Quantile Selection Models

Manuel Arellano
CEMFI, Madrid

Stéphane Bonhomme
CEMFI, Madrid

April 2011

Work in progress - Please do not cite without permission

Presentation prepared for the Banco de Portugal Labour Market Conference, Douro, June 4-5, 2011

Abstract

We propose a method to correct for sample selection in quantile regression models. Selection is modelled via the cumulative distribution function, or copula, of the percentile error in the outcome equation and the error in the participation decision. Copula parameters are estimated using an iterated nonlinear least squares procedure. Given copula parameters, the percentile levels of the outcome are re-ajusted to correct for selection, and quantile parameters are estimated by minimizing a rotated "check" function. We apply the method to correct wage percentiles for selection into employment, using data for the UK for the period 1978-2000. We find that correcting for sample selection magnifies the increase in wage inequality over the period.

JEL CODE: C13, J31.
KEYWORDS: Quantiles, sample selection, copula, gender wage gap.

1 Introduction

Sample selection

- Non-random sample selection is a major issue in empirical work.
- Important example (and our application): wages and employment (Gronau 74, Heckman 76).
- In regression models, the solution involves adding a selection factor as a control (Heckman 79, Das, Newey and Vella 03).
- Excluded variables (e.g., determinants of employment that do not affect wages directly) are key to achieve credible identification.

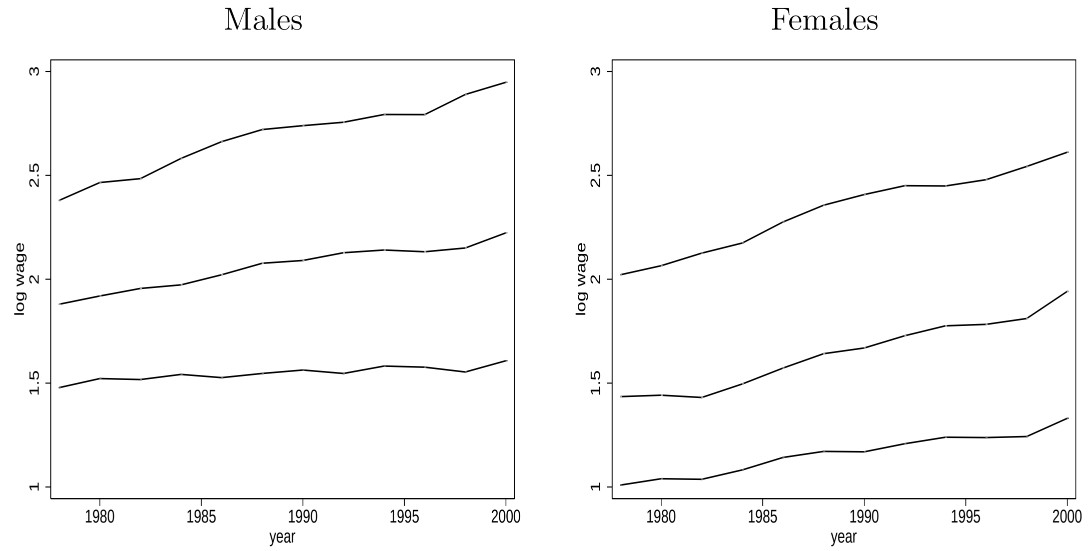
Distributions

- Most selection-correction approaches focus on estimating (conditional) mean models.
- In many applications, however, a flexible specification of the entire distribution of outcomes is of interest.
- Example 1: when employment rates vary over time, the evolution of observed wage inequality and latent wage inequality may differ.
- Example 2: if men and women have different employment rates, the gender wage gap conditional on employment may over/underestimate the latent gap.

Quantile regression

- Quantile regression is widely used to estimate conditional wage distributions.
- Linear model: each percentile $\tau \in (0,1)$ is associated with a conditional quantile $q_\tau(y|x) = x'\beta_\tau$.
- β_τ can be estimated by minimizing a convex (check) function (Koenker & Bassett 78).
- However: to our knowledge there is yet no widely accepted quantile regression approach in the presence of sample selection.
- Thus, this paper aims at contributing to fill this gap.

Figure 1: Wage inequality and gender wage gap in the UK



Note: 10%, 50% and 90%-percentiles of log-hourly wages.
Source: Family Expenditure Survey, 1978-2000.

Figure 2: Employment rates in the UK

Source: Family Expenditure Survey, 1978-2000.

This paper

- We propose a selection correction method for quantile regression.
  - In the absence of sample selection, our estimator coincides with standard quantile regression.
  - In the presence of selection, our approach consists in shifting the percentile levels as a function of the amount of selection.
- Our approach connects with:
  - Bounds approaches (Manski 94, Blundell, Gosling, Ichimura, and Meghir 07).
  - Parametric and semiparametric versions of the Heckman (79) sample selection model.

Literature I: Quantiles, distributions, and treatment effects

- Chernozhukov and Hansen (05, 06): Instrumental variables quantile regression. Need to observe potential outcomes for treated and non-treated, and a rank invariance or similarity assumption.
- See also Torgovitsky (10) for a model with continuous endogenous regressors.
- Imbens and Rubin (97): Identification and estimation of unconditional distributions of potential outcomes in treatment effects models. Binary instrument, identification for compliers (as in Abadie 03, and in Abadie, Angrist and Imbens 02).
- Carneiro and Lee (09): Uses the framework of Heckman and Vytlacil (05) to identify and estimate distributions of potential outcomes.

Literature II: Quantile selection models

- Buchinsky (98, 01): Proposed an additive approach to correct for sample selection in quantile regression.
- However: difficult to specify a data-generating process that is consistent with this approach (Albrecht, van Vuuren and Vroman 09).

- Picchio and Mussida (10): Flexible parametric model to correct the gender wage gap for selection into employment ($\neq$ quantile regression).
- Huber and Melly (10): Deal with a similar model as we do. They focus on testing for additivity. In contrast, our focus is on providing a practical estimation method.

Outline

- Introduction
- Model
- Identification
- Estimation
- Properties and extensions
- Empirical application
- Conclusion

2 Model

2.1 The quantile selection model

- We assume:

$$
\begin{aligned}
Y^* &= Q_U\left(Y^* \mid X\right), \\
D &= \mathbf{1}\left\{V \leq p\left(Z\right)\right\}, \\
Y &= Y^* \quad \text{if} \quad D = 1.
\end{aligned}
$$

- Y^* is the latent outcome (e.g. market wages).
- D is the participation indicator (employment).
- $Z = (W, X)$ strictly contains X, so W is the exclusion restriction.
- In estimation, we will restrict the conditional U-quantiles of Y^* given X to be linear:

$$
Q_U\left(Y^* \mid X = x\right) = x'\beta_U.
$$

Assumptions

- **A1** (exclusion restriction): (U, V) is statistically independent of Z given X.
- **A2** (normalization): (U, V) given $X = x$ has uniform marginals. We denote its cumulative distribution function (or "copula") as $C_x\left(u, v\right)$.
- **A3** (continuous outcomes): The conditional *cdf* $F_{Y^*|X}\left(y|x\right)$ is strictly increasing in y. In addition, $u \mapsto C_x(u, v)$ is strictly increasing.
- **A4** (propensity score): $p\left(Z\right) = \Pr\left(D = 1|Z\right) > 0$ with probability one.

2.2 Examples

Special case I: additive selection model

- Suppose that outcomes are additive in the unobservables:

$$
\begin{aligned}
Y^* &= g\left(X\right) + \varepsilon, \\
D &= \mathbf{1}\left\{V \leq p\left(Z\right)\right\},
\end{aligned}
$$

  where (ε, V) is statistically independent of Z.

- Then, the following restrictions hold (Das *et al.* 03):

$$\begin{aligned}\mathbb{E}(Y|D=1,Z) &= g(X)+\mathbb{E}(\varepsilon|Z,V\leq p(Z)) \\ &= g(X)+\lambda(p(Z)),\end{aligned}$$

  where $\lambda(p)=\mathbb{E}(\varepsilon|V\leq p)$.

- Assumption A1 is satisfied, with $U=F_{\varepsilon}(\varepsilon)$.

Special case II: reservation wage

- Suppose the following reservation rule:

$$D = \mathbf{1}\{Y^{*}\geq R(Z)+\eta\},$$

  where (Y^{*},η) is statistically independent of Z given X.

- Equivalently:

$$D = \mathbf{1}\left\{V\leq F_{\eta-Y^{*}|Z}(-R(Z)|Z)\right\},$$

  where $V=F_{\eta-Y^{*}|Z}(\eta-Y^{*}|Z)$ is standard uniform, independent of Z.

- Let $Y^{*}=Q_{U}(Y^{*}\mid X)$, non-additive in U. In this case, U and V are independent of Z given X, but are not jointly independent of X.

Extension: treatment effects with selection on unobservables.

- Consider the system:

$$\begin{aligned}Y_0^{*} &= Q_{U_0}(Y_0^{*}\mid X), \\ Y_1^{*} &= Q_{U_1}(Y_1^{*}\mid X), \\ D &= \mathbf{1}\{V\leq p(Z)\}, \\ Y &= (1-D)Y_0^{*}+DY_1^{*},\end{aligned}$$

  where (U_0,U_1,V) is independent of Z given X.

- The model coincides with the standard potential outcomes framework in the treatment effects literature (Vytlacil 02).

- We make no assumption of rank invariance or rank similarity (unlike Chernozhukov and Hansen 05). $\Rightarrow$ Our approach recovers the distributions of Y_0^{*} and Y_1^{*} given X.

3 Identification

3.1 Restrictions and conditions for point-identification

Latent quantiles

- The τ-quantile of latent outcomes is $q_\tau(x) \equiv Q_\tau(Y^* \mid X = x)$. That is:

$$\begin{aligned}
\Pr(Y^* \leq q_\tau(x)|X = x) &= \Pr\left[q_U(x) \leq q_\tau(x)|X = x\right] \\
&= \Pr(U \leq \tau|X = x) \\
&= \tau.
\end{aligned}$$

- In the absence of sample selection, $q_\tau(x)$ could be estimated using standard (parametric or nonparametric) quantile regression.

- Unfortunately, we do not observe Y^* when $D = 0$.

Observed quantiles

- Conditioning on participation, and letting $z = (x, w)$:

$$\Pr(Y^* \leq q_\tau(x)|D = 1, Z = z) = \underbrace{\Pr(U \leq \tau|V \leq p(z), Z = z)}_{\equiv G_x(\tau, p(z))},$$

  where $G_x(\tau, p(z)) = C_x(\tau, p(z))/p(z)$.

