BANCA D'ITALIA
EUROSISTEMA

Temi di Discussione

(Working Papers)

Estimation of counterfactual distributions with a continuous endogenous treatment

by Santiago Pereda Fernández

February 2016

Number 1053

BANCA D'ITALIA
EUROSISTEMA

Temi di discussione

(Working papers)

Estimation of counterfactual distributions with a continuous endogenous treatment

by Santiago Pereda Fernández

Number 1053 - February 2016

The purpose of the Temi di discussione *series is to promote the circulation of working papers prepared within the Bank of Italy or presented in Bank seminars by outside economists with the aim of stimulating comments and suggestions.*

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

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

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

ESTIMATION OF COUNTERFACTUAL DISTRIBUTIONS WITH A CONTINUOUS ENDGENOUS TREATMENT

by Santiago Pereda Fernández*

Abstract

Policy makers are often interested in the distributional effects of a policy. In this paper I propose a method to estimate the actual and counterfactual distributions of an outcome variable when the treatment variable is endogenous, continuous, and its effect is heterogeneous. The model is 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. I conduct three kinds of counterfactual experiments: changing the distribution of the treatment, changing the distribution of the instrument, and changing the rule that determines the treatment, and I discuss the estimation of each of these counterfactuals. To illustrate these methods, I estimate several counterfactuals that affect the distribution of the share of food consumption.

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

Contents

* Bank of Italy, Structural Economic Analysis Directorate.

1 Introduction*

Estimation of the effect of a policy is usually one of the main objectives of experiments in economics. If the treatment effect is homogeneous, IV can be enough to identify the causal effect of the treatment, and the mean impact of the policy on the variable of interest is simply the treatment effect multiplied by the mean change in the treatment. However, these methods do not capture the distributional effects that take place if the treatment effect is heterogeneous.[1] Moreover, even if the treatment affects the outcome variable linearly, the mean effect of a policy may not be the treatment effect multiplied by the mean increase in the treatment, and the endogeneity of the treatment poses another problem that needs to be addressed, as depending on how the policy maker can enforce the treatment, the amount of treatment may differently depend on its effect.

In this paper I consider a triangular system of equations in which both the outcome and the treatment monotonically depend on a single unobservable variable each. These unobservables isolate the endogeneity of the treatment, which can be characterized by a copula. I discuss the identification and the estimation of the different components of the distribution of an outcome variable Y: its conditional distribution, which depends on both equations of the triangular system, the distribution of the instrument, and the copula. Then, I use the estimators of these functionals to estimate the actual distribution of the outcome, and the counterfactual distribution that would result from changing some of them. These estimators of the counterfactual distribution can in turn be used to estimate other functionals, such as the unconditional quantile treatment effect, or the Gini index.

I consider three types of policy counterfactuals: *(i)* a change in the distribution of the treatment and the exogenous covariates; *(ii)* a change in the distribution of the instrument

*Banca d'Italia, via Nazionale 91, 00184 Roma, Italy. I would like to thank Manuel Arellano, Stéphane Bonhomme, Domenico Depalo, Bryan Graham, Michael Jansson, James Powell, Demian Pouzo, Enrique Sentana, and seminar participants at CEMFI and University of California, Berkeley for their helpful comments and discussion. All remaining errors are my own. I can be reached via email at santiago.pereda@bancaditalia.it

[1] A policy maker interested in inequality would also like to know what the impact of the policy would be on the variance, a particular quantile of the distribution or any measure of inequality such as the Gini index.

and the exogenous covariates, for which the endogeneity needs to be taken into account in the estimation of the counterfactuals;[2] and *(iii)* a change in the structural relation between the treatment and the unobservables, which can happen when the policy maker can only partially enforce the treatment.

This paper is thus related to several others in the literature of estimation of distributional effects. Machado and Mata (2005) and Melly (2006) proposed estimators of such effects using quantile regression under exogeneity, which Chernozhukov et al. (2013) generalized by a proposing a method to estimate any functional of interest, given an initial estimator of the conditional quantile curve or the conditional distribution function. The estimator I propose extends these methods in the presence of endogeneity with an instrumental variables approach, similarly to Pereda-Fernández (2010). On the other hand, Martinez-Sanchis et al. (2012) adapted Melly (2006) to the presence of endogeneity using a control function approach. Firpo et al. (2009) proposed a different method to estimate distributional effects under exogeneity, based on the influence function rather than on quantile regression methods as in this paper. Frölich and Melly (2013) proposed a nonparametric estimator of the unconditional quantile treatment effect for the subpopulation of compliers when the treatment is an endogenous binary variable. However, in this paper I assume both the outcome and the treatment to be continuous, which allows me to nonparametrically identify the copula that captures the endogeneity of the treatment.

The estimation of the unconditional distribution of Y that I propose is valid for any consistent estimator of the two functions that conform the triangular model, and the copula. For the first two, I use existing quantile methods.[3] For the copula, I propose two estimators: one nonparametric, that requires the copula to be invariant to the covariates, and one parametric. Both estimators require the inversion of the estimated quantile processes that conform the triangular model, and they achieve $\sqrt{n}$ convergence rate, which is also achieved by the estimators of the unconditional distribution based on them.

[2]Alternatively, one could model the relation between the instrument and the outcome variable,*i.e.* a reduced form equation. This possibility is discussed in section 3.4.

[3]In particular I use instrumental variables quantile regression (Chernozhukov and Hansen, 2006) for the estimation of the second stage equation, and quantile regression (Koenker, 2005) for the first stage equation.

The identification of these type of triangular models has received a lot of attention in the literature, using either instrumental variable or control function approaches. Early works are Chesher (2003) or Imbens and Newey (2009), who study the nonparametric identification of nonseparable models using a control function approach. Other papers propose methods that could be referred to as semiparametric, 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) propose a parametric model of Chesher (2003). D'Haultfœuille and Février (2015) and Torgovitsky (2015) are recent papers that establish the identification of nonseparable triangular models when the support of the instrument is small, for which they use the monotonicity of the unobservables. On the other hand, Hoderlein and Mammen (2007) discusses the identification of such models without monotonicity.

Examples of empirical works that fit into framework presented in this paper include 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), or the impact of class size on scholastic achievement (Angrist and Lavy, 1999) or on long-term outcomes (Fredriksson et al., 2013). All these studies could benefit by studying the effects on inequality of an intervention that results in a different assignment of the amount of treatment for the whole population. In this paper I consider the estimation of Engel curves using data on Italian households. I estimate the distribution of the proportion of food consumption as a function of total expenditure, and four counterfactual distributions that involve either a redistribution of total income from high to low income households, or an equal number of children by household. The results show that the redistribution of income, despite substantially reducing income inequality, does not translate into a reduction of consumption inequality. On the other hand, if all households had the same number of children, inequality would be reduced, and the more children, the largest the reduction.

The rest of the paper is organized as follows. In section 2 I 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 carry out a Monte Carlo experiment, and in section 5 I apply the methodology presented in this paper to the estimation of Engel curves. Finally, section 6 concludes.

2 Identification

Let Y be the outcome variable of interest, $X \equiv (X_1\, X_2')'$ be the vector composed by the treatment, X_1, and a set of exogenous covariates, X_2, $Z \equiv (Z_1\, X_2')'$ be the vector composed by the instrumental variable, Z_1 and the exogenous covariates, and U and V be uniformly distributed random variables that are not observed by the econometrician. These variables conform the following triangular model:

$$Y = g\left(X_1, X_2, U\right) \tag{1}$$

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

$$U, V|Z \sim C_{UV|X_2} \tag{3}$$

where $g\left(\cdot,\cdot,\cdot\right)$ and $h\left(\cdot,\cdot,\cdot\right)$ are nonseparable and strictly increasing in their last argument, and $C_{UV|X_2}$ is the copula of (U, V), conditional on the vector of exogenous covariates.[4] The Skorohod representation allows us to isolate the endogeneity of the treatment, captured by the copula, from the structural equations of the outcome and the treatment.[5] In this setup, $h\left(\cdot,\cdot,\cdot\right)$ represents the conditional quantile function (CQF) of X_1, which satisfies $\mathbb{P}\left(X_1 \leq h\left(Z_1, X_2, \tau\right)|Z_1, X_2\right) = \tau$, but because of the endogeneity of the treatment, $g\left(\cdot,\cdot,\cdot\right)$ represents instead the structural quantile function (SQF) of Y, which is different from its CQF, and for which $\mathbb{P}\left(Y \leq g\left(X_1, X_2, \tau\right)|Z_1, X_2\right) = \tau$.[6]

[4] By definition, a copula is the multivariate distribution of $(U_1, ..., U_d)$ such that their marginal distributions are uniformly distributed on the unit interval. Sklar (1959) showed that any multivariate distribution of the continuously distributed variables $X_1, ..., X_d$ there exists a unique cdf C, such that $\mathbb{P}\left(X_1 \leq x_1, ..., X_d \leq x_d\right) = C\left(F_1\left(x_1\right), ..., F_d\left(x_d\right)\right)$.

[5] The Skorohod representation states that a random variable φ_i can be written in terms of its quantile function: $\varphi_i = q\left(U_i\right)$, where $U_i \sim U\left(0,1\right)$.

[6] See Chernozhukov and Hansen (2013) for a more detailed discussion on the difference between the SQF and the CQF of Y.

2.1 Identification of the Actual Distribution of Y

Before focusing on the counterfactual distribution of Y, I consider its actual distribution and discuss the identification of the different components upon which it depends. Denote the conditional distribution of Y by $F_{Y|Z}$, and the conditional copula by $C_{U|VX_2}$. Then,

$$\begin{aligned} F_{Y|Z}(y|z) &= \int_{[0,1]^2} \mathbf{1}\left(g\left(h\left(z_1, x_2, v\right), x_2, u\right) \leq y\right) dC_{UV|X_2}\left(u, v|x_2\right) \\ &= \int_{[0,1]^2} \mathbf{1}\left(u \leq g^{-1}\left(h\left(z_1, x_2, v\right), x_2, y\right)\right) dC_{U|VX_2}\left(u|v, x_2\right) dv \\ &= \int_0^1 C_{U|VX_2}\left(g^{-1}\left(h\left(z_1, x_2, v\right), x_2, y\right)|v, x_2\right) dv \end{aligned} \tag{4}$$

where $\mathbf{1}(\cdot)$ is the indicator function. This implies that by conditioning on (V, Z), there is a bijection between Y and U, given by $F_{Y|ZV}(y|z, v) = C_{U|VX}\left(g^{-1}\left(h\left(z_1, x_2, v\right), x_2, y\right)|v, x_2\right)$. To better understand this relation, consider the exogenous case, *i.e.* U and V are independent of each other, conditional on Z. The conditional copula would simplify to $C_{U|VX_2}(u|v, x_2) = u$. Moreover, by equation 2, when Z and V are known, so is X_1. Thus, it follows

$$F_{Y|ZV}(y|z, v) = g^{-1}\left(h\left(z_1, x_2, v\right), x_2, y\right) = g^{-1}\left(x_1, x_2, y\right) \equiv u \tag{5}$$

whereas under endogeneity

$$F_{Y|ZV}(y|z, v) = C_{U|VX_2}(u|v, x_2) \neq u \tag{6}$$

Consequently, the identification of $F_{Y|Z}$ requires the identification of three components: the SQF of Y, the CQF of X_1 and the copula of (U, V) conditional on X_2. Identification of $g(\cdot, \cdot, \cdot)$ has been an active area of research, with recent works by D'Haultfœuille and Février (2015) and Torgovitsky (2015) establishing its identification with a continuous treatment even when the support of the instrument Z_1 are two points. Although the assumptions required for identification are different, both of them require the instrument to be strongly exogenous, strict monotonicity and continuity of both $g(\cdot, \cdot, \cdot)$ and $h(\cdot, \cdot, \cdot)$, and a normalization of U. The identification of $h(\cdot,\cdot,\cdot)$, was established by Matzkin (2003), and it follows by the normalization that V is uniformly distributed, the strict monotonicity, and the continuity of $h(\cdot,\cdot,\cdot)$. As for the identification of the copula, it is obtained by inverting the SQF and the CQF, which is possible by the continuity and the monotonicity of both functions in their last argument:

$$U = g^{-1}(X_1, X_2, Y)$$
$$V = h^{-1}(Z_1, X_2, X_1)$$

Hence, it follows that $C_{UV|X_2}(u, v|x_2) = \mathbb{P}(U \leq u, V \leq v|z)$. Finally, to obtain the unconditional distribution of Y, integrate $F_{Y|Z}$ over the actual distribution of Z:

$$F_Y(y) = \int_0^1 F_{Y|Z}(y|z)\, dF_Z(z) \tag{7}$$

2.2 Identification of the Counterfactual Distributions of Y

Deriving the unconditional counterfactual distribution of Y under exogeneity (Chernozhukov et al., 2013) is relatively straightforward, as there is no dependence between the regressors and the unobservable. With an endogenous treatment there exist conceptually different types of counterfactuals, and in this paper I differentiate between three of them: those that change the distribution of X, which is randomly assigned to the population, those that change the distribution of Z, and those that change the determination of the treatment, *i.e.* $h(\cdot,\cdot,\cdot)$.[7] If the policy maker were able to enforce the values of X for each individual (without knowing their particular values of (u, v)), then the endogeneity would no longer be relevant for the derivation of the unconditional distribution of Y, which would be given by

$$F_Y^{cf}(y) = \int_{\mathcal{X}} \int_0^1 \mathbf{1}\left(g(x_1, x_2, u) \leq y\right) du dF_X^{cf}(x)$$

[7] Throughout this paper I denote the counterfactual distributions by adding the superscript cf.

On the other hand, if the policy maker cannot directly enforce X, but rather Z, then equations 4 and 7 suggest the appropriate way to obtain the counterfactual distribution:

$$F_Y^{cf}(y) = \int_{\mathcal{Z}} \int_{[0,1]^2} \mathbf{1}\left(g\left(h\left(z_1, x_2, v\right), x_2, u\right) \leq y\right) dC_{UV|X_2}\left(u, v|x_2\right) dF_Z^{cf}(z)$$

Finally, there is the possibility that the policy maker can only partially enforce the distribution of X by changing the determination of X_1. For instance, in a Mincer equation, if a policy maker decided to set a minimum level of compulsory schooling, it would not affect those students who would have attained an education level above this cap, but it would increase the education level of those below it, whose values of their unobservables are different. In such cases, the copula distribution is necessary to derive the counterfactuals. Denote by $h^{cf}(z_1, x_2, v)$ the counterfactual function that determines X_1.[8] Then,

$$F_Y^{cf}(y) = \int_{\mathcal{Z}} \int_{[0,1]^2} \mathbf{1}\left(g\left(h^{cf}\left(z_1, x_2, v\right), x_2, u\right) \leq y\right) dC_{UV|X_2}\left(u, v|x_2\right) dF_Z(z)$$

3 Estimation

I use the sample analogue to estimate the actual and counterfactual distributions of Y. First I present the estimators of these distributions based on estimators of the SQF of Y and the conditional copula distribution. Then, I propose two methods to estimate the latter. For notational simplicity, define $S_Y(u|z,v) \equiv g\left(h\left(z_1, x_2, v\right), x_2, u\right)$, and $u(y,z,v) \equiv C_{U|VX_2}^{-1}\left(F_{Y|ZV}(y|z,v)|v, x_2\right)$.

3.1 Estimation of the Actual Distribution of Y

As shown in equations 4 and 7, the distribution of Y depends on three functionals: $S_Y(u|z,v)$, $C_{U|VX_2}(u|v,x_2)$ and $F_Z(z)$. In this paper I work with the following assumptions:

Assumption 1. $(y_i, x_{1i}, x_{2i}', z_{1i})'$ *are iid for* $i = 1, ..., n$*, defined on the probability space*

[8] In the previous example, $h^{cf} = \max\left(h\left(z_1, x_2, v\right), x^{\min}\right)$.

$(\Omega, \mathcal{F}, \mathbb{P})$ *and take values in a compact set.*

Assumption 2. *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 $C_{UV|X_2}(u, v|x_2)$ is bounded above and away from zero on $[0,1]^2$, and it is uniformly continuous and differentiable with respect to its arguments a.e. Moreover, the marginals are uniformly distributed on the unit interval and therefore $c_{UV|X_2}(u, v|x_2) = c_{U|VX_2}(u|vx_2) = c_{V|UX_2}(v|ux_2)$.*

Assumption 3. *Let $\hat{S}_Y(u|z, v)$ and $\hat{C}_{U|VX_2}(u|v, x_2)$ respectively denote an estimator of the structural function of Y given Z and V, and of the conditional copula. These estimators are strictly monotone in their first argument and jointly asymptotically Gaussian, i.e.*

$$\sqrt{n}\begin{pmatrix} \hat{S}_Y(u|z, v) - S_Y(u|z, v) \\ \hat{C}_{U|VX_2}(u|v, x_2) - C_{U|VX_2}(u|v, x_2) \end{pmatrix} \Rightarrow \mathbb{G}_{\mathcal{N}}(u, v, z)$$

where $\mathbb{G}_{\mathcal{N}}(u, v, z)$ is a Gaussian process with zero mean and covariance $\Sigma_{\mathcal{N}}(u, v, z, \tilde{u}, \tilde{v}, \tilde{z})$.

Let the estimator of $F_Y(y)$ be given by[9]

$$\begin{aligned} \hat{F}_Y(y) &= \frac{1}{n}\sum_{i=1}^{n}\int_{[0,1]^2} \mathbf{1}\left(\hat{S}_Y(u|z_i, v) \le y\right) d\hat{C}_{UV|X_2}(u, v|x_{2i}) \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \hat{C}_{U|VX_2}\left(\hat{S}_Y^{-1}(y|z_i, v)\,|v, x_{2i}\right) dv \end{aligned} \tag{8}$$

The following theorem characterizes its asymptotic distribution:

Theorem 1. *Let assumptions 1 to 3 hold. The asymptotic distribution of $\hat{F}_Y$ is given by*

$$\sqrt{n}\left(\hat{F}_Y(y) - F_Y(y)\right) \Rightarrow \mathbb{G}_{\mathcal{O}}(y)$$

where $\mathbb{G}_{\mathcal{O}}(y) \equiv \int_{\mathcal{Z}}\int_0^1 O(y, v, z)\,\mathbb{G}_{\mathcal{N}}(u(y, z, v), v, z)\, dv dF_Z(z)$ is a Gaussian process, and

[9] With this estimator, it is straightforward to estimate the unconditional quantile function of Y, or any other function that depends on $F_Y(y)$ by plugging in this estimator, as in Chernozhukov et al. (2013).

$O(y,v,z) = \begin{bmatrix} -f_{Y|ZV}(y|z,v) & 1 \end{bmatrix}$. $\mathbb{G}_{\mathcal{O}}(y)$ *has zero mean and covariance function given by:*

$$\Sigma_{\mathcal{O}}(y,\tilde{y}) = \int_{\mathcal{Z}^2}\int_{[0,1]^2} O(y,z,v)\,\Sigma_{\mathcal{N}}(u(y,z,v),v,z,u(\tilde{y},\tilde{z},\tilde{v}),\tilde{v},\tilde{z})\,O(\tilde{y},\tilde{z},\tilde{v})'\,dvd\tilde{v}dF_Z(z)\,dF_Z(\tilde{z})$$

3.2 Estimation of the Counterfactual Distributions of Y

The estimator of the first type of counterfactual, changing the distribution of X exogenously, fits the framework of Chernozhukov et al. (2013), so the estimator is

$$\hat{F}_Y^{cf}(y) = \frac{1}{N}\sum_{i=1}^{N}\int_0^1 \mathbf{1}\left(\hat{S}_Y\left(u|x_i^{cf}\right) \leq y\right) du \tag{9}$$

If the policy maker cannot directly fix the distribution of X, but that of Z, then the estimator actually requires using the copula:

$$\hat{F}_Y^{cf}(y) = \frac{1}{N}\sum_{i=1}^{N}\int_{[0,1]^2} \mathbf{1}\left(\hat{S}_Y\left(u|z_i^{cf},v\right) \leq y\right) d\hat{C}_{UV|X_2}\left(u,v|x_{2i}^{cf}\right) \tag{10}$$

Regarding the counterfactual in which the policy maker changes the way the treatment is determined, the estimator is given by

$$\hat{F}_Y^{cf}(y) = \frac{1}{N}\sum_{i=1}^{N}\int_{[0,1]^2} \mathbf{1}\left(\hat{g}\left(\hat{h}^{cf}(z_{1i},x_{2i},v),x_{2i},u\right) \leq y\right) d\hat{C}_{UV|X_2}(u,v|x_{2i}) \tag{11}$$

Notice that these estimators can be used as an argument to estimate other functionals of interest, such as the unconditional quantile treatment effect, or the Gini index. See Chernozhukov et al. (2013) for a detailed discussion on those estimators and their asymptotic properties.

