $\hat{Q}_{Y}^{m}(\tau)$. *Moreover,*

$$\sqrt{\frac{n}{\omega_0}}\left(\hat{F}_{Y|Z}^{m,*}(y|z) - \hat{F}_{Y|Z}^{m}(y|z)\right) \rightsquigarrow \mathbb{Z}_{F_m|Z}(y, z)$$

$$\sqrt{\frac{n}{\omega_0}}\left(\hat{F}_{Y}^{m,*}(y) - \hat{F}_{Y}^{m}(y)\right) \rightsquigarrow \mathbb{Z}_{F_m}(y)$$

$$\sqrt{\frac{n}{\omega_0}}\left(\hat{Q}_{Y}^{*,m}(\tau) - \hat{Q}_{Y}^{m}(\tau)\right) \rightsquigarrow \mathbb{Z}_{Q_m}(\tau)$$

which are stochastic processes in metric spaces $\ell^\infty(\mathcal{YZM})$, $\ell^\infty(\mathcal{YM})$, *and* $\ell^\infty(\mathcal{TM})$, *respectively.*

The estimation of the asymptotic variance of these estimators is based on Algorithm 3 in Chernozhukov et al. (2013):[19]

1. For each repetition $t = 1, ..., T$, compute the bootstrapped estimator of the desired estimator, *e.g.*, $\hat{F}_{Y,t}^{m,*}(y)$.

[18] The bootstrap validity for the second kind of counterfactual estimators is analogously proved and therefore omitted.

[19] Note that the bootstrapped covariance matrix requires additional conditions to be valid. See Kato (2011) for further details.

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

It is also possible to construct Kolmogorov-Smirnov test statistics to carry out uniformly valid inference using these bootstrap estimators. See, *e.g.*, Pereda-Fernández (2023).

Remark 5. *The same bootstrap algorithm can be used to compute the confidence intervals of the counterfactuals when they are Hadamard differentiable. If, in contrast, the counterfactual is Hadamard directionally differentiable, but not Hadamard differentiable, this bootstrap estimator is not valid. In principle, one could apply the consistent alternative proposed by Fang and Santos (2019). However, in the current case it would not be applicable in practice because the derivative function would depend on several density functions and integrals that would need to be estimated. Alternatively, one could obtain the intervals by first obtaining the intervals of the primitive parameters, and then compute the counterfactuals for the draws that lie inside the confidence region for these parameters.*[20] *This may also be infeasible due to the sheer number of parameters computed with the methods presented here.*

4 Empirical Application

To illustrate the methods presented in this paper, I consider the estimation of the effect of smoking during pregnancy on child weight. Specifically, I explore some counterfactual scenarios to estimate the proportion of newborns with *Low Birth Weight* (LBW, less than 2,500 grams), since newborns falling into this category have a higher chance of developing several medical conditions later in life, including cognitive development and chronic diseases (Case and Paxson, 2009; Almond and Currie, 2011).

I combine two datasets: the 1990 Natality Data from the National Vital Statistics System of the National Center for Health Statistics, which records every birth in the United States

[20] See, *e.g.*, Woutersen and Ham (2013).

during 1990, and the Tax Burden of Tobacco, which includes the percentage of state taxes over the final price of cigarettes. This variable is the instrument I use.[21]

I separately run the regressions for white and black mothers. The covariates I include in the regression are a quadratic polynomial of the mother's age, the number of years of education, the number of gestation weeks, and dummy variables for marital status (1 if married), and the sex of the newborn (1 if female). I restrict the data to firstborn children of mothers aged 18-35 who smoked during pregnancy.[22] Moreover, I exclude multiple births, since they have a higher chance of having a lower birth weight. This leaves us with 144,478 births, of which over 90% correspond to white mothers. Regarding the copulas, I consider three parametric copulas (Gaussian, Clayton and Frank), as well as the nonparametric estimator of the copula.

4.1 Estimation of the Structural Functions

Table 1 presents some descriptive statistics of the variables used. Children of white mothers weight on average over 200 grams more than their black counterparts. In addition, white mothers tend to live in states with a slightly larger tax on tobacco. Despite this, daily cigarette consumption is larger for white mothers.

Figures 1-2 show the IVQR estimates for white and black mothers, which display a fair amount of heterogeneity. Most importantly, note that the coefficients for smoking are substantially negative and significant for white mothers, as common wisdom would suggest. For them, the daily consumption of an extra cigarette is associated with a weight decrease of the baby of between 10 to 60 grams, depending on the quantile.[23] On the other hand, for black mothers it would be associated with a decrease of 30 grams, up to an increase of at most 65 grams. In contrast, the effect for black mothers is positive at most quantiles. However, this effect is very imprecisely estimated, so it is not a significant effect. This counterintuitive

[21] For a more detailed description of the datasets, see Evans and Ringel (1999).

[22] The age variable is introduced in the regressions as actual age minus 18.

[23] Note that the extreme quantiles take values outside this range. Because these are less precisely estimated, I do not comment them in the text.

Table 1: Descriptive statistics

	White	Black	Difference
Birth weight	3,208.8	2,976.7	232.1
	(1.45)	(4.98)	(5.19)
Mean number of daily cigarettes smoked	12.16	8.81	3.35
	(0.02)	(0.06)	(0.06)
State tax as a percentage of price	0.19	0.17	0.02
	(0.00)	(0.00)	(0.00)
Age	23.54	24.02	-0.48
	(0.01)	(0.04)	(0.04)
Years of Education	11.98	12.02	-0.04
	(0.00)	(0.01)	(0.02)
Married	0.60	0.18	0.41
	(0.00)	(0.00)	(0.00)
Gestation weeks	39.55	38.77	0.79
	(0.01)	(0.02)	(0.03)
Female baby	0.49	0.49	0.00
	(0.00)	(0.00)	(0.00)
N	132.029	12.449	

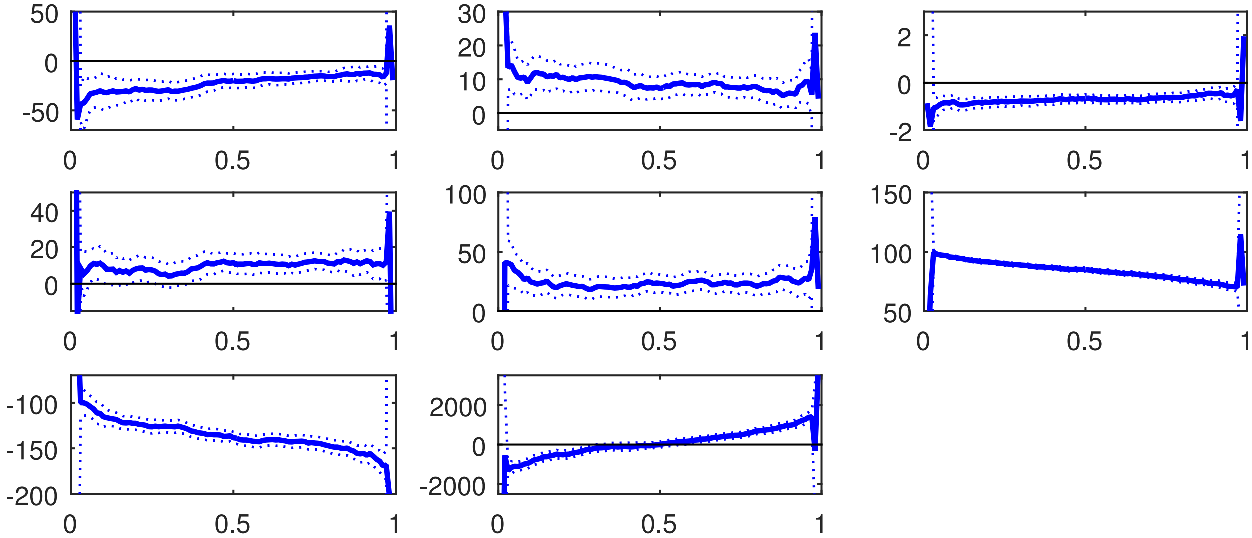
Note: Standard errors in parentheses.

finding could be the result of some model misspecification or that the instrument is not valid.

Most of the remaining coefficients have the same sign for mothers of both races, with the exception of the age polynomial. Education, being married, and the number of gestation weeks have all positive coefficients, whereas female babies weighed over 100 grams less than their male counterparts for the majority of the population. The difference of these coefficients between black and white mothers are smaller and, in most cases, not significant.

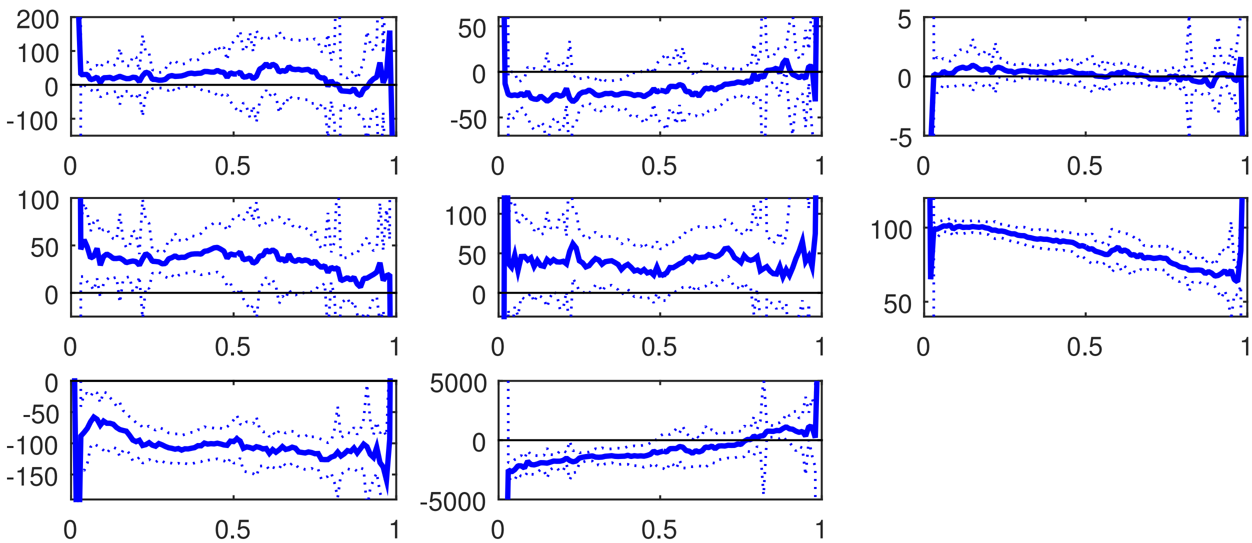
To compare the estimates of the copula parameters, I report the value of the Kendall's τ statistics for each them in Table 2. For white mothers, the estimated parameter has a correlation statistic of around 0.14-0.15, consistently for all estimators. In contrast, the correlation statistics for black mothers are negative, with the exception of the Clayton copula. The latter follows because this copula can only take positive correlation values. For the remaining copulas, the correlation statistic is around 0.21-0.24. An interpretation of this result is that white mothers who smoke more are those whose children weight more at birth conditional on their observed characteristics, whereas the opposite is true for black mothers.

Figure 1: IVQR estimates (white mothers)



Notes: the solid lines denote the IVQR estimates, and the dashed lines denote the bootstrapped 95% uniform confidence bands. From left to right, the IVQR estimates shown are: in the first row, cigarettes consumption, mother's age, and mother's age to the square; in the second row, years of education, married, and gestation weeks; in the third row, sex of the baby, and the intercept.

Figure 2: IVQR estimates (black mothers)



Notes: the solid lines denote the IVQR estimates, and the dashed lines denote the bootstrapped 95% uniform confidence bands. From left to right, the IVQR estimates shown are: in the first row, cigarettes consumption, mother's age, and mother's age to the square; in the second row, years of education, married, and gestation weeks; in the third row, sex of the baby, and the intercept.

Thus, smoking habits of pregnant women and birth weight of their children are related by potentially different reasons for women of each race.

Table 2: Kendall's τ statistics of estimated copula parameters

	Gaussian	Clayton	Frank	Gumbel
White	0.148	0.135	0.144	0.156
	(0.007)	(0.006)	(0.007)	(0.007)
Black	-0.221	0.005	-0.214	-0.243
	(0.038)	(0.000)	(0.023)	(0.042)

Note: Bootstrapped standard errors in parentheses.

To assess the fit of the copulas to the data, I report the values of the likelihood with the estimated copulas in Table 3. In addition, I report two measures of fit of the estimated unconditional distribution of the outcome: the mean and maximum difference between the estimated distributions and the empirical cdf. The results show that the parametric copula that attained the maximum likelihood for white mothers was the Gaussian, although the two measures suggest that the other two parametric distributions led to a similar fit. On the other hand, the estimates for black mothers show a more differentiated level of fit: because the estimated copula is negative, the Clayton estimate is much worse than the others, and in this case the maximum likelihood is attained for the Frank copula. Still, the measures of fit for the Frank and Gaussian copulas are similar. Note also that the fit for the nonparametric copulas are much better than for the parametric ones.

4.2 Estimation of the Actual Distributions

The estimates of the unconditional distribution for the four specifications considered are presented in Table 4. Those with the parametric copula for white mothers are almost indistinguishable from each other, whereas those with the nonparametric copula are slightly different, particularly on the left tail of the distribution. As such, the former estimate roughly a 10% of LBW babies, whereas the latter the estimate is slightly below 9%. For black mothers, the estimate with the Clayton copula is much different from the others, owing to the fact that its correlation cannot be negative. On the other hand, the estimate with the

Table 3: Fit of the copulas

		Gaussian	Clayton	Frank	Nonparametric
White	$\int_{\mathcal{Y}} \left\| Q_{0.5}\left(\hat{F}_Y(y)\right) - F_Y(y) \right\| dy$	0.0172	0.0171	0.0175	0.0097
	$\sup_y \left\| \hat{F}_Y(y) - F_Y(y) \right\|$	0.0264	0.0266	0.0265	0.0224
	$\mathcal{L}$	3515.6	3185.7	3418.8	
Black	$\int_{\mathcal{Y}} \left\| Q_{0.5}\left(\hat{F}_Y(y)\right) - F_Y(y) \right\| dy$	0.0217	0.0288	0.0225	0.0101
	$\sup_y \left\| \hat{F}_Y(y) - F_Y(y) \right\|$	0.0431	0.0547	0.0430	0.0258
	$\mathcal{L}$	745.9	-29.6	799.0	

Notes: The first and fourth rows represent the integral of the difference between the median across repetitions of the estimated counterfactual cdf and the true cdf; the second and fifth rows represents the maximum of this difference; the third and sixth rows represent the estimated likelihood function.

nonparametric copula is slightly lower than those with the other parametric copulas on the left tail, and slightly higher on the right tail. The estimated fractions of LBW babies range between 18% (nonparametric copula) and 19.5% (Gaussian and Frank copulas).

4.3 Estimation of the Counterfactual Distributions

To assess the impact of smoking on birth weight, I consider three types of counterfactuals. First, I swap the copula between mothers of both races. This way, one can assess how differences in birth weight can be linked to differences in the amount of self-selection into smoking, *i.e.*, how smoking is correlated to other unobserved factors that may affect birth weight. Second, I reduce smoking intensity to each mother by half. Because this is correlated to other unobserved factors, the change in birth weight is heterogeneous. Finally, I limit the daily amount of smoked cigarettes to 10, which reduces smoking only for heavier smokers.[24]

The counterfactual estimates with the Gaussian and nonparametric copulas and the differences with respect to the estimates of the actual distributions are respectively shown in Tables 5 and 6.[25] The first counterfactual results in an increase of LWB. For white mothers,

[24] As stated in the text, this counterfactual is not Hadamard differentiable, so the bootstrap does not consistently approximate the asymptotic distribution of the estimator. Regardless, because the alternative method reported in Remark 5 is infeasible, I report these inconsistent estimates.

[25] I only report the estimates with these two copulas because the Frank and Clayton copulas either yield very similar results or fit the actual data much worse. The results with them are available upon request.

Table 4: Actual unconditional distribution estimates

		2,500	3,000	3,500	4,000	4,500
White	Gaussian	0.099	0.325	0.710	0.936	0.971
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	Clayton	0.098	0.328	0.710	0.933	0.970
		(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
	Frank	0.100	0.323	0.710	0.935	0.970
		(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
	Nonparametric	0.087	0.336	0.717	0.933	0.972
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Black	Gaussian	0.195	0.507	0.804	0.958	0.976
		(0.003)	(0.004)	(0.006)	(0.004)	(0.004)
	Clayton	0.210	0.509	0.793	0.946	0.970
		(0.008)	(0.003)	(0.011)	(0.006)	(0.004)
	Frank	0.195	0.507	0.804	0.956	0.974
		(0.004)	(0.004)	(0.006)	(0.005)	(0.004)
	Nonparametric	0.181	0.490	0.827	0.962	0.983
		(0.001)	(0.002)	(0.002)	(0.003)	(0.003)

Note: Bootstrapped standard errors in parentheses.

this increase ranges between 3.5 and 4.3 percentage points, whereas for black mothers, it ranges between 2.4 and 4.5 percentage points. In both cases, the estimate with the parametric copula is smaller than with the nonparametric copula. The reason for these unusual changes are the fact that the estimates of the copula and the effect of smoking on birth weight have the reverse sign for both groups. Thus, while we estimate that white mothers who tend to smoke more are those whose babies would be heavier, a reversal in the amount of self-selection would lead to a weight increase to babies in the right tail of the distribution and a decrease for those on the left tail.

The second counterfactual reports more meaningful estimates for white women: halving smoking intensity would reduce the fraction of LWB by over 4 percentage points, regardless of the copula employed in the estimation. For the estimate with the nonparametric copula it represents over a half in the incidence of LWB. This counterfactual would also have an impact on other parts of the distribution (Figure 3). For instance, the fraction of babies weighing less than 3,000 grams would decrease by over 10 percentage points, to about 22%.

Table 5: Counterfactual unconditional distribution estimates

			2,500	3,000	3,500	4,000	4,500
CF 1	White	Gaussian	0.134	0.349	0.691	0.923	0.971
			(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
		Nonparametric	0.130	0.353	0.680	0.937	0.975
			(0.004)	(0.005)	(0.005)	(0.004)	(0.003)
	Black	Gaussian	0.219	0.511	0.788	0.940	0.966
			(0.010)	(0.003)	(0.013)	(0.008)	(0.005)
		Nonparametric	0.226	0.505	0.785	0.942	0.971
			(0.010)	(0.002)	(0.014)	(0.007)	(0.005)
CF 2	White	Gaussian	0.057	0.223	0.631	0.916	0.970
			(0.001)	(0.004)	(0.004)	(0.002)	(0.002)
		Nonparametric	0.042	0.227	0.637	0.915	0.972
			(0.001)	(0.003)	(0.003)	(0.002)	(0.002)
	Black	Gaussian	0.250	0.600	0.833	0.968	0.980
			(0.016)	(0.020)	(0.011)	(0.006)	(0.004)
		Nonparametric	0.239	0.595	0.850	0.975	0.991
			(0.013)	(0.021)	(0.011)	(0.004)	(0.004)
CF 3	White	Gaussian	0.069	0.264	0.665	0.923	0.970
			(0.001)	(0.003)	(0.002)	(0.001)	(0.002)
		Nonparametric	0.062	0.272	0.668	0.921	0.972
			(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
	Black	Gaussian	0.215	0.538	0.823	0.963	0.977
			(0.007)	(0.006)	(0.005)	(0.004)	(0.004)
		Nonparametric	0.206	0.525	0.839	0.966	0.984
			(0.006)	(0.005)	(0.004)	(0.003)	(0.003)

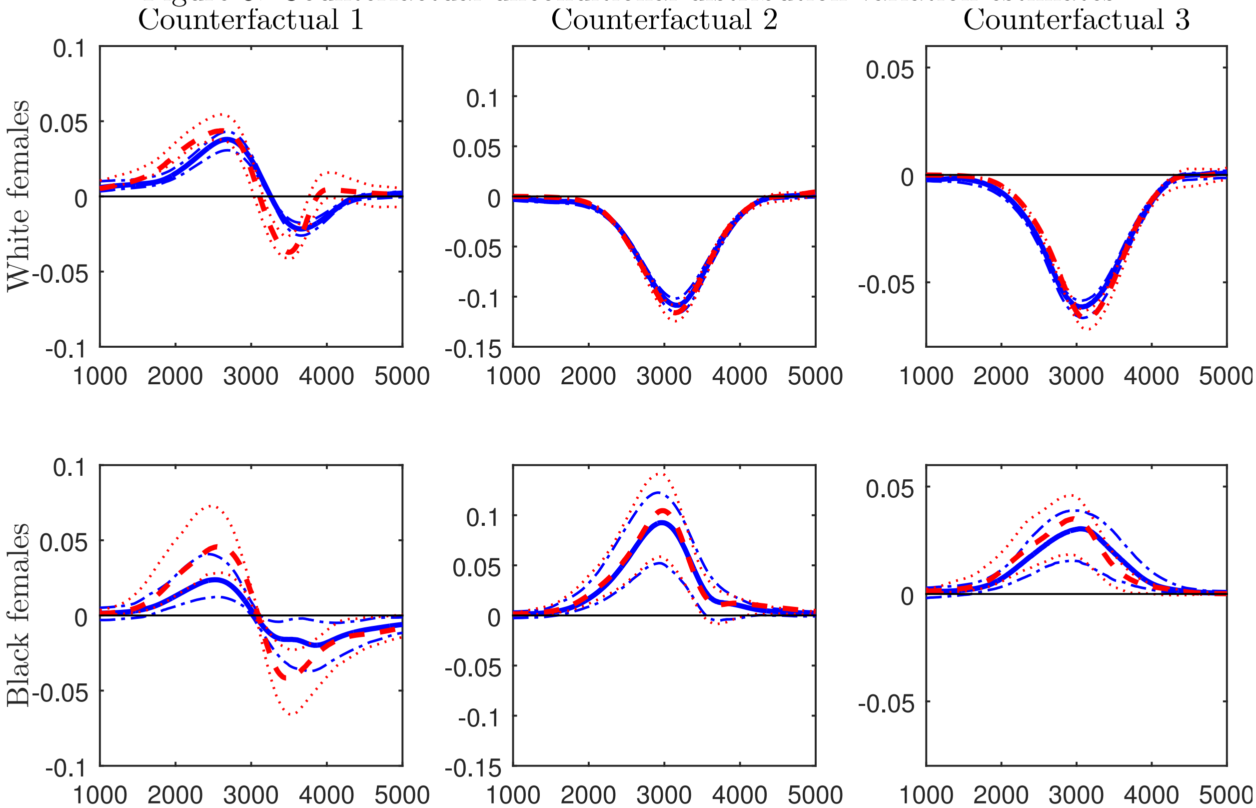
Note: Bootstrapped standard errors in parentheses.