- Hence, the conditional τ-quantile of Y^* coincides with the conditional $G_x(\tau, p(z))$-quantile of Y given $D = 1$.

  $\Rightarrow$ If we knew the mapping G_x from latent to observed ranks, one could estimate $q_\tau(x)$ using rotated quantile regression.

- Key insight: the exclusion restriction provides information about the mapping G_x.

Identifying restrictions

- Given two values $z = (x, w)$ and $\widetilde{z} = (x, \widetilde{w})$ we have:

$$\underbrace{F_{Y|D=1,Z}\left(F^{-1}_{Y|D=1,Z}\left(\tau\middle|\widetilde{z}\right)\middle|z\right)}_{\text{data}} = G_x\left(G_x^{-1}(\tau, p(\widetilde{z})), p(z)\right). \qquad (1)$$

- These restrictions are uninformative in the absence of an exclusion restriction (i.e., when $z = \widetilde{z} = x$).

- Moreover, these are the only restrictions on G_x: For any G_x satisfying (1), one can find a distribution of latent outcomes $F_{Y^*|X}$ that rationalizes the data.

Point-identification

- Two simple conditions lead to point-identification of $G_x\left(., p(z)\right)$:

  1. Identification at infinity: there exists some $z_x = (x, w_x)$ such that $p(z_x) = 1$.
  2. Flexible parametric identification: the function G_x is real analytic.
     $\Rightarrow$ extrapolated outside an open neighborhood.

  $\Rightarrow$ Two sources of identification: variation in the excluded variable, and functional forms.

- In either of these two cases, the quantiles $q_\tau(x)$ are also point-identified, for all $\tau \in (0, 1)$.

3.2 Partial identification

Bounds on G_x

- Note that, if $G_x(., p(\widetilde{z}))$ were known for some $\widetilde{z}$, then by (1) $G_x(., p(z))$ would be known for all z values.
  $\Rightarrow$ G_x is identified up to a monotone transformation of its first argument.

- Fix $\widetilde{z}$, and bound $G_x(., p(\widetilde{z}))$ by the worst-case ("Fréchet") bounds:

$$\max\left(\frac{\tau + p(\widetilde{z}) - 1}{p(\widetilde{z})}, 0\right) \leq G_x\left(\tau, p(\widetilde{z})\right) \leq \min\left(\frac{\tau}{p(\widetilde{z})}, 1\right).$$

- This implies the following bounds on G_x:

$$\begin{aligned} G_x\left(\tau, p(z)\right) &\leq \inf_{\widetilde{z}} F_{Y|D=1,Z}\left[F^{-1}_{Y|D=1,Z}\left(\min\left(\frac{\tau}{p(\widetilde{z})}, 1\right) \Big| \widetilde{z}\right) \Big| z\right] \\ G_x\left(\tau, p(z)\right) &\geq \sup_{\widetilde{z}} F_{Y|D=1,Z}\left[F^{-1}_{Y|D=1,Z}\left(\max\left(\frac{\tau + p(\widetilde{z}) - 1}{p(\widetilde{z})}, 0\right) \Big| \widetilde{z}\right) \Big| z\right]. \end{aligned}$$

Bounds on $q_\tau(x)$

- Using the above bounds on G_x, we can bound the quantiles of latent outcomes as:

$$
\begin{aligned}
q_\tau(x) &\leq \inf_{\widetilde{z}} F^{-1}_{Y|D=1,Z}\left(\min\left(\frac{\tau}{p(\widetilde{z})},1\right)\Big|\,\widetilde{z}\right)\\
q_\tau(x) &\geq \sup_{\widetilde{z}} F^{-1}_{Y|D=1,Z}\left(\max\left(\frac{\tau+p(\widetilde{z})-1}{p(\widetilde{z})},0\right)\Big|\,\widetilde{z}\right),
\end{aligned}
$$

  where the infimum and supremum are taken with respect to elements of the support of Z given $X=x$.

- These bounds are sharp, and coincide with the ones derived in Manski (94).

4 Estimation

4.1 Preliminaries

Parametric specification

- Quantile functions are linear: $q_\tau(x)=x'\beta_\tau$.

- $G_x(\tau,p)\equiv G(\tau,p;\rho)=\frac{C(\tau,p;\rho)}{p}$ is indexed by a parameter vector ρ (remark: could depend on x).

  Many convenient parsimonious specifications are available (e.g., based on Gaussian, Frank, or Gumbel copulas). More flexible alternatives exist (e.g., based on Bernstein copulas).

- $p(z;\theta)$ is a known function of a parameter θ (this may be relaxed).

- We will denote as $\widehat{\theta}$ a consistent (maximum-likelihood) estimate of θ.

Digression: quantile curves are non-additive in x and $p(z)$

- The τ-quantile of observed outcomes given $z=(x,w)$ is:

$$q^d_\tau(z)\equiv F^{-1}_{Y|D=1,Z}(\tau|\,z)=x'\beta_{G^{-1}(\tau,p(z))}.$$

- We show that $q_\tau^d(z)$ is non-additive in x and $p(z)$, unless:

  i) All coefficients of β_τ but the constant are independent of τ, or

  ii) U and V are statistically independent.

- However, additive specifications such as $q_\tau^d(z) = x'\beta_\tau + \lambda_\tau(p(z))$ (for a smooth function $\lambda_\tau(p)$) are often used in applied work.

4.2 Three-step estimation strategy

- We propose to compute selection-corrected quantile regression estimates in three steps:

  1. Compute $\widehat{\theta}$, and predict $p\left(z_i; \widehat{\theta}\right)$.
  2. Compute $\widehat{\rho}$. This is the computationally expensive step.
  3. For any given τ, compute $G\left(\tau, p\left(z_i; \widehat{\theta}\right); \widehat{\rho}\right)$, and compute $\widehat{\beta}_\tau$.

- This may be done using a standard quantile regression algorithm.

Step 3: estimation of β_τ

- Given $\widehat{\rho}$ (and given $\widehat{\theta}$), $\widehat{\beta}_\tau$ can be estimated for any given τ as:

$$\widehat{\beta}_\tau = \underset{b}{\operatorname{argmin}} \sum_{i=1}^N D_i \left[\widehat{G}_{\tau i} \left(y_i - x_i'b\right)^+ + \left(1 - \widehat{G}_{\tau i}\right) \left(y_i - x_i'b\right)^-\right],$$

  where $a^+ = \max(a, 0)$, $a^- = \max(-a, 0)$, and:

$$\widehat{G}_{\tau i} = G\left(\tau, p\left(z_i; \widehat{\theta}\right); \widehat{\rho}\right).$$

- This amounts to minimizing a *rotated check* function.

  $\Rightarrow$ Estimation using a simple linear program.

- Compare with the (infeasible) quantile regression estimate:

$$\widetilde{\beta}_\tau = \underset{b}{\operatorname{argmin}} \sum_{i=1}^N \left[\tau \left(y_i^* - x_i'b\right)^+ + (1 - \tau) \left(y_i^* - x_i'b\right)^-\right].$$

Step 2: estimation of ρ, Method A

- The moment restrictions are:

$$\mathbb{E}\left[\mathbf{1}\left\{Y \leq x'\beta_\tau\right\} - G\left(\tau, p(z;\theta);\rho\right) \,\middle|\, D=1, Z=z\right] \quad = \quad 0.$$

- We propose estimating ρ by:

$$\widehat{\rho} = \underset{c}{\operatorname{argmin}} \left\| \sum_{i=1}^{N} \sum_{\ell} D_i \varphi_{\tau_\ell}\left(z_i\right) \left[\mathbf{1}\left\{y_i \leq x_i' \widehat{\beta}_{\tau_\ell}\left(c\right)\right\} - G\left(\tau_\ell, p(z_i;\widehat{\theta});c\right)\right] \right\|,$$

where $\{\tau_\ell\}$ is a finite grid, $\varphi_\tau\left(z_i\right)$ are instrument functions, and:

$$\begin{aligned} \widehat{\beta}_\tau\left(c\right) \quad = \quad \underset{b}{\operatorname{argmin}} \sum_{i=1}^{N} D_i \Big[& G\left(\tau, p(z_i;\widehat{\theta});c\right)\left(y_i - x_i'b\right)^{+} \\ & + \left(1 - G\left(\tau, p(z_i;\widehat{\theta});c\right)\right)\left(y_i - x_i'b\right)^{-} \Big]. \end{aligned}$$

Step 2: estimation of ρ, Method B

- From the identification discussion we have:

$$\mathbb{E}\left(\mathbf{1}\left\{Y \leq q_\tau^d\left(\widetilde{z}\right)\right\} \,\middle|\, D=1, Z=z\right) = G\left[G^{-1}\left(\tau, p(\widetilde{z};\theta);\rho\right), p(z;\theta);\rho\right].$$

- Given a consistent estimate $\widehat{q}_\tau^d\left(\widetilde{z}\right)$ (and given $\widehat{\theta}$), we minimize the following objective with respect to c:

$$\sum_{i \neq j} \sum_{\ell} D_i \left(\mathbf{1}\left\{y_i \leq \widehat{q}_{\tau_\ell}^d\left(x_i, w_j\right)\right\} - G\left[G^{-1}\left(\tau_\ell, p(x_i, w_j;\widehat{\theta});c\right), p(x_i, w_i;\widehat{\theta});c\right]\right)^2.$$

- q_τ^d may be estimated as a sample quantile (cell-by-cell, as in Chamberlain 93), or using nonparametric quantile regression methods (when covariates are continuous).

Method B: iteration

- Recall that the observed quantiles satisfy:

$$q_\tau^d\left(z\right) = x'\beta_{G^{-1}(\tau, p(z))}.$$

- Given an estimate $\widehat{\rho}$, we can estimate:

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

and update ρ by minimizing:

$$\sum_{i\neq j}\sum_{\ell} D_i\Big(\mathbf{1}\left\{y_i \leq \widetilde{q}_{\tau_\ell}^{d}(x_i,w_j)\right\} - G\left[G^{-1}\left(\tau_\ell, p(x_i,w_j;\widehat{\theta});c\right), p(x_i,w_i;\widehat{\theta});c\right]\Big)^2.$$

- This procedure may be iterated until convergence.

