# Estimation of Wage Inequality in the UK by Quantile Regression with Censored Selection 

Songnian Chen*, Nianqing Liu ${ }^{\dagger}$, Hanghui Zhang ${ }^{\dagger}$, and Yahong Zhou ${ }^{\ddagger}$

November 12, 2023


#### Abstract

Arellano and Bonhomme (2017) proposed a quantile selection model to study the evolution of wage inequality in the UK, which specifies a binary selection equation and requires an exclusion restriction. In this paper we propose a quantile selection model with a more informative censored selection equation. Following Heckman (1974, 1979), Heckman and Sedlacek (1990), and Blundell et al. (2003), among others, the employment selection equation could be equivalently modelled by an hours worked equation through a censored selection. In our model, both the outcome and selection equations are specified as semiparametric quantile regressions, and no exclusion restriction is needed. We propose a quantile selection estimator that was applied to study wage inequality using the same data as in Arellano and Bonhomme (2017). Among our major findings based on our method, after adjusting for sample selection, (i) there is significant negative selection among males, in contrast to the finding of significant positive selection by Arellano and Bonhomme (2017); (ii) similar to Arellano and Bonhomme (2017), we also find positive selection for females, but our selection effects are more significant than those of Arellano and Bonhomme (2017) (See Section 5 for more details); (iii) the gender wage gap has remained large and accounting for selection leads to much smaller reduction in the gender wage gap over time, compared with the observed wage distribution and that of Arellano and Bonhomme (2017).


Key words: Wage inequality, exclusion restriction, quantile regression, censored selection JEL classification: C14, C31, J31

## 1 Introduction

The quantile regression framework developed by Koenker and Bassett (1978) has become important empirical tools for economic analysis. By allowing varying quantile regression coefficients across different regions of the conditional distribution of the outcome variable, the quantile regression provides a comprehensive characterization of the entire conditional distribution. In labor economics, quantile regression techniques have been extensively applied to study the evolution of wage inequality and wage gaps, see, e.g., Juhn et

[^0]al. (1993), Chamberlain (1994), Buchinsky (1994), Gosling et al. (2000), Blau and Kahn (2017), Maasoumi and Wang (2019), Chernozhukov et al. (2019), among others.

Non-random sample selection has played an important role in economics since the pioneering work of Gronau (1974) and Heckman $(1974,1979)$. For example, for the study of wages and employment, Arellano and Bonhomme (2017) noted that only the wages of employed individuals are observed, so conventional measures of wage gaps or wage inequality may be biased and wage inequality for those at work may provide a distorted picture of market-level wage inequality. To overcome such bias and recover the latent wage distribution, Arellano and Bonhomme (2017) proposed a quantile regression model subject to sample selection where sample selection is modelled via the bivariate cumulative distribution function, or copula, of the errors in the outcome and the selection equations. Furthermore, they proposed a two-step method for the estimation of the copula parameter, and the final step for the estimation for the entire family of quantile regression coefficient process given a consistent estimator for the copula parameter. Arellano and Bonhomme (2017) then applied their method to study the evolution of wage inequality and wage gaps in the UK for the period 1978-2000 using the data from the Family Expenditure Survey (FES). Among their major findings are the strong bias correction effects for males at the bottom of the wage distribution and significant reduction in gender wage gap, especially at the lower part of the wage distribution, compared with the observed wage distribution without correcting for sample selection.

There are two important features associated with the main estimation procedure in Arellano and Bonhomme (2017). First, the binary selection equation is specified as a parametric model in Arellano and Bonhomme's (2017) estimation. ${ }^{1}$ Such a parametric specification is largely due to convenience, which in general is not justified by economic theory. On the other hand, nonparametric treatment of the selection equation is not feasible due to the large number of regressors in most studies on wage gaps and wage inequality, including that of Arellano and Bonhomme (2017).

Another feature of their approach is the requirement of an exclusion restriction that some relevant variables in the selection equation to be excluded from the outcome equation; in particular, they impose the exclusion restriction that the out-of-work benefit income is independent of the unobservables that determine wages. However, Blundell et al. (2007) found evidence against this particular exclusion restriction, and argued that the welfare benefits system in the UK would lead to a positive relationship between wages and this particular excluded variable. Indeed, it is well recognized in the literature that exclusion restrictions typically have little justification in most empirical studies, see, e.g., Krueger and Whitmore (2001), Lee (2009), among others. In a related study, Chernozhukov et al. (2022) also investigated the wage evolution in the UK based on the FES data using distribution regression with sample selection. They also adopted the probit specification for the binary selection equation and relied on the same exclusion restriction.

Issues arising from the above two features can be addressed by a censored selection mechanism, a more informative selection mechanism than the binary selection. In this paper we consider the censored quantile selection model by exploiting extra information provided by the censored selection equation. Specifically,

[^1]following Heckman (1974, 1979), Heckman and Sedlacek (1990), and Blundell et al. (2003), among others, the employment selection equation could be equivalently modelled by an hours worked equation through a censored selection equation when the information on hours worked is available. According to the optimal labor supply model in Blundell, Reed and Stoker (2003) and Blundell, MaCurdy, and Meghir (2007), the structural residuals in the participation and hours equations are the same in those canonical labor supply models. ${ }^{2}$ In practice, the information on hours worked is typically available in common data sets used to study wage inequality, such as the FES and the CPS, our quantile selection model consists of a latent quantile wage equation and a censored hours worked quantile selection equation. Consequently, we consider the estimation of the censored selection model where both the outcome equation and the selection equation are specified as semiparametric quantile regressions, without imposing the exclusion restriction.

Similar to Arellano and Bonhomme (2017), we propose a two-step estimator for the copula parameter. But we differ from Arellano and Bonhomme (2017) in the way the moment conditions are constructed. In particular, for a pair of quantile indices that correspond to the selection equation and outcome equation respectively, we select a subsample for which the moment conditions correspond to a rotated quantile function. Once we obtain an estimate for the copula parameter, the quantile regression coefficients for the outcome equation can be obtained by solving standard quantile regressions. In addition, unlike Arellano and Bonhomme (2017), with censored selection equation, identification based on moment conditions in general does not require an exclusion restriction. This feature makes our approach particularly appealing since excluded variables are generally difficult to find in empirical applications.

We apply our method to study wage inequality in the UK using the same data as in Arellano and Bonhomme (2017). ${ }^{3}$ Among our major findings based on our method, after adjusting for sample selection, (i) there is significant negative selection among males; (ii) our results provide significant correction to females, with the magnitude of corrections two to four times as in Arellano and Bonhomme (2017); (iii) the gender wage gap has remained large and the wage gap reduction is not significant, compared with the observed wage distribution without correcting for selection or the results based on Arellano and Bonhomme (2017). On the other hand, in a survey article, Blau and Khan (2017) noted that the long-term trend has been a substantial reduction in the gender wage gap in advanced nations, whereas Mulligan and Rubinstein (2008) did not find much of a reduction in gender wage gap after selection correction. Regarding the selection pattern, Maasoumi and Wang (2019) noted that it could vary in magnitude and sign over time. Ermisch and Wright (1994) argued that negative selection into employment is very plausible when there is relatively high positive correlation between the wage offer and reservation wage of a potential worker. Negative selection has also been observed by Dolton and Makepeace (1987), Steinberg (1989), Wright and Ermisch (1991), Mulligan and Rubinstein (2008), and Mocan and Unel (2017).

Sample selection model with censored selection was classified by Amemiya (1985) as the type 3 Tobit model. This model has also been studied by Amemiya (1978, 1979), Vella (1993) and Wooldridge (1998) in a parametric context, whereas Lee (1994), Honoré et al. (1997), Chen (1997), Christofides et al. (2003) and Lee and Vella (2006) considered semiparametric estimation. More recently, Fernández-Val et al. (2021) studied nonseparable sample selection models with censored selection rules.

[^2]Maasoumi and Wang (2019) applied Arellano and Bonhomme's (2017) method to study the evolution wage inequality in the U.S. for the period 1976-2013 using the data from the Current Population Survey (CPS). Aside from Arellano and Bonhomme (2017), Buchinsky $(1998,2001)$ also considered quantile regression subject to sample selection, but with an additive structure. In general, however, the additive structure is very restrictive and Buchinsky's quantile regression model does not allow for general heterogeneity. Huber and Melly (2015), on the other hand, focused on testing for the additive structure in a quantile selection model.

The rest of the paper is organized as follows. Section 2 presents the quantile selection model subject to censored quantile selection, discusses model identification and outlines our estimation procedure. ${ }^{4}$ Section 3 presents the large sample properties of our estimator. Simulation results are contained in Section 4. We apply our method to the FES data in Section 5. We then conclude in Section 6. The Appendix contains proofs of the main theorems and the tables for simulation.

## 2 Model, Identification and Estimation

### 2.1 Model

In this paper we consider quantile selection models with nonparametric and semiparametric specification for the outcome equation and the selection equation. In both cases the selection bias is modelled through a parametric copula function.

The quantile selection model with nonparametric specification for the outcome and selection equations take the following form:

$$
\begin{align*}
Y_{1}^{*} & =q_{1}(U, X)  \tag{2.1}\\
Y_{2} & =\max \left\{Y_{2}^{*}, 0\right\}=\max \left\{q_{2}(V, X), 0\right\}  \tag{2.2}\\
D & =1\left\{Y_{2}^{*}>0\right\} \tag{2.3}
\end{align*}
$$

where $Y_{1}^{*}$ and $Y_{2}^{*}$ are the latent outcome variables (e.g, market wage and hours worked), $D$ is the participation indicator (employment), $X$ contains observed characteristics, $U$ and $V$ are unobserved characteristics with standard uniform marginals. We observe $\left(Y_{1}, Y_{2}, D, X\right)$ where $Y_{1}=D Y_{1}^{*}$, so that the potential outcome $Y_{1}^{*}$ is observed when $D=1$, or equivalently $Y_{2}^{*}>0$; see $\operatorname{Heckman}(1974,1979)$, Heckman and Sedlacek (1990), and Blundell et al. (2003), for more discussions. Note that the participation is modelled through a censored regression and consequently we do not need to impose the usual exclusion restriction, unlike Arellano and Bonhomme (2017). However, we assume a parametric specification on the copula function for the joint distribution of $(U, V)$.

For the case with semiparametric specification for the outcome equation and the selection equations, we consider the following model:

[^3]\[

$$
\begin{align*}
& Y_{1}^{*}=X^{\prime} \beta(U)  \tag{2.4}\\
& Y_{2}=\max \left\{Y_{2}^{*}, 0\right\}=\max \left\{X^{\prime} \gamma(V), 0\right\}  \tag{2.5}\\
& D=1\left\{Y_{2}^{*}>0\right\} \tag{2.6}
\end{align*}
$$
\]

where both the outcome and selection equations follow a linear-in-parameter structure for the quantile regression functions.

### 2.2 Identification

In this subsection, we study identification of the models stated above. We first consider nonparametric identification of the outcome equation. We do not rely on identification at infinity or the exclusion restriction. We make the following assumption:

Assumption 1: $\left\{Y_{1 i}, Y_{2 i}, D_{i}, X_{i}, U_{i}, V_{i}\right\}_{i=1}^{n}$ is a random sample generated from (2.1-2.3).
A1 (Unobservables) The bivariate distribution of $(U, V)$ given $X=x$ is absolutely continuous with respect to the Lebesgue measure, with standard uniform marginals and rectangular support. We denote its cumulative distribution function (c.d.f.) as $C_{x}^{*}\left(u, v, \rho_{0}\right)$ for some $\rho_{0}$ with finite dimension.

A2 (Continuous outcomes) The conditional distributions $F_{Y_{1}^{*} \mid X}\left(y_{1} \mid x\right), F_{Y_{2}^{*} \mid X}\left(y_{2} \mid x\right)$ and their inverses are strictly increasing for any given $x$. In addition, $C_{x}^{*}\left(u, v, \rho_{0}\right)$ is strictly increasing in $u$ and $v$.