Table 6: Counterfactual unconditional distribution variation estimates

			2,500	3,000	3,500	4,000	4,500
CF 1	White	Gaussian	0.035	0.025	-0.018	-0.013	0.000
			(0.003)	(0.002)	(0.002)	(0.001)	(0.001)
		Nonparametric	0.043	0.017	-0.037	0.005	0.003
			(0.004)	(0.004)	(0.005)	(0.004)	(0.003)
	Black	Gaussian	0.024	0.004	-0.016	-0.018	-0.010
			(0.008)	(0.002)	(0.009)	(0.007)	(0.004)
		Nonparametric	0.045	0.015	-0.042	-0.020	-0.012
			(0.011)	(0.002)	(0.013)	(0.008)	(0.005)
CF 2	White	Gaussian	-0.042	-0.102	-0.079	-0.020	0.000
			(0.001)	(0.002)	(0.003)	(0.001)	(0.001)
		Nonparametric	-0.044	-0.109	-0.080	-0.018	0.000
			(0.001)	(0.003)	(0.003)	(0.002)	(0.002)
	Black	Gaussian	0.055	0.092	0.029	0.010	0.004
			(0.014)	(0.020)	(0.013)	(0.004)	(0.002)
		Nonparametric	0.058	0.104	0.024	0.012	0.008
			(0.014)	(0.023)	(0.011)	(0.005)	(0.001)
CF 3	White	Gaussian	-0.030	-0.061	-0.045	-0.012	0.000
			(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
		Nonparametric	-0.025	-0.064	-0.049	-0.012	0.000
			(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
	Black	Gaussian	0.021	0.030	0.019	0.005	0.001
			(0.005)	(0.006)	(0.006)	(0.002)	(0.000)
		Nonparametric	0.025	0.035	0.013	0.003	0.001
			(0.006)	(0.007)	(0.004)	(0.001)	(0.000)

Note: Bootstrapped standard errors in parentheses.

Figure 3: Counterfactual unconditional distribution variation estimates



Notes: The solid thick blue lines denote the estimates with the parametric copula, the dashed-dotted thin blue lines denote their 95% confidence intervals, the dashed thick red lines denote the estimates with the nonparametric copula, and the dotted thin red lines denote their 95% confidence intervals; uniform confidence intervals computed using the multiplicative bootstrap.

On the other hand, the counterfactual effect on black mothers would yield a generalized decrease in birth weight, and the proportion of LBW would increase up to one quarter of all births.

Finally, the third counterfactual would decrease smoking for heavier smokers, which for white women implies a larger increase in birth weight for babies with relatively low weight. This is confirmed by the estimates in Table 6, and the proportion of LWB would decrease by 2.5 to 3 percentage points, whereas those weighing less than 3,000 grams would decrease by over 6 percentage points. On the other hand, the variation for black females would be the opposite, and the fraction of LBW would increase by over 2.5 percentage points.

4.4 Estimation of the MTE

The last estimand of interest is the MTE, which captures the variation in LBW induced by a marginal increase in smoking intensity for a person with unobserved characteristic V. The estimates are reported in Table 7, and they show the opposite behavior for white and black mothers. For the former, the effect is negative throughout the entire distribution of V and increasing. As such, mothers more likely to smoke heavily (high V) would have a smaller reduction on their child's weight from smoking less than those more likely to smoke little (low V). On the other hand, the MTE estimates for black women are positive and increasing, so heavier smoker mothers would have a higher decrease in their child's weight from smoking less.

Table 7: MTE estimates

	Quantile				
	0.1	0.25	0.5	0.75	0.9
White	-10.697	-9.779	-8.775	-7.802	-6.960
	(0.788)	(0.718)	(0.726)	(0.644)	(0.612)
Black	6.657	8.270	10.172	12.126	13.878
	(2.460)	(2.939)	(3.637)	(4.458)	(5.093)

Note: Bootstrapped standard errors in parentheses.

5 Conclusions

In this paper I propose an estimator of counterfactual unconditional distribution functions in the presence of an endogenous continuous treatment with heterogeneous effects. This estimator is based on the estimators of the quantile processes that characterize a triangular system of equations, and the estimator of the distribution of the copula of the conditional ranks, which captures the endogeneity of the treatment. The latter is nonparametrically identified by inverting the quantile processes of the triangular system, and it can be estimated either parametrically, resulting in an estimator that is asymptotically Gaussian with the usual $\sqrt{n}$ convergence rate, or nonparametrically using the empirical cdf of the estimated values of the copula of the conditional ranks. The counterfactuals I consider involve either a change of one of the structural functions of the individuals of one group with that of another group, or a change in the intensity of the treatment that is not independent of the unobservables.

As an empirical application I estimate the effect of birth smoking of newborns' birth weight, and I carry out three counterfactuals in which I respectively swap the copula between mothers of both races, reduce the number of smoked cigarettes during pregnancy to one half of the actual quantity, and I limit their consumption to a maximum of ten per day. The second and third counterfactuals would increase the birth weight of newborns, though this effect would be heterogeneous, and would substantially reduce the percentage of newborns with low birth weight.

References

Acemoglu, D., S. Johnson, and J. A. Robinson (2012). The colonial origins of comparative development: An empirical investigation: Reply. *The American Economic Review 102*(6), 3077–3110.

Almond, D. and J. Currie (2011). Killing me softly: The fetal origins hypothesis. *The Journal of Economic Perspectives 25*(3), 153–172.

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

Angrist, J. D. and V. Lavy (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics 114*(2), 533–575.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica 69*(5), 1127–1160.

Case, A. and C. Paxson (2009). Early life health and cognitive function in old age. *The American economic review 99*(2), 104.

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

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

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

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

Chernozhukov, V. and C. Hansen (2013). Quantile models with endogeneity. *Annual Review of Economics 5*, 57–81.

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

Dahl, G. B. and L. Lochner (2012). The impact of family income on child achievement: Evidence from the earned income tax credit. *The American Economic Review 102*(5), 1927–1956.

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

Einmahl, U., D. M. Mason, et al. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *The Annals of Statistics 33*(3), 1380–1403.

Evans, W. N. and J. S. Ringel (1999). Can higher cigarette taxes improve birth outcomes? *Journal of public Economics 72*(1), 135–154.

Fang, Z. and A. Santos (2019). Inference on directionally differentiable functions. *The Review of Economic Studies 86*(1), 377–412.

Firpo, S., N. M. Fortin, and T. Lemieux (2009). Unconditional quantile regressions. *Econometrica 77*(3), 953–973.

Florens, J.-P., J. J. Heckman, C. Meghir, and E. Vytlacil (2008). Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects. *Econometrica 76*(5), 1191–1206.

Fredriksson, P., B. Ockert, and H. Oosterbeek (2013). Long-term effects of class size. *The Quarterly Journal of Economics 128*(1), 249–285.

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

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

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

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

Hoderlein, S. and E. Mammen (2007). Identification of marginal effects in nonseparable models without monotonicity. *Econometrica 75*(5), 1513–1518.

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

Ichimura, H. and W. K. Newey (2022). The influence function of semiparametric estimators. *Quantitative Economics 13*(1), 29–61.

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.

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

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

Koenker, R. (2005). *Quantile regression*, Volume 38 of *Econometric Society Monograph Series*. Cambridge University Press.

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

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

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

Lleras-Muney, A. (2005). The relationship between education and adult mortality in the united states. *The Review of Economic Studies 72*(1), 189–221.

Ma, L. and R. Koenker (2006). Quantile regression methods for recursive structural equation models. *Journal of Econometrics 134*(2), 471–506.

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

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

Martinez-Sanchis, E., J. Mora, and I. Kandemir (2012). Counterfactual distributions of wages via quantile regression with endogeneity. *Computational Statistics & Data Analysis 56*(11), 3212–3229.

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

Matzkin, R. L. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica 71*(5), 1339–1375.

Melly, B. (2006). Estimation of counterfactual distributions using quantile regression. *Review of Labor Economics 68*(4), 543–572.

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

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

Pereda-Fernández, S. (2010). Quantile regression discontinuity: Estimating the effect of class size on scholastic achievement. Technical report, Master Thesis CEMFI No. 1002.

Pereda-Fernández, S. (2023). Identification and estimation of triangular models with a binary treatment. *Journal of Econometrics 234*(2), 585–623.

Pratt, J. W. (1960). On interchanging limits and integrals. *The Annals of Mathematical Statistics 31*(1), 74–77.

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.

Woutersen, T. and J. C. Ham (2013). Calculating confidence intervals for continuous and discontinuous functions of parameters. Technical report, cemmap working paper.

Let $E \equiv (Y, X_1, X_2, Z_1, D)$. The following notation is used throughout the appendix:[26]

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

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

$$\hat{R}_m(E,\beta,\iota,\gamma,\tau,\nu) \equiv \begin{bmatrix} \varphi_\tau\left(Y - X'\beta - \hat{\Phi}(\tau)'\iota\right)\hat{\Psi}_p(\tau)\mathbf{1}(D=p) \\ \varphi_\nu(X_1 - Z'\gamma)Z\mathbf{1}(D=q) \end{bmatrix}$$

$$R_m(E,\beta,\iota,\gamma,\tau,\nu) \equiv \begin{bmatrix} \varphi_\tau\left(Y - X'\beta - \Phi(\tau)'\iota\right)\Psi_p(\tau)\mathbf{1}(D=p) \\ \varphi_\nu(X_1 - Z'\gamma)Z\mathbf{1}(D=q) \end{bmatrix}$$

$$\hat{G}_m(E,\beta,\iota,\gamma,\tau,\nu) \equiv \begin{bmatrix} \rho_\tau\left(Y - X'\beta - \hat{\Phi}(\tau)'\iota\right)\mathbf{1}(D=p) \\ \rho_\nu(X_1 - Z'\gamma)\mathbf{1}(D=q) \end{bmatrix}$$

$$G_m(E,\beta,\iota,\gamma,\tau,\nu) \equiv \begin{bmatrix} \rho_\tau\left(Y - X'\beta - \Phi(\tau)'\iota\right)\mathbf{1}(D=p) \\ \rho_\nu(X_1 - Z'\gamma)\mathbf{1}(D=q) \end{bmatrix}$$

$$Q_{m,n}(\beta,\iota,\gamma,\tau,\nu) \equiv \mathbb{E}_n\left[\hat{G}_m(E,\beta,\iota,\gamma,\tau,\nu)\right]$$

$$Q_m(\beta,\iota,\gamma,\tau,\nu) \equiv \mathbb{E}\left[G_m(E,\beta,\iota,\gamma,\tau,\nu)\right]$$

$\varepsilon_d = Y - X'\beta_d$, $\varepsilon_d(\tau) = Y - X'\beta_d(\tau)$, $\hat{\varepsilon}_d(\tau) = Y - X'\hat{\beta}_d(\tau)$, $\eta_d = X_1 - Z'\gamma_d$, $\eta_d(\nu) = X_1 - Z'\gamma_d(\nu)$, $\hat{\eta}_d(\nu) = X_1 - Z'\hat{\gamma}_d(\nu)$, $\Psi_d(\tau) \equiv \left(\Phi_d(\tau)', X_2\right)'$, $\hat{\Psi}_d(\tau) \equiv \left(\hat{\Phi}_d(\tau)', X_2\right)'$, $\Phi_d(\tau) \equiv \Phi_d(\tau, Z)$, $\hat{\Phi}_d(\tau) \equiv \hat{\Phi}_d(\tau, Z)$, $\upsilon_m(\tau,\nu) = \left(\beta_{p,2}(\tau)', \iota_p(\tau), \gamma_q(\nu)'\right)'$, $\hat{\upsilon}_m(\tau,\nu) = \left(\hat{\beta}_{p,2}(\tau)', \hat{\iota}_p(\tau), \hat{\gamma}_q(\nu)'\right)'$, and $\varphi_\tau(u) \equiv (\mathbf{1}(u<0) - \tau)$.

A Mathematical Proofs

This appendix collects the proofs of the propositions and theorems stated in the text. These are in turn based on some lemmas, which are stated and proved in Appendix B.

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

A.1 Proof of Proposition 1

The proof is split into three steps. First, I show the identification of k_d; second, I show the identification of g_d; finally, I show the identification of $C^d_{UV|X_2}$.

The identification of k_d follows Matzkin (2003). By definition, $k_d(\cdot,\cdot,\cdot)$ is continuous in all its arguments and strictly increasing in its third. Therefore, it can be inverted:

$$k^I_d(z_1,x_2,y) \equiv \inf\{v : x_1 \geq k_d(z_1,x_2,v)\}$$

Let g be a strictly increasing function and define $\tilde{v} \equiv g(v)$ and

$$\tilde{k}^I_d(z_1,x_2,x_1) \equiv g\left(k^I_d(z_1,x_2,x_1)\right)$$

Then, it is possible to write

$$k_d(z_1,x_2,v) = k_d\left(z_1,x_2,g^{-1}(\tilde{v})\right) = \tilde{k}_d(z_1,x_2,\tilde{v})$$

where $\tilde{k}_d(z_1,x_2,v)$ is the inverse of $\tilde{k}^I_d(z_1,x_2,x_1)$. By Lemma 1 in Matzkin (2003), k_d and $\tilde{k}_d$ are observationally equivalent. Therefore, k_d and $\tilde{k}_d$ are also observationally equivalent, so k_d is identified up to a monotone transformation. By Assumption 1, $V \sim U(0,1)$, so k_d is identified.

The identification of g_d is instead based on Torgovitsky (2015). By the definition of g_d and k_d, Assumptions G.C, G.S, and FS.S in Torgovitsky (2015) are satisfied. Moreover, by Assumption 1, Assumptions G.N and FS.E are satisfied. These, together with Assumption 2 imply that $g(\cdot,\cdot,\cdot)$ is identified by Theorem 2 in Torgovitsky (2015).

Lastly, the identification of the copula is achieved by inverting the SQF and the CQF, which is possible by the continuity and the monotonicity of both functions in their last argument: $u = g^{-1}_d(x_1,x_2,y)$, $v = k^{-1}_d(z_1,x_2,x_1)$. Hence, the copula is identified by

$$C^d_{UV|X_2}(u,v|x_2) = \mathbb{P}\left(g^{-1}_d(X_1,X_2,Y) \leq u, k^{-1}_d(Z_1,X_2,X_1) \leq v|z\right)$$

A.2 Proof of Proposition 2

This proof is split into several steps. By Lemma 1 and the continuity of g_p and $c^r_{U|VX_2}$ (Assumptions 3 and 10), $\sqrt{n}\left(\hat{F}^m_{Y|Z}(y|z) - F^{m,\epsilon}_{Y|Z}(y|z)\right)$ can be written as:

$$
\begin{aligned}
&\sqrt{n}\left(\hat{F}^m_{Y|Z}(y|z) - F^{m,\epsilon}_{Y|Z}(y|z)\right) \\
&\quad = \sqrt{n}\int_{\mathcal{U}^2} \mathbf{1}\left((z'\hat{\gamma}_q(v), x_2)'\hat{\beta}_p(u) \le y\right)\left[c^r_{UV|X_2}\left(u, v|x_2; \hat{\theta}_r\right) - c^r_{UV|X_2}(u, v|x_2; \theta_r)\right] dudv \\
&\quad + \sqrt{n}\int_{\mathcal{U}^2} \left[\mathbf{1}\left((z'\hat{\gamma}_q(v), x_2)'\hat{\beta}_p(u) \le y\right) - \mathbf{1}\left((z'\hat{\gamma}_q(v), x_2)'\beta_p(u) \le y\right)\right] dC^r_{UV|X_2}(u, v|x_2; \theta_r) \\
&\quad + \sqrt{n}\int_{\mathcal{U}^2} \left[\mathbf{1}\left((z'\hat{\gamma}_q(v), x_2)'\beta_p(u) \le y\right) - \mathbf{1}\left((z'\gamma_q(v), x_2)'\beta_p(u) \le y\right)\right] dC^r_{UV|X_2}(u, v|x_2; \theta_r) \\
&\quad = \sqrt{n}\int_{\mathcal{U}^2} \mathbf{1}\left((z'\gamma_q(v), x_2)'\beta_p(u) \le y\right)\left[c^r_{UV|X_2}\left(u, v|x_2; \hat{\theta}_r\right) - c^r_{UV|X_2}(u, v|x_2; \theta_r)\right] dudv \\
&\quad + \sqrt{n}\int_{\mathcal{U}^2} \left[\mathbf{1}\left((z'\gamma_q(v), x_2)'\hat{\beta}_p(u) \le y\right) - \mathbf{1}\left((z'\gamma_q(v), x_2)'\beta_p(u) \le y\right)\right] dC^r_{UV|X_2}(u, v|x_2; \theta_r) \\
&\quad + \sqrt{n}\int_{\mathcal{U}^2} \left[\mathbf{1}\left((z'\hat{\gamma}_q(v), x_2)'\beta_p(u) \le y\right) - \mathbf{1}\left((z'\gamma_q(v), x_2)'\beta_p(u) \le y\right)\right] dC^r_{UV|X_2}(u, v|x_2; \theta_r) \\
&\quad + o^*_P(1)
\end{aligned}
$$

Using Lemmas 4-6, the chain rule for Hadamard differentiable mappings, Lemma 1, and the functional delta method, it follows that

$$
\begin{aligned}
&\sqrt{n}\left(\hat{F}^m_{Y|Z}(y|z) - F^{m,\epsilon}_{Y|Z}(y|z)\right) \\
&\quad \rightsquigarrow \int_\epsilon^{1-\epsilon}\left[\frac{\partial}{\partial\theta}C^r_{U|VX_2}\left(g_p^{-1}(z'\gamma_q(v), x_2, y)\,|v, x_2; \theta_r\right)\right] dv\mathbb{Z}_{\theta_r} \\
&\quad - \int_\epsilon^{1-\epsilon} f^m_{Y|ZV}(y|z, v)\,(z'\gamma_q(v), x_2)'\,\mathbb{Z}_{\beta_p}\left(C^{r,-1}_{U|VX_2}\left(F^m_{Y|ZV}(y|z, v)\,|v, x_2; \theta_r\right)\right) dv \\
&\quad - \int_\epsilon^{1-\epsilon} f^m_{Y|ZV}(y|z, v)\,\beta_{p,1}\left(C^{r,-1}_{U|VX_2}\left(F^m_{Y|ZV}(y|z, v)\,|v, x_2; \theta_r\right)\right) z'\mathbb{Z}_{\gamma_q}(v)\, dv \\
&\quad \equiv \mathbb{Z}_{F_m|Z}(y, z)
\end{aligned}
$$

in $\ell^\infty(\mathcal{YZM})$. The result holds uniformly in $m \in \mathcal{M}$ because $\mathcal{M}$ is a finite set.

Moreover, by Assumptions 4-5, apply Lemma E.4 in Chernozhukov et al. (2013) to obtain $\sqrt{n_s}\int_{\mathcal{Z}} fd\left(\hat{F}^s_Z - F^s_Z\right) \rightsquigarrow \mathbb{Z}_{Z_s}(f)$ in $\ell^\infty(\mathcal{F})$. Taking these results together, the desired result follows.