Comments on estimation of ρ

- <u>Method A:</u> the objective function is not continuous, and non-convex.
- May use grid search (low dimensional ρ), or simulation-based optimization.
- <u>Method B:</u> based on the identification argument. Robust to misspecification of $q_\tau(x)$ (when covariates are discrete).
- Fast and straightforward way to obtain good starting values for method A.

5 Properties and extensions

Asymptotic properties (method A)

- We impose standard assumptions on the quantile model for potential outcomes.
- We also impose regularity conditions on the copula $C(u,v;\rho)$.

  $\Rightarrow$ application of GMM with non-smooth moment functions.
- We show that, for any $\tau \in (\varepsilon, 1-\varepsilon)$:

$$\sqrt{N}\begin{pmatrix} \widehat{\beta}_\tau - \beta_\tau \\ \widehat{\rho} - \rho \end{pmatrix} \xrightarrow{d} \mathcal{N}\left[0, V_\tau\right],$$

where a consistent estimator of V_τ can be constructed following Powell (86).

Estimating bounds

- Let $\overline{p}_x = \sup_w p(x,w)$. Quantile linearity implies the following (not necessarily sharp) bounds:

$$\underbrace{x'\beta_{G^{-1}\left(\max\left(\frac{\tau+\overline{p}_x-1}{\overline{p}_x},0\right),\overline{p}_x\right)}}_{\underline{q}_\tau(x)} \leq q_\tau(x) \leq \underbrace{x'\beta_{G^{-1}\left(\min\left(\frac{\tau}{\overline{p}_x},1\right),\overline{p}_x\right)}}_{\overline{q}_\tau(x)}.$$

- Under the assumption that the support of W given $X = x$ is independent of x, $\overline{p}_x$ can be consistently estimated by $\widehat{p}_x = \sup_{i=1}^N p(x, w_i; \widehat{\theta})$.

  $\Rightarrow$ Consistent estimates of $\underline{q}_\tau(x)$ and $\overline{q}_\tau(x)$ are easily obtained.

- We are thus using our model as a flexible specification of the bounds, which are nonparametrically identified.

Unconditional quantiles

- Once $\widehat{\beta}_\tau$ has been computed (for a grid of τ's), the unconditional *cdf* of Y^* (a discretized or simulated version of) may be estimated as in Machado & Mata (05):

$$\widehat{F}_{Y^*}(y) = \frac{1}{N}\sum_{i=1}^N \int_0^1 \mathbf{1}\left\{x_i'\widehat{\beta}_\tau \leq y\right\} d\tau,$$

  and its unconditional quantiles as $\widehat{q}_\tau = \inf\left\{y, \widehat{F}_{Y^*}(y) \geq \tau\right\}$.

- Confidence bands for unconditional effects may be derived using the results in Chernozhukov, Fernandez-Val and Melly (10).

- Note that unconditional quantiles are monotone by construction. If needed, conditional quantiles may be rearranged (Chernozhukov, Fernandez-Val and Galichon 10).

Extensions

- Nonparametric $p(z)$: When Z is continuous, extending our approach requires dividing by $p(z)$. A trimming approach may be used along the lines of Buchinsky & Hahn (98).

- Censoring and endogeneity: Our approach may be easily combined with Powell (86) and Chernozhukov and Hansen (05, 06).

- Nonparametric G: One may want to let the dimension of ρ increase with the sample size (see Chen and Pouzo 10 for general results).

6 Application to wage inequality and gender gaps in the UK

6.1 Data

- FES, 1978-2000, aged 23 to 59. $77,830$ males, $89,848$ females.
- Time, cohort, and regional effects, education and number of kids (26 controls). Sample split by gender and marital status.
- Exclusion restriction: we follow Blundell, Reed and Stoker (03) and use their measure of potential out-of-work income as our excluded regressor w in $z = (w, x)$.
- This measure is constructed for each individual in the sample using the IFS tax and welfare-benefit simulation model.

Figure 3: Descriptive statistics (conditional on employment)

	Mean	Min	Max	q10	q50	q90
			Males			
			Married			
Log-wage	2.10	.172	4.30	1.56	2.06	2.71
Propensity score	.879	.021	1.00	.766	.893	.979
			Single			
Log-wage	1.99	.319	4.28	1.45	1.95	2.58
Propensity score	.753	.259	1.00	.574	.765	.916
			Females			
			Married			
Log-wage	1.64	-.378	3.59	1.11	1.57	2.32
Propensity score	.681	.006	.998	.512	.699	.844
			Single			
Log-wage	1.78	-.465	3.58	1.20	1.76	2.42
Propensity score	.718	.019	1.00	.475	.735	.933

Source: Family Expenditure Survey, 1978-2000.
Note: The propensity score is estimated using a probit model.

6.2 Results

Implementation

- We model the propensity score using a probit.
- In the benchmark results, we use a Frank copula (Frank 79), yielding:

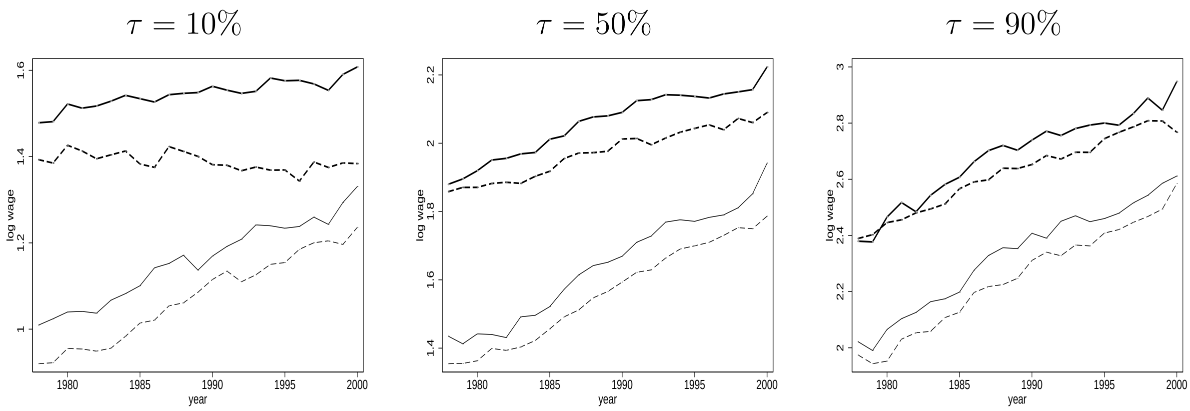
$$G(u, v; \rho) = -\frac{1}{\rho v} \ln \left[1 - \frac{(1 - e^{-\rho u})(1 - e^{-\rho v})}{1 - e^{-\rho}} \right].$$

  We also use an encompassing two-parameter generalized Frank family (called "BB8" in Joe 97).

- To estimate ρ (by gender and marital status), we take the grid:

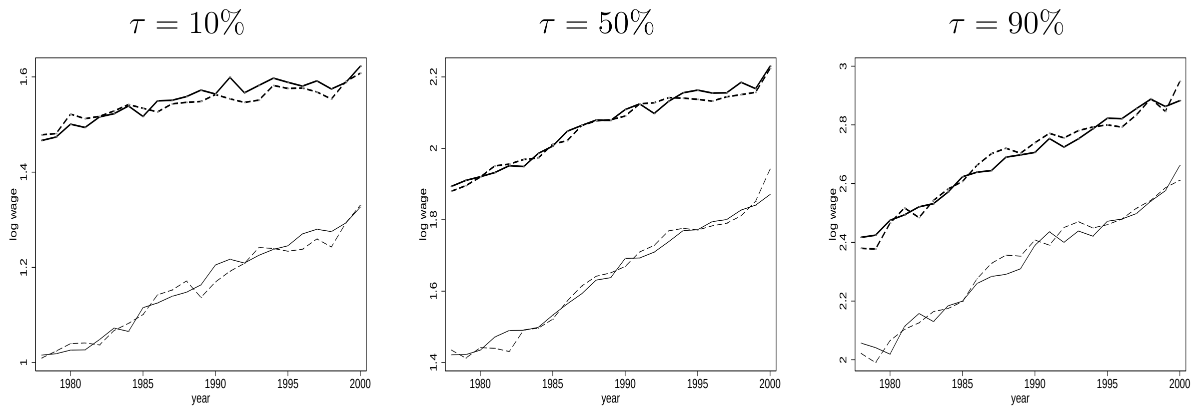
$$\tau \in \left\{ \frac{1}{10}, \frac{2}{10}, ..., \frac{9}{10} \right\}.$$

Figure 4: Wage quantiles (latent and observed), by gender



Percentiles of hourly wages, conditional on employment (solid lines) and corrected for selection (dashed). Male wages (at the top) are plotted in thick lines, while female wages are in thin lines.

Figure 5: Fit to wage quantiles, by gender

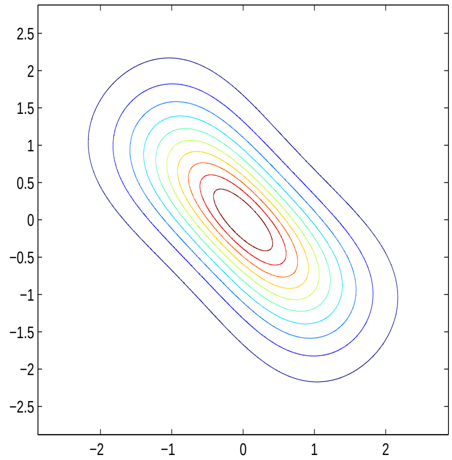


Percentiles of hourly wages, conditional on employment. Data (solid lines) and predicted (dashed). Male wages (at the top) are plotted in thick lines, while female wages are in thin lines.