3.3 Estimation of the SQF and the Copula

The estimators of the distribution of Y presented in this paper depend on the estimators of the SQF of Y and the Copula (U,V), which need to be estimated in a first step. I propose to estimate the former using the instrumental variables quantile regression estimator (Chernozhukov and Hansen, 2006), and the quantile regression estimator (Koenker and Bassett, 1978).[10] To derive their asymptotic properties, I impose the following regularity conditions:

Assumption 4.

$$g(u, x_1, x_2) = x'\beta(u)$$
$$h(v, z_1, x_2) = z'\gamma(v)$$

where $\beta(u)$ and $\gamma(v)$ are uniformly continuous and g and h are strictly increasing in their first argument.

Assumption 5. *For all (τ, θ), $\left(\beta(\tau)', \gamma(\theta)'\right)' \in int\mathcal{B} \times \mathcal{G}$, where $\mathcal{B} \times \mathcal{G}$ is compact and convex.*

Assumption 6.

$$\Pi(\beta, \iota, \gamma, \tau, \theta) \equiv \mathbb{E}\begin{bmatrix} \left(\tau - \mathbf{1}\left(Y < X'\beta + \Phi(\tau)'\iota\right)\right)\Psi(\tau) \\ \left(\theta - \mathbf{1}\left(X_1 < Z'\gamma\right)\right)Z \end{bmatrix}$$
$$\Pi(\beta, \gamma, \tau, \theta) \equiv \mathbb{E}\begin{bmatrix} \left(\tau - \mathbf{1}\left(Y < X'\beta\right)\right)\Psi(\tau) \\ \left(\theta - \mathbf{1}\left(X_1 < Z'\gamma\right)\right)Z \end{bmatrix}$$

where $\Psi(\tau) \equiv \left[\Phi(\tau)', X_2'\right]'$, $\Phi(\tau)$ is a vector a transformation of instruments, Jacobian matrices $\frac{\partial}{\partial(\beta', \gamma')}\Pi(\beta, \gamma, \tau, \theta)$ and $\frac{\partial}{\partial\left(\beta_2', \iota', \gamma'\right)}\Pi(\beta, \iota, \gamma, \tau, \theta)$ are continuous and have full rank, uniformly over $\mathcal{B} \times \mathcal{I} \times \mathcal{G} \times \mathcal{T} \times \mathcal{C}$ and the image of $\mathcal{B} \times \mathcal{G}$ under the mapping $(\beta, \gamma) \mapsto \Pi(\beta, \gamma, \tau, \theta)$ is simply-connected.[11]

[10] Chernozhukov and Hansen (2005) does no longer constitute the state of the art in the identification of a triangular model. Torgovitsky (2015) and D'Haultfœuille and Février (2015) are the two most recent contributions to this literature and, as far as I know, no estimator of the structural quantile function is based on the identification results of these two papers. Proposing such an estimator that is also easily implementable in current applied research is beyond the scope of this paper.

[11] Notice that I changed Chernozhukov and Hansen (2006) notation and ι denotes the parameter γ in their paper.

Assumption 7. *$wp \to 1$, the function $\hat{\Phi}(\tau, z) \in \mathcal{F}$ and $\hat{\Phi}(\tau, z) \xrightarrow{p} \Phi(\tau, z)$ uniformly in (τ, z) over compact sets, where $\Phi(\tau, z) \in \mathcal{F}$; the functions $f(\tau, z) \in \mathcal{F}$ are uniformly smooth functions in z with the uniform smoothness order $\omega > \dim\left((x_1, z')'\right)/2$, and moreover $\|f(\tau', z) - f(\tau, z)\| < C|\tau - \tau'|^a$, $C > 0, a > 0$, for all (z, τ, τ').*

Let $\hat{\beta}(\cdot)$ and $\hat{\gamma}(\cdot)$ be the IVQR and QR estimators of the parameters of the processes defined in assumption 4.[12] Then, define the following estimator of the SQF:

$$\hat{S}_Y(u|z, v) \equiv \hat{x}(v)' \hat{\beta}(u) = \begin{pmatrix} z'\hat{\gamma}(v) & x_2 \end{pmatrix} \hat{\beta}(u) \tag{12}$$

This estimator has the following limiting distribution:

Proposition 1. *Let $\hat{S}_Y(u|z, v) \equiv \hat{x}(v)' \hat{\beta}(u) = \begin{pmatrix} z'\hat{\gamma}(v) & x_2 \end{pmatrix} \hat{\beta}(u)$. Under assumptions 1, 2, and 4 to 7, the asymptotic distribution of $\hat{S}_Y(u|z, v)$ is given by:*

$$\sqrt{n}\left(\hat{S}_Y(u|z, v) - S_Y(u|z, v)\right) \Rightarrow \mathbb{G}_{\mathcal{K}}(u, v, z)$$

where $\mathbb{G}_{\mathcal{K}}(u, v, z) \equiv K(u, v, z)\, \mathbb{G}_{\mathcal{J}}(u, v)$ is a Gaussian Process, $\mathbb{G}_{\mathcal{J}}(u, v)$ is the joint asymptotic distribution of the IVQR and QR estimators, a Gaussian process with zero mean and covariance function $\Sigma_{\mathcal{J}}(u, v, \tilde{u}, \tilde{v})$, and $K(u, v, z) \equiv \begin{bmatrix} x(v)' & \beta_1(u)\, z' \end{bmatrix}$. $\mathbb{G}_{\mathcal{K}}(u, v, z)$ has zero mean and covariance function given by:

$$\Sigma_{\mathcal{K}}(u, v, z, \tilde{u}, \tilde{v}, \tilde{z}) \equiv K(u, v, z)\, \Sigma_{\mathcal{J}}(u, v, \tilde{u}, \tilde{v})\, K(\tilde{u}, \tilde{v}, \tilde{z})'$$

Regarding the copula, I propose two alternative estimators: one parametric and another nonparametric. Both estimators require the inversion of the quantile processes to obtain the

[12]For simplicity, in this paper I assume constant weights in the estimation. Generalizing the estimation to have non-constant weights is straightforward.

fitted values of (u_i, v_i) for each individual:

$$\hat{v}_i = \int_0^1 \mathbf{1}\left(z_i'\hat{\gamma}(v) \leq x_{1i}\right) dv$$
$$\hat{u}_i = \int_0^1 \mathbf{1}\left(x_i'\hat{\beta}(u) \leq y_i\right) du$$

For each of the estimators I work with either one of the following two assumptions:

Assumption 8. *The copula $C_{UV|X_2}(U, V|x_2; \xi)$ is known up to the vector of parameters $\xi \in int(\mathcal{R})$, where $\mathcal{R}$ is bounded and of finite dimension. Moreover, its pdf, denoted by $c(u, v|x_2; \xi)$, is three times continuously differentiable with respect to its arguments on $[0, 1]^2$.*[13]

Assumption 9. *The copula $C_{UV}(u, v)$ is independent of X_2.*

If the copula satisfies assumption 8, estimate the parameters upon which it depends by Quasi Maximum Likelihood, yielding the vector $\hat{\xi}$:

$$\hat{\xi} = \arg\max_{\xi} \frac{1}{n}\Sigma_{j=1}^{n} \log\left(c\left(u_j, v_j|x_2; \xi\right)\right) + \frac{1}{n}\Sigma_{j=1}^{n} \log\left(\frac{c\left(\hat{u}_j, \hat{v}_j|x_2; \xi\right)}{c\left(u_j, v_j|x_2; \xi\right)}\right) \tag{13}$$

The first term in equation 13 is the log likelihood function. However, because the actual values of the copula are not observed, the function that is maximized differs from the actual log likelihood function by the second term. Finally, the estimator of the copula is given by

$$\hat{C}_{UV|X_2}(u, v|x_2) \equiv C_{UV|X_2}\left(u, v|x_2; \hat{\xi}\right) \tag{14}$$

Alternatively, if assumption 9 holds, the nonparametric estimator of the copula is the sample analog based on the fitted values of (u_i, v_i):

$$\check{C}_{UV}(u, v) \equiv \frac{1}{n}\Sigma_{i=1}^{n}\mathbf{1}\left(\hat{u}_i \leq u\right)\mathbf{1}\left(\hat{v}_i \leq v\right) \tag{15}$$

[13] Notice that this rules out the cases of perfect correlation, since in those, either $\mathbb{P}(U = u|V = v) = \mathbf{1}(u = v)$ or $\mathbb{P}(U = u|V = v) = \mathbf{1}(u = 1 - v)$, implying that the joint pdf takes a value of zero in a large subspace of $[0, 1]^2$.

Both estimators of the copula suffer from the unobservability of u_i and v_i, which requires the inversion of the estimated quantile processes. If the copula is sufficiently smooth, both estimators are consistent, although only the parametric estimator can be shown to be asymptotically Gaussian. On the other hand, the nonparametric estimator of the copula does not impose the parametric distribution, but it requires the copula to be invariant with respect to the covariates, which may constitute a strong assumption for some applications. The asymptotic distribution of the parametric estimator of the copula, together with the structural quantile function is given by:

Proposition 2. *Under assumptions 1, 2, and 4 to 8, the joint asymptotic distribution of* $\hat{S}_Y(u|z,v)$ *and* $\hat{C}_{U|VX_2}(u|v,x_2)$ *is given by*

$$\sqrt{n}\begin{pmatrix}\hat{S}_Y(u|z,v)-S_Y(u|z,v)\\ \hat{C}_{U|VX_2}(u|v,x_2)-C_{U|VX_2}(u|v,x_2)\end{pmatrix}\Rightarrow\mathbb{G}_{\mathcal{N}}(u,v,z)$$

where $\mathbb{G}_{\mathcal{N}}(u,v,z)\equiv N(u,v,z)\,\mathbb{G}_{\mathcal{M}}(u,v)$ *is a Gaussian process,* $\mathbb{G}_{\mathcal{M}}(u,v)$ *is the joint asymptotic distribution of the IVQR and QR estimators, and the estimator of the copula parameters, a Gaussian process with zero mean and covariance matrix* $\Sigma_{\mathcal{M}}(u,v,\tilde{u},\tilde{v})$*, and* $N(u,v,z)\equiv\begin{bmatrix}x(v)' & \beta_1(u)z' & 0\\ 0 & 0 & \frac{\partial}{\partial\xi}C_{U|VX_2}(u|v,x_2;\xi)\end{bmatrix}$. *The process* $\mathbb{G}_{\mathcal{N}}(u,v,z)$ *has zero mean and covariance function given by:*

$$\Sigma_{\mathcal{N}}(u,v,z,\tilde{u},\tilde{v},\tilde{z})\equiv N(u,v,z)'\,\Sigma_{\mathcal{M}}(u,v,\tilde{u},\tilde{v})\,N(\tilde{u},\tilde{v},\tilde{z})'$$

Let the estimator of the distribution of Y based on the nonparametric estimator of the copula be given by

$$\check{F}_{Y|Z}(y|z)\equiv\int_0^1\mathbf{1}\left(\hat{x}(v)'\hat{\beta}(u)\leq y\right)d\check{C}_{UV}(u,v)=\frac{1}{n}\Sigma_{j=1}^{n}\mathbf{1}\left(\hat{x}(\hat{v}_j)'\hat{\beta}(\hat{u}_j)\leq y\right)\tag{16}$$

The uniform convergence of $\check{C}_{UV}(u,v)$ can be shown to be at a rate $\sqrt{n}$, which in turn allows to show that $\check{F}_{Y|Z}(y|z)$ is indeed uniformly consistent at that rate. However, it is not possible to obtain asymptotic Gaussianity by the usual arguments: the nonlinearity of the indicator function prevents us from using the extended continuous mapping theorem, which is required because (u_i, v_i) are estimated. This issue could be overcome by using a smooth function that converges uniformly to the indicator function, but even in this scenario it would not be possible to establish the asymptotic normality based on theorem 1, as the estimator of the *conditional* copula converges at rate slower than $\sqrt{n}$.[14] Nevertheless, $\check{F}_{Y|Z}(y|z)$ is a uniformly consistent estimator, as shown by the following proposition:

Proposition 3. *Let lemma 3 and assumptions 1, 2, 4 to 7, and 9 hold. Then,*

$$\sup_{y,z} \sqrt{n} \left| \check{F}_{Y|Z}(y|z) - F_{Y|Z}(y|z) \right| = O_p(1)$$

Estimation of the asymptotic variance of the different estimators is feasible, but computationally cumbersome: many of these variances need to be computed for a large number of values, making it particularly impractical. For example, $\Sigma_{\mathcal{O}}(y, \tilde{y})$ would need to be computed for every possible combination of (y_i, y_j), for $i, j = 1, ..., n$, and $\Sigma_{\mathcal{N}}(u, v, z, \tilde{u}, \tilde{v}, \tilde{z})$ would require the computation of an even larger number of combinations of the arguments upon which it depends. Nevertheless, the estimators of the variance can be found in appendix C.

3.4 Discussion of Alternative Methods and their Validity

The method presented in this paper is not the only one to estimate the counterfactual distribution when the treatment is endogenous. When the counterfactual involves changing the distribution of Z, then it is possible to directly estimate the CQF of Y given Z, *i.e.* a reduced form regression, and then apply Chernozhukov et al. (2013) to obtain the counterfactual unconditional distribution of Y. However, this approach does not provide an estimator of the distribution of the treatment in the counterfactual, which may be of interest for the policy maker if, for example, the treatment is costly to implement. Moreover, this strategy is not feasible to implement for the third type of counterfactuals, when the

[14] See appendix B.8.

distribution of the treatment is only partially affected. When this type of counterfactuals are more relevant, such as an increase in compulsory education, this method is not an alternative to the one proposed in this paper.

Another possibility is to 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) proposes 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, he assumes that $Q_{U|XZ}(\tau|x,z) = Q_{U|V}(\tau|v) \equiv \lambda_\tau(v)$, so this model and the one I use in this paper are not nested, and the joint distribution of U and V, as defined in this model is not the copula given in equation 3. Martinez-Sanchis et al. (2012) propose an estimator of the unconditional distribution of Y based on Lee (2007) estimator.[15] This estimator can consistently estimate the actual distribution of Y, and the counterfactual distribution when the distribution of Z is changed, but it fails to do so when the distribution of the treatment is partially affected. The reason is that, by definition, U and V are heteroskedastic in the covariates, and changing the way the determination of the treatment implies a different conditional distribution of (U, V) given Z, which is not captured by the fitted values of (U, V). On the other hand, the copula is invariant to such counterfactuals, and it is thus appropriate to estimate the distribution of Y.

[15]Note that Martinez-Sanchis et al. (2012) do not show the asymptotic distribution of their estimator, thus not providing a way to carry out uniform inference.

4 Monte Carlo

To evaluate the finite sample performance of the estimator, I carried out a simulation study with the following data generating process:

$$X_{1i} = Z_{1i}\gamma_1(v_i) + X_{2i}\gamma_2(v_i) + \gamma_3(v_i)$$
$$Y_i = X_{1i}\beta_1(u_i) + X_{2i}\beta_2(u_i) + \beta_3(u_i)$$

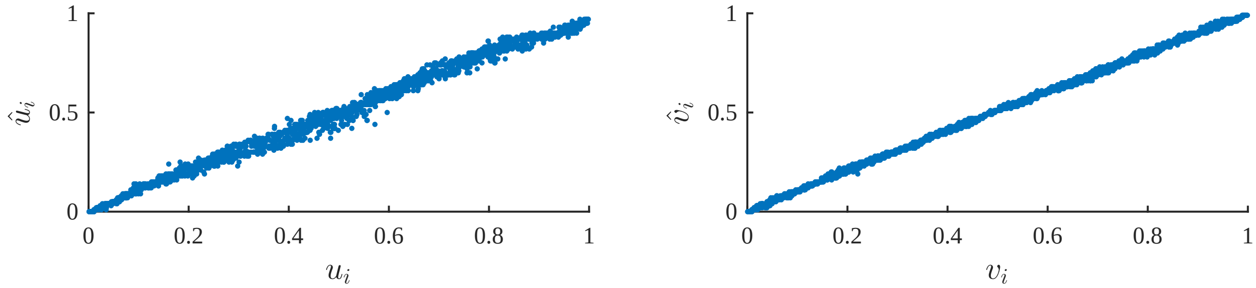
where the parameters are given by $\gamma(\theta) = \left[4 + 2\tan(\theta), 2 + 3(\theta - 0.5)^3, 4F_{t_5}^{-1}(\theta)\right]'$, and $\beta(\tau) = \left[1 + 4\log(1+\tau), 3 + 4e^{\tau}(1+\tau)^{-1}, 5\Phi^{-1}(\tau)\right]'$, the instrument and the exogenous variables are drawn from $Z_{1i} \sim U(1,3)$, $X_{2i} \sim U(10,15)$, and the copula is drawn from $(u_i, v_i) \sim Clayton(2)$. The sample size equals $N = 2000$, the number of repetitions is $M = 200$ and the quantile grid for both the first and second stage equations estimation was made out of $H = K = 99$ evenly spaced quantiles.

Figure 1 shows the scatter plot of the fitted values of (u_i, v_i) and their actual values, since they are required to estimate the copula. Both estimates are reasonably close to the 45 degree line, and therefore to their true values, but the estimates of v_i are more accurate than those of u_i. Figure 2 compares the performance of the different estimators of the actual distribution: the estimator with the parametric copula, the estimator with the nonparametric copula, the estimator proposed by Martinez-Sanchis et al. (2012), and the estimator proposed by Chernozhukov et al. (2013) using the reduced form regression of Y on Z. All four estimators approximate the true distribution reasonably well, as shown in table 1, particularly around the center of the distribution. Regarding their precision, the parametric estimator performs slightly worse than the other three. Increasing the number of quantiles used in the estimation to approximate the integrals results in a better approximation at the tails.[16]

Now consider a counterfactual in which the policy maker sets a compulsory minimum treatment, *i.e.* $x_1 = \max\{z'\gamma(v), 100\}$. Figure 3 compares the performance of the two estimators proposed in this paper and Martinez-Sanchis et al. (2012) estimator in the

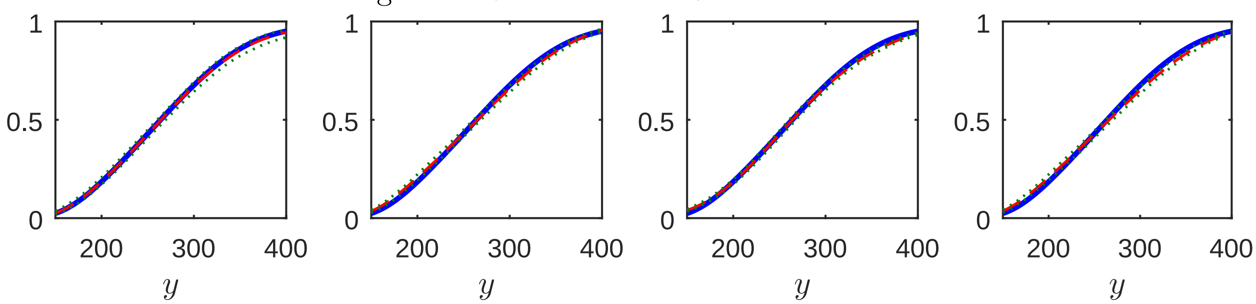
[16] Results available upon request.

Figure 1: Fitted Values of the Copula Realizations



Each graph is the scatterplot of the true values of conditional quantiles and their fitted values.