A.3 Proof of Proposition 3

By the results in Proposition 2, Lemma 1, and Assumption 11, and using the same arguments as in Proposition 2, it follows that

$$
\begin{aligned}
&\sqrt{n}\left(\hat{F}^{d,cf}_{Y|Z}\left(y|z\right)-F^{d,cf,\epsilon}_{Y|Z}\left(y|z\right)\right)\\
&=\sqrt{n}\int_{\mathcal{U}^2}\mathbf{1}\left(\left(k^{cf}_{d}\left(v\right),x_2\right)'\beta_d\left(u\right)\leq y\right)\left[c^{d}_{UV|X_2}\left(u,v|x_2;\hat{\theta}_d\right)-c^{d}_{UV|X_2}\left(u,v|x_2;\theta_d\right)\right]dudv\\
&+\sqrt{n}\int_{\mathcal{U}^2}\left[\mathbf{1}\left(\left(k^{cf}_{d}\left(v\right),x_2\right)'\hat{\beta}_d\left(u\right)\leq y\right)-\mathbf{1}\left(\left(k^{cf}_{d}\left(v\right),x_2\right)'\beta_d\left(u\right)\leq y\right)\right]dC^{d}_{UV|X_2}\left(u,v|x_2;\theta_d\right)\\
&+\sqrt{n}\int_{\mathcal{U}^2}\left[\mathbf{1}\left(\left(\hat{k}^{cf}_{d}\left(v\right),x_2\right)'\beta_d\left(u\right)\leq y\right)-\mathbf{1}\left(\left(k^{cf}_{d}\left(v\right),x_2\right)'\beta_d\left(u\right)\leq y\right)\right]dC^{d}_{UV|X_2}\left(u,v|x_2;\theta_d\right)\\
&+o^{*}_{P}\left(1\right)\\
&\rightsquigarrow\int_{\epsilon}^{1-\epsilon}\left[\frac{\partial}{\partial\theta}C^{d}_{U|VX_2}\left(g^{-1}_{d}\left(k^{cf}_{d}\left(v\right),x_2,y\right)|v,x_2;\theta_d\right)\right]dv\mathbb{Z}_{\theta_d}\\
&-\int_{\epsilon}^{1-\epsilon}f^{d}_{Y|ZV}\left(y|z,v\right)\left(k^{cf}_{d}\left(v\right),x_2\right)'\mathbb{Z}_{\beta_d}\left(C^{d,-1}_{U|VX_2}\left(F^{d}_{Y|ZV}\left(y|z,v\right)|v,x_2;\theta_d\right)\right)dv\\
&-\int_{\epsilon}^{1-\epsilon}f^{d}_{Y|ZV}\left(y|z,v\right)\beta_{d,1}\left(C^{d,-1}_{U|VX_2}\left(F^{d}_{Y|ZV}\left(y|z,v\right)|v,x_2;\theta_d\right)\right)z'\psi'_{k_d}\left(\mathbb{Z}_{\gamma_d}\left(v\right)\right)dv\\
&\equiv\mathbb{Z}_{F^{cf}_{d}|Z}\left(y,z\right)
\end{aligned}
$$

in $\ell^{\infty}\left(\mathcal{YZD}\right)$, and where we have used Theorem 2.1 in Fang and Santos (2019) rather than the functional delta method, to account for the potentially Hadamard directional differentiable counterfactual. The result holds uniformly in $m\in\mathcal{M}$ because $\mathcal{M}$ is a finite set. The desired result follows immediately.

Moreover, using the same arguments as in Theorem 1, it is immediate to show that

$$
\sqrt{n}\left(\hat{F}^{d,cf}_{Y}\left(y\right)-F^{d,cf,\epsilon}_{Y}\left(y\right)\right)\rightsquigarrow\int_{\mathcal{Z}}\mathbb{Z}_{F^{cf}_{d}|Z}\left(y|z\right)dF^{d}_{Z}\left(z\right)+\sqrt{\lambda_s}\mathbb{Z}_{Z_d}\left(F^{d,cf,\epsilon}_{Y|Z}\left(y|z\right)\right)\equiv\mathbb{Z}_{F^{cf}_{d}}\left(y\right)
$$

uniformly in $\ell^{\infty}(\mathcal{YD})$, and

$$\sqrt{n}\left(\hat{Q}_Y^{d,cf}(\tau) - Q_Y^{d,cf,\epsilon}(\tau)\right) \rightsquigarrow -\frac{\mathbb{Z}_{F_d^{cf}}\left(Q_Y^{d,cf,\epsilon}(\tau)\right)}{f_Y^d\left(Q_Y^{d,cf,\epsilon}(\tau)\right)} \equiv \mathbb{Z}_{Q_d^{cf}}(\tau)$$

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

A.4 Proof of Proposition 4

$$\begin{aligned}
&\sup_{y,z,m}\sqrt{n}\left|\check{F}^m_{Y|Z}(y|z) - F^{m,\epsilon}_{Y|Z}(y|z)\right| \\
&\quad\le \sup_{y,z,m}\sqrt{n}\left|\check{F}^m_{Y|Z}(y|z) - \tilde{F}^m_{Y|Z}(y|z)\right| + \sup_{y,z}\sqrt{n}\left|\tilde{F}^m_{Y|Z}(y|z) - F^{m,\epsilon}_{Y|Z}(y|z)\right| \\
&\quad= \sup_{y,z,m}\sqrt{n}\left|\int_{\epsilon}^{1-\epsilon}\mathbf{1}\left(\hat{g}_p\left(\hat{x}_1^q(v), x_2, u\right) \le y\right) d\left(\check{C}^r_{UV}(u,v) - C^r_{UV}(u,v)\right)\right| + O_P(1) \\
&\quad\le \sup_{r}\sqrt{n}\int_{\epsilon}^{1-\epsilon} d\left|\check{C}^r_{UV}(u,v) - C^r_{UV}(u,v)\right| + O_P(1) \\
&\quad\le \sup_{u,v,r}\sqrt{n}\left|\check{C}^r_{UV}(u,v) - C^r_{UV}(u,v)\right| + O_P(1) = O_P(1)
\end{aligned}$$

where the first inequality follows from the triangle inequality, the first equality from the definition of the estimators and the uniform consistency of $\tilde{F}^m_{Y|Z}(y|z)$ shown in Lemma 3, the second inequality from the fact that the indicator function is no larger than one, the third inequality by taking the supremum of the difference, and the last equality by Lemma 9.

By Lemma 4, it follows that $\sup_{y,m}\sqrt{n}\left|\check{F}^m_Y(y) - F^{m,\epsilon}_Y(y)\right| = O_P(1)$. Also, because the inverse map is Hadamard differentiable uniformly with respect to an index (Chernozhukov et al., 2010), it follows that $\sup_{\tau,m}\sqrt{n}\left|\check{Q}^m_Y(\tau) - Q^{m,\epsilon}_Y(\tau)\right| = O_P(1)$. Both results hold uniformly in $m \in \mathcal{M}$ because $\mathcal{M}$ is a finite set.

A.5 Proof of Theorem 1

First, I show the asymptotic distribution of $\hat{F}_Y^m(y)$. By Proposition 2,

$$\sqrt{n}\begin{pmatrix}\hat{F}_{Y|Z}^m(y|z) - F_{Y|Z}^{m,\epsilon}(y|z) \\ \int_{\mathcal{Z}} f d\left(\hat{F}_Z^s(z) - F_Z^s(z)\right)\end{pmatrix} \rightsquigarrow \begin{pmatrix}\mathbb{Z}_{F_m|Z}(y,z) \\ \sqrt{\lambda_s}\mathbb{Z}_{Z_s}(f)\end{pmatrix}$$

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

By Lemma D.1 in Chernozhukov et al. (2013),

$$\begin{aligned}\sqrt{n}\left(\hat{F}_Y^m(y) - F_Y^{m,\epsilon}(y)\right) &= \sqrt{n}\int_{\mathcal{Z}}\left(\hat{F}_{Y|Z}^m(y|z) - F_{Y|Z}^{m,\epsilon}(y|z)\right)dF_Z^s(z) \\ &+ \sqrt{n}\int_{\mathcal{Z}} F_{Y|Z}^m(y|z)\, d\left(\hat{F}_Z^s(z) - F_Z^s(z)\right) + o_P^*(1) \\ &\rightsquigarrow \int_{\mathcal{Z}}\mathbb{Z}_{F_m|Z}(y|z)\, dF_Z^s(z) + \sqrt{\lambda_s}\mathbb{Z}_{Z_s}\left(F_{Y|Z}^{m,\epsilon}(y|z)\right) \equiv \mathbb{Z}_{F_m}(y)\end{aligned}$$

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

Next, I show the asymptotic distribution of $\hat{Q}_Y^m(\tau)$. By the functional delta method,

$$\begin{aligned}\sqrt{n}\left(\hat{Q}_Y^m(\tau) - Q_Y^{m,\epsilon}(\tau)\right) &= -\frac{\sqrt{n}\left(\hat{F}_Y^m\left(Q_Y^{m,\epsilon}(\tau)\right) - F_Y^{m,\epsilon}\left(Q_Y^{m,\epsilon}(\tau)\right)\right)}{f_Y^m\left(Q_Y^{m,\epsilon}(\tau)\right)} + o_P^*(1) \\ &\rightsquigarrow -\frac{\mathbb{Z}_{F_m}\left(Q_Y^{m,\epsilon}(\tau)\right)}{f_Y^m\left(Q_Y^{m,\epsilon}(\tau)\right)} \equiv \mathbb{Z}_{Q_m}(\tau)\end{aligned}$$

in $\ell^\infty(\mathcal{TM})$, where I have used the Hadamard differentiability of the quantile operator (Chernozhukov et al., 2010). By Assumption 6, $\tau \to Q_Y^{m,\epsilon}(\tau)$ is a.s. uniformly continuous, and together with the a.s. uniform continuity of $\mathbb{Z}_{F_m}(y)$, it follows that $\mathbb{Z}_{Q_m}(\tau)$ is a.s. uniformly continuous with respect to τ. The result holds jointly in $m \in \mathcal{M}$ because $\mathcal{M}$ is a finite set.

A.6 Proof of Theorem 2

Using the same arguments used in Lemma 1 and Assumption 13, it follows that

$$\sqrt{n}\left(\hat{\vartheta}_m^*(\cdot,\cdot\cdot)-\vartheta_m(\cdot,\cdot\cdot)\right)\equiv\sqrt{n}\begin{bmatrix}\hat{\beta}_p^*(\cdot)-\beta_p(\cdot)\\ \hat{\gamma}_q^*(\cdot\cdot)-\gamma_q(\cdot\cdot)\\ \hat{\theta}_r^*-\theta_r\end{bmatrix}$$

$$=\begin{bmatrix}\begin{bmatrix}J_p(\cdot)^{-1} & 0\\ 0 & H_q(\cdot\cdot)^{-1}\end{bmatrix}\mathbb{G}_n^* f_m\left(E,\beta_1(\cdot),\upsilon(\cdot,\cdot\cdot),\cdot,\cdot\cdot\right)\\ L_r^{-1}\left[\mathbb{G}_n^*\left[\frac{\partial\ell_r(U_r,V,\theta_r)}{\partial\theta}\right]\right.\\ \left.+\int_{\mathcal{Z}}\int_{\mathcal{U}^2}\frac{\partial^2\ell_r(U_r,V,\theta_r)}{\partial\theta\partial(u,v)}\sqrt{n}W\begin{pmatrix}\hat{U}_r-U_r\\ \hat{V}-V\end{pmatrix}dC^r_{UV|X_2}\left(U_r,V|X_2;\theta_r\right)dF_Z^r(z)\right]\end{bmatrix}+o_P^*(1)$$

in $\ell^\infty(\mathcal{UUM})$, and where $f\mapsto\mathbb{G}_n^*[f(E)]\equiv\frac{1}{\sqrt{n}}\sum_{i=1}^n W_i\left[f(E_i)-\mathbb{E}\left(f(E_i)\right)\right]$. Consequently, $\sqrt{n}\left(\hat{\vartheta}_m^*(\cdot,\cdot\cdot)-\vartheta_m(\cdot,\cdot\cdot)\right)\rightsquigarrow\mathbb{Z}_{\vartheta_m}^*(\cdot,\cdot\cdot)\equiv\sqrt{\omega_0}\mathbb{Z}_{\vartheta_m}(\cdot,\cdot\cdot)$, a zero-mean Gaussian process with covariance $\omega_0\Sigma_{\vartheta_m}(\tau,\nu,\tilde{\tau},\tilde{\nu})$. By the functional delta method, Proposition 2, and Theorem 1, it is straightforward to show that

$$\sqrt{\frac{n}{\omega_0}}\begin{pmatrix}\hat{F}_{Y|Z}^{m,*}(y|z)-F_{Y|Z}^{m,\epsilon}(y|z)\\ \int_{\mathcal{Z}}fd\left(\hat{F}_Z^{s,*}(z)-F_Z^s(z)\right)\end{pmatrix}\rightsquigarrow\begin{pmatrix}\mathbb{Z}_{F_m|Z}(y,z)\\ \sqrt{\lambda_s}\mathbb{Z}_{Z_s}(f)\end{pmatrix}\text{ in }\ell^\infty(\mathcal{YZFM})$$

$$\sqrt{\frac{n}{\omega_0}}\left(\hat{Q}_Y^{m,*}(\tau)-Q_Y^{m,\epsilon}(\tau)\right)\rightsquigarrow\mathbb{Z}_{Q_m}(\tau)\text{ in }\ell^\infty(\mathcal{TM})$$

$$\sqrt{\frac{n}{\omega_0}}\left(\hat{F}_Y^{m,*}(y)-F_Y^{m,\epsilon}(y)\right)\rightsquigarrow\mathbb{Z}_{F_m}(y)\text{ in }\ell^\infty(\mathcal{YM})$$

B Auxiliary Lemmas

B.1 Asymptotic Distribution of the Structural Estimators

Lemma 1. *Let $\hat{\gamma}_p(\nu)$ and $\hat{\beta}_q(\tau)$ denote the conditional QR and conditional IVQR estimators of quantiles ν and τ of Equations 2 and 1, respectively. Under Assumptions 3 to 10, their* *joint asymptotic distribution is given by:*

$$\sqrt{n}\left(\hat{\vartheta}_m(\cdot,\cdot\cdot)-\vartheta_m(\cdot,\cdot\cdot)\right)\equiv\begin{pmatrix}\hat{\beta}_p(\cdot)-\beta_p(\cdot)\\ \hat{\gamma}_q(\cdot\cdot)-\gamma_q(\cdot\cdot)\\ \hat{\theta}_r-\theta_r\end{pmatrix}\rightsquigarrow\begin{pmatrix}\mathbb{Z}_{\beta_p}(\cdot)\\ \mathbb{Z}_{\gamma_q}(\cdot\cdot)\\ \mathbb{Z}_{\theta_r}\end{pmatrix}\equiv\mathbb{Z}_{\vartheta_m}(\cdot,\cdot\cdot)$$

a stochastic process in metric space $\ell^{\infty}(\mathcal{UUM})$, where $\mathbb{Z}_{\vartheta_m}(\tau,\nu)$ is a zero-mean tight Gaussian process with a.s. uniformly continuous paths in $\mathcal{UUM}$.

Proof. The first step is to show the joint asymptotic distribution of the IVQR and QR estimators. By Assumption 6, $Q_m(\beta,\iota,\gamma,\tau,\nu)$ is continuous over $\mathcal{B}\times\mathcal{I}\times\mathcal{G}\times\mathcal{U}\times\mathcal{U}$. Furthermore, by Lemma 8, $\sup_{(\beta,\iota,\gamma)\in\mathcal{B}\times\mathcal{I}\times\mathcal{G}}\|Q_{m,n}(\beta,\iota,\gamma,\tau,\nu)-Q_m(\beta,\iota,\gamma,\tau,\nu)\|\xrightarrow{P}0$.

Hence, the uniform convergence of $\sup_{(\beta_1,\tau,\nu)\in\mathcal{B}_1\times\mathcal{U}\times\mathcal{U}}\left\|\hat{\zeta}_m(\beta_1,\tau,\nu)-\zeta_m(\beta_1,\tau,\nu)\right\|\xrightarrow{P}0$ follows by Lemma 7, which implies that $\sup_{(\beta_1,\tau)\in\mathcal{B}_1\times\mathcal{U}}\left\|\|\hat{\iota}_p(\beta_1,\tau)\|_{A_p(\tau)}-\|\iota_p(\beta_1,\tau)\|_{A_p(\tau)}\right\|\xrightarrow{P}0$, where $\|x\|_A=\sqrt{x'Ax}$ for some uniformly positive definite matrix $A_p(\tau)$, such as $A_p(\tau)=I$ or $A_p(\tau)=\frac{1}{n_d}\sum_{i=1}^{n}\hat{\Phi}_{d,i}(\tau)\hat{\Phi}_{d,i}(\tau)'$.

By Lemma 7, $\sup_{\tau\in\mathcal{U}}\left\|\hat{\beta}_{p,1}(\tau)-\beta_{p,1}(\tau)\right\|\xrightarrow{P}0$, and hence $\sup_{\tau\in\mathcal{U}}\left\|\hat{\beta}_{p,2}(\tau)-\beta_{p,2}(\tau)\right\|\xrightarrow{P}0$, $\sup_{\tau\in\mathcal{U}}\left\|\hat{\iota}_p\left(\hat{\beta}_{p,1}(\tau)\tau\right)-0\right\|\xrightarrow{P}0$, and $\sup_{\nu\in\mathcal{U}}\|\hat{\gamma}_q(\nu)-\gamma_q(\nu)\|\xrightarrow{P}0$, which proves their uniform consistency.

Consider a collection of closed balls $B_{\delta_n}(\beta_{p,1}(\tau))$ centered at $\beta_{p,1}(\tau)\,\forall\tau$, with radius δ_n independent of τ and $\delta_n\to0$ slowly enough. Let $\beta_{1n}(\tau)$ be any value inside $B_{\delta_n}(\beta(\tau))$. By Theorem 3.3 in Koenker and Bassett (1978),

$$O\left(\frac{1}{\sqrt{n}}\right)=\sqrt{n}\mathbb{E}_n\hat{R}_m\left(E,\beta_{1n}(\cdot),\hat{\zeta}_m(\beta_{1n}(\cdot),\cdot,\cdot\cdot),\cdot,\cdot\cdot\right)$$

By Lemma 8, the following expansion holds for any $\sup_{\tau\in\mathcal{U}}\|\beta_{1n}(\tau)-\beta_{p,1}(\tau)\|\xrightarrow{P}0$

$$\begin{aligned}
O\left(\frac{1}{\sqrt{n}}\right) &= \mathbb{G}_n\hat{R}_m\left(E,\beta_{1n}(\cdot),\hat{\zeta}_m(\beta_{1n}(\cdot),\cdot,\cdot\cdot),\cdot,\cdot\cdot\right)\\
&+\sqrt{n}\mathbb{E}\hat{R}_m\left(E,\beta_{1n}(\cdot),\hat{\zeta}_{m,n}(\beta_{1n}(\cdot),\cdot,\cdot\cdot),\cdot,\cdot\cdot\right)\\
&= \mathbb{G}_nR_m(E,\beta_{p,1}(\cdot),\zeta_m(\beta_1(\cdot),\cdot,\cdot\cdot),\cdot,\cdot\cdot)+o_P(1)\\
&+\sqrt{n}\mathbb{E}\hat{R}_m\left(E,\beta_{1n}(\cdot),\hat{\zeta}_{m,n}(\beta_{1n}(\cdot),\cdot,\cdot\cdot),\cdot,\cdot\cdot\right)\ \textit{in}\ \ell^\infty(\mathcal{UU})\\
&= \mathbb{G}_nR_m(E,\beta_{p,1}(\cdot),\zeta_m(\cdot,\cdot\cdot),\cdot,\cdot\cdot)+o_P(1)\\
&+(J_{m,\zeta}(\cdot,\cdot\cdot)+o_P(1))\sqrt{n}\left(\hat{\zeta}_m(\beta_{1n}(\cdot),\cdot,\cdot\cdot)-\zeta_m(\cdot,\cdot\cdot)\right)\\
&+(J_{m,\beta_1}(\cdot)+o_P(1))\sqrt{n}(\beta_{1n}(\cdot)-\beta_1(\cdot))\ \textit{in}\ \ell^\infty(\mathcal{UU})
\end{aligned}$$

where

$$J_{m,\zeta}(\cdot,\cdot\cdot)\equiv\frac{\partial}{\partial(\beta_2',\iota',\gamma')}\mathbb{E}\left.\left[\begin{array}{c}\varphi_\cdot\left(Y-X_1'\beta_1(\cdot)-X_2'\beta_2-\Phi_p(\cdot)'\iota\right)\Psi_p(\cdot)\mathbf{1}(D=p)\\ \varphi_{\cdot\cdot}(X_1-Z'\gamma)z\mathbf{1}(D=q)\end{array}\right]\right|_{\zeta=\zeta_m(\cdot,\cdot\cdot)}$$

$$J_{m,\beta_1}(\cdot)\equiv\left[\begin{array}{c}\frac{\partial}{\partial\beta_1}\mathbb{E}\left[\varphi_\cdot(Y-X_1'\beta_1-X_2'\beta_2(\cdot))\Psi_p(\cdot)\mathbf{1}(D=p)\right]\Big|_{\beta_1=\beta_1(\cdot)}\\ 0_{d_X\times 1}\end{array}\right]$$

For any $\sup_{\tau\in\mathcal{U}}\|\beta_{1n}(\tau)-\beta_1(\tau)\|\xrightarrow{P}0$

$$\begin{aligned}
\sqrt{n}\left(\hat{\zeta}_m(\beta_{1n}(\cdot),\cdot,\cdot\cdot)-\zeta_m(\cdot,\cdot\cdot)\right) &= -J_{m,\zeta}^{-1}(\cdot,\cdot\cdot)\mathbb{G}_nR_m(E,\beta_{p,1}(\cdot),\zeta_m(\cdot,\cdot\cdot),\cdot,\cdot\cdot)\\
&-J_{m,\zeta}^{-1}(\cdot,\cdot\cdot)J_{m,\beta_{p,1}}(\cdot)[1+o_P(1)]\sqrt{n}(\beta_{1n}(\cdot)-\beta_1(\cdot))+o_P(1)
\end{aligned}$$

in $\ell^\infty(\mathcal{UU})$. Consequently,

$$\begin{aligned}
\sqrt{n}(\hat{\iota}_p(\beta_{1n}(\cdot),\cdot)-0) &= -\bar{J}_{m,\iota}(\cdot,\cdot\cdot)\mathbb{G}_nR_m(E,\beta_{p,1}(\cdot),\zeta_m(\cdot,\cdot\cdot),\cdot,\cdot\cdot)\\
&-\bar{J}_{m,\iota}(\cdot,\cdot\cdot)J_{m,\beta_1}(\cdot)[1+o_P(1)]\sqrt{n}(\beta_{1n}(\cdot)-\beta_{p,1}(\cdot))+o_P(1)
\end{aligned}$$

in $\ell^\infty(\mathcal{UU})$, where $\left[\bar{J}_{m,\beta_2}(\cdot,\cdot\cdot)':\bar{J}_{m,\iota}(\cdot,\cdot\cdot)':\bar{J}_{m,\gamma}(\cdot,\cdot\cdot)'\right]$ is the conformable partition of $J_{m,\zeta}^{-1}(\cdot,\cdot\cdot)$.

