SUPPLEMENTAL MATERIAL TO "QUANTILE SELECTION MODELS WITH AN APPLICATION TO UNDERSTANDING CHANGES IN WAGE INEQUALITY"

(*Econometrica*, Vol. 85, No. 1, January 2017, 1–28)

BY MANUEL ARELLANO AND STÉPHANE BONHOMME

S1. BOUNDS ANALYSIS

WE SHOW THAT the quantile bounds (10) and (11) cannot be improved upon. In the analysis, we omit the x subscript for conciseness. Throughout, we work under the assumption that the model is correctly specified. Hence there exists a copula C_0 (with conditional copula G_0) and a c.d.f. F_0, which are the true copula and c.d.f. of (U, V) and Y^*, respectively. Let $\mathcal{P}$ denote the support of $p(Z)$, and let $\overline{p} = \sup_{\mathcal{P}} p$.

Let $\widetilde{G}$ be a conditional copula strictly increasing in its first argument, and let us define the following subcopula:

$$C(\tau, p) \equiv C_0\big(G_0^{-1}\big(\widetilde{G}(\tau, \overline{p}), \overline{p}\big), p\big), \quad \text{for all } (\tau, p) \in (0,1) \times \mathcal{P}. \tag{S1}$$

It is simple to see that C is a subcopula.[1] It can thus be extended to a copula on $(0,1) \times (0,1)$ (e.g., Lemma 2.3.5 in Nelsen (1999)). With some abuse of notation, we denote the extension as C, and denote $G(\tau, p) = C(\tau, p)/p$.

Lastly, we assume that the supports of Y^* and Y coincide, denote the support as $\mathcal{Y}$, and we let

$$F(y) \equiv \widetilde{G}^{-1}\big(G_0\big(F_0(y), \overline{p}\big), \overline{p}\big), \quad \text{for all } y \in \mathcal{Y}. \tag{S2}$$

Note that F is a c.d.f.

Let $(\widetilde{U}, \widetilde{V})$ be a bivariate random variable drawn from C, independently of Z. Let $\widetilde{D} = \mathbf{1}\{\widetilde{V} \leq p(Z)\}$, $\widetilde{Y}^* = F^{-1}(\widetilde{U})$, and $\widetilde{Y} = \widetilde{Y}^*$ if $\widetilde{D} = 1$. We start by showing that the distributions of $(\widetilde{Y}, \widetilde{D}, Z)$ and (Y, D, Z) coincide. To see this, note that

$$\begin{aligned}
\Pr(\widetilde{Y} \leq y | \widetilde{D} = 1, Z = z) &= G\big(F(y), p(z)\big) \\
&= G\big(\widetilde{G}^{-1}\big(G_0\big(F_0(y), \overline{p}\big), \overline{p}\big), p(z)\big) \\
&= G_0\big(F_0(y), p(z)\big) \\
&= \Pr(Y \leq y | D = 1, Z = z),
\end{aligned}$$

where we have used (S2) and (S1) in the second and third equalities, respectively.

Finally, to see that F in (S2) can get arbitrarily close to the bounds in (10) and (11), we take $\widetilde{G}$ to be arbitrarily close to the lower and upper Fréchet copula bounds. For the upper bound, we take a conditional copula $\widetilde{G}$ that satisfies Assumption A3 and is arbitrarily close to $(\tau, p) \mapsto \min(\frac{\tau}{p}, 1)$. Similarly, for the lower bound, we take a $\widetilde{G}$ that satisfies Assumption A3 and is arbitrarily close to $(\tau, p) \mapsto \max(\frac{\tau+p-1}{p}, 0)$.[2]

[1] This is because $C(\tau, 0) = C(0, p) = 0$, and C is two-increasing; that is, $C(\tau_2, p_2) - C(\tau_2, p_1) - C(\tau_1, p_2) + C(\tau_1, p_1) \geq 0$ for $\tau_1 \leq \tau_2$ and $p_1 \leq p_2$.

[2] For example, one may take $\widetilde{G}(\tau, p) = C_\theta(\tau, p)/p$ for $\theta > 0$, where

$$C_\theta(\tau, p) \equiv \frac{1}{2(\theta - 1)}\Big(1 + (\tau + p)(\theta - 1) - \sqrt{\big(1 + (\tau + p)(\theta - 1)\big)^2 - 4\tau p \theta(\theta - 1)}\Big)$$

 DOI: 10.3982/ECTA14030

S2. NONPARAMETRIC SPECIFICATION WITH DISCRETE COVARIATES

Consider a model where covariates X and Z are discrete, with a nonparametric quantile specification:

$$q(\tau, X) = X'\beta_\tau = \sum_{k=1}^{K} \beta_{\tau k}\mathbf{1}\{X = x_k\},$$

with x_k denoting the points of support of X. Let $\overline{G}_k(\tau, c)$ denote the mean of $G(\tau, p(Z_i; \widehat{\theta}); c)$ for participants in cell $X_i = x_k$. Let also $\widehat{r}_i$ denote the empirical rank of Y_i in the outcome distribution, conditional on $(D_i = 1, X_i)$. By (16), $x_k'\widehat{\beta}_{\tau k}(c)$ is simply the empirical $\overline{G}_k(\tau, c)$-quantile of Y_i conditional on $(D_i = 1, X_i = x_k)$. It follows that, conditional on $(D_i = 1, X_i = x_k)$, $Y_i \leq X_i'\widehat{\beta}_\tau(c)$ is equivalent to $\widehat{r}_i \leq \overline{G}_k(\tau, c)$.