A3 (Propensity score) $p(X) \equiv \operatorname{Pr}(D=1 \mid X)>0$ with probability 1.
Let $\tau_{x}$ denote $F_{Y_{2}^{*} \mid X}(0 \mid x)$ for $x \in \mathcal{X}$, where $\mathcal{X}$ denotes the support of $X$.
Proposition 1: Let Assumption 1 hold. Also, suppose that (i) there exists some $x_{0} \in \mathcal{X}$ such that $\tau_{x_{0}}=\inf _{x \in \mathcal{X}} \tau_{x}$; (ii) for $\tau, \tilde{\tau} \in(0,1)$ and some $\tilde{\rho}, C^{*}\left(\tau, \tau_{0}, \rho_{0}\right)=C^{*}\left(\tilde{\tau}, \tau_{0}, \tilde{\rho}\right)$ holds for $\tau_{0} \geq \tau_{0 x_{0}}$ if and only if $\tau=\tilde{\tau}$ and $\rho_{0}=\tilde{\rho}$, then $q_{1}(x, \tau)$ is identified for $\tau \in(0,1)$.

Proof: Note that for any $\tau_{0} \in(0,1)$ and $x \in \mathcal{X}$, if $q_{2}\left(\tau_{0}, x\right)>0$, then $q_{2}\left(\tau_{0}, x\right)$ can be identified since for such $\tau_{0}$ the $\tau_{0}$ th conditional quantiles of $Y_{2}$ and $Y_{2}^{*}$ given $X=x$ coincide, and in particular, we have

$$
P\left(Y_{1}<t_{1}, Y_{2}<q_{2}\left(\tau_{0}, x\right) \mid X=x\right)=C_{x}^{*}\left(F_{Y_{1} \mid X}\left(t_{1} \mid x\right), \tau_{0}, \rho_{0}\right)
$$

for any $t_{1} \in \mathbb{R}$. Define

$$
I(F, \rho)=\iint_{\tau_{x}}^{1} \int_{-\infty}^{+\infty}\left[C_{x}^{*}\left(F_{Y_{1} \mid X}\left(t_{1} \mid x\right), \tau_{0}, \rho_{0}\right)-C_{x}^{*}\left(F\left(t_{1} \mid x\right), \tau_{0}, \rho\right)\right]^{2} 1\left\{q_{2}\left(\tau_{0}, x\right)>0\right\} d t_{1} d \tau_{0} d F_{X}(x)
$$

where $F_{X}(\cdot)$ denotes the distribution of $X$.
Note that $I\left(F_{Y_{1} \mid X}(\cdot \mid \cdot), \rho_{0}\right)=0$. Suppose $I(\tilde{F}, \tilde{\rho})=0$ for some $(\tilde{F}, \tilde{\rho})$. Then, we have

$$
C_{x}^{*}\left(F_{Y_{1} \mid X}\left(t_{1} \mid x\right), \tau_{0}, \rho_{0}\right)=C_{x}^{*}\left(\tilde{F}_{Y_{1} \mid X}\left(t_{1} \mid x\right), \tau_{0}, \tilde{\rho}\right)
$$

for all $t_{1} \in \mathbb{R}$, and $x \in \mathcal{X}$, if $q_{2}\left(\tau_{0}, x\right)>0$. In particular, for $x=x_{0}$, we have

$$
C_{x}^{*}\left(F_{Y_{1} \mid X}\left(t_{1} \mid x_{0}\right), \tau_{0}, \rho_{0}\right)=C_{x}^{*}\left(\tilde{F}_{Y_{1} \mid X}\left(t_{1} \mid x_{0}\right), \tau_{0}, \tilde{\rho}\right)
$$

for all $t_{1} \in \mathbb{R}$ and $\tau_{0}>\tau_{0 x_{0}}$. Then Condition (ii) implies that $\tilde{\rho}=\rho_{0}$ and $\tilde{F}_{Y_{1} \mid X}\left(t_{1} \mid x_{0}\right)=F_{Y_{1} \mid X}\left(t_{1} \mid x_{0}\right)$. As a result, once $\rho_{0}$ is identified, we have, for any $x \in \mathcal{X}$,

$$
C_{x}^{*}\left(F_{Y_{1} \mid X}\left(t_{1} \mid x\right), \tau_{0}, \rho_{0}\right)=C_{x}^{*}\left(\tilde{F}_{Y_{1} \mid X}\left(t_{1} \mid x\right), \tau_{0}, \rho_{0}\right)
$$

for all $t_{1} \in \mathbb{R}$ and $\tau_{0}>\tau_{0 x}$. Then, from Assumption A2, we can deduce

$$
F_{Y_{1} \mid X}\left(t_{1} \mid x\right)=\tilde{F}_{Y_{1} \mid X}\left(t_{1} \mid x\right)
$$

for all $t_{1} \in \mathbb{R}$, which implies that $q_{1}(\tau, x)=\tilde{q}_{1}(\tau, x)$, for all $\tau \in(0,1)$ and $x \in \mathcal{X}$ where $\tilde{q}_{1}(\tau, x)=$ $\tilde{F}_{Y_{1} \mid X}^{-1}(\tau \mid x)$.

Remark 1: Proposition 1 provides an identification result with nonparametrically specified outcome and selection equations, but with a parametric specification for the copula function. On the other hand, by imposing an exclusion restriction, Arellano and Bonhomme (2017) considered the identification with nonparametric specification for the outcome and selection equations as well as the copula function, by relying on identification at infinity or analytic extrapolation.

Remark 2: Our identification strategy, in spirit, follows the common identification strategy used for nonlinear regression models. Essentially we view $C_{x}^{*}\left(\tau, \tau_{0}, \rho\right)$ as a function of $\tau_{0}$ indexed by $(\tau, \rho)$. Let $\tilde{\theta}=(\tilde{\tau}, \tilde{\rho})$ and $\theta^{*}=\left(\tau^{*}, \rho^{*}\right)$. Also, let $C_{x}^{*}\left(\tilde{\tau}, \tau_{0}, \tilde{\rho}\right)=\bar{C}_{x}\left(\tilde{\theta}, \tau_{0}\right)$ and $C_{x}^{*}\left(\tau^{*}, \tau_{0}, \rho^{*}\right)=\bar{C}_{x}\left(\theta^{*}, \tau_{0}\right)$. Then the Condition (ii) in Proposition 1 basically states that $\bar{C}_{x}\left(\theta^{*}, \tau_{0}\right)=\bar{C}_{x}\left(\tilde{\theta}, \tau_{0}\right)$ holds for $\tau_{0} \geq \tau_{0 x}$ if and only if $\theta^{*}=\tilde{\theta}$.

Remark 3: Note that for the quantile selection model with a binary selection equation, we can only recover $F_{Y_{2} \mid X}(0 \mid x)$ and $C_{x}^{*}\left(F_{Y_{1} \mid X}\left(t_{1} \mid x\right), F_{Y_{2} \mid X}(0 \mid x), \rho_{0}\right)$, for $t_{1} \in \mathbb{R}$ and $x \in \mathcal{X}$. Without an exclusion restriction, lack of variation of $C_{x}^{*}\left(F_{Y_{1} \mid X}\left(t_{1} \mid x\right), F_{Y_{2} \mid X}(0 \mid x), \rho_{0}\right)$ prevents the identification of $F_{Y_{1} \mid X}\left(t_{1} \mid x\right)$ and $\rho_{0}$; in contrast, for the quantile selection model with a censored selection, we are able to recover $C_{x}^{*}\left(F_{Y_{1} \mid X}\left(t_{1} \mid x\right), \tau_{0}, \rho_{0}\right)$ for $t_{1} \in \mathbb{R}$ and $\tau_{0} \geq \tau_{0 x}$, and consequently the variation of $C_{x}^{*}\left(F_{Y_{1} \mid X}\left(t_{1} \mid x\right), \tau_{0}, \rho_{0}\right)$ in $\tau_{0}$ and the parametric structure of the copula function provides important information for the identification of $F_{Y_{1} \mid X}\left(t_{1} \mid x\right)$ and $\rho_{0}$, for $t_{1} \in \mathbb{R}$ and $x \in \mathcal{X}$, without resorting to the exclusion restriction.

We now consider the semiparametric case. We make some slight adjustment to Assumption 1.
Assumption 1': $\left\{Y_{1 i}, Y_{2 i}, D_{i}, X_{i}, U_{i}, V_{i}\right\}_{i=1}^{n}$ is a random sample generated from (2.4-2.6).
A1 (Unobservables) The bivariate distribution of $(U, V)$ given $X=x$ is absolutely continuous with respect to the Lebesgue measure, with standard uniform marginals and rectangular support. We denote its cumulative distribution function (c.d.f.) as $C^{*}\left(u, v, \rho_{0}\right)$.

A2 (Continuous outcomes) The conditional distributions $F_{Y_{1}^{*} \mid X}\left(y_{1} \mid x\right), F_{Y_{2}^{*} \mid X}\left(y_{2} \mid x\right)$ and their inverses are strictly increasing for any given $x$. In addition, $C^{*}\left(u, v, \rho_{0}\right)$ is strictly increasing in $u$ and $v$.

A3 (Propensity score) $p(X) \equiv \operatorname{Pr}(D=1 \mid X)>0$ with probability 1.
Note that compared with Assumption 1, we assume that the copula function is independent of the regressors. ${ }^{5}$ Define $C\left(\tau, \tau_{0}, \rho_{0}\right)=\frac{C^{*}\left(\tau, 1, \rho_{0}\right)-C^{*}\left(\tau, \tau_{0}, \rho_{0}\right)}{1-\tau_{0}}$.

[^4]Proposition 2: Let Assumption 1' hold. In addition, suppose that (i) $\gamma\left(\tau_{0}\right)$ is identified for any $\tau_{0} \in$ $\left(\tau_{01}, \tau_{02}\right)$ with $E\left[1\left\{X^{\prime} \gamma\left(\tau_{01}\right)>0\right\} X X^{\prime}\right]$ being nonsingular; (ii) $C\left(\tau, \tau_{0}, \rho_{0}\right)=C\left(\bar{\tau}, \tau_{0}, \bar{\rho}\right)$ holds for $\tau_{0} \in$ $\left(\tau_{01}, \tau_{02}\right)$ if and only if $(\bar{\tau}, \bar{\rho})=\left(\tau, \rho_{0}\right)$; then $\beta(\tau)$ is identified.

Proof: Define for $\tau \in(0,1)$,

$$
I(b, \tau, \rho)=\int_{\tau_{01}}^{\tau_{02}} E\left[1\left\{X^{\prime} \gamma\left(\tau_{0}\right)>0\right\}\left\{E\left[D 1\left\{Y_{2}>X^{\prime} \gamma\left(\tau_{0}\right)\right\} 1\left\{Y_{1}<X^{\prime} b\right\} \mid X\right]-\left(1-\tau_{0}\right) C\left(\tau, \tau_{0}, \rho\right)\right\}\right]^{2} d \tau_{0}
$$

It is straightforward to show that $I\left(\beta(\tau), \tau, \rho_{0}\right)=0$. Now suppose $I(b, \tau, \rho)=0$ for some $(b, \rho)$, then we can show

$$
\left(1-\tau_{0}\right)^{2} \int_{\tau_{01}}^{\tau_{02}} E\left\{1\left\{X^{\prime} \gamma\left(\tau_{0}\right)>0\right\}\left[C\left(F_{Y_{1}^{*} \mid X}\left(X^{\prime} b \mid X\right), \tau_{0}, \rho_{0}\right)-C\left(\tau, \tau_{0}, \rho\right)\right]^{2}\right\} d \tau_{0}=0
$$

and in particular, for any $x$ such that $x^{\prime} \gamma\left(\tau_{0}\right)>0$, we have

$$
C\left(F_{Y_{1}^{*} \mid X}\left(X^{\prime} b \mid x\right), \tau_{0}, \rho_{0}\right)=C\left(\tau, \tau_{0}, \rho\right)
$$

for $\tau_{0} \in\left[\tau_{01}, \tau_{02}\right]$. Then by Assumption (ii) we can deduce that $\rho=\rho_{0}$ and $x^{\prime} b=x^{\prime} \beta(\tau)$, which, in turn, implies that $b=\beta(\tau)$ if $E\left[1\left\{X^{\prime} \gamma\left(\tau_{0}\right)>0\right\} X X^{\prime}\right]$ is nonsingular.