By the uniform consistency of $\hat{\beta}_{p,1}(\tau)$, $wp \to 1$,

$$\hat{\beta}_{p,1}(\tau) = \arg \min_{\beta_{1n}(\tau) \in B_n(\beta_1(\tau))} \|\hat{\iota}_p(\beta_{1n}(\tau), \tau)\|_{A_p(\tau)} \;\; \forall \tau \in \mathcal{U}$$

By Lemma 8, $\mathbb{G}_n R_m(E, \beta_{p,1}(\cdot), \zeta_m(\cdot, \cdot\cdot), \cdot, \cdot\cdot) = O_P(1)$, so it follows that

$$\sqrt{n}\|\hat{\iota}_p(\beta_{1n}(\cdot), \cdot)\|_{A_p(\cdot)} = \left\|O_P(1) - \bar{J}_{m,\iota}(\cdot, \cdot\cdot) J_{m,\beta_1}(\cdot)[1 + o_P(1)]\sqrt{n}(\beta_{1n}(\cdot) - \beta_{p,1}(\cdot))\right\|_{A_p(\cdot)}$$

in $\ell^\infty(\mathcal{U}\mathcal{U})$, since $A_p(\cdot)$ and $\overline{J}_{m,\iota}(\cdot, \cdot\cdot) J_{m,\beta_1}(\cdot)$ have full rank uniformly in $\mathcal{U} \times \mathcal{U}$. Hence,

$$\begin{aligned}
&\sqrt{n}\left(\hat{\beta}_{p,1}(\cdot) - \beta_{p,1}(\cdot)\right) \\
&\quad = \arg\min_{\mu \in \mathbb{R}} \left\|-\bar{J}_{m,\iota}(\cdot)\mathbb{G}_n R_m(E, \beta_{p,1}(\cdot), \zeta_m(\cdot, \cdot\cdot), \cdot, \cdot\cdot) - \bar{J}_{m,\iota}(\cdot, \cdot\cdot)\bar{J}_{m,\beta_1}(\cdot)\mu\right\|_{A_p(\cdot)} + o_P(1)
\end{aligned}$$

in $\ell^\infty(\mathcal{U}\mathcal{U})$. So jointly in $\ell^\infty(\mathcal{U}\mathcal{U})$

$$\begin{aligned}
\sqrt{n}\left(\hat{\beta}_{p,1}(\cdot) - \beta_{p,1}(\cdot)\right) &= -\left(J_{m,\beta_1}(\cdot)' \bar{J}_{m,\iota}(\cdot, \cdot\cdot)' A_p(\cdot) \bar{J}_{m,\iota}(\cdot, \cdot\cdot) J_{m,\beta_1}(\cdot)\right)^{-1} \\
&\cdot \left(J_{m,\beta_1}(\cdot)' \bar{J}_{m,\iota}(\cdot, \cdot\cdot)' A_p(\cdot) \bar{J}_{m,\iota}(\cdot, \cdot\cdot)\right)\mathbb{G}_n R_m(E, \beta_{p,1}(\cdot), \zeta_m(\cdot, \cdot\cdot), \cdot, \cdot\cdot) + o_P(1) \\
&= O_P(1)
\end{aligned}$$

$$\begin{aligned}
&\sqrt{n}\left(\hat{\zeta}_m\left(\hat{\beta}_{p,1}(\cdot), \cdot, \cdot\cdot\right) - \zeta_m(\cdot, \cdot\cdot)\right) \\
&\quad = -J_{m,\zeta}^{-1}(\cdot, \cdot\cdot)\left[I - J_{m,\beta_1}(\cdot)\left(J_{m,\beta_1}(\cdot)' \bar{J}_{m,\iota}(\cdot, \cdot\cdot)' A_p(\cdot) \bar{J}_{m,\iota}(\cdot, \cdot\cdot) J_{m,\beta_1}(\cdot)\right)^{-1}\right. \\
&\quad \left.\cdot\; J_{\beta_{m,1}}(\cdot)' \bar{J}_{m,\iota}(\cdot, \cdot\cdot)' A_p(\cdot) \bar{J}_{m,\iota}(\cdot, \cdot\cdot)\right]\mathbb{G}_n R_m(E, \beta_{p,1}(\cdot), \zeta_m(\cdot, \cdot\cdot), \cdot, \cdot\cdot) + o_P(1) \\
&\quad = O_P(1)
\end{aligned}$$

Due to the invertibility of $J_{m,\beta_1}(\tau)\,\bar{J}_{m,\iota}(\tau,\nu)$,

$$\begin{aligned}\sqrt{n}\left(\hat{\iota}_p\left(\hat{\beta}_{p,1}(\cdot),\cdot\right)-0\right) &= -\bar{J}_{m,\iota}(\cdot,\cdot\cdot)\left[I-J_{m,\beta_1}(\cdot)\left[J_{m,\beta_1}(\cdot)'\,\bar{J}_{m,\iota}(\cdot,\cdot\cdot)'\right]^{-1}\bar{J}_{m,\iota}(\cdot,\cdot\cdot)\right]\\ &\quad\cdot\mathbb{G}_n R_m\left(E,\beta_{p,1}(\cdot),\zeta_m(\cdot,\cdot\cdot),\cdot,\cdot\cdot\right)+o_P(1)\\ &= 0\times O_P(1)+o_P(1)\end{aligned}$$

in $\ell^\infty(\mathcal{U}\mathcal{U})$. Because $\left(\beta_{1n}(\cdot),\hat{\zeta}_m(\beta_{1n}(\cdot),\cdot,\cdot\cdot)\right)=\left(\hat{\beta}_{p,1}(\cdot),\hat{\beta}_{p,2}(\cdot),0+o_P\left(\frac{1}{\sqrt{n}}\right),\hat{\gamma}_q(\cdot\cdot)\right)$, and substituting it into the expansion yields

$$-\mathbb{G}_n R_m\left(E,\beta_{p,1}(\cdot),\zeta(\cdot,\cdot\cdot),\cdot,\cdot\cdot\right)=\begin{bmatrix}J_p(\cdot) & 0_{d_X}\\ 0_{d_X} & H_q(\cdot\cdot)\end{bmatrix}\sqrt{n}\begin{pmatrix}\hat{\beta}_p(\cdot)-\beta_p(\cdot)\\ \hat{\gamma}_q(\cdot\cdot)-\gamma_q(\cdot\cdot)\end{pmatrix}+o_P(1)$$

in $\ell^\infty(\mathcal{U}\mathcal{U})$. By Lemma 8, $\mathbb{G}_n R_m\left(E,\beta_{p,1}(\cdot),\zeta_m(\cdot,\cdot\cdot),\cdot,\cdot\cdot\right)\rightsquigarrow\mathbb{Z}_{R_m}(\cdot,\cdot\cdot)$ *in* $\ell^\infty(\mathcal{U}\mathcal{U})$, a Gaussian process with covariate function $\Sigma_{R_{m,\tilde{m}}}(\tau,\nu,\tilde{\tau},\tilde{\nu})=\mathbb{E}\left[\mathbb{Z}_{R_m}(\tau,\nu)\,\mathbb{Z}_{R_{\tilde{m}}}(\tilde{\tau},\tilde{\nu})'\right]$, which yields

$$\sqrt{n}\begin{pmatrix}\hat{\beta}_p(\cdot)-\beta_p(\cdot)\\ \hat{\gamma}_q(\cdot\cdot)-\gamma_q(\cdot\cdot)\end{pmatrix}\rightsquigarrow\begin{bmatrix}J_p(\cdot)^{-1} & 0_{d_X}\\ 0_{d_X} & H_q(\cdot\cdot)^{-1}\end{bmatrix}\mathbb{Z}_{R_m}(\cdot,\cdot\cdot)\equiv\mathbb{Z}_{L_m}(\cdot,\cdot\cdot)\qquad(17)$$

where

$$H_q(\nu)\equiv\mathbb{E}\left[f_{X_1}\left(Z'\gamma_q(\nu)\,|Z\right)ZZ'\mathbf{1}(D=q)\right]=\frac{1}{\lambda_q}\mathbb{E}\left[f_{X_1}\left(Z'\gamma_q(\nu)\,|Z\right)ZZ'|D=q\right]\equiv\frac{1}{\lambda_q}\overline{H}_q(\nu)$$

$$\begin{aligned}J_p(\tau)&\equiv\mathbb{E}\left[f_Y\left(X'\beta_p(\tau)\,|X,Z_1\right)\Psi_p(\tau)\,X'\mathbf{1}(D=p)\right]\\ &=\frac{1}{\lambda_p}\mathbb{E}\left[f_Y\left(X'\beta_p(\tau)\,|X,Z_1\right)\Psi_p(\tau)\,X'|D=p\right]\equiv\frac{1}{\lambda_p}\overline{J}_p(\tau)\end{aligned}$$

where the expectations are with respect to $F_Z^d(z)$.

The covariance function of the process is given by

$$\Sigma_{L_{m\tilde{m}}}(\tau,\nu,\tilde{\tau},\tilde{\nu}) = \begin{bmatrix} \Sigma^{1,1}_{L_{m\tilde{m}}}(\tau,\tilde{\tau}) & \Sigma^{2,1}_{L_{m\tilde{m}}}(\tau,\tilde{\nu})' \\ \Sigma^{2,1}_{L_{\tilde{m}m}}(\tilde{\tau},\nu) & \Sigma^{2,2}_{L_{m\tilde{m}}}(\nu,\tilde{\nu}) \end{bmatrix}$$

where

$$\begin{aligned}
\Sigma^{1,1}_{L_{m\tilde{m}}}(\tau,\tilde{\tau}) &= J_p(\tau)^{-1}(\tau\wedge\tilde{\tau}-\tau\tilde{\tau})\,\mathbb{E}\left[\Psi_p(\tau,Z)\,\Psi_p(\tilde{\tau},Z)'\,\mathbf{1}(D=p)\right]J_p(\tilde{\tau})^{-1}\,\mathbf{1}(p=\tilde{p}) \\
&= \lambda_p\overline{J}_p(\tau)^{-1}(\tau\wedge\tilde{\tau}-\tau\tilde{\tau})\,\mathbb{E}\left[\Psi_p(\tau)\,\Psi_p(\tilde{\tau})'\,|D=p\right]\overline{J}_p(\tilde{\tau})^{-1}\,\mathbf{1}(p=\tilde{p}) \\
\Sigma^{2,1}_{L_{\tilde{m}m}}(\tilde{\tau},\nu) &= H_q(\nu)^{-1}\,\mathbb{E}\left[\left(\mathbf{1}(Y\le X'\beta_q(\tilde{\tau}))\,\mathbf{1}(X_1\le Z'\gamma_q(\nu))-\tilde{\tau}\nu\right)Z\Psi_q(\tilde{\tau},z)'\,\mathbf{1}(D=q)\right]' \\
&\quad\cdot J_q(\tilde{\tau})^{-1}\,\mathbf{1}(q=\tilde{p}) \\
&= \lambda_q\overline{H}_q^{-1}(\nu)\,\mathbb{E}\left[\left(\mathbf{1}(Y\le X'\beta_q(\tilde{\tau}))\,\mathbf{1}(X_1\le Z'\gamma_q(\nu))-\tilde{\tau}\nu\right)Z\Psi_q(\tilde{\tau},z)'\,|D=q\right]' \\
&\quad\cdot \overline{J}_q^{-1}(\tilde{\tau})\,\mathbf{1}(q=\tilde{p}) \\
\Sigma^{2,2}_{L_{m\tilde{m}}}(\nu,\tilde{\nu}) &= H_q(\nu)^{-1}(\nu\wedge\tilde{\nu}-\nu\tilde{\nu})\,\mathbb{E}\left[ZZ'\mathbf{1}(D=q)\right]H_q(\tilde{v})^{-1}\,\mathbf{1}(q=\tilde{q}) \\
&= \lambda_q\overline{H}_q^{-1}(\nu)\,\mathbb{E}\left[ZZ'|D=q\right]\overline{H}_q(\tilde{v})^{-1}\,\mathbf{1}(q=\tilde{q})
\end{aligned}$$

The second step is to express the estimator of the copulas parameters as a function of the IVQR and QR estimators. To do so, apply the mean value theorem to the score:

$$0=\mathbb{E}_n\left[\frac{\partial\ell_r\left(\hat{U}_r,\hat{V},\hat{\theta}_r\right)}{\partial\theta}\right]=\mathbb{E}_n\left[\frac{\partial\ell_r\left(\hat{U}_r,\hat{V},\theta_r\right)}{\partial\theta}\right]+\mathbb{E}_n\left[\frac{\partial^2\ell_r\left(\hat{U}_r,\hat{V},\overline{\theta}_r\right)}{\partial\theta\partial\theta'}\right]\left(\hat{\theta}_r-\theta_r\right)$$

where $\overline{\theta}_r$ lies between $\hat{\theta}_r$ and θ_r. Rearranging the previous equation yields

$$\sqrt{n}\left(\hat{\theta}_r-\theta_r\right)=\left[\mathbb{E}_n\left[-\frac{\partial^2\ell_r\left(\hat{U}_r,\hat{V},\overline{\theta}_r\right)}{\partial\theta\partial\theta'}\right]\right]^{-1}\sqrt{n}\mathbb{E}_n\left[\frac{\partial\ell_r\left(\hat{U}_r,\hat{V},\theta_r\right)}{\partial\theta}\right] \tag{18}$$

Define the Hessian as

$$L_r\equiv\mathbb{E}\left[-\frac{\partial^2\ell_r(U_r,V,\theta_r)}{\partial\theta\partial\theta'}\right]$$

Now I show its uniform convergence:

$$\begin{aligned}
\left\|\frac{\partial^2 \ell_r\left(\hat{U}_r,\hat{V},\overline{\theta}_r\right)}{\partial\theta\partial\theta'}-\frac{\partial^2 \ell_r\left(U_r,V,\theta_r\right)}{\partial\theta\partial\theta'}\right\| &= \left\|vec\left(\frac{\partial^2 \ell_r\left(\hat{U}_r,\hat{V},\overline{\theta}_r\right)}{\partial\theta\partial\theta'}-\frac{\partial^2 \ell_r\left(U_r,V,\theta_r\right)}{\partial\theta\partial\theta'}\right)\right\| \\
&= \left\|\nabla^3 \ell_r\left(\overline{U}_r,\overline{V},\overline{\overline{\theta}}_r\right)\begin{pmatrix}\hat{U}_r-U_r\\ \hat{V}-V\\ \overline{\theta}_r-\theta_r\end{pmatrix}\right\| \\
&\leq \left\|\nabla^3 \ell_r\left(\overline{U}_r,\overline{V},\overline{\overline{\theta}}_r\right)\right\|\left\|\begin{pmatrix}\hat{U}_r-U_r\\ \hat{V}-V\\ \overline{\theta}_r-\theta_r\end{pmatrix}\right\| \\
&\leq K\cdot o_P^*\left(1\right)=o_P^*\left(1\right)
\end{aligned}$$

where $U_r=\epsilon+\int_\epsilon^{1-\epsilon}\mathbf{1}\left(X'\beta_r\left(u\right)\leq Y\right)du,\ V=\int_\epsilon^{1-\epsilon}\mathbf{1}\left(Z'\gamma_r\left(v\right)\leq X_1\right)dv,\ \nabla^3\ell_r\left(u,v,\theta_r\right)$ is a matrix such that its (i,k)-th element is the partial derivative of the i-th element of $vec\left(\frac{\partial^2\ell_r(U_r,V,\theta_r)}{\partial\theta\partial\theta'}\right)$ writh respect to the l-th element of $\left(u,v,\theta_r'\right)'$, vec denotes the vectorization operator, and $\|\cdot\|$ the Euclidean norm. The first equality follows by the mean value theorem, the first inequality by the triangular inequality, and the last inequality by Assumption 10. Using this result,

$$\begin{aligned}
\left\|\mathbb{E}_n\left[-\frac{\partial^2 \ell_r\left(\hat{U}_r,\hat{V},\overline{\theta}_r\right)}{\partial\theta\partial\theta'}\right]-L_r\right\| \leq{}& \mathbb{E}_n\left\|-\frac{\partial^2 \ell_r\left(\hat{U}_r,\hat{V},\overline{\theta}_r\right)}{\partial\theta\partial\theta'}+\frac{\partial^2 \ell_r\left(U_r,V,\theta_r\right)}{\partial\theta\partial\theta'}\right\| \\
&+\left\|\mathbb{E}_n\left[-\frac{\partial^2 \ell_r\left(U_r,V,\theta_r\right)}{\partial\theta\partial\theta'}\right]-L_r\right\|=o_P^*\left(1\right)
\end{aligned}$$

where the inequality follows by the triangular inequality, the first term is $o_P^*\left(1\right)$ by the argument above, and the second term by uniform law of large numbers.

Then, I show the asymptotic distribution of $\sqrt{n}\mathbb{E}_n\left[\frac{\partial\ell_r\left(\hat{U}_r,\hat{V},\theta_r\right)}{\partial\theta}\right]$. Apply the mean value theorem to $\left(\hat{U}_r, \hat{V}\right)$:

$$\sqrt{n}\mathbb{E}_n\left[\frac{\partial \ell_r\left(\hat{U}_r, \hat{V}, \theta_r\right)}{\partial \theta}\right] = \mathbb{G}_n\left[\frac{\partial \ell_r\left(U_r, V, \theta_r\right)}{\partial \theta}\right] + \sqrt{n}\mathbb{E}_n\left[\frac{\partial^2 \ell_r\left(\overline{U}_r, \overline{V}, \theta_r\right)}{\partial \theta \partial\left(u, v\right)}\begin{pmatrix}\hat{U}_r - U_r \\ \hat{V} - V\end{pmatrix}\right] \tag{19}$$

The first term is simply the usual term that appears in the maximization of the log likelihood function, while the second term accounts for the fact that (U_r, V) are estimated rather than observed. Focusing on the second term, it follows that

$$\begin{aligned}
\sqrt{n}\mathbb{E}_n\left[\frac{\partial^2 \ell_r\left(\overline{U}_r, \overline{V}, \theta_r\right)}{\partial \theta \partial\left(u, v\right)}\begin{pmatrix}\hat{U}_r - U_r \\ \hat{V} - V\end{pmatrix}\right] &= \mathbb{E}_n\left[\frac{\partial^2 \ell_r\left(U_r, V, \theta_r\right)}{\partial \theta \partial\left(u, v\right)}\sqrt{n}\begin{pmatrix}\hat{U}_r - U_r \\ \hat{V} - V\end{pmatrix}\right] + o_P^*\left(1\right) \\
= \mathbb{E}_n\left[\frac{\partial^2 \ell_r\left(U_r, V, \theta_r\right)}{\partial \theta \partial\left(u, v\right)}\sqrt{n}\begin{pmatrix}\hat{U}_r - U_r \\ \hat{V} - V\end{pmatrix}\right] &- \mathbb{E}\left[\frac{\partial^2 \ell_r\left(U_r, V, \theta_r\right)}{\partial \theta \partial\left(u, v\right)}\sqrt{n}\begin{pmatrix}\hat{U}_r - U_r \\ \hat{V} - V\end{pmatrix}\right] \qquad (20) \\
+ \mathbb{E}\left[\frac{\partial^2 \ell_r\left(U_r, V, \theta_r\right)}{\partial \theta \partial\left(u, v\right)}\sqrt{n}\begin{pmatrix}\hat{U}_r - U_r \\ \hat{V} - V\end{pmatrix}\right] &+ o_P^*\left(1\right)
\end{aligned}$$

where the first equality follows by the extended continuous mapping theorem and the uniform consistency of $\hat{\beta}_r\left(\cdot\right)$ and $\hat{\gamma}_r\left(\cdot\right)$.