Let us replace the finite sum in (15) by an integral with respect to a continuous function $\kappa(\tau)$. The above shows that, in the model with discrete covariates, $\widehat{\rho}$ minimizes

$$\begin{aligned}\Bigg\| \sum_{i=1}^{N}\sum_{k=1}^{K}\int_0^1 & D_i\mathbf{1}\{X_i = x_k\}\varphi(\tau, Z_i) \\ & \times \big[\mathbf{1}\big\{\widehat{r}_i \leq \overline{G}_k(\tau, c)\big\} - G\big(\tau, p(Z_i; \widehat{\theta}); c\big)\big]\kappa(\tau)\, d\tau\Bigg\|.\end{aligned}$$

Using the change in variables $u \equiv \overline{G}_k(\tau, c)$, we equivalently have that $\widehat{\rho}$ minimizes the following objective:

$$\begin{aligned}\Bigg\| \sum_{i=1}^{N}\sum_{k=1}^{K}\int_0^1 & D_i\mathbf{1}\{X_i = x_k\}\varphi\big(\overline{G}_k^{-1}(u, c), Z_i\big) \\ & \times \big[\mathbf{1}\{\widehat{r}_i \leq u\} - G\big(\overline{G}_k^{-1}(u, c), p(Z_i; \widehat{\theta}); c\big)\big]\kappa\big(\overline{G}_k^{-1}(u, c)\big)\frac{\partial \overline{G}_k^{-1}(u, c)}{\partial u}\, du\Bigg\|,\end{aligned}$$

which is continuously differentiable with respect to c as long as φ, κ, G, and $\overline{G}_k^{-1}$, $\frac{\partial \overline{G}_k^{-1}}{\partial u}$, are continuously differentiable with respect to τ and c, respectively.

S3. AN ALTERNATIVE ESTIMATOR FOR THE COPULA PARAMETER

From (6) we have, for all $x \in \mathcal{X}$ and $(z_1, z_2) \in \mathcal{Z}_x \times \mathcal{Z}_x$,

$$\mathbb{E}\big(\mathbf{1}\big\{Y \leq q^d(\tau, z_2)\big\}|D = 1, Z = z_1\big) = G\big[G^{-1}\big(\tau, p(z_2; \theta); \rho\big), p(z_1; \theta); \rho\big],$$

where $q^d(\tau, z_2)$ denotes the τ-quantile of Y conditional on $(D = 1, Z = z_2)$.

is the Plackett copula family (e.g., Smith (2003)). Lower and upper Fréchet bounds correspond to $\theta \to 0$ and $\theta \to +\infty$, respectively.

Given consistent estimates $\widehat{q}^d(\tau, z)$ and $\widehat{\theta}$, we thus propose estimating ρ by minimizing the following objective with respect to c:

$$\sum_{i=1}^{N}\sum_{j\neq i}\sum_{\ell=1}^{L} D_i\big(\mathbf{1}\{Y_i \leq \widehat{q}^d(\tau_\ell, B_j, X_i)\} - G\big[G^{-1}\big(\tau_\ell, p(B_j, X_i; \widehat{\theta}); c\big), p(B_i, X_i; \widehat{\theta}); c\big]\big)^2.$$

In case covariates are discrete, the $q^d(\tau, z)$ may be estimated as sample quantiles, cell-by-cell, as in Chamberlain (1993). Alternatively, when covariates are continuous, nonparametric quantile regression methods may be used, such as the series-based quantile regression estimator of Belloni, Chernozhukov, and Fernández-Val (2011). The asymptotic properties of such estimators of ρ could be characterized using U-process techniques (e.g., Jochmans (2013)), although we leave this analysis to future work.

The method can be iterated (possibly multiple times). Recall that the observed quantiles satisfy $q^d(\tau, z) = x'\beta_{G^{-1}(\tau, p(z);\rho)}$. Hence, given estimates $\widehat{\rho}$ and $\widehat{\beta}$, one could estimate

$$\widetilde{q}^d(\tau, z) \equiv x'\widehat{\beta}_{G^{-1}(\tau, p(z);\widehat{\rho})},$$

and update ρ by minimizing

$$\sum_{i=1}^{N}\sum_{j\neq i}\sum_{\ell=1}^{L} D_i\big(\mathbf{1}\{Y_i \leq \widetilde{q}^d(\tau_\ell, B_j, X_i)\} - G\big[G^{-1}\big(\tau_\ell, p(B_j, X_i; \widehat{\theta}); c\big), p(B_i, X_i; \widehat{\theta}); c\big]\big)^2.$$

S4. ASYMPTOTIC PROPERTIES

In this section, we start by deriving the asymptotic distribution of $\widehat{\beta}_\tau$ given a consistent and asymptotically normal estimator of the copula parameter ρ. Then, in the second part of the section, we derive the joint asymptotic distribution of $\widehat{\beta}_\tau$ and $\widehat{\rho}$, for $\widehat{\rho}$ given by (15). The derivations are standard (e.g., Section 7 in Newey and McFadden (1994)).

S4.1. *Analysis Conditional on a Consistent and Asymptotically Normal Estimator of ρ*

Let

$$g_{i\tau} \equiv D_i\big(\mathbf{1}\{Y_i \leq X_i'\beta_\tau\} - G\big(\tau, p(Z_i; \theta); \rho\big)\big).$$

We make the following assumptions.

ASSUMPTION S1:

(i) *There exists a positive definite matrix Σ_τ such that*

$$\sqrt{N}\begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N} X_i g_{i\tau} \\ \widehat{\theta}-\theta \\ \widehat{\rho}-\rho \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma_\tau).$$

(ii) *The c.d.f. of Y given $Z = Z_i$ and $D_i = 1$ is absolutely continuous, with continuous density f_i bounded away from zero and infinity at the points $X_i'\beta_\tau$, $i = 1, \dots, N$.*

(iii) *The function G is continuously differentiable with respect to its second and third arguments, with derivatives $\partial_p G$ and $\partial_\rho G$, respectively. The propensity score $p(\cdot; \theta)$ is continuously differentiable with respect to its second argument, with derivative $\partial_\theta p$.*

(iv) *There exist a positive definite matrix J_τ, and matrices $P_{1\tau}$ and $P_{2\tau}$, such that*

$$J_\tau = \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} p(Z_i;\theta) X_i X_i' f_i\big(X_i'\beta_\tau\big),$$

$$P_{1\tau} = \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} p(Z_i;\theta) X_i \big(\partial_\theta p(Z_i;\theta)\big)' \partial_p G\big(\tau, p(Z_i;\theta);\rho\big),$$

$$P_{2\tau} = \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} p(Z_i;\theta) X_i \big(\partial_\rho G\big(\tau, p(Z_i;\theta);\rho\big)\big)'.$$