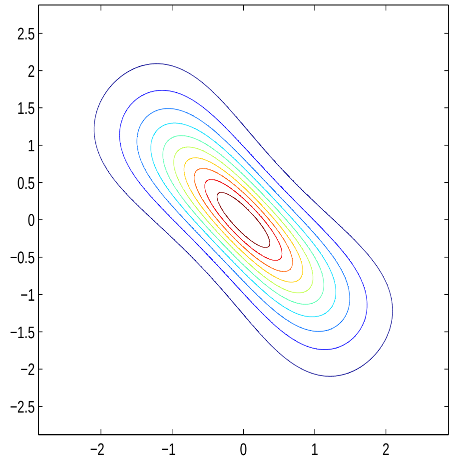
Figure 6: Sample selection: estimated copulas (Frank family), by marital status

Males

Married
(Correlation -.63)

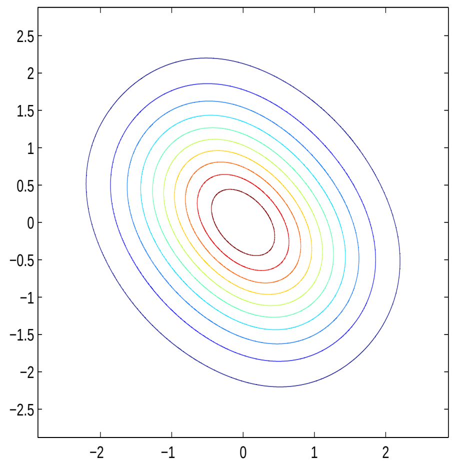


Single
(Correlation -.77)

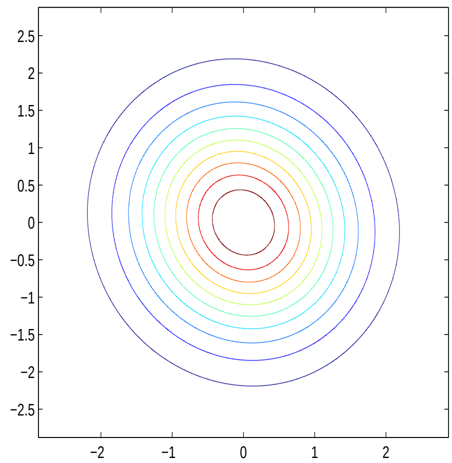


Females

Married
(Correlation -.25)



Single
(Correlation -.06)

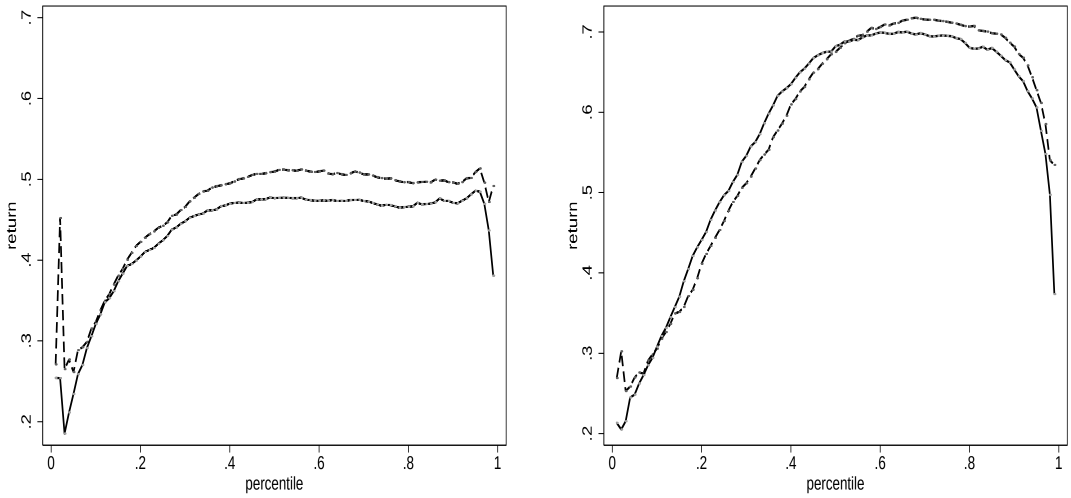


Note: Contour plots of the copula density, Gaussian margins. Negative (rank) correlation indicates positive selection into employment.

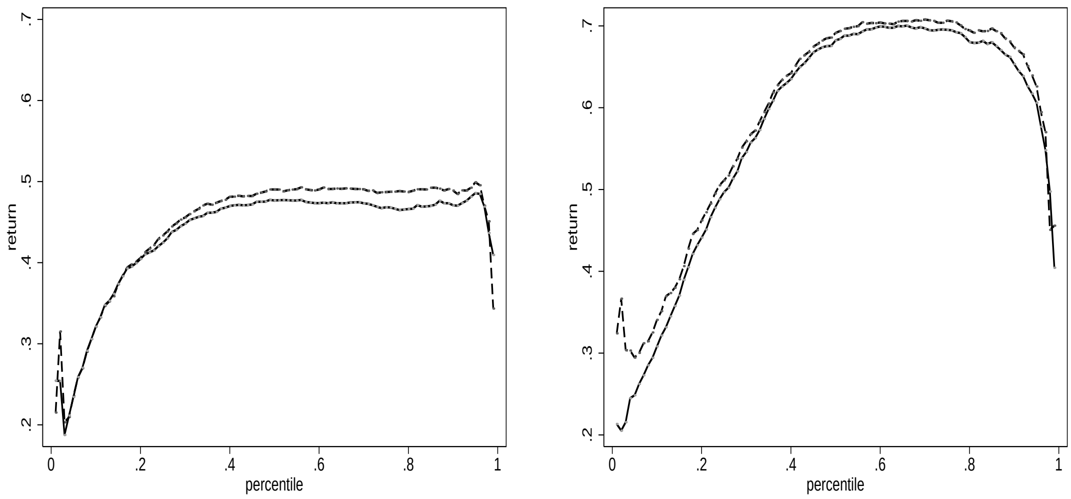
Figure 7: Return to college (married), by gender

Males Females

Quantile selection

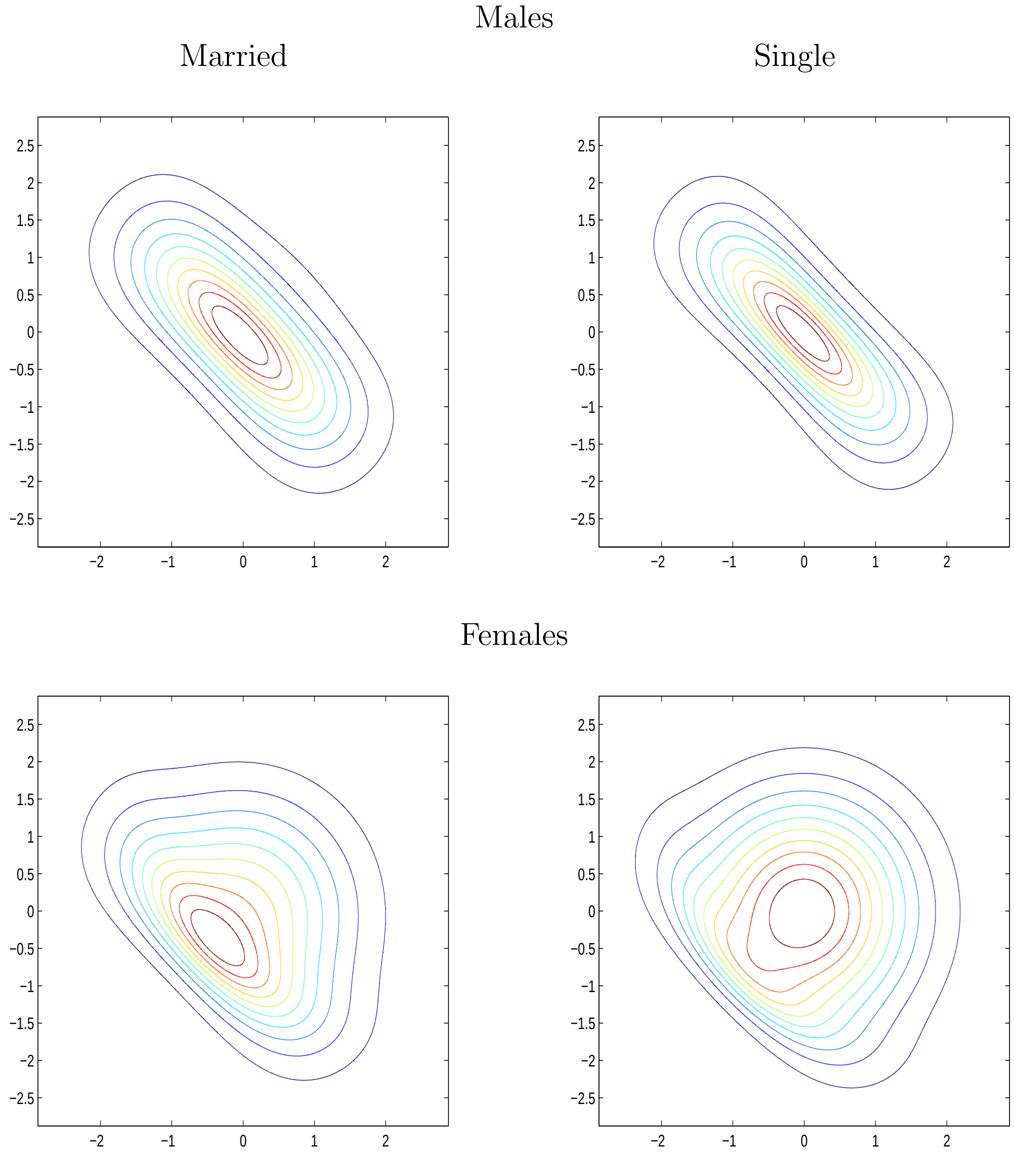


Additive specification



Note: Quantile slope coefficients of college education. Conditional on employment (solid), and corrected for selection (dashed).

Figure 8: Estimated copulas (generalized Frank family), by gender and marital status



Note: Contour plots of the copula density, Gaussian margins.

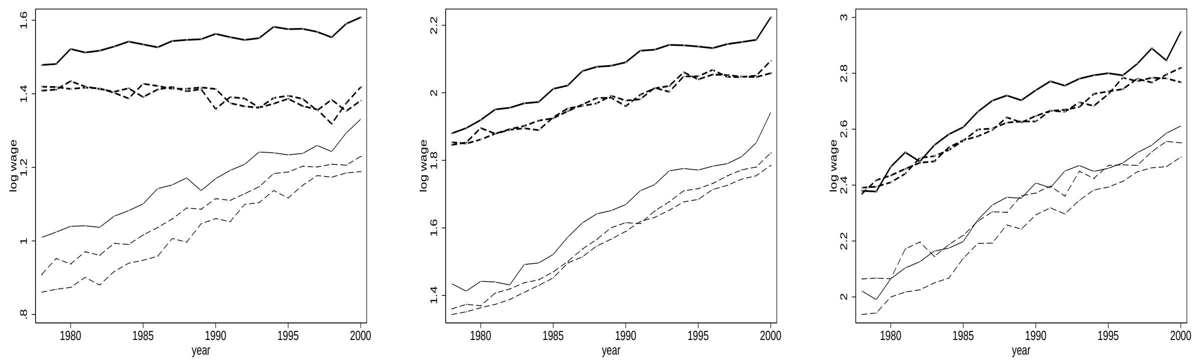
Figure 9: Wage quantiles, generalized Frank family

Percentiles of hourly wages, conditional on employment (solid lines) and corrected for selection (dashed). Male wages (at the top) are plotted in thick lines, while female wages are in thin lines.

