Quantile selection models: with an application to understanding changes in wage inequality

Manuel Arellano
Stephane Bonhomme

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP75/15

Quantile Selection Models

With an Application to Understanding Changes in Wage Inequality*

Manuel Arellano
CEMFI, Madrid

Stéphane Bonhomme
University of Chicago

December 2015

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 by minimizing a method-of-moments criterion. Given these parameter estimates, the percentile levels of the outcome are re-adjusted 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 also extend the method to account for the presence of equilibrium effects when performing counterfactual exercises.

JEL code: C13, J31.
Keywords: Quantiles, sample selection, copula, gender wage gap.

*This paper was the basis for Arellano's Walras-Bowley lecture given at the North American Summer Meeting of the Econometric Society in 2011. We thank Xiaohong Chen, David Cox, Ivan Fernández-Val, Toru Kitagawa, Roger Koenker, José Machado, Costas Meghir, and seminar audiences at various venues for comments. We also thank Costas for sharing the data.

1 Introduction

Non-random sample selection is a major issue in empirical work. Most selection-correction approaches focus on estimating conditional mean models.[1] In many applications, however, a flexible specification of the entire distribution of outcomes is of interest. In this paper we propose a selection correction method for quantile models.

Quantile regression is widely used to estimate conditional distributions. In a linear quantile model, each percentile is associated with a percentile-specific parameter. Conveniently, quantile parameters can be estimated by minimizing a convex ("check") function (Koenker and Bassett, 1978). Quantile regression has proved to be a valuable tool to analyze changes in distributions. However, to our knowledge there is yet no widely accepted quantile regression approach in the presence of sample selection.

A classic example where sample selection features prominently is the study of wages and employment (Gronau, 1974, Heckman, 1974). Only the wages of employed individuals are observed, so conventional measures of wage gaps or wage inequality may be biased. For example, in our empirical application we study the evolution of wage inequality and employment in the UK. Over the past three decades wage inequality has sharply increased. This change in the wage distribution, similar to the one experienced in the US, has motivated a large literature.[2] At the same time, employment rates have also varied during the period, especially for males. In this context, our method to correct for selection allows documenting the evolution of distributions of latent wages, and to separate them from changes in employment composition.

In regression models, correcting for sample selection involves adding a selection factor as a control. In quantile regression models, we show that selection-corrected estimates can be obtained by suitably shifting the percentile levels as a function of the amount of selection. In practice, this amounts to rotating the "check" function that is optimized in standard quantile regression. This preserves the linear programming structure, and thus the computational simplicity, of quantile regression methods.

In our quantile model, sample selection is modelled via the bivariate cumulative distri-

[1] A very short list of references includes the parametric approaches of Heckman (1979) and Heckman and Sedlacek (1985), and the semi/nonparametric approaches of Heckman (1990), Ahn and Powell (1993), and Das, Newey, and Vella (2003).

[2] Gosling, Machin, and Meghir (2000) use quantile regression to study the evolution of wage inequality in the UK. Some studies for the US are Autor, Katz and Kearney (2005), Chamberlain (1993), Buchinsky (1994), and Angrist, Chernozhukov and Fernández-Val (2006).

bution function, or *copula*, of the errors in the outcome and the selection equation. Our identification analysis covers the case where the copula is left unrestricted. However, in practice, one may wish to let the copula depend on a low-dimensional vector of parameters.[3] As in linear sample selection models, excluded variables (e.g., determinants of employment that do not affect wages directly) are key to achieve credible identification. We show how to estimate the parameters of the copula by minimizing a method-of-moments criterion that exploits variation in excluded regressors.

Our estimation algorithm consists of three steps: estimation of the propensity score of participation, the copula parameter, and the quantile parameters, in turn. We derive the asymptotic distribution of the estimator. We also analyze a number of extensions of the method. In particular, we propose a bounds method to assess the influence on the quantile estimates of the parametric restrictions imposed on the copula.

We apply the method to study the evolution of wage inequality in the UK in the last quarter of the twentieth century, and find that correcting for selection into employment strongly affects male wages at the bottom of the distribution. This is consistent with low-skilled males being progressively driven out of the labor market. Sample selection has smaller effects for females. As a result, correcting for sample selection accentuates the decrease in the gender wage gap at the bottom (though not at the top) of the distribution. We also perform several robustness checks, in particular regarding the specification of the copula. Lastly, we propose a method to account for the presence of equilibrium effects and apply it to a counterfactual exercise.

Literature and outline. Our approach connects with two complementary approaches that have been used to deal with sample selection: bounds methods (Manski, 1994, Blundell, Gosling, Ichimura, and Meghir, 2007), and parametric and semiparametric versions of the Heckman (1979) sample selection model. It allows one to perform distributional decomposition exercises (e.g., DiNardo, Fortin and Lemieux, 1996, Firpo, Fortin and Lemieux, 2011) while accounting for sample selection.

The paper also connects to the large literature on quantiles, distributions, and treatment effects. Chernozhukov and Hansen (2005, 2006) develop an instrumental variables quantile

[3]Copulas have been extensively used in statistics and financial econometrics (e.g., Joe, 1997, and Nelsen, 1999). Single-parameter copula families have been shown to yield satisfactory fit to empirical data in various contexts. For example, Bonhomme and Robin (2009) use a Plackett copula to model year-to-year earnings mobility.

regression approach. Unlike in this paper, they need to observe outcomes for the treated and non-treated, and rely on a rank invariance or similarity assumption. See also Torgovitsky (2015) and D'Haultfoeuille and Février (2015) for models with continuous endogenous regressors. Imbens and Rubin (1997) study identification and estimation of unconditional distributions of potential outcomes in a treatment effects model with a binary instrument, and achieve identification for compliers (as in Abadie, 2003, and in Abadie, Angrist and Imbens, 2002). Carneiro and Lee (2009) uses the framework of Heckman and Vytlacil (2005) to identify and estimate distributions of potential outcomes.

The literature on quantile selection models, in contrast, is scarce. Buchinsky (1998, 2001) proposes an additive approach to correct for sample selection in quantile regression. Huber and Melly (2015) consider a more general, non-additive quantile model, as we do. They focus on testing for additivity. In contrast, our focus is on providing a practical estimation method. Also related are Neal (2004), who develops imputation methods to correct the black/white wage gap among women, Olivetti and Petrongolo (2008), who apply similar methods to the gender wage gap, and Picchio and Mussida (2010), who propose a parametric model to correct the gender wage gap for selection into employment.

The rest of the paper is as follows. In Section 2 we present the quantile selection model and discuss identification. In Section 3 we construct the estimator and derive its asymptotic properties. The empirical analysis is shown in Section 4. Lastly, Section 5 concludes.

2 Model and identification

2.1 Model and assumptions

We consider the following sample selection model:

$$Y^* = q(U, X), \tag{1}$$

$$D = \mathbf{1}\{V \leq p(Z)\}, \tag{2}$$

$$Y = Y^* \quad \text{if} \quad D = 1, \tag{3}$$

where Y^* is the latent outcome (e.g., market wage), D is the participation indicator (employment), and $Z = (B, X)$ strictly contains X, so B are the excluded covariates. Potential outcomes $Y^* = Y$ are observed only when $D = 1$ (e.g., if the individual is a labor market participant).

We make four assumptions.

Assumption 1

A1 *(exclusion restriction)* (U, V) *is jointly statistically independent of* Z *given* X.

A2 *(unobservables) The bivariate distribution of* (U, V) *given* $X = x$ *has uniform marginals. We denote its cumulative distribution function (cdf) as* $C_x(u, v)$.

A3 *(continuous outcomes) The conditional cdf* $F_{Y^*|X}(y|x)$ *is strictly increasing in* y*. In addition,* $C_x(u, v)$ *is strictly increasing in* u.

A4 *(propensity score)* $p(Z) \equiv \Pr(D = 1|Z) > 0$ *with probability one.*

Assumption A1 is satisfied if $Z = (B, X)$ strictly contains X, and (U, V) is jointly independent of B given X. In the example of wages and employment, B may measure opportunity costs of participation in the labor market. Following Blundell *et al.* (2003), our empirical application will use a measure of potential out-of-work welfare income as exclusion restriction.

Model (1)-(3) depends on two sources of unobserved heterogeneity: the outcome rank U and the percentile rank V. In Assumption A2 we normalize their marginal distributions to be uniform on the unit interval, independent of Z. In particular, $\tau \mapsto q(\tau, x)$ is the conditional quantile function of Y^* given $X = x$. A special case is the linear quantile model $Y^* = X'\beta_U$, which is widely used in applied work since Koenker and Bassett (1978). The Skorohod representation (1) is without loss of generality.[4]

Joint independence between (U, V) and Z given X, as stated in Assumption A2, is stronger than marginal independence. This requires the conditional cdf (that is, the *copula*) of the pair (U, V) given (B, X) to solely depend on X. The presence of dependence between U and V is the source of sample selection bias.

Lastly, A3 restricts the analysis to absolutely continuous outcomes, and A4 is a support assumption on the propensity score often made in sample selection models.

Examples. Before discussing identification of model (1)-(3) we briefly outline two special cases. In Appendix F we also describe an extension to a treatment effects setup with selection on unobservables.

A first special case is obtained when outcomes are additive in unobservables: $Y^* = g(X) + \varepsilon$, where (ε, V) is independent of Z. Note that Assumption A1 is satisfied, with

[4]Indeed, $U = F_{Y^*|X}(Y^*|X)$, where $F_{Y^*|X}$ is the conditional cdf of Y^* given X. Moreover, U being independent of Z given X is equivalent to potential outcomes Y^* being independent of Z given X.

$U = F_\varepsilon(\varepsilon)$, for F_ε the cdf of ε. Moreover, the following restrictions hold (as in Das *et al.*, 2003):

$$\mathbb{E}(Y \mid D = 1, Z) = g(X) + \mathbb{E}(\varepsilon \mid Z, V \leq p(Z)) = g(X) + \lambda(p(Z)),$$

where $\lambda(p) \equiv \mathbb{E}(\varepsilon \mid V \leq p)$.

As a second special case suppose the following reservation rule:

$$D = \mathbf{1}\{Y^* \geq R(Z) + \eta\}, \tag{4}$$

where (Y^*, η) is statistically independent of Z given X. In a labor market application, (4) may represent the participation decision of an individual, who compares her potential wage Y^* with a reservation wage $R(Z) + \eta$. Note that (4) may equivalently be written as:

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

where $V \equiv F_{\eta-Y^*|Z}(\eta - Y^*|Z)$ is uniformly distributed on the unit interval, and independent of Z. Letting $Y^* = q(U, X)$, (U, V) is independent of Z given X, so Assumption A1 is satisfied. At the same time, however, U and V are not jointly independent of X. Thus, in this reservation value model the copula $C_x(\cdot, \cdot)$ depends on x in general.

2.2 Main restrictions and identification

We have, conditional on participation and for all $\tau \in (0, 1)$:

$$\begin{aligned} \Pr(Y^* \leq q(\tau, x) \mid D = 1, Z = z) &= \Pr(U \leq \tau \mid V \leq p(z), Z = z), \\ &= G_x(\tau, p(z)), \end{aligned} \tag{5}$$

where $G_x(\tau, p) \equiv C_x(\tau, p)/p$, and we have used Assumptions A1 to A4. The conditional copula $G_x(\cdot, \cdot)$ measures the dependence between U and V, which is the source of sample selection bias. As a special case, if U and V are conditionally independent given $X = x$ then $G_x(\tau, p(z)) = \tau$. More generally, (5) shows that G_x maps ranks τ in the distribution of latent outcomes (given $X = x$) to ranks $G_x(\tau, p(z))$ in the distribution of observed outcomes conditional on participation (given $Z = z$).

An implication of (5) is that, for each $\tau \in (0, 1)$, the conditional τ-quantile of Y^* coincides with the conditional $G_x(\tau, p(z))$-quantile of Y given $D = 1$. Hence, if we knew the mapping G_x from latent to observed ranks, one could recover $q(\tau, x)$ as a quantile of observed outcomes, at a suitably shifted percentile rank. Figure 1 illustrates the effect of selection on latent and observed wage quantiles. The shaded areas under the dashed density correspond to latent outcomes, and have probability mass 10% each. In contrast, the masses under the solid density, which corresponds to observed selected outcomes, generally differ from 10%.

Equation (5) is instrumental to correct quantile functions from selection. Given knowledge of the mapping G_x, latent quantiles can readily be recovered. Moreover, the exclusion restriction provides information about G_x. The intuition for this is that (5) holds for all z in the support of Z given $X = x$, thus generating restrictions on G_x.

The following result spells out the restrictions on the conditional copula G_x. We denote as $\mathcal{X}$ the support of X, and as $\mathcal{Z}_x$ the support of Z given $X = x$. G_x^{-1} and $F_{Y|D=1,Z}^{-1}$ denote the inverses of G_x and $F_{Y|D=1,Z}$ with respect to their first arguments, which exist by Assumption A3. Proofs are given in Appendix A.

Lemma 1 *Let $x \in \mathcal{X}$. Then, under Assumptions A1 to A4:*

$$F_{Y|D=1,Z}\left(F_{Y|D=1,Z}^{-1}\left(\tau\middle|z_2\right)\middle|z_1\right) = G_x\left(G_x^{-1}\left(\tau, p(z_2)\right), p(z_1)\right), \quad \text{for all } (z_1, z_2) \in \mathcal{Z}_x \times \mathcal{Z}_x. \tag{6}$$

Moreover, for any G_x satisfying (6), one can find a distribution of latent outcomes $F_{Y^*|X}$ such that $G_x\left(F_{Y^*|X}(y|x), p(z)\right) = F_{Y|D=1,Z}(y|z)$ for all (z, y) in the support of (Z, Y) given $X = x$.