Part (i) requires that $\frac{1}{N}\sum_{i=1}^{N} X_i g_{i\tau}$, $\widehat{\theta}$, and $\widehat{\rho}$ jointly satisfy an asymptotic normality result. In particular, this requires ρ to be point-identified from (18). Under weak regularity conditions, it is easy to show that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_i g_{i\tau} \xrightarrow{d} \mathcal{N}\big(0, \mathbb{E}\big[G_{\tau i}(1 - G_{\tau i}) p(Z_i;\theta) X_i X_i'\big]\big),$$

where we have denoted

$$G_{\tau i} \equiv G\big(\tau, p(Z_i;\theta);\rho\big). \tag{S3}$$

Part (ii) is standard in quantile regression (e.g., Theorem 4.2 in Koenker and Bassett (1978)). The only difference here is that we work with the c.d.f. of Y given Z, and not given X. Part (iii) requires that the copula and propensity score be differentiable. Most of the usual parametric families of copulas are differentiable in both their arguments. Exceptions are piecewise-constant empirical copulas, which are not continuous. Lastly, part (iv) requires the existence of moments.

THEOREM S1: *Let $\tau \in (0,1)$, and let Assumptions* A1 *to* A4 *and* S1 *hold. Then, as N tends to infinity,*

$$\sqrt{N}(\widehat{\beta}_\tau - \beta_\tau) \xrightarrow{d} \mathcal{N}\big(0, J_\tau^{-1} P_\tau \Sigma_\tau P_\tau' J_\tau^{-1}\big),$$

where $P_\tau \equiv [I_{\dim\beta}, -P_{1\tau}, -P_{2\tau}]$, and J_τ, $P_{1\tau}$, $P_{2\tau}$ are given in Assumption S1.

Theorem S1 provides the asymptotic distribution of quantile estimates, corrected for the fact that $\widehat{\theta}$ and $\widehat{\rho}$ have been estimated. Note that, in the absence of sample selection, the formula boils down to a well-known expression (Koenker (2005, p. 120)).

PROOF: By a standard result in quantile regression, the following approximate moment condition is satisfied (see, e.g., Theorem 3.3. in Koenker and Bassett (1978)):

$$\frac{1}{N}\sum_{i=1}^{N} X_i g_i(\tau, \widehat{\beta}_\tau, \widehat{\theta}, \widehat{\rho}) = O_p\left(\frac{1}{N}\right), \tag{S4}$$

where

$$g_i(\tau, b, a, c) \equiv D_i\big(\mathbf{1}\big\{Y_i \le X_i'b\big\} - G\big(\tau, p(Z_i; a); c\big)\big).$$

Under standard conditions, we have

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N} X_i g_i(\tau, \widehat{\beta}_\tau, \widehat{\theta}, \widehat{\rho}) &= O_p\left(\frac{1}{N}\right) \\ &= \widehat{\mathbb{E}}[X_i g_{i\tau}] + \frac{\partial \mathbb{E}[X_i g_{i\tau}]}{\partial \beta'}(\widehat{\beta}_\tau - \beta_\tau) + \frac{\partial \mathbb{E}[X_i g_{i\tau}]}{\partial \theta'}(\widehat{\theta} - \theta) \\ &\quad + \frac{\partial \mathbb{E}[X_i g_{i\tau}]}{\partial \rho'}(\widehat{\rho} - \rho) + o_p\left(\frac{1}{\sqrt{N}}\right),\end{aligned}$$

where $J_\tau = \frac{\partial \mathbb{E}[X_i g_{i\tau}]}{\partial \beta'}$, $P_{1\tau} = -\frac{\partial \mathbb{E}[X_i g_{i\tau}]}{\partial \theta'}$, and $P_{2\tau} = -\frac{\partial \mathbb{E}[X_i g_{i\tau}]}{\partial \rho'}$ exist by Assumption S1, parts (ii), (iii), and (iv), and $\widehat{\mathbb{E}}[Z_i] = \frac{1}{N}\sum_{i=1}^{N} Z_i$ denotes a sample mean. Hence, as J_τ is non-singular,

$$\begin{aligned}\widehat{\beta}_\tau - \beta_\tau &= -J_\tau^{-1}\big[\widehat{\mathbb{E}}[X_i g_{i\tau}] - P_{1\tau}(\widehat{\theta} - \theta) - P_{2\tau}(\widehat{\rho} - \rho)\big] + o_p\left(\frac{1}{\sqrt{N}}\right) \\ &= -J_\tau^{-1} P_\tau \begin{pmatrix} \widehat{\mathbb{E}}[X_i g_{i\tau}] \\ \widehat{\theta} - \theta \\ \widehat{\rho} - \rho \end{pmatrix} + o_p\left(\frac{1}{\sqrt{N}}\right).\end{aligned} \tag{S5}$$

The result then comes from part (i) in Assumption S1. *Q.E.D.*

S4.2. *Joint Analysis of $\widehat{\beta}_\tau$ and $\widehat{\rho}$*

We now derive the joint asymptotic distribution of $\widehat{\beta}_\tau$ and $\widehat{\rho}$, for $\widehat{\rho}$ given by (15). For simplicity, we focus on the just-identified case, where ρ and φ have the same dimensions.[3]

The estimation of θ, ρ, and $\beta_{\tau_1}, \dots, \beta_{\tau_L}$ is based on the following just-identified system of moment restrictions (in addition to the score equations for θ):

$$\begin{aligned}\sum_{\ell=1}^{L} \mathbb{E}\big[\varphi(\tau_\ell, Z_i) g_i(\tau_\ell, \beta_{\tau_\ell}, \theta, \rho)\big] &= 0, \\ \mathbb{E}\big[X_i g_i(\tau_1, \beta_{\tau_1}, \theta, \rho)\big] &= 0, \\ \cdots \cdots \cdots& \\ \mathbb{E}\big[X_i g_i(\tau_L, \beta_{\tau_L}, \theta, \rho)\big] &= 0.\end{aligned}$$

Throughout this subsection, we assume that the conditions of Theorem 7.2 in Newey and McFadden (1994) are satisfied, so the estimators are root-N consistent and jointly asymptotically normal. We gather relevant notation in the following assumption, with the aim of deriving explicit expressions for asymptotic variances.