Similar to Proposition 1, here we adopt a nonlinear least squares type identification strategy.
Remark 4: Similar to the discussion in Remark 2, write $C\left(\tau, \tau_{0}, \rho\right)$ as $g\left(\tau_{0}, \theta\right)$, which is a function of $\tau_{0}$ with a parameter $\theta$, where $\theta=(\tau, \rho)$, then Condition (ii) in Proposition 2 states that $g\left(\tau_{0}, \theta\right)=g\left(\tau_{0}, \theta^{*}\right)$ for $\tau_{0} \in\left(\tau_{01}, \tau_{02}\right)$, if and only if $\theta=\theta^{*}$, where $\theta^{*}=(\bar{\tau}, \bar{\rho})$. This is a common identification strategy for nonlinear regression analysis. See Section 2.2.2 in Newey and McFadden (1994) for more details.

Remark 5: The main arguments of Proposition 2 are essentially based on the following conditional moments:

$$
E\left[D \cdot 1\left\{Y_{2}>X^{\prime} \gamma\left(\tau_{0}\right)\right\} 1\left\{Y_{1}<X^{\prime} \beta(\tau)\right\} \mid X, X^{\prime} \gamma\left(\tau_{0}\right)>0\right]-\left(1-\tau_{0}\right) \cdot C\left(\tau, \tau_{0}, \rho_{0}\right)=0
$$

for $\tau \in(0,1)$ and $\tau_{0} \in\left(\tau_{01}, \tau_{02}\right)$. On the other hand, it is well known conditional moment restrictions are equivalent to infinite number of unconditional moment restrictions. In practice, for estimation purpose we could adopt increasing number of moment conditions as the sample size increases to take full advantage of the conditional moments. However, for ease of empirical implementation, in the next subsection we consider semiparametric estimation based on a finite number of unconditional moments, similar to Arellano and Bonhomme (2017, 2017S).

### 2.3 Estimation

We now consider the estimation of the semiparametric model $(2.4-2.6)$. To construct appropriate moment conditions, define

$$
D_{2}\left(\tau_{0}\right)=1\left\{Y_{2}>X^{\prime} \gamma\left(\tau_{0}\right)\right\}
$$

for any $\tau_{0} \in(0,1)$. Note that

$$
D_{2}\left(\tau_{0}\right)=1\left\{Y_{2}^{*}>X^{\prime} \gamma\left(\tau_{0}\right)\right\}=1\left\{V>\tau_{0}\right\}
$$

if $X^{\prime} \gamma\left(\tau_{0}\right)>0$. Let $\left[\tau_{l}, \tau_{u}\right]$ be a range of $\tau_{0}$ for which we can obtain reasonably precise estimates for $\gamma\left(\tau_{0}\right)$. Then, for those observations with $X^{\prime} \gamma\left(\tau_{0}\right)>0$, the subsample selection indicator $D_{2}$ no longer depends on regressors, unlike $D$. Appropriate moment conditions are constructed based on this observation. Note that for the observation with $X^{\prime} \gamma\left(\tau_{0}\right)>0$, for any pair of $\tau$ and $\tau_{0}$, we have

$$
\begin{aligned}
& D_{2}\left(\tau_{0}\right) 1\left\{Y_{1}<X^{\prime} \beta(\tau)\right\} \\
& =1\left\{Y_{1}^{*}<X^{\prime} \beta(\tau), Y_{2}^{*}>X^{\prime} \gamma\left(\tau_{0}\right)\right\} \\
& =1\left\{X^{\prime} \beta(U)<X^{\prime} \beta(\tau), X^{\prime} \gamma(V)>X^{\prime} \gamma\left(\tau_{0}\right)\right\} \\
& =1\left\{U<\tau, V>\tau_{0}\right\}
\end{aligned}
$$

which, in turn, yields the following conditional moment conditions,

$$
\begin{equation*}
E\left[D_{2}\left(\tau_{0}\right) 1\left\{X^{\prime} \gamma\left(\tau_{0}\right)>0\right\}\left(1\left\{Y_{1}<X^{\prime} \beta(\tau)\right\}-C\left(\tau, \tau_{0}, \rho_{0}\right)\right) \mid X\right]=0 \tag{2.7}
\end{equation*}
$$

In particular, given $\gamma\left(\tau_{0}\right)$, we work with the following unconditional moment conditions for the estimation of $\beta(\tau)$ and $\rho_{0}$ :

$$
\begin{align*}
& E g_{1}\left(\xi_{i}, b, \gamma\left(\tau_{0}\right), \rho, \tau, \tau_{0}\right)=0  \tag{2.8}\\
& E g_{2}\left(\xi_{i}, b, \gamma\left(\tau_{0}\right), \rho, \tau, \tau_{0}\right)=0 \tag{2.9}
\end{align*}
$$

where

$$
g_{1}\left(\xi_{i}, b, \gamma, \rho, \tau, \tau_{0}\right)=1\left\{X_{i}^{\prime} \gamma>0\right\} 1\left\{Y_{2 i}>X_{i}^{\prime} \gamma\right\}\left(1\left\{Y_{1 i}<X_{i}^{\prime} b\right\}-C\left(\tau, \tau_{0}, \rho\right)\right) X_{i}
$$

and

$$
g_{2}\left(\xi_{i}, b, \gamma, \rho, \tau, \tau_{0}\right)=1\left\{X_{i}^{\prime} \gamma>0\right\} 1\left\{Y_{2 i}>X_{i}^{\prime} \gamma\right\}\left(1\left\{Y_{1 i}<X_{i}^{\prime} b\right\}-C\left(\tau, \tau_{0}, \rho\right)\right) \varphi\left(X_{i}\right)
$$

with $\xi_{i}=\left(X_{i}, Y_{1 i}, Y_{2 i}\right)$ and $\varphi$ is some instrumental function of $X$. Also note that (2.8) can be viewed as the first order condition for the following minimization problem:

$$
\min _{b \in B} E q_{1}\left(\xi_{i}, b, \gamma\left(\tau_{0}\right), \rho, \tau, \tau_{0}\right)
$$

where

$$
\begin{aligned}
q_{1}\left(\xi_{i}, b, \gamma, \rho, \tau, \tau_{0}\right) & =1\left\{X_{i}^{\prime} \gamma>0\right\} 1\left\{Y_{2 i}>X_{i}^{\prime} \gamma\right\} \\
& \times\left[C\left(\tau, \tau_{0}, \rho\right)\left(Y_{1 i}-X_{i}^{\prime} b\right)^{+}+\left(1-C\left(\tau, \tau_{0}, \rho\right)\right)\left(Y_{1 i}-X_{i}^{\prime} b\right)^{-}\right]
\end{aligned}
$$

Based on the above observation, we are now ready to propose a two-step estimator for $\rho_{0}$. We replace $\gamma\left(\tau_{0}\right)$ for $\tau_{0} \in(0,1)$, by some existing estimator $\hat{\gamma}\left(\tau_{0}\right)$ such as Powell (1986) and Chen (2018). Let $\mathcal{J}_{0}$ and $\mathcal{J}_{1}$ be two sets of quantile indices such that $\tau_{0} \in \mathcal{J}_{0}$ and $\tau \in \mathcal{J}_{1}$ then $\tau_{0} \in\left[\tau_{l}, \tau_{u}\right]$ and $\tau \in\left[\tau_{l}^{*}, \tau_{u}^{*}\right]$.

Step 1: For a given $\rho, \tau_{0} \in \mathcal{J}_{0}$ and $\tau \in \mathcal{J}_{1}$, define $\hat{\beta}\left(\tau, \tau_{0}, \rho\right)$ as a solution to the minimization problem,

$$
\min _{b \in B} Q_{1 n}\left(b, \hat{\gamma}\left(\tau_{0}\right), \rho, \delta_{n}, \tau, \tau_{0}\right)
$$

where

$$
Q_{1 n}\left(b, \gamma, \rho, \delta, \tau, \tau_{0}\right)=\frac{1}{n} \sum_{i=1}^{n} q_{1}\left(\xi_{i}, b, \gamma, \rho, \delta, \tau, \tau_{0}\right)
$$

and here, with a slight abuse of notation, we define

$$
\begin{aligned}
q_{1}\left(\xi_{i}, b, \gamma, \rho, \delta, \tau, \tau_{0}\right) & =1\left\{X_{i}^{\prime} \gamma>\delta\right\} 1\left\{Y_{2 i}>X_{i}^{\prime} \gamma\right\} \\
& \times\left[C\left(\tau, \tau_{0}, \rho\right)\left(Y_{1 i}-X_{i}^{\prime} b\right)^{+}+\left(1-C\left(\tau, \tau_{0}, \rho\right)\right)\left(Y_{1 i}-X_{i}^{\prime} b\right)^{-}\right]
\end{aligned}
$$

Here $\delta_{n}$ is a sequence of positive numbers converging to 0 . Such a sequence is adopted to avoid possible boundary issues. For any $\tau \in \mathcal{J}_{1}$, once $\hat{\beta}\left(\tau, \tau_{0}, \rho\right)$ is available for $\tau_{0} \in \mathcal{J}_{0}$, we define

$$
\hat{\beta}(\tau, \rho)=\frac{1}{\# \mathcal{J}_{0}} \sum_{\tau_{0} \in \mathcal{J}_{0}} \hat{\beta}\left(\tau, \tau_{0}, \rho\right) .
$$

Step 2: Define

$$
G_{2 n}\left(b, \gamma, \rho, \delta, \tau, \tau_{0}\right)=\frac{1}{n} \sum_{i=1}^{n} g_{2}\left(\xi_{i}, b, \gamma, \rho, \delta, \tau, \tau_{0}\right)
$$

where

$$
g_{2}\left(\xi_{i}, b, \gamma, \rho, \delta, \tau, \tau_{0}\right)=1\left\{X_{i}^{\prime} \gamma>\delta\right\} 1\left\{Y_{2 i}>X_{i}^{\prime} \gamma\right\}\left(1\left\{Y_{1 i}<X_{i}^{\prime} b\right\}-C\left(\tau, \tau_{0}, \rho\right)\right) \varphi\left(X_{i}\right)
$$

Now, given $\hat{\beta}(\tau, \rho)$ for $\tau \in \mathcal{J}_{1}$, we estimate $\rho_{0}$ by $\hat{\rho}$, which solves

$$
\min _{\rho \in \varrho} G_{2 n}(\rho)
$$

where

$$
G_{2 n}(\rho)=\sum_{\tau_{0} \in \mathcal{J}_{0}, \tau \in \mathcal{J}_{1}}\left\|G_{2 n}\left(\hat{\beta}(\tau, \rho), \hat{\gamma}\left(\tau_{0}\right), \rho, \delta_{n}, \tau, \tau_{0}\right)\right\|
$$

and $\|\cdot\|$ denotes the Euclidean norm.
Finally, given our estimator for $\rho_{0}, \hat{\rho}$, for any $\tau \in(0,1)$, we can estimate $\beta(\tau)$ by $\hat{\beta}(\tau, \hat{\rho})$, where

$$
\hat{\beta}(\tau, \hat{\rho})=\frac{1}{\# \mathcal{J}_{0}} \sum_{\tau_{0} \in \mathcal{J}_{0}} \hat{\beta}\left(\tau, \tau_{0}, \hat{\rho}\right)
$$

with $\hat{\beta}\left(\tau, \tau_{0}, \hat{\rho}\right)$ being defined as in Step 1.
Remark 6: Note that in estimating $\beta\left(\tau, \tau_{0}, \rho\right)$ we would use the subsample $1\left\{X^{\prime} \gamma\left(\tau_{0}\right)>0\right\}$ when $\gamma\left(\tau_{0}\right)$ is known. When we replace $\gamma\left(\tau_{0}\right)$ by $\hat{\gamma}\left(\tau_{0}\right)$, we make some slight adjustment using $1\left\{X^{\prime} \hat{\gamma}\left(\tau_{0}\right)>\delta_{n}\right\}$ to guarantee that with large probability $1\left\{X^{\prime} \gamma\left(\tau_{0}\right)>0\right\}$ holds for the selected subsample.