Figure 10: Estimated bounds on latent wage quantiles, by gender

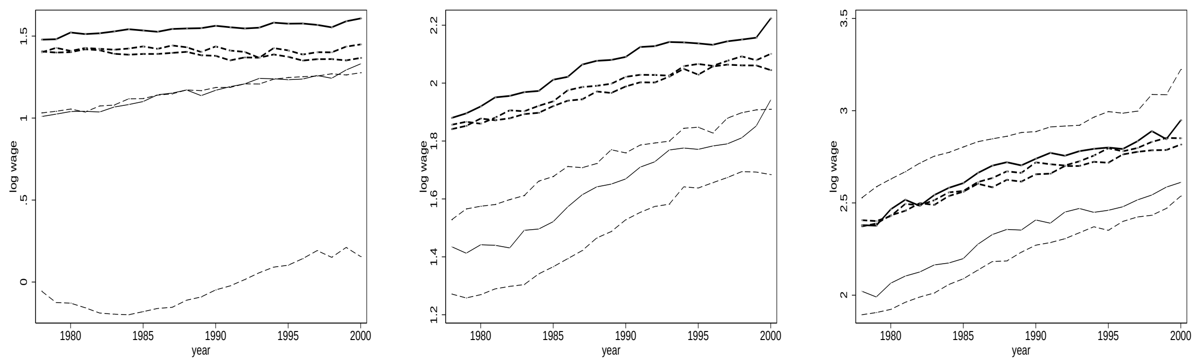
All observations

$\tau = 10\%$ $\tau = 50\%$ $\tau = 90\%$



Trimming 1% of extreme observations in out-of-work income

$\tau = 10\%$ $\tau = 50\%$ $\tau = 90\%$



Estimated bounds on percentiles of hourly wages (dashed). The solid lines show the percentiles conditional on employment. Male wages (at the top) are plotted in thick lines, while female wages are in thin lines.

7 Conclusion

Summary

- Selection-corrected quantile regression ("QR-S"!) is an interesting tool for empirical research, as shown by our application.
- As in the well-known additive model, correcting for sample selection reveals relevant features of the data.
- We have proposed a simple multi-step algorithm, based on a rotated "check" function, to correct QR estimates for selection.
- The approach may be combined with recent advances on quantile regression (such as unconditional quantiles, rearrangement, censoring, or endogeneity).

Future work

- Selection-corrected quantile regression requires 1) parametric specification of the copula, and 2) large support of the propensity score.
- When 2) is satisfied, one could take advantage of the fact that $q_\tau(x)$ is the limit of $q_\tau^d(z)$ as $p(z) \to 1$.
- In practice, $q_\tau^d(z)$ could be estimated using penalized QR methods (e.g., Belloni and Chernozhukov 09). The properties of this "estimator at infinity" are yet to be studied.
- When 2) is not satisfied, one could use a similar approach to flexibly estimate the bounds on $q_\tau(x)$.

References

[1] Ahn, H., and J. L. Powell (1993): "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58, 3–29.

[2] Albrecht, J., A. van Vuuren, and S. Vroman (2009): "Counterfactual Distributions with Sample Selection Adjustments: Econometric Theory and an Application to the Netherlands," *Labour Economics*, 16(4), 383–396.

[3] Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006): "Quantile Regression Under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica*, 74, 539–563.

[4] Blundell, R., H. Reed, and T. M. Stoker (2003): "Interpreting Aggregate Wage Growth: The Role of Labor Market Participation," *American Economic Review*, 93(4), 1114–1131.

[5] Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007): "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds," *Econometrica*, 75, 323–364.

[6] Bonhomme, S., and J. M. Robin (2009): "Assessing the Equalizing Force of Mobility Using Short Panels: France, 1990-2000," *Review of Economic Studies*, 76(1), 63–92.

[7] Buchinsky, M. (1994): "Changes in the U.S. Wage Structure 1963 to 1987; An Application of Quantile Regressions," *Econometrica*, 62, 405–458.

[8] Buchinsky, M. (1998): "The dynamics of changes in the female wage distribution in the USA: a quantile regression approach," *Journal of Applied Econometrics*, 13, 1–30.

[9] Buchinsky, M. (2001): "Quantile Regression with Sample Selection: Estimating Women's Return to Education in the US," *Empirical Economics*, 26, 87–113.

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

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

[12] Chen, X., and D. Pouzo (2008): "Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Moments," Cowles Foundation Discussion Paper 1650.

[13] Chen, X., and D. Pouzo (2009): "Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Residuals," *Journal of Econometrics*, 152, 46–60.

[14] Chernozhukov, V., I. Fernández-Val, and A. Galichon (2009): "Quantile and Probability Curves without Crossing," *mimeo*, MIT.

[15] Chernozhukov, V., and C. Hansen (2005): "An IV model of Quantile Treatment Effects," *Econometrica*, 73, 245–262.

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

[17] Chernozhukov, V., and H. Hong (2003): "An MCMC Approach to Classical Estimation", *Journal of Econometrics*, 115, 293–346.

[18] Das, M., W. K. Newey, and F. Vella (2003): "Nonparametric Estimation of Sample Selection Models," *Review of Economic Studies*, 70, 33–58.

[19] Evans, L. C. (1998): *Partial Differential Equations*, vol. 19 of Graduate Studies in Mathematics, American Mathematical Society.

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

[21] Gosling, A., S. Machin, and C. Meghir (2000): "The Changing Distribution of Male Wages in the UK," *Review of Economic Studies*, 67, 635–666.

[22] Gronau, R. (1974): "Wage Comparison - A Selectivity Bias," *Journal of Political Economy*, 82, 1119–1143.

[23] Heckman, J. J. (1974): "Shadow Prices, Market Wages and Labour Supply," *Econometrica*, 42, 679–694.

[24] Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161.

[25] Heckman, J. J. (1990): “Varieties of Selection Bias,” *The American Economic Review*, 80, 313–318. (Papers and Proceedings of the Hundred and Second Annual Meeting of the American Economic Association.)

[26] Heckman, J., and G. Sedlacek (1985): “Heterogeneity, Aggregation and Market Wage Functions; an Empirical Model of Self Selection into the Labor Market,” *Journal of Political Economy*, 98, 1077–1125.

[27] Imbens, G., and J. Angrist (1994): “Identification and Estimation of Local Average Treatment Effects,” *Econometrica*, 62, 467–476.

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

[29] Judd, K. (1998): *Numerical Methods in Economics*, MIT Press. Cambridge, London.

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

[31] Koenker, R., and G. Bassett (1978): “Regression Quantiles,” *Econometrica*, 46, 33–50.

[32] Koenker, R., and Z. Xiao (2002): “Inference on the quantile regression process,” *Econometrica*, 70, 1583–1612.

[33] Machado, J. A. F., and J. Mata (2005): “Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regression,” *Journal of Applied Econometrics*, 20, 445–465.

[34] Manski, C. (1994): “The Selection Problem” in *Advances in Econometrics*, Sixth World Congress, Vol 1, ed. by C. Sims. Cambridge, U.K.: Cambridge University Press, 143–170.

[35] Melly, B. (2007): “Estimation of Counterfactual Distributions Using Quantile Regression,” *mimeo*.

[36] Neal, D. (2004): “The Measured Black-White Wage Gap among Women Is too Small,” *Journal of Political Economy*, 112, 1–28.

[37] Nelsen, R. B. (1998): *An Introduction to Copulas*. New-York: Springer Verlag.

[38] Newey, W., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," *in* R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics* vol 4: 2111-245. Amsterdan: Elsevier Science.

[39] Olivetti, C., and B. Petrongolo (2008): "Unequal Pay or Unequal Employment? A Cross-Country Analysis of Gender Gaps," *Journal of Labor Economics*, 26(4), 621–654.

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

[41] Vytlacil, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331–341.

APPENDIX

A Asymptotic properties

A.1 Asymptotic properties

To derive the asymptotic distribution of the estimator $\widehat{\beta}_\tau = \widehat{\beta}_\tau(\widehat{\rho})$, where $\widehat{\beta}_\tau(c)$ is given by (**??**), let us first define:

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

We make the following assumptions.

Assumption 1 *The following conditions are satisfied.*

i) There exists a positive definite matrix Σ_τ such that:

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

ii) The cdf 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 $X_i'\beta_\tau$, $i=1,...,N$.

iii) The function G is almost surely differentiable with respect to its second and third components, with derivatives $\partial_p G$ and $\partial_\rho G$, respectively.

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

$$\begin{aligned}
J_\tau &= \underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^N p\left(Z_i;\theta\right)X_iX_i'f_i\left[X_i'\beta_\tau\right], \\
P_{1\tau} &= \underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^N p\left(Z_i;\theta\right)X_i\left(\partial_\theta p\left(Z_i;\theta\right)\right)'\partial_p G\left(\tau,p\left(Z_i;\theta\right)|X_i;\rho\right), \\
P_{2\tau} &= \underset{N\to\infty}{\text{plim}}\frac{1}{N}\sum_{i=1}^N p\left(Z_i;\theta\right)X_i\left(\partial_\rho G\left(\tau,p\left(Z_i;\theta\right)|X_i;\rho\right)\right)'.
\end{aligned}$$

Condition i) requires that $\frac{1}{N}\sum_{i=1}^N X_i g_{i\tau}$, $\widehat{\theta}$, and $\widehat{\rho}$ jointly satisfy a central limit theorem. Under weak regularity conditions, it is easy to show that:

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

where we have denoted:

$$G_{i\tau} \equiv G\left(\tau,p\left(Z_i;\theta\right)|X_i;\rho\right). \tag{A1}$$

The asymptotic covariance matrix of $\widehat{\theta}$ will typically be straightforward to compute, e.g. if $\widehat{\theta}$ is a maximum likelihood estimator. The variance of $\widehat{\rho}$ and the covariance terms are harder to calculate. Below we will compute the joint asymptotic distribution of $\widehat{\beta}_\tau$ and $\widehat{\rho}$, when $\widehat{\rho}$ is the GMM estimator (**??**).