[3]Note that the instrument function $\varphi(\tau, Z_i) = p(Z_i; \widehat{\theta})$ used in Section 5 depends on $\widehat{\theta}$. This slightly affects the formula for the asymptotic variance. For simplicity, here we do not account for this dependence.

ASSUMPTION S2:

(i) *There exists a positive definite matrix H, and a function $S_i \equiv s(D_i, Z_i)$, such that*

$$\widehat{\theta} - \theta = -H^{-1}\widehat{\mathbb{E}}[S_i] + o_p\left(\frac{1}{\sqrt{N}}\right). \tag{S6}$$

(ii) *For all ℓ, there exist a positive definite matrix $\widetilde{J}_{\tau_\ell}$, and matrices $\widetilde{P}_{1\tau_\ell}$ and $\widetilde{P}_{2\tau_\ell}$, such that*

$$\widetilde{J}_{\tau_\ell} = \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} p(Z_i;\theta)\varphi(\tau_\ell, Z_i)X_i' f_i\big(X_i'\beta_{\tau_\ell}\big),$$

$$\widetilde{P}_{1\tau_\ell} = \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} p(Z_i;\theta)\varphi(\tau_\ell, Z_i)\big(\partial_\theta p(Z_i;\theta)\big)' \partial_p G\big(\tau_\ell, p(Z_i;\theta);\rho\big),$$

$$\widetilde{P}_{2\tau_\ell} = \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} p(Z_i;\theta)\varphi(\tau_\ell, Z_i)\big(\partial_\rho G\big(\tau_\ell, p(Z_i;\theta);\rho\big)\big)'.$$

(iii) *The following matrix inverse exists*:

$$A_\rho \equiv \left[\sum_{\ell=1}^{L}\big(\widetilde{P}_{2\tau_\ell} - \widetilde{J}_{\tau_\ell} J_{\tau_\ell}^{-1} P_{2\tau_\ell}\big)\right]^{-1}. \tag{S7}$$

Part (i) will be satisfied if $\widehat{\theta}$ is asymptotically linear, for example when it is a regular maximum likelihood estimator. Parts (ii) and (iii) require that some moments exist.

Define the following matrices:

$$B_\rho \equiv -A_\rho\big[\widetilde{J}_{\tau_1} J_{\tau_1}^{-1}, \dots, \widetilde{J}_{\tau_L} J_{\tau_L}^{-1}\big], \tag{S8}$$

$$C_\rho \equiv A_\rho\left(\sum_{\ell=1}^{L}\big[\widetilde{P}_{1\tau_\ell} - \widetilde{J}_{\tau_\ell} J_{\tau_\ell}^{-1} P_{1\tau_\ell}\big]H^{-1}\right), \tag{S9}$$

and, for a given $\tau \in (0,1)$:

$$A_\beta(\tau) \equiv J_\tau^{-1} P_{2\tau} A_\rho, \tag{S10}$$

$$B_\beta(\tau) \equiv J_\tau^{-1} P_{2\tau} B_\rho, \tag{S11}$$

$$C_\beta(\tau) \equiv J_\tau^{-1}\big(P_{2\tau} C_\rho - P_{1\tau} H^{-1}\big). \tag{S12}$$

Then, let

$$\Delta_\tau \equiv \begin{pmatrix} A_\beta(\tau) & -J_\tau^{-1} & B_\beta(\tau) & C_\beta(\tau) \\ A_\rho & 0 & B_\rho & C_\rho \end{pmatrix}.$$

Lastly, let

$$\sigma_{i\ell m} \equiv \min\{G_{\tau_\ell i}, G_{\tau_m i}\} - G_{\tau_\ell i} G_{\tau_m i},$$

$$\sigma_{i\ell}(\tau) \equiv \min\{G_{\tau_\ell i}, G_{\tau i}\} - G_{\tau_\ell i} G_{\tau i},$$

$$\sigma_i(\tau) \equiv G_{\tau i}(1 - G_{\tau i}),$$

where $G_{\tau i}$ is given by (S3), and define

$$\Omega_\tau \equiv \begin{pmatrix} \Omega_\tau^{1,1} & \Omega_\tau^{1,2} & \cdots & \Omega_\tau^{1,L+2} & 0 \\ \Omega_\tau^{2,1} & \Omega_\tau^{2,2} & \cdots & \Omega_\tau^{2,L+2} & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \Omega_\tau^{L+2,1} & \Omega_\tau^{L+2,2} & \cdots & \Omega_\tau^{L+2,L+2} & 0 \\ 0 & 0 & \cdots & 0 & \mathbb{E}[S_i S_i'] \end{pmatrix}, \tag{S13}$$

where Ω_τ is symmetric, and

$$\begin{aligned}
\Omega_\tau^{1,1} &\equiv \sum_{\ell=1}^{L}\sum_{m=1}^{L}\mathbb{E}\big[\sigma_{i\ell m}p(Z_i;\theta)\varphi(\tau_\ell,Z_i)\varphi(\tau_m,Z_i)'\big],\\
\Omega_\tau^{1,2} &\equiv \sum_{\ell=1}^{L}\mathbb{E}\big[\sigma_{i\ell}(\tau)p(Z_i;\theta)\varphi(\tau_\ell,Z_i)X_i'\big],\\
\Omega_\tau^{1,2+m} &\equiv \sum_{\ell=1}^{L}\mathbb{E}\big[\sigma_{i\ell m}p(Z_i;\theta)\varphi(\tau_\ell,Z_i)X_i'\big], \quad m=1,\dots,L,\\
\Omega_\tau^{2,2} &\equiv \mathbb{E}\big[\sigma_i(\tau)p(Z_i;\theta)X_iX_i'\big],\\
\Omega_\tau^{2,2+m} &\equiv \mathbb{E}\big[\sigma_{im}(\tau)p(Z_i;\theta)X_iX_i'\big], \quad m=1,\dots,L,\\
\Omega_\tau^{2+\ell,2+m} &\equiv \mathbb{E}\big[\sigma_{i\ell m}p(Z_i;\theta)X_iX_i'\big], \quad \ell=1,\dots,L, m=1,\dots,L.
\end{aligned}$$

We have the following result.