Note that the restrictions in (6) are uninformative in the absence of an exclusion restriction. They become informative as soon as the conditional support of Z given $X = x$ contains two or more values. Moreover, the second part of Lemma 1 shows that these are the only restrictions on G_x, in the sense that, for any G_x satisfying (6), one can find a distribution of latent outcomes that rationalizes the data.

Nonparametric point-identification. Two simple conditions lead to nonparametric point identification of G_x, and hence to point-identification of $q(\cdot, x)$ as well. We denote as $\mathcal{P}_x$ the conditional support of the propensity score $p(Z)$ given $X = x$.

Proposition 1 *Let Assumptions A1 to A4 hold. Let $x \in \mathcal{X}$. Suppose that one of the two following conditions holds:*

i) (identification at infinity) There exists some $z_x \in \mathcal{Z}_x$ such that $p(z_x) = 1$.

ii) (analytic extrapolation) $\mathcal{P}_x$ contains an open interval and, for all $\tau \in (0,1)$, the function $p \mapsto G_x(\tau, p)$ is real analytic on $\mathcal{P}_x$.

Then the functions $(\tau, p) \mapsto G_x(\tau, p)$ and $\tau \mapsto q(\tau, x)$ are nonparametrically identified.

Both conditions in Proposition 1 allow to point-identify the dependence mapping G_x and the quantile function $q(\cdot, x)$ using an extrapolation strategy. Under *i*), identification is achieved at the boundary of the support of the propensity score ("at infinity"). Under *ii*), extrapolation is based on the property that real analytic functions that coincide on an open neighborhood coincide everywhere. Absent conditions *i*) and *ii*) of Proposition 1, the model is nonparametrically partially identified in general.

Partial identification Let $x \in \mathcal{X}$ and $\widetilde{z} \in \mathcal{Z}_x$. Using the worst-case Fréchet bounds (e.g., Heckman, Smith and Clements, 1997) on the copula C_x we can bound:

$$\max\left(\frac{\tau + p(\widetilde{z}) - 1}{p(\widetilde{z})}, 0\right) \leq G_x\left(\tau, p(\widetilde{z})\right) \leq \min\left(\frac{\tau}{p(\widetilde{z})}, 1\right), \quad \text{for all } \tau \in (0,1). \tag{7}$$

Let now $z \in \mathcal{Z}_x$. Evaluating (6) at $(z_1, z_2) = (z, \widetilde{z})$, and using (7) to bound $G_x\left(\tau, p(\widetilde{z})\right)$, we obtain the following bounds on $G_x\left(\tau, p(z)\right)$:

$$G_x\left(\tau, p(z)\right) \leq \inf_{\widetilde{z} \in \mathcal{Z}_x} F_{Y|D=1,Z}\left[F^{-1}_{Y|D=1,Z}\left(\min\left(\frac{\tau}{p(\widetilde{z})}, 1\right) \Big| \widetilde{z}\right) \Big| z\right] \tag{8}$$

$$G_x\left(\tau, p(z)\right) \geq \sup_{\widetilde{z} \in \mathcal{Z}_x} F_{Y|D=1,Z}\left[F^{-1}_{Y|D=1,Z}\left(\max\left(\frac{\tau + p(\widetilde{z}) - 1}{p(\widetilde{z})}, 0\right) \Big| \widetilde{z}\right) \Big| z\right]. \tag{9}$$

Moreover, using (5) and (7) we have the following bounds on the quantiles of latent outcomes:

$$q(\tau, x) \leq \inf_{\widetilde{z} \in \mathcal{Z}_x} F^{-1}_{Y|D=1,Z}\left(\min\left(\frac{\tau}{p(\widetilde{z})}, 1\right) \Big| \widetilde{z}\right) \tag{10}$$

$$q(\tau, x) \geq \sup_{\widetilde{z} \in \mathcal{Z}_x} F^{-1}_{Y|D=1,Z}\left(\max\left(\frac{\tau + p(\widetilde{z}) - 1}{p(\widetilde{z})}, 0\right) \Big| \widetilde{z}\right). \tag{11}$$

The quantile bounds in (10) and (11) were first derived by Manski (1994, 2003) in a slightly more general selection model. In related work, Kitagawa (2009) provides a comprehensive study of the role of independence and first-stage monotonicity restrictions in a LATE context. The quantile bounds in (10) and (11) coincide with the choice of the upper or lower Fréchet bounds for the copula of (U, V). In this sense, these are worst-case bounds.[5] In Appendix B we show that these bounds cannot be improved upon.

[5]Note however that the Fréchet copula bounds do not satisfy (6) in general. By (8) and (9), the bounds on G_x are generally tighter than the Fréchet bounds.

3 Estimation

We adopt a flexible semi-parametric specification. Following a large literature on quantile regression, we assume that quantile functions are linear, that is:

$$q\,(\tau, x) = x'\beta_\tau, \qquad \text{for all } \tau \in (0,1) \text{ and } x \in \mathcal{X}. \tag{12}$$

Although our estimation strategy could conceptually be extended to deal with nonlinear specifications, the linear quantile model is convenient for computation.

We assume that the copula function, and hence the function G_x, is indexed by a parameter vector ρ; that is:

$$G_x(\tau, p) \equiv G(\tau, p; \rho) = \frac{C(\tau, p; \rho)}{p}.$$

The statistical literature offers a number of convenient parsimonious specifications, including the Gaussian, Frank, or Gumbel copulas. See Nelsen (1999) and Joe (1997) for comprehensive references. Flexible families may be constructed, for example by relying on the Bernstein family of polynomials (Sancetta and Satchell, 2004). In all these examples, one may let the vector ρ depend on x. For example, for scalar $\rho \in (-1, 1)$ one may specify $\rho(x) = \left(e^{x'\gamma} - 1\right) / \left(e^{x'\gamma} + 1\right)$, where γ is a vector of parameters. For simplicity we omit the dependence of ρ on x in the following.

The parametric assumptions on the copula are substantive. Restricting the analysis to a finite-dimensional ρ allows us to focus on the case where ρ is point-identified and to propose a simple estimation method. In addition, below we propose a bounds approach to assess the influence on quantile estimates of the parametric assumptions made on the copula.

Lastly, the propensity score $p(z; \theta)$ is specified as a known function of a parameter θ. This assumption may be relaxed, at the cost of making the asymptotic analysis more involved; see the extensions at the end of this section.

The functional form of selected quantiles. Before describing the estimator, we first comment on the form of the conditional quantiles given participation, when quantile functions of latent outcomes are linear as in (12). The τ-quantile of outcomes of participants given $z = (b, x)$ is, by (5):

$$q^d\,(\tau, z) \equiv F^{-1}_{Y|Z,D=1}\,(\tau|z) = x'\beta_{G^{-1}(\tau, p(z); \rho)}. \tag{13}$$

Equation (13) makes it clear that sample selection affects all quantiles, and that quantile functions of observed outcomes are generally non-additive in x and $p(z)$. We have the following result, where it is assumed that ρ does not depend on x.

Proposition 2 *Let $\tau \in (0,1)$. Suppose that ρ does not depend on x. Then $z \mapsto q^d(\tau, z)$ is non-additive in x and $p(z)$, unless:*

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

ii) U and V are statistically independent.

Additive specifications such as $q^d(\tau, z) = x'\beta_\tau + \lambda_\tau(p(z))$, for a smooth function $\lambda_\tau(p)$, are sometimes used in applied work (e.g., Buchinsky, 2001, Albrecht *et al.*, 2009). In contrast, in our framework, conditional quantiles of participants are non-additive. Huber and Melly (2015) make a related point in a testing context. Correcting for sample selection thus requires shifting the percentile ranks of individual observations. We now explain how this can be done in estimation.

3.1 Three-step estimation strategy

Let (Y_i, D_i, B_i, X_i), $i = 1, ..., N$, be an i.i.d. sample, with $Z_i \equiv (B_i, X_i)$. We propose to compute selection-corrected quantile regression estimates in three steps. In the first step, we compute $\widehat{\theta}$, a consistent estimate of the propensity score parameter θ. In the second step, we compute a consistent estimator $\widehat{\rho}$ of the copula parameter vector ρ. Lastly, given $\widehat{\theta}$ and $\widehat{\rho}$, for any given $\tau \in (0,1)$ we compute $\widehat{\beta}_\tau$, a consistent estimator of the τth quantile regression coefficient.

The first step can be done using maximum likelihood. We now present the third and second steps in turn.

Step 3: quantile regression. Let us suppose that consistent estimators $\widehat{\theta}$ and $\widehat{\rho}$ are available. Then, for any given $\tau \in (0,1)$ we compute:

$$\widehat{\beta}_\tau = \underset{b \in \mathcal{B}}{\operatorname{argmin}} \sum_{i=1}^{N} D_i \left[\widehat{G}_{\tau i} (Y_i - X_i'b)^+ + \left(1 - \widehat{G}_{\tau i}\right) (Y_i - X_i'b)^- \right], \tag{14}$$

where $\mathcal{B}$ is the parameter space for β_τ, $a^+ = \max(a, 0)$, $a^- = \max(-a, 0)$, and:

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

Solving (14) amounts to minimizing a *rotated* check function. As with standard quantile regression, the optimization problem takes the form of a simple linear program, and can thus be solved in a fast and reliable way. It is instructive to compare the rotated quantile regression estimate $\widehat{\beta}_\tau$ with the following infeasible quantile regression estimate based on the latent outcomes:

$$\widetilde{\beta}_\tau \quad = \quad \underset{b\in\mathcal{B}}{\operatorname{argmin}} \sum_{i=1}^{N} \left[\tau \left(Y_i^* - X_i'b\right)^+ + (1-\tau)\left(Y_i^* - X_i'b\right)^-\right].$$

We see that, in order to correct for selection in (14), τ is replaced by the selection-corrected, individual-specific percentile rank $\widehat{G}_{\tau i}$.

Step 2: copula parameter. From (5), we obtain the following conditional moment restrictions:

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

We propose to estimate the copula parameter ρ as:

$$\widehat{\rho} = \underset{c\in\mathcal{C}}{\operatorname{argmin}} \left\| \sum_{i=1}^{N}\sum_{\ell=1}^{L} D_i \varphi\left(\tau_\ell, Z_i\right) \left[\mathbf{1}\left\{Y_i \le X_i'\widehat{\beta}_{\tau_\ell}(c)\right\} - G\left(\tau_\ell, p(Z_i;\widehat{\theta});c\right)\right]\right\|, \tag{15}$$

where $\tau_1 < \tau_2 < ... < \tau_L$ is a finite grid on $]0,1[$, $\varphi(\tau, Z_i)$ are instrument functions with $\dim\varphi \ge \dim\rho$, and:

$$\begin{aligned}\widehat{\beta}_\tau(c) \quad \equiv \quad \underset{b\in\mathcal{B}}{\operatorname{argmin}} \sum_{i=1}^{N} D_i \Big[& G\left(\tau, p(Z_i;\widehat{\theta});c\right)\left(Y_i - X_i'b\right)^+ \\ & + \left(1 - G\left(\tau, p(Z_i;\widehat{\theta});c\right)\right)\left(Y_i - X_i'b\right)^-\Big]. \end{aligned} \tag{16}$$

Effectively, in this step we are estimating ρ together with $\beta_{\tau_1}, ..., \beta_{\tau_L}$.

This step is computationally more demanding than Step 3. In particular, the objective function in (15) is not continuous, due to the presence of the indicator functions, and generally non-convex. In practice, for low-dimensional ρ one may use grid search, as in our application. For higher-dimensional ρ, simulation-based methods such as simulated annealing (see, e.g., Judd, 1998), or the pseudo-Bayesian approach of Chernozhukov and Hong (2003), could be used. Importantly, evaluating the objective function is usually fast and straightforward. The reason is that (16) is a linear programming problem, for which there exist fast algorithms.[6]

[6] For example, the Matlab version of Morillo, Koenker and Eilers is directly applicable to the problem at hand. Available at: http://www.econ.uiuc.edu/∼roger/research/rq/rq.m

In addition, in experiments we observed that using a large number of percentile values τ_ℓ in (15) tends to smooth the objective function. In Appendix C we consider a model with discrete covariates, and show that in this case an integrated version of the objective function in (15), with a continuum of τ values, is differentiable with respect to the copula parameter c under weak conditions.

Finally, solving (15) is only one possibility to estimate the copula parameter. In Appendix D we describe an alternative estimator of ρ that relies on the copula restrictions (6). The method provides a fast and straightforward way to obtain good starting values to minimize the objective function in (15). Another possibility would be to estimate ρ using a likelihood approach, based on the semi-parametric structure of the model. An interesting question, which we do not address in this paper, would be to construct a semi-parametric efficient estimator for ρ by exploiting the continuum of moment restrictions in (6).