Condition ii) is standard in ordinary quantile regression (e.g., Theorem 4.2 in Koenker and Bassett, 1978). The only difference here is that we work with the *cdf* of Y given Z, and not given X. Condition iii) requires that the copula be differentiable. Most of the usual parametric families of copulas are differentiable in both their arguments. An exception are piecewise-constant empirical copulas, which are not continuous. Lastly, Condition iv) imposes the existence of moments.

Theorem A1 *Let $\tau \in]0,1[$, and let Assumptions* **??** *and 1 hold. Then, as N tends to infinity:*

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

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

Theorem A1 provides the asymptotic distribution of quantile estimates, corrected for the fact that $\widehat{\theta}$ and $\widehat{\rho}$ have been estimated.[1] Remark that, in the absence of sample selection, the formula boils down to the standard expression (Koenker, 2005, p.120).

The theorem can be easily generalized to derive the asymptotic distribution of a finite number of quantiles $\left[\widehat{\beta}_{\tau_1}, ..., \widehat{\beta}_{\tau_L}\right]$. An interesting extension is to derive the large sample theory of the quantile process: $\sqrt{N}\left(\widehat{\beta}_{(.)} - \beta_{(.)}\right)$ in a suitable metric space. This can be done along the lines of Koenker and Xiao (2002) or Chernozhukov and Hansen (2006).

Asymptotic distribution of $\widehat{\rho}$. We now compute the joint asymptotic distribution of $\sqrt{N}\left(\widehat{\beta}_\tau - \beta_\tau\right)$ and $\sqrt{N}\left(\widehat{\rho} - \rho\right)$, in the special case where $\widehat{\rho}$ is given by (**??**). We use a finite counting measure:

$$K(\tau) = \sum_{\ell=1}^{L} \pi_\ell \mathbf{1}\left\{\tau = \tau_\ell\right\},$$

where $\pi_\ell > 0$, $\sum_{\ell=1}^{L} \pi_\ell = 1$, and $\tau_1 < ... < \tau_L$ is a set of elements of $]0,1[$.

Assumption 2 *The following conditions are satisfied.*

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

$$\widehat{\theta} - \theta = -H^{-1}\widehat{\mathbb{E}}\left[s_i\right] + O_p\left(\frac{1}{N}\right).$$

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

$$\begin{aligned}
\widetilde{J}_{\tau_\ell} &= \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} p(Z_i;\theta)\,\varphi(Z_i)\,\varphi(Z_i)'\, f_i\left[X_i'\beta_{\tau_\ell}\right],\\
\widetilde{P}_{1\tau_\ell} &= \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} p(Z_i;\theta)\,\varphi(Z_i)\left(\partial_\theta p(Z_i;\theta)\right)'\partial_p G\left(\tau_\ell, p(Z_i;\theta)\,|X_i;\rho\right),\\
\widetilde{P}_{2\tau_\ell} &= \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} p(Z_i;\theta)\,\varphi(Z_i)\left(\partial_\rho G\left(\tau_\ell, p(Z_i;\theta)\,|X_i;\rho\right)\right)'.
\end{aligned}$$

iii) The following matrix inverse exists:

$$A_\rho = \left[\sum_{\ell=1}^{L}\pi_\ell\left(\widetilde{P}_{2\tau_\ell} - \widetilde{J}_{\tau_\ell} J_{\tau_\ell}^{-1} P_{2\tau_\ell}\right)\right]^{-1}. \tag{A2}$$

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

We now state the main theorem.

[1] Proofs of the results in this section are given in Appendix D.

Theorem A2 *Let Assumptions* **??***, 1, and 2 hold. Then:*

$$\sqrt{N}\begin{pmatrix} \widehat{\beta}_\tau - \beta_\tau \\ \widehat{\rho} - \rho \end{pmatrix} \xrightarrow{d} \mathcal{N}\left[0, V_\tau\right],$$

where the expression of V_τ is given in Appendix D.

Estimating the asymptotic variance. To construct empirical counterparts of the asymptotic variance appearing in Theorem A1, 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:

$$\widehat{J}_\tau = \frac{1}{2Nh_N}\sum_{i=1}^{N}\mathbf{1}\left\{\left|\widehat{\varepsilon}_i\left(\tau\right)\right| \leq h_N\right\} p\left(Z_i; \widehat{\theta}\right) X_i X_i',$$

where $\widehat{\varepsilon}_i = Y_i - X_i'\widehat{\beta}_\tau$, and where 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 V_τ that appears in Theorem A2.

B Proofs

Proof of Proposition ??. We have:

$$\begin{aligned} P\left(q_\tau\left(x\right), w, x\right) &= \Pr\left(Y^* \leq q_\tau\left(x\right), D = 1 \mid W = w, X = x\right) \\ &= \Pr\left(Q_U\left(Y^* \mid X = x\right) \leq q_\tau\left(x\right), V \leq p\left(w, x\right) \mid W = w, X = x\right) \\ &= \Pr\left(U \leq \tau, V \leq p\left(w, x\right) \mid W = w, X = x\right) \\ &= \Pr\left(U \leq \tau, V \leq p\left(w, x\right)\right), \end{aligned} \tag{B3}$$

where we have used Assumption **??**a to derive the third equality, and Assumption **??**b to derive the fourth equality.

Let us denote as $C\left(\tau, p\right) = \Pr\left(U \leq \tau, V \leq p\right)$ the *cdf* of (U, V). Taking derivatives in (B3) with respect to x and w', respectively, we obtain:

$$\begin{aligned} \frac{\partial q_\tau\left(x\right)}{\partial x}\frac{\partial P\left(q_\tau\left(x\right), w, x\right)}{\partial r} + \frac{\partial P\left(q_\tau\left(x\right), w, x\right)}{\partial x} &= \frac{\partial p\left(w, x\right)}{\partial x}\frac{\partial C\left(\tau, p\left(w, x\right)\right)}{\partial p}, \\ \frac{\partial P\left(q_\tau\left(x\right), w, x\right)}{\partial w'} &= \frac{\partial C\left(\tau, p\left(w, x\right)\right)}{\partial p}\frac{\partial p\left(w, x\right)}{\partial w'}. \end{aligned}$$

Lastly, (**??**) follows from right-multiplying the first equality by $\frac{\partial p(w,x)}{\partial w}$, left-multiplying the second by $\frac{\partial p(w,x)}{\partial x}$, and taking differences.

Proof of Proposition ??. As in the proof of Proposition **??** we have:

$$\Pr\left(Y \leq X'\beta_\tau, D = 1 \mid Z\right) = C\left(\tau, p\left(Z; \theta\right) | X; \rho\right),$$

where C depends on X because Assumption **??**a– instead of **??**b– holds. Hence:

$$\begin{aligned} \Pr\left(Y \leq X'\beta_\tau | Z, D = 1\right) &= \frac{C\left(\tau, p\left(Z; \theta\right) | X; \rho\right)}{p\left(Z; \theta\right)} \\ &= G\left(\tau, p\left(Z; \theta\right) | X; \rho\right), \end{aligned}$$

where, by Assumption **??**b, $p\left(Z; \theta\right) \neq 0$ almost surely.

Proof of Corollary ??. Denote as G^{-1} the function such that:

$$G\left(G^{-1}\left(u,p\right),p\right)=G^{-1}\left(G\left(u,p\right),p\right)=u.$$

Note that, by Assumption **??**b, G does not depend on X.

The τ-quantile of $Y|Z,D=1$ is $X'\beta_{G^{-1}(\tau,p(Z;\theta);\rho)}$. So it will be additive in X and $p\left(Z;\theta\right)$ if and only if all coefficients of β_τ but the constant are independent of τ, or $G^{-1}\left(\tau,p\left(Z;\theta\right);\rho\right)$ is independent of $p\left(Z;\theta\right)$. The latter situation happens only if $C\left(\tau,p;\rho\right)=pH\left(\tau\right)$, for some function H. But because C is a *cdf* with uniform marginals, it follows that $C\left(\tau,1;\rho\right)=\tau$, hence $H\left(\tau\right)=\tau$. Therefore, C is the independent copula and U and V are independent. The "if" part of the proposition is immediate.

Proof of Proposition ??. As in the proof of Proposition **??**.

C Extensions

In this section of the appendix we describe three extensions of our approach.

C.1 Endogeneity

To proceed, let us assume that the latent outcome is given by the following linear quantile model:

$$Y^*=E'\alpha_U+X'\beta_U, \tag{C4}$$

where the percentile level U is independent of X, but may be correlated with the endogenous regressors E. As before, the participation equation is given by (**??**).

We have the following result, which is an immediate corollary of Proposition **??**.

Corollary 1 *Suppose that (U,V) is independent of Z given X. Assume also that $q\left(u,x,e\right)=e'\alpha\left(u\right)+x'\beta\left(u\right)$ is strictly increasing in its first argument. Then, for any $\tau\in[0,1]$:*

$$\mathbb{E}\left[\mathbf{1}\left\{Y\leq E'\alpha_\tau+X'\beta_\tau\right\}-G\left(\tau,p\left(Z;\theta\right)|X;\rho\right)\mid Z,D=1\right]=0. \tag{C5}$$

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 (2006)'s estimator to control for selection.