THEOREM S2: *Let Assumptions* A1 *to* A4, S1, *and* S2 *hold*. *Suppose that* $\dim\varphi = \dim\rho$. *Then*,

$$\sqrt{N}\begin{pmatrix}\widehat{\beta}_\tau - \beta_\tau \\ \widehat{\rho}-\rho\end{pmatrix} \xrightarrow{d} \mathcal{N}\big(0, \Delta_\tau\Omega_\tau\Delta_\tau'\big).$$

PROOF: As in the proof of Theorem S1, we start with an approximate moment equation:

$$\sum_{\ell=1}^{L}\widehat{\mathbb{E}}\big[\varphi(\tau_\ell,Z_i)g_i(\tau_\ell,\widehat{\beta}_{\tau_\ell},\widehat{\theta},\widehat{\rho})\big] = o_p\bigg(\frac{1}{\sqrt{N}}\bigg).$$

Moreover, we have

$$\begin{aligned}
\sum_{\ell=1}^{L}\widehat{\mathbb{E}}\big[\varphi(\tau_\ell,Z_i)g_i(\tau_\ell,\widehat{\beta}_{\tau_\ell},\widehat{\theta},\widehat{\rho})\big] &= \sum_{\ell=1}^{L}\big\{\widehat{\mathbb{E}}\big[\varphi(\tau_\ell,,Z_i)g_{i\tau_\ell}\big] + \widetilde{J}_{\tau_\ell}(\widehat{\beta}_{\tau_\ell}-\beta_{\tau_\ell})\\
&\quad - \widetilde{P}_{1\tau_\ell}(\widehat{\theta}-\theta) - \widetilde{P}_{2\tau_\ell}(\widehat{\rho}-\rho)\big\} + o_p\bigg(\frac{1}{\sqrt{N}}\bigg).
\end{aligned}$$

So, by (S5),

$$\begin{aligned}
o_p\bigg(\frac{1}{\sqrt{N}}\bigg) &= \sum_{\ell=1}^{L}\big\{\widehat{\mathbb{E}}\big[\varphi(\tau_\ell,Z_i)g_{i\tau_\ell}\big] - \widetilde{P}_{1\tau_\ell}(\widehat{\theta}-\theta) - \widetilde{P}_{2\tau_\ell}(\widehat{\rho}-\rho)\\
&\quad - \widetilde{J}_{\tau_\ell}\big(J_{\tau_\ell}^{-1}\big[\widehat{\mathbb{E}}[X_ig_{i\tau_\ell}] - P_{1\tau_\ell}(\widehat{\theta}-\theta) - P_{2\tau_\ell}(\widehat{\rho}-\rho)\big]\big)\big\} + o_p\bigg(\frac{1}{\sqrt{N}}\bigg).
\end{aligned}$$

So, by (S6),

$$\widehat{\rho}-\rho=\left[\sum_{\ell=1}^{L}\left(\widetilde{P}_{2\tau_\ell}-\widetilde{J}_{\tau_\ell}J_{\tau_\ell}^{-1}P_{2\tau_\ell}\right)\right]^{-1}$$
$$\times\left\{\sum_{\ell=1}^{L}\widehat{\mathbb{E}}\left[\varphi(\tau_\ell,Z_i)g_{i\tau_\ell}\right]-\sum_{\ell=1}^{L}\widetilde{J}_{\tau_\ell}J_{\tau_\ell}^{-1}\widehat{\mathbb{E}}[X_i g_{i\tau_\ell}]\right.$$
$$\left.+\left(\sum_{\ell=1}^{L}\left[\widetilde{P}_{1\tau_\ell}-\widetilde{J}_{\tau_\ell}J_{\tau_\ell}^{-1}P_{1\tau_\ell}\right]H^{-1}\right)\widehat{\mathbb{E}}[S_i]\right\}+o_p\left(\frac{1}{\sqrt{N}}\right).$$

Hence:

$$\widehat{\rho}-\rho=A_\rho\left(\sum_{\ell=1}^{L}\widehat{\mathbb{E}}\left[\varphi(\tau_\ell,Z_i)g_{i\tau_\ell}\right]\right)+B_\rho\widehat{\mathbb{E}}[X_i g_i]+C_\rho\widehat{\mathbb{E}}[S_i]+o_p\left(\frac{1}{\sqrt{N}}\right),$$

where A_ρ, B_ρ, and C_ρ are given by (S7)–(S9), and

$$\widehat{\mathbb{E}}[X_i g_i]=\begin{pmatrix}\widehat{\mathbb{E}}[X_i g_{i\tau_1}]\\ \cdots\\ \widehat{\mathbb{E}}[X_i g_{i\tau_L}]\end{pmatrix}.$$

Let now $\tau\in(0,1)$. Using (S5),

$$\widehat{\beta}_\tau-\beta_\tau=-J_\tau^{-1}\left[\widehat{\mathbb{E}}[X_i g_{i\tau}]-P_{1\tau}(\widehat{\theta}-\theta)-P_{2\tau}(\widehat{\rho}-\rho)\right]+o_p\left(\frac{1}{\sqrt{N}}\right)$$
$$=-J_\tau^{-1}\Bigg[\widehat{\mathbb{E}}[X_i g_{i\tau}]+P_{1\tau}H^{-1}\widehat{\mathbb{E}}[S_i]$$
$$-P_{2\tau}\left(A_\rho\left(\sum_{\ell=1}^{L}\widehat{\mathbb{E}}\left[\varphi(\tau_\ell,Z_i)g_{i\tau_\ell}\right]\right)+B_\rho\widehat{\mathbb{E}}[X_i g_i]+C_\rho\widehat{\mathbb{E}}[S_i]\right)\Bigg]$$
$$+o_p\left(\frac{1}{\sqrt{N}}\right).$$

So

$$\widehat{\beta}_\tau-\beta_\tau=A_\beta(\tau)\left(\sum_{\ell=1}^{L}\widehat{\mathbb{E}}\left[\varphi(\tau_\ell,Z_i)g_{i\tau_\ell}\right]\right)-J_\tau^{-1}\widehat{\mathbb{E}}[X_i g_{i\tau}]$$
$$+B_\beta(\tau)\widehat{\mathbb{E}}[X_i g_i]+C_\beta(\tau)\widehat{\mathbb{E}}[S_i]+o_p\left(\frac{1}{\sqrt{N}}\right),$$

where $A_\beta(\tau)$, $B_\beta(\tau)$, and $C_\beta(\tau)$ are given by (S10)–(S12).