## 3 Large Sample Properties

This section provides the large sample properties of our estimator. We make following assumptions.
Assumption A.1: $\left\{X_{i}, Y_{1 i}, Y_{2 i}, U_{i}, V_{i}\right\}_{i=1}^{n}$ is a random sample generated from equations (2.4-2.6) .
Assumption A.2: $(U, V)$ is jointly statistically independent of $X$, and the bivariate distribution of $(U, V)$, $C^{*}\left(u, v, \rho_{0}\right)$, is absolutely continuous with respect to the Lebesgue measure with standard uniform marginals and rectangular support. In addition, $C^{*}\left(u, v, \rho_{0}\right)$ is strictly increasing in $u$ and $v$.

Assumption A.3: The conditional density of $\left(Y_{1}^{*}, Y_{2}^{*}\right)$ given $X=x, f\left(y_{1}, y_{2} \mid x\right)$ is continuously differentiable in $\left(y_{1}, y_{2}\right)$ with positive density on $\mathbb{R}^{2}$.

Assumption A.4: $\beta(\tau), \gamma\left(\tau_{0}\right)$ and $\rho_{0}$ are interior point of a compact set $B \times \Gamma \times \varrho$, respectively, for any $\tau_{0} \in\left[\tau_{l}, \tau_{u}\right]$ and $\tau \in\left[\tau_{l}^{*}, \tau_{u}^{*}\right]$.

Assumption A.5: $\hat{\gamma}\left(\tau_{0}\right)$ satisfies

$$
\max _{\tau_{0} \in\left[\tau_{\tau}, \tau_{u}\right]}\left|\hat{\gamma}\left(\tau_{0}\right)-\gamma_{0}\left(\tau_{0}\right)\right|=O\left(n^{-1 / 2} \log \log n\right)
$$

almost surely and uniformly over $\tau_{0} \in\left[\tau_{l}, \tau_{u}\right]$ and

$$
\sqrt{n}\left(\hat{\gamma}\left(\tau_{0}\right)-\gamma_{0}\left(\tau_{0}\right)\right)=\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{\gamma_{i}}\left(\tau_{0}\right)+o_{p}(1)
$$

with $E \phi_{\gamma i}\left(\tau_{0}\right)=0$.
Assumption A.6: $\delta_{n} \propto n^{-t}$ for some $t<1 / 2$.
Assumption A.7: the matrices $E\left[1\left\{X^{\prime} \gamma\left(\tau_{0}\right)>0\right\} 1\left\{V>\tau_{0}\right\} f_{1}\left(X^{\prime} \beta(\tau) \mid X\right) X X^{\prime}\right]$ are positive definite for all $\tau \in\left[\tau_{l}^{*}, \tau_{u}^{*}\right], \tau_{0} \in \mathcal{J}_{0}$, where $f_{1}\left(y_{1} \mid x\right)$ denotes the conditional density function of $Y_{1}^{*}$ at $y_{1}$ given $X=x$; In addition,

$$
\lim _{\varepsilon \rightarrow 0} \sup _{\tau_{0} \in\left[\tau_{1}, \tau_{u}\right]} \operatorname{Pr}\left(\left|X^{\prime} \gamma\left(\tau_{0}\right)\right|<\varepsilon\right)=0
$$

Let

$$
\bar{G}_{2}(\rho)=\sum_{\tau_{0} \in \mathcal{J}_{0}, \tau \in \mathcal{J}_{1}}\left\|E\left[G_{2 n}\left(\beta(\tau, \rho), \gamma\left(\tau_{0}\right), \rho, 0, \tau_{0}, \tau\right)\right]\right\|
$$

Assumption A.8: $\rho_{0}$ is the unique minimizer of $\bar{G}_{2}(\rho)$.
Define

$$
\begin{aligned}
& \partial G_{1 \beta}\left(\tau, \tau_{0}, \rho\right)=\frac{\partial}{\partial b} \bar{G}_{1}\left(\beta\left(\tau, \tau_{0}, \rho\right), \gamma\left(\tau_{0}\right), \rho, \tau, \tau_{0}\right) \\
& \partial G_{1 \gamma}\left(\tau, \tau_{0}, \rho\right)=\frac{\partial}{\partial \gamma} \bar{G}_{1}\left(\beta(\tau), \gamma\left(\tau_{0}\right), \rho, \tau, \tau_{0}\right) \\
& \partial G_{1 \rho}\left(\tau, \tau_{0}, \rho\right)=\frac{\partial}{\partial \rho} \bar{G}_{1}\left(\beta(\tau), \gamma\left(\tau_{0}\right), \rho, \tau, \tau_{0}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
& \partial G_{2 \beta}\left(\tau, \tau_{0}, \rho\right)=\frac{\partial}{\partial b} \bar{G}_{2}\left(\beta\left(\tau, \tau_{0}, \rho\right), \gamma\left(\tau_{0}\right), \rho, \tau, \tau_{0}\right) \\
& \partial G_{2 \gamma}\left(\tau, \tau_{0}, \rho\right)=\frac{\partial}{\partial \gamma} \bar{G}_{2}\left(\beta(\tau), \gamma\left(\tau_{0}\right), \rho, \tau, \tau_{0}\right) \\
& \partial G_{2 \rho}\left(\tau, \tau_{0}, \rho\right)=\frac{\partial}{\partial \rho} \bar{G}_{2}\left(\beta(\tau), \gamma\left(\tau_{0}\right), \rho, \tau, \tau_{0}\right)
\end{aligned}
$$

where

$$
\bar{G}_{1}\left(b, \gamma\left(\tau_{0}\right), \rho, \tau, \tau_{0}\right)=E\left[1\left\{X_{i}^{\prime} \gamma\left(\tau_{0}\right)>0\right\} 1\left\{Y_{2 i}^{*}>X_{i}^{\prime} \gamma\left(\tau_{0}\right)\right\}\left(1\left\{Y_{1 i}<X_{i}^{\prime} b\right\}-C\left(\tau, \tau_{0}, \rho\right)\right) X_{i}\right]
$$

and

$$
\bar{G}_{2}\left(b, \gamma\left(\tau_{0}\right), \rho, \tau, \tau_{0}\right)=E\left[1\left\{X_{i}^{\prime} \gamma\left(\tau_{0}\right)>0\right\} 1\left\{Y_{2 i}^{*}>X_{i}^{\prime} \gamma\left(\tau_{0}\right)\right\}\left(1\left\{Y_{1 i}<X_{i}^{\prime} b\right\}-C\left(\tau, \tau_{0}, \rho\right)\right) \varphi\left(X_{i}\right)\right]
$$

In addition, let

$$
L_{2}\left(\tau_{0}, \tau\right)=\partial G_{2 \beta}\left(\tau, \tau_{0}, \rho_{0}\right) \frac{\partial}{\partial \rho} \beta\left(\tau, \rho_{0}\right)+\partial G_{2 \rho}\left(\tau, \tau_{0}, \rho_{0}\right)
$$

Assumption A.9: (i) $\partial G_{1 \beta}\left(\tau, \tau_{0}, \rho\right)$ is nonsingular for all $\tau \in\left[\tau_{l}^{*}, \tau_{u}^{*}\right], \tau_{0} \in\left[\tau_{l}, \tau_{u}\right], \rho \in \varrho$ and (ii) the matrix $\left(L_{2}\left(\tau_{0}, \tau\right)\right)_{\tau_{0} \in \mathcal{J}_{0}, \tau \in \mathcal{J}_{1}}$ is of full rank.

Assumption A. 1 describes the data generating mechanism. Assumption A. 2 describes the copula function and requires the unobserved error terms are statistically independent of all the covariates $X$ as in Arellano and Bonhomme (2017). Assumption A. 3 provides some smoothness and boundedness conditions on the joint conditional distribution of $Y_{1}^{*}$ and $Y_{2}^{*}$ given the exogenous variables. The compactness condition in Assumption A. 4 is standard for extremum estimators. Assumption A. 5 describes the large sample properties of the first step estimator $\hat{\gamma}\left(\tau_{0}\right)$, which are satisfied by Chen's (2018) sequential quantile regression estimator and Powell's (1986) estimator. Assumption A. 6 states that $\delta_{n}$ can go to zero at any rate slower than $\sqrt{n}$. Assumption A. 7 contains the full rank condition for quantile regression by taking into sample selection. Assumption A. 8 is the global identification condition. Similar to typical moment-based estimators, here we assume that finite number of moments are sufficient for parameter identification. As discussed above, Proposition 2 essentially provides conditions for global identification based on infinite number of moment conditions. While we could adopt increasing number of moment conditions as the sample size increases, for practical empirical implementation, we consider semiparametric estimation based on a finite number of moment equations here. The non-singularity of $\partial G_{1 \beta}\left(\tau, \tau_{0}, \rho\right)$ in Assumption A. 9 is to ensures that $\hat{\beta}(\tau, \rho)$ satisfies an asymptotic linear representation, and the full rankness of $\left(L_{2}\left(\tau_{0}, \tau\right)\right)_{\tau_{0} \in \mathcal{J}_{0}, \tau \in \mathcal{J}_{1}}$ is the local identification condition for $\rho_{0}$.

The following theorem establishes the consistency and asymptotic normality of our estimator for copula coefficient $\rho_{0}$ and quantile coefficients $\beta(\tau)$ for $\tau \in\left[\tau_{l}^{*}, \tau_{u}^{*}\right]$.

Before presenting the theorem, with a slight abuse of notation, we first define

$$
G_{1 n}\left(b, \gamma, \rho, \delta, \tau, \tau_{0}\right)=\frac{1}{n} \sum_{i=1}^{n} g_{1}\left(\xi_{i}, b, \gamma, \rho, \delta, \tau, \tau_{0}\right)
$$

with

$$
g_{1}\left(\xi_{i}, b, \gamma, \rho, \delta, \tau, \tau_{0}\right)=1\left\{X_{i}^{\prime} \gamma>\delta\right\} 1\left\{Y_{2 i}>X_{i}^{\prime} \gamma\right\}\left(1\left\{Y_{1 i}<X_{i}^{\prime} b\right\}-C\left(\tau, \tau_{0}, \rho\right)\right) X_{i}
$$

and further define

$$
\phi_{\beta i}(\tau)=\tilde{\phi}_{\beta i}(\tau)+\frac{\partial \beta\left(\tau, \rho_{0}\right)}{\partial \rho} \phi_{\rho i}
$$

and

$$
\phi_{\rho i}=\left(\sum_{\tau_{0} \in \mathcal{J}_{0}, \tau \in \mathcal{J}_{1}} L_{2}^{\prime}\left(\tau_{0}, \tau\right) L_{2}\left(\tau_{0}, \tau\right)\right)^{-1} \sum_{\tau_{0} \in \mathcal{J}_{0}, \tau \in \mathcal{J}_{1}} L_{2}^{\prime}\left(\tau_{0}, \tau\right) \tilde{L}_{2 i}\left(\tau_{0}, \tau\right)
$$

where $\tilde{L}_{2 i}\left(\tau_{0}, \tau\right)=g_{2}\left(\tilde{\xi}_{i}, \beta\left(\tau, \rho_{0}\right), \gamma\left(\tau_{0}\right), \rho_{0}, 0, \tau, \tau_{0}\right)+\partial G_{2 \beta}\left(\tau, \tau_{0}, \rho_{0}\right) \tilde{\phi}_{\beta i}(\tau)+\partial G_{2 \gamma}\left(\tau, \tau_{0}, \rho_{0}\right) \phi_{\gamma i}\left(\tau_{0}\right)$, and $\tilde{\phi}_{\beta i}(\tau)=\left(1 / \# \mathcal{J}_{0}\right) \cdot \sum_{\tau_{0} \in \mathcal{J}_{0}} \partial G_{1 \beta}^{-1}\left(\tau, \tau_{0}, \rho_{0}\right) g_{1}\left(\xi_{i}, \beta\left(\tau, \tau_{0}, \rho_{0}\right), \gamma\left(\tau_{0}\right), \rho_{0}, 0, \tau, \tau_{0}\right)+\partial G_{1 \beta}^{-1}\left(\tau, \tau_{0}, \rho_{0}\right) \cdot \partial G_{1 \gamma}\left(\tau, \tau_{0}, \rho_{0}\right) \phi_{\gamma i}\left(\tau_{0}\right)$.