Figure 2: Unconditional CDF Estimators



In each of the four graphs, the solid line represents the actual distribution of Y, the dashed line represents the median (pointwise) across repetitions of the estimator, and the dotted line represent the 2.5 and 97.5 percentiles (pointwise) across repetitions. The first estimator uses the parametric estimator of the copula, the second one uses the nonparametric estimator of the copula, the third one is the estimator proposed by Martinez-Sanchis et al. (2012), and the fourth one is the estimator proposed by Chernozhukov et al. (2013).

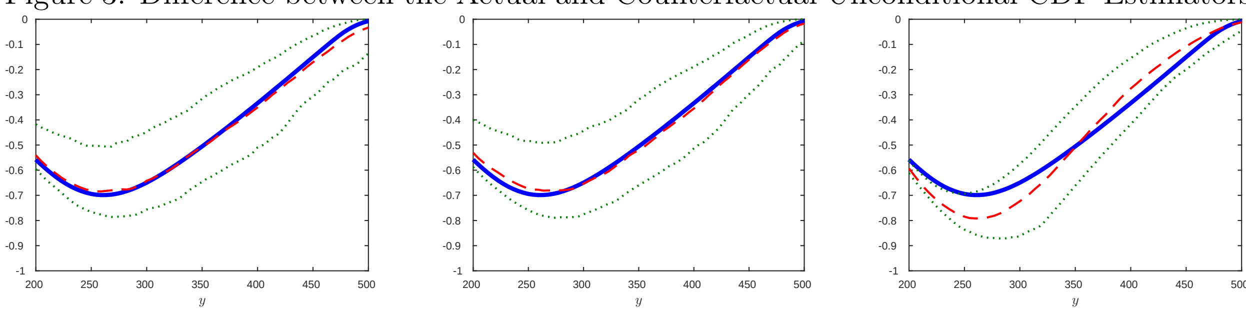
Table 1: Fit of the Copula Distributions

	Parametric	Nonparametric	*Martinez et al.*	$CQF\left(y\vert z\right)$
$\int_{\mathcal{Y}} \left\vert Q_{0.5}\left(\hat{F}_Y\left(y\right)\right) - F_Y\left(y\right)\right\vert dy$	0.002	0.016	0.010	0.013
$\sup_y \left\vert \hat{F}_Y\left(y\right) - F_Y\left(y\right)\right\vert$	0.010	0.022	0.022	0.031
$\int_{\mathcal{Y}} \nabla_{0.025}^{0.975} Q\left(\hat{F}_Y\left(y\right)\right) dy$	0.021	0.015	0.014	0.014
$\sup_y \nabla_{0.025}^{0.975} Q\left(\hat{F}_Y\left(y\right)\right)$	0.054	0.042	0.040	0.042

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.

estimation of the difference between the counterfactual and the actual distributions. The estimators proposed in this paper do a good job at estimating the counterfactual distribution, with the true difference of the distributions being inside the 95% confidence bands. However, the converse is not true for Martinez-Sanchis et al. (2012) estimator, which is particularly biased at the lower tail of the distribution, *i.e.* the part of the distribution most affected by the counterfactual. This is confirmed by difference between the actual counterfactual distribution and the median estimate across repetitions in table 2. Regarding their accuracy, it is very similar for all three estimators.

Figure 3: Difference between the Actual and Counterfactual Unconditional CDF Estimators



In each of the three graphs, the solid line represents the actual distribution of Y, the dashed line represents the median (pointwise) across repetitions of the estimator, and the dotted line represent the 2.5 and 97.5 percentiles (pointwise) across repetitions. The first estimator uses the parametric estimator of the copula, the second one uses the nonparametric estimator of the copula, and the third one is the estimator proposed by Martinez-Sanchis et al. (2012)

Table 2: Fit of the Copula Distributions

	Parametric	Nonparametric	*Martinez et al.*
$\int_{\mathcal{Y}} \left\| \hat{F}_Y(y) - F_Y(y) \right\| dy$	0.009	0.007	0.042
$\sup_y \left\| \hat{F}_Y(y) - F_Y(y) \right\|$	0.031	0.020	0.115
$\int_{\mathcal{Y}} \nabla_{0.025}^{0.975} Q\left(\hat{F}_Y(y)\right) dy$	0.210	0.204	0.137
$\sup_y \nabla_{0.025}^{0.975} Q\left(\hat{F}_Y(y)\right)$	0.337	0.348	0.326

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.

5 Empirical Application

To illustrate this estimation method, I consider the estimation of Engel curves.[17] I evaluate the effect on the distribution of consumption under different counterfactuals involving a redistribution of income or a change in the number of children of the different families. I use the 2012 wave of the Bank of Italy's Survey of Households' Income and Wealth (SHIW), focusing on the subsample of married couples in which the head of family is between 30 and 65 years old, giving us a cross section of size 3238. Following Deaton and Muellbauer (1980), I model the share of food consumption as a function linear in the logarithm of total consumption, plus a set of dummies for the number of children in the household,[18] and three macro regions of Italy[19] To address the endogeneity between food consumption and total consumption I instrument the latter using the logarithm of the household's income.

In line with previous findings in this literature, an increase in total consumption is associated with a decrease in the share of food consumption, as shown in figure 4. This effect, however, is far from constant across quantiles, and it is closer to zero at low quantiles and it increases significantly at high quantiles. Having one child results in an increase of the food consumption share of about two percentage points, and families with two or more children see that share increase by almost four percentage points. This effect does not vary much across quantiles, with the largest variation taking place at quantiles above 0.9. The precision of the estimates at those quantiles is however quite poor. Geographically, the North and Center macroregions have a smaller food share consumption than the South and Isles.

Before explicitly stating some counterfactuals and looking at their distributional effects, let us compare the fit of the estimator of the unconditional cdf of the share of food consumption when the copula is estimated both parametrically and nonparametrically.[20] Figure 5 plots both estimators, as well as the 95% confidence bands which were computed using the

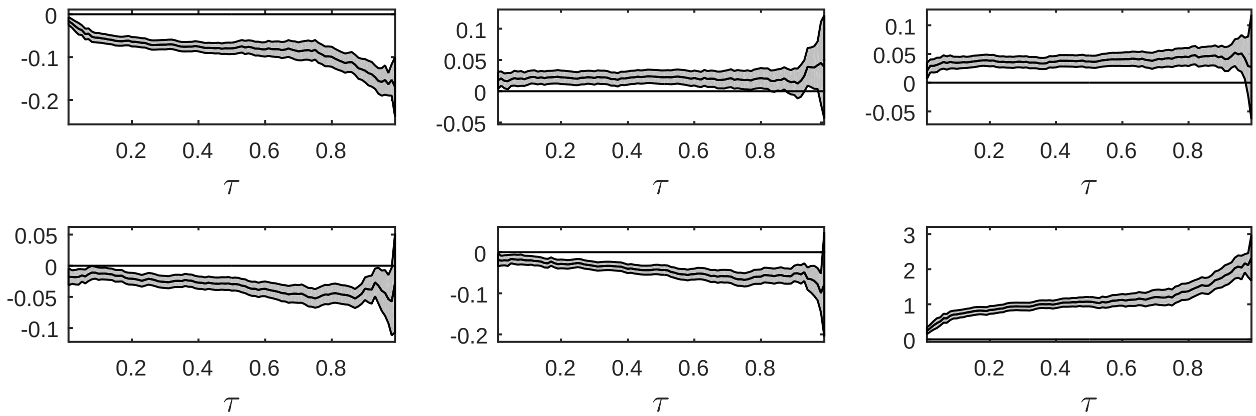
[17]The study of the relation between consumption on a particular set of goods and total expenditure, can be traced back to Engel (1857). Lewbel (2006) describes Engel curves in detail, as well as some of the challenges in their estimation. For more recent developments in the estimation of Engel curves see, for example, Blundell et al. (2007) or Battistin and Nadai (2013).

[18]No children, one child, and at least two children.

[19]North, Center, and South and Isles.

[20]The parametric estimator uses a Clayton copula. See appendix D for details on its choice.

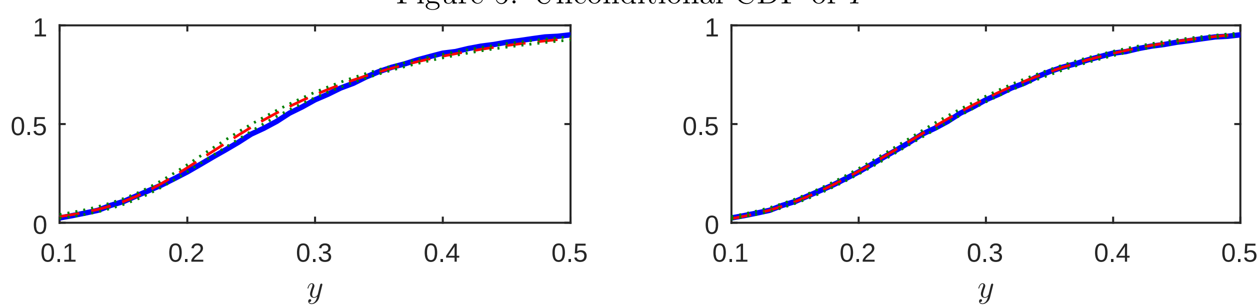
Figure 4: IVQR Estimates



From left to right, the coefficients shown in the first row are for log consumption, a dummy for households with one children, and a dummy for households with two or more children, and those in the second row are dummies for the North and Center regions, and the constant term.

bootstrap. Clearly, the fit of the estimator based on the nonparametric estimator of the copula is the best. The empirical cdf lies always inside the confidence bands, and it is very close to the estimator $\check{F}_Y(y)$. On the other hand, the estimator $\hat{F}_Y(y)$ fails to be a good fit at around the center of the distribution and the right tail.

Figure 5: Unconditional CDF of Y



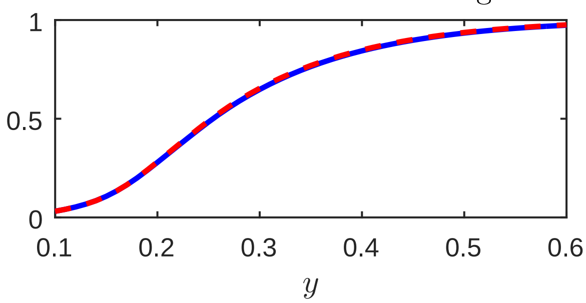
The solid blue line represents the actual distribution of Y; the dashed red line represents the median across repetitions of the estimated distribution of Y with the parametric estimator (left) and the nonparametric estimator (right); the green dotted lines represent the bootstrapped 2.5 and 97.5 percentiles across repetitions of the estimated distribution of Y.

Now consider the following counterfactual: a social planner redistributes income among agents. In particular, each household is taxed a 10% of their income and then they receive a transfer that is an equal share of the total collected income. This counterfactual would reduce the Gini index of income from 0.316 to 0.284.[21] Figure 6 shows the effect of such

[21]Notice that this is the Gini index of the subpopulation considered in this exercise, not of the whole

counterfactual distribution of income on the distribution of food consumption share. The effect of this policy is so tiny, that the counterfactual distribution is almost indistinguishable from actual one. To give some numbers to this change, the Gini index of the share of food consumption would decrease from 0.241 to 0.239 (using the parametric estimator of the copula), and from 0.221 to 0.219 (using the nonparametric estimator of the copula).[22]

Figure 6: Counterfactual 1



The solid blue line represents the estimated distribution of Y; the dashed red line represents the estimated distribution of Y with the parametric estimator (left) and the nonparametric estimator (right) of the counterfactual 1.

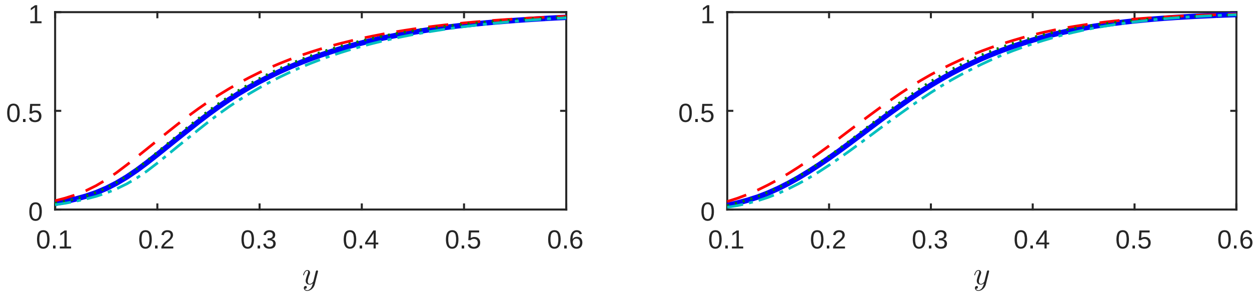
A different counterfactual would be to consider the distribution of food share consumption if all households had the same number of children. Given the classification of households into households with no children, one child, and two or more children, there are three more counterfactuals to consider. Figure 7 shows the estimated distribution of food share consumption for each of the three different counterfactuals. Simple inspection of the figure reveals that increasing the number of children leads to a right displacement of the distribution of food share consumption, *i.e.* an overall increase in food share consumption. This shift, however, is neither parallel, nor of the same magnitude for each of the counterfactuals. Table 3 shows the changes in the Gini index. If every couple had no children, then not only food consumption share would be the lowest, but also the most unequal. On the other hand, if every couple had at least two children, the inequality would be the minimum attained by the considered counterfactuals. Notice that these estimates are quite noisy, and the 95%

population.

[22]The Gini index is not computed for the empirical cdf of the share of food consumption, but for the estimators $\hat{F}_Y(y)$ and $\check{F}_Y(y)$. This is done so to make the counterfactual Gini index comparable to the factual Gini index in both cases.

confidence interval of each of the counterfactuals would include the estimated Gini index with the actual distribution of X.

Figure 7: Counterfactuals 2 to 4



The solid blue line represents the estimated distribution of Y; the dashed red line, the dotted green line, and the dashed-dotted light blue line represent the estimated distributions of Y with the parametric estimator (left) and the nonparametric estimator (right), of the counterfactuals 2 to 4, respectively.

Table 3: Gini Index Estimates

	Actual	Counterfactual 2	Counterfactual 3	Counterfactual 4
$\hat{G}_Y(y)$	0.241	0.252	0.240	0.233
	$(0.230, 0.252)$	$(0.239, 0.264)$	$(0.226, 0.253)$	$(0.221, 0.244)$
$\check{G}_Y(y)$	0.221	0.233	0.220	0.212
	$(0.216, 0.229)$	$(0.224, 0.246)$	$(0.212, 0.230)$	$(0.207, 0.222)$

Notes: The first row represents the actual and counterfactual Gini indices using the parametric estimator of the copula, whereas the second row represents the same using the nonparametric estimator of the copula. Bootstrapped 95% confidence intervals reported in parentheses.

6 Conclusions

In this paper I propose an estimator of actual and 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 that capture 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. I consider three types of counterfactuals: exogenously changing the distribution of the treatment, exogenously changing the distribution of the instrument, and partially affecting the distribution of the treatment by changing the way it is determined, *i.e.* by affecting the structural relation between the treatment and the instrument.

I show the performance of these estimators in a Monte Carlo simulation, comparing them to alternative estimators of the unconditional distribution of the outcome variable. In the empirical application I estimated the Engel curve for food consumption and I considered four different counterfactuals: the first one involved a redistribution of households' income; the other three assumed that every couple had no children, one child, and at least two children, respectively. The first counterfactual had little impact on the inequality in food share consumption, whereas the other three showed that, the more children the couples had, the largest the share of food consumption over total expenditure, and also the most equal the distribution of food share consumption is.

References

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.

Battistin, E. and M. D. Nadai (2013). Identification and estimation of engel curves with endogenous and unobserved expenditures. *Journal of Applied Econometrics 30*, 487–508.

Blundell, R., X. Chen, and D. Kristensen (2007). Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica 75*(6), 1613–1669.

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

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.

Deaton, A. and J. Muellbauer (1980). An almost ideal demand system. *The American economic review 70*(3), 312–326.

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.

Engel, E. (1857). Die vorherrschenden gewerbezweige in den gerichtsämtern mit beziehung auf die productions-und consumtionsverhältnisse des königreichs sachsen. *Zeitschrift des sächsischen Statistischen Bureaus 3*, 129–182.

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

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.

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

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.

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

Lewbel, A. (2006). Engel curves. *The New Palgrave Dictionary of Economics 2*, 1–7.

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.

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.

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.

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

Powell, J. L. (1986). Censored regression quantiles. *Journal of econometrics 32*(1), 143–155.

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.

Appendix

Let $W \equiv (Y, X_1, X_2, Z_1)$. The following notation is used throughout the appendix:[23]

$$
\begin{aligned}
f \mapsto \mathbb{E}_n [f(W)] &\equiv \frac{1}{n} \sum_{i=1}^{n} f(W_i) \\
f \mapsto \mathbb{G}_n [f(W)] &\equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f(W_i) - \mathbb{E}(f(W_i)) \\
\hat{f}(W, \beta, \iota, \gamma, \tau, \theta) &\equiv \begin{bmatrix} \varphi_\tau \left(Y - X'\beta - \hat{\Phi}(\tau)' \iota \right) \hat{\Psi}(\tau) \\ \varphi_\theta (X_1 - Z'\gamma) Z \end{bmatrix} \\
f(W, \beta, \iota, \gamma, \tau, \theta) &\equiv \begin{bmatrix} \varphi_\tau \left(Y - X'\beta - \Phi(\tau)' \iota \right) \Psi(\tau) \\ \varphi_\theta (X_1 - Z'\gamma) Z \end{bmatrix} \\
\hat{g}(W, \beta, \iota, \gamma, \tau, \theta) &\equiv \begin{bmatrix} \rho_\tau \left(Y - X'\beta - \hat{\Phi}(\tau)' \iota \right) \hat{\Psi}(\tau) \\ \rho_\theta (X_1 - Z'\gamma) Z \end{bmatrix} \\
g(W, \beta, \iota, \gamma, \tau, \theta) &\equiv \begin{bmatrix} \rho_\tau \left(Y - X'\beta - \Phi(\tau)' \iota \right) \Psi(\tau) \\ \rho_\theta (X_1 - Z'\gamma) Z \end{bmatrix} \\
Q_n(\beta, \iota, \gamma, \tau, \theta) &\equiv \mathbb{E}_n [\hat{g}(Y, W, \beta, \iota, \gamma, \tau, \theta)] \\
Q(\beta, \iota, \gamma, \tau, \theta) &\equiv \mathbb{E} [g(Y, W, \beta, \iota, \gamma, \tau, \theta)]
\end{aligned}
$$

$\varepsilon = Y - X'\beta$, $\varepsilon(\tau) = Y - X'\beta(\tau)$, $\hat{\varepsilon}(\tau) = Y - X'\hat{\beta}(\tau)$, $\eta = X_1 - Z'\gamma$, $\eta(\theta) = X_1 - Z'\gamma(\theta)$, $\hat{\eta}(\theta) = X_1 - Z'\hat{\gamma}(\theta)$, $\Psi(\tau) \equiv \left(\Phi(\tau)', X_2\right)'$, $\hat{\Psi}(\tau) \equiv \left(\hat{\Phi}(\tau)', X_2\right)'$, $\Phi(\tau) \equiv \Phi(\tau, Z)$, $\hat{\Phi}(\tau) \equiv \hat{\Phi}(\tau, Z)$, $\varphi_\tau(u) \equiv (\mathbf{1}(u < 0) - \tau)$, $\rho_\tau(u) \equiv (\tau - \mathbf{1}(u < 0))\, u$, and $\ell_j(u, v, \xi) \equiv \log(c(u, v | x_{2j}; \xi))$.