Remark: unconditional quantiles. Once θ and ρ have been estimated, the parameters β_τ are estimated by simple quantile regression using the rescaled percentile levels $\widehat{G}_{\tau i} = G\left(\tau, p\left(Z_i; \widehat{\theta}\right); \widehat{\rho}\right)$ in place of τ. So, the techniques developed in the context of ordinary quantile regression can be used in the presence of sample selection. As an example, counterfactual distributions may be constructed as explained in Machado and Mata (2005) and Chernozhukov, Fernández-Val and Melly (2013). Specifically, the unconditional cdf of Y^* may be estimated as a discretized or simulated version of:

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

and unconditional quantiles can be estimated as $\widehat{q}(\tau) = \inf\left\{y, \widehat{F}_{Y^*}(y) \geq \tau\right\}$. Also, a pervasive problem in quantile regression is that estimated quantile curves may cross each other because of sampling error. The approach proposed by Chernozhukov, Fernández-Val and Galichon (2010), based on quantiles rearrangement, may also be applied in our context.[7]

3.2 Properties and extensions

Asymptotic properties. In Appendix E we derive the asymptotic distributions of $\widehat{\rho}$ and $\widehat{\beta}_\tau$ for given $\tau \in]0,1[$. We show that, under standard conditions for quantile regression

[7] A difference with standard quantile regression concerns inference, as one needs to take into account that ρ and θ have already been estimated when computing asymptotic confidence intervals.

estimators (as in Koenker, 2005), and under suitable differentiability conditions on G, the estimators satisfy:

$$\sqrt{N}\begin{pmatrix} \widehat{\beta}_{\tau}-\beta_{\tau} \\ \widehat{\rho}-\rho \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, V_{\tau}\right), \tag{17}$$

where ρ and β_{τ} denote true parameter values. We provide an explicit expression for the asymptotic variance V_{τ}, and show that it can be estimated using an approach similar to the one in Powell (1986). Alternatively, given the distributional characterization in (17), confidence intervals may be estimated using subsampling (Politis, Romano and Wolf, 1999). We will use subsampling in the empirical application.

Estimating bounds. The above method to estimate the copula parameter ρ relies on the assumption that the copula, and hence the quantile functions, are point-identified. In the absence of functional form assumptions on the copula, both G and $q(\tau, x)$ are partially identified in general. In particular, the quantiles of latent outcomes are bounded by (10) and (11).[8] In practice, a simple way to informally assess the influence of functional form assumptions on the results is to compute estimates of the bounds in (10) and (11), obtained from the semi-parametric model.

Denoting $\overline{p}_x = \sup_b p(x,b)$ the supremum of the support of the excluded variable B for given $X = x$, the model implies the following bounds:[9]

$$\underline{q}(\tau, x) \equiv x'\beta_{G^{-1}\left(\max\left(\frac{\tau+\overline{p}_x-1}{\overline{p}_x},0\right),\overline{p}_x;\rho\right)} \leq q(\tau, x) \leq x'\beta_{G^{-1}\left(\min\left(\frac{\tau}{\overline{p}_x},1\right),\overline{p}_x;\rho\right)} \equiv \overline{q}(\tau, x). \tag{18}$$

Under the assumption that the support of B given $X = x$ is independent of x, $\overline{p}_x$ can be consistently estimated by $\widehat{p}_x = \sup_{i\in\{1,\ldots,N\}} p(x, B_i; \widehat{\theta})$. As these estimates may be sensitive to outliers, in the application we will also consider alternative estimates based on a trimming approach. Consistent estimates of $\underline{q}(\tau, x)$ and $\overline{q}(\tau, x)$ are then obtained by replacing $\overline{p}_x$, β_{τ}, and ρ, by $\widehat{p}_x$, $\widehat{\beta}_{\tau}$, and $\widehat{\rho}$, respectively.

We are thus using our model as a semi-parametric specification for the self-selected conditional quantiles, and therefore for the bounds, which themselves are nonparametrically identified. An alternative strategy, robust to violation of the parametric assumptions on the

[8] Note that (10) and (11) do not impose a linear representation of the quantile functions as in $q(\tau, X) = X'\beta_{\tau}$. Under linearity one could in principle derive tighter bounds.

[9] In fact, one can show that, given that $G(\cdot, \cdot; \rho)$ is a conditional copula, $p \mapsto G^{-1}\left(\min\left(\frac{\tau}{p}, 1\right), p\right)$ is non-increasing, and $p \mapsto G^{-1}\left(\max\left(\frac{\tau+p-1}{p}, 0\right), p\right)$ is non-decreasing, for all $\tau \in (0, 1)$.

copula, would be to construct estimators and confidence sets for the identified sets of the copula and quantile functions. We will return to this possibility in the conclusion.

Extensions. Finally, in Appendix F we outline several extensions of the framework. The first one is to allow for a nonparametric propensity score, instead of a parametric specification. The second one is the construction of a test statistic to test for the absence of sample selection. We also outline how to adapt the method to allow for some regressors to be endogenous (as in Chernozhukov and Hansen, 2005, 2006), and for outcomes to be partially censored (as in Powell, 1986).

4 Wages and labor market participation in the UK

In this section, we apply our method to measure market-level changes in wage inequality in the UK. Moreover, we compare wages of males and females in the UK at different quantiles, correcting for selection into work. Due to changes in employment rates, wage inequality for those at work may provide a distorted picture of market-level inequality. Our exercise decomposes actual changes in the aggregate wage distribution into different interpretable sources (selection and non-selection components). Our procedure could be standardized into building economic statistics, similar to other decomposition-based statistics such as price indices adjusted for changes in quality.

In this application, the latent variable Y^* represents the opportunity cost of working for each person, whether employed or not, *at* given employment rates. It is not a potential outcome in the conventional treatment-effect sense, because Y^* depends on the market price of skill, which may be affected by changes in participation rates. In order to account for equilibrium effects on skill prices we also propose an extension of the method, and we apply it to a counterfactual exercise.

4.1 Data and methodology

We use data from the Family Expenditure Survey (FES) from 1978 to 2000. To construct the sample, we closely follow previous work using these data: Gosling *et al.* (2000) and Blundell *et al.* (2003), who focus on males, and Blundell *et al.* (2007), who consider both males and females. We select individuals aged 23 to 59 who are not in full-time education, and drop observations for which education is not reported, or for which wages are missing but the individual is working. Hourly wages are constructed by dividing usual weekly pre-tax earnings by usual weekly hours worked. In addition, we drop the self-employed from the sample. We end up with $77,630$ observations for males, and $89,848$ observations for females.

During the period of analysis, wage inequality increased sharply in the UK. For example, in our sample, the logarithm of the 90/10 percentile ratio of male hourly wages increased from .90 in 1978 to 1.34 in 2000. This is in line with previous evidence on wage inequality (Gosling *et al.*, 2000). Moreover, a comparison of mean log-wages between males and females shows a mean log-wage gap of .44 in 1978, and a mean gap of .30 in 2000. During the same period the overall employment rate of males fell from 92% to 80%. The mean employment rate of females also changed over the period, though not in a monotone way. This suggests that correcting for selection into employment might be important. We now use our approach to provide selection-corrected measures of wage inequality and gender wage gaps.

We use the quantile selection model to model log-hourly wages Y and employment status D. Our controls X include linear, quadratic, and cubic time trends, four cohort dummies (born in 1919-34, 1935-44, 1955-64, and 1965-77, the baseline category being 1945-54), two education dummies (end of schooling at 17 or 18, and end of schooling after 18), and 11 regional dummies. In addition, we include as regressors the marital status and the number of kids split by age categories (six dummies, from 1 year old to 17-18 years old). Our sample contains 75% of married men and 74% of married women.

We follow Blundell *et al.* (2003) and use their measure of potential out-of-work (welfare) income, interacted with marital status, as our excluded regressor B. This variable is constructed for each individual in the sample using the Institute of Fiscal Studies (IFS) tax and welfare-benefit simulation model. We estimate the propensity score using a probit model. In Table 1 we report several descriptive statistics on the distribution of log-wages, and on the distribution of the estimated propensity score, by gender and marital status. Out-of work income is a strong determinant of labor market participation. For example, in the sample of married (respectively, single) males the likelihood of the probit model of participation increases from $-21,454$ to $-20,438$ (resp., $-10,480$ to $-10,275$) when out-of-work income is added.

The main sources of variation in out-of-work income are the demographic composition of households (age, household size) and the housing costs that households face, as well as changes in policy over time. Our maintained assumption is that those determinants are exogenous to the latent wage equation. Though not uncontroversial,[10] out-of-work income provides a natural choice for an excluded variable in this context. Moreover, variations in out-of-work income over time are partly due to changes in policy, motivating the counterfactual analysis that we will present at the end of this section.

Implementation. We specify the copula $C(.,.;\rho)$ as a member of the one-parameter Frank family (Frank, 1979). We provide details on Frank copulas in Appendix G. We let the copula parameter be gender and marital-status specific. We will return to the choice of the copula below. In addition, to compute $\widehat{\rho}$ in (15) we take $\tau_\ell = \ell/10$ for $\ell = 1, ..., 9$, and $\varphi(\tau_\ell, Z_i) = \varphi(Z_i) = p\left(Z_i; \widehat{\theta}\right)$.[11] Finally, we use grid search for computation of $\widehat{\rho}$, and take 200 grid points.

4.2 Selection-corrected wage distributions

On the nine panels of Figure 2 we plot the evolution of the log-wage deciles for men (thick lines), and women (thin lines). The solid lines show the deciles of observed log-wages, conditional on employment. The dashed lines show the selection-corrected deciles, by gender. To compute the latter, we estimated the selection-corrected quantile regression coefficients using our method, and we then simulated the wage distribution using the method of Machado and Mata (2005), re-adjusting the percentile levels in order to correct for sample selection.

Focusing first on male wages, we see that correcting for sample selection makes a strong difference at the bottom of the wage distribution. For example, at the 10% percentile male wages increased by 10% conditional on employment, while latent wages remained broadly flat. We also see sizable differences between latent and observed wages at the 20% and 30% percentiles. There are smaller differences in the middle and at the top of the distribution. In addition, differences across quantiles illustrate the sharp increase in male wage inequality in the UK over the period.

The results for male wages are consistent with low-skilled individuals being progressively

[10] For example, as argued by Blundell *et al.* (2007), the way the out-of-work income variable operates may imply a positive correlation with potential wages, if individuals who earn more on the labor market have better housing, hence a higher out-of-work income.

[11] When considering a two-parameter copula we take $p\left(Z_i; \widehat{\theta}\right)$ and $p\left(Z_i; \widehat{\theta}\right)^2$ as instrument functions. We also estimated the model with $\varphi(\tau_\ell, Z_i) = \sqrt{\tau_\ell(1-\tau_\ell)}\, p\left(Z_i; \widehat{\theta}\right)$, in order to give more weight to central quantiles, and obtained very similar results. As already mentioned, here we do not attempt to address the question of efficient estimation of ρ.

driven out of the labor market. Our estimated copula has a rank correlation of $-.24$ for married males, and of $-.79$ for singles,[12] which means that individuals with higher wages (higher U) tend to participate more (lower V). Thus, associated with the fall in participation over time, positive selection into employment implies that individuals at the bottom of the latent wage distribution tend to become increasingly non-employed. Selection into employment is stronger for singles than for married males. The 95% confidence intervals for the rank correlation coefficients are $(-.35, -.06)$ for married males, and $(-.84, -.42)$ for singles, respectively.[13]

Looking now at female wages, we observe less difference between wages conditional on employment and latent wages. Indeed, we estimate a copula with rank correlation of $-.17$ for married females, and of $-.08$ for singles, suggesting that there is less positive selection into employment for women than for men. A tentative explanation could be that for females non-economic factors play a bigger role in participation decisions. The confidence intervals for the correlation coefficients are $(-.30, -.01)$ for married females, and $(-.24, .16)$ for singles.

As a result of this evolution, the selection-corrected gender wage gap tends to decrease over time. This is especially true at the bottom of the wage distribution. For example, at the 10% percentile, the difference in log wages between men and women decreases from 43% at the beginning of the period to 28% at the end. A comparable decrease can be seen at the 20% and 30% percentiles. Hence, correcting for sample selection magnifies the reduction in the wage gap in this part of the distribution. However, at the top of the distribution the gap seems to decrease less, from 38% to roughly 32% at the end of the period, and it is virtually unaffected by the selection correction.

Model fit. Figure 3 shows the model fit to the wage percentiles of employed workers. To predict wage percentiles, we simulated wages using our parameter estimates. The results show that the fit to wage quantiles is accurate at the top of the distribution for both genders. At the bottom of the distribution we observe some discrepancies, particularly for females. In addition, we estimated the model allowing the Frank copula parameter to vary with calendar time, on subsamples before and after 1990, in addition to gender and marital status

[12] The rank (or "Spearman") correlation of a copula C is given by: $12 \int_0^1 \int_0^1 uvdC\,(u,v) - 3$.

[13] We computed the confidence intervals using subsampling. Following Chernozhukov and Fernández-Val (2005) we chose the subsample size as a constant plus the square-root of the sample size, where the constant (≈ 1000) was taken to ensure reasonable finite sample performance of the estimator. In our application subsampling is an attractive option given the large sample sizes.

(not reported). We found some evidence of increasingly positive selection into employment for females.[14] The fit to the selected wage quantiles improved slightly. At the same time, quantiles of latent wages were comparable to the ones in Figure 2.

Choice of copula. We then investigate the robustness of our results to the choice of the copula. The symmetry properties of the Frank copula are apparent in the first two rows of Figure 4, which shows the contour plots of the copula densities that we estimated on the FES data.[15] As a specification check, we consider an encompassing two-parameter family, which we call the "generalized Frank copula". This family may capture different degrees of dependence in different regions of the (U, V) plane, as we explain in Appendix G. The estimated copula densities in the generalized Frank family are shown in the last two rows of Figure 4. We see that, for both males and females, the differences between the estimated Frank and generalized Frank copulas are relatively small. Moreover, as shown by Figure 5, the quantiles of latent wages are similar for both genders when using a Frank or a generalized Frank copula.