Theorem 1 If Assumptions A.1-A. 9 hold, then (i) $\hat{\rho}$ is consistent for $\rho_{0}$ and for any $\tau \in\left[\tau_{l}^{*}, \tau_{u}^{*}\right]$, and $\hat{\beta}(\tau, \hat{\rho})$ is consistent for $\beta(\tau)$; (ii) furthermore, $\hat{\rho}$ and $\hat{\beta}(\tau, \hat{\rho})$ have the asymptotic linear representation:

$$
\sqrt{n}\left(\hat{\rho}-\rho_{0}\right)=\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{\rho i}+o_{p}(1)
$$

and

$$
\sqrt{n}(\hat{\beta}(\tau, \hat{\rho})-\beta(\tau))=\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{\beta i}(\tau)+o_{p}(1)
$$

uniformly in $\tau \in\left[\tau_{l}^{*}, \tau_{u}^{*}\right]$. Thus, $\hat{\rho}$ and $\hat{\beta}(\tau, \hat{\rho})$ are asymptotic normal with

$$
\sqrt{n}\left(\hat{\rho}-\rho_{0}\right) \xrightarrow{d} N\left(0, E\left[\phi_{\rho i} \phi_{\rho i}^{\prime}\right]\right),
$$

and

$$
\sqrt{n}(\hat{\beta}(\tau, \hat{\rho})-\beta(\tau)) \xrightarrow{d} N\left(0, E\left[\phi_{\beta i}(\tau) \phi_{\beta i}^{\prime}(\tau)\right]\right) .
$$

In order to conduct statistical inference, it is important to provide consistent estimates for asymptotic covariance matrices. We can follow Arellano and Bonhomme (2017) to construct analytical estimates. Alternatively, resampling methods are also useful since direct estimation of the limiting covariance matrix can be difficult when the sample size is only moderately large, as it involves the estimation of the conditional density function, typical in a context of quantile regression. In fact, in our case, there are two density terms involved. Similar to Chen et al. (2003), Chernozhukov et al. (2015) and Chen (2018), we consider the multiplier bootstrap. Specifically, let $\left\{\eta_{i}\right\}_{1}^{n}$ be i.i.d. draws of positive random variables with $E \eta=\operatorname{Var}(\eta)=1$, independent of the data. First, we use the multiplier bootstrap estimator $\hat{\gamma}^{*}\left(\tau_{0}\right)$ for $\gamma\left(\tau_{0}\right)$ in Chen (2018), then the following assumption is satisfied.

Assumption A. $5^{\prime}: \hat{\gamma}^{*}\left(\tau_{0}\right)$ satisfies

$$
\max _{\tau_{0} \in\left[\tau_{1}, \tau_{u}\right]}\left|\hat{\gamma}^{*}\left(\tau_{0}\right)-\gamma_{0}\left(\tau_{0}\right)\right|=O\left(n^{-1 / 2} \log \log n\right)
$$

almost surely and uniformly over $\tau_{0} \in\left[\tau_{l}, \tau_{u}\right]$

$$
\sqrt{n}\left(\hat{\gamma}^{*}\left(\tau_{0}\right)-\gamma_{0}\left(\tau_{0}\right)\right)=\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \eta_{i} \phi_{\gamma_{i}}\left(\tau_{0}\right)+o_{p}(1)
$$

jointly in space $P \times P_{\eta}$.

Now given the first step estimator $\hat{\gamma}^{*}\left(\tau_{0}\right)$, for a fixed $\rho$, and $\tau_{0} \in \mathcal{J}_{0}$ and $\tau \in \mathcal{J}_{1}$ we define $\hat{\beta}^{*}\left(\tau, \tau_{0}, \rho\right)$ as a solution to

$$
\min _{b \in B} Q_{1 n}^{*}\left(b, \hat{\gamma}^{*}\left(\tau_{0}\right), \rho, \delta_{n}, \tau, \tau_{0}\right)
$$

where

$$
Q_{1 n}^{*}\left(b, \gamma, \rho, \delta, \tau, \tau_{0}\right)=\frac{1}{n} \sum_{i=1}^{n} \eta_{i} q_{1}\left(\xi_{i}, b, \gamma, \rho, \delta, \tau, \tau_{0}\right)
$$

For any $\tau \in \mathcal{J}_{1}$, once $\hat{\beta}^{*}\left(\tau, \tau_{0}, \rho\right)$ is available for $\tau_{0} \in \mathcal{J}_{0}$, we define

$$
\hat{\beta}^{*}(\tau, \rho)=\frac{1}{\# \mathcal{J}_{0}} \sum_{\tau_{0} \in \mathcal{J}_{0}} \hat{\beta}^{*}\left(\tau, \tau_{0}, \rho\right)
$$

Next, define

$$
G_{2 n}^{*}\left(b, \gamma, \rho, \delta, \tau, \tau_{0}\right)=\frac{1}{n} \sum_{i=1}^{n} \eta_{i} g_{2}\left(\xi_{i}, b, \gamma, \rho, \delta, \tau, \tau_{0}\right)
$$

Now, given $\hat{\beta}^{*}(\tau, \rho)$ for $\tau \in \mathcal{J}_{1}$, we define

$$
\hat{\rho}^{*}=\min _{\rho \in \varrho} G_{2 n}^{*}(\rho)
$$

where

$$
G_{2 n}^{*}(\rho)=\sum_{\tau_{0} \in \mathcal{J}_{0}, \tau \in \mathcal{J}_{1}}\left\|G_{2 n}^{*}\left(\hat{\beta}^{*}(\tau, \rho), \hat{\gamma}^{*}\left(\tau_{0}\right), \rho, \delta_{n}, \tau, \tau_{0}\right)\right\|
$$

Finally, given $\hat{\rho}^{*}$, for any $\tau$, define

$$
\hat{\beta}^{*}\left(\tau, \hat{\rho}^{*}\right)=\frac{1}{\# \mathcal{J}_{0}} \sum_{\tau_{0} \in \mathcal{J}_{0}} \hat{\beta}^{*}\left(\tau, \tau_{0}, \hat{\rho}^{*}\right)
$$

We will show that the asymptotic distribution of $\sqrt{n}(\hat{\beta}(\tau, \hat{\rho})-\beta(\tau))$ can be approximated by the limiting distribution of $\sqrt{n}\left(\hat{\beta}^{*}\left(\tau, \hat{\rho}^{*}\right)-\hat{\beta}(\tau, \hat{\rho})\right)$ conditional on the data, which in practice can be implemented through numerical simulation. We make the following additional assumption.

Assumption A.10: The weights $\left\{\eta_{i}\right\}_{1}^{n}$ are i.i.d. draws from a positive random variable $\eta$ with $E \eta=\operatorname{Var}(\eta)=$ 1 and it possesses $2+c_{0}$ moment for some $c_{0}>0$ that lives in a probability space $\left(\Omega_{\eta}, F_{\eta}, P_{\eta}\right)$, independent of the data $\left\{X_{i}, Y_{1 i}^{*}, Y_{2 i}^{*}\right\}$.

Theorem 2 If Assumptions A.1-A.4, A.5' and A.6-A.10 hold, then conditional on the data,

$$
\sqrt{n}\left(\hat{\beta}^{*}\left(\tau, \hat{\rho}^{*}\right)-\hat{\beta}(\tau, \hat{\rho})\right) \xrightarrow{d} N\left(0, E\left[\phi_{\beta i}(\tau) \phi_{\beta i}^{\prime}(\tau)\right]\right) .
$$

## 4 A Simulation Study

In this section we report the results of some Monte Carlo experiments to demonstrate the finite sample performance of our estimator. The first four designs of our simulation investigate the performance of our estimation approach when there exists some excluded covariate. We adopt the following data generating processes with sample size of $n=250$ and $n=500$, each replicated 400 times:

$$
\begin{aligned}
& Y_{1}^{*}=-1+X_{1}+X_{2}+\sigma(X) \cdot \Phi^{-1}(U) \\
& Y_{2}^{*}=1+Z_{1}+Z_{2}+Z_{3}+Z_{4}+\Phi^{-1}(V) \\
& Y_{2}=\max \left\{Y_{2}^{*}, 0\right\}, D=1\left\{Y_{2}^{*}>0\right\}, Y_{1}=D \cdot Y_{1}^{*}
\end{aligned}
$$

where $\sigma(X)=1$ for the homoscedastic case and $\sigma(X)=1+0.4 \cdot X_{1}$ for the heteroscedastic case, $X_{1}, X_{2}$, $Z_{3}$, and $Z_{4}$ are independent standard normal $N(0,1)$ with $Z_{1}=X_{1}$ and $Z_{2}=X_{2},(U, V)$ are independent of all regressors $\left(X\right.$ and $Z$ ) and distributed according to either Gaussian copula $C_{G}^{*}(\cdot, \cdot, 0.7)$ or Frank copula $C_{F}^{*}(\cdot, \cdot, 5.628),{ }^{6}$ and $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution. Consequently, there are four types of Data Generating Processes (DGPs) specified by the combinations of $\sigma(X)$ and copula (of $(U, V))$ in our experiments. The censoring percentages are about $33 \%$ in all of four DGPs. In the last four experiments, we look into the performance of our estimation method when there is no excluded covariates. Their DGPs are specified as

$$
\begin{aligned}
& Y_{1}^{*}=-1+X_{1}+X_{2}+\sigma(X) \cdot \Phi^{-1}(U) \\
& Y_{2}^{*}=1+X_{1}+X_{2}+\Phi^{-1}(V) \\
& Y_{2}=\max \left\{Y_{2}^{*}, 0\right\}, D=1\left\{Y_{2}^{*}>0\right\}, Y_{1}=D \cdot Y_{1}^{*}
\end{aligned}
$$

where $\sigma(X)=1$ for the homoscedastic case and $\sigma(X)=1+0.4 \cdot X_{1}$ for the heteroscedastic case, $X_{1}$ and $X_{2}$ are independent standard normal $N(0,1)$, and $(U, V)$ are from either Gaussian copula $C_{G}^{*}(\cdot, \cdot, 0.7)$ or Frank copula $C_{F}^{*}(\cdot, \cdot, 5.628)$.

Before discussing the performance of our estimator, we first describe the implementation details of our estimation procedure in the simulation study. The set $\mathcal{J}_{0}$ of quantile indices is given by $\left\{\tilde{\tau}_{0 j}\right\}_{j=1}^{L_{0}}$ where $\tilde{\tau}_{01}=0.9, \tilde{\tau}_{0 L_{0}}=0.3$, and $\tilde{\tau}_{0 j}-\tilde{\tau}_{0 j+1}=0.05$ for $j=1, \ldots, L_{0}-1$; and the set $\mathcal{J}_{1}$ is specified as $\left.\left\{\tilde{\tau}_{j}\right\}\right\}_{j=1}^{L}$ where $\tilde{\tau}_{1}=0.9, \tilde{\tau}_{L}=0.1$, and $\tilde{\tau}_{j}-\tilde{\tau}_{j+1}=0.05$ for $j=1, \ldots, L-1$. We follow Chen's (2018) three-step estimation procedure to obtain the estimates of $\gamma\left(\tau_{0}\right)$ for all $\tau_{0} \in \mathcal{J}_{0}$. Similar to Chernozhukov and Hong (2002), we choose the subsample selector parameter $\delta_{n}$ to be the $1 \%$ quantile of those positive $Z_{i}^{\prime} \hat{\gamma}\left(\tau_{0}\right)$ or $X_{i}^{\prime} \hat{\gamma}\left(\tau_{0}\right)$ in both steps of our estimation. ${ }^{7}$ Moreover, in the second step, we choose $\|\cdot\|$ in $G_{2 n}(\cdot)$ to be Euclidean norm and $\varphi(Z)=\left(1, Z_{1}, Z_{2}, Z_{3}, Z_{4}\right)$ in the first four experiments and $\varphi(X)=\left(1, X_{1}, X_{2}, X_{1}^{2}, X_{2}^{2}, X_{1} \cdot X_{2}\right)$ in the last four designs. We then estimate the quantile coefficients $\beta(\tau)$ for $\tau \in \mathcal{J}_{1}$ and copula parameter $\rho$ by our two-step estimation procedure in those eight DGP scenarios with sample size of $n=250$ and $n=500$.