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

A Mathematical Proofs

A.1 Proof of Theorem 1

Begin by showing the asymptotic distribution of $\hat{F}_{Y|Z}(y|z)$, for which I add and subtract the unfeasible estimator: $\tilde{F}_{Y|Z}(y|z) \equiv \int_{[0,1]^2} \mathbf{1}\left(\hat{S}_Y(u|z,v) \le y\right) dC_{UV|X_2}(u,v|x_2)$.

$$
\begin{aligned}
&\sqrt{n}\left(\hat{F}_{Y|Z}(y|z) - F_{Y|Z}(y|z)\right) \\
&= \sqrt{n}\left(\hat{F}_{Y|Z}(y|z) - \tilde{F}_{Y|Z}(y|z)\right) + \sqrt{n}\left(\tilde{F}_{Y|Z}(y|z) - F_{Y|Z}(y|z)\right)
\end{aligned}
$$

The first term can be expressed as

$$
\begin{aligned}
&\sqrt{n}\left(\hat{F}_{Y|Z}(y|z) - \tilde{F}_{Y|Z}(y|z)\right) \\
&= \sqrt{n}\int_0^1 \left[\hat{C}_{U|VX_2}\left(\hat{S}_Y^{-1}(y|z,v)|v,x_2\right) - C_{U|VX_2}\left(\hat{S}_Y^{-1}(y|z,v)|v,x_2\right)\right] dv \\
&= \sqrt{n}\int_0^1 \left[\hat{C}_{U|VX_2}\left(S_Y^{-1}(y|z,v)|v,x_2\right) - C_{U|VX_2}\left(S_Y^{-1}(y|z,v)|v,x_2\right)\right] dv + o_P^*(1)
\end{aligned}
$$

where I have used the extended continuous mapping theorem, the uniform continuity of $C_{UV|X_2}(u,v|x_2)$, and the uniform consistency of $\hat{S}_Y(y|z,v)$. As for the second term, by assumption 3, lemmas 1 and 2, the functional chain rule, the extended continuous mapping theorem and the functional delta method

$$
\begin{aligned}
&\sqrt{n}\left(\tilde{F}_{Y|Z}(y|z) - F_{Y|Z}(y|z)\right) \\
&= \sqrt{n}\int_0^1 \left[C_{U|VX_2}\left(\hat{S}_Y^{-1}(y|z,v)|v,x_2\right) - C_{U|VX_2}\left(S_Y^{-1}(y|z,v)|v,x_2\right)\right] dv \\
&= -\sqrt{n}\int_0^1 f_{Y|ZV}(y|z,v)\left[\hat{S}_Y(u(y,z,v)|z,v) - S_Y(u(y,z,v)|z,v)\right] dv + o_P^*(1)
\end{aligned}
$$

Therefore,

$$
\sqrt{n}\left(\hat{F}_{Y|Z}(y|z) - F_{Y|Z}(y|z)\right) \Rightarrow \int_0^1 N(y,z,v)\,\mathbb{G}_{\mathcal{M}}(u(y,z,v),z,v)\,dv
$$

By assumption 1 and the functional delta method,

$$\sqrt{n}\left(\hat{F}_Y(y)-F_Y(y)\right)\Rightarrow\int_{\mathcal{Z}}\int_0^1 N(y,z,v)\,\mathbb{G}_{\mathcal{M}}(u(y,z,v),z,v)\,dvdF_Z(z)$$

□

A.2 Proof of Proposition 1

Start by expanding $\sqrt{n}\left(\hat{S}_Y(u|z,v)-S_Y(u|z,v)\right)$ around $(\gamma(v),\beta(u))$

$$\begin{aligned}\sqrt{n}\left(\hat{S}_Y(u|z,v)-S_Y(u|z,v)\right)&=\sqrt{n}\left[x(v)'\left(\hat{\beta}(u)-\beta(u)\right)+(\hat{x}(v)-x(v))'\hat{\beta}(u)\right]\\&=\sqrt{n}\left[x(v)'\left(\hat{\beta}(u)-\beta(u)\right)+\beta_1(u)\,z'(\hat{\gamma}(v)-\gamma(v))\right]\\&+o_P^*(1)\end{aligned}$$

By lemma 3 and the functional delta method it follows that

$$\sqrt{n}\left(\hat{x}(v)'\hat{\beta}(u)\,x(v)'\beta(u)\right)\Rightarrow\mathbb{G}_{\mathcal{K}}(u,v,z)$$

□

A.3 Proof of Proposition 2

For the estimator of the copula I have that

$$\begin{aligned}\sqrt{n}\left(\hat{C}_{U|VX_2}(u|v,x_2)-C_{U|VX_2}(u|v,x_2)\right)&\equiv\sqrt{n}\left(C_{U|VX_2}\left(u|v,x_2;\hat{\xi}\right)-C_{U|VX_2}(u|v,x_2;\xi)\right)\\&=\frac{\partial}{\partial\xi}C_{U|VX_2}\left(u|v,x_2;\overline{\xi}\right)\sqrt{n}\left(\hat{\xi}-\xi\right)\\&=\frac{\partial}{\partial\xi}C_{U|VX_2}(u|v,x_2;\xi)\sqrt{n}\left(\hat{\xi}-\xi\right)+o_P^*(1)\end{aligned}$$

where the first equality follows by a mean value expansion around ξ, and the second by the consistency of $\hat{\xi}$ and the continuous mapping theorem. Together with proposition 1, apply the functional delta method to lemma 6 and it follows that

$$\sqrt{n}\begin{pmatrix}\hat{S}_Y\left(u|z,v\right)-S_Y\left(u|z,v\right)\\ \hat{C}_{U|VX_2}\left(u|v,x_2\right)-C_{U|VX_2}\left(u|v,x_2\right)\end{pmatrix}\Rightarrow\mathbb{G}_{\mathcal{N}}\left(u,v,z\right)=N\left(u,v,z\right)\mathbb{G}_{\mathcal{M}}\left(u,v\right)$$

□

A.4 Proof of Proposition 3

$$\begin{aligned}
&\sup_{y,z}\sqrt{n}\left|\check{F}_Y\left(y|z\right)-F_Y\left(y|z\right)\right|\\
&\leq\sup_{y,z}\sqrt{n}\left|\check{F}_{Y|Z}\left(y|z\right)-\tilde{F}_{Y|Z}\left(y|z\right)\right|+\sup_{y,z}\sqrt{n}\left|\tilde{F}_{Y|Z}\left(y|z\right)-F_{Y|Z}\left(y|z\right)\right|\\
&=\sup_{y,z}\sqrt{n}\left|\int_0^1\mathbf{1}\left(\hat{S}_Y\left(u|z,v\right)\leq y\right)d\left(\check{C}_{UV}\left(u,v\right)-C_{UV}\left(u,v\right)\right)\right|+O_P^*\left(1\right)\\
&\leq\sqrt{n}\int_0^1 d\left|\check{C}_{UV}\left(u,v\right)-C_{UV}\left(u,v\right)\right|+O_P^*\left(1\right)\\
&\leq\sup_{u,v}\sqrt{n}\left|\check{C}_{UV}\left(u,v\right)-C_{UV}\left(u,v\right)\right|+O_P^*\left(1\right)=O_P^*\left(1\right)
\end{aligned}$$

where the first inequality follows from the triangle inequality, the first equality from the definition of the estimators and the uniform consistency of $\tilde{F}_{Y|Z}\left(y|z\right)$ shown in theorem 1, 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 7.

□

B Auxiliary Lemmas

B.1 Hadamard Derivative of $F_{Y|ZV}(y|z,v)$ with Respect to $S_Y(u|z,v)$

Lemma 1. *Define* $F_Y(y|z,v,h_t) \equiv \int_0^1 \mathbf{1}(S_Y(u|z,v)+th_t(u|z,v)\leq y)\,dC_{U|VX_2}(u|v,x_2)$. *Under assumption 2, as* $t\searrow 0$,

$$D_{h_t}(y|z,v,h_t)=\frac{F_{Y|ZV}(y|z,v,h_t)-F_{Y|ZV}(y|z,v)}{t}\to D_h(y|z,v)$$

where $D_h(y|z,v)\equiv -f_{Y|ZV}(y|z,v)\,h\left(C^{-1}_{U|VX_2}\left(F_{Y|ZV}(y|z,v)\,|v,x_2\right)|z,v\right)$. *The convergence holds uniformly in any compact subset of* $\mathcal{YZV}$ *for any* $h_t:\|h_t-h\|_\infty\to 0$, *where* $\mathcal{YZV}\equiv\{(y,z,v):y\in\mathcal{Y}_z,z\in\mathcal{Z},\ \ v\in[0,1]\}$ *and* $h_t\in\ell^\infty(\mathcal{UZV})$ *and* $h\in C(\mathcal{UZV})$.

Proof. $\forall\delta>0\exists\epsilon>0$ such that if $u\in B_\epsilon\left(C^{-1}_{U|VX_2}(F_Y(y|z,v)\,|v,x_2)\right)$ and $t\geq 0$ small enough,

$$\begin{aligned}&\mathbf{1}(S_Y(u|z,v)+th_t(u|z,v)\leq y)\\&\leq\mathbf{1}\left(S_Y(u|z,v)+t\left[h\left(C^{-1}_{U|VX_2}\left(F_{Y|ZV}(y|z,v)\,|v,x_2\right)|z,v\right)-\delta\right]\leq y\right)\end{aligned}$$

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

$$\mathbf{1}(S_Y(u|z,v)+th_t(u|z,v)\leq y)=\mathbf{1}(S_Y(u|z,v)\leq y)$$

So for small enough $t\geq 0$,

$$\begin{aligned}&\tfrac{1}{t}\int_0^1[\mathbf{1}(S_Y(u|z,v)+th_t(u|z,v)\leq y)-\mathbf{1}(S_Y(u|z,v)\leq y)]\,C_{U|VX_2}(u|v,x_2)\,du\\\leq&\tfrac{1}{t}\int_{B_\epsilon(F_Y(y|z,v))}[\mathbf{1}(S_Y(u|z,v)+th_t(u|z,v)\leq y)-\mathbf{1}(S_Y(u|z,v)\leq y)]\,C_{U|VX_2}(u|v,x_2)\,du\end{aligned}\tag{17}$$

Let $\tilde{y}=S_Y(u|z,v)$, so that $u=C^{-1}_{U|VX_2}\left(F_{Y|ZV}(\tilde{y}|z,v,)\,|v,x_2\right)$ and J be the image of

$B_\epsilon\left(F_{Y|ZV}\left(y|z,v\right)\right)$ under $u\rightarrow S_Y\left(u|z,v\right)$. Then, equation 17 equals

$$\frac{1}{t}\int_{J\cap\left[y,y-t\left(h\left(F_{Y|ZV}(y|z,v)|z,v\right)-\delta\right)\right]}f_{Y|ZV}\left(\tilde{y}|z,v\right)d\tilde{y}$$

For fixed ϵ and $t\searrow 0$

$$\begin{aligned}&J\cap\left[y,y-t\left(h\left(C^{-1}_{U|VX_2}\left(F_{Y|ZV}\left(y|z,v\right)|v,x_2\right)|z,v\right)-\delta\right)\right]\\&=\left[y,y-t\left(h\left(C^{-1}_{U|VX_2}\left(F_{Y|ZV}\left(y|z,v\right)|v,x_2\right)|z,v\right)-\delta\right)\right]\end{aligned}$$

$$f_{Y|ZV}\left(\tilde{y}|z,v\right)\rightarrow f_{Y|ZV}\left(y|z,v\right)$$

as $F_{Y|ZV}\left(\tilde{y}|z,v\right)\rightarrow F_{Y|ZV}\left(y|z,v\right)$. Therefore, the right hand term in equation 17 is no greater than

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

Similarly, $-f_{Y|ZV}\left(y|z,v\right)\left(h\left(C^{-1}_{U|VX_2}\left(F_{Y|ZV}\left(y|z,v\right)|v,x_2\right)|z,v\right)+\delta\right)+o\left(1\right)$ bounds equation 17 from below. Since δ can be arbitrarily small, the result follows.

To show uniformity of this result, apply Lemma 5 in Chernozhukov et al. (2013). Let $(y,z,v)\in K$, where K is a compact subset of $\mathcal{YZV}$. Take a sequence (y_t,z_t,v_t) in K that converges to $(y,z,v)\in K$, since the function

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

is uniformly continuous on K it follows that the preceding argument applies to this sequence. This result follows by the assumed continuity of $h\left(u|z,v\right)$, $F_{Y|ZV}\left(y|z,v\right)$ and $f_{Y|ZV}\left(y|z,v\right)$ in all of its arguments, and the compactness of K. □

B.2 Hadamard Derivative of $F_{Y|Z}(y|z)$ with Respect to $F_{Y|Z}(y|z,v)$

Lemma 2. *Define* $F_{Y|Z}(y|z,h_t) \equiv \int_0^1 \left[F_{Y|Z}(y|z,v) + th_t(y|z,v)\right] dv$*. As* $t \searrow 0$*,*

$$D_{h_t}(y|z,h_t) = \frac{F_{Y|Z}(y|z,h_t) - F_{Y|Z}(y|z)}{t} \to D_h(y|z)$$

where $D_h(y|z) \equiv \int_0^1 h(y|z,v)\,dv$*. The convergence holds uniformly in any compact subset of* $\mathcal{YZ} \equiv \{(y,z) : y \in \mathcal{Y}_z, z \in \mathcal{Z},\}$ *for any* $h_t : \|h_t - h\|_\infty \to 0$*, where* $h_t \in \ell^\infty(\mathcal{UZ})$ *and* $h \in C(\mathcal{UZ})$*.*

Proof.

$$\begin{aligned} D_{h_t}(y|z,h_t) &= \frac{F_{Y|Z}(y|z,h_t) - F_{Y|Z}(y|z)}{t} \\ &= \frac{1}{t}\int_0^1 \left[F_{Y|Z}(y|z,v) + th_t(y|z,v) - F_{Y|Z}(y|z,v)\right] dv \\ &= \int_0^1 h_t(y|z,v)\,dv \to \int_0^1 h(y|z,v)\,dv \end{aligned}$$

□

B.3 Asymptotic Distribution of the IVQR and QR Estimators

The proof of the following lemma is an extension of the proof in Chernozhukov and Hansen (2006) to account for the joint distribution of both estimators.

Lemma 3. *Let* $\hat{\gamma}(v)$ *and* $\hat{\beta}(u)$ *denote the conditional QR and conditional IVQR estimators of quantiles* v *and* u *of equations 2 and 1, respectively. Under assumptions 1, 2 and 4 to 7, their joint asymptotic distribution is given by:*

$$\sqrt{n}\left[\begin{pmatrix}\hat{\beta}(u) \\ \hat{\gamma}(v)\end{pmatrix} - \begin{pmatrix}\beta(u) \\ \gamma(v)\end{pmatrix}\right] \Rightarrow \mathbb{G}_{\mathcal{J}}(u,v)$$

where $\mathbb{G}_{\mathcal{J}}(u,v)$ *is a zero-mean Gaussian process with covariance function* $\Sigma_{\mathcal{J}}(u,v,\tilde{u},\tilde{v})$*,* *given by:*

$$\Sigma_{\mathcal{J}}\left(u,v,\tilde{u},\tilde{v}\right)=\begin{bmatrix}\Sigma_{\mathcal{J}}^{11}\left(u,\tilde{u}\right) & \Sigma_{\mathcal{J}}^{21}\left(u,\tilde{v}\right)' \\ \Sigma_{\mathcal{J}}^{21}\left(\tilde{u},v\right) & \Sigma_{\mathcal{J}}^{22}\left(v,\tilde{v}\right)\end{bmatrix}$$

where

$$\begin{aligned}
\Sigma_{\mathcal{J}}^{11}\left(u,\tilde{u}\right) &\equiv J\left(u\right)^{-1}\left(u\vee\tilde{u}-u\tilde{u}\right)\mathbb{E}\left[\Psi\left(u,z\right)\Psi\left(\tilde{u}',z\right)\right]J\left(\tilde{u}\right)^{-1} \\
\Sigma_{\mathcal{J}}^{21}\left(\tilde{u},v\right) &\equiv H\left(v\right)^{-1}\mathbb{E}\left[\left(\mathbf{1}\left(y\leq x'\beta\left(\tilde{u}\right)\right)\mathbf{1}\left(x_1\leq z'\gamma\left(v\right)\right)-\tilde{u}v\right)z\Psi\left(\tilde{u},z\right)'\right]'J\left(\tilde{u}\right)^{-1} \\
\Sigma_{\mathcal{J}}^{22}\left(v,\tilde{v}\right) &\equiv H\left(v\right)^{-1}\left(v\vee\tilde{v}-v\tilde{v}\right)\mathbb{E}\left[zz'\right]H\left(\tilde{v}\right)^{-1} \\
H\left(v\right) &\equiv \mathbb{E}\left[f_{X_1}\left(z'\gamma\left(v\right)|z\right)zz'\right] \\
J\left(u\right) &\equiv \mathbb{E}\left[f_Y\left(x'\beta\left(u\right)|x,z_1\right)\Psi\left(u,z\right)x'\right]
\end{aligned}$$

Proof. **Step 1** (Consistency) By assumption 2, $Q\left(\beta,\iota,\gamma,\tau,\theta\right)$ is continuous over $\mathcal{B}\times\mathcal{I}\times\mathcal{G}\times\mathcal{T}\times\mathcal{C}$. Furthermore, by lemma 5, $\sup_{(\beta,\iota,\gamma)\in\mathcal{B}\times\mathcal{I}\times\mathcal{G}}\left\|Q_n\left(\beta,\iota,\gamma,\tau,\theta\right)-Q\left(\beta,\iota,\gamma,\tau,\theta\right)\right\|\xrightarrow{p}0$.

By lemma 4, have uniform convergence of $\sup_{(\beta_1,\tau,\theta)\in\mathcal{B}_1\times\mathcal{T}\times\mathcal{C}}\left\|\hat{\vartheta}\left(\beta_1,\tau,\theta\right)-\vartheta\left(\beta_1,\tau,\theta\right)\right\|\xrightarrow{p}0$, which by lemma 4 implies that $\sup_{(\beta_1,\tau)\in\mathcal{B}_1\times\mathcal{T}}\left\|\left\|\hat{\iota}\left(\beta_1,\tau\right)\right\|_{B_1(\tau)}-\left\|\iota\left(\beta_1,\tau\right)\right\|_{B_1(\tau)}\right\|\xrightarrow{p}0$.

By lemma 4, $\sup_{\tau\in\mathcal{T}}\left\|\hat{\beta}_1\left(\tau\right)-\beta_1\left(\tau\right)\right\|\xrightarrow{p}0$, and therefore $\sup_{\tau\in\mathcal{T}}\left\|\hat{\beta}_2\left(\tau\right)-\beta_2\left(\tau\right)\right\|\xrightarrow{p}0$, $\sup_{\tau\in\mathcal{T}}\left\|\hat{\iota}\left(\hat{\beta}_1\left(\tau\right)\tau\right)-0\right\|\xrightarrow{p}0$ and $\sup_{\theta\in\mathcal{C}}\left\|\hat{\gamma}\left(\theta\right)-\gamma\left(\theta\right)\right\|\xrightarrow{p}0$.