By the uniform law of large numbers, (20) is $o_P^*\left(1\right)$. Moreover,

$$\begin{aligned}
&\mathbb{E}\left[\frac{\partial^2 \ell_r\left(U_r, V, \theta_r\right)}{\partial \theta \partial\left(u, v\right)}\sqrt{n}\begin{pmatrix}\hat{U}_r - U_r \\ \hat{V} - V\end{pmatrix}\right] \\
&= \int_{\mathcal{Z}}\int_{\mathcal{U}^2}\frac{\partial^2 \ell_r\left(U_r, V, \theta_r\right)}{\partial \theta \partial\left(u, v\right)}\sqrt{n}\begin{pmatrix}\hat{U}_r - U_r \\ \hat{V} - V\end{pmatrix} dC_{UV|X_2}\left(U_r, V|x_2; \theta_r\right) dF_Z^r\left(z\right) \\
&\rightsquigarrow \int_{\mathcal{Z}}\int_{\mathcal{U}^2}\frac{\partial^2 \ell_r\left(u, v, \theta_r\right)}{\partial \theta \partial\left(u, v\right)} M_r\left(y, x_1, z\right)\mathbb{Z}_{L_r}\left(u, v\right) dC_{UV|X_2}\left(u, v|x_2; \theta_r\right) dF_Z^r\left(z\right)
\end{aligned}$$

by the extended continuous mapping theorem, Lemma 2, and Equation 17.

Then, one can apply the functional delta method to Equation 18 to obtain

$$\begin{aligned}
\sqrt{n}\left(\hat{\theta}_r-\theta_r\right) &= L_r^{-1}\left[\mathbb{G}_n\left[\frac{\partial \ell_r\left(U_r,V,\theta_r\right)}{\partial\theta}\right]\right.\\
&\quad\left.+\int_{\mathcal{Z}}\int_{\mathcal{U}^2}\frac{\partial^2\ell_r\left(U_r,V,\theta_r\right)}{\partial\theta\partial\left(u,v\right)}\sqrt{n}\begin{pmatrix}\hat{U}_r-U_r\\ \hat{V}-V\end{pmatrix}dC^r_{UV|X_2}\left(U_r,V|x_2;\theta_r\right)dF^r_Z\left(z\right)\right]+o^*_P\left(1\right)\\
&\rightsquigarrow L_r^{-1}\left[\mathbb{Z}_{\ell_r}+\int_{\mathcal{Z}}\int_{\mathcal{U}^2}\frac{\partial^2\ell_r\left(u,v,\theta_r\right)}{\partial\theta\partial\left(u,v\right)}M_r\left(y,x_1,z\right)\mathbb{Z}_{L_r}\left(u,v\right)dC^r_{UV|X_2}\left(u,v|x_2;\theta_r\right)dF^r_Z\left(z\right)\right]\\
&\equiv\mathbb{Z}_{\theta_r}
\end{aligned}$$

where

$$\mathbb{G}_n\left[\frac{\partial\ell_r\left(U_r,V,\theta_r\right)}{\partial\theta}\right]\rightsquigarrow\mathbb{Z}_{\ell_r}$$

The desired result follows immediately, and it holds uniformly in $m\in\mathcal{M}$, as $\mathcal{M}$ is a finite set. □

B.2 Asymptotic Distribution of the Fitted Values

Lemma 2. *Under Assumptions 3 to 10,*

$$\sqrt{n}\begin{pmatrix}\hat{U}_d-U_d\\ \hat{V}-V\end{pmatrix}=M_d\left(Y,X_{1i},Z\right)\sqrt{n}\begin{pmatrix}\hat{\beta}_d\left(U_d\right)-\beta_d\left(U_d\right)\\ \hat{\gamma}_d\left(V\right)-\gamma_d\left(V\right)\end{pmatrix}+o^*_P\left(1\right)\rightsquigarrow M_d\left(Y,X_1,Z\right)\begin{bmatrix}\mathbb{Z}_{\beta_d}\left(U_d\right)\\ \mathbb{Z}_{\gamma_d}\left(V\right)\end{bmatrix}$$

a stochastic process in metric space in $\ell^\infty\left(\mathcal{Y}\mathcal{X}_1\mathcal{Z}\mathcal{U}\mathcal{U}\mathcal{D}\right)$*, and where*

$$\begin{aligned}
M_d\left(y,x_1,z\right)&\equiv-\begin{pmatrix}g^d_Y\left(y|x,d\right)&0\\0&f^d_{X_1}\left(x_1|z,d\right)\end{pmatrix}\\
g^d_Y\left(y|x,d\right)&\equiv\frac{\partial}{\partial y}\int_0^1\mathbf{1}\left(x'\beta_d\left(u\right)\leq y\right)du\\
f^d_{X_1}\left(x_1|z,d\right)&\equiv\frac{\partial}{\partial x_1}\int_0^1\mathbf{1}\left(z'\gamma_d\left(v\right)\leq x_1\right)dv
\end{aligned}$$

Proof.

$$\sqrt{n}\left(\hat{U}_d - U_d\right) = \sqrt{n}\left(\int_{\epsilon}^{1-\epsilon} \mathbf{1}\left(X'\hat{\beta}_d(u) \leq Y\right) du - \int_{\epsilon}^{1-\epsilon} \mathbf{1}\left(X'\beta_d(u) \leq Y\right) du\right)$$
$$= -g_Y^d(Y|X, D=d) X' \sqrt{n}\left(\hat{\beta}_d(U_d) - \beta_d(U_d)\right) + o_P^*(1)$$

where we have used Proposition 2 in Chernozhukov et al. (2010). Note that $g_{Y^d}(y|x)$ can be interpreted as the conditional density of Y if there was no endogeneity. By the same argument, it can be shown that

$$\sqrt{n}\left(\hat{V} - V\right) = -f_{X_1}^d(X_1|Z, D=d) Z' \sqrt{n}\left(\hat{\gamma}_d(V) - \gamma_d(V)\right) + o_P^*(1)$$

Then, we can write

$$\sqrt{n}\begin{pmatrix} \hat{U}_d - U_d \\ \hat{V} - V \end{pmatrix} = M_d(Y, X_1, Z) \sqrt{n}\begin{bmatrix} \hat{\beta}_d(U_d) - \beta_d(U_d) \\ \hat{\gamma}_d(V) - \gamma_d(V) \end{bmatrix} + o_P^*(1) \rightsquigarrow M_d(Y, X_1, Z) \mathbb{Z}_{L_d}(U_d, V)$$

where the final result follows by the functional delta method and the asymptotic distribution of the IVQR and QR estimators shown in Lemma 1. □

B.3 Asymptotic Distribution of the Unfeasible Estimator

Lemma 3. *Let Assumptions 3 to 10 hold. The asymptotic distribution of* $\tilde{F}_{Y|Z}^m(y|z)$ *is given by*

$$\sqrt{n}\left(\tilde{F}_{Y|Z}^m(y|z) - F_{Y|Z}^{m,\epsilon}(y|z)\right) \rightsquigarrow \mathbb{Z}_{\tilde{F}_m|Z}(y, z)$$

a stochastic process in metric space $\ell^\infty(\mathcal{YZM})$*, where* $\mathbb{Z}_{\tilde{F}_m|Z}(y, z)$ *is a zero-mean tight Gaussian processes, defined in the proof, with a.s. uniformly continuous paths in* $\mathcal{YZM}$*.*

Proof. Note that $\tilde{F}_{Y|Z}^m(y|z)$ is a particular case of $\hat{F}_{Y|Z}^m(y|z)$ when the copula is known.

Therefore, using the same arguments as in Proposition 2, it is immediate to show that

$$\begin{aligned}
&\sqrt{n}\left(\tilde{F}^{m}_{Y|Z}\left(y|z\right)-F^{m,\epsilon}_{Y|Z}\left(y|z\right)\right)\rightsquigarrow\\
&\quad-\int_{\epsilon}^{1-\epsilon}f^{m}_{Y|ZV}\left(y|z,v\right)\left(z'\gamma_{q}\left(v\right),x_{2}\right)'\mathbb{Z}_{\beta_{p}}\left(C^{r,-1}_{U|VX_{2}}\left(F^{m}_{Y|ZV}\left(y|z,v\right)|v,x_{2};\hat{\theta}_{r}\right)\right)dv\\
&\quad-\int_{\epsilon}^{1-\epsilon}f^{m}_{Y|ZV}\left(y|z,v\right)\beta_{p,1}\left(C^{r,-1}_{U|VX_{2}}\left(F^{m}_{Y|ZV}\left(y|z,v\right)|v,x_{2};\hat{\theta}_{r}\right)\right)z'\mathbb{Z}_{\gamma_{q}}\left(v\right)dv\\
&\quad\equiv\mathbb{Z}_{\tilde{F}_{m}|Z}\left(y,z\right)
\end{aligned}$$

in $\ell^{\infty}\left(\mathcal{YZM}\right)$. The result holds uniformly in $m\in\mathcal{M}$ because $\mathcal{M}$ is a finite set. □

B.4 Hadamard Derivatives

Lemma 4. *Let* $F^{\epsilon}_{Y|Z}\left(y|z\right)\equiv\phi_{F}\left(F_{Y|ZV}\left(y|z,v\right)\right)=\epsilon+\int_{\epsilon}^{1-\epsilon}F_{Y|ZV}\left(y|z,v\right)dv$ *and* $F_{Y|ZV}\left(y|z,v,h_{n}\right)\equiv\phi_{F}\left(F_{Y|ZV}\left(y|z,v\right)+t_{n}h_{n}\left(y,z,v\right)\right)$*, where* $\phi_{F}:\mathbb{D}_{\phi_{F}}\subseteq\ell^{\infty}\left(\mathcal{YZU}\right)\mapsto\ell^{\infty}\left(\mathcal{YZ}\right)$*. Then, for all sequences* $\{h_{n}\}\subset\ell^{\infty}\left(\mathcal{YZU}\right)$ *and* $t_{n}\subset\mathbb{R}$ *such that* $t_{n}\to0$*,* $h_{n}\to h\in\mathcal{C}\left(\mathcal{YZU}\right)$ *as* $n\to\infty$*,* $F_{Y|ZV}\left(y|z,v\right)+t_{n}h_{n}\left(y,z,v\right)\in\mathbb{D}_{\phi_{F}}$ *for all* n*, the mapping*

$$\phi'_{F}\left(h\right)\equiv\int_{\epsilon}^{1-\epsilon}h\left(y,z,v\right)dv$$

is continuous, linear, and satisfies

$$\lim_{n\to\infty}\left\|\frac{\phi_{F}\left(F_{Y|ZV}\left(y|z,v\right)+t_{n}h_{n}\left(y,z,v\right)\right)-\phi_{F}\left(F_{Y|ZV}\left(y|z,v\right)\right)}{t_{n}}-\phi'_{F}\left(h\left(y,z,v\right)\right)\right\|_{\ell^{\infty}\left(\mathcal{YZ}\right)}=0$$

Proof.

$$\begin{aligned}
&\frac{\phi_F\left(F_{Y|ZV}(y|z,v)+t_nh_n(y,z,v)\right)-\phi_F\left(F_{Y|ZV}(y|z,v)\right)}{t_n}\\
&=\frac{\int_\epsilon^{1-\epsilon}\left[F_{Y|ZV}(y|z,v)+t_nh_n(y,z,v)-F_{Y|ZV}(y|z,v)\right]dv}{t_n}\\
&=\frac{\int_\epsilon^{1-\epsilon}t_nh_n(y,z,v)\,dv}{t_n}\\
&=\int_\epsilon^{1-\epsilon}h_n(y,z,v)\,dv\\
&\to\int_\epsilon^{1-\epsilon}h(y,z,v)\,dv=\phi_F'(h)
\end{aligned}$$

in $\ell^\infty(\mathcal{YZ})$ as $n\to\infty$. The last step follows by the pointwise convergence of $h(y,z,v)$ and the fact that it is uniformly bounded.[27] The desired result follows. □

Lemma 5. *Let* $F_{Y|ZV}(y|z,v)\equiv\phi_{g,c}\left(g(u,v,z),c_{UV|X_2}(u,v|x_2)\right)=\int_\epsilon^{1-\epsilon}\mathbf{1}(g(u,v,z)\le y)\cdot c_{UV|X_2}(u,v|x_2)\,du$ *and* $F_{Y|ZV}(y|z,v,h_n)\equiv\phi_{g,c}\left(g(u,v,z)+t_nh_{g,n}(u,v,z),c_{UV|X_2}(u,v|x_2)+t_nh_{c,n}\right)$*, where* $\phi_{g,c}:\mathbb{D}_{\phi_{g,c}}\subseteq\mathbb{D}\mapsto\ell^\infty(\mathcal{YZU})\equiv\mathbb{E}$*. Let the copula density,* $c_{UV|X_2}(u,v|x_2)$*, exist and be continuous with respect to* (u,v)*. Then, for all sequences* $\{h_n\}\subset\mathbb{D}$ *and* $t_n\subset\mathbb{R}$ *such that* $t_n\to0$*,* $h_n\equiv\left(h_{g,n}',h_{c,n}'\right)'\to\left(h_g',h_c'\right)'\equiv h\in\mathbb{D}_0$ *as* $n\to\infty$*,* $\left(g(u,v,z)+t_nh_{g,n}(u,v,z),c_{UV|X_2}(u,v|x_2)+t_nh_{c,n}(u,v,z)\right)\in\mathbb{D}_{\phi_{g,c}}$ *for all* n*, the mapping*

$$\begin{aligned}
\phi_{g,c}'(h)\equiv&-f_{Y|ZV}(y|z,v)\,h_g\left(C_{U|VX_2}^{-1}\left(F_{Y|ZV}(y|z,v)\,|v,x_2\right),v,z\right)\\
&+\int_\epsilon^{1-\epsilon}\mathbf{1}(g(u,v,z)\le y)\,h_c(u,v,z)\,du
\end{aligned}$$

is continuous, linear, and satisfies

$$\lim_{n\to\infty}\left\|\frac{\phi_{g,c}\left(g(u,v,z)+t_nh_{g,n}(u,v,z),c_{UV|X_2}(u,v|x_2)+t_nh_{c,n}(u,v,z)\right)}{t_n}\right.\left.\frac{-\phi_{g,c}\left(g(u,v,z),c_{UV|X_2}(u,v|x_2)\right)}{t_n}-\phi_{g,c}'(h(u,v,z))\right\|_{\mathbb{E}}=0$$

[27] See, *e.g.*, Theorem 1 in Pratt (1960).

Proof. The first step of the proof is partly based on the proof to Proposition 2 in Chernozhukov et al. (2010).

Let $B_\epsilon(x)$ be a closed ball of radius ϵ centered at x. For any $\delta > 0 \exists \epsilon > 0$: for $u \in B_\epsilon\left(C^{-1}_{U|VX_2}\left(F_{Y|ZV}(y|z,v)\,|v,x_2\right)\right)$ and for small enough $t_n \geq 0$

$$\mathbf{1}\left(g(u,v,z) + t_n h_n(u,v,z) \leq y\right) \leq \mathbf{1}\left(g(u,v,z) + t_n\left(h_g\left(C^{-1}_{U|VX_2}\left(F_{Y|ZV}(y|z,v)\,|v,x_2\right),v,z\right) - \delta\right) \leq y\right)$$

whereas $\forall u \notin B_\epsilon\left(C^{-1}_{U|VX_2}\left(F_{Y|ZV}(y|z,v)\,|v,x_2\right)\right)$

$$\mathbf{1}\left(g(u,v,z) + t_n h_n(u,v,z) \leq y\right) \leq \mathbf{1}\left(g(u,v,z) \leq y\right)$$

Therefore, for small enough $t_n \geq 0$

$$\begin{aligned}
&\frac{1}{t_n}\left[\int_\epsilon^{1-\epsilon} \mathbf{1}\left(g(u,v,z) + t_n h_n(u,v,z) \leq y\right) c_{UV|X_2}(u,v|x_2)\,du\right.\\
&\qquad \left. - \int_\epsilon^{1-\epsilon} \mathbf{1}\left(g(u,v,z) \leq y\right) c_{UV|X_2}(u,v|x_2)\,du\right] \qquad (21)\\
&\qquad \leq \frac{1}{t_n}\int_{B_\epsilon}\left[\mathbf{1}\left(g(u,v,z) + t_n\left(h_g\left(C^{-1}_{U|VX_2}\left(F_{Y|ZV}(y|z,v)\,|v,x_2\right),v,z\right) - \delta\right) \leq y\right)\right.\\
&\qquad \left. - \mathbf{1}\left(g(u,v,z) \leq y\right)\right] c_{UV|X_2}(u,v|x_2)\,du \qquad (22)
\end{aligned}$$

where B_ϵ is shorthand for $B_\epsilon\left(C^{-1}_{U|VX_2}\left(F_{Y|ZV}(y|z,v)\,|v,x_2\right)\right)$. By the change of variable $\tilde{y} = F^{-1}_{Y|ZV}\left(C_{U|VX_2}(u|v,x_2)\,|z,v\right)$, then $d\tilde{y} = {c_{UV|X_2}(u,v|x_2)\,du}/{f_{Y|ZV}\left(F^{-1}_{Y|ZV}\left(C_{U|VX_2}(u|v,x_2)\,|z,v\right)|z,v\right)}$. Thus, Equation 22 equals

$$\frac{1}{t_n}\int_{J\cap\left[y,\,y-t_n\left(h_g\left(C^{-1}_{U|VX_2}\left(F_{Y|ZV}(y|z,v)|v,x_2\right),v,z\right)-\delta\right)\right]} f_{Y|ZV}(\tilde{y}|z,v)\,d\tilde{y}$$

where J is the image of $B_\epsilon\left(C^{-1}_{U|VX_2}\left(F_{Y|ZV}(y|z,v)\,|v,x_2\right)\right)$ under $u \mapsto g(u,v,z)$. The change of variables is possible because $g(u,v,z)$ is a bijection between $B_\epsilon\left(C^{-1}_{U|VX_2}\left(F_{Y|ZV}(y|z,v)\,|v,x_2\right)\right)$ and J.

Fix $\epsilon > 0$ for $t_n \to 0$. Then, $J\cap\left[y, y - t_n\left(h_g\left(C^{-1}_{U|VX_2}\left(F_{Y|ZV}(y|z,v)\,|v,x_2\right),v,z\right) - \delta\right)\right] =$

$\left[y, y - t_n \left(h_g \left(C^{-1}_{U|VX_2}\left(F_{Y|ZV}(y|z,v)|v,x_2\right),v,z\right) - \delta\right)\right]$ and $f_{Y|ZV}(\tilde{y}|z,v) \rightarrow f_{Y|ZV}(y|z,v)$ as $F_{Y|ZV}(\tilde{y}|z,v) \rightarrow F_{Y|ZV}(y|z,v)$ in $\mathbb{E}$. Thus, Equation 22 is no greater than

$$-f_{Y|ZV}(y|z,v)\left(h_g\left(C^{-1}_{U|VX_2}\left(F_{Y|ZV}(y|z,v)|v,x_2\right),v,z\right) - \delta\right)$$

By a similar argument,

$$-f_{Y|ZV}(y|z,v)\left(h_g\left(C^{-1}_{U|VX_2}\left(F_{Y|ZV}(y|z,v)|v,x_2\right),v,z\right) + \delta\right)$$

bounds Equation 21 from below. Since $\delta > 0$ can be made arbitrarily small, the desired result follows.

To show that the result holds uniformly in $(y,z,v) \in K$, a compact subset of $\mathcal{YZU}$, we use Lemma B.4 in Chernozhukov et al. (2013). Take a sequence (y_t, z_t, v_t) in K that converges to $(y,z,v) \in K$. Then, the preceding argument applies to this sequence, since the function $f_{Y|ZV}(y|z,v)\, h_g\left(C^{-1}_{U|VX_2}\left(F_{Y|ZV}(y|z,v)|v,x_2\right),v,z\right)$ is uniformly continuous on K. This result follows by the assumed continuity of $h_g(u,v,z)$, $F_{Y|ZV}(y|z,v)$, and $f_{Y|ZV}(y|z,v)$ in all its arguments, as well as the compactness of K.