Bounds estimates. As a further check of the influence of functional forms on the estimates, in Figure 6 we report estimates of the bounds derived in equation (18). We see that bounds on wage quantiles for males (in dashed lines) are essentially on top of each other. The bounds for females are wider, though still informative. However, the results for females are sensitive to the estimator of the supremum of the propensity score ($\overline{p}_x$) that we use. Larger participation rates are associated with smaller values of out-of-work income. In Figure 7 we report estimates of the bounds when trimming 1% of extreme observations in out-of-work income. We see that, while the results for males are very stable, those for females are very different, showing extremely wide bounds throughout the wage distribution. This reflects the fact that the selection problem is more severe for females, as their employment rates are lower.

[14] On US data, Mulligan and Rubinstein (2008) document that women's selection into participation shifted from being negative in the 1970s to being positive in the 1990s.

[15] As a graphical convention (common in the literature on copulas), we plot the copula density by rescaling the margins so that they are standard normal. That is, if $C(u, v)$ denotes the copula, we plot the contours of:

$$(x, y) \mapsto \phi(x)\,\phi(y)\,\frac{\partial^2 C}{\partial u \partial v}\left(\Phi(x), \Phi(y)\right),$$

where ϕ and Φ denote the standard normal density and cdf, respectively.

In Figure 8 we compare the bounds, for males, for two education groups: statutory schooling (71% of the sample, in thin lines) and high-school and college (29%, in thick lines). We use a trimmed estimator of the supremum of the propensity score. We see that the bounds are narrow for more educated individuals, and that they are wider for the low educated whose employment rates are lower. We observe some evidence of an increase in the education gap over time, particularly at the median, although the evidence after correcting for selection is more mixed. The graphs also show evidence of an inequality increase within the education groups that we consider (similarly as in Blundell *et al.*, 2007).

4.3 Counterfactuals in the presence of equilibrium effects

In this section we consider a simple equilibrium model of wage quantile functions and non-random selection into work as a flexible tool for examining changes in the distribution of wages over time. We show how the simplicity of linear quantiles can be essentially preserved while embedding wage functions in a model of human capital, employment decisions, and labor demand. We then use the model to recompute wage and employment distributions in a counterfactual scenario where potential out-of-work income is kept at its 1978 value.

Wages and participation. We abstract from hours of work and dynamics. Let r_t^s be the skill price of a worker of education level s in time period t. Let also $h(s,x,u)$ be the amount of human capital of a worker with education (or "skill level") s, observed characteristics x (such as cohort and gender), and unobserved ability u. The wage rate for an individual i of schooling level S_i in period t is:

$$W_{it} = r_t^{S_i} \cdot h\left(S_i, X_{it}, U_{it}\right),$$

where there are two skill levels ($S_i \in \{1,2\}$). Note that the human capital function h is time-invariant. This assumption is called the "proportionality hypothesis" in Heckman and Sedlacek (1985).

The individual work decision is:

$$D_{it} = 1\left\{r_t^{S_i} h\left(S_i, X_{it}, U_{it}\right) \geq W^R\left(S_i, Z_{it}, \eta_{it}\right)\right\},$$

where $Z_{it} = (B_{it}, X_{it})$. We mimic the setting of standard selection models consisting of an outcome equation and a participation equation.

Let $\overline{X}_{it} \equiv (S_i, X_{it})$. The log-human capital function and log-reservation wage are specified as: $\ln h\left(S_i, X_{it}, U_{it}\right) \equiv \overline{X}'_{it}\beta(U_{it})$, and: $\ln W^R\left(S_i, Z_{it}, \eta_{it}\right) \equiv \overline{X}'_{it}\gamma\left(\eta_{it}\right) + B'_{it}\varphi$, so that the participation decision is:

$$D_{it} = 1\left\{\overline{X}'_{it}\gamma\left(\eta_{it}\right) - \overline{X}'_{it}\beta\left(U_{it}\right) \leq \ln r_t^{S_i} - B'_{it}\varphi\right\} = 1\left\{V_{it} \leq \overline{F}\left(\ln r_t^{S_i} - B'_{it}\varphi, \overline{X}_{it}\right)\right\},$$

where the composite error $\overline{X}'_{it}\left(\gamma\left(\eta_{it}\right) - \beta\left(U_{it}\right)\right)$ is assumed independent of Z_{it} given $\overline{X}_{it} = \overline{x}$ with cdf $\overline{F}\left(., \overline{x}\right)$, and V_{it} is its uniform transformation. In practice, we approximate the propensity score by a single-index (probit) model of the form $F\left(\ln r_t^{S_i} - Z'_{it}\psi\right)$.

Using wage and participation equations, our quantile selection approach allows one to perform partial equilibrium counterfactual exercises where skill prices r_t^s are kept constant. In order to allow for equilibrium responses in skill prices, we now introduce a model for labor demand. See Heckman, Lochner and Taber (1998) and Lee and Wolpin (2006) for related approaches in structural settings.

Labor demand. Consider a one-sector economy with one physical capital input (which we assume fixed) and two types of human capital. We assume a standard aggregate production function: $F_t\left(L_t, K_t\right) = A_t L_t^{\alpha} K_t^{1-\alpha}$, where L_t is a CES aggregator of the human capital inputs: $L_t = \left[a_t H_{1t}^{\phi} + \left(1 - a_t\right) H_{2t}^{\phi}\right]^{1/\phi}$. If $\phi = 1$ the two labor skills are perfect substitutes, in which case an increase in the supply of one type of human capital does not affect the relative skill prices. The scope for equilibrium effects critically depends on the structure of production.

From the first-order conditions we obtain:

$$\ln\left(\frac{r_t^1}{r_t^2}\right) = \ln\left(\frac{a_t}{1 - a_t}\right) + (\phi - 1)\ln\left(\frac{H_{1t}}{H_{2t}}\right). \tag{19}$$

In Appendix H we discuss how to recover estimates of H_{1t}, H_{2t}, ϕ, and a_t from micro-data based on (19). In practice, due to weak identification from our time-series, we calibrate $\phi = .4$ using Card and Lemieux (2001)'s estimate on UK data. We also import $\alpha = .6$ from the literature.

Counterfactual equilibrium skill prices. Suppose we are interested in estimating the counterfactual equilibrium skill prices, $\ln \widetilde{r}_t^s$ say, that would have prevailed under technology conditions in period t and the labor force composition or the welfare policy in some other period.

Equilibrium log skill prices satisfy the equations:

$$\ln r_t^s = \ln A_t + \ln \alpha + (1-\alpha)\ln\left(\frac{K_t}{L_t}\right) + \ln a_{st} + (\phi - 1)\ln\left(\frac{H_{st}}{L_t}\right),$$

where $a_{1t} \equiv a_t$ and $a_{2t} \equiv 1 - a_t$. In addition, the labor supply equations imply:

$$H_{st}\left(r_t^s\right) = \sum_{S_i = s} F\left(\ln r_t^s - Z_{it}'\psi\right) \int_0^1 e^{\overline{X}_{it}'\beta(u)} dG\left[u, F\left(\ln r_t^s - Z_{it}'\psi\right); \rho\right], \quad s = 1, 2, \quad (20)$$

$$L_t = \left(a_{1t}\left[H_{1t}\left(r_t^1\right)\right]^{\phi} + a_{2t}\left[H_{2t}\left(r_t^2\right)\right]^{\phi}\right)^{1/\phi}. \quad (21)$$

The log-difference between observed and counterfactual skill prices is given by:

$$\ln \widetilde{r}_t^s - \ln r_t^s = (1-\alpha)\ln\left(\frac{L_t}{\widetilde{L}_t}\right) + (\phi - 1)\left[\ln\left(\frac{\widetilde{H}_{st}}{\widetilde{L}_t}\right) - \ln\left(\frac{H_{st}}{L_t}\right)\right], \quad s = 1, 2, \quad (22)$$

where the counterfactual skill aggregates $\widetilde{H}_{st}$ and $\widetilde{L}_t$ satisfy (20)-(21) at prices $(\widetilde{r}_t^1, \widetilde{r}_t^2)$. Note that capital (which is fixed) and neutral technical progress are common to both sets of prices and thus cancel out in (22).

Counterfactual log-skill prices $\ln \widetilde{r}_t^1$ and $\ln \widetilde{r}_t^2$ are then obtained as the solution to the two nonlinear equations in (22), using (20)-(21). This fixed-point problem depends on the following inputs: the parameters β, ψ, ρ, and r_t^s (estimated using our quantile selection method), the aggregate quantities H_{st} and L_t and the technological shocks a_t (estimated as explained in Appendix H), and the parameters ϕ and α (which we take from the literature). As starting value for the counterfactual $\widetilde{r}_t^s$ we take the estimated r_t^s, and we solve for the fixed point iteratively.

Results. Figure 9 shows the estimates of latent wage quantiles in two scenarios: when out-of-work income is as in the data (solid lines), and in a counterfactual scenario when out-of-work income is kept at its 1978 value (dashed). The specification that we use has some differences compared to the one in Figure 2. In particular, here the two education groups are college and non-college, the specification is pooled across genders, and controls are interacted with gender.[16] We present the results by gender.

We see that accounting for general equilibrium responses tends to lower latent counterfactual quantiles throughout the distribution. This is due to the fact that in the counterfactual scenario out-of-work income is lower, thus increasing employment rates, and as a result

[16]The fit of the model used in this subsection is shown in Figures 12 and 13.

pushing skill prices down. General equilibrium effects appear to be relatively small for both genders, although they seem more sizable at the bottom of the distribution.

Figure 11 shows actual employment rates (as predicted by the model), and employment rates in the partial equilibrium and general equilibrium counterfactuals. We see that in the counterfactual employment rates tend to increase (dashed lines). The dampening effect on employment that comes from the general equilibrium response of skill prices is quantitatively small (dotted lines).

Lastly, Figure 11 shows the actual evolution of wages conditional on employment as predicted by the model (solid lines), and the evolution in the counterfactual scenario where out-of-work income is kept at its 1978 value, with skill prices fixed (dashed) and with skill prices adjusting through general equilibrium (dotted). We see that, in the partial equilibrium counterfactual, wages of male workers tend to be lower at the bottom of the distribution, due to positive selection into employment. In addition, general equilibrium responses imply further reduction in wages. In the middle and at the top of the distribution, and for females, differences between actual and counterfactual evolution appear to be smaller.

5 Conclusion

We have presented a three-step method to correct quantile regression estimates for sample selection. In a first step, the parameters of the participation equation are estimated. In a second step, the parameters of the copula linking the percentile error of the outcome equation to the participation error are computed by minimizing a method-of-moments objective function. In a third step, quantile parameters are computed by minimizing a weighted check function, using a fast linear programming routine. The method provides a simple and intuitive way to compute selection-adjusted quantile parameters. Moreover, our application shows that such selection corrections for quantiles may be as empirically relevant as in the standard regression context of the popular Heckman (1979) sample selection model.

An important issue is the choice of the copula. An approach that treats the copula nonparametrically is conceptually attractive, for example a sieve approach based on conditional moment restrictions as in Chen and Pouzo (2009, 2012). It would be desirable to allow the copula to be partially identified, and to conduct inference on the identified set of quantile functions. The empirical application suggests that nonparametric bounds might be informative when selection is not too severe (as in the case of men in the application).

References

[1] Abadie, A. (2003): "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, 113(2), 231–263.

[2] Abadie, A., J. Angrist, and G. Imbens (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70(1), 91–117.

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

[4] 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.

[5] 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.

[6] Autor, D., L. Katz, and M. Kearney (2005): "Residual Wage Inequality: The Role of Composition and Prices," NBER Working Paper 11628.

[7] Belloni, A., V. Chernozhukov, and I. Fernández-Val (2011): "Conditional Quantile Processes based on Series or Many Regressors," unpublished manuscript.

[8] 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.

[9] 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.

[10] 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.

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

[12] 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.

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

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

[15] Card, D., and T. Lemieux (2001): "Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis," *Quarterly Journal of Economics*, 116, 705–746.

[16] Carneiro, P., and S. Lee (2009): "Estimating Distributions of Potential Outcomes using Local Instrumental Variables with an Application to Changes in College Enrollment and Wage Inequality," *Journal of Econometrics*, 149, 191–208.

[17] 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).

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

[19] Chen, X., and D. Pouzo (2012): "Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Moments," *Econometrica*, 80(1), 277–322.

[20] Chernozhukov, V., and I. Fernández-Val (2005): "Subsampling Inference on Quantile Regression Processes," *Sankhya: The Indian Journal of Statistics*, 67(2), 253–276.

[21] Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010): "Quantile and Probability Curves without Crossing," *Econometrica*, 78(3), 1093–1125.

[22] Chernozhukov, V., I. Fernández-Val, and B. Melly (2013): "Inference on Counterfactual Distributions," *Econometrica*, 81(6), 2205-68.

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

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

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

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

[27] D'Haultfoeuille, X., and P. Février (2015): "Identification of Nonseparable Triangular Models With Discrete Instruments," *Econometrica*, 3, 1199–1210.

[28] DiNardo, J., N. Fortin, and T. Lemieux (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 64, 1001–1044.

[29] Firpo, S., N. Fortin, and T. Lemieux (2011): "Decomposition Methods in Economics," in O. Ashenfelter and D. Card, eds., *Handbook of Economics*, Amsterdam: North-Holland, Vol. IV.A, 1–102.

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

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

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

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

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

[35] 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.)

[36] Heckman, J. J., L. Lochner, and C. Taber (1998): "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Earnings with Heterogeneous Agents," *Review of Economic Dynamics*, 1, 1–58.