Step 2 (Asymptotics) Consider a collection of closed balls $B_{\delta_n}\left(\beta_1\left(\tau\right)\right)$ centered at $\beta_1\left(\tau\right)\forall\tau$, δ_n independent of τ and $\delta_n\to 0$ slowly enough. Let $\beta_{1n}\left(\tau\right)$ be any value inside $B_{\delta_n}\left(\beta\left(\tau\right)\right)$. By Theorem 3.3 in Koenker and Bassett (1978),

$$O\left(\frac{1}{\sqrt{n}}\right)=\sqrt{n}\mathbb{E}\hat{f}\left(W,\beta_{1n}\left(\cdot\right),\hat{\vartheta}\left(\beta_{1n}\left(\cdot\right),\cdot,\cdot\cdot\right),\cdot,\cdot\cdot\right)$$

By lemma 5, the following expansion holds for any $\sup_{\tau\in\mathcal{T}}\|\beta_{1n}(\tau)-\beta_1(\tau)\| \xrightarrow{p} 0$

$$
\begin{aligned}
O\left(\frac{1}{\sqrt{n}}\right) &= \mathbb{G}_n\hat{f}\left(W,\beta_{1n}(\cdot),\hat{\vartheta}\left(\beta_{1n}(\cdot),\cdot,\cdot\cdot\right),\cdot,\cdot\cdot\right)+\sqrt{n}\mathbb{E}\hat{f}\left(W,\beta_{1n}(\cdot),\hat{\vartheta}_n\left(\beta_{1n}(\cdot),\cdot,\cdot\cdot\right),\cdot,\cdot\cdot\right)\\
&= \mathbb{G}_n\hat{f}\left(W,\beta_1(\cdot),\vartheta\left(\beta_1(\cdot),\cdot,\cdot\cdot\right),\cdot,\cdot\cdot\right)+o_P(1)\\
&+\sqrt{n}\mathbb{E}\hat{f}\left(W,\beta_{1n}(\cdot),\hat{\vartheta}_n\left(\beta_{1n}(\cdot),\cdot,\cdot\cdot\right),\cdot,\cdot\cdot\right)\ in\ \ell^\infty\left(\mathcal{T}\times\mathcal{C}\right)\\
&= \mathbb{G}_n\hat{f}\left(W,\beta_1(\cdot),\vartheta\left(\beta_1(\cdot),\cdot,\cdot\cdot\right),\cdot,\cdot\cdot\right)+o_P(1)\\
&+\left(J_\vartheta(\cdot,\cdot\cdot)+o_P(1)\right)\sqrt{n}\left(\hat{\vartheta}\left(\beta_{1n}(\cdot),\cdot,\cdot\cdot\right)-\vartheta(\cdot,\cdot\cdot)\right)\\
&+\left(J_{\beta_1}(\cdot)+o_P(1)\right)\sqrt{n}\left(\beta_{1n}(\cdot)-\beta_1(\cdot)\right)\ in\ \ell^\infty\left(\mathcal{T}\times\mathcal{C}\right)
\end{aligned}
$$

where

$$
J_\vartheta(\cdot,\cdot\cdot)\equiv\frac{\partial}{\partial\left(\beta_2',\iota',\gamma'\right)}\mathbb{E}\left.\begin{bmatrix}\varphi_\cdot\left(Y-X_1'\beta_1(\cdot)-X_2'\beta_2-\Phi(\cdot)'\iota\right)\Psi(\cdot)\\ \varphi_{\cdot\cdot}\left(X_1-Z'\gamma\right)z\end{bmatrix}\right|_{\vartheta=\vartheta(\cdot,\cdot\cdot)}
$$

$$
J_{\beta_1}(\cdot)\equiv\begin{bmatrix}\frac{\partial}{\partial\beta_1}\mathbb{E}\left[\varphi_\cdot\left(Y-X_1'\beta_1-X_2'\beta_2(\cdot)\right)\Psi(\cdot)\right]\Big|_{\beta_1=\beta_1(\cdot)}\\ 0_{dim(\mathcal{X})\times 1}\end{bmatrix}
$$

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

$$
\begin{aligned}
\sqrt{n}\left(\hat{\vartheta}\left(\beta_{1n}(\cdot),\cdot,\cdot\cdot\right)-\vartheta(\cdot,\cdot\cdot)\right) &= -J_\vartheta^{-1}(\cdot,\cdot\cdot)\,\mathbb{G}_nf\left(Y,W,\beta_1(\cdot),\vartheta(\cdot,\cdot\cdot),\cdot,\cdot\cdot\right)\\
&-J_\vartheta^{-1}(\cdot,\cdot\cdot)\,J_{\beta_1}(\cdot)\left[1+o_P(1)\right]\sqrt{n}\left(\beta_{1n}(\cdot)-\beta_1(\cdot)\right)+o_P(1)
\end{aligned}
$$

in $\ell^\infty\left(\mathcal{T}\times\mathcal{C}\right)$. So I have

$$
\sqrt{n}\left(\hat{\iota}\left(\beta_{1n}(\cdot),\cdot\right)-0\right)-\bar{J}_\iota(\cdot,\cdot\cdot)\,\mathbb{G}_nf\left(Y,W,\beta_1(\cdot),\vartheta(\cdot,\cdot\cdot),\cdot,\cdot\cdot\right)-\bar{J}_\iota(\cdot,\cdot\cdot)\,J_{\beta_1}(\cdot)\left[1+o_P(1)\right]
$$

in $\ell^\infty\left(\mathcal{T}\times\mathcal{C}\right)$, where $\left[\bar{J}_{\beta_2}(\cdot,\cdot\cdot)':\bar{J}_\iota(\cdot,\cdot\cdot)':\bar{J}_\gamma(\cdot,\cdot\cdot)'\right]$ is the comfortable partition of $J_\vartheta^{-1}(\cdot,\cdot\cdot)$.

By step 1, $wp \to 1$,

$$\hat{\beta}_1(\tau) = \arg \inf_{\beta_{1n}(\tau) \in B_n(\beta_1(\tau))} \|\hat{\iota}(\beta_{1n}(\tau), \tau)\|_{B_1(\tau)} \ \forall \tau \in \mathcal{T}$$

By lemma 5, $\mathbb{G}_n f(Y, W, \beta_1(\cdot), \vartheta(\cdot, \cdot\cdot), \cdot, \cdot\cdot) = O_p(1)$, so it follows that

$$\sqrt{n}\|\hat{\iota}(\beta_{1n}(\cdot), \cdot)\|_{B_1(\cdot)} = \left\|O_p(1) - \bar{J}_\iota(\cdot, \cdot\cdot) J_{\beta_1}(\cdot)[1 + o_P(1)]\sqrt{n}(\beta_{1n}(\cdot) - \beta_1(\cdot))\right\|_{B_1(\cdot)}$$

in $\ell^\infty(\mathcal{T} \times \mathcal{C})$. Hence,

$$\begin{aligned}\sqrt{n}\left(\hat{\beta}_1(\cdot) - \beta_1(\cdot)\right) &= \arg\inf_{\mu \in \mathbb{R}} \left\|-\bar{J}_\iota(\cdot)\mathbb{G}_n f(Y, W, \beta_1(\cdot), \vartheta(\cdot, \cdot\cdot), \cdot, \cdot\cdot) - \bar{J}_\iota(\cdot, \cdot\cdot)\bar{J}_{\beta_1}(\cdot)\mu\right\|_{B_1(\cdot)} \\ &\quad + o_P(1)\end{aligned}$$

in $\ell^\infty(\mathcal{T} \times \mathcal{C})$. So jointly in $\ell^\infty(\mathcal{T} \times \mathcal{C})$

$$\begin{aligned}\sqrt{n}\left(\hat{\beta}_1(\cdot) - \beta_1(\cdot)\right) &= -\left(J_{\beta_1}(\cdot)' \bar{J}_\iota(\cdot, \cdot\cdot)' B_1(\cdot) \bar{J}_\iota(\cdot, \cdot\cdot) J_{\beta_1}(\cdot)\right)^{-1} \\ &\quad \cdot \left(J_{\beta_1}(\cdot)' \bar{J}_\iota(\cdot, \cdot\cdot)' B_1(\cdot) \bar{J}_\iota(\cdot, \cdot\cdot)\right)\mathbb{G}_n f(Y, W, \beta_1(\cdot), \vartheta(\cdot, \cdot\cdot), \cdot, \cdot\cdot) + o_P(1) \\ &= O_p(1)\end{aligned}$$

$$\begin{aligned}&\sqrt{n}\left(\hat{\vartheta}\left(\hat{\beta}_1(\cdot), \cdot, \cdot\cdot\right) - \vartheta(\cdot, \cdot\cdot)\right) \\ &= -J_\vartheta^{-1}(\cdot, \cdot\cdot)\Big[I - J_{\beta_1}(\cdot)\left(J_{\beta_1}(\cdot)' \bar{J}_\iota(\cdot, \cdot\cdot)' B_1(\cdot) \bar{J}_\iota(\cdot, \cdot\cdot) J_{\beta_1}(\cdot)\right)^{-1} \\ &\cdot J_{\beta_1}(\cdot)' \bar{J}_\iota(\cdot, \cdot\cdot)' B_1(\cdot) \bar{J}_\iota(\cdot, \cdot\cdot)\Big]\mathbb{G}_n f(Y, W, \beta_1(\cdot), \vartheta(\cdot, \cdot\cdot), \cdot, \cdot\cdot) + o_P(1) = O_P(1)\end{aligned} \tag{18}$$

Due to invertibility of $J_{\beta_1}(\tau)\,\bar{J}(\tau,\theta)$,

$$
\begin{aligned}
&\sqrt{n}\left(\hat{\iota}\left(\hat{\beta}_1(\cdot),\cdot\right)-0\right)\\
&=-\bar{J}_\iota(\cdot,\cdot\cdot)\left[I-J_{\beta_1}(\cdot)\left[J_{\beta_1}(\cdot)'\,\bar{J}_\iota(\cdot,\cdot\cdot)'\right]^{-1}\bar{J}_\iota(\cdot,\cdot\cdot)\right]\mathbb{G}_n f\left(W,\beta_1(\cdot),\vartheta(\cdot,\cdot\cdot),\cdot,\cdot\cdot\right)+o_P(1)\\
&=0\times O_P(1)+o_P(1) \qquad (19)
\end{aligned}
$$

in $\ell^\infty(\mathcal{T}\times\mathcal{C})$. Because $\left(\beta_{1n}(\cdot),\hat{\vartheta}(\beta_{1n}(\cdot),\cdot,\cdot\cdot)\right)=\left(\hat{\beta}_1(\cdot),\hat{\beta}_2(\cdot),0+o_P\left(\frac{1}{\sqrt{n}}\right),\hat{\gamma}(\cdot\cdot)\right)$, and if I substitute it into the expansion, I have:

$$
-\mathbb{G}_n f\left(W,\beta_1(\cdot),\vartheta(\cdot,\cdot\cdot),\cdot,\cdot\cdot\right)=\begin{bmatrix} J(\cdot) & 0_{dim(\mathcal{X})}\\ 0_{dim(\mathcal{X})} & H(\cdot\cdot)\end{bmatrix}\sqrt{n}\begin{pmatrix}\hat{\beta}(\cdot)-\beta(\cdot)\\ \hat{\gamma}(\cdot\cdot)-\gamma(\cdot\cdot)\end{pmatrix}+o_P(1)
$$

in $\ell^\infty(\mathcal{T}\times\mathcal{C})$. By lemma 5, $\mathbb{G}_n f\left(W,\beta_1(\cdot),\vartheta(\cdot,\cdot\cdot),\cdot,\cdot\cdot\right)\Rightarrow\mathbb{G}_{\mathcal{G}}(\cdot,\cdot\cdot)$ *in* $\ell^\infty(\mathcal{T}\times\mathcal{C})$, a Gaussian process with covariate function $S(\tau,\theta,\tau',\theta')=\mathbb{E}\left[\mathbb{G}_{\mathcal{G}}(\tau,\theta)\,\mathbb{G}_{\mathcal{G}}(\tau',\theta')'\right]$, which yields

$$
\sqrt{n}\begin{pmatrix}\hat{\beta}(\cdot)-\beta(\cdot)\\ \hat{\gamma}(\cdot\cdot)-\gamma(\cdot\cdot)\end{pmatrix}\Rightarrow\begin{bmatrix} J(\cdot)^{-1} & 0_{dim(\mathcal{X})}\\ 0_{dim(\mathcal{X})} & H(\cdot)^{-1}\end{bmatrix}\mathcal{G}(\cdot,\cdot\cdot)=\mathbb{G}_{\mathcal{J}}(\tau,\theta)\ \ in\ \ell^\infty(\mathcal{T}\times\mathcal{C})
$$

□

B.4 Argmax Process

Lemma 4. ***(Chernozhukov and Hansen, 2004)*** *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 (2005). □

B.5 Stochastic Expansion

Lemma 5. *Under assumptions 1 and 4 to 7, the following statements hold:*

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

2. $\mathbb{G}_n f\left(W,\beta\left(\cdot\right),0,\gamma\left(\cdot\cdot\right),\cdot,\cdot\cdot\right) \Rightarrow \mathbb{G}\mathcal{G}\left(\cdot,\cdot\cdot\right)$ *in* $\ell^\infty\left(\mathcal{T},\mathcal{C}\right)$*, where* $\mathbb{G}_\mathcal{G}$ *is a Gaussian process with covariance function* $S\left(\left(\tau,\theta\right),\tau',\theta'\right)$ *defined below in the proof.*

   Furthermore, for any $\sup_{(\tau,\theta)\in\mathcal{T}\times\mathcal{C}}\left\|\left(\hat{\beta}\left(\tau\right),\hat{\iota}\left(\tau\right),\hat{\gamma}\left(\theta\right)\right)-\left(\beta\left(\tau\right),0,\gamma\left(\theta\right)\right)\right\| = o_P\left(1\right)$, $\sup_{(\tau,\theta)\in\mathcal{T}\times\mathcal{C}}\left\|\mathbb{G}_n\hat{f}\left(W,\hat{\beta}\left(\tau\right),\hat{\iota}\left(\tau\right),\hat{\gamma}\left(\theta\right),\tau,\theta\right)-\mathbb{G}_n f\left(W,\beta\left(\tau\right),0,\gamma\left(\theta\right),\tau,\theta\right)\right\| = o_P\left(1\right)$

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

$$\mathcal{H}\equiv\left\{h=\left(\Phi,\Psi,\pi,\tau,\theta\right)\mapsto\begin{bmatrix}\varphi_\tau\left(Y-X'\beta-\Phi\left(Z\right)'\iota\right)\Psi\left(Z\right)\\ \varphi_\theta\left(X_1-Z'\gamma\right)Z\end{bmatrix}\pi\in\Pi,\Phi,\Psi\in\mathcal{F}\right\}$$

where $\mathcal{F}$ is the class of uniformly smooth functions in z with the uniform smoothness order $\omega<\frac{\dim(w)}{2}$ and $\left\|f\left(\tau',z\right)-f\left(\tau,z\right)\right\|<C\left(\tau-\tau'\right)^a, C>0, a>0\,\forall\left(z,\tau,\tau'\right)\forall f\in\mathcal{F}$. $\mathcal{H}$ is Donsker, and the bracketing number of $\mathcal{F}$, by Corollary 2.7.4 in van der Vaart and Wellner (1996) satisfies

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

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

$$\mathcal{D}_1\equiv\left\{\left(\Phi,\pi\right)\to\left(X'\beta+\Phi\left(X,Z\right)'\iota\right),\pi\in\Pi,\Phi\in\mathcal{F}\right\}$$

satisfies

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

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

$$\mathcal{D}_2 \equiv \{(\pi) \to (Z'\gamma), \pi \in \Pi\}$$

satisfies

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

for some $\delta''' < 0$ such that $\delta''' < \delta''$. Since the indicator function is bounded and monotone, and the density functions $f_{Y|X_1Z}(y)$ and $f_{X_1|Z}(x_1)$ are bounded by assumption 2, then I have that the bracketing number of

$$\mathcal{E} \equiv \left\{(\Phi, \pi) \to \mathbf{1}\left(Y < X'\beta + \Phi(X, Z)'\iota\right) + \mathbf{1}(X_1 < Z'\gamma), \pi \in \Pi, \Phi \in \mathcal{F}\right\}$$

satisfies

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

Since $\mathcal{E}$ has a constant envelope, it is Donsker. Let $\mathcal{T} \equiv \{\tau \to \tau\}$ and $\mathcal{C} \equiv \{\theta \mapsto \theta\}$. Then I have that $\mathcal{H} \equiv \mathcal{T} \times \mathcal{F} + \mathcal{C} \times \mathcal{F} - \mathcal{E} \times \mathcal{F}$. Since $\mathcal{H}$ is Lipschitz over $(\mathcal{T} \times \mathcal{C} \times \mathcal{F} \times \mathcal{E})$, it follows that it is Donsker by Theorem 2.10.6 in van der Vaart and Wellner (1996).

Define

$$h \equiv (\Phi, \Psi, \pi, \tau, \theta) \mapsto \mathbb{G}_n \begin{bmatrix} \varphi_\tau\left(\varepsilon - \Phi(Z)'\iota\right)\Psi(Z) \\ \varphi_\theta(\eta)Z) \end{bmatrix}$$

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

$$(\tau,\theta)\mapsto\mathbb{G}_n\begin{bmatrix}\varphi_\tau\left(\epsilon-\Phi(Z)'\iota\right)\Psi(Z)\\ \varphi_\theta(\eta)Z\end{bmatrix}$$

By the uniform Hölder continuity of $(\tau,\theta)\mapsto\left(\tau,\beta(\tau)',\Phi(\tau,Z)',\Psi(\tau,Z)',\theta,\gamma(\theta)'\right)'$ in (τ,θ) with respect to the supremum norm, it is also Donsker in $\ell^\infty(\mathcal{H})$. Therefore, I have

$$\mathbb{G}_n\begin{bmatrix}\varphi_{\cdot}(\varepsilon(\cdot))\Psi(\cdot,Z)\\ \varphi_{\cdot\cdot}(\eta(\cdot\cdot))Z\end{bmatrix}\Rightarrow\mathbb{G}_{\mathcal{G}}(\cdot,\cdot\cdot)$$

with covariate function

$$S(\tau,\theta,\tau',\theta')=\mathbb{E}\left[\mathbb{G}_{\mathcal{G}}(\tau,\theta)\,\mathbb{G}_{\mathcal{G}}(\tau',\theta')'\right]\equiv\begin{bmatrix}S^{11}(\tau,\tilde{\tau}) & S^{21}\left(\tau,\tilde{\theta}\right)'\\ S^{21}(\tilde{\tau},\theta) & S^{22}\left(\theta,\tilde{\theta}\right)\end{bmatrix}$$

where

$$\begin{aligned}S^{11}(\tau,\tilde{\tau})&=(\tau\vee\tilde{\tau}-\tau\tilde{\tau})\,\mathbb{E}\left[\Psi(\tau,Z)\Psi(\tilde{\tau},Z)'\right]\\ S^{21}(\tilde{\tau},\theta)&=\mathbb{E}\left[\left(\mathbf{1}(y\le x'\beta(\tilde{\tau}))\,\mathbf{1}(x_1\le z'\gamma(\theta))-\tilde{\tau}\theta\right)Z\Psi(\tilde{\tau},Z)'\right]\\ S^{22}\left(\theta,\tilde{\theta}\right)&=\left(\theta\vee\tilde{\theta}-\theta\tilde{\theta}\right)\mathbb{E}\left[ZZ'\right]\end{aligned}$$

Since $\hat{\Psi}(\cdot)\xrightarrow{p}\Psi(\cdot)$, and $\hat{\Phi}(\cdot)\xrightarrow{p}\Phi(\cdot)$ uniformly over compact sets and $\hat{\pi}(\tau,\theta)\xrightarrow{p}\pi(\tau,\theta)$ uniformly in (τ,θ). $\delta_n\equiv\sup_{(\tau,\theta)\in\mathcal{T}\times\mathcal{C}}\xi\left(h'(\tau,\theta),h(\tau,\theta)\right)|\xrightarrow{p}0$ by assumptions 6 and 7, for $h'(\tau,\theta)=\hat{h}(\tau,\theta)$, where

$$\xi(h,h')\equiv\sqrt{\mathbb{E}\left\|\begin{bmatrix}\rho_\tau\left(\varepsilon-\Phi(Z)'\iota\right)\Psi(Z)\\ \rho_\theta(\eta)Z\end{bmatrix}-\begin{bmatrix}\rho_{\tilde{\tau}}\left(\tilde{\varepsilon}-\tilde{\Phi}(Z)'\tilde{\iota}\right)\tilde{\Psi}(Z)\\ \rho_{\tilde{\theta}}(\tilde{\eta})\tilde{Z}\end{bmatrix}\right\|^2}$$