In the first estimation step, we compute $\widehat{\theta}$. In the second step, we compute $\widehat{\rho}$ as:

$$\begin{aligned}\widehat{\rho} \;=\; \arg\min_c \Bigg\| \int_0^1 \sum_{i=1}^N D_i\varphi_\tau\left(Z_i\right)\Big(\mathbf{1}\left\{Y_i\leq E_i'\widetilde{\alpha}_\tau\left(c\right)+X_i'\widetilde{\beta}_\tau\left(\widetilde{\alpha}_\tau\left(c\right);c\right)\right\}\\ -G\left(\tau,p\left(Z_i;\widehat{\theta}\right)|X_i;c\right)\Big)dK\left(\tau\right)\Bigg\|,\end{aligned}$$

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

$$\begin{aligned}\left(\widetilde{\beta}_\tau\left(\alpha;c\right),\widetilde{\gamma}_\tau\left(\alpha;c\right)\right) \;=\; \arg\min_{(b,g)}\sum_{i=1}^N D_i\Big\{G\left(\tau,\widehat{p}\left(Z_i;\widehat{\theta}\right)|X_i;c\right)\left(Y_i-X_i'b-\mu_\tau\left(Z_i\right)'g\right)^+\\ +\left(1-G\left(\tau,\widehat{p}\left(Z_i;\widehat{\theta}\right)|X_i;c\right)\right)\left(Y_i-X_i'b-\mu_\tau\left(Z_i\right)'g\right)^-\Big\},\end{aligned}$$

and:

$$\widetilde{\alpha}_\tau\left(c\right)=\arg\min_a\left\|\widetilde{\gamma}_\tau\left(a;c\right)\right\|.$$

Lastly, once $\widehat{\rho}$ has been estimated, we compute

$$\widehat{\alpha}_\tau=\widetilde{\alpha}_\tau\left(\widehat{\rho}\right),\quad\text{and}\quad\widehat{\beta}_\tau=\widetilde{\beta}_\tau\left(\widehat{\alpha}_\tau;\widehat{\rho}\right).$$

C.2 Censoring

Suppose that Y^* is censored when $Y^* \leq y_0$, 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, the following identity holds under Assumptions **??**a and **??** (the proof being as in that of Proposition **??**):

$$\Pr\left(Y \leq \max\left\{X'\beta_\tau, y_0\right\}|Z, D = 1\right) = G\left(\tau, p\left(Z;\theta\right)|X;\rho\right). \tag{C6}$$

In particular, this implies that the $G\left(\tau, p\left(Z;\theta\right)|X;\rho\right)$-quantile of observed outcomes coincides with $\max\{X'\beta_\tau, y_0\}$. The quantile regression coefficients can be estimated by the algorithm described in Subsection **??**, when replacing $X_i'b$ and $X_i'\widehat{\beta}_\tau(c)$ by $\max\{X_i'b, y_0\}$ and $\max\left\{X_i'\widehat{\beta}_\tau(c), y_0\right\}$, respectively.

In particular:

$$\begin{aligned}\widehat{\beta}_\tau(c) &= \arg\min_b \sum_{i=1}^N D_i\Big\{G\left(\tau, \widehat{p}\left(Z_i;\widehat{\theta}\right)|X_i;c\right)\left(Y_i - \max\left\{X_i'b, y_0\right\}\right)^+ \\ &\qquad + \left(1 - G\left(\tau, \widehat{p}\left(Z_i;\widehat{\theta}\right)|X_i;c\right)\right)\left(Y_i - \max\left\{X_i'b, y_0\right\}\right)^-\Big\}. \end{aligned}\tag{C7}$$

The optimization problem in (C7) is a selection-corrected version of Powell's (1986) censored quantile estimator.

C.3 An alternative estimator for ρ

Here we introduce an alternative estimator of the copula parameter ρ. We assume that G is strictly increasing with respect to its first argument, and denote as G^{-1} its inverse. The estimator is based on the fact that the τ-quantile of observed outcomes is, by (**??**):

$$X'\beta_{G^{-1}(\tau, p(Z;\theta)|X;\rho)}.$$

When covariates are discrete we follow a standard minimum approach. Let $\widehat{q}_{km}(\tau)$ denote the τ-sample quantile of observed outcomes in the Z_{km}-cell. Motivated by (**??**) we estimate ρ by:

$$\widehat{\rho} = \arg\min_c \int_0^1 \sum_{k=1}^K \sum_{m=1}^M \varphi_{\tau km}\left(\widehat{q}_{km}(\tau) - X_m'\widehat{\beta}_{G^{-1}(\tau, \widehat{p}_{km}|X_m;c)}(c)\right)^2 dK(\tau), \tag{C8}$$

where $\widehat{\beta}_\tau(c)$ is given by (**??**).

When a cell-by-cell approach is not tractable, we may use the following double check function estimator:

$$\begin{aligned}\widehat{\rho} &= \arg\min_c \int_0^1 \sum_{i=1}^N D_i\varphi_\tau(Z_i)\Big\{\tau\left(Y_i - X_i'\widehat{\beta}\left[G^{-1}\left(\tau, \widehat{p}(Z_i)|X_i;c\right);c\right]\right)^+ \\ &\qquad + (1-\tau)\left(Y_i - X_i'\widehat{\beta}\left[G^{-1}\left(\tau, \widehat{p}(Z_i)|X_i;c\right);c\right]\right)^-\Big\}dK(\tau),\end{aligned}$$

where now $\widehat{\beta}_\tau(c)$ is given by (**??**).

D Asymptotic distribution

Proof of Theorem A1. Here we provide a simple sketch of the proof, using standard asymptotic arguments (see e.g. Newey and McFadden, 1994, p.2185).

First note that, by a standard result in ordinary quantile regression, the following approximate moment condition is satisfied (Theorem 3.3. in Koenker and Bassett, 1978):

$$\frac{1}{N}\sum_{i=1}^{N} X_i g_i\left(\widehat{\beta}_\tau, \widehat{\theta}, \widehat{\rho}\right) = O_p\left(\frac{1}{N}\right), \tag{D9}$$

where

$$g_i(b,a,c) \equiv D_i\left(\mathbf{1}\left\{Y_i \leq X_i' b\right\} - G\left(\tau, p\left(Z_i; a\right)|X; c\right)\right).$$

An expansion around the truth yields, evaluating the functions and their derivatives at true values :

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

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 1 parts *ii*), *iii*), and *iv*), and where $\widehat{\mathbb{E}} Z = \frac{1}{N}\sum_{i=1}^{N} Z_i$ denotes a sample mean. Hence, as J_τ is non-singular:

$$\begin{aligned}
\widehat{\beta}_\tau - \beta_\tau &\approx -J_\tau^{-1}\left[\widehat{\mathbb{E}}\left[X_i g_{i\tau}\right] - P_{1\tau}\left(\widehat{\theta} - \theta\right) - P_{2\tau}\left(\widehat{\rho} - \rho\right)\right] \\
&= -J_\tau^{-1} P_\tau \begin{pmatrix} \widehat{\mathbb{E}}\left[X_i g_{i\tau}\right] \\ \widehat{\theta} - \theta \\ \widehat{\rho} - \rho \end{pmatrix}.
\end{aligned} \tag{D10}$$

The result then comes from part *i*) in Assumption 1.

Asymptotic distribution of $\widehat{\rho}$. Define the following matrices:

$$B_\rho = -A_\rho\left[\pi_1 \widetilde{J}_{\tau_1} J_{\tau_1}^{-1}, ..., \pi_L \widetilde{J}_{\tau_L} J_{\tau_L}^{-1}\right], \tag{D11}$$

$$C_\rho = A_\rho\left(\sum_{\ell=1}^{L} \pi_\ell\left[\widetilde{P}_{1\tau_\ell} - \widetilde{J}_{\tau_\ell} J_{\tau_\ell}^{-1} P_{1\tau_\ell}\right] H^{-1}\right), \tag{D12}$$

and, for a given $\tau \in]0,1[$:

$$A_\beta(\tau) = J_\tau^{-1} P_{2\tau} A_\rho, \tag{D13}$$

$$B_\beta(\tau) = J_\tau^{-1} P_{2\tau} B_\rho, \tag{D14}$$

$$C_\beta(\tau) = J_\tau^{-1}\left(P_{2\tau} C_\rho - P_{1\tau} H^{-1}\right). \tag{D15}$$

Lastly, let:

$$\begin{aligned}
\sigma_{i\ell m} &= \min\left\{G_{i\tau_\ell}, G_{i\tau_m}\right\} - G_{i\tau_\ell} G_{i\tau_m}, \\
\sigma_{i\ell}(\tau) &= \min\left\{G_{i\tau_\ell}, G_{i\tau}\right\} - G_{i\tau_\ell} G_{i\tau}, \\
\sigma_i(\tau) &= G_{i\tau}\left(1 - G_{i\tau}\right),
\end{aligned}$$

where $G_{i\tau}$ is given by (A1), and define:

$$\Pi_i(\tau) = \begin{pmatrix} \sum_{\ell,m=1}^{L}\pi_\ell\pi_m\sigma_{i\ell m} & \sum_\ell^L \pi_\ell\sigma_{i\ell}(\tau) & \sum_\ell^L \pi_\ell\sigma_{i\ell 1} & \ldots & \sum_\ell^L \pi_\ell\sigma_{i\ell L} \\ \sum_\ell^L \pi_\ell\sigma_{i\ell}(\tau) & \sigma_i(\tau) & \sigma_{i1}(\tau) & \ldots & \sigma_{iL}(\tau) \\ \sum_\ell^L \pi_\ell\sigma_{i\ell 1} & \sigma_{i1}(\tau) & \sigma_{i11} & \ldots & \sigma_{i1L} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \sum_\ell^L \pi_\ell\sigma_{i\ell L} & \sigma_{iL}(\tau) & \sigma_{iL1} & \ldots & \sigma_{iLL} \end{pmatrix}.$$

The main theorem is as follows.

Theorem D3 *Let Assumptions **??**, 1, and 2 hold. Then:*

$$\sqrt{N}\begin{pmatrix} \widehat{\beta}_\tau - \beta_\tau \\ \widehat{\rho} - \rho \end{pmatrix} \xrightarrow{d} \mathcal{N}\left[0, \Delta_\tau\Omega_\tau\Delta_\tau'\right],$$

where

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

and:

$$\Omega_\tau = \begin{pmatrix} \mathbb{E}\left[\Pi_i(\tau)\odot\left\{p(Z_i;\theta)\begin{pmatrix}\varphi(Z_i) \\ \iota_{L+1}\otimes X_i\end{pmatrix}\begin{pmatrix}\varphi(Z_i) \\ \iota_{L+1}\otimes X_i\end{pmatrix}'\right\}\right] & 0 \\ 0 & \mathbb{E}\left[s_is_i'\right] \end{pmatrix}, \tag{D16}$$

where $\odot$ denotes the element-wise (Hadamard) matrix product, and ι_{L+1} is a $(L+1)\times 1$ vector of ones.