Next, denote

$$\psi_{i\tau} \equiv \begin{pmatrix} \sum_{\ell=1}^{L} \varphi(\tau_\ell, Z_i) g_{i\tau_\ell} \\ X_i g_{i\tau} \\ X_i g_{i\tau_1} \\ \cdots \\ X_i g_{i\tau_L} \\ S_i \end{pmatrix}.$$

From the above, we have

$$\sqrt{N} \begin{pmatrix} \widehat{\beta}_\tau - \beta_\tau \\ \widehat{\rho} - \rho \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, V_\tau),$$

with

$$V_\tau = \begin{pmatrix} A_\beta(\tau) & -J_\tau^{-1} & B_\beta(\tau) & C_\beta(\tau) \\ A_\rho & 0 & B_\rho & C_\rho \end{pmatrix} \times \mathbb{E}\big(\psi_{i\tau}\psi_{i\tau}'\big) \begin{pmatrix} A_\beta(\tau) & -J_\tau^{-1} & B_\beta(\tau) & C_\beta(\tau) \\ A_\rho & 0 & B_\rho & C_\rho \end{pmatrix}'.$$

Finally, we check that $\mathbb{E}(\psi_{i\tau}\psi_{i\tau}') = \Omega_\tau$ given by (S13):

$$\mathbb{E}\left[\left(\sum_{\ell=1}^{L} \varphi(\tau_\ell, Z_i) g_{i\tau_\ell}\right)\left(\sum_{m=1}^{L} \varphi(\tau_m, Z_i) g_{i\tau_m}\right)'\right] = \sum_{\ell=1}^{L}\sum_{m=1}^{L} \mathbb{E}\big[\sigma_{i\ell m} p(Z_i;\theta)\varphi(\tau_\ell, Z_i)\varphi(\tau_m, Z_i)'\big],$$

and similarly,

$$\mathbb{E}\left[\left(\sum_{\ell=1}^{L} \varphi(\tau_\ell, Z_i) g_{i\tau_\ell}\right)(X_i g_{i\tau_m})'\right] = \sum_{\ell=1}^{L} \mathbb{E}\big[\sigma_{i\ell m} p(Z_i;\theta)\varphi(\tau_\ell, Z_i) X_i'\big],$$

$$\mathbb{E}\big[(X_i g_{i\tau_\ell})(X_i g_{i\tau_m})'\big] = \mathbb{E}\big[\sigma_{i\ell m} p(Z_i;\theta) X_i X_i'\big],$$

and, as S_i is a function of (D_i, Z_i), we have $\mathbb{E}[g_{i\tau_\ell} S_i'] = 0$.

This completes the proof of Theorem S2. *Q.E.D.*

Estimating the Asymptotic Variance

To construct an empirical counterpart of the asymptotic variance appearing in Theorem S1, note that all matrices but J_τ can be estimated by sample analogs, replacing the population expectations by empirical means. Moreover, following Powell (1986), a consistent estimator of J_τ is

$$\widehat{J}_\tau = \frac{1}{2Nh_N} \sum_{i=1}^{N} \mathbf{1}\big\{\big|\widehat{\varepsilon}_i(\tau)\big| \leq h_N\big\} D_i X_i X_i',$$

where $\widehat{\varepsilon}_i \equiv Y_i - X_i'\widehat{\beta}_\tau$, and h_N is a bandwidth that satisfies $h_N \to 0$ and $Nh_N^2 \to +\infty$ as N tends to infinity. We may proceed similarly to estimate $\widetilde{J}_\tau$ that appears in Theorem S2.

S5. EXTENSIONS

Nonparametric Propensity Score

Although the paper focuses on the case where the propensity score is parametrically specified, our approach can accommodate a nonparametric modeling of $p(Z)$ as well. A difficulty is that the G function has $p(Z)$ in the denominator. A similar problem arises in Buchinsky and Hahn's (1998) censored quantile regression estimator. Similarly as in Buchinsky and Hahn, one could trim out the observations for which $\widehat{p}(Z_i) < c$, where $\widehat{p}(Z)$ is a nonparametric estimate (e.g., a kernel-based Nadaraya–Watson estimator) and $c > 0$ is a vanishing trimming threshold. We leave this extension to future work.

Testing for the Absence of Sample Selection

Under the null hypothesis of absence of sample selection, we have $G(\tau, p(Z;\theta);\rho) = \tau$. So, β_τ satisfies

$$\mathbb{E}\big[\mathbf{1}\big\{Y \le X'\beta_\tau\big\} - \tau | D = 1, Z = z\big] = 0, \quad \text{for all } \tau \in (0,1).$$

This motivates using a test statistic of the form

$$S = \left\| \sum_{\ell=1}^{L} \sum_{i=1}^{N} D_i \varphi(\tau_\ell, Z_i)\big(\mathbf{1}\big\{Y_i \le X_i'\widehat{\beta}_{\tau_\ell}\big\} - \tau_\ell\big)\right\|^2,$$

where $\varphi(\tau, Z_i)$ are instrument functions, and $\widehat{\beta}_\tau$ is the quantile regression estimate of the τ-specific slope coefficient, computed on the sample of participants ($D_i = 1$).