As $\delta_n \xrightarrow{p} 0$

$$\sup_{(\tau,\theta)} \left\| \mathbb{G}_n \begin{bmatrix} \rho_\tau \left(\hat{\varepsilon}(\tau) - \hat{\Phi}(\tau, Z)' \hat{\iota}(\tau) \right) \hat{\Psi}(\tau, Z) \\ \rho_\theta \left(\hat{\eta}(\theta) \right) Z \end{bmatrix} - \mathbb{G}_n \begin{bmatrix} \rho_\tau \left(\varepsilon(\tau) - \Phi(\tau, Z)' \iota(\tau) \right) \Psi(\tau, Z) \\ \rho_\theta \left(\eta(\theta) \right) Z \end{bmatrix} \right\|$$

$$\leq \sup_{\xi\left(\tilde{h}, h\right) \leq \delta_n, \tilde{h}, h \in \mathcal{H}} \left\| \mathbb{G}_n \begin{bmatrix} \rho_\tau \left(\varepsilon - \tilde{\Phi}(Z)' \tilde{\iota} \right) \tilde{\Psi}(Z) \\ \rho_\theta(\eta) Z \end{bmatrix} \mathbb{G}_n \begin{bmatrix} \rho_\tau \left(\varepsilon - \Phi(Z)' \iota \right) \Psi(Z) \\ \rho_\theta(\eta) Z \end{bmatrix} \right\| = o_P(1)$$

by stochastic equicontinuity of $h \mapsto \mathbb{G}_n \begin{bmatrix} \rho_\tau \left(\varepsilon - \Phi(Z)' \iota \right) \Psi(Z) \\ \rho_\theta(\eta) Z \end{bmatrix}$, which proves claim 2. To prove claim 1, define

$$\mathcal{A} \equiv \left\{ (\Phi, \beta, \iota, \gamma, \tau, \theta) \mapsto \begin{bmatrix} \rho_\tau \left(\varepsilon - \Phi(Z)' \iota \right) \\ \rho_\theta(\eta) \end{bmatrix} \right\}$$

This class of functions is uniformly Lipschitz over $(\mathcal{F} \times \mathcal{B} \times \mathcal{I} \times \mathcal{G} \times \mathcal{T} \times \mathcal{C})$ 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:

$$\sup_{h \in \mathcal{H}} \left| \mathbb{E}_n \begin{bmatrix} \rho_\tau \left(\varepsilon - \Phi(Z)' \iota \right) \\ \rho_\theta(\eta) \end{bmatrix} - \mathbb{E} \begin{bmatrix} \rho_\tau \left(\varepsilon - \Phi(Z)' \iota \right) \\ \rho_\theta(\eta) \end{bmatrix} \right| \xrightarrow{p} 0$$

which gives,

$$\sup_{(\beta,\iota,\gamma,\tau,\theta)} \left| \mathbb{E}_n \begin{bmatrix} \rho_\tau \left(\varepsilon - \tilde{\Phi}(\tau, Z)' \iota \right) \\ \rho_\theta(\eta) \end{bmatrix} - \mathbb{E} \begin{bmatrix} \rho_\tau \left(\varepsilon - \tilde{\Phi}(\tau, Z)' \iota \right) \\ \rho_\theta(\eta) \end{bmatrix} \right|_{\tilde{\Phi} = \hat{\Phi}} \xrightarrow{p} 0$$

By uniform consistency of $\hat{\Phi}(\cdot)$ and assumption 7, I have that

$$\sup_{(\beta,\iota,\gamma,\tau,\theta)} \left| \mathbb{E} \begin{bmatrix} \rho_\tau \left(\varepsilon - \tilde{\Phi}(\tau, Z)' \iota \right) \\ \rho_\theta(\eta) \end{bmatrix} - \mathbb{E} \begin{bmatrix} \rho_\tau \left(\varepsilon - \Phi(\tau, Z)' \iota \right) \\ \rho_\theta(\eta) \end{bmatrix} \right| \xrightarrow{p} 0$$

which implies claim 1. $\square$

B.6 Asymptotic Distribution of the IVQR and QR Estimators and the Estimator of the Copula Parameters

Lemma 6. *Under assumptions 1, 2, and 4 to 8,*

$$\sqrt{n}\begin{pmatrix}\hat{\beta}(u)-\beta(u)\\ \hat{\gamma}(v)-\gamma(v)\\ \hat{\xi}-\xi\end{pmatrix}\Rightarrow \mathbb{G}_{\mathcal{M}}(u,v)$$

where $\mathbb{G}_{\mathcal{M}}(u,v)$ is a Gaussian process with covariance matrix $\Sigma_{\mathcal{M}}(u,v,\tilde{u},\tilde{v})$ equal to

$$\Sigma_{\mathcal{M}}(u,v,\tilde{u},\tilde{v})\equiv\begin{bmatrix}\Sigma_{\mathcal{J}}(u,v,\tilde{u},\tilde{v}) & \Sigma_{\mathcal{M}}^{21}(u,\tilde{v})'\\ \Sigma_{\mathcal{M}}^{21}(\tilde{u},v) & \Sigma_{\xi}\end{bmatrix}$$

$$\begin{aligned}\Sigma_{\mathcal{M}}^{21}(u,v)=&H_1^{-1}\mathbb{E}\left[\frac{\partial\ell_j(u_j,v_j,\xi)}{\partial\xi}\begin{bmatrix}\left(\mathbf{1}\left(y_j\le x_j'\beta(u)\right)-u\right)\Psi(u,z_j)'J(u)^{-1}\\ \left(\mathbf{1}\left(x_{1j}\le z_j'\gamma(v)\right)-v\right)z_j'H(v)^{-1}\end{bmatrix}\right]\\&+H_1^{-1}\mathbb{E}\left[\frac{\partial^2\ell_j(u_j,v_j,\xi)}{\partial\xi\partial(u,v)}M(y_j,x_{1j},z_j)\Sigma_{\mathcal{J}}(u,v,u_j,v_j)\right]\end{aligned}$$

where the expectation is taken with respect to $F_Z(z_j)C_{UV|X_2}(u_j,v_j|x_{2j};\xi)$, and where I have used $M(y_j,x_{1j},z_j)\equiv-\begin{pmatrix}g_Y(y_j|x_j)x_j' & 0\\ 0 & f_{X_1}(x_{1j}|z_j)z_j'\end{pmatrix}$, $\Sigma_\xi\equiv H_1^{-1}(H_1+H_2)H_1^{-1}$, and

$$\begin{aligned}H_1&\equiv\mathbb{E}\left[-\frac{\partial^2\ell_j(u_j,v_j,\xi)}{\partial\xi\partial\xi'}\right]\\ H_2&\equiv\mathbb{E}\left[\frac{\partial^2\ell_j(u_j,v_j,\xi)}{\partial\xi\partial(u,v)}M(y_j,x_{1j},z_j)\Sigma_{\mathcal{J}}(u_j,v_j,u_h,v_h)M(y_j,x_{1j},z_j)'\frac{\partial^2\ell_h(u_h,v_h,\xi)}{\partial(u,v)'\partial\xi'}\right]\end{aligned}$$

Proof. Begin by writing $\hat{\xi}$ in terms of the influence function. To do so, apply the mean value theorem to the score:

$$0=\mathbb{E}_n\left[\frac{\partial \ell_j\left(\hat{u}_j,\hat{v}_j,\hat{\xi}\right)}{\partial \xi}\right]=\mathbb{E}_n\left[\frac{\partial \ell_j\left(\hat{u}_j,\hat{v}_j,\xi\right)}{\partial \xi}\right]+\mathbb{E}_n\left[\frac{\partial^2 \ell_j\left(\hat{u}_j,\hat{v}_j,\overline{\xi}\right)}{\partial \xi \partial \xi'}\right]\left(\hat{\xi}-\xi\right)$$

where $\overline{\xi}$ lies between $\hat{\xi}$ and ξ. Rearranging the previous equation yields

$$\sqrt{n}\left(\hat{\xi}-\xi\right)=\left[\mathbb{E}_n\left[\frac{\partial^2 \ell_j\left(\hat{u}_j,\hat{v}_j,\overline{\xi}\right)}{\partial \xi \partial \xi'}\right]\right]^{-1}\sqrt{n}\mathbb{E}_n\left[\frac{\partial \ell_j\left(\hat{u}_j,\hat{v}_j,\xi\right)}{\partial \xi}\right] \tag{20}$$

Now show the uniform convergence of the Hessian:

$$\begin{aligned}
\left|\frac{\partial^2 \ell_j\left(\hat{u}_j,\hat{v}_j,\overline{\xi}\right)}{\partial \xi \partial \xi'}-\frac{\partial^2 \ell_j\left(u_j,v_j,\xi\right)}{\partial \xi \partial \xi'}\right| &\leq \left|\nabla^3 \ell_j\left(\overline{u}_j,\overline{v}_j,\overline{\overline{\xi}}\right)\begin{pmatrix}\hat{u}_j-u_j\\ \hat{v}_j-v_j\\ \overline{\xi}-\xi\end{pmatrix}\right| \\
&= \left|\nabla^3 \ell_j\left(\overline{u}_j,\overline{v}_j,\overline{\overline{\xi}}\right)\right|\left|\begin{pmatrix}\hat{u}_j-u_j\\ \hat{v}_j-v_j\\ \overline{\xi}-\xi\end{pmatrix}\right| \\
&\leq K\cdot o_P^*(1)=o_P^*(1)
\end{aligned}$$

where $\nabla^3 \ell_j(u,v,\xi)$ is a three dimensional array whose (i,j,k) element is the partial derivative of $\log\left(c\left(u,v|x_j;\xi\right)\right)$ with respect to the ith element of ξ, its jth element of ξ and its kth element of $(u,v,\xi')'$. The first equality follows by the mean value theorem, and the last equality follows by assumptions 2 and 8. Using this result,

$$\begin{aligned}
&\left|\mathbb{E}_n\left[\frac{\partial^2 \ell_j\left(\hat{u}_j,\hat{v}_j,\xi\right)}{\partial \xi \partial \xi'}\right]-\mathbb{E}\left[\frac{\partial^2 \ell_j\left(u_j,v_j,\xi\right)}{\partial \xi \partial \xi'}\right]\right| \\
&\leq \mathbb{E}_n\left|\frac{\partial^2 \ell_j\left(\hat{u}_j,\hat{v}_j,\xi\right)}{\partial \xi \partial \xi'}-\frac{\partial^2 \ell_j\left(u_j,v_j,\xi\right)}{\partial \xi \partial \xi'}\right|+\left|\mathbb{E}_n\left[\frac{\partial^2 \ell_j\left(u_j,v_j,\xi\right)}{\partial \xi \partial \xi'}\right]-\mathbb{E}\left[\frac{\partial^2 \ell_j\left(u_j,v_j,\xi\right)}{\partial \xi \partial \xi'}\right]\right| \\
&= o_P^*(1)
\end{aligned}$$

where the inequality follows by the triangular inequality, the first term is $o_P^*(1)$ by the argument above, and the second term by uniform law of large numbers. Then, show the asymptotic distribution of $\sqrt{n}\mathbb{E}_n\left[\frac{\partial \ell_j(\hat{u}_j,\hat{v}_j,\xi)}{\partial \xi}\right]$. Apply the mean value theorem to $(\hat{u}_j, \hat{v}_j)$ for all $j = 1, ..., n$.

$$\sqrt{n}\mathbb{E}_n\left[\frac{\partial \ell_j\left(\hat{u}_j,\hat{v}_j,\xi\right)}{\partial \xi}\right] = \mathbb{G}_n\left[\frac{\partial \ell_j\left(u_j,v_j,\xi\right)}{\partial \xi}\right] + \sqrt{n}\mathbb{E}_n\left[\frac{\partial^2 \ell_j\left(\overline{u}_j,\overline{v}_j,\xi\right)}{\partial \xi \partial\left(u,v\right)}\begin{pmatrix}\hat{u}_j - u_j\\ \hat{v}_j - v_j\end{pmatrix}\right] \tag{21}$$

The first term is simply the usual term that appears in the maximization of the log likelihood function, and the second term takes into account that (u_j, v_j) are estimated, but not observed. Leaving aside the first term and focusing on the second, it follows that

$$\sqrt{n}\begin{pmatrix}\hat{u}_j - u_j\\ \hat{v}_j - v_j\end{pmatrix} = \sqrt{n}\begin{pmatrix}\int_0^1 \mathbf{1}\left(\hat{S}_Y\left(u|x_j\right)\leq y_j\right)du - \int_0^1 \mathbf{1}\left(S_Y\left(u|x_j\right)\leq y_j\right)du\\ \int_0^1 \mathbf{1}\left(\hat{h}\left(z_{1j},x_{2j},v_j\right)\leq x_{1j}\right)dv - \int_0^1 \mathbf{1}\left(h\left(z_{1j},x_{2j},v_j\right)\leq x_{1j}\right)dv\end{pmatrix}$$

Define $G_Y\left(y|x\right) \equiv \int_0^1 \mathbf{1}\left(S_Y\left(u|x\right)\leq y\right)du = S_Y^{-1}\left(y|x\right)$, and $g_Y\left(y|x\right) \equiv \frac{\partial}{\partial y}G_Y\left(y|x\right)$.[24] Apply Lemma 4 in Chernozhukov et al. (2013) to get

$$\sqrt{n}\begin{pmatrix}\hat{u}_j - u_j\\ \hat{v}_j - v_j\end{pmatrix} = \sqrt{n}M\left(y_j,x_{1j},z_j\right)\begin{pmatrix}\hat{\beta}\left(u_j\right)-\beta\left(u_j\right)\\ \hat{\gamma}\left(v_j\right)-\gamma\left(v_j\right)\end{pmatrix} + o_P^* \tag{22}$$

where the $*$ denotes that the convergence in probability is uniform in (u_j, v_j).

By the extended continuous mapping theorem, assumption 8, and the uniform consistency of $(\hat{u}_j, \hat{v}_j)$, it follows that

$$\frac{\partial^2 \ell_j\left(\overline{u}_j,\overline{v}_j,\xi\right)}{\partial \xi \partial\left(u,v\right)} = \frac{\partial^2 \ell_j\left(u_j,v_j,\xi\right)}{\partial \xi \partial\left(u,v\right)} + o_P^*\left(1\right)$$

By the information equality, the asymptotic variance of the first term equals H_1. After

[24]These would be the conditional cdf and pdf of Y if U and X were independent. These functions are different from the actual conditional cdf and pdf of Y, which are given by $F_Y\left(y|x\right) \equiv \int_0^1 \mathbf{1}\left(S_Y\left(u|x\right)\leq y\right)f\left(u|x\right)du$, and $f_Y\left(y|x\right) \equiv \frac{\partial}{\partial y}F_Y\left(y|x\right)$. Under endogeneity $f\left(u|x\right) \neq 1$, and hence $G_Y \neq F_Y$. Even though the actual data is not going to depend on G_Y, the way u_j is identified makes it convenient for inference.

some tedious algebra, it is possible to show that the asymptotic variance of the second term equals H_2, and the asymptotic covariance equals zero. Hence, using this result and equation 20, I can rewrite the estimators as

$$\sqrt{n}\begin{pmatrix}\hat{\beta}(u)-\beta(u)\\ \hat{\gamma}(v)-\gamma(v)\\ \hat{\xi}-\xi\end{pmatrix}=\begin{bmatrix}\sqrt{n}\begin{pmatrix}\hat{\beta}(u)-\beta(u)\\ \hat{\gamma}(v)-\gamma(v)\end{pmatrix}\\ H_1^{-1}\mathbb{E}_n\left[\frac{\partial^2\ell_j(u_j,v_j,\xi)}{\partial\xi\partial(u,v)}M(y_j,x_{1j},z_j)\sqrt{n}\begin{pmatrix}\hat{\beta}(u_j)-\beta(u_j)\\ \hat{\gamma}(v_j)-\gamma(v_j)\end{pmatrix}\right]+\\ +H_1^{-1}\mathbb{G}_n\left[\frac{\partial\ell_j(u_j,v_j,\xi)}{\partial\xi}\right]+o_P^*(1)\end{bmatrix}$$

By lemma 3, the extended continuous mapping theorem, and the functional delta method, it follows that

$$\sqrt{n}\begin{pmatrix}\hat{\beta}(u)-\beta(u)\\ \hat{\gamma}(v)-\gamma(v)\\ \hat{\xi}-\xi\end{pmatrix}\Rightarrow\mathbb{G}_{\mathcal{M}}(u,v)$$

□

B.7 Uniform consistency of $\check{C}_{UV}(u,v)$

Lemma 7. *Let assumptions 1, 2, and 9 hold, and $(\hat{u}_i,\hat{v}_i)$ be uniformly consistent estimators for (u_i,v_i). Then, $\sqrt{n}\sup_{u,v}\left|\check{C}_{UV}(u,v)-C_{UV}(u,v)\right|=O_P(1)$.*

Proof. Define $\tilde{C}_{UV}(u,v)\equiv\mathbb{E}_n[\mathbf{1}(u_i\le u)\mathbf{1}(v_i\le v)]$ and split the proof into showing the probability limit of $\tilde{C}_{UV}(u,v)$ and $\check{C}_{UV}(u,v)$ is the same, and then that $\tilde{C}_{UV}(u,v)$ is a consistent estimator of $C_{UV}(u,v)$.

$$\begin{aligned}\sqrt{n}\left|\check{C}_{UV}(u,v)-\tilde{C}_{UV}(u,v)\right| &= \frac{1}{\sqrt{n}}\left|\Sigma_{i=1}^n\mathbf{1}(\hat{u}_i\le u)\mathbf{1}(\hat{v}_i\le v)-\mathbf{1}(u_i\le u)\mathbf{1}(v_i\le v)\right|\\ &\le\mathbb{G}_n\left|\mathbf{1}(\hat{u}_i\le u)-\mathbf{1}(u_i\le u)\right|+\mathbb{G}_n\left|\mathbf{1}(\hat{v}_i\le v)-\mathbf{1}(v_i\le v)\right|\end{aligned}$$

Consider the sequence s_n that satisfies $s_n \to 0$ and $s_n\sqrt{n} \to \infty$ as $n \to \infty$.

$$\begin{aligned}
\sup_u \mathbb{P}\left(\sqrt{n}\left|\mathbf{1}\left(\hat{u}_i \le u\right) - \mathbf{1}\left(u_i \le u\right)\mathbf{1}\right| > \varepsilon\right) &= \sup_u \mathbb{P}\left(\mathbf{1}\left(\hat{u}_i \le u\right) \neq \mathbf{1}\left(u_i \le u\right)\right) \\
&\le \sup_u \mathbb{P}\left(\left|u_i - u\right| \le s_n\right) + \mathbb{P}\left(\left|u_i - \hat{u}_i\right| > s_n\right) \\
&\le 2s_n + \mathbb{P}\left(\left|u_i - \hat{u}_i\right| > s_n\right)
\end{aligned}$$

Take limits to conclude that $\lim_{n\to\infty} \sup_u \mathbb{P}\left(\sqrt{n}\left|\mathbf{1}\left(\hat{u}_i \le u\right) - \mathbf{1}\left(u_i \le u\right)\right| > \varepsilon\right) = 0$. By a parallel argument, $\lim_{n\to\infty} \sup_v \mathbb{P}\left(\sqrt{n}\left|\mathbf{1}\left(\hat{v}_i \le v\right) - \mathbf{1}\left(v_i \le v\right)\right| > \varepsilon\right) = 0$. Consequently, $\lim_{n\to\infty} \sup_{u,v} \mathbb{P}\left(\sqrt{n}\left|\check{C}_{UV}\left(u,v\right) - \tilde{C}_{UV}\left(u,v\right)\right| > \varepsilon\right) = 0$

As for the second step, consider the class $\mathcal{C}_{UV} \equiv \left\{\left\{\left(x_1,x_2\right) : x_1 \le u, x_2 \le v\right\}, u,v \in [0,1]\right\}$. This is a VC class with VC dimension $V\left(\mathcal{C}_{UV}\right) = 3$. Therefore, by Theorem 2.6.4 in van der Vaart and Wellner (1996), its covering number is bounded: $N\left(\varepsilon, \mathcal{C}_{UV}, L_2\left(P\right)\right) \le 3 \cdot 4^3 K e^3 \varepsilon^{-4} < \infty$ for some constant K and $0 < \varepsilon < 1$. By theorem 2.5.2 in van der Vaart and Wellner (1996), it is $\mathbb{P}$-Donsker, so $\sqrt{n}\sup_{u,v}\left|\tilde{C}_{UV}\left(u,v\right) - C_{UV}\left(u,v\right)\right| = O_P\left(1\right)$. Hence,

$$\begin{aligned}
&\sqrt{n}\left(\check{C}_{UV}\left(u,v\right) - C_{UV}\left(u,v\right)\right) \\
&= \sqrt{n}\left(\check{C}_{UV}\left(u,v\right) - \tilde{C}_{UV}\left(u,v\right)\right) + \sqrt{n}\left(\tilde{C}_{UV}\left(u,v\right) - C_{UV}\left(u,v\right)\right) = O_P^*\left(1\right)
\end{aligned}$$

□

B.8 Uniform consistency of $\check{C}_{U|V}\left(u|v\right)$

Consider the estimator $\check{C}_{UV}\left(u,v\right)$ defined by equation 15. This estimator can be seen as the integration over $[0,1]$ of a nonparametric estimator of the conditional copula distribution $C_{U|V}\left(u|v\right)$, given by

$$\check{C}_{U|V}\left(u|v\right) \equiv \frac{H_n+1}{n}\Sigma_{i=1}^{n}\mathbf{1}\left(\hat{u}_i \le u\right)\mathbf{1}\left(\underline{\theta}\left(v\right) \le \hat{v}_i < \overline{\theta}\left(v\right)\right)$$

where H_n denotes the number of evenly spaced quantiles that are used in the estimation of the quantile process $h\left(z_1, x_2, v\right)$,[25] and $\overline{\theta}\left(v\right)$ and $\underline{\theta}\left(v\right)$ are defined as $\left\{\max_i \theta_i : \theta_i < v\right\}$ and $\left\{\min_i \theta_i : \theta_i \geq v\right\}$. It can be checked that $\check{C}_{UV}\left(u, v\right) = \frac{1}{H_n+1}\Sigma_{h=0}^{\left(H_n+1\right)\underline{\theta}\left(v\right)}\check{C}_{U|V}\left(u, \theta_h\right)$. Geometrically, I am splitting the $[0,1]$ interval into $H_n + 1$ intervals of equal length, and each v_i belong to any of these intervals almost surely. The probability of v_i being in any of these intervals is equal to $\frac{1}{H_n+1}$, since $v_i \sim U\left(0,1\right)$. H_n is 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 8. *Let assumptions 1, 2, and 9 hold, $\left(\hat{u}_i, \hat{v}_i\right)$ converge uniformly in probability to $\left(u_i, v_i\right)$ at a rate $\sqrt{n}$, $H_n \to \infty$, and $\frac{na_n}{\log(n)} \to \infty$ as $n \to \infty$, where $a_n \leq \frac{1}{H_n}$. Then,*

$$sup_{u,v}\left|\check{C}_{U|V}\left(u|v\right) - C_{U|V}\left(u|v\right)\right| = o_P\left(1\right)$$

Proof. The proof is split into two steps: first show the consistency of the unfeasible estimator $\tilde{C}_{U|V}\left(u|v\right) \equiv \frac{H_n+1}{n}\Sigma_{i=1}^{n}\mathbf{1}\left(u_i \leq u\right)\mathbf{1}\left(\underline{\theta}\left(v\right) \leq v_i < \overline{\theta}\left(v\right)\right)$, and then show that $\check{C}_{U|V}\left(u|v\right)$ and $\tilde{C}_{U|V}\left(u|v\right)$ converge to the same limit.