Proof of Theorem D3. As in the proof of Theorem A1, we start with an approximate moment equation:

$$\sum_{\ell=1}^{L}\pi_\ell\widehat{\mathbb{E}}\left[\varphi(Z_i)\,g_i\left(\widehat{\beta}_{\tau_\ell},\widehat{\theta},\widehat{\rho}\right)\right] = O_p\left(\frac{1}{N}\right).$$

Expanding around the truth:

$$\begin{aligned}\sum_{\ell=1}^{L}\pi_\ell\widehat{\mathbb{E}}\left[\varphi(Z_i)\,g_i\left(\widehat{\beta}_{\tau_\ell},\widehat{\theta},\widehat{\rho}\right)\right] \;\approx\; & \sum_{\ell=1}^{L}\pi_\ell\Big\{\widehat{\mathbb{E}}\left[\varphi(Z_i)\,g_{i\tau_\ell}\right] + \widetilde{J}_{\tau_\ell}\left(\widehat{\beta}_{\tau_\ell}-\beta_{\tau_\ell}\right) \\ & -\widetilde{P}_{1\tau_\ell}\left(\widehat{\theta}-\theta\right) - \widetilde{P}_{2\tau_\ell}\left(\widehat{\rho}-\rho\right)\Big\}.\end{aligned}$$

So, by (D10):

$$\begin{aligned} O_p\left(\frac{1}{N}\right) \;\approx\; & \sum_{\ell=1}^{L}\pi_\ell\Big\{\widehat{\mathbb{E}}\left[\varphi(Z_i)\,g_{i\tau_\ell}\right] - \widetilde{P}_{1\tau_\ell}\left(\widehat{\theta}-\theta\right) - \widetilde{P}_{2\tau_\ell}\left(\widehat{\rho}-\rho\right) \\ & -\widetilde{J}_{\tau_\ell}\left(J_{\tau_\ell}^{-1}\left[\widehat{\mathbb{E}}\left[X_ig_{i\tau_\ell}\right] - P_{1\tau_\ell}\left(\widehat{\theta}-\theta\right) - P_{2\tau_\ell}\left(\widehat{\rho}-\rho\right)\right]\right)\Big\}.\end{aligned}$$

So:

$$\begin{aligned}\widehat{\rho}-\rho \;\approx\; & \left[\sum_{\ell=1}^{L}\pi_\ell\left(\widetilde{P}_{2\tau_\ell} - \widetilde{J}_{\tau_\ell}J_{\tau_\ell}^{-1}P_{2\tau_\ell}\right)\right]^{-1}\times \\ & \Big\{\sum_{\ell=1}^{L}\pi_\ell\widehat{\mathbb{E}}\left[\varphi(Z_i)\,g_{i\tau_\ell}\right] - \sum_{\ell=1}^{L}\pi_\ell\widetilde{J}_{\tau_\ell}J_{\tau_\ell}^{-1}\widehat{\mathbb{E}}\left[X_ig_{i\tau_\ell}\right] \\ & + \left(\sum_{\ell=1}^{L}\pi_\ell\left[\widetilde{P}_{1\tau_\ell} - \widetilde{J}_{\tau_\ell}J_{\tau_\ell}^{-1}P_{1\tau_\ell}\right]H^{-1}\right)\widehat{\mathbb{E}}\left[s_i\right]\Big\}.\end{aligned}$$

Hence:

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

where A_{ρ}, B_{ρ}, and C_{ρ} are given by (A2)-(D12), and:

$$\widehat{\mathbb{E}}\left[X_i g_i\right]=\left(\begin{array}{c}\widehat{\mathbb{E}}\left[X_i g_{i\tau_1}\right]\\ \ldots\\ \widehat{\mathbb{E}}\left[X_i g_{i\tau_L}\right]\end{array}\right).$$

Let now $\tau\in]0,1[$. Using (D10):

$$\begin{aligned}\widehat{\beta}_{\tau}-\beta_{\tau} &\approx -J_{\tau}^{-1}\left[\widehat{\mathbb{E}}\left[X_i g_{i\tau}\right]-P_{1\tau}\left(\widehat{\theta}-\theta\right)-P_{2\tau}\left(\widehat{\rho}-\rho\right)\right]\\ &\approx -J_{\tau}^{-1}\Big[\widehat{\mathbb{E}}\left[X_i g_{i\tau}\right]+P_{1\tau}H^{-1}\widehat{\mathbb{E}}\left[s_i\right]\\ &\qquad -P_{2\tau}\left(A_{\rho}\left(\sum_{\ell=1}^{L}\pi_{\ell}\widehat{\mathbb{E}}\left[\varphi\left(Z_i\right)g_{i\tau_{\ell}}\right]\right)+B_{\rho}\widehat{\mathbb{E}}\left[X_i g_i\right]+C_{\rho}\widehat{\mathbb{E}}\left[s_i\right]\right)\Big].\end{aligned}$$

So:

$$\begin{aligned}\widehat{\beta}_{\tau}-\beta_{\tau} &\approx A_{\beta}\left(\tau\right)\left(\sum_{\ell=1}^{L}\pi_{\ell}\widehat{\mathbb{E}}\left[\varphi\left(Z_i\right)g_{i\tau_{\ell}}\right]\right)-J_{\tau}^{-1}\widehat{\mathbb{E}}\left[X_i g_{i\tau}\right]\\ &\qquad +B_{\beta}\left(\tau\right)\widehat{\mathbb{E}}\left[X_i g_i\right]+C_{\beta}\left(\tau\right)\widehat{\mathbb{E}}\left[s_i\right],\end{aligned}$$

where $A_{\beta}\left(\tau\right)$, $B_{\beta}\left(\tau\right)$, and $C_{\beta}\left(\tau\right)$ are given by (D13)-(D15).

Next, denote:

$$\psi_{i\tau}=\left(\begin{array}{c}\varphi\left(Z_i\right)\sum_{\ell=1}^{L}\pi_{\ell}g_{i\tau_{\ell}}\\ X_i g_{i\tau}\\ X_i g_{i\tau_1}\\ \ldots\\ X_i g_{i\tau_L}\\ s_i\end{array}\right).$$

From the above, we have:

$$\sqrt{N}\left(\begin{array}{c}\widehat{\beta}_{\tau}-\beta_{\tau}\\ \widehat{\rho}-\rho\end{array}\right)\xrightarrow{d}\mathcal{N}\left[0,V_{\tau}\right],$$

with:

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

Finally, we check that $\mathbb{E}\left(\psi_{i\tau}\psi_{i\tau}'\right) = \Omega_\tau$ given by (D16).

$$\begin{aligned}
\mathbb{E}\left(\psi_{i1}\psi_{i1}'\right) &\equiv \mathbb{E}\left[\left(\varphi\left(Z_i\right)\sum_{\ell=1}^{L}\pi_\ell g_{i\tau_\ell}\right)\left(\varphi\left(Z_i\right)\sum_{m=1}^{L}\pi_m g_{i\tau_m}\right)'\right] \\
&= \mathbb{E}\left[\varphi\left(Z_i\right)\varphi\left(Z_i\right)'\sum_{\ell,m=1}^{L}\pi_\ell\pi_m\mathbb{E}\left(g_{i\tau_\ell}g_{i\tau_m}'|Z_i\right)\right] \\
&= \sum_{\ell,m=1}^{L}\pi_\ell\pi_m\mathbb{E}\left[\left(\min\left\{G_{i\tau_\ell}, G_{i\tau_m}\right\} - G_{i\tau_\ell}G_{i\tau_m}\right)p\left(Z_i;\theta\right)\varphi\left(Z_i\right)\varphi\left(Z_i\right)'\right] \\
&= \sum_{\ell,m=1}^{L}\pi_\ell\pi_m\mathbb{E}\left[\sigma_{i\ell m}p\left(Z_i;\theta\right)\varphi\left(Z_i\right)\varphi\left(Z_i\right)'\right],
\end{aligned}$$

and similarly:

$$\begin{aligned}
\mathbb{E}\left[\left(\varphi\left(Z_i\right)\sum_{\ell=1}^{L}\pi_\ell g_{i\tau_\ell}\right)\left(X_i g_{i\tau_m}\right)'\right] &= \sum_{\ell}^{L}\pi_\ell\mathbb{E}\left[\sigma_{i\ell m}p\left(Z_i\right)\varphi\left(Z_i\right)X_i'\right], \\
\mathbb{E}\left[\left(X_i g_{i\tau_\ell}\right)\left(X_i g_{i\tau_m}\right)'\right] &= \mathbb{E}\left[\sigma_{i\ell m}p\left(Z_i\right)X_iX_i'\right],
\end{aligned}$$

and, as $s_i \equiv s\left(D_i, Z_i\right)$: $\mathbb{E}\left[g_{i\tau_\ell}s_i'\right] = 0$.

Hence the result.

E 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 $\gamma \geq 1$ and $\theta \in]0,1]$, and is given by:

$$C\left(u,v;\gamma,\theta\right) = \frac{1}{\delta}\left[1 - \left\{1 - \frac{1}{\gamma}\left[1 - \left(1-\delta u\right)^\theta\right]\left[1 - \left(1-\delta v\right)^\theta\right]\right\}^{\frac{1}{\theta}}\right], \tag{E17}$$

where $\delta = 1 - \left(1-\gamma\right)^{\frac{1}{\theta}}$. Joe (1997) refers to (E17) as the "BB8" copula.

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

$$C_1 \prec C_2 \quad \text{if and only if} \quad C_1\left(u,v\right) \leq C_2\left(u,v\right), \quad \forall (u,v).$$

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. Importantly for interpretation, the concordance of the generalized Frank copula given by (E17) 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

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

where:

$$C_F\left(u,v;\gamma\right) = \frac{1}{\ln\left(1-\gamma\right)}\ln\left[1 - \frac{1}{\gamma}\left\{1 - \exp\left[\ln\left(1-\gamma\right)u\right]\right\}\left\{1 - \exp\left[\ln\left(1-\gamma\right)v\right]\right\}\right]. \tag{E18}$$

C_F given by (E18) 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 $\left(\frac{1}{2}, \frac{1}{2}\right)$ in the (U, V) space. In comparison, the generalized Frank copula (E17) 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_F.[2] 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.

Figure **??** pictures various shapes of copula densities that can be obtained using a Frank or a generalized Frank copula. On the figure, the marginal distributions are standard normal, and contour plots are shown. The first two rows picture the Frank and Gaussian copula densities for the same measure of rank (or "Spearman") correlation. We see that the two densities are symmetric with respect to the origin, and their shapes are quite similar although the Gaussian is somewhat "rounder".

The next two rows show the density of the generalized Frank copula, conveniently rotated in the plane. We see that the generalized Frank can capture some interesting asymmetries. For example, in the third row of the figure, there is more dependence on the upper-right quarter of the graph than on the lower-left quarter. In addition, asymmetries tend to disapppear as θ tends to infinity, in which case the generalized Frank tends to the original Frank copula.

[2] This is because:

$$\begin{aligned}
\Pr(1-U \leq u, V \leq v) &= \Pr(V \leq v) - \Pr(1-U > u, V \leq v) \\
&= v - \Pr(U < 1-u, V \leq v) \\
&= v - C(1-u, v; \gamma, \theta).
\end{aligned}$$