[^5]In particular, after obtaining an estimate of the copula parameter $\rho$, i.e. $\hat{\rho}$, the quantile coefficients can be recovered by $\hat{\beta}(\tau, \hat{\rho})$ for all $\tau \in \mathcal{J}_{1}$.

We now report the performance of our estimator $\hat{\rho}$ and $\hat{\beta}(\tau, \hat{\rho})$ for $\tau=0.9,0.75,0.50,0.25$, and 0.10 to demonstrate how well our estimation procedure can recover the quantile coefficients at different quantile levels. In particular, we report the bias (Bias), the standard deviation (SD), and the root mean square error (RMSE) for our estimator. With excluded covariates, Tables 2-3 are for the homoscedastic case, and Tables 4-5 focus on the heteroscedastic case. Tables 6-7 and 8-9 report the results for the homoscedastic and heteroscedastic cases, respectively, without excluded covariates. All the tables are placed in Appendix B.

Table 2 provides the simulation results for the homoscedastic case (i.e., $\sigma(X)=1$ ) with Gaussian copula which has a correlation coefficient of $\rho_{0}=0.7$. It shows that our estimator of quantile coefficients performs reasonably well in the entire quantile range even under the moderate sample size of $n=250$. Moreover, when the sample size increases to 500, the root mean square error (RMSE) of our estimator for both quantile coefficients and correlation parameter (of Gaussian copula) decreases significantly; in addition, there are effectively little or no biases across the board.

We report the simulation results for the homoscedastic case with the Frank copula having a parameter of $\rho_{0}=5.628$ in Table 3. ${ }^{8}$ Similar to Table 2, it shows that, with Frank copula, (i) our estimator for quantile coefficients is reasonably satisfactory for all quantile levels, and (ii) our estimator becomes closer to its true value as sample size increases. In addition, comparing with the results of Gaussian copula in Table 2, the estimator of quantile coefficients with Frank copula performs slightly better (in terms of RMSE) overall for the upper quantiles such as $\tau=0.75$ and 0.90 , but slightly worse for the lower quantiles such as $\tau=0.10$ and 0.25 , with overall very good performance. For the estimation of the copula parameter, one might get the impression that both the biases and standard errors are quite sizable, especially when $n=250$. However, this is quite misleading and mainly due to the peculiar nature of the Frank copula. To present a better picture, we convert the copula parameter to the Kendall's tau, which provides a much more intuitive description for the copula structure since Kendall's tau is a more concrete and intuitive summary of the dependence measure of any particular copula. Then, it turns out our estimator for Kendall's tau performs very well, with very small biases and standard errors, even when $n=250$.

Tables 4 and 5 report the results for the heteroscedastic cases (i.e., $\sigma(X)=1+0.4 \cdot X_{1}$ ) with Gaussian copula $C_{G}^{*}(\cdot, \cdot, 0.7)$ and Frank copula $C_{F}^{*}(\cdot, \cdot, 5.628)$, respectively. Heteroscedastic model captures the nature of quantile regression where both slope and intercept coefficients are varying over quantiles. In general we observe a similar pattern of performances of our estimator to the homoscedastic designs; our estimator for both the quantile coefficients and copula parameter largely performs very well.

Tables 6 to 9 report the results for the cases with no exclusion restriction, with both homoscedastic and heteroscedastic designs and the Gaussian copula $C_{G}^{*}(\cdot, \cdot, 0.7)$ and Frank copula $C_{F}^{*}(\cdot, \cdot, 5.628)$. Compared with the first four designs that impose the exclusion restriction, our estimator continue to perform very well, with somewhat larger RMSEs. This is quite encouraging and suggests that our method is robust to absence of excluded covariates in the selection equation. Such a feature makes our approach empirically appealing, since excluded covariates are typically difficult to justify and find in many real data sets.

[^6]
## 5 An Application to the UK Labor Market

In this section, we apply our estimation procedure to study the wage inequality within and between genders in the UK labor market. The wage inequality within gender is demonstrated by the difference of potential wages at different quantiles, while the wage inequality between genders is measured by the differences of male and female potential wages at the same quantile levels. It is well known in labor economics that the estimation of wage distribution, and hence gender wage gap, is biased if the nonrandom selection (into work) issue is not addressed. We use our quantile censored selection model to correct for the (nonrandom) sample selection bias by using the censored data of working hours which are available in most, if not for all, labor economic data sets (see, e.g., Table 2.25 of Killingsworth and Heckman, 1986 for summary of typical data sets in the literature). Our method recovers the given quantiles of potential wage distributions for both males and females, and can hence estimates the gender wage gap at given quantile levels.

We first describe the data set which is used for our empirical study. It explains that how our data is constructed from the original data source, and what characteristics are used to control for the observed heterogeneity in the wage and selection equations. It also provides the descriptive statistics of those characteristics as well as log wages and hours of work. Second, we explain how our estimation approach is implemented in the empirical application. In particular, we do not need any instrumental variable which appears in the selection equation but is excluded from the wage equation. Indeed, as argued by Blundell et al. (2007), in general, such instruments derived from economic theory are hard to find. Finally, we show our estimation results of potential wage quantiles for single/married male and female as well as gender wage gap at different quantile levels after correcting the sample selection bias in the observed wage distribution.

### 5.1 Data Description

The data we used for the analysis is from the UK Family Expenditure Survey (FES) during the period of 1978-2000. ${ }^{9}$ The sample is constructed by following the previous papers analyzing this data set, see, e.g., Gosling et al. (2000), Blundell et al. (2003), Blundell et al. (2007), and Arellano and Bonhomme (2017). It includes men and women with an age range of 23 to 59 who were not in full time education. It drops the individuals who were self employed or had missing values for the demographic variables in Table 1. It also excludes from the sample the observations who reported positive working hours but did not report any wages. ${ }^{10}$ This gives us a sample of 162,532 individuals in total.

The hourly wage, the log of which is $Y_{1}$ in our model, of each working individual is constructed by the ratio of usual weekly earnings to usual weekly working hour (namely $Y_{2}$ ). Besides gender and marital status, we follow Blundell et al. (2003) to control the heterogeneity of wage distribution and sample selection by including regressors of education (three categories: leaving school before 17, at 17 or 18 , after 18), cohort (five categories: born 1919-1934, 1935-1944, 1945-1954, 1955-1964, 1965-1977), region (12 standard regions), time trend up to cubic term. To compare our results with Arellano and Bonhomme (2017), we also add the regressors of number of kids split by age categories (with six dummies) and the measure of out-of-work

[^7]income. The latter is evaluated using the IFS tax and benefit simulation model TAXBEN and was initially used by Blundell et al. (2003) to interpret aggregate wage growth in the UK labor market. ${ }^{11}$ Table 1 provides the summary statistics of the log wage, usual weekly working hour, and the selected characteristics of observations in the sample.

Table 1: Summary statistics

| Variable | Obs. | Mean | Std. Dev. | Min | Max |
| :--- | :--- | :--- | :--- | :--- | :--- |
| log wage | 121,805 | 1.887 | 0.517 | -0.465 | 4.299 |
| hours | 162,532 | 27.356 | 19.466 | 0 | 163 |
| sex | 162,532 | 0.540 | 0.498 | 0 | 1 |
| married | 162,532 | 0.748 | 0.434 | 0 | 1 |
| \# kids | 162,532 | 0.974 | 1.143 | 0 | 12 |
| out-of-work income | 162,532 | 4.760 | 0.814 | -6.127 | 8.893 |
| left school $\leq 16$ | 162,532 | 0.703 | 0.457 | 0 | 1 |
| left school 17-18 | 162,532 | 0.157 | 0.364 | 0 | 1 |
| left school 19+ | 162,532 | 0.140 | 0.347 | 0 | 1 |
| cohort 1919-1934 | 162,532 | 0.145 | 0.353 | 0 | 1 |
| cohort 1935-1944 | 162,532 | 0.216 | 0.412 | 0 | 1 |
| cohort 1945-1954 | 162,532 | 0.294 | 0.455 | 0 | 1 |
| cohort 1955-1964 | 162,532 | 0.242 | 0.428 | 0 | 1 |
| cohort 1965-1977 | 162,532 | 0.103 | 0.304 | 0 | 1 |
| region 1 | 162,532 | 0.060 | 0.237 | 0 | 1 |
| region 2 | 162,532 | 0.089 | 0.285 | 0 | 1 |
| region 3 | 162,532 | 0.111 | 0.314 | 0 | 1 |
| region 4 | 162,532 | 0.073 | 0.260 | 0 | 1 |
| region 5 | 162,532 | 0.094 | 0.291 | 0 | 1 |
| region 6 | 162,532 | 0.037 | 0.188 | 0 | 1 |
| region 7 | 162,532 | 0.104 | 0.305 | 0 | 1 |
| region 8 | 162,532 | 0.187 | 0.390 | 0 | 1 |
| region 9 | 162,532 | 0.078 | 0.268 | 0 | 1 |
| region 10 | 162,532 | 0.051 | 0.221 | 0 | 1 |
| region 11 | 162,532 | 0.091 | 0.288 | 0 | 1 |
| region 12 | 162,532 | 0.025 | 0.156 | 0 | 1 |
| Rots 1 (1) | 1251 |  |  |  |  |

Notes: (i) sex $=0$ for male and $=1$ for female;
(ii) married $=0$ for single and $=1$ for married;
(iii) the summary statistics for $\log$ wage are only for working subsample.

The wage inequality within and between genders changed dramatically in the UK during the last two decades of twentieth century. Figure 1(a) demonstrates the within-gender inequality in terms of interdecile range (IDR) of (log) wage distribution for working individuals over time. ${ }^{12}$ It shows that the within-gender inequality for both males and females increases dramatically over time, with IDR from about 0.90 (1.01) in 1978 increasing to about 1.34 (1.28) in 2000 for males (females). Figure 1(b) provides the gender wage gap

[^8]for 1st, 5 th, and 9 th deciles of (log) wage distributions of workers over time. We can see that the gender wage gap for these three deciles has similar dynamic pattern over time: for instance, the median drops from 0.44 in the year of 1978, with an initial significant increase around 1980 and occasional increases later, to 0.28 in 2000. During the same period, the employment rates of males and females have varied much. Figure 1(c) shows how the employment rates evolved over time. The employment rate of males drops from 0.95 in 1978 to 0.80 in 1993, and then increases slowly back to 0.84 in 2000. Moreover, the employment rate of females increases slowly from 0.64 in 1978 to 0.69 in 2000, with a significant growth during 1982-1990 and non-monotonic fluctuations otherwise; it reaches lowest at 0.59 in 1982 and highest at 0.70 in 1999. Given these facts about the changes of employment rates over time, it is crucial to correct the sample selection bias when we evaluate the dynamics of (latent) wage inequality within and between genders in the sample.


Figure 1: UK workers: IDR, gender wage gap, and employment rate, 1978-2000.

### 5.2 Estimation Procedure

We first describe the specification and implementation details of our estimation procedure in the application. Our sample is divided into four subsamples according to the gender and marital status, i.e., the parameters of our model, such as copula parameter and all quantile coefficients, are assumed to be gender and marital-status specific. In each subsample, we use the potential out-of-work income as well as other characteristics (i.e. education, cohort, region, time trend up to cubic term, and number of kids) as regressors to control the heterogeneity of both wage distribution and sample selection process. In particular, we do not exclude the potential out-of-work income from the wage equation. In other words, both the wage and (censored) selection equations share the same set of regressors in our application. This is in contrast with Arellano and Bonhomme (2017) which excludes the out-of-work income from the wage equation but keep it in the (binary) selection equation, while all other characteristics appear in both wage and selection equations.