[37] Heckman, J. 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.

[38] Heckman, J. J., J. Smith, and N. Clements (1997): "Making the Most out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64, 487–535.

[39] Heckman, J. J., and E. Vytlacil (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73(3), 669–738.

[40] Huber, M., and B. Melly (2015): "A Test of the Conditional Independence Assumption in Sample Selection Models," to appear in the *Journal of Applied Econometrics*.

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

[42] Imbens, G., and D. Rubin (1997): "Estimating Outcome Distributions for Compliers in Instrumental Variables Models," *Review of Economic Studies*, 64, 555–574.

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

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

[45] Katz, L. F., and K. M. Murphy (1992): "Changes in the Wage Structure, 1963-1987: Supply and Demand Factors," *Quarterly Journal of Economics*, 107, 35–78.

[46] Kitagawa, T. (2009): "Identification Region of the Potential Outcome Distributions under Instrument Independence," Cemmap Working Paper CWP30/09.

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

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

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

[50] Lee, D., and K. I. Wolpin (2006): "Intersectoral Labor Mobility and the Growth of the Services Sector," *Econometrica*, 74(1), 1–46.

[51] 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.

[52] Manski, C. F. (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.

[53] Manski, C. F. (2003): *Partial Identification of Probability Distributions*, Springer Series in Statistics, Vol. 12. Berlin: Springer Verlag.

[54] Mulligan, C., and Y. Rubinstein (2008): "Selection, Investment, and Women's Relative Wages Over Time," *Quarterly Journal of Economics*, 123(3), 1061–1110.

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

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

[57] 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.

[58] Picchio, M., and C. Mussida (2010): "Gender Wage Gap: a Semi-Parametric Approach with Sample Selection Correction," IZA Discussion Paper No. 4783. Bonn: IZA.

[59] Politis, D. N., J. P. Romano, and M. Wolf (1999): *Subsampling*. Springer, New York.

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

[61] Sancetta,A., and S. Satchell (2004): "The Bernstein Copula and its Applications to Modeling and Approximations of Multivariate Distributions," *Econometric Theory*, 20, 535–562.

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

[63] Torgovitsky, A. (2015): "Identification of Nonseparable Models Using Instruments with Small Support," *Econometrica*, 83(3), 1185–1197.

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

APPENDIX

A Proofs

Proof of Lemma 1. Equation (6) is a direct application of (5), using the fact that by A3 both G_x and $F_{Y|D=1,Z}$ are strictly increasing in their first argument.

To show the second part, let $x \in \mathcal{X}$ and let G_x satisfy (6). Pick a $z_x \in \mathcal{Z}_x$, and define:

$$F_{Y^*|X}(y|x) \equiv G_x^{-1}\left(F_{Y|D=1,Z}(y|z_x), p(z_x)\right).$$

For all (z, y) in the support of (Z, Y) given $X = x$ we have:

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

where we have used (6) to obtain the second equality.

Proof of Proposition 1. Let us start with i). Evaluating (6) at $z_1 = z$ and $z_2 = z_x$, and noting that $G_x^{-1}(\tau, 1) = \tau$, we have that $G_x(\tau, p(z)) = F_{Y|D=1,Z}\left(F_{Y|D=1,Z}^{-1}\left(\tau|z_x\right)|z\right)$. Hence G_x is identified. The identification of q then comes from (5) and Assumption A3.

Let us now suppose ii). Let G_x and $\widetilde{G}_x$ satisfy model (1)-(3), and let Assumptions A1 to A4 hold. Then, by (6) we have:

$$G_x\left[G_x^{-1}(\tau, p_2), p_1\right] - \widetilde{G}_x\left[\widetilde{G}_x^{-1}(\tau, p_2), p_1\right] = 0, \quad \text{for all } (p_1, p_2) \in \mathcal{P}_x \times \mathcal{P}_x.$$

Hence, for each $\tau \in (0, 1)$, the function:

$$(p_1, p_2) \mapsto G_x\left[G_x^{-1}(\tau, p_2), p_1\right] - \widetilde{G}_x\left[\widetilde{G}_x^{-1}(\tau, p_2), p_1\right],$$

which is real analytic, is zero on a product of two open neighborhoods. As a result it is zero everywhere on $(0, 1) \times (0, 1)$, and evaluating it at $p_2 = 1$ leads to:

$$G_x(\tau, p_1) - \widetilde{G}_x(\tau, p_1) = 0, \quad \text{for all } p_1 \in \mathcal{P}_x.$$

Hence, as G_x and $\widetilde{G}_x$ are real analytic in their second argument, they coincide on $(0, 1) \times (0, 1)$. This implies that G_x, and hence q (as in the first part of the proof), are identified.

Proof of Proposition 2. For clarity here we denote $x = (\widetilde{x}, 1)$, where $\widetilde{x}$ contains all covariates but the constant term. Let also $\widetilde{\beta}$ contain all β coefficients except the intercept. Finally, let $\widetilde{q}^d(x, p) = x'\beta_{G^{-1}(\tau,p;\rho)}$. For $\widetilde{q}^d(x, p)$ to be additive in $\widetilde{x}$ and p, it is necessary and sufficient that $\widetilde{\beta}_{G^{-1}(\tau,p;\rho)}$ does not depend on p. This happens only if $\widetilde{\beta}_\tau$ does not depend on τ, or if $G^{-1}(\tau, p; \rho)$ does not depend on p. In the second case, taking $p = 1$ implies that $G^{-1}(\tau, p; \rho) = \tau$ for all (τ, p), so U and V are independent.

B 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. We start by noting that, as the model is correctly specified, there exists a copula C_0 (with conditional copula G_0) and a cdf F_0, which are the true copula and cdf of (U, V) and Y^*, respectively. Let $\mathcal{P}$ denotes 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:

$$C(\tau, p) \equiv C_0\left(G_0^{-1}\left(\widetilde{G}(\tau, \overline{p}), \overline{p}\right), p\right), \quad \text{for all } (\tau, p) \in (0,1) \times \mathcal{P}. \tag{B1}$$

It is simple to see that C is a subcopula.[17] 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:

$$F(y) \equiv \widetilde{G}^{-1}\left(G_0\left(F_0(y), \overline{p}\right), \overline{p}\right), \quad \text{for all } y \in \mathcal{Y}. \tag{B2}$$

Note that F is a cdf.

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:

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

where we have used (B2) and (B1) in the second and third equalities, respectively.

Finally, to see that F in (B2) 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 $(\tau, p) \mapsto \min\left(\frac{\tau}{p}, 1\right)$. Similarly, for the lower bound we take a $\widetilde{G}$ that satisfies Assumption A3 and is arbitrarily close to $(\tau, p) \mapsto \max\left(\frac{\tau+p-1}{p}, 0\right)$.[18]

C Estimation with discrete covariates

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

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

[17] 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 $p_1 \leq p_2$.

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

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

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

with x_k denoting the points of support of X. Let $\overline{G}_k(\tau,c)$ denote the mean of $G\left(\tau, p(Z_i;\widehat{\theta});c\right)$ 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:

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

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

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

which is continuously differentiable with respect to c as soon 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.

D An alternative estimator for the copula parameter

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

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

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

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

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

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 method can be iterated. Once an estimator of ρ is available one can update it as follows. Recall that the observed quantiles satisfy:

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

Hence, given estimates $\widehat{\rho}$ and $\widehat{\beta}$, we can estimate:

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

and update ρ by minimizing:

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

This procedure may be iterated further. We leave the study of the asymptotic properties of this alternative estimator of the copula parameter to future work.

E 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).

E.1 Analysis conditional on a consistent and asymptotically normal estimator of ρ

Let:

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

We make the following assumptions.

Assumption E1

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 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

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

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. In particular, this requires ρ to be point-identified from the population counterpart of (15). Under weak regularity conditions, it is easy to show that:

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

where we have denoted:

$$G_{\tau i} \equiv G\left(\tau, p\left(Z_i;\theta\right);\rho\right). \tag{E3}$$

Condition *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 cdf of Y given Z, and not given X. Condition *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, Condition *iv*) requires the existence of moments.

Theorem E1 *Let $\tau \in]0,1[$, and let Assumptions A1 to A4 and E1 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 $P_\tau \equiv [I_{\dim\beta}, -P_{1\tau}, -P_{2\tau}]$, and $J_\tau, P_{1\tau}, P_{2\tau}$ are given in Assumption E1.

Theorem E1 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 the standard 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):

$$\frac{1}{N}\sum_{i=1}^{N} X_i g_i\left(\widehat{\beta}_\tau, \widehat{\theta}, \widehat{\rho}\right) = O_p\left(\frac{1}{N}\right), \tag{E4}$$

where

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

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

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

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 E1 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:

$$\begin{aligned}
\widehat{\beta}_\tau - \beta_\tau &= -J_\tau^{-1}\left[\widehat{\mathbb{E}}\left[X_i g_{i\tau}\right] - P_{1\tau}\left(\widehat{\theta} - \theta\right) - P_{2\tau}\left(\widehat{\rho} - \rho\right)\right] + o_p\left(\frac{1}{\sqrt{N}}\right) \qquad \text{(E5)} \\
&= -J_\tau^{-1}P_\tau \begin{pmatrix} \widehat{\mathbb{E}}\left[X_i g_{i\tau}\right] \\ \widehat{\theta} - \theta \\ \widehat{\rho} - \rho \end{pmatrix} + o_p\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}$$

The result then comes from part *i*) in Assumption E1.

■

E.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.[19]

Assumption E2

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

$$\widehat{\theta} - \theta = -H^{-1}\widehat{\mathbb{E}}[S_i] + o_p\left(\frac{1}{\sqrt{N}}\right). \tag{E6}$$

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

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

iii) The following matrix inverse exists:

$$A_\rho \equiv \left[\sum_{\ell=1}^{L}\left(\widetilde{P}_{2\tau_\ell} - \widetilde{J}_{\tau_\ell} J_{\tau_\ell}^{-1} P_{2\tau_\ell}\right)\right]^{-1}. \tag{E7}$$

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.

Define the following matrices:

$$B_\rho \equiv -A_\rho\left[\widetilde{J}_{\tau_1} J_{\tau_1}^{-1}, ..., \widetilde{J}_{\tau_L} J_{\tau_L}^{-1}\right], \tag{E8}$$

$$C_\rho \equiv A_\rho\left(\sum_{\ell=1}^{L}\left[\widetilde{P}_{1\tau_\ell} - \widetilde{J}_{\tau_\ell} J_{\tau_\ell}^{-1} P_{1\tau_\ell}\right] H^{-1}\right), \tag{E9}$$

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

$$A_\beta(\tau) \equiv J_\tau^{-1} P_{2\tau} A_\rho, \tag{E10}$$

$$B_\beta(\tau) \equiv J_\tau^{-1} P_{2\tau} B_\rho, \tag{E11}$$

$$C_\beta(\tau) \equiv J_\tau^{-1}\left(P_{2\tau} C_\rho - P_{1\tau} H^{-1}\right). \tag{E12}$$

Then, let:

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

[19] Note that the instrument function $\varphi(\tau, Z_i) = p\left(Z_i;\widehat{\theta}\right)$ used in Section 4 depends on $\widehat{\theta}$. This slightly affects the formula for the asymptotic variance. For simplicity here we do not account for this dependence.

Lastly, let:

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

where $G_{\tau i}$ is given by (E3), and define:

$$\Omega_\tau \equiv \left(\begin{array}{ccccc}
\Omega_\tau^{1,1} & \Omega_\tau^{1,2} & ... & \Omega_\tau^{1,L+2} & 0 \\
\Omega_\tau^{2,1} & \Omega_\tau^{2,2} & ... & \Omega_\tau^{2,L+2} & 0 \\
... & ... & ... & ... & ... \\
\Omega_\tau^{L+2,1} & \Omega_\tau^{L+2,2} & ... & \Omega_\tau^{L+2,L+2} & 0 \\
0 & 0 & ... & 0 & \mathbb{E}\left[S_iS_i'\right]
\end{array}\right), \tag{E13}$$

where Ω_τ is symmetric, and:

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

We have the following result.

Theorem E2 *Let Assumptions A1 to A4, E1, and E2 hold. Suppose that* $\dim\varphi = \dim\rho$. *Then:*

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

Proof.
As in the proof of Theorem E1, we start with an approximate moment equation:

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

Expanding around true parameter values:

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

So, by (E5):

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

So, by (E6):

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

Hence:

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

where A_ρ, B_ρ, and C_ρ are given by (E7)-(E9), and:

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

Let now $\tau\in]0,1[$. Using (E5):

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

So:

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

where $A_\beta\left(\tau\right)$, $B_\beta\left(\tau\right)$, and $C_\beta\left(\tau\right)$ are given by (E10)-(E12).

Next, denote:

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

From the above, we have:

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

with:

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

Finally, we check that $\mathbb{E}\left(\psi_{i\tau}\psi_{i\tau}'\right)=\Omega_{\tau}$ given by (E13):

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

and similarly:

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

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

This completes the proof of Theorem E2.

■

Estimating the asymptotic variance. To construct an empirical counterpart of the asymptotic variance appearing in Theorem E1, 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:

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

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 V_{τ} that appears in Theorem E2.

Extensions. The results can be easily generalized to derive the asymptotic distribution for 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 $\tau\mapsto\sqrt{N}\left(\widehat{\beta}_{\tau}-\beta_{\tau}\right)$. This can be done along the lines of Koenker and Xiao (2002) or Chernozhukov and Hansen (2006). Confidence bands for unconditional effects may be derived using the results in Chernozhukov, Fernández-Val and Melly (2013).