Consider the class $\mathcal{C}_{UV} \equiv \left\{\left\{\left(x_1, x_2\right) : x_1 \leq u, v_l \leq x_2 < v_u\right\}, u, v_l, v_u \in \left[0,1\right], v_l < v_u\right\}$. It is a VC class with VC dimension $V\left(\mathcal{C}_{UV}\right) = 4$. Therefore, by Theorem 2.6.4 in van der Vaart and Wellner (1996), its covering number is bounded: $N\left(\varepsilon, \mathcal{C}_{UV}, L_2\left(P\right)\right) \leq 4^5 K e^4 \varepsilon^{-6} < \infty$ for some constant K and $0 < \varepsilon < 1$. By Corollary 1 in Einmahl et al. (2005)

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

This result implies that $\sup_{(u,v)\in[0,1]}\left|\tilde{C}_{U|V}\left(u|v\right) - C_{U|V}\left(u|v\right)\right| = o_P\left(1\right)$.

Regarding the second step, notice that it is not possible to apply the extended continuous mapping theorem to conclude that if $\hat{u}_i \xrightarrow{p} u_i$ and $\hat{v}_i \xrightarrow{p} v_i$, then $\mathbf{1}\left(\hat{u}_i \leq u\right) \xrightarrow{p} \mathbf{1}\left(u_i \leq u\right)$

[25] These quantiles are denoted by $0 = \theta_0, \theta_1, ..., \theta_{H_n}, \theta_{H_n+1} = 1$.

or $\mathbf{1}\left(\underline{\theta}(v) \leq \hat{v}_i < \overline{\theta}(v)\right) \xrightarrow{p} \mathbf{1}\left(\underline{\theta}(v) \leq v_i < \overline{\theta}(v)\right)$ uniformly in (u, v). Hence, a different argument is required for the proof:

$$
\begin{aligned}
&\sup_{u,v} \mathbb{P}\left(\left|\mathbf{1}(\hat{u}_i \leq u)\,\mathbf{1}\left(\underline{\theta}(v) \leq \hat{v}_i < \overline{\theta}(v)\right) - \mathbf{1}(u_i \leq u)\,\mathbf{1}\left(\underline{\theta}(v) \leq v_i < \overline{\theta}(v)\right)\right| \geq \varepsilon r_n\right) \\
&\leq \sup_{u} \mathbb{P}\left(\left|\mathbf{1}(\hat{u}_i \leq u) - \mathbf{1}(u_i \leq u)\right| \geq \varepsilon r_n\right) \\
&+ \sup_{v} \mathbb{P}\left(\left|\mathbf{1}\left(\underline{\theta}(v) \leq \hat{v}_i < \overline{\theta}(v)\right) - \mathbf{1}\left(\underline{\theta}(v) \leq v_i < \overline{\theta}(v)\right)\right| \geq \varepsilon r_n\right)
\end{aligned}
$$

Examine the convergence of each term separately:

$$
\begin{aligned}
\sup_{u} \mathbb{P}\left(\left|\mathbf{1}(\hat{u}_i \leq u) - \mathbf{1}(u_i \leq u)\right| > \varepsilon r_n\right) &= \sup_{u} \mathbb{P}\left(\mathbf{1}(\hat{u}_i \leq u) \neq \mathbf{1}(u_i \leq u)\right) \\
&\leq \sup_{u} \mathbb{P}\left(|u_i - u| \leq s_n\right) + \mathbb{P}\left(|u_i - \hat{u}_i| > s_n\right) \\
&\leq 2s_n + \mathbb{P}\left(|u_i - \hat{u}_i| > s_n\right)
\end{aligned}
$$

where s_n is a sequence that satisfies $s_n \to 0$ and $\sqrt{n}s_n \to \infty$ as $n \to \infty$. The second inequality follows from the fact that $u_i \sim U(0,1)$. Since $u_i - \hat{u}_i = O_P^*\left(\frac{1}{\sqrt{n}}\right)$, it follows that, if $\sqrt{n}s_n \to \infty$ as $n \to \infty$, $u_i - \hat{u}_i = o_P^*(s_n)$, and therefore

$$
\lim_{n\to\infty} \sup_{u} \mathbb{P}\left(\left|\mathbf{1}(\hat{u}_i \leq u) - \mathbf{1}(u_i \leq u)\right| > \varepsilon r_n\right) = 0
$$

Similarly,

$$
\begin{aligned}
&\sup_{v} \mathbb{P}\left(\left|\mathbf{1}\left(\underline{\theta}(v) \leq \hat{v}_i < \overline{\theta}(v)\right) - \mathbf{1}\left(\underline{\theta}(v) \leq v_i < \overline{\theta}(v)\right)\right| \geq \varepsilon r_n\right) \\
&= \sup_{v} \mathbb{P}\left(\mathbf{1}\left(\underline{\theta}(v) \leq \hat{v}_i < \overline{\theta}(v)\right) \neq \mathbf{1}\left(\underline{\theta}(v) \leq v_i < \overline{\theta}(v)\right)\right) \\
&\leq \sup_{v} \mathbb{P}\left(\left|v_i - \underline{\theta}(v)\right| \leq s_n\right) + \sup_{v} \mathbb{P}\left(\left|v_i - \overline{\theta}(v)\right| \leq s_n\right) + \mathbb{P}\left(|u_i - \hat{u}_i| > s_n\right) \\
&\leq 4s_n + \mathbb{P}\left(|v_i - \hat{v}_i| > s_n\right)
\end{aligned}
$$

And under the same conditions as before, I get that

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\mathbf{1}\left(\underline{\theta}\left(v\right) \leq \hat{v}_i < \overline{\theta}\left(v\right)\right) - \mathbf{1}\left(\underline{\theta}\left(v\right) \leq v_i < \overline{\theta}\left(v\right)\right)\right| \geq \varepsilon r_n\right) = 0$$

Hence, $\mathbf{1}\left(\hat{u}_i \leq u\right)\mathbf{1}\left(\underline{\theta}\left(v\right) \leq \hat{v}_i < \overline{\theta}\left(v\right)\right) - \mathbf{1}\left(u_i \leq u\right)\mathbf{1}\left(\underline{\theta}\left(v\right) \leq v_i < \overline{\theta}\left(v\right)\right) = o_P^*\left(r_n^{-1}\right)$. As a consequence, $\sup_{(u,v)}\left|\check{C}_{U|V}\left(u|v\right) - \tilde{C}_{U|V}\left(u|v\right)\right| = o_P\left(H_n r_n\right)$. Since r_n can converge to zero at any speed, it follows that the previous quantity equals $o_P\left(1\right)$ if $H_n \to \infty$ at any speed. By the triangle inequality,

$$\left|\check{C}_{U|V}\left(u|v\right) - C_{U|V}\left(u|v\right)\right| \leq \left|\check{C}_{U|V}\left(u|v\right) - \tilde{C}_{U|V}\left(u|v\right)\right| + \left|\tilde{C}_{U|V}\left(u|v\right) - C_{U|V}\left(u|v\right)\right|$$

Therefore, it follows that $\sup_{u,v}\left|\check{C}_{U|V}\left(u|v\right) - C_{U|V}\left(u|v\right)\right| = o_P\left(1\right)$. □

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

If instead of using indicator functions, one used functions that are (uniformly) smooth in u and v, then one could use the extended continuous mapping theorem. Consequently, it would be possible to estimate $C_{U|V}\left(u|v\right)$ by $\grave{C}_{U|V}\left(u|v\right) = \frac{1}{nh_n}\Sigma_{i=1}^{n}\grave{f}\left(u_i, u, n\right)\grave{K}\left(\frac{v_i - v}{h_n}\right)$, where $\grave{f}\left(u_i, u, n\right)$ is a function that is uniformly smooth in u and that converges to $\mathbf{1}\left(u_i \leq u\right)$ as $n \to \infty$, 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 Estimator of the Asymptotic Variance of $\hat{F}_Y(y)$

Begin by estimating the asymptotic variance of the estimator given by equation 8:

$$\hat{\Sigma}_{\mathcal{O}}(y,\tilde{y}) = \frac{1}{n^2 H_n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{H_n} \sum_{h=1}^{H_n} \hat{O}(y, z_i, v_k)\, \Sigma_{\mathcal{N}}(u(y, z_i, v_k), z_i, v_k, u(\tilde{y}, z_j, v_h)\, z_j, v_h)\, \hat{O}(\tilde{y}, z_j, v_h)'$$

where $\hat{O}(y,z,v) = \begin{pmatrix} -\hat{f}_{Y|ZV}(y|z,v) & 1 \end{pmatrix}$ and

$$\hat{f}_{Y|ZV}(y|z,v) = \Sigma_{k=1}^{K_n} \frac{(\tau_{k+1}-\tau_k)\, c_{U|VX_2}\left(\tau_k|v, x_2; \hat{\xi}\right)}{\hat{x}(v)'\left(\hat{\beta}(\tau_{k+1}) - \hat{\beta}(\tau_k)\right)} \mathbf{1}\left(\hat{x}(v)'\hat{\beta}(\tau_k) \leq y \leq \hat{x}(v)'\hat{\beta}(\tau_{k+1})\right)$$

If the SQF and the copula are estimated by equations 12 and 14, then the central term of the variance equals

$$\hat{\Sigma}_{\mathcal{N}}(u,v,\tilde{u},\tilde{v}) = \hat{N}(u,v,\tilde{u},\tilde{v})\, \hat{\Sigma}_{\mathcal{M}}(u,v,\tilde{u},\tilde{v})\, \hat{N}(u,v,\tilde{u},\tilde{v})'$$

where $\hat{N}(u,v,z) = \begin{pmatrix} \hat{x}(v)' & \hat{\beta}_1(u)\, z' & 0 \\ 0 & 0 & \frac{\partial}{\partial \xi} C_{U|VX_2}\left(u|v, x_2; \hat{\xi}\right) \end{pmatrix}$ and

$$\hat{\Sigma}_{\mathcal{M}}(u,v,\tilde{u},\tilde{v}) = \begin{pmatrix} \hat{\Sigma}_{\mathcal{J}}(u,v,\tilde{u},\tilde{v}) & \hat{\Sigma}_{\mathcal{M}}^{21}(u,\tilde{v})' \\ \hat{\Sigma}_{\mathcal{M}}^{21}(\tilde{u},v) & \hat{\Sigma}_{\xi} \end{pmatrix}$$

$$\hat{\Sigma}_{\xi} = \hat{H}_1^{-1}\left(\hat{H}_1 + \hat{H}_2\right)\hat{H}_1^{-1}$$

$$\hat{H}_1 = -\frac{1}{n}\Sigma_{i=1}^{n} \frac{\partial^2 \ell_i\left(\hat{u}_i, \hat{v}_i, \hat{\xi}\right)}{\partial \xi \partial \xi'}$$

$$\hat{H}_2 = \frac{1}{n^2}\Sigma_{i=1}^{n}\Sigma_{j=1}^{n} \frac{\partial^2 \ell_i\left(\hat{u}_i, \hat{v}_i, \hat{\xi}\right)}{\partial \xi \partial (u,v)} \hat{M}(y_i, x_{1i}, z_i)\, \Sigma_{\mathcal{J}}(\hat{u}_i, \hat{v}_i, \hat{u}_j, \hat{v}_j)\, \hat{M}(y_j, x_{1j}, z_j)' \frac{\partial^2 \ell_j\left(\hat{u}_j, \hat{v}_j, \hat{\xi}\right)}{\partial (u,v)' \partial \xi'}$$

$$\hat{M}(y,x_1,z)=\begin{pmatrix}\hat{g}_Y(y|x)\,x' & 0\\ 0 & \hat{f}_{X_1}(x_1|z)\,z'\end{pmatrix}$$

$$\begin{aligned}\hat{g}_Y(y|x)&=\Sigma_{k=1}^{K_n}\frac{\tau_{k+1}-\tau_k}{x'\left(\hat{\beta}(\tau_{k+1})-\hat{\beta}(\tau_k)\right)}\mathbf{1}\left(x'\hat{\beta}(\tau_k)\le y\le x'\hat{\beta}(\tau_{k+1})\right)\\ \hat{f}_{X_1}(x_1|z)&=\Sigma_{h=1}^{H_n}\frac{\theta_{h+1}-\theta_h}{z'(\hat{\gamma}(\theta_{h+1})-\hat{\gamma}(\theta_h))}\mathbf{1}(z'\hat{\gamma}(\theta_h)\le x_1\le z'\hat{\gamma}(\theta_{h+1}))\end{aligned}$$

$$\begin{aligned}\hat{\Sigma}_{\mathcal{M}}^{21}(u,v)=\hat{H}_1^{-1}\frac{1}{n}\Sigma_{i=1}^{n}\frac{\partial\ell_j\left(\hat{u}_i,\hat{v}_i,\hat{\xi}\right)}{\partial\xi}&\begin{bmatrix}\left(\mathbf{1}\left(y_i\le x_i'\hat{\beta}(u)\right)-u\right)\hat{\Psi}(u,z_i)'\hat{J}(u)\\ (\mathbf{1}(x_{1i}\le z_i'\hat{\gamma}(v))-v)\,z_i'\hat{H}(v)\end{bmatrix}\\ +\hat{H}_1^{-1}\frac{1}{n}\Sigma_{i=1}^{n}\frac{\partial\ell_j\left(\hat{u}_i,\hat{v}_i,\hat{\xi}\right)}{\partial\xi\partial(u,v)}&\hat{M}(y_j,x_{1j},z_j)\,\hat{\Sigma}_{\mathcal{J}}(u,v,\hat{u}_i,\hat{v}_i)\end{aligned}$$

$$\begin{aligned}\hat{\Sigma}_{\mathcal{J}}(u,v,\tilde{u},\tilde{v})&=\begin{bmatrix}\hat{\Sigma}_{\mathcal{J}}^{11}(u,\tilde{u}) & \hat{\Sigma}_{\mathcal{J}}^{21}(u,\tilde{v})'\\ \hat{\Sigma}_{\mathcal{J}}^{21}(\tilde{u},v) & \hat{\Sigma}_{\mathcal{J}}^{22}(v,\tilde{v})\end{bmatrix}\\ \hat{\Sigma}_{\mathcal{J}}^{11}(u,\tilde{u})&=\hat{J}(u)^{-1}(\min\{u,\tilde{u}\}-u\tilde{u})\frac{1}{n}\Sigma_{i=1}^{n}\hat{\Psi}(u,z_i)'\hat{\Psi}(\tilde{u},z_i)\hat{J}(\tilde{u})^{-1}\\ \hat{\Sigma}_{\mathcal{J}}^{21}(u,v)&=\hat{H}(v)^{-1}\frac{1}{n}\Sigma_{i=1}^{n}\left(\mathbf{1}\left(y_i\le x_i'\hat{\beta}(u)\right)\mathbf{1}(x_{1i}\le z_i'\hat{\gamma}(v))-uv\right)z_i\hat{\Psi}(u,z_i)'\hat{J}(u)^{-1}\\ \hat{\Sigma}_{\mathcal{J}}^{22}(v,\tilde{v})&=\hat{H}(u)^{-1}(\min\{v,\tilde{v}\}-v\tilde{v})\frac{1}{n}\Sigma_{i=1}^{n}z_iz_i'\hat{H}(\tilde{v})^{-1}\end{aligned}$$

and the matrices $\hat{J}(u)$ and $\hat{H}(v)$ are estimated using Powell (1986) estimator:

$$\begin{aligned}\hat{J}(u)&=\frac{1}{2nh_n}\Sigma_{i=1}^{n}\mathbf{1}(|\hat{\varepsilon}_i(u)|\le h_n)\left(\hat{\Phi}(u,z_i),x_{2i}'\right)'\left(\hat{\Phi}(u,z_i),x_{2i}'\right)\\ \hat{H}(v)&=\frac{1}{2nh_n}\Sigma_{i=1}^{n}\mathbf{1}(|\hat{\eta}_i(v)|\le h_n)\,z_iz_i'\end{aligned}$$

for some appropriately chosen bandwidth h_n.

D Fit of the Parametric Copulas to the Data

Table 4 compares the performance of the estimators based on different parametric copulas. The first row represents the integral of the difference between the estimated cdf and the empirical cdf, and the second row represents the largest difference between the two cdfs. Among the parametric copulas, the Clayton copula has the best fit in both cases. However, the fit of the estimator based on the nonparametric estimator of the copula is remarkably better than the fit of any of the estimators based on the parametric copulas.

Table 4: Fit of the Copula Distributions

	Gaussian	Clayton	Frank	Gumbel	Nonparametric
$\int_{\mathcal{Y}} \left\| \hat{F}_Y(y) - F_Y(y) \right\| dy$	0.012	0.009	0.015	0.014	0.003
$\sup_y \left\| \hat{F}_Y(y) - F_Y(y) \right\|$	0.056	0.043	0.070	0.055	0.013

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.

RECENTLY PUBLISHED "TEMI" (*)

N. 1028 – *The impact of CCPs' margin policies on repo markets*, by Arianna Miglietta, Cristina Picillo and Mario Pietrunti (September 2015).

N. 1029 – *European structural funds during the crisis: evidence from Southern Italy*, by Emanuele Ciani and Guido de Blasio (September 2015).

N. 1030 – *Female employment and pre-kindergarten: on the uninteded effects of an Italian reform*, by Francesca Carta and Lucia Rizzica (September 2015).