Now note that

$$\begin{aligned}
\frac{1}{t_n}\left[F_{Y|ZV}(y,z,v,h_n) - F_{Y|ZV}(y|z,v)\right] &= \frac{1}{t_n}\int_{\epsilon}^{1-\epsilon} \mathbf{1}\left(g(u,v,z) + t_n h_{g,n}(u,v,z) \leq y\right) c_{UV|X_2}(u,v|x_2)\, du \\
&- \frac{1}{t_n}\int_{\epsilon}^{1-\epsilon} \mathbf{1}\left(g(u,v,z) \leq y\right) c_{UV|X_2}(u,v|x_2)\, du \\
&+ \frac{1}{t_n}\int_{\epsilon}^{1-\epsilon} \mathbf{1}\left(g(u,v,z) + t_n h_{g,n}(u,v,z) \leq y\right)\left(c_{UV|X_2}(u,v|x_2) + t_n h_{c,n}(u,v,z)\right) du \\
&- \frac{1}{t_n}\int_{\epsilon}^{1-\epsilon} \mathbf{1}\left(g(u,v,z) + t_n h_{g,n}(u,v,z) \leq y\right) c_{UV|X_2}(u,v|x_2)\, du \\
&= \frac{1}{t_n}\int_{\epsilon}^{1-\epsilon} \left[\mathbf{1}\left(g(u,v,z) + t_n h_{g,n}(u,v,z) \leq y\right) - \mathbf{1}\left(g(u,v,z) \leq y\right)\right] c_{UV|X_2}(u,v|x_2)\, du \\
&+ \int_{\epsilon}^{1-\epsilon} \mathbf{1}\left(g(u,v,z) + t_n h_{g,n}(u,v,z) \leq y\right) h_{c,n}(u,v,z)\, du \\
&\rightarrow \phi'_{g,c}(h)
\end{aligned}$$

in $\mathbb{E}$, where the convergence follows by the first step of the proof, $h_n \to h$, and $t_n h_{g,n}(u,v,z) \to 0$ as $n \to \infty$. The desired result follows. $\square$

For the structural functions used in this paper, $g(u,v,z) = (z'\gamma_q(v), x_2)' \beta_p(u)$ and $c_{UV|X_2}(u,v|x_2) = c^r_{UV|X_2}(u,v|x_2;\theta_r)$, so $\mathbb{D} \equiv \ell^\infty(\mathcal{U})^{d_z} \times \ell^\infty(\mathcal{U})^{d_x} \times \mathbb{R}^{d_{\theta_r}} \times \ell^\infty(\mathcal{F})$ and $\mathbb{D}_0 \equiv \mathcal{C}(\mathcal{U})^{d_z} \times \mathcal{C}(\mathcal{U})^{d_x} \times \mathbb{R}^{d_{\theta_r}} \times \mathcal{C}(\mathcal{F})$.

Lemma 6. *Let* $g(u,v,z) \equiv \phi_{\beta,\gamma}(\beta(u), \gamma(v), z) = (z'\gamma(v), x_2)' \beta(u)$ *and* $g(u,v,z,h_n) \equiv \phi_{\beta,\gamma}(\beta(u) + t_n h_{\beta,n}(u,v,z), \gamma(v) + t_n h_{\gamma,n}(u,v,z), z)$, *where* $\phi_{\beta,\gamma} : \mathbb{D}_{\phi_{\beta,\gamma}} \subseteq \ell^\infty(\mathcal{U})^{d_z} \times \ell^\infty(\mathcal{U})^{d_x} \times \ell^\infty(\mathcal{Z}) \mapsto \ell^\infty(\mathcal{U}\mathcal{U}\mathcal{Z})$. *Then, for all sequences* $\{h_n\} \subset \ell^\infty(\mathcal{U})^{d_z} \times \ell^\infty(\mathcal{U})^{d_x} \times \ell^\infty(\mathcal{Z})$ *and* $t_n \subset \mathbb{R}$ *such that* $t_n \to 0$, $h_n \equiv \left(h'_{\beta,n}, h'_{\gamma,n}\right)' \to \left(h'_\beta, h'_\gamma\right)' \equiv h \in \mathbb{D}_0$ *as* $n \to \infty$, $(\beta(u) + t_n h_{\beta,n}(u,v,z), \gamma(v) + t_n h_{\gamma,n}(u,v,z), z) \in \mathbb{D}_{\phi_{\beta,\gamma}}$ *for all* n, *the mapping*

$$\phi'_{\beta,\gamma}(h) \equiv z' h_\gamma(u,v,z)\, \beta_1(u) + (z'\gamma(v), x_2)' h_\beta(u,v,z)$$

is continuous, linear, and satisfies

$$\lim_{n\to\infty} \left\| \frac{\phi_{\beta,\gamma}(\beta(u) + t_n h_{\beta,n}(u,v,z), \gamma(v) + t_n h_{\gamma,n}(u,v,z), z) - \phi_{\beta,\gamma}(\beta(u), \gamma(v), z)}{t_n} \right.$$
$$\left. - \phi'_{\beta,\gamma}(h(u,v,z)) \right\|_{\ell^\infty(\mathcal{U}\mathcal{U}\mathcal{Z})} = 0$$

Proof.

$$\begin{aligned}
&\frac{g(u,v,z,h_n) - g(u,v,z)}{t_n} \\
&= \frac{(z'(\gamma(v) + t_n h_{\gamma,n}(u,v,z)), x_2)' (\beta(u) + t_n h_{\beta,n}(u,v,z)) - (z'\gamma(v), x_2)' \beta(u)}{t_n} \\
&= \frac{(z'(\gamma(v) + t_n h_{\gamma,n}(u,v,z)), x_2)' (\beta(u) + t_n h_{\beta,n}(u,v,z)) - (z'\gamma(v), x_2)' (\beta(u) + t_n h_{\beta,n}(u,v,z))}{t_n} \\
&+ \frac{(z'\gamma(v), x_2)' (\beta(u) + t_n h_{\beta,n}(u,v,z)) - (z'\gamma(v), x_2)' \beta(u)}{t_n} \\
&= (z' h_{\gamma,n}(u,v,z), 0)' (\beta(u) + t_n h_{\beta,n}(u,v,z)) + (z'\gamma(v), x_2)' h_{\beta,n}(u,v,z) \\
&\to z' h_\gamma(u,v,z)\, \beta_1(u) + (z'\gamma(v), x_2)' h_\beta(u,v,z) = \phi'_{\beta,\gamma}(h)
\end{aligned}$$

in $\ell^\infty(\mathcal{UUZ})$ as $n \to \infty$. The convergence follows because $h_n \to h$, and $t_n h_{\beta,n}(u,v,z) \to 0$ as $n \to \infty$.

□

B.5 Argmax Process

Lemma 7. *(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) $Q_n(\cdot|\cdot) \xrightarrow{p} Q_\infty(\cdot|\cdot)$ in $\ell^\infty(K\times\Pi)$, where $Q_\infty(\cdot|\cdot)$ is continuous. Then $Z_n(\cdot) = Z_\infty(\cdot) + o_P(1)$ in $\ell^\infty(\Pi)$*

Proof. See Chernozhukov and Hansen (2006). □

B.6 Stochastic Expansion

Lemma 8. *Under Assumptions 3 to 9, the following statements hold:*

1. $\sup_{(\beta,\iota,\gamma,m)\in\mathcal{B}\times\mathcal{I}\times\mathcal{G}\times\mathcal{M}} \left|\mathbb{E}_n\left[\hat{G}_m(E,\beta,\iota,\gamma,\tau,\nu)\right] - \mathbb{E}\left[G_m(E,\beta,\iota,\gamma,\tau,\nu)\right]\right| = o_P(1)$

2. $\mathbb{G}_n R_m(E,\beta_p(\cdot),0,\gamma_q(\cdot\cdot),\cdot,\cdot\cdot) \rightsquigarrow \mathbb{Z}_{R_m}(\cdot,\cdot\cdot)$ *in* $\ell^\infty(\mathcal{UUM})$*, where* $\mathbb{Z}_{R_m}$ *is a Gaussian process with covariance function* $\Sigma_{R_m,\tilde{m}}(\tau,\nu,\tilde{\tau},\tilde{\nu})$ *defined below in the proof.*

   Furthermore, for any $\sup_{(\tau,\nu,m)\in\mathcal{U}\times\mathcal{U}\times\mathcal{M}} \left\|\left(\hat{\beta}_p(\tau),\hat{\iota}_p(\tau),\hat{\gamma}_q(\nu)\right) - (\beta_p(\tau),0,\gamma_q(\nu))\right\| = o_P(1)$, $\sup_{(\tau,\nu,m)\in\mathcal{U}\times\mathcal{U}\times\mathcal{M}} \left\|\mathbb{G}_n\hat{R}_m\left(E,\hat{\beta}_p(\tau),\hat{\iota}_p(\tau),\hat{\gamma}_q(\nu),\tau,\nu\right) - \mathbb{G}_n R_m(E,\beta_p(\tau),0,\gamma_q(\nu),\tau,\nu)\right\| = o_P(1)$

Proof. Let $\upsilon = (\beta,\iota,\gamma)$ and $\Upsilon = \mathcal{B}\times\mathcal{I}\times\mathcal{G}$, where $\mathcal{I}$ is a closed ball around 0. Define the class of functions $\mathcal{H}$ as

$$\mathcal{H} \equiv \left\{ h = (\Phi,\Psi,\upsilon,\tau,\nu,m) \mapsto \begin{bmatrix} \varphi_\tau\left(Y - X'\beta - \Phi(Z)'\iota\right)\Psi(Z)\mathbf{1}(D=p) \\ \varphi_\nu(X_1 - Z'\gamma)Z\mathbf{1}(D=q) \end{bmatrix} \upsilon \in \Upsilon, \Phi,\Psi \in \mathcal{F}_Z \right\}$$

where $\mathcal{F}_Z$ is defined in Assumption 9. The bracketing number of $\mathcal{F}_Z$, by Corollary 2.7.4 in van der Vaart and Wellner (1996) satisfies

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

for some $\delta' < 0$. Therefore, $\mathcal{F}_Z$ is Donsker with a constant envelope. By Corollary 2.7.4 in van der Vaart and Wellner (1996), the bracketing number of

$$\mathcal{J}_1 \equiv \left\{(\Phi, \upsilon) \to \left(X'\beta + \Phi\left(Z\right)'\iota\right), \upsilon \in \Upsilon, \Phi \in \mathcal{F}_Z\right\}$$

satisfies

$$\log N_{[\cdot]}\left(\epsilon, \mathcal{J}_1, L_2\left(P\right)\right) = O\left(\epsilon^{-\frac{\dim(x_1,z)}{\omega}}\right) = O\left(\epsilon^{-2-\delta''}\right)$$

for some $\delta'' < 0$. Also, by Corollary 2.7.4 in van der Vaart and Wellner (1996), the bracketing number of

$$\mathcal{J}_2 \equiv \left\{\upsilon \to \left(Z'\gamma\right), \upsilon \in \Upsilon\right\}$$

satisfies

$$\log N_{[\cdot]}\left(\epsilon, \mathcal{J}_2, L_2\left(P\right)\right) = O\left(\epsilon^{-2-\delta'}\right)$$

Since the indicator function is bounded and monotone, and the density functions $f_{Y|X_1Z}\left(y|x_1, z\right)$ and $f_{X_1|Z}\left(x_1|z\right)$ are bounded by Assumption 6, the bracketing number of

$$\mathcal{E} \equiv \left\{(\Phi, \upsilon) \to \mathbf{1}\left(Y < X'\beta + \Phi\left(X, Z\right)'\iota\right)\mathbf{1}\left(D = p\right) + \mathbf{1}\left(X_1 < Z'\gamma\right)\mathbf{1}\left(D = q\right), \upsilon \in \Upsilon, \Phi \in \mathcal{F}_Z\right\}$$

satisfy

$$\log N_{[\cdot]}\left(\epsilon,\mathcal{E},L_2\left(P\right)\right)=O\left(\epsilon^{-\frac{\dim(d,x_1,z)}{\omega}}\right)=O\left(\epsilon^{-2-\delta'''}\right)$$

for some δ''' such that $\delta''' < \delta''$. Since $\mathcal{E}$ has a constant envelope, it is Donsker. Then, $\mathcal{H}\equiv\mathcal{U}\times\mathcal{F}_Z+\mathcal{U}\times\mathcal{F}_Z-\mathcal{E}\times\mathcal{F}_Z$. Since $\mathcal{H}$ is Lipschitz over $(\mathcal{U}\times\mathcal{F}_Z\times\mathcal{E})$, it follows that it is Donsker by Theorem 2.10.6 in van der Vaart and Wellner (1996).

Define

$$h\equiv\left(\Phi,\Psi,\upsilon,\tau,\nu\right)\mapsto\mathbb{G}_n\begin{bmatrix}\varphi_\tau\left(\varepsilon-\Phi\left(Z\right)'\iota\right)\Psi\left(Z\right)\mathbf{1}\left(D=p\right)\\ \varphi_\nu\left(\eta\right)Z\mathbf{1}\left(D=q\right)\end{bmatrix}$$

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

$$\left(\tau,\nu\right)\mapsto\mathbb{G}_n\begin{bmatrix}\varphi_\tau\left(\varepsilon_p-\Phi_p\left(Z\right)'\iota_p\right)\Psi_p\left(Z\right)\mathbf{1}\left(D=p\right)\\ \varphi_\nu\left(\eta_q\right)Z\mathbf{1}\left(D=q\right)\end{bmatrix}$$

By the uniform Hölder continuity of $\left(\tau,\nu\right)\mapsto\left(\tau,\beta_p\left(\tau\right)',\Phi_p\left(\tau,Z\right)',\Psi_p\left(\tau,Z\right)',\nu,\gamma_q\left(\nu\right)'\right)'$ in $\left(\tau,\nu\right)$ with respect to the supremum norm, it is also Donsker in $\ell^\infty\left(\mathcal{H}\right)$. This, together with Assumption 5 implies

$$\mathbb{G}_n\begin{bmatrix}\varphi_\cdot\left(\varepsilon_p\left(\cdot\right)\right)\Psi_p\left(\cdot,Z\right)\mathbf{1}\left(D=p\right)\\ \varphi_{\cdot\cdot}\left(\eta_q\left(\cdot\cdot\right)\right)Z\mathbf{1}\left(D=q\right)\end{bmatrix}\rightsquigarrow\mathbb{Z}_{R_m}\left(\cdot,\cdot\cdot\right)$$

with covariance function

$$\Sigma_{R_{m\tilde{m}}}\left(\tau,\nu,\tilde{\tau},\tilde{\nu}\right)=\mathbb{E}\left[\mathbb{Z}_{R_m}\left(\tau,\nu\right)\mathbb{Z}_{R_{\tilde{m}}}\left(\tilde{\tau},\tilde{\nu}\right)'\right]\equiv\begin{bmatrix}\Sigma^{11}_{R_{m\tilde{m}}}\left(\tau,\tilde{\tau}\right) & \Sigma^{21}_{R_{m\tilde{m}}}\left(\tau,\tilde{\nu}\right)'\\ \Sigma^{21}_{R_{m\tilde{m}}}\left(\tilde{\tau},\nu\right) & \Sigma^{22}_{R_{m\tilde{m}}}\left(\nu,\tilde{\nu}\right)\end{bmatrix}$$

where

$$
\begin{aligned}
\Sigma_{R_{m\tilde{m}}}^{11}(\tau,\tilde{\tau}) &= (\tau\wedge\tilde{\tau}-\tau\tilde{\tau})\,\mathbb{E}\left[\Psi_p(\tau)\,\Psi_p(\tilde{\tau})'\,\mathbf{1}(D=p)\right]\mathbf{1}(p=\tilde{p}) \\
&= \frac{1}{\lambda_p}(\tau\wedge\tilde{\tau}-\tau\tilde{\tau})\,\mathbb{E}\left[\Psi_p(\tau)\,\Psi_p(\tilde{\tau})'\,|D=p\right]\mathbf{1}(p=\tilde{p}) \\
\Sigma_{R_{m\tilde{m}}}^{21}(\tilde{\tau},\nu) &= \mathbb{E}\left[\left(\mathbf{1}\left(Y\le X'\beta_q(\tilde{\tau})\right)\mathbf{1}\left(X_1\le Z'\gamma_q(\nu)\right)-\tilde{\tau}\nu\right)Z\Psi_q(\tilde{\tau})'\,\mathbf{1}(D=q)\right]\mathbf{1}(q=\tilde{p}) \\
&= \frac{1}{\lambda_q}\mathbb{E}\left[\left(\mathbf{1}\left(Y\le X'\beta_q(\tilde{\tau})\right)\mathbf{1}\left(X_1\le Z'\gamma_q(\nu)\right)-\tilde{\tau}\nu\right)Z\Psi_q(\tilde{\tau})'\,|D=q\right]\mathbf{1}(q=\tilde{p}) \\
\Sigma_{R_{m\tilde{m}}}^{22}(\nu,\tilde{\nu}) &= (\nu\wedge\tilde{\nu}-\nu\tilde{\nu})\,\mathbb{E}\left[ZZ'\mathbf{1}(D=q)\right]\mathbf{1}(q=\tilde{q}) \\
&= \frac{1}{\sqrt{\lambda_q\lambda_{\tilde{q}}}}(\nu\wedge\tilde{\nu}-\nu\tilde{\nu})\,\mathbb{E}\left[ZZ'|D=q\right]\mathbf{1}(q=\tilde{q})
\end{aligned}
$$

Let $h,h'\in\mathcal{H}$. Since $\hat{\Psi}_p(\cdot)\overset{p}{\to}\Psi_p(\cdot)$, and $\hat{\Phi}_p(\cdot)\overset{p}{\to}\Phi_p(\cdot)$ uniformly over compact sets and $\hat{\upsilon}_m(\tau,\nu)\overset{P}{\to}\upsilon_m(\tau,\nu)$ uniformly in (τ,ν), it follows by Assumptions 8 and 9 that $\delta_n\equiv\sup_{(\tau,\nu)\in\mathcal{U}\times\mathcal{U}}\xi\left(h'(\tau,\nu),h(\tau,\nu)\right)|_{h'=\hat{h}}\overset{P}{\to}0$, where ξ is the $L_2(P)$ pseudometric on $\mathcal{H}$:

$$
\xi\left(h,\tilde{h}\right)\equiv\sqrt{\mathbb{E}\left\|\begin{bmatrix}\varphi_\tau\left(\varepsilon_p-\Phi_p(Z)'\iota_p\right)\Psi_p(Z)\,\mathbf{1}(D=p)\\ \varphi_\nu(\eta_q)\,Z\mathbf{1}(D=q)\end{bmatrix}-\begin{bmatrix}\varphi_{\tilde{\tau}}\left(\tilde{\varepsilon}_p-\tilde{\Phi}_p(Z)'\tilde{\iota}\right)\tilde{\Psi}_p(Z)\,\mathbf{1}(D=p)\\ \varphi_{\tilde{\nu}}(\tilde{\eta}_q)\,Z\mathbf{1}(D=q)\end{bmatrix}\right\|^2}
$$

As $\delta_n\overset{P}{\to}0$

$$
\begin{aligned}
\sup_{(\tau,\nu)\in\mathcal{U}\times\mathcal{U}}&\left\|\mathbb{G}_n\begin{bmatrix}\varphi_\tau\left(\hat{\varepsilon}_p(\tau)-\hat{\Phi}_p(\tau,Z)'\hat{\iota}_p(\tau)\right)\hat{\Psi}_p(\tau,Z)\,\mathbf{1}(D=p)\\ \varphi_\nu\left(\hat{\eta}_q(\nu)\right)Z\mathbf{1}(D=q)\end{bmatrix}\right.\\
&\left.-\,\mathbb{G}_n\begin{bmatrix}\varphi_\tau\left(\varepsilon_p(\tau)-\Phi_p(\tau,Z)'\iota_p(\tau)\right)\Psi_p(\tau,Z)\,\mathbf{1}(D=p)\\ \varphi_\nu\left(\eta_q(\nu)\right)Z\mathbf{1}(D=q)\end{bmatrix}\right\|\\
&\le\sup_{\xi\left(\tilde{h},h\right)\le\delta_n,\tilde{h},h\in\mathcal{H}}\left\|\mathbb{G}_n\begin{bmatrix}\varphi_\tau\left(\varepsilon_p-\tilde{\Phi}_p(Z)'\tilde{\iota}_p\right)\tilde{\Psi}_p(Z)\,\mathbf{1}(D=p)\\ \varphi_\nu(\eta_q)\,Z\mathbf{1}(D=q)\end{bmatrix}\right.\\
&\left.-\,\mathbb{G}_n\begin{bmatrix}\varphi_\tau\left(\varepsilon_p-\Phi_p(Z)'\iota_p\right)\Psi_p(Z)\,\mathbf{1}(D=p)\\ \varphi_\nu(\eta_q)\,Z\mathbf{1}(D=q)\end{bmatrix}\right\|=o_P(1)
\end{aligned}
$$

by stochastic equicontinuity of $h \mapsto \mathbb{G}_n \begin{bmatrix} \varphi_\tau \left(\varepsilon_p - \Phi_p (Z)' \iota_p\right) \Psi_p (Z) \mathbf{1} (D = p) \\ \varphi_\nu (\eta_q) Z \mathbf{1} (D = q) \end{bmatrix}$, which proves claim 2. Note that the result is holds uniformly in $m \in \mathcal{M}$ because $\mathcal{M}$ is a finite set. To prove claim 1, define

$$\mathcal{A} \equiv \left\{ (\Phi, \beta, \iota, \gamma, \tau, \nu) \mapsto \begin{bmatrix} \rho_\tau \left(\varepsilon - \Phi (Z)' \iota\right) \mathbf{1} (D = p) \\ \rho_\nu (\eta) \mathbf{1} (D = q) \end{bmatrix} \right\}$$