F Extensions

Treatment effects with selection on unobservables. As a direct extension of model (1)-(3) consider the following system of equations:

$$Y_0^*=q\left(U_0,X\right),\quad Y_1^* = q\left(U_1,X\right),\quad Y=(1-D)Y_0^*+DY_1^*,$$

where, in the spirit of Assumption A1, (U_0, U_1, V) is assumed independent of Z given X. This model coincides with the standard potential outcomes framework in the treatment effects literature (Vytlacil, 2002). In the context of the empirical application, $Y_0^* = 0$, and Y_1^* is the partial equilibrium causal effect of working. In this framework, the quantile IV method of Chernozhukov and Hansen (2005) relies on an assumption of rank invariance or rank similarity which restricts the dependence between U_0 and U_1. Specifically, rank invariance (respectively, similarity) requires that U_0 and U_1 be identically distributed (resp., given V), thus ruling out most patterns of sample selection. In contrast, in the identification analysis our approach leaves the conditional distribution of U_0, U_1 and V given X unrestricted.

Nonparametric propensity score. Although the paper focuses on the case where the propensity score is parametrically specified, our approach can accommodate a nonparametric modelling 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 (1998)'s censored quantile regression estimator. When using a nonparametric estimate $\widehat{p}(Z)$ (for example, a kernel-based Nadaraya-Watson estimator), Buchinsky and Hahn's construction relies on trimming out the observations for which $\widehat{p}(Z_i) < c$, where $c > 0$ is a small number. A suitable choice of c guarantees root-N consistency of the quantile regression coefficients. We conjecture that a similar device could work in our case, although 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:

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

This motivates using a test statistic of the form:

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

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:

$$Y^* = E'\alpha_U + X'\beta_U, \tag{F14}$$

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 $\tau \in (0,1)$:

$$\mathbb{E}\left[\mathbf{1}\left\{Y \leq E'\alpha_\tau + X'\beta_\tau\right\} - G(\tau, p(Z;\theta); \rho) \mid D = 1, Z = z\right] = 0. \tag{F15}$$

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

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

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

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

and:

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

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:

$$\Pr\left(Y \le \max\left\{X'\beta_\tau, y_0\right\} \mid D = 1, Z = z\right) = G\left(\tau, p(z;\theta); \rho\right). \tag{F16}$$

In particular, 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 $\max\left\{X_i'\widetilde{\beta}_\tau(c), y_0\right\}$, respectively, where:

$$\begin{aligned}\widetilde{\beta}_\tau(c) \quad \equiv \quad & \underset{b}{\operatorname{argmin}} \sum_{i=1}^{N} D_i \Big\{ G\left(\tau, \widehat{p}\left(Z_i;\widehat{\theta}\right); c\right) \left(Y_i - \max\left\{X_i'b, y_0\right\}\right)^+ \\ & + \left(1 - G\left(\tau, \widehat{p}\left(Z_i;\widehat{\theta}\right); c\right)\right) \left(Y_i - \max\left\{X_i'b, y_0\right\}\right)^- \Big\}.\end{aligned} \tag{F17}$$

The optimization problem in (F17) is a selection-corrected version of Powell's (1986) censored quantile estimator.

G 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:

$$C(u,v;\gamma,\theta) = \frac{1}{\delta}\left[1 - \left\{1 - \frac{1}{\gamma}\left[1 - (1-\delta u)^\theta\right]\left[1 - (1-\delta v)^\theta\right]\right\}^{\frac{1}{\theta}}\right], \tag{G18}$$

where $\delta = 1 - (1-\gamma)^{\frac{1}{\theta}}$. Joe (1997) refers to (G18) 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. Importantly for interpretation, the concordance of the generalized Frank copula given by (G18) 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:

$$C_F(u, v; \gamma) = \frac{1}{\ln(1-\gamma)} \ln\left[1 - \frac{1}{\gamma}\{1 - \exp[\ln(1-\gamma)u]\}\{1 - \exp[\ln(1-\gamma)v]\}\right]. \tag{G19}$$

C_F given by (G19) 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) plane. In comparison, the generalized Frank copula (G18) 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.[20] 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.

H Estimating the elasticity of substitution

The estimation of equation (19) is based on time series aggregate data. We use the microdata to construct time series of the relevant aggregates. The time series of the log-relative price of skill $\ln\left(\widehat{r}_t^1/\widehat{r}_t^2\right)$ is obtained from the estimation of the wage functions. Time series of relative aggregate labor supplies can be estimated by aggregation of individual units of human capital of employed workers:

$$\ln\widehat{\left(\frac{H_{1t}}{H_{2t}}\right)} = \ln\sum_{S_i=1}\frac{W_{it}}{\widehat{r}_t^1} - \ln\sum_{S_i=2}\frac{W_{it}}{\widehat{r}_t^2} = \ln\left(\sum_{S_i=1}W_{it}/\sum_{S_i=2}W_{it}\right) - \ln\left(\widehat{r}_t^1/\widehat{r}_t^2\right).$$

The log ratio of factor-specific productivities $\ln\left(\frac{a_t}{1-a_t}\right)$ is allowed to vary over time to capture skill-biased technical change. It is specified as a trend $\lambda(t)$ plus an unobservable shock ε_t. The equation to be estimated is therefore:

$$\ln\left(\widehat{r}_t^1/\widehat{r}_t^2\right) = \lambda(t) + (\phi - 1)\ln\left(\widehat{H}_{1t}/\widehat{H}_{2t}\right) + \varepsilon_t. \tag{H20}$$

[20] This is because:

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

This equation was estimated on aggregate US data by Katz and Murphy (1992), who obtained $\widehat{\phi} = 0.3$. A comparable estimate on UK data in Card and Lemieux (2001) is $\widehat{\phi} = 0.4$. We then estimate a_t as:

$$\widehat{a}_t \equiv \Lambda\left(\ln\left(\widehat{r}_t^1/\widehat{r}_t^2\right) - \left(\widehat{\phi} - 1\right)\ln\left(\widehat{H}_{1t}/\widehat{H}_{2t}\right)\right),$$

where $\Lambda(r) = \exp(r)/(1 + \exp(r))$.

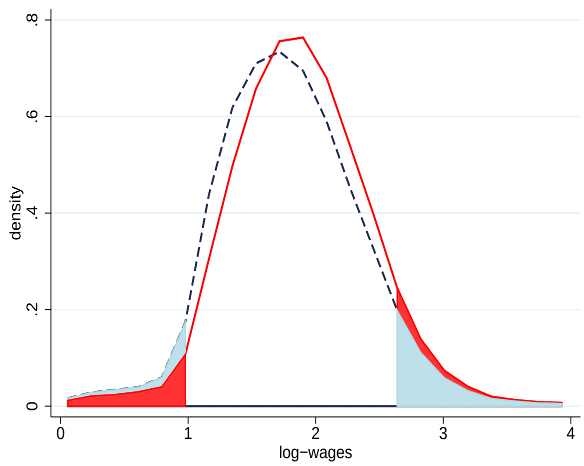
Finally, note that the explanatory variable $\ln\left(\widehat{H}_{1t}/\widehat{H}_{2t}\right)$ is likely to be correlated with ε_t in (H20), in which case OLS estimates are inconsistent. Natural instrumental variables would be aggregates (by skill) of labor supply shifters such as potential out-of-work welfare income.

Table 1: 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

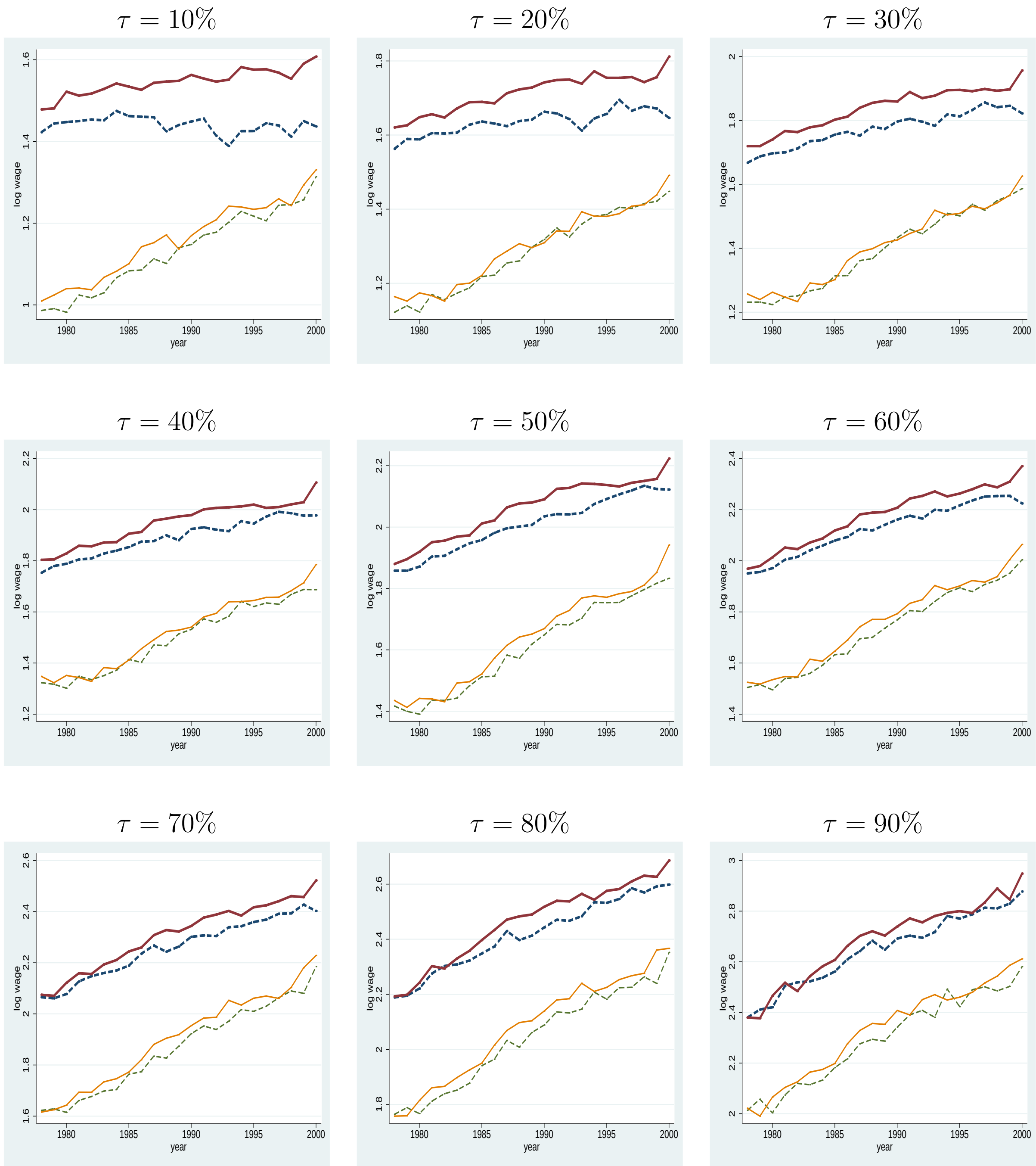
Source: Family Expenditure Survey, 1978-2000.
Note: The propensity score is estimated using a probit model.

Figure 1: Sample selection shifts percentile ranks



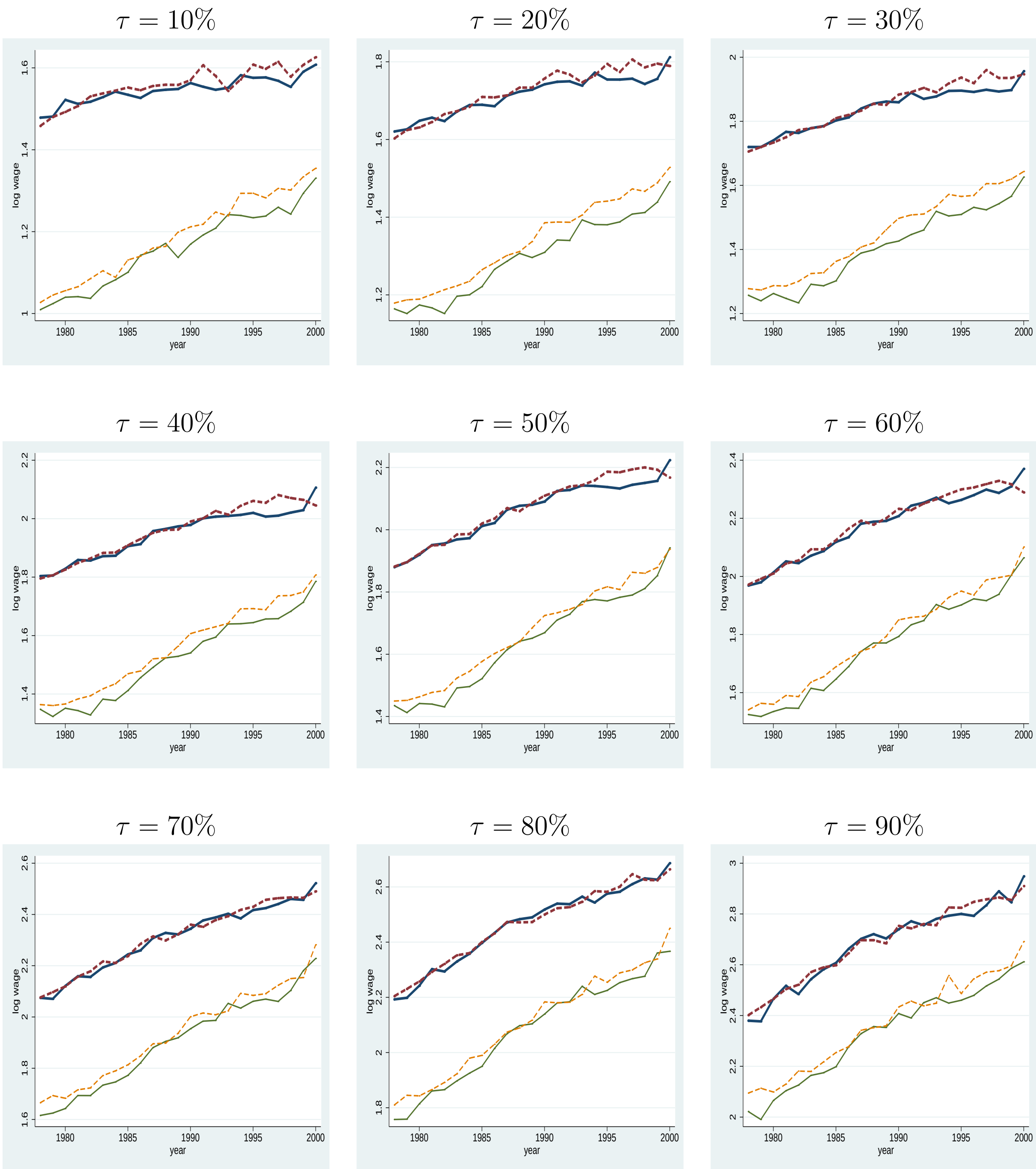
Note: Latent (dashed) and observed (solid) log-wages. Indicated are the $\tau = 10\%$ *and* 90% *percentiles of latent log-wages.*