N. 1031 – *The predictive content of business survey indicators: evidence from SIGE*, by Tatiana Cesaroni and Stefano Iezzi (September 2015).

N. 1032 – *Sovereign debt exposure and the bank lending channel: impact on credit supply and the real economy*, by Margherita Bottero, Simone Lenzu and Filippo Mezzanotti (September 2015).

N. 1033 – *Does trend inflation make a difference?*, by Michele Loberto and Chiara Perricone (September 2015).

N. 1034 – *Procyclicality of credit rating systems: how to manage it*, by Tatiana Cesaroni (September 2015).

N. 1035 – *The time varying effect of oil price shocks on euro-area exports*, by Marianna Riggi and Fabrizio Venditti (September 2015).

N. 1036 – *Domestic and international macroeconomic effects of the Eurosystem expanded asset purchase programme*, by Pietro Cova, Patrizio Pagano and Massimiliano Pisani (September 2015).

N. 1037 – *Deconstructing the gains from trade: selection of industries vs. reallocation of workers*, by Stefano Bolatto and Massimo Sbracia (November 2015).

N. 1038 – *Young adults living with their parents and the influence of peers*, by Effrosyni Adamopoulou and Ezgi Kaya (November 2015).

N. 1039 – *Shoe-leather costs in the euro area and the foreign demand for euro banknotes*, by Alessandro Calza and Andrea Zaghini (November 2015).

N. 1040 – *The macroeconomic effects of low and falling inflation at the zero lower bound*, by Stefano Neri and Alessandro Notarpietro (November 2015).

N. 1041 – *The use of fixed-term contracts and the (adverse) selection of public sector workers*, by Lucia Rizzica (November 2015).

N. 1042 – *Multitask agents and incentives: the case of teaching and research for university professors*, by Marta De Philippis (November 2015).

N. 1043 – *Exposure to media and corruption perceptions*, by Lucia Rizzica and Marco Tonello (November 2015).

N. 1044 – *The supply side of household finance*, by Gabriele Foà, Leonardo Gambacorta, Luigi Guiso and Paolo Emilio Mistrulli (November 2015).

N. 1045 – *Optimal inflation weights in the euro area*, by by Daniela Bragoli, Massimiliano Rigon and Francesco Zanetti (January 2016).

N. 1046 – *Carry trades and exchange rate volatility: a TVAR approach*, by Alessio Anzuini and Francesca Brusa (January 2016).

N. 1047 – *A new method for the correction of test scores manipulation* by Santiago Pereda Fernández (January 2016).

N. 1048 – *Heterogeneous peer effects in education* by by Eleonora Patacchini, Edoardo Rainone and Yves Zenou (January 2016).

N. 1049 – *Debt maturity and the liquidity of secondary debt markets*, by Max Bruche and Anatoli Segura (January 2016).

N. 1050 – *Contagion and fire sales in banking networks*, by Sara Cecchetti, Marco Rocco and Laura Sigalotti (January 2016).

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

2013

A. MERCATANTI, *A likelihood-based analysis for relaxing the exclusion restriction in randomized experiments with imperfect compliance*, Australian and New Zealand Journal of Statistics, v. 55, 2, pp. 129-153, **TD No. 683 (August 2008).**

F. CINGANO and P. PINOTTI, *Politicians at work. The private returns and social costs of political connections*, Journal of the European Economic Association, v. 11, 2, pp. 433-465, **TD No. 709 (May 2009).**

F. BUSETTI and J. MARCUCCI, *Comparing forecast accuracy: a Monte Carlo investigation*, International Journal of Forecasting, v. 29, 1, pp. 13-27, **TD No. 723 (September 2009).**

D. DOTTORI, S. I-LING and F. ESTEVAN, *Reshaping the schooling system: The role of immigration*, Journal of Economic Theory, v. 148, 5, pp. 2124-2149, **TD No. 726 (October 2009).**

A. FINICELLI, P. PAGANO and M. SBRACIA, *Ricardian Selection*, Journal of International Economics, v. 89, 1, pp. 96-109, **TD No. 728 (October 2009).**

L. MONTEFORTE and G. MORETTI, *Real-time forecasts of inflation: the role of financial variables*, Journal of Forecasting, v. 32, 1, pp. 51-61, **TD No. 767 (July 2010).**

R. GIORDANO and P. TOMMASINO, *Public-sector efficiency and political culture*, FinanzArchiv, v. 69, 3, pp. 289-316, **TD No. 786 (January 2011).**

E. GAIOTTI, *Credit availablility and investment: lessons from the "Great Recession"*, European Economic Review, v. 59, pp. 212-227, **TD No. 793 (February 2011).**

F. NUCCI and M. RIGGI, *Performance pay and changes in U.S. labor market dynamics*, Journal of Economic Dynamics and Control, v. 37, 12, pp. 2796-2813, **TD No. 800 (March 2011).**

G. CAPPELLETTI, G. GUAZZAROTTI and P. TOMMASINO, *What determines annuity demand at retirement?,* The Geneva Papers on Risk and Insurance – Issues and Practice, pp. 1-26, **TD No. 805 (April 2011).**

A. ACCETTURO e L. INFANTE, *Skills or Culture? An analysis of the decision to work by immigrant women in Italy,* IZA Journal of Migration, v. 2, 2, pp. 1-21, **TD No. 815 (July 2011).**

A. DE SOCIO, *Squeezing liquidity in a "lemons market" or asking liquidity "on tap"*, Journal of Banking and Finance, v. 27, 5, pp. 1340-1358, **TD No. 819 (September 2011).**

S. GOMES, P. JACQUINOT, M. MOHR and M. PISANI, *Structural reforms and macroeconomic performance in the euro area countries: a model-based assessment,* International Finance, v. 16, 1, pp. 23-44, **TD No. 830 (October 2011).**

G. BARONE and G. DE BLASIO, *Electoral rules and voter turnout,* International Review of Law and Economics, v. 36, 1, pp. 25-35, **TD No. 833 (November 2011).**

O. BLANCHARD and M. RIGGI, *Why are the 2000s so different from the 1970s? A structural interpretation of changes in the macroeconomic effects of oil prices*, Journal of the European Economic Association, v. 11, 5, pp. 1032-1052, **TD No. 835 (November 2011).**

R. CRISTADORO and D. MARCONI, *Household savings in China,* in G. Gomel, D. Marconi, I. Musu, B. Quintieri (eds), The Chinese Economy: Recent Trends and Policy Issues, Springer-Verlag, Berlin, **TD No. 838 (November 2011).**

A. ANZUINI, M. J. LOMBARDI and P. PAGANO, *The impact of monetary policy shocks on commodity prices*, International Journal of Central Banking, v. 9, 3, pp. 119-144, **TD No. 851 (February 2012).**

R. GAMBACORTA and M. IANNARIO, *Measuring job satisfaction with CUB models,* Labour, v. 27, 2, pp. 198-224, **TD No. 852 (February 2012).**

G. ASCARI and T. ROPELE, *Disinflation effects in a medium-scale new keynesian model: money supply rule versus interest rate rule,* European Economic Review, v. 61, pp. 77-100, **TD No. 867 (April 2012)**

E. BERETTA and S. DEL PRETE, *Banking consolidation and bank-firm credit relationships: the role of geographical features and relationship characteristics,* Review of Economics and Institutions, v. 4, 3, pp. 1-46, **TD No. 901 (February 2013).**

M. ANDINI, G. DE BLASIO, G. DURANTON and W. STRANGE, *Marshallian labor market pooling: evidence from Italy,* Regional Science and Urban Economics, v. 43, 6, pp.1008-1022, **TD No. 922 (July 2013).**

G. SBRANA and A. SILVESTRINI, *Forecasting aggregate demand: analytical comparison of top-down and bottom-up approaches in a multivariate exponential smoothing framework,* International Journal of Production Economics, v. 146, 1, pp. 185-98, **TD No. 929 (September 2013).**

A. FILIPPIN, C. V, FIORIO and E. VIVIANO, *The effect of tax enforcement on tax morale,* European Journal of Political Economy, v. 32, pp. 320-331, **TD No. 937 (October 2013).**

2014

G. M. TOMAT, *Revisiting poverty and welfare dominance*, Economia pubblica, v. 44, 2, 125-149, **TD No. 651 (December 2007).**

M. TABOGA, *The riskiness of corporate bonds*, Journal of Money, Credit and Banking, v.46, 4, pp. 693-713, **TD No. 730 (October 2009).**

G. MICUCCI and P. ROSSI, *Il ruolo delle tecnologie di prestito nella ristrutturazione dei debiti delle imprese in crisi*, in A. Zazzaro (a cura di), Le banche e il credito alle imprese durante la crisi, Bologna, Il Mulino, **TD No. 763 (June 2010).**

F. D'AMURI, *Gli effetti della legge 133/2008 sulle assenze per malattia nel settore pubblico,* Rivista di politica economica, v. 105, 1, pp. 301-321, **TD No. 787 (January 2011).**

R. BRONZINI and E. IACHINI, *Are incentives for R&D effective? Evidence from a regression discontinuity approach,* American Economic Journal : Economic Policy, v. 6, 4, pp. 100-134, **TD No. 791 (February 2011).**

P. ANGELINI, S. NERI and F. PANETTA, *The interaction between capital requirements and monetary policy*, Journal of Money, Credit and Banking, v. 46, 6, pp. 1073-1112, **TD No. 801 (March 2011).**

M. BRAGA, M. PACCAGNELLA and M. PELLIZZARI, *Evaluating students' evaluations of professors,* Economics of Education Review, v. 41, pp. 71-88, **TD No. 825 (October 2011).**

M. FRANCESE and R. MARZIA, *Is there Room for containing healthcare costs? An analysis of regional spending differentials in Italy,* The European Journal of Health Economics, v. 15, 2, pp. 117-132, **TD No. 828 (October 2011).**

L. GAMBACORTA and P. E. MISTRULLI, *Bank heterogeneity and interest rate setting: what lessons have we learned since Lehman Brothers?,* Journal of Money, Credit and Banking, v. 46, 4, pp. 753-778, **TD No. 829 (October 2011).**

M. PERICOLI, *Real term structure and inflation compensation in the euro area*, International Journal of Central Banking, v. 10, 1, pp. 1-42, **TD No. 841 (January 2012).**

E. GENNARI and G. MESSINA, *How sticky are local expenditures in Italy? Assessing the relevance of the flypaper effect through municipal data,* International Tax and Public Finance, v. 21, 2, pp. 324-344, **TD No. 844 (January 2012).**

V. DI GACINTO, M. GOMELLINI, G. MICUCCI and M. PAGNINI, *Mapping local productivity advantages in Italy: industrial districts, cities or both?*, Journal of Economic Geography, v. 14, pp. 365–394, **TD No. 850 (January 2012).**

A. ACCETTURO, F. MANARESI, S. MOCETTI and E. OLIVIERI, *Don't Stand so close to me: the urban impact of immigration,* Regional Science and Urban Economics, v. 45, pp. 45-56, **TD No. 866 (April 2012).**

M. PORQUEDDU and F. VENDITTI, *Do food commodity prices have asymmetric effects on euro area inflation,* Studies in Nonlinear Dynamics and Econometrics, v. 18, 4, pp. 419-443, **TD No. 878 (September 2012).**

S. FEDERICO, *Industry dynamics and competition from low-wage countries: evidence on Italy*, Oxford Bulletin of Economics and Statistics, v. 76, 3, pp. 389-410, **TD No. 879 (September 2012).**

F. D'AMURI and G. PERI, *Immigration, jobs and employment protection: evidence from Europe before and during the Great Recession,* Journal of the European Economic Association, v. 12, 2, pp. 432-464, **TD No. 886 (October 2012).**

M. TABOGA, *What is a prime bank? A euribor-OIS spread perspective,* International Finance, v. 17, 1, pp. 51-75, **TD No. 895 (January 2013).**

G. CANNONE and D. FANTINO, *Evaluating the efficacy of european regional funds for R&D,* Rassegna italiana di valutazione, v. 58, pp. 165-196, **TD No. 902 (February 2013).**

L. GAMBACORTA and F. M. SIGNORETTI, *Should monetary policy lean against the wind? An analysis based on a DSGE model with banking,* Journal of Economic Dynamics and Control, v. 43, pp. 146-74, **TD No. 921 (July 2013).**

M. BARIGOZZI, CONTI A.M. and M. LUCIANI, *Do euro area countries respond asymmetrically to the common monetary policy?,* Oxford Bulletin of Economics and Statistics, v. 76, 5, pp. 693-714, **TD No. 923 (July 2013).**

U. ALBERTAZZI and M. BOTTERO, *Foreign bank lending: evidence from the global financial crisis,* Journal of International Economics, v. 92, 1, pp. 22-35, **TD No. 926 (July 2013).**

R. DE BONIS and A. SILVESTRINI, *The Italian financial cycle: 1861-2011,* Cliometrica, v.8, 3, pp. 301-334, **TD No. 936 (October 2013).**

G. BARONE and S. MOCETTI, *Natural disasters, growth and institutions: a tale of two earthquakes,* Journal of Urban Economics, v. 84, pp. 52-66, **TD No. 949 (January 2014).**

D. PIANESELLI and A. ZAGHINI, *The cost of firms' debt financing and the global financial crisis,* Finance Research Letters, v. 11, 2, pp. 74-83, **TD No. 950 (February 2014).**

J. LI and G. ZINNA, *On bank credit risk: sytemic or bank-specific? Evidence from the US and UK,* Journal of Financial and Quantitative Analysis, v. 49, 5/6, pp. 1403-1442, **TD No. 951 (February 2015).**

A. ZAGHINI, *Bank bonds: size, systemic relevance and the sovereign,* International Finance, v. 17, 2, pp. 161-183, **TD No. 966 (July 2014).**

G. SBRANA and A. SILVESTRINI, *Random switching exponential smoothing and inventory forecasting,* International Journal of Production Economics, v. 156, 1, pp. 283-294, **TD No. 971 (October 2014).**

M. SILVIA, *Does issuing equity help R&D activity? Evidence from unlisted Italian high-tech manufacturing firms,* Economics of Innovation and New Technology, v. 23, 8, pp. 825-854, **TD No. 978 (October 2014).**

2015

M. BUGAMELLI, S. FABIANI and E. SETTE, *The age of the dragon: the effect of imports from China on firm-level prices*, Journal of Money, Credit and Banking, v. 47, 6, pp. 1091-1118, **TD No. 737 (January 2010).**

R. BRONZINI, *The effects of extensive and intensive margins of FDI on domestic employment: microeconomic evidence from Italy*, B.E. Journal of Economic Analysis & Policy, v. 15, 4, pp. 2079-2109, **TD No. 769 (July 2010).**

G. BULLIGAN, M. MARCELLINO and F. VENDITTI, *Forecasting economic activity with targeted predictors,* International Journal of Forecasting, v. 31, 1, pp. 188-206, **TD No. 847 (February 2012).**

A. CIARLONE, *House price cycles in emerging economies,* Studies in Economics and Finance, v. 32, 1, **TD No. 863 (May 2012).**

D. FANTINO, A. MORI and D. SCALISE, *Collaboration between firms and universities in Italy: the role of a firm's proximity to top-rated departments,* Rivista Italiana degli economisti, v. 1, 2, pp. 219-251, **TD No. 884 (October 2012).**

D. DEPALO, R. GIORDANO and E. PAPAPETROU, *Public-private wage differentials in euro area countries: evidence from quantile decomposition analysis,* Empirical Economics, v. 49, 3, pp. 985-1115, **TD No. 907 (April 2013).**

G. BARONE and G. NARCISO, *Organized crime and business subsidies: Where does the money go?,* Journal of Urban Economics, v. 86, pp. 98-110, **TD No. 916 (June 2013).**

P. ALESSANDRI and B. NELSON, *Simple banking: profitability and the yield curve,* Journal of Money, Credit and Banking, v. 47, 1, pp. 143-175, **TD No. 945 (January 2014).**

M. TANELI and B. OHL, *Information acquisition and learning from prices over the business cycle,* Journal of Economic Theory, 158 B, pp. 585–633, **TD No. 946 (January 2014).**

R. AABERGE and A. BRANDOLINI, *Multidimensional poverty and inequality,* in A. B. Atkinson and F. Bourguignon (eds.), Handbook of Income Distribution, Volume 2A, Amsterdam, Elsevier, **TD No. 976 (October 2014).**

V. CUCINIELLO and F. M. SIGNORETTI, *Large banks,loan rate markup and monetary policy,* International Journal of Central Banking, v. 11, 3, pp. 141-177, **TD No. 987 (November 2014).**

M. FRATZSCHER, D. RIMEC, L. SARNOB and G. ZINNA, *The scapegoat theory of exchange rates: the first tests,* Journal of Monetary Economics, v. 70, 1, pp. 1-21, **TD No. 991 (November 2014).**

A. NOTARPIETRO and S. SIVIERO, *Optimal monetary policy rules and house prices: the role of financial frictions,* Journal of Money, Credit and Banking, v. 47, S1, pp. 383-410, **TD No. 993 (November 2014).**

R. ANTONIETTI, R. BRONZINI and G. CAINELLI, *Inward greenfield FDI and innovation,* Economia e Politica Industriale, v. 42, 1, pp. 93-116, **TD No. 1006 (March 2015).**

T. CESARONI, *Procyclicality of credit rating systems: how to manage it,* Journal of Economics and Business, v. 82. pp. 62-83, **TD No. 1034 (October 2015).**

M. RIGGI and F. VENDITTI, *The time varying effect of oil price shocks on euro-area exports,* Journal of Economic Dynamics and Control, v. 59, pp. 75-94, **TD No. 1035 (October 2015).**

FORTHCOMING

G. DE BLASIO, D. FANTINO and G. PELLEGRINI, *Evaluating the impact of innovation incentives: evidence from an unexpected shortage of funds,* Industrial and Corporate Change, **TD No. 792 (February 2011).**

A. DI CESARE, A. P. STORK and C. DE VRIES, *Risk measures for autocorrelated hedge fund returns,* Journal of Financial Econometrics, **TD No. 831 (October 2011).**

E. BONACCORSI DI PATTI and E. SETTE, *Did the securitization market freeze affect bank lending during the financial crisis? Evidence from a credit register,* Journal of Financial Intermediation, **TD No. 848 (February 2012).**

M. MARCELLINO, M. PORQUEDDU and F. VENDITTI, *Short-Term GDP Forecasting with a mixed frequency dynamic factor model with stochastic volatility,* Journal of Business & Economic Statistics, **TD No. 896 (January 2013).**

M. ANDINI and G. DE BLASIO, *Local development that money cannot buy: Italy's Contratti di Programma,* Journal of Economic Geography, **TD No. 915 (June 2013).**

F BRIPI, *The role of regulation on entry: evidence from the Italian provinces,* World Bank Economic Review, **TD No. 932 (September 2013).**

G. ALBANESE, G. DE BLASIO and P. SESTITO, *My parents taught me. evidence on the family transmission of values,* Journal of Population Economics, **TD No. 955 (March 2014).**

A. L. MANCINI, C. MONFARDINI and S. PASQUA, *Is a good example the best sermon? Children's imitation of parental reading,* Review of Economics of the Household, **TD No. 958 (April 2014).**

R. BRONZINI and P. PISELLI, *The impact of R&D subsidies on firm innovation,* Research Policy, **TD No. 960 (April 2014).**

L. BURLON, *Public expenditure distribution, voting, and growth,* Journal of Public Economic Theory, **TD No. 961 (April 2014).**

L. BURLON and M. VILALTA-BUFI, *A new look at technical progress and early retirement,* IZA Journal of Labor Policy, **TD No. 963 (June 2014).**

A. BRANDOLINI and E. VIVIANO, *Behind and beyond the (headcount) employment rate,* Journal of the Royal Statistical Society: Series A, **TD No. 965 (July 2015).**

G. ZINNA, *Price pressures on UK real rates: an empirical investigation,* Review of Finance, **TD No. 968 (July 2014).**

A. CALZA and A. ZAGHINI, *Shoe-leather costs in the euro area and the foreign demand for euro banknotes,* International Journal of Central Banking, **TD No. 1039 (December 2015).**