This class of functions is uniformly Lipschitz over $(\mathcal{F}_Z \times \mathcal{B} \times \mathcal{I} \times \mathcal{G} \times \mathcal{U} \times \mathcal{U})$ and bounded by Assumption 4, so by Theorem 2.10.6 in van der Vaart and Wellner (1996), $\mathcal{A}$ is Donsker. Therefore, the following Uniform Law of Large Numbers hold:

$$\sup_{h \in \mathcal{H}} \left| \mathbb{E}_n \begin{bmatrix} \rho_\tau \left(\varepsilon_p - \Phi (Z)' \iota\right) \mathbf{1} (D = p) \\ \rho_\nu (\eta_q) \mathbf{1} (D = q) \end{bmatrix} - \mathbb{E} \begin{bmatrix} \rho_\tau \left(\varepsilon_p - \Phi (Z)' \iota\right) \mathbf{1} (D = p) \\ \rho_\nu (\eta_q) \mathbf{1} (D = q) \end{bmatrix} \right| \xrightarrow{P} 0$$

which gives,

$$\sup_{(\beta,\iota,\gamma,\tau,\nu)} \left| \mathbb{E}_n \begin{bmatrix} \rho_\tau \left(\varepsilon_p - \tilde{\Phi}_p (\tau, Z)' \iota\right) \mathbf{1} (D = p) \\ \rho_\nu (\eta_q) \mathbf{1} (D = q) \end{bmatrix} - \mathbb{E} \begin{bmatrix} \rho_\tau \left(\varepsilon_p - \tilde{\Phi}_p (\tau, Z)' \iota\right) \mathbf{1} (D = p) \\ \rho_\nu (\eta_q) \mathbf{1} (D = q) \end{bmatrix} \right|_{\tilde{\Phi}_p = \hat{\Phi}_p} \xrightarrow{P} 0$$

By uniform consistency of $\hat{\Phi}_p (\cdot)$ and Assumption 9, I have that

$$\sup_{(\beta,\iota,\gamma,\tau,\nu)} \left| \mathbb{E} \begin{bmatrix} \rho_\tau \left(\varepsilon_p - \tilde{\Phi}_p (\tau, Z)' \iota\right) \mathbf{1} (D = p) \\ \rho_\nu (\eta_q) \mathbf{1} (D = q) \end{bmatrix} - \mathbb{E} \begin{bmatrix} \rho_\tau \left(\varepsilon_p - \Phi_p (\tau, Z)' \iota \mathbf{1} (D = p)\right) \\ \rho_\nu (\eta_q) \mathbf{1} (D = q) \end{bmatrix} \right|_{\tilde{\Phi}_p = \hat{\Phi}_p} \xrightarrow{P} 0$$

These two results imply claim 1. □

B.7 Uniform Consistency of Nonparametric Copula Estimators

Lemma 9. *Let Assumptions 4 to 9, and 12 hold, and $\left(\hat{U}_{d,i}, \hat{V}_i\right)$ be uniformly consistent estimators for $(U_{d,i}, V_i)$. Then, $\sup_{u,v,d} \left| \check{C}^d_{UV} (u, v) - C^d_{UV} (u, v) \right| = O_P \left(\frac{1}{\sqrt{n}} \right)$.*

Proof. Define $\tilde{C}_{UV}^{d}(u,v) \equiv \frac{n}{n_d}\mathbb{E}_n\left[\mathbf{1}(U_{d,i} \leq u)\mathbf{1}(V_i \leq v)\mathbf{1}(D_i = d)\right]$ and split the proof into showing the probability limit of $\tilde{C}_{UV}^{d}(u,v)$ and $\check{C}_{UV}^{d}(u,v)$ is the same at the $\sqrt{n}$ rate, and then that $\tilde{C}_{UV}^{d}(u,v)$ is a consistent estimator of $C_{UV}^{d}(u,v)$ at the faster parametric rate.[28]

By the triangular inequality,

$$\begin{aligned}\left|\check{C}_{UV}^{d}(u,v) - \tilde{C}_{UV}^{d}(u,v)\right| &= \frac{n}{n_d}\mathbb{E}_n\left|\mathbf{1}\left(\hat{U}_{d,i} \leq u\right)\mathbf{1}\left(\hat{V}_i \leq v\right) - \mathbf{1}(U_{d,i} \leq u)\mathbf{1}(V_i \leq v)\right|\mathbf{1}(D_i = d)\\ &\leq \frac{n}{n_d}\mathbb{E}_n\left|\mathbf{1}\left(\hat{U}_{d,i} \leq u\right) - \mathbf{1}(U_{d,i} \leq u)\right|\mathbf{1}(D_i = d)\\ &+ \frac{n}{n_d}\mathbb{E}_n\left|\mathbf{1}\left(\hat{V}_i \leq v\right) - \mathbf{1}(V_i \leq v)\right|\mathbf{1}(D_i = d)\end{aligned}$$

Define the following function:

$$\sigma(U,u;w) \equiv \begin{cases} 1 & U \leq u \\ 1 - w(U-u) & u < U \leq u + \frac{1}{w} \\ 0 & U > u + \frac{1}{w} \end{cases}$$

for some $w > 0$. Note that, as $w \to \infty$, $\sigma(U,u;w) \to \mathbf{1}(U \leq u)$. Now write

$$\begin{aligned}&\sup_{u,d}\mathbb{E}_n\left|\mathbf{1}\left(\hat{U}_{d,i} \leq u\right) - \mathbf{1}(U_{d,i} \leq u)\right|\mathbf{1}(D_i = d)\\ &\leq \sup_{u,d}\mathbb{E}_n\left|\mathbf{1}\left(\hat{U}_{d,i} \leq u\right) - \sigma\left(\hat{U}_{d,i},u;w\right)\right|\mathbf{1}(D_i = d) \qquad (23)\\ &+ \sup_{u,d}\mathbb{E}_n\left|\sigma\left(\hat{U}_{d,i},u;w\right) - \sigma(U_{d,i},u;w)\right|\mathbf{1}(D_i = d) \qquad (24)\\ &+ \sup_{u,d}\mathbb{E}_n\left|\sigma(U_{d,i},u;w) - \mathbf{1}(U_{d,i} \leq u)\right|\mathbf{1}(D_i = d) \qquad (25)\end{aligned}$$

For any w, (23) equals $\sup_{u,d}\mathbb{E}_n\mathbf{1}\left(u < \hat{U}_{d,i} \leq u + \frac{1}{w}\right)\mathbf{1}(D_i = d) \xrightarrow{P} \frac{1}{w}\mathbb{P}(D_i = d)$ by Lemma 2 and the ULLN; (24)$\xrightarrow{P} 0$ by the extended continuous mapping theorem and Lemma 2; (25)

[28] Notice that it is not possible to apply the extended continuous mapping theorem to conclude that if $\hat{U}_{d,i} \xrightarrow{P} U_{d,i}$ and $\hat{V}_i \xrightarrow{P} V_i$, then $\mathbf{1}\left(\hat{U}_{d,i} \leq u\right) \xrightarrow{P} \mathbf{1}(U_{d,i} \leq u)$ or $\mathbf{1}\left(\hat{V}_i < v\right) \xrightarrow{P} \mathbf{1}(V_i < v)$ uniformly in (u,v). Hence, a different argument is required for the proof.

equals $\sup_{u,d}\mathbb{E}_n\mathbf{1}\left(u<U_{d,i}\leq u+\frac{1}{w}\right)\mathbf{1}\left(D_i=d\right)\xrightarrow{P}\frac{1}{w}\mathbb{P}\left(D_i=d\right)$ by the ULLN. Therefore,

$$\begin{aligned}&\sup_{u,d}\mathbb{E}_n\left|\mathbf{1}\left(\hat{U}_{d,i}\leq u\right)-\mathbf{1}\left(U_{d,i}\leq u\right)\right|\mathbf{1}\left(D_i=d\right)\\&\quad=\lim_{w\to\infty}\sup_{u,d}\mathbb{E}_n\left|\mathbf{1}\left(\hat{U}_{d,i}\leq u\right)-\mathbf{1}\left(U_{d,i}\leq u\right)\right|\mathbf{1}\left(D_i=d\right)=O_P\left(\frac{1}{\sqrt{n}}\right)\end{aligned}$$

Hence, by Assumption 5, $\sup_{u,d}\frac{n}{n_d}\mathbb{E}_n\left|\mathbf{1}\left(\hat{U}_{d,i}\leq u\right)-\mathbf{1}\left(U_{d,i}\leq u\right)\right|\mathbf{1}\left(D_i=d\right)=O_P\left(\frac{1}{\sqrt{n}}\right)$.

Using the same argument, we can conclude that $\sup_{v,d}\frac{n}{n_d}\mathbb{E}_n\left|\mathbf{1}\left(\hat{V}_{d,i}\leq v\right)-\mathbf{1}\left(V_{d,i}\leq v\right)\right|\cdot\mathbf{1}\left(D_i=d\right)=O_P\left(\frac{1}{\sqrt{n}}\right)$. As a consequence, $\sup_{u,v,d}\left|\check{C}^d_{UV}\left(u,v\right)-\tilde{C}^d_{UV}\left(u,v\right)\right|=O_P\left(\frac{1}{\sqrt{n}}\right)$.

As for the second step, consider the class $\mathcal{C}_{UVD}\equiv\{\{(x_1,x_2,x_3):x_1\leq u,x_2\leq v,x_3=d\},u,v\in[0,1],d\in\mathcal{D}\}$. This is a VC class with VC dimension $V\left(\mathcal{C}_{UVD}\right)=3$. Therefore, by Theorem 2.6.4 in van der Vaart and Wellner (1996), its covering number is bounded: $N\left(\epsilon,\mathcal{C}_{UVD},L_2\left(P\right)\right)\leq 3\cdot 4^3\kappa e^3\epsilon^{-4}<\infty$ for some constant κ and $0<\epsilon<1$. By Theorem 2.5.2 in van der Vaart and Wellner (1996), it is $\mathbb{P}$-Donsker, so $\sqrt{n}\sup_{u,v}\left|\tilde{C}^d_{UV}\left(u,v\right)-C^d_{UV}\left(u,v\right)\right|=O_P\left(1\right)$. Moreover, the result holds uniformly in $d\in\mathcal{D}$ because $\mathcal{D}$ is a finite set. Hence,

$$\begin{aligned}\sup_{u,v,d}\left|\check{C}^d_{UV}\left(u,v\right)-C^d_{UV}\left(u,v\right)\right|&\leq\sup_{u,v,d}\left|\check{C}^d_{UV}\left(u,v\right)-\tilde{C}^d_{UV}\left(u,v\right)\right|+\sup_{u,v,d}\left|\tilde{C}^d_{UV}\left(u,v\right)-C^d_{UV}\left(u,v\right)\right|\\&=O_P\left(\frac{1}{\sqrt{n}}\right)\end{aligned}$$

□

Consider the estimator $\check{C}^d_{UV}\left(u,v\right)$ defined by Equation 12. This estimator can be seen as the integration over $[0,1]$ of a nonparametric estimator of the conditional copula distribution $C^d_{U|V}\left(u|v\right)$, given by

$$\check{C}^d_{U|V}\left(u|v\right)\equiv\frac{H_n+1}{n_d}\Sigma_{i=1}^n\mathbf{1}\left(\hat{U}_{d,i}\leq u\right)\mathbf{1}\left(\underline{\nu}\left(v\right)\leq\hat{V}_i<\overline{\nu}\left(v\right)\right)\mathbf{1}\left(D_i=d\right)$$

where H_n denotes the number of evenly spaced quantiles that are used in the estimation of the quantile process $h\left(z_1,x_2,v\right)$, denoted by $0=\nu_0,\nu_1,...,\nu_{H_n},\nu_{H_n+1}=1$, and $\overline{\nu}\left(v\right)$ and $\underline{\nu}\left(v\right)$ are defined as $\{\max_i\nu_i:\nu_i<v\}$ and $\{\min_i\nu_i:\nu_i\geq v\}$. It can be checked that

$\check{C}^d_{UV}(u,v) = \frac{1}{H_n+1}\Sigma_{h=0}^{(H_n+1)\underline{\nu}(v)}\check{C}^d_{U|V}(u,\nu_h)$. Geometrically, I am splitting the $[0,1]$ interval into H_n+1 intervals of equal length, and each V_i belongs to one of them with probability $\frac{1}{H_n+1}$, since $V_i \sim U(0,1)$. H_n is to be interpreted as the inverse of the bandwidth of this kernel estimator, and $H_n \to \infty$ as $n \to \infty$. For each of the cells, compute the conditional distribution of the copula. The following lemma establishes the uniform consistence of this conditional estimator of the copula, which unlike the conditional estimator, converges at a rate slower than $\sqrt{n}$.

Lemma 10. *Let Assumptions 4 to 9, and 12 hold,* $\left(\hat{U}_{d,i},\hat{V}_i\right)$ *converge uniformly in probability to* $(U_{d,i},V_i)$ *at a rate* $\sqrt{n}$, $H_n \to \infty$, $\frac{\sqrt{n}}{H_n} \to \infty$, *and* $\frac{na_n}{\log(n)} \to \infty$ *as* $n \to \infty$, *and* $0 < a_n < \frac{1}{H_n+1} < b_n < 1$ *for sequences* a_n, b_n *such that* $\frac{na_n}{\log(n)} \to \infty$ *and* $b_n \to 0$ *as* $n \to \infty$.

$$\sup_{u,v,d}\left|\check{C}^d_{U|V}(u|v) - C^d_{U|V}(u|v)\right| = O_P\left(\frac{H_n}{\sqrt{n}}\right)$$

Proof. The proof is split into two steps: first I show that $\check{C}^d_{U|V}(u|v)$ and the unfeasible estimator $\tilde{C}^d_{U|V}(u|v) \equiv \frac{H_n+1}{n_d}\Sigma_{i=1}^n \mathbf{1}(U_{d,i} \le u)\mathbf{1}(\underline{\nu}(v) \le V_i < \overline{\nu}(v))\mathbf{1}(D_i = d)$ converge to the same limit at the $\frac{H_n}{\sqrt{n}}$ rate of convergence, and then I show the consistency of $\tilde{C}^d_{U|V}(u|v)$ at the usual parametric rate of convergence.

As shown in Lemma 9, $\sup_{u,d}\frac{n}{n_d}\mathbb{E}_n\left|\mathbf{1}\left(\hat{U}_{d,i} \le u\right) - \mathbf{1}(U_{d,i} \le u)\right|\mathbf{1}(D_i = d) = O_P\left(\frac{1}{\sqrt{n}}\right)$. Define the following function:

$$\tilde{\sigma}(V,\underline{v},\overline{v};w) \equiv \begin{cases} 0 & V < \underline{v} - \frac{1}{w} \\ 1 + w(V - \underline{v}) & \underline{v} - \frac{1}{w} < V \le \underline{v} \\ 1 & \underline{v} \le V \le \overline{v} \\ 1 - w(V - \overline{v}) & \overline{v} < V \le \overline{v} + \frac{1}{w} \\ 0 & V > \overline{v} + \frac{1}{w} \end{cases}$$

for some $w > 0$. As $w \to \infty$, $\tilde{\sigma}(V,\underline{v},\overline{v};w) \to \mathbf{1}(\underline{v} \le V \le \overline{v})$. Now write

$$
\begin{aligned}
&\sup_{v,d} \mathbb{E}_n \left| \mathbf{1}\left(\underline{\nu}(v) \le \hat{V}_i \le \overline{\nu}(v)\right) - \mathbf{1}\left(\underline{\nu}(v) \le V_i \le \overline{\nu}(v)\right) \right| \mathbf{1}(D_i = d) \\
&\qquad \le \sup_{v,d} \mathbb{E}_n \left| \mathbf{1}\left(\underline{\nu}(v) \le \hat{V}_i \le \overline{\nu}(v)\right) - \tilde{\sigma}\left(\hat{V}_i, \underline{\nu}(v), \overline{\nu}(v); w\right) \right| \mathbf{1}(D_i = d) && (26) \\
&\qquad + \sup_{v,d} \mathbb{E}_n \left| \tilde{\sigma}\left(\hat{V}_i, \underline{\nu}(v), \overline{\nu}(v); w\right) - \tilde{\sigma}\left(V_i, \underline{\nu}(v), \overline{\nu}(v); w\right) \right| \mathbf{1}(D_i = d) && (27) \\
&\qquad + \sup_{v,d} \mathbb{E}_n \left| \tilde{\sigma}\left(V_i, \underline{\nu}(v), \overline{\nu}(v); w\right) - \mathbf{1}\left(\underline{\nu}(v) \le V_i \le \overline{\nu}(v)\right) \right| \mathbf{1}(D_i = d) && (28)
\end{aligned}
$$

Using the same arguments as in Lemma 9, it follows that

$$
\begin{aligned}
&\sup_{v,d} \mathbb{E}_n \left| \mathbf{1}\left(\underline{\nu}(v) \le \hat{V}_i \le \overline{\nu}(v)\right) - \mathbf{1}\left(\underline{\nu}(v) \le V_i \le \overline{\nu}(v)\right) \right| \mathbf{1}(D_i = d) \\
&\qquad = \lim_{w \to \infty} \sup_{v,d} \mathbb{E}_n \left| \mathbf{1}\left(\underline{\nu}(v) \le \hat{V}_i \le \overline{\nu}(v)\right) - \mathbf{1}\left(\underline{\nu}(v) \le V_i \le \overline{\nu}(v)\right) \right| \mathbf{1}(D_i = d) = O_P\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}
$$

Hence, by Assumption 5, $\sup_v \sqrt{n} \frac{n}{n_d} \mathbb{E}_n \left| \mathbf{1}\left(\underline{\nu}(v) \le \hat{V}_i \le \overline{\nu}(v)\right) - \mathbf{1}\left(\underline{\nu}(v) \le V_i \le \overline{\nu}(v)\right) \right| \cdot \mathbf{1}(D_i = d) = O_P(1)$. The result holds uniformly in $d \in \mathcal{D}$ because $\mathcal{D}$ is a finite set.

Consequently, using the results in Lemma 9 yields

$$
\begin{aligned}
\sup_{u,v,d} \left| \check{C}^d_{U|V}(u|v) - \tilde{C}^d_{U|V}(u|v) \right| &= \frac{H_n + 1}{n} \frac{n}{n_d} \left| \sum_{i=1}^{n} \mathbf{1}\left(\hat{U}_{d,i} \le u\right) \mathbf{1}\left(\underline{\nu}(v) \le \hat{V}_i < \overline{\nu}(v)\right) \right. \\
&\quad \left. - \sum_{i=1}^{n} \mathbf{1}\left(U_{d,i} \le u\right) \mathbf{1}\left(\underline{\nu}(v) \le V_i < \overline{\nu}(v)\right) \right| \\
&\le (H_n + 1) \frac{n}{n_d} \mathbb{E} \left| \mathbf{1}\left(\hat{U}_{d,i} \le u\right) - \mathbf{1}\left(U_{d,i} \le u\right) \right| \\
&\quad + (H_n + 1) \frac{n}{n_d} \mathbb{E} \left| \mathbf{1}\left(\underline{\nu}(v) \le \hat{V}_i < \overline{\nu}(v)\right) - \mathbf{1}\left(\underline{\nu}(v) \le V_i < \overline{\nu}(v)\right) \right| \\
&\le (H_n + 1) O_P\left(\frac{1}{\sqrt{n}}\right) = O_P\left(\frac{H_n}{\sqrt{n}}\right)
\end{aligned}
$$

Consider the class $\mathcal{C}'_{UVD} \equiv \{\{(x_1, x_2, x_3) : x_1 \le u, v_l \le x_2 < v_u, x_3 = d\}, u, v_l, v_u \in [0,1], v_l < v_u, d \in \mathcal{D}\}$. It is a VC class with VC dimension $V(\mathcal{C}'_{UVD}) = 4$. Hence, by Theorem 2.6.4 in van der Vaart and Wellner (1996), its covering number is bounded: $N(\epsilon, \mathcal{C}'_{UVD}, L_2(P)) \le$

$4^5\kappa e^4\epsilon^{-6} < \infty$ for some constant κ and $0 < \epsilon < 1$. By Corollary 1 in Einmahl et al. (2005)

$$\lim_{n\to\infty} \sup_{a_n \le \frac{1}{H_n+1} \le b_n} \sup_{(u,v)\in[0,1]} \left| \tilde{C}^d_{U|V}(u|v) - C^d_{U|V}(u|v) \right| = 0$$

This result implies that $\sup_{(u,v)\in[0,1]} \left| \tilde{C}^d_{U|V}(u|v) - C^d_{U|V}(u|v) \right| = o_P(1)$. By the triangle inequality,

$$\left| \check{C}^d_{U|V}(u|v) - C^d_{U|V}(u|v) \right| \le \left| \check{C}^d_{U|V}(u|v) - \tilde{C}^d_{U|V}(u|v) \right| + \left| \tilde{C}^d_{U|V}(u|v) - C^d_{U|V}(u|v) \right|$$

Because $\mathcal{D}$ is a finite set, it follows that $\sup_{u,v,d} \left| \check{C}^d_{U|V}(u|v) - C^d_{U|V}(u|v) \right| = O_P\left(\frac{H_n}{\sqrt{n}}\right)$. $\square$

Some remarks are in order. First, this lemma limits the rate of growth of the number of cells of the unit interval, which has to satisfy $H_n = o_P\left(\min\left\{\frac{n}{\log(n)}, \sqrt{n}\right\}\right)$. This, however, does not imply that the estimator achieves the maximum possible convergence rate. The kernel $K(V_i, v, H_n) \equiv (H_n+1)\,\mathbf{1}(\underline{\nu}(v) \le V_i < \overline{\nu}(v))$ is not symmetric around zero, which would improve the convergence rate of the estimator. Furthermore, it depends on two nonlinear functions of v: $\underline{\nu}(v)$ and $\overline{\nu}(v)$, which means that one cannot use a Taylor expansion around v to establish the asymptotic normality of this estimator.