To implement our estimation procedure, we choose the Frank copula to model the dependence between $U$ and $V$ (which are the unobserved factors in the log-wage and working hours equations, respectively). We obtain very similar results (not reported) by adopting the Gaussian copula instead of Frank copula. In particular, we obtain an estimate of rank correlation very close to the one with Frank copula. Thus our empirical findings seem not sensitive to the choice of copula in our estimation. After obtaining the estimates of $\gamma(\cdot)$ from the three-step estimation procedure of Chen (2018), we use our two-step estimation procedure
to recover the quantile and copula coefficients $\beta(\cdot)$ and $\rho$. The subsample selector $\delta_{n}$ is chosen to be the $1 \%$ quantile of those positive $X_{i}^{\prime} \hat{\gamma}(\cdot)$ in both steps of our estimation, similar to Chernozhukov and Hong (2002) and Chen (2018). In the second step, we use $\varphi(X)=(1, X, X \otimes X)$ as the instruments ${ }^{13}$ to search for the best $\rho$ among 200 grid points. ${ }^{14}$

### 5.3 Main Results

We next present our main empirical findings. Figures 2 and 3 show, respectively, the dynamics of deciles of log-wages for both males and females and the evolution of IDR, which measures the within-gender inequality, and figures 4-6 provide the dynamics of gender wage gap. The lines for males (females) are in black (red). We display the quantities (e.g., deciles of log-wages, IDR, and gender wage gap) of workers in solid lines, and the ones given by our method in dotted lines. Our two-step procedure is used to estimate the quantile coefficients $\beta(\cdot)$ (in wage equation) and copula parameter $\rho$. The approach of Machado and Mata (2005) is then adopted to simulate the (unconditional) wage distribution after correcting for the sample selection bias. The results of Arellano and Bonhomme (2017) are included for comparison. ${ }^{15}$ They are shown in dashed lines and labeled as AB17.

We first examine the results of males. Regarding the copula estimates of males, we obtain a Spearman's rank correlation (implied by the copula estimate) of -0.33 (with s.e. of 0.018 ) for married males and -0.28 (with s.e. of 0.027) for singles. In contrast, Arellano and Bonhomme (2017) obtained a rank correlation of -0.24 (with $95 \%$ confidence interval of ( $-0.35,-0.06$ )) for married males and -0.79 (with $95 \%$ confidence interval of ( $-0.84,-0.42$ ) for singles. Consequently, with much smaller standard errors compared with Arellano and Bonhomme (2017), ${ }^{16}$ our copula estimates imply that both the married and single males have negative selection into employment, whereas Arellano and Bonhomme's (2017) estimates imply significant positive selections into employment for single males, ${ }^{17}$ less so for married males. One of our most striking findings is the negative selection into employment for males. Ermisch and Wright (1994) argued that negative selection into employment is very plausible when there is relatively high positive correlation between the wage offer and reservation wage of a potential worker. Such a (relatively high) positive correlation is possible since a person with higher productivity in outside jobs tends to be more competent in tasks at home. For other papers finding evidence of negative selection into employment, see, e.g., Dolton and Makepeace (1987), Steinberg (1989), Wright and Ermisch (1991), Mulligan and Rubinstein (2008), and Mocan and Unel (2017).

Figure 2 shows the effect to correct the aforementioned selection bias in the wage distribution of males. It suggests that our selection correction is significant for males at lower deciles. At the 10th percentile,

[^9]the observed wage of males increases by $8.8 \%$ during the period of 1978-2000, but our latent male wage increases by $13.1 \%$ in contrast with almost no increase in the latent male wage of Arellano and Bonhomme (2017). Sizable differences among those three wages can also be observed for the 20th and 30th percentiles. For the middle and high deciles, our latent wage differs less (resp. significantly) from the observed wage before (resp. after) 1993 but deviates significantly from the Arellano and Bonhomme's (2017) estimates of latent wage during the whole period.


Figure 2: First nine deciles of wage distribution after no, AB17, and our corrections (Black for males and red for females).

Second, we present the results of females. We obtain an estimate of copula with a rank correlation of 0.28 (with s.e. of 0.027 ) for married females and -0.06 (with s.e. of 0.041 ) for singles, while Arellano and Bonhomme's (2017) estimates of rank correlations are -0.17 (with $95 \%$ confidence interval of ( $-0.30,-0.01$ ) ) for married females and -0.08 (with confidence interval of $(-0.24,0.16)$ ) for singles. We thus obtain similar direction of selection to Arellano and Bonhomme (2017), i.e., positive selection for married females but
insignificant selection for singles, except that our estimates are more significant quantitatively with much smaller standard errors.

Figure 2 also displays the effect of selection bias correction for females. It shows that our correction of sample selection is more significant than Arellano and Bonhomme (2017) in all nine deciles of wage distribution for females. Specifically, our results show a further correction from the outcome of Arellano and Bonhomme (2017), and have a magnitude of correction as large as twice to four times of theirs in most cases. On average, our result provides bias corrections of $0.09(10 \%), 0.06(50 \%)$, and $0.12(90 \%)$, while Arellano and Bonhomme (2017) has bias corrections of 0.03 (10\%), 0.03 ( $50 \%$ ), and 0.03 ( $90 \%$ ).

We examine within-gender inequality (in terms of IDR) and gender wage gap shown in figure 3 and in figures 4-6 respectively. With all three estimates of within-gender wage inequality increasing over time, figure 3 shows that our within-gender inequality is significantly smaller than those of Arellano and Bonhomme (2017) for both males and females. In addition, our increase is significantly less than those that correspond to the observed and the estimates of Arellano and Bonhomme (2017) for males, but is comparable for females.


Figure 3: The IDR for both males and females after no, AB17, and our corrections.

Regarding gender wage gap dynamics, figure 4 shows that all of the three estimates of gender wage gap decline over time. Specifically, our estimate of wage gap is largest among the three in all deciles, and the other two estimates almost coincide in the middle and high deciles with Arellano and Bonhomme's (2017) estimate below the one with no correction in the low deciles. Nevertheless, ours has the smallest reduction among the three in all deciles with a few exceptions starting from 1978 (but with no exceptions starting from 1981). For example, during the period of 1978-2000, the average gender wage gaps are 0.51 (ours), 0.31 (AB17), and 0.39 (no correction) in the 10th percentile, and 0.50 (ours), 0.40 (AB17), and 0.42 (no correction)


Figure 4: The gender wage gap after no, $A B 17$, and our corrections.
in the 50 th percentile, and 0.49 (ours), 0.35 (AB17), and 0.36 (no correction) in the 90th percentile. Figures 5 and 6 further show the gender wage gaps for single and married groups, respectively. Our estimate of gender wage gap still has the largest value but minimal reduction among the three estimates for both single and married groups; while Arellano and Bonhomme's (2017) estimate of gender wage gap is much smaller than the one with no correction in low and middle deciles for singles (in the 10th percentile, the averages are 0.01 for AB 17 and 0.26 for no correction; in the median case, they are 0.10 for AB17 and 0.20 for no correction; in the 90th percentile, they become 0.11 for AB17 and 0.16 for no correction) but is very close to the one with no correction in all deciles for married (in the 10th percentile, the averages are 0.45 for AB17 and 0.43 for no correction; in the median case, they are 0.50 for AB17 and 0.49 for no correction; in the 90th percentile, they become 0.43 for AB 17 and 0.41 for no correction). To better understand the above findings on the gender wage gap reductions, note that in Panel c) of Figure 1, the striking pattern is that the male participation rate has decreased a lot, from $95 \%$ to around $80-85 \%$, while the female participation rate has slightly increased. As the participation and wage unobservables $V$ and $U$ are negatively correlated, we observe the wages of "worse" males at the end than at the start of the period, and we observe the wages of "better" females at the end than at the start of the period. Thus, the reduction in the the gender wage gap over the period stems from this compositional change, which is why accounting for selection leads to a much lower reduction of the gender-wage gap in our estimate. ${ }^{18}$ As Arellano and Bonhomme (2017) found that the men in employment are positively selected, accounting for selection with their method leads to a

[^10]

Figure 5: The gender wage gap for singles after no, AB17, and our corrections.


Figure 6: The gender wage gap for married after no, AB17, and our corrections.
larger, not lower, reduction of the gender-wage gap in their estimate. ${ }^{19}$
In summary, among our most important results, our analysis reveals that (i) there is a significant negative selection for males; (ii) the bias correction is significant for females, with the magnitude of corrections two to four times as in Arellano and Bonhomme (2017); (iii) the gender wage gap has remained large and the wage gap reduction has been very limited, in particular, accounting for selection bias leads to a much smaller reduction over time, compared with the observed wage gap and the selection corrected version of Arellano and Bonhomme (2017).

## 6 Conclusion

In this paper, we have proposed a quantile selection model with censored selection where both the latent outcome and participation equations are modelled as semiparametric quantile regressions, and the selection bias is modelled through a copula function. With a censored selection, no exclusion restriction is needed for identification. We have further developed estimators for the copula parameter and quantile regression coefficients in the outcome equation. We applied our method to study the evolution of wage inequality in the UK.

In the paper we focus on the outcome equation with exogenous regressors. Vella (1993) and Lee and Vella (2006) studied sample selection model with endogeneity, which is also subject to censored selection rules. Quantile regression with endogeneity has been studied by Chernozhukov and Hansen $(2006,2008)$ and Chen (2018). As endogeneity arises frequently in empirical analysis of economic data, it is interesting to incorporate endogeneity in our framework. This is an interesting topic for future research.

## Appendix A. Proofs of Theorems

Lemma A1. Suppose Assumptions A.1-A. 6 and A. 9 are satisfied, then $\hat{\beta}(\tau, \rho)$ is uniformly consistent and has uniform asymptotic linear representation for any given $\tau \in\left[\tau_{l}^{*}, \tau_{u}^{*}\right]$ and $\rho \in \varrho$,

$$
\begin{equation*}
\sup _{\rho \in \varrho}\|\hat{\beta}(\tau, \rho)-\beta(\tau, \rho)\|=o_{p}(1) \tag{A1}
\end{equation*}
$$

and

$$
\begin{equation*}
\sqrt{n}(\hat{\beta}(\tau, \rho)-\beta(\tau, \rho))=\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\phi}_{\beta i}(\tau, \rho)+o_{p}(1) \tag{A2}
\end{equation*}
$$

[^11]where $\beta(\tau, \rho)=\frac{1}{\# \mathcal{J}_{0}} \sum_{\tau_{0} \in \mathcal{J}_{0}} \beta\left(\tau, \tau_{0}, \rho\right)$ and $\tilde{\phi}_{\beta i}(\tau, \rho)$ is defined in the proof below in (A.4); Furthermore, $\hat{\beta}(\tau, \rho)$ satisfies the asymptotic tightness property
\[

$$
\begin{equation*}
(\hat{\beta}(\tau, \rho)-\beta(\tau, \rho))-(\hat{\beta}(\tau, \bar{\rho})-\beta(\tau, \bar{\rho}))=o_{p}\left(n^{-1 / 2}\right) \tag{A3}
\end{equation*}
$$

\]

uniformly in $\rho$ and $\bar{\rho}$ in the $o_{p}(1)$ neighborhood of $\rho_{0}$.
Proof of Lemma A1: Recall that

$$
Q_{1 n}\left(b, \gamma, \rho, \delta, \tau_{0}, \tau\right)=\frac{1}{n} \sum_{i=1}^{n} q_{1}\left(\xi_{i}, b, \gamma, \rho, \delta, \tau, \tau_{0}\right)
$$

and

$$
Q_{1}\left(b, \gamma\left(\tau_{0}\right), \rho, \delta, \tau_{0}, \tau\right)=E q_{1}\left(\xi, b, \gamma\left(\tau_{0}\right), \rho, \delta, \tau, \tau_{0}\right)
$$

with

$$
Q_{1}\left(b, \gamma\left(\tau_{0}\right), \rho, \tau_{0}, \tau\right)=E q_{1}\left(\xi, b, \gamma\left(\tau_{0}\right), \rho, 0, \tau, \tau_{0}\right)
$$

Define the class of functions

$$
\mathcal{M}_{1}=\left\{q_{1}\left(\xi, b, \gamma, \rho, \delta, \tau, \tau_{0}\right):(b, \gamma, \rho) \in B \times \Gamma \times \varrho, \delta, \tau \in\left[\tau_{l}^{*}, \tau_{u}^{*}\right], \tau_{0} \in\left[\tau_{l}, \tau_{u}\right]\right\}
$$

where

$$
\begin{aligned}
q_{1}\left(\xi, b, \gamma, \rho, \delta, \tau, \tau_{0}\right)= & 1\left\{Z^{\prime} \gamma>\delta\right\} 1\left\{Y_{2}>Z^{\prime} \gamma\right\} \\
& \times\left[C\left(\tau, \tau_{0}, \rho\right)\left(Y-X^{\prime} b\right)^{+}+\left(1-C\left(\tau, \tau_{0}, \rho\right)\right)\left(Y-X^{\prime} b\right)^{-}\right]
\end{aligned}
$$