Figure 2: Wage quantiles, by gender



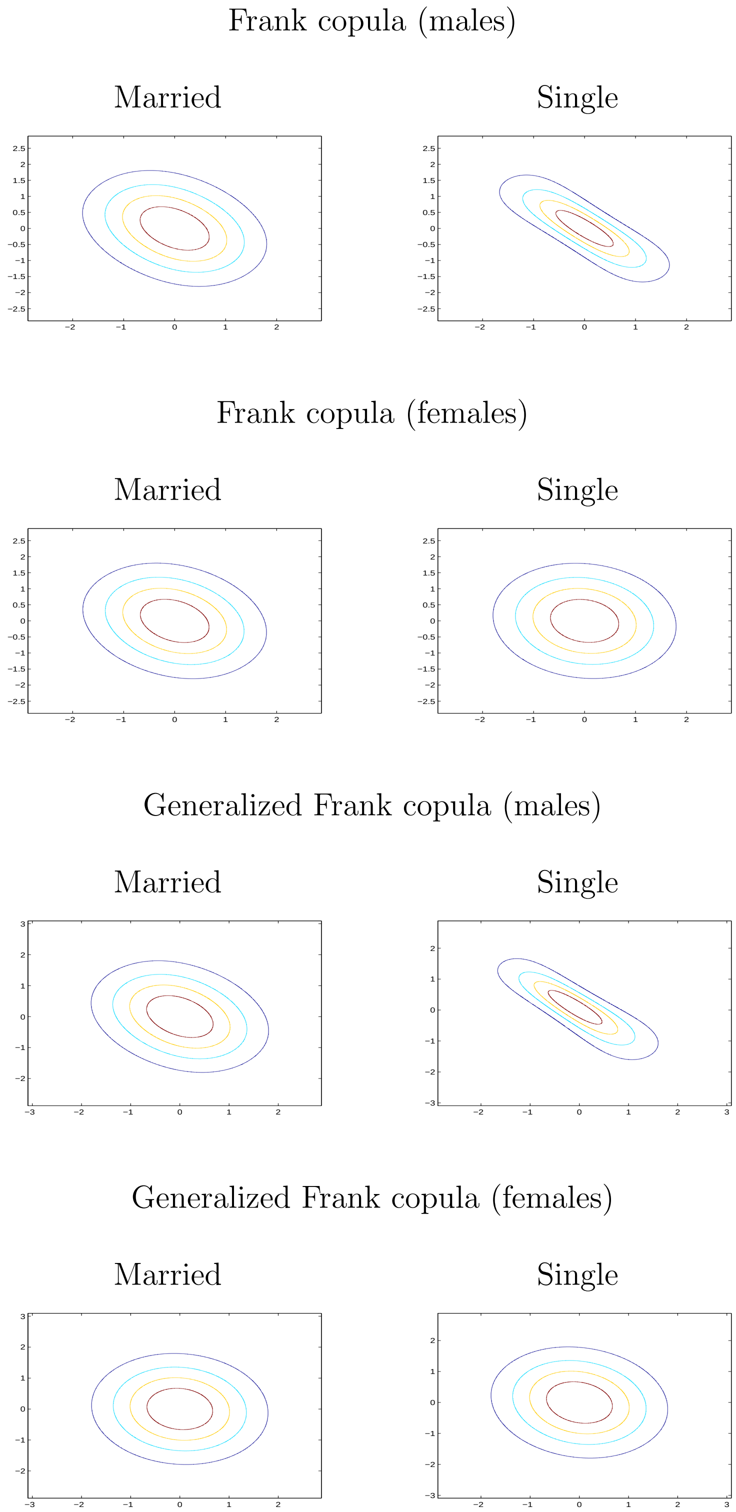
Note: FES data for 1978-2000. Quantiles of log-hourly wages, conditional on employment (solid lines) and corrected for selection (dashed). Male wages are plotted in thick lines (top lines in each graph), while female wages are in thin lines (bottom lines).

Figure 3: Fit to wage quantiles, by gender (employed individuals)



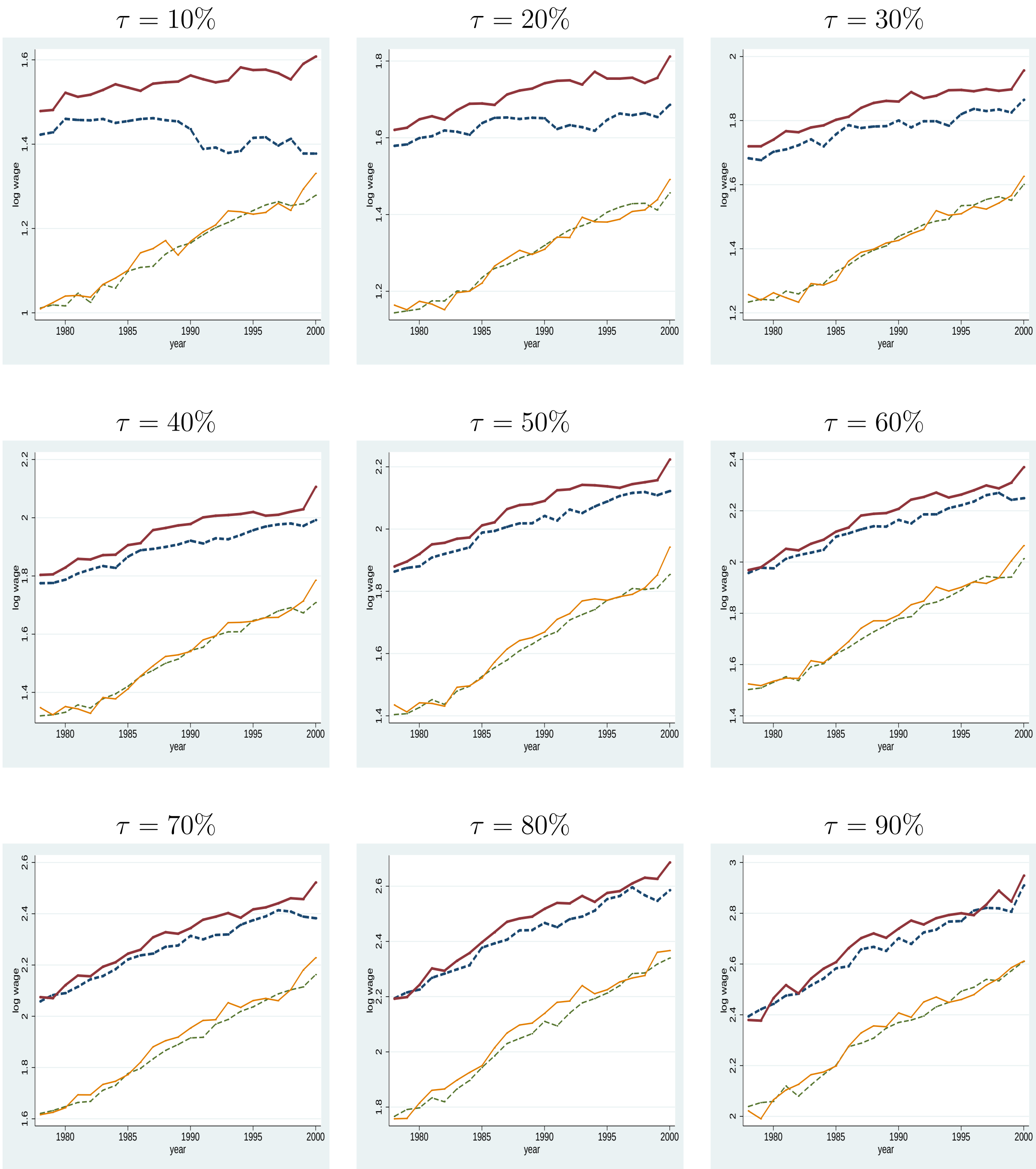
Note: FES data for 1978-2000. Quantiles of log-hourly wages conditional on employment, data (solid lines) and model fit (dashed). Male wages are plotted in thick lines (top lines in each graph), while female wages are in thin lines (bottom lines).

Figure 4: Contour plots of the copula



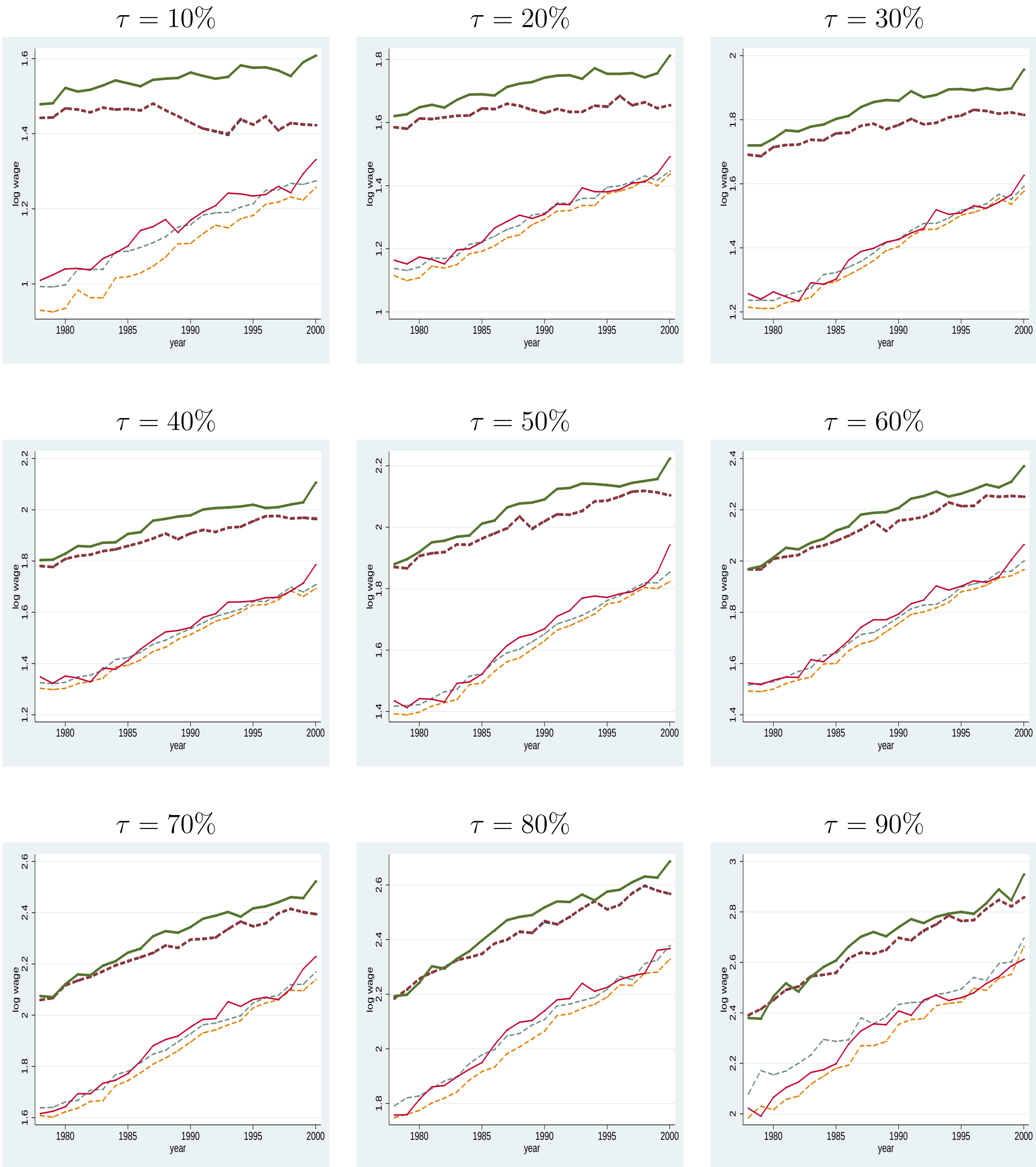
Note: FES data for 1978-2000. Contour plots of the estimated copula between the percentile level in the wage equation and the participation error. Negative correlation indicates positive selection into employment. The first two rows show the Frank copula, while the last two rows show the generalized Frank copula; see Appendix G.