Second, one could regularize the indicator function as suggested in Section 3.3 to be able to apply the extended continuous mapping theorem. Thus, it would be possible to estimate $C^d_{U|V}(u|v)$ by $\grave{C}^d_{U|V}(u|v) = \frac{1}{nh_n}\Sigma_{i=1}^n \grave{f}(U_{d,i}, u, n)\,\grave{K}\left(\frac{V_i - v}{h_n}\right)$, where $\grave{f}(U_{d,i}, u, n)$ is the regularized indicator function, and $\grave{K}\left(\frac{V_i - v}{h_n}\right)$ is a kernel function that is continuous in its argument and that, in order to improve the convergence rate, is symmetric around zero. Studying the asymptotic properties of such estimator is beyond the scope of this paper.

C Nonparametric First Stage Equation

Assume that the first stage equation is an unknown function, rather than linear in the covariates as in Assumption 3. The IVQR estimator of $\hat{\beta}(\cdot)$ does not require the linearity of h to be consistent and asymptotically Gaussian the $\sqrt{n}$ convergence rate. However, the estimators of the counterfactual distributions $\hat{F}_Y^{cf}(y)$ and $\check{F}_Y^{cf}(y)$ use an estimator of h as an input, which has an impact on their asymptotic convergence.

Let $\hat{h} \equiv \hat{h}(z_1, x_2, v)$ denote the estimator of the CQF and $\hat{F}_Y^{cf}\left(y, \hat{h}\right)$ the estimator of the counterfactual distribution. Newey (1991, 1994), Andrews (1994), or Ichimura and Newey (2022) have already studied such estimators, establishing conditions under which they are consistent at the $\sqrt{n}$ convergence rate. One of these conditions is the asymptotic linearity of the semiparametric estimator, *i.e.*, it can be rewritten as the sum of the sample average of an influence function whose variance is finite and a stochastic term that is $O_P\left(\frac{1}{\sqrt{n}}\right)$.

For the counterfactual estimators considered in this paper, which are non-linear functions of h, asymptotic linearity requires the estimator $\hat{h}$ to converge at a rate faster than $\sqrt[4]{n}$. Even when the dimension of X_2 is small, most popular nonparametric estimators of h would typically converge at a slower rate. For example, the Nadayara-Watson kernel regression estimator would converge at most at the $n^{\frac{1}{d_{X_2}+4}}$ unless one is willing to use higher-order kernels, which are not typically used in empirical work. Consequently, although it is possible to use nonparametric estimators of the first stage equation that would allow the estimator of the counterfactual distribution of Y to be asymptotically linear, deriving sufficient conditions for these nonparametric estimators lies beyond the scope of this paper.

D Comparison with Alternative Methods

In principle, one could estimate the triangular equation model using a control function approach, and then estimate the counterfactual distribution of Y based on these estimates. For example, Lee (2007) proposed a control function quantile regression estimator for the following triangular model:

$$Y = X\beta(\tau) + Z_1'\gamma(\tau) + U$$

$$X = \mu(\alpha) + Z'\pi(\alpha) + V$$

The identification of this model is based on different conditions than those considered in this paper. In particular, Lee (2007) assumes that $Q_{U|XZ}(\tau|x,z) = Q_{U|V}(\tau|v) \equiv \lambda_\tau(v)$, so this and the baseline model of this paper are not nested. In this model, the distribution of U is independent of both the endogenous treatment and the remaining covariates, after controlling for V. Further, he imposes additivity of a function of V in the second stage equation, and therefore the joint distribution of U and V differs from the copula given in Assumption 1.

Martinez-Sanchis et al. (2012) proposed an estimator of the unconditional distribution of Y based on Lee (2007). This estimator can consistently estimate the actual distribution of Y, and the counterfactual distribution when the distribution of Z is changed. However, it is not possible to consistently estimate the types of counterfactuals considered in this paper: by definition, U and V are heteroskedastic in the covariates, and the structural change in the determination of the treatment implies a different conditional distribution of (U,V) given Z, which is not captured by the estimated values of (U,V). On the other hand, the copula approach proposed in this paper explicitly accounts for the structural relation between the dependent variable, the treatment, and the unobservables. Therefore, it can be used to estimate the counterfactual distribution of Y when this structural dependence is changed.

Finally, one could follow Chernozhukov et al. (2013) to compute the counterfactuals based on the IVQR estimator and the counterfactual distribution of the treatment. However, this estimator is biased as long as the counterfactual treatment intensity is correlated with the unobservables. Hence, this method would be appropriate if the policy maker could assign treatment intensity at random or if treatment intensity was the same for everyone, but not if individuals have the ability to choose the treatment intensity.

E Monte Carlo

To evaluate the finite sample performance of the estimator, I carried out a simulation study. I split the sample into two groups, and the data generating process for each of them is as follows:

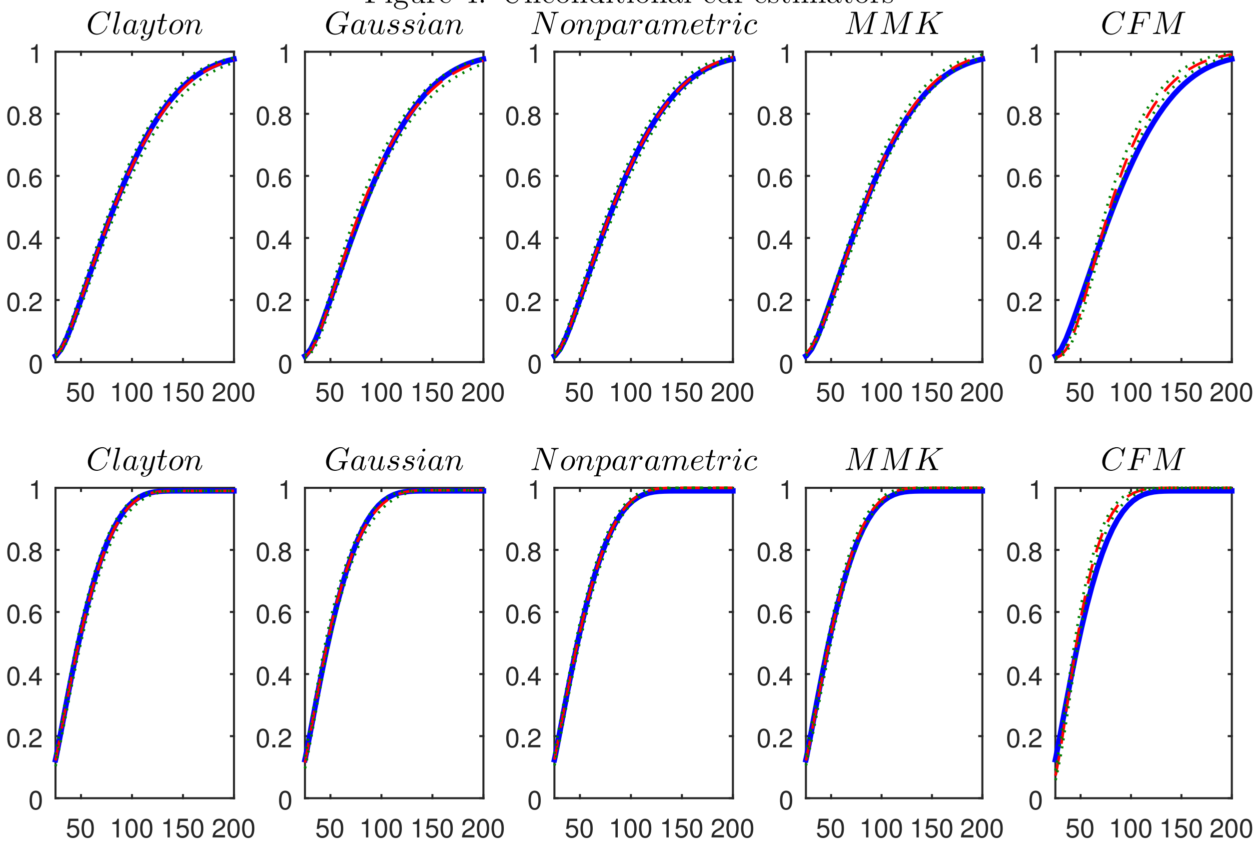
$$X_{1i} = Z_{1i}\gamma_1(V_i) + X_{2i}\gamma_2(V_i) + \gamma_3(V_i)$$
$$Y_i = X_{1i}\beta_1(U_{d,i}) + X_{2i}\beta_2(U_{d,i}) + \beta_3(U_{d,i})$$

For the first group, the parameters equal $\gamma(\nu) = \left[2+\nu, 1+1.5\log(1+\nu), F_{t_5}^{-1}(\nu)\right]'$, and $\beta(\tau) = \left[1+2\tan(\tau), 1+2(\tau-0.5)^3, \Phi^{-1}(\tau)\right]'$, the instrument and the exogenous variables are drawn from $Z_{1i} \sim U(1,10)$, $X_{2i} \sim U(10,15)$, and the copula is drawn from $(U_{1,i}, V_i) \sim Clayton(1.5)$. For the second group, all parameters are the same with two exceptions: the correlation parameter equals 3, and $\beta_1(\tau) = 0.5 + \tan(\tau)$. The sample size equals $N = 2000$, the number of repetitions is $R = 1000$, and the quantile grid for both the first and second stage equations estimation was made out of 99 evenly spaced quantiles.

Figure 4 compares the performance of the different estimators of the actual distribution: two estimators with a parametric copula, one correctly specified and one misspecified (with a Gaussian copula), the estimator with the nonparametric copula, the estimator proposed by Martinez-Sanchis et al. (2012) (*MMK*), and the estimator proposed by Chernozhukov et al. (2013) (*CFM*). All three proposed estimators, as well as the MMK estimator approximate the actual distributions reasonably well. As shown in Table 8, among these, the correctly specified estimator provides the best approximation, whereas the incorrectly specified one and the MMK perform slightly worse. On the other hand, the CFM estimator is biased in a large part of the distribution. Moreover, at certain points of the distribution it is slightly less precise than the other five estimators, as highlighted by the maximum interquantile range.

I consider two different counterfactuals. In the first one, I swap the IVQR of group 1 for those of group 2, whilst keeping the remaining parameters constant. The difference with respect to the baseline estimates would represent the coefficients component in an Oaxaca-Blinder decomposition. The second one sets a compulsory minimum treatment: $x_1 = \max\{z'\gamma(v), 40\}$. Figure 5 shows, for each of the five estimators, the difference between the counterfactual and the actual distributions, *i.e.*, the unconditional distributional effect of the

Figure 4: Unconditional cdf estimators



Notes: in each graph, the solid blue line represents the actual distribution of Y, the dashed red line represents the median (pointwise) across repetitions of the estimator, and the dotted green lines represent the 2.5 and 97.5 percentiles (pointwise) across repetitions. Clayton, Gaussian, Nonparametric, MMK and CFM denote the estimator with a Clayton copula, the estimator with a Gaussian copula, the estimator with a nonparametric copula, the estimator proposed by Martinez-Sanchis et al. (2012), and the estimator proposed by Chernozhukov et al. (2013).

Table 8: Fit of the actual distributions

		Cla	Gau	NP	MMK	CFM
Group 1	$\int_{\mathcal{Y}} \left\lvert Q_{0.5}\left(\hat{F}_Y(y)\right) - F_Y(y)\right\rvert dy$	0.002	0.006	0.003	0.006	0.026
	$\sup_y \left\lvert \hat{F}_Y(y) - F_Y(y)\right\rvert$	0.005	0.015	0.010	0.010	0.055
	$\int_{\mathcal{Y}} \nabla_{0.025}^{0.975} Q\left(\hat{F}_Y(y)\right) dy$	0.029	0.029	0.028	0.027	0.028
	$\sup_y \nabla_{0.025}^{0.975} Q\left(\hat{F}_Y(y)\right)$	0.055	0.055	0.058	0.057	0.068
Group 2	$\int_{\mathcal{Y}} \left\lvert Q_{0.5}\left(\hat{F}_Y(y)\right) - F_Y(y)\right\rvert dy$	0.002	0.004	0.006	0.008	0.021
	$\sup_y \left\lvert \hat{F}_Y(y) - F_Y(y)\right\rvert$	0.006	0.011	0.010	0.012	0.065
	$\int_{\mathcal{Y}} \nabla_{0.025}^{0.975} Q\left(\hat{F}_Y(y)\right) dy$	0.017	0.017	0.016	0.015	0.016
	$\sup_y \nabla_{0.025}^{0.975} Q\left(\hat{F}_Y(y)\right)$	0.058	0.059	0.059	0.061	0.076

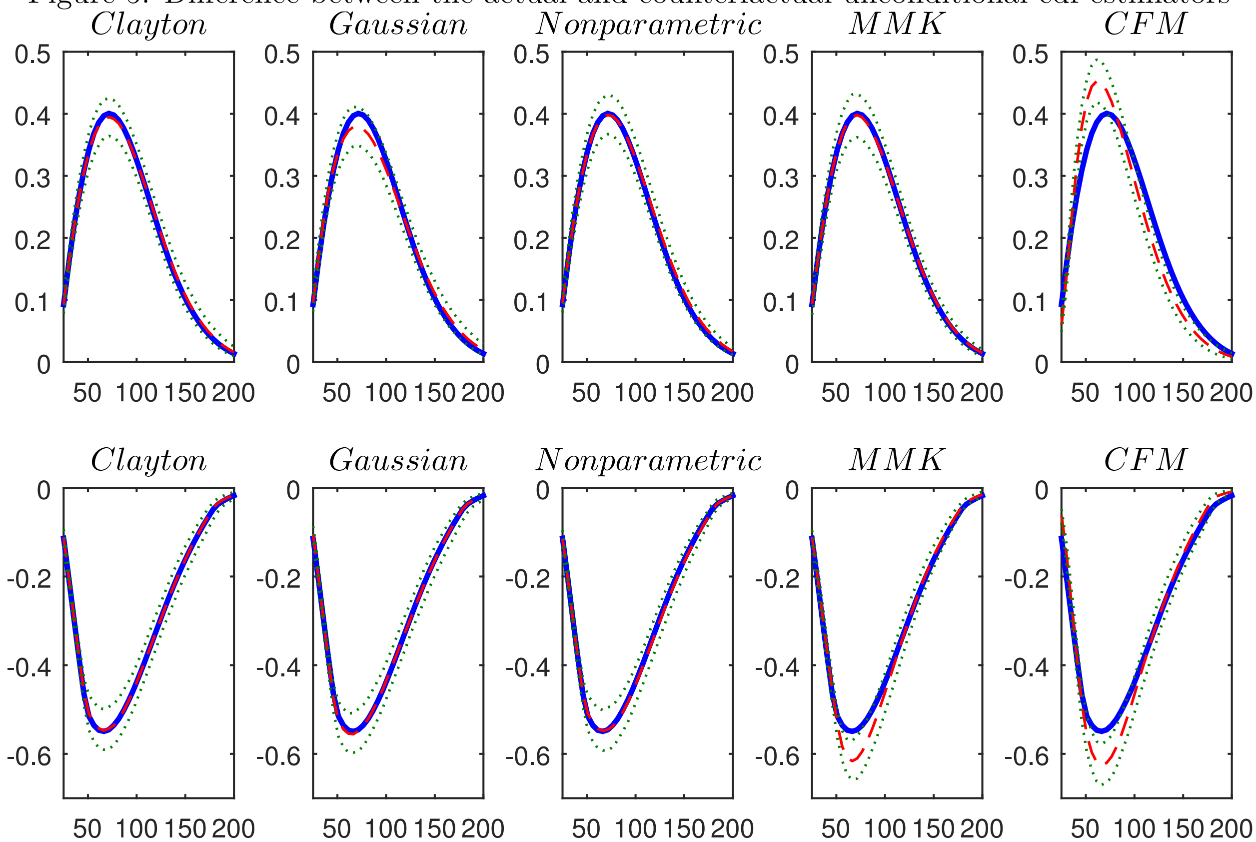
Notes: for each group, the first row represents the integral of the difference between the median across repetitions of the estimated counterfactual cdf and the true cdf; the second row represents the maximum of this difference; the third and fourth rows represent the same differences between the 97.5 and 2.5 percentiles. Cla, Gau, NP, MMK, and CFM denote the estimator with a Clayton copula, the estimator with a Gaussian copula, the estimator with a nonparametric copula, the estimator proposed by Martinez-Sanchis et al. (2012), and the estimator proposed by Chernozhukov et al. (2013).

policy. For the first counterfactual, correctly specified parametric, the nonparametric, and the MMK estimators correctly estimate the actual difference. The misspecified parametric estimator is slightly biased, although the true difference between the distributions lies inside the 95% confidence bands. On the other hand, the CFM estimator is biased at large segments of the distribution, with the sign of the bias varying at different points. Table 9 shows that the level of accuracy is similar for all estimators, although the lowest is attained by the correctly specified parametric one.

Regarding the second counterfactual, the MMK and CFM estimators are biased. This is not surprising, since in this counterfactual the structural relation between the treatment variable and the unobservables is changed. Since the counterfactual sets a minimum treatment value, it mostly affects the lower part of the distribution, which is where these two estimators display the greatest amount of bias.

Therefore, one can conclude that the CFM estimator does a good job whenever treatment is randomly or equally assigned, as well as when one fits the actual distribution. On the other hand, the MMK estimator can approximate well counterfactuals that do not break

Figure 5: Difference between the actual and counterfactual unconditional cdf estimators



Notes: in each graph, the solid blue line represents the difference between the counterfactual and the actual distribution of Y, the dashed red line represents the median (pointwise) across repetitions of the estimated difference, and the dotted green lines represent the 2.5 and 97.5 percentiles (pointwise) across repetitions. Clayton, Gaussian, Nonparametric, MMK, and CFM denote the estimator with a Clayton copula, the estimator with a Gaussian copula, the estimator with a nonparametric copula, the estimator proposed by Martinez-Sanchis et al. (2012), and the estimator proposed by Chernozhukov et al. (2013).

Table 9: Fit of the difference between the counterfactual and actual distributions

		Cla	Gau	NP	MMK	CFM
Counterfactual 1	$\int_{\mathcal{Y}} \left\lvert Q_{0.5}\left(\hat{F}_Y(y)\right) - F_Y(y)\right\rvert dy$	0.003	0.007	0.004	0.002	0.023
	$\sup_y \left\lvert \hat{F}_Y(y) - F_Y(y)\right\rvert$	0.006	0.023	0.010	0.008	0.078
	$\int_{\mathcal{Y}} \nabla_{0.025}^{0.975} Q\left(\hat{F}_Y(y)\right) dy$	0.033	0.033	0.032	0.033	0.032
	$\sup_y \nabla_{0.025}^{0.975} Q\left(\hat{F}_Y(y)\right)$	0.062	0.066	0.063	0.072	0.070
Counterfactual 2	$\int_{\mathcal{Y}} \left\lvert Q_{0.5}\left(\hat{F}_Y(y)\right) - F_Y(y)\right\rvert dy$	0.001	0.003	0.003	0.013	0.021
	$\sup_y \left\lvert \hat{F}_Y(y) - F_Y(y)\right\rvert$	0.006	0.009	0.007	0.067	0.077
	$\int_{\mathcal{Y}} \nabla_{0.025}^{0.975} Q\left(\hat{F}_Y(y)\right) dy$	0.047	0.047	0.045	0.043	0.047
	$\sup_y \nabla_{0.025}^{0.975} Q\left(\hat{F}_Y(y)\right)$	0.095	0.094	0.097	0.098	0.094

Notes: the first row represents the integral of the difference between the median across repetitions of the estimated counterfactual cdf and the true cdf; the second row represents the maximum of this difference; the third and fourth rows represent the same differences between the 97.5 and 2.5 percentiles. Cla, Gau, NP, MMK, and CFM denote the estimator with a Clayton copula, the estimator with a Gaussian copula, the estimator with a nonparametric copula, the estimator proposed by Martinez-Sanchis et al. (2012), and the estimator proposed by Chernozhukov et al. (2013).

the structural relation between the treatment and the conditional ranks, such as a change in the parameters. On the other hand, the proposed estimators perform well in all these counterfactual scenarios, even if the copula estimator is not correctly specified.