It is straightforward to verify that $\mathcal{M}_{1}$ is Euclidean (Pakes and Pollard, 1989) with a square integrable envelop, then we have

$$
\begin{aligned}
& Q_{1 n}\left(b, \hat{\gamma}\left(\tau_{0}\right), \rho, \delta_{n}, \tau_{0}, \tau\right)-Q_{1}\left(b, \gamma\left(\tau_{0}\right), \rho, \tau_{0}, \tau\right) \\
= & {\left[Q_{1 n}\left(b, \hat{\gamma}\left(\tau_{0}\right), \rho, \delta_{n}, \tau_{0}, \tau\right)-Q_{1}\left(b, \hat{\gamma}\left(\tau_{0}\right), \rho, \delta_{n}, \tau_{0}, \tau\right)\right] } \\
& +\left[Q_{1}\left(b, \hat{\gamma}\left(\tau_{0}\right), \rho, \delta_{n}, \tau_{0}, \tau\right)-Q_{1}\left(b, \gamma\left(\tau_{0}\right), \rho, \tau_{0}, \tau\right)\right] \\
= & o_{p}(1)
\end{aligned}
$$

uniformly in $\left(b, \rho, \tau_{0}, \tau\right) \in B \times \varrho \times\left[\tau_{l}, \tau_{u}\right] \times\left[\tau_{l}^{*}, \tau_{u}^{*}\right]$, where the difference in the first bracket on the right hand side of first equality is $o_{p}(1)$ following from the uniform law of large numbers (Pakes and Pollard, 1989) and the difference in the second bracket is also $o_{p}(1)$ following from the Assumptions A.5-A.7.

For any given $\left(\rho, \tau_{0}, \tau\right) \in \varrho \times\left[\tau_{l}, \tau_{u}\right] \times\left[\tau_{l}^{*}, \tau_{u}^{*}\right]$, since $\beta\left(\tau, \tau_{0}, \rho\right)$ is the unique minimizer of $Q_{1}\left(b, \gamma\left(\tau_{0}\right), \rho, 0, \tau_{0}, \tau\right)$, which is continuous in $b$, together with the compactness for the parameter space, we obtain

$$
\hat{\beta}\left(\tau, \tau_{0}, \rho\right)-\beta\left(\tau, \tau_{0}, \rho\right)=o_{p}(1)
$$

In addition, note that $\beta\left(\tau, \tau_{0}, \rho\right)$ is continuous in $\left(\tau, \tau_{0}, \rho\right)$. Then, by Lemma A. 1 in Carroll et. al., (1997), it follows that

$$
\hat{\beta}\left(\tau, \tau_{0}, \gamma, \rho\right)-\beta\left(\tau, \tau_{0}, \gamma, \rho\right)=o_{p}(1)
$$

holds uniformly in $\left(\rho, \tau_{0}, \tau\right) \in \varrho \times\left[\tau_{l}, \tau_{u}\right] \times\left[\tau_{l}^{*}, \tau_{u}^{*}\right]$.
Next, following the subgradient argument in Powell (1984), Honoré (1992) and Chen (2018), with probability approaching one we have

$$
G_{1 n}\left(\hat{\beta}\left(\tau, \tau_{0}, \rho\right), \hat{\gamma}\left(\tau_{0}\right), \rho, \delta_{n}, \tau_{0}, \tau\right)=o_{p}\left(n^{-1 / 2}\right)
$$

uniformly in $\left(\tau_{0}, \tau, \rho\right)$, where

$$
G_{1 n}\left(b, \gamma, \rho, \delta, \tau, \tau_{0}\right)=\frac{1}{n} \sum_{i=1}^{n} g_{1}\left(\xi_{i}, b, \gamma, \rho, \delta, \tau, \tau_{0}\right)
$$

with

$$
g_{1}\left(\xi_{i}, b, \gamma, \rho, \delta, \tau, \tau_{0}\right)=1\left\{X_{i}^{\prime} \gamma>\delta\right\} 1\left\{Y_{2 i}>X_{i}^{\prime} \gamma\right\}\left(1\left\{Y_{1 i}<X_{i}^{\prime} b\right\}-C\left(\tau, \tau_{0}, \rho\right)\right) X_{i}
$$

Let $\mathcal{G}_{1}$ be the class of functions

$$
\mathcal{G}_{1}=\left\{g_{1}\left(\xi_{i}, b, \gamma, \delta, \rho, \tau, \tau_{0}\right): \tau, \tau_{0} \in(0,1), \delta, b \in B, \gamma \in \Gamma, \rho \in \varrho\right\}
$$

Similar to $\mathcal{M}_{1}, \mathcal{G}_{1}$ is also Euclidean with a square integrable envelop; then by Lemma 2.17 (Pakes and Pollard, 1989), we obtain

$$
\begin{aligned}
o_{p}\left(n^{-1 / 2}\right)= & G_{1 n}\left(\hat{\beta}\left(\tau, \tau_{0}, \rho\right), \hat{\gamma}\left(\tau_{0}\right), \delta_{n}, \rho, \tau_{0}, \tau\right) \\
= & G_{1 n}\left(\beta\left(\tau, \tau_{0}, \rho\right), \gamma\left(\tau_{0}\right), \delta_{n}, \rho, \tau_{0}, \tau\right) \\
& +\left[G_{1}\left(\hat{\beta}\left(\tau, \tau_{0}, \rho\right), \hat{\gamma}\left(\tau_{0}\right), \delta_{n}, \rho, \tau_{0}, \tau\right)-G_{1}\left(\beta\left(\tau, \tau_{0}, \rho\right), \gamma\left(\tau_{0}\right), \delta_{n}, \rho, \tau_{0}, \tau\right)\right]+o_{p}\left(n^{-1 / 2}\right)
\end{aligned}
$$

uniformly in $\left(\tau_{0}, \tau, \rho\right)$, where

$$
G_{1}\left(b, \gamma, \beta, \rho, \delta, \tau_{0}, \tau\right)=E\left[g_{1}\left(\xi_{i}, b, \gamma, \rho, \delta, \tau, \tau_{0}\right)\right]
$$

Then, it is straightforward to show that

$$
\begin{aligned}
& {\left[G_{1}\left(\hat{\beta}\left(\tau, \tau_{0}, \rho\right), \hat{\gamma}\left(\tau_{0}\right), \rho, \delta_{n}, \tau_{0}, \tau\right)-G_{1}\left(\beta\left(\tau, \tau_{0}, \rho\right), \gamma\left(\tau_{0}\right), \rho, \delta_{n}, \tau_{0}, \tau\right)\right] } \\
= & \partial G_{1 \beta}\left(\tau, \tau_{0}, \rho\right)\left(\hat{\beta}\left(\tau, \tau_{0}, \rho\right)-\beta\left(\tau, \tau_{0}, \rho\right)\right) \\
& +\partial G_{1 \gamma}\left(\tau, \tau_{0}, \rho\right)\left(\hat{\gamma}\left(\tau_{0}\right)-\gamma\left(\tau_{0}\right)\right)+o_{\rho}\left(n^{-1 / 2}+\left(\hat{\beta}\left(\tau, \tau_{0}, \rho\right)-\beta\left(\tau, \tau_{0}, \rho\right)\right)\right)
\end{aligned}
$$

uniformly in $\left(\rho, \tau_{0}, \tau\right) \in \varrho \times\left[\tau_{l}, \tau_{u}\right] \times\left[\tau_{l}^{*}, \tau_{u}^{*}\right]$. From the above results, we can deduce that

$$
\begin{aligned}
\sqrt{n}\left(\hat{\beta}\left(\tau, \tau_{0}, \rho\right)-\beta\left(\tau, \tau_{0}, \rho\right)\right)= & \partial G_{1 \beta}^{-1}\left(\tau, \tau_{0}, \rho\right) \sqrt{n} G_{1 n}\left(\beta\left(\tau, \tau_{0}, \rho\right), \gamma\left(\tau_{0}\right), \rho, 0, \tau_{0}, \tau\right) \\
& +\partial G_{1 \beta}^{-1}\left(\tau, \tau_{0}, \rho\right) \partial G_{1 \gamma}\left(\tau, \tau_{0}, \rho\right) \sqrt{n}\left(\hat{\gamma}\left(\tau_{0}\right)-\gamma\left(\tau_{0}\right)\right)+o_{p}(1) \\
= & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\phi}_{\beta i}\left(\tau, \tau_{0}, \rho\right)+o_{p}(1)
\end{aligned}
$$

where

$$
\begin{align*}
\tilde{\phi}_{\beta i}\left(\tau, \tau_{0}, \rho\right)= & \partial G_{1 \beta}^{-1}\left(\tau, \tau_{0}, \rho\right) g_{1}\left(\xi_{i}, \beta\left(\tau, \tau_{0}, \rho\right), \gamma\left(\tau_{0}\right), \rho, 0, \tau, \tau_{0}\right)  \tag{A.4}\\
& +\partial G_{1 \beta}^{-1}\left(\tau, \tau_{0}, \rho\right) \partial G_{1 \gamma}\left(\tau, \tau_{0}, \rho\right) \phi_{\gamma i}\left(\tau_{0}\right)
\end{align*}
$$

Thus, we have

$$
\sqrt{n}(\hat{\beta}(\tau, \rho)-\beta(\tau, \rho))=\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\phi}_{\beta i}(\tau, \rho)+o_{p}(1)
$$

uniformly in $\left(\rho, \tau_{0}, \tau\right) \in \varrho \times\left[\tau_{l}, \tau_{u}\right] \times\left[\tau_{l}^{*}, \tau_{u}^{*}\right]$, where

$$
\tilde{\phi}_{\beta i}(\tau, \rho)=\frac{1}{\# \mathcal{J}_{0}} \sum_{\tau_{0} \in \mathcal{J}_{0}} \tilde{\phi}_{\beta i}\left(\tau, \tau_{0}, \rho\right)
$$

Then from Lemma 2.17 of Pakes and Pollard (1989), we can further deduce that

$$
(\hat{\beta}(\tau, \rho)-\beta(\tau, \rho))-(\hat{\beta}(\tau, \bar{\rho})-\beta(\tau, \bar{\rho}))=o_{p}\left(n^{-1 / 2}\right)
$$

uniformly in $\rho$ and $\bar{\rho}$ in the $o_{p}(1)$ neighborhood of $\rho_{0}$.
Proof of Theorem 1: We first establish consistency. Similar to the arguments in the proof of Lemma A1, we can establish the following uniform convergence result

$$
\sup _{\rho \in \varrho} \sup _{\tau_{0}, \tau \in\left[\tau_{\nu}, \tau_{u}\right] \times\left[\tau_{1}^{*}, \tau_{u i}^{*}\right]}\left\|G_{2 n}\left(\hat{\beta}(\tau, \rho), \hat{\gamma}\left(\tau_{0}\right), \rho, \delta_{n}, \tau_{0}, \tau\right)-G_{2}\left(\hat{\beta}(\tau, \rho), \hat{\gamma}\left(\tau_{0}\right), \rho, \delta_{n}, \tau_{0}, \tau\right)\right\|=o_{p}(1)
$$

In addition, from Lemma A1 and Assumption A. 5 , we can deduce that

$$
\sup _{\rho \in \varrho} \sup _{\tau_{0}, \tau \in\left[\tau_{l}, \tau_{u}\right] \times\left[\tau_{l}^{*}, \tau_{u}^{*}\right]}\left\|G_{2}\left(\hat{\beta}(\tau, \rho), \hat{\gamma}\left(\tau_{0}\right), \rho, \delta_{n}, \tau_{0}, \tau\right)-G_{2}\left(\beta(\tau, \rho), \gamma\left(\tau_{0}\right), \rho, 0, \tau_{0}