Endogeneity

Let us assume that the latent outcome is given by the following linear quantile model:

$$Y^* = E'\alpha_U + X'\beta_U, \tag{S14}$$

where the percentile level U is independent of X, but may be correlated with the endogenous regressor E. As before, the participation equation is given by (2). Suppose that (U, V) is independent of Z given X. Assume also that $q(\tau, X, E) \equiv E'\alpha_\tau + X'\beta_\tau$ is strictly increasing in its first argument. Then, for any $\tau \in (0,1)$,

$$\mathbb{E}\big[\mathbf{1}\big\{Y \le E'\alpha_\tau + X'\beta_\tau\big\} - G\big(\tau, p(Z;\theta);\rho\big) \mid D = 1, Z = z\big] = 0. \tag{S15}$$

To estimate ρ, θ, and $\{\alpha_\tau, \beta_\tau\}$ for any $\tau \in (0,1)$, one can use the following three-step estimation method, which extends Chernozhukov and Hansen's (2006) estimator to correct for selection. In the first step, we compute $\widehat{\theta}$. In the second step, we compute $\widehat{\rho}$ as

$$\widehat{\rho} = \underset{c}{\operatorname{argmin}} \left\| \sum_{\ell=1}^{L} \sum_{i=1}^{N} D_i \varphi(\tau_\ell, Z_i)\big(\mathbf{1}\big\{Y_i \le E_i'\widetilde{\alpha}_{\tau_\ell}(c) + X_i'\widetilde{\beta}_{\tau_\ell}\big(\widetilde{\alpha}_{\tau_\ell}(c); c\big)\big\} \right.$$
$$\left. - G\big(\tau_\ell, p(Z_i;\widehat{\theta}); c\big)\big) \right\|,$$

where, for $\mu_\tau(Z_i)$ a $\dim\alpha \times 1$ vector of instruments, we have defined

$$\big(\widetilde{\beta}_\tau(\alpha; c), \widetilde{\gamma}_\tau(\alpha; c)\big) \equiv \underset{(b,g)}{\operatorname{argmin}} \sum_{i=1}^{N} D_i\big\{G\big(\tau, \widehat{p}(Z_i; \widehat{\theta}); c\big)\big(Y_i - X_i'b - \mu_\tau(Z_i)'g\big)^+ + \big(1 - G\big(\tau, \widehat{p}(Z_i; \widehat{\theta}); c\big)\big)\big(Y_i - X_i'b - \mu_\tau(Z_i)'g\big)^-\big\},$$

and

$$\widetilde{\alpha}_\tau(c) \equiv \underset{a}{\operatorname{argmin}} \big\| \widetilde{\gamma}_\tau(a; c)\big\|.$$

Lastly, once $\widehat{\rho}$ has been estimated, we compute $\widehat{\alpha}_\tau \equiv \widetilde{\alpha}_\tau(\widehat{\rho})$, and $\widehat{\beta}_\tau \equiv \widetilde{\beta}_\tau(\widehat{\alpha}_\tau; \widehat{\rho})$.

Censoring

Suppose that Y^* is censored when $Y^* < y_0$, where y_0 is a known threshold, so that we observe $Y = \max\{Y^*, y_0\}$ when $D = 1$. From the equivariance property of quantiles, the τ-quantile of $\max\{Y^*, y_0\}$ is $\max\{X'\beta_\tau, y_0\}$. So, under Assumptions A1 to A4,

$$\Pr\big(Y \le \max\{X'\beta_\tau, y_0\} | D = 1, Z = z\big) = G\big(\tau, p(z; \theta); \rho\big). \tag{S16}$$

This implies that the $G(\tau, p(Z; \theta); \rho)$-quantile of observed outcomes coincides with $\max\{X'\beta_\tau, y_0\}$. The β coefficients can thus be estimated as in the main text, replacing $X_i'b$ and $X_i'\widehat{\beta}_\tau(c)$ by $\max\{X_i'b, y_0\}$ and $\max\{X_i'\widetilde{\beta}_\tau(c), y_0\}$, respectively, where

$$\widetilde{\beta}_\tau(c) \equiv \underset{b}{\operatorname{argmin}} \sum_{i=1}^{N} D_i\big\{G\big(\tau, \widehat{p}(Z_i; \widehat{\theta}); c\big)\big(Y_i - \max\{X_i'b, y_0\}\big)^+ + \big(1 - G\big(\tau, \widehat{p}(Z_i; \widehat{\theta}); c\big)\big)\big(Y_i - \max\{X_i'b, y_0\}\big)^-\big\}. \tag{S17}$$

The optimization problem in (S17) is a selection-corrected version of Powell's (1986) censored quantile estimator.

S6. FRANK AND GENERALIZED FRANK COPULAS

Let us consider the following two-parameter family of copulas, which we call the "generalized Frank" family for reasons that will be clear below. The copula depends on two parameters $\theta \ge 1$ and $\gamma \in (0, 1)$, and is given by

$$C(u, v; \gamma, \theta) = \frac{1}{\delta}\left[1 - \left\{1 - \frac{1}{\gamma}\big[1 - (1 - \delta u)^\theta\big]\big[1 - (1 - \delta v)^\theta\big]\right\}^{\frac{1}{\theta}}\right], \tag{S18}$$

where $\delta = 1 - (1 - \gamma)^{\frac{1}{\theta}}$. Joe (1997) referred to (S18) as the "BB8" copula.

It is convenient to introduce the following *concordance* ordering $\prec$ on copulas:

$$C_1 \prec C_2 \quad \text{if and only if} \quad C_1(u, v) \le C_2(u, v) \text{ for all } (u, v).$$

As $\prec$ is the first-order stochastic dominance ordering, $C_1 \prec C_2$ unambiguously indicates that C_1 induces less correlation than C_2. The concordance of the generalized Frank copula given by (S18) increases in θ and γ. In particular, $\theta = 1$ or $\gamma \to 0$ correspond to the independent copula.