Figure 5: Wage quantiles, by gender (generalized Frank copula)



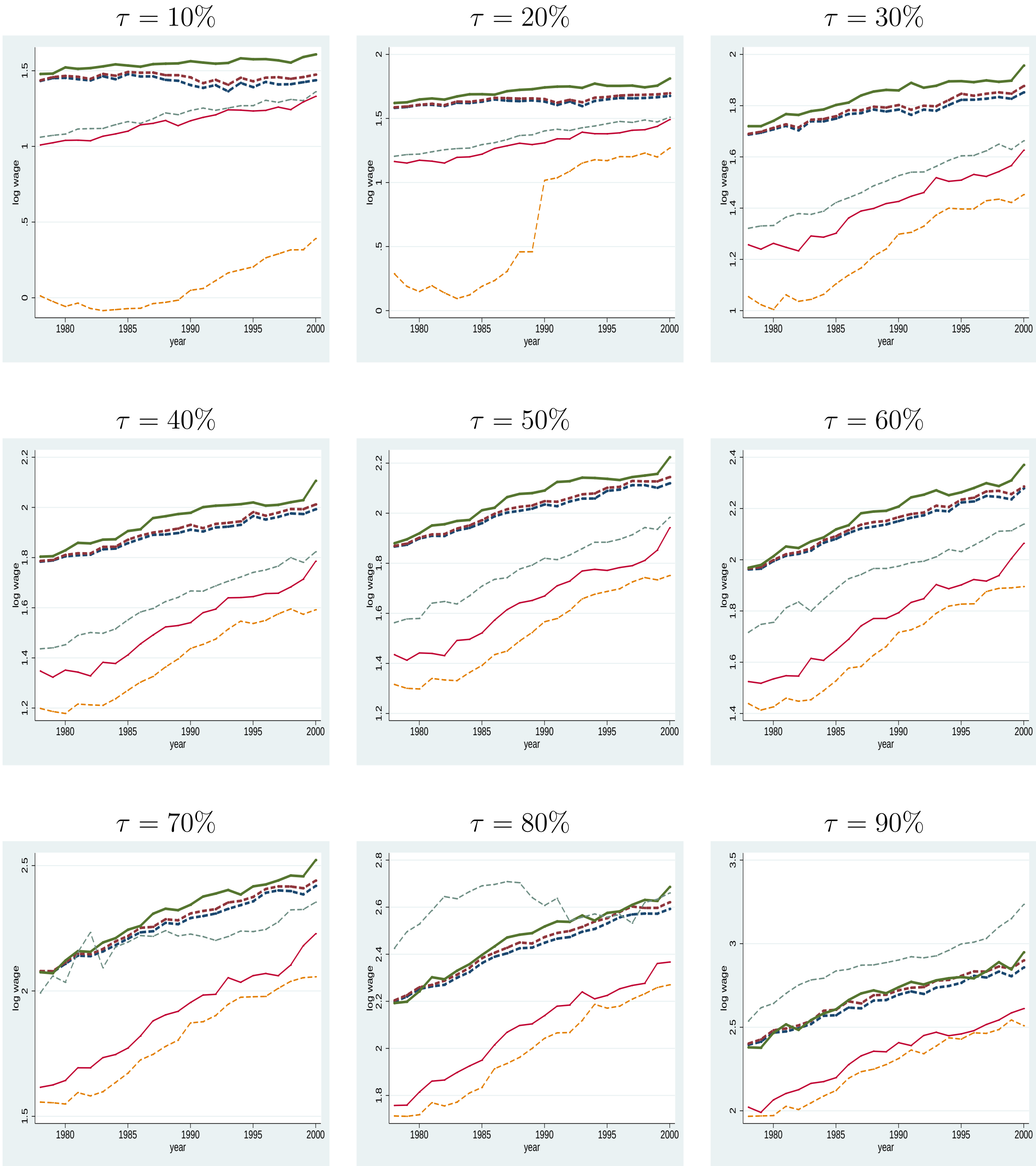
Note: FES data for 1978-2000. Percentiles of log-hourly wages, conditional on employment (solid lines) and corrected for selection (dashed). Male wages are plotted in thick lines (top lines in each graph), while female wages are in thin lines (bottom lines).

Figure 6: Estimated bounds on latent wage quantiles, by gender



Note: FES data for 1978-2000. Estimated bounds on quantiles of log-hourly wages (dashed). The solid lines show the quantiles conditional on employment. Male wages are plotted in thick lines, while female wages are in thin lines.

Figure 7: Estimated bounds on latent wage quantiles, by gender (trimming 1% of extreme observations in out-of-work income)

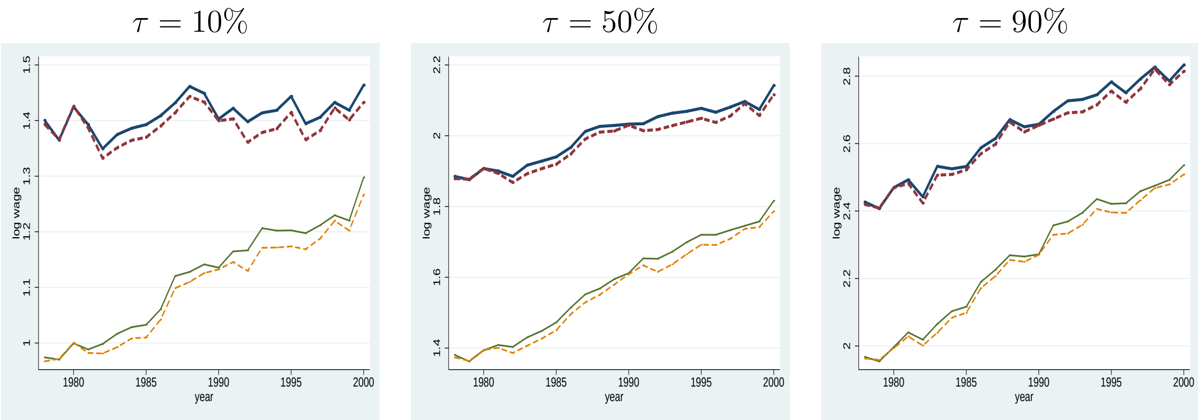


Note: FES data for 1978-2000. Estimated bounds on quantiles of log-hourly wages (dashed). The solid lines show the quantiles conditional on employment. Male wages are plotted in thick lines, while female wages are in thin lines.

Figure 8: Estimated bounds on latent wage quantiles (males), by education (trimming 1% of extreme observations in out-of-work income)

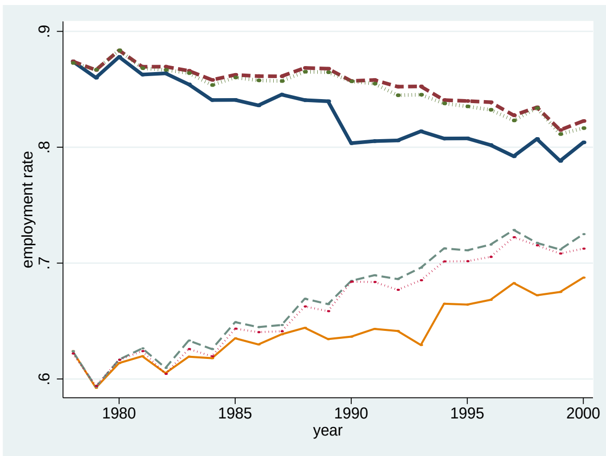
Note: FES data for 1978-2000. Estimated bounds on quantiles of log-hourly wages (dashed). The solid lines show the quantiles conditional on employment. Wages for high-school and college are plotted in thick lines (at the top), while wages for statutory schooling are in thin lines (bottom).

Figure 9: Latent wage quantiles and counterfactual equilibrium latent wage quantiles, by gender



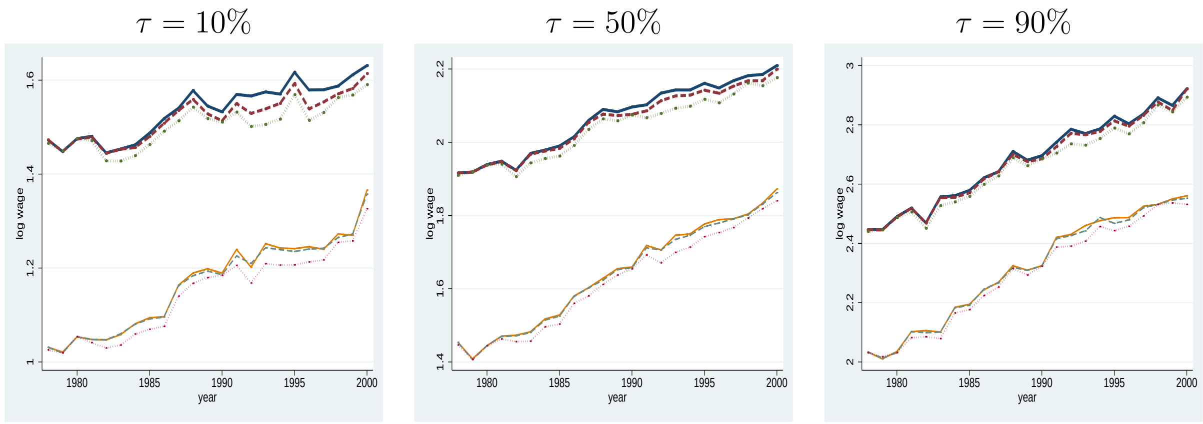
Note: FES data for 1978-2000. Quantiles of log-hourly wages corrected for selection. Latent wage quantiles (solid lines) and counterfactual general equilibrium latent wage quantiles (dashed). Male wages are plotted in thick lines (top lines in each graph), while female wages are in thin lines (bottom lines).

Figure 10: Employment (actual and counterfactual), by gender



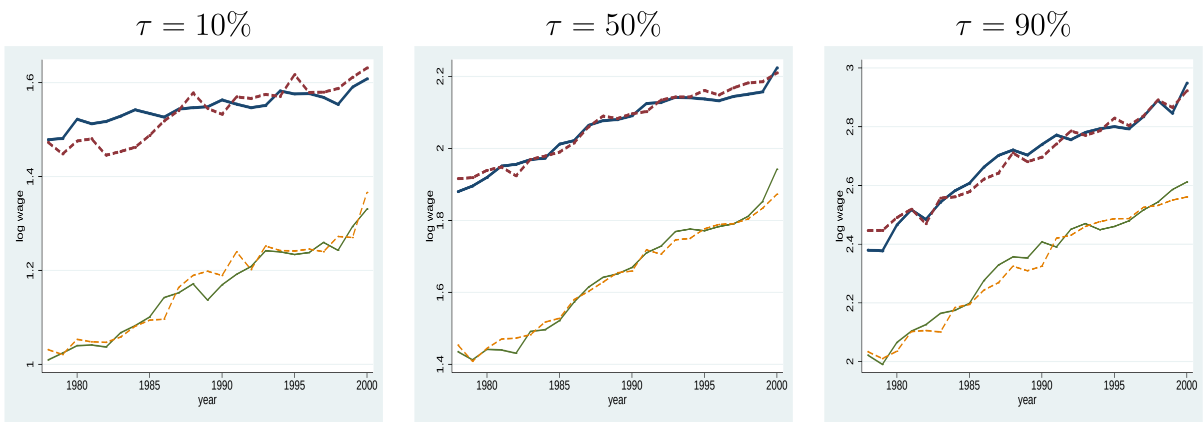
Note: FES data for 1978-2000. Actual employment rate predicted by the model (solid lines), counterfactual employment rate at constant prices (dashed), and counterfactual employment rate at equilibrium prices (dotted). Male employment is plotted in thick lines (top lines), while female employment is in thin lines (bottom lines).

Figure 11: Wage quantiles conditional on employment (actual and counterfactual), by gender



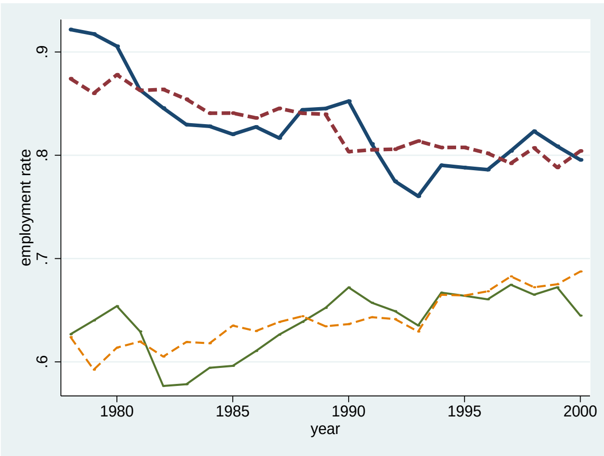
Note: FES data for 1978-2000. Quantiles of log-hourly wages conditional on employment. Actual quantiles predicted by the model (solid lines), counterfactual quantiles in partial equilibrium (dashed), and counterfactual quantiles in general equilibrium (dotted). Male wages are plotted in thick lines (top lines in each graph), while female wages are in thin lines (bottom lines).

Figure 12: Fit to wage quantiles, by gender



Note: FES data for 1978-2000. Specification used in Subsection 4.3. 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.

Figure 13: Fit to employment, by gender



Note: FES data for 1978-2000. Specification used in Subsection 4.3. 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.