An interesting special case is obtained when $\theta \to \infty$, for fixed γ. Then

$$C(u, v; \gamma, \theta) \underset{\theta\to\infty}{\to} C_F(u, v; \gamma),$$

where

$$C_F(u, v; \gamma) = \frac{1}{\ln(1-\gamma)} \ln\left[1 - \frac{1}{\gamma}\{1 - \exp[\ln(1-\gamma)u]\}\{1 - \exp[\ln(1-\gamma)v]\}\right]. \tag{S19}$$

C_F given by (S19) is the Frank copula (Frank (1979)), with parameter $\eta = -\ln(1-\gamma)$. Here also, concordance increases with η.

The density of the Frank copula is symmetric with respect to the point $(\frac{1}{2}, \frac{1}{2})$ in the (U, V) plane. In comparison, the generalized Frank copula (S18) permits some asymmetries, by allowing the dependence to increase on the main diagonal. However, the generalized Frank copula treats symmetrically u and v, so that it is symmetric with respect to the main diagonal.

Taking negative η, the Frank copula exhibits negative dependence. This is important in our empirical application, as we estimate that U and V are negatively correlated. To allow for negative dependence in the generalized Frank copula, we simply consider

$$\widetilde{C}(u, v; \gamma, \theta) = v - C(1-u, v; \gamma, \theta),$$

which is the copula of $(1-U, V)$ where (U, V) is distributed as C.[4] In addition, by taking instead the copula of $(U, 1-V)$, we obtain

$$\widetilde{C}(u, v; \gamma, \theta) = u - C(u, 1-v; \gamma, \theta).$$

In this way, we may allow for decreasing dependence along the second diagonal.

S7. ADDITIONAL FIGURES: FIT OF THE MODEL USED IN THE EQUILIBRIUM COUNTERFACTUAL EXERCISE

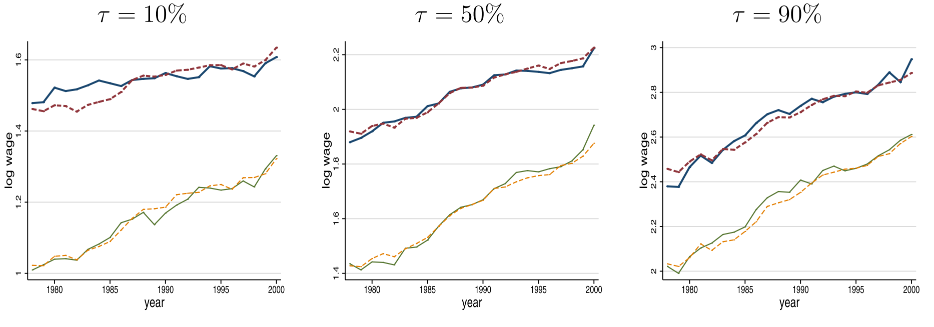

FIGURE S1.—Fit to wage quantiles, by gender. *Note*: FES data for 1978–2000. Specification used in Section 6. Quantiles of log-hourly wages conditional on employment. Data (solid lines) and predicted by the model (dashed). Male wages (at the top) are plotted in thick lines, while female wages are in thin lines.

[4]This is because $\Pr(1-U \le u, V \le v) = \Pr(V \le v) - \Pr(1-U > u, V \le v) = v - C(1-u, v; \gamma, \theta)$.

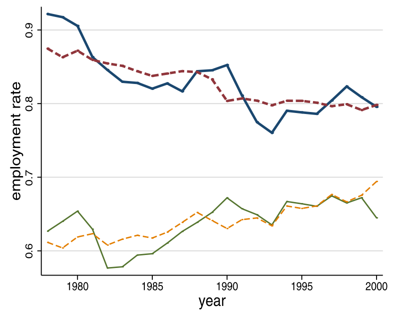

FIGURE S2.—Fit to employment, by gender. *Note*: FES data for 1978–2000. Specification used in Section 6. Employment rate in the data (solid lines) and predicted by the model (dashed). Male employment (at the top) is plotted in thick lines, while female employment is in thin lines.

REFERENCES

BELLONI, A., V. CHERNOZHUKOV, AND I. FERNÁNDEZ-VAL (2011): "Conditional Quantile Processes Based on Series or Many Regressors," Unpublished Manuscript. [3]

BUCHINSKY, M., AND J. HAHN (1998): "An Alternative Estimator for the Censored Regression Model," *Econometrica*, 66, 653–671. [10]

CHAMBERLAIN, G. (1993): "Quantile Regressions, Censoring and the Structure of Wages," in *Advances in Econometrics: Proceedings of the 6th World Congress in Barcelona, Vol. I*, ed. by C. Sims. Cambridge: Cambridge University Press. [3]

CHERNOZHUKOV, V., AND C. HANSEN (2006): "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models," *Journal of Econometrics*, 132, 491–525. [10]

FRANK, M. J. (1979): "On the Simultaneous Associativity of $F(x, y)$ and $x + y - F(x, y)$," *Aequationes Mathematicae*, 19, 194–226. [12]

JOCHMANS, K. (2013): "Pairwise-Comparison Estimation With Non-parametric Controls," *Econometrics Journal*, 16 (3), 340–372. [3]

JOE, H. (1997): *Multivariate Models and Dependence Concepts*. London: Chapman & Hall. [11]

KOENKER, R. (2005): *Quantile Regression*. Econometric Society Monograph Series. Cambridge: Cambridge University Press. [4]

KOENKER, R., AND G. BASSETT (1978): "Regression Quantiles," *Econometrica*, 46, 33–50. [4]

NELSEN, R. B. (1999): *An Introduction to Copulas*. New-York: Springer Verlag. [1]

NEWEY, W. K., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics*, 4, 2111–2245. [3,5]

POWELL, J. L. (1986): "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143–155. [9,11]

SMITH, M. D. (2003): "Modelling Sample Selection Using Archimedean Copulas," *The Econometrics Journal*, 6, 99–123. [2]

CEMFI, Casado del Alisal, 5, Madrid, 28014, Spain; arellano@cemfi.es

and

Dept. of Economics, University of Chicago, 1126 East 59th Street, Chicago, Illinois 60637, U.S.A.; sbonhomme@uchicago.edu.

Co-editor Elie Tamer handled this manuscript.

Manuscript received 18 December, 2015; final version accepted 26 September, 2016; available online 3 October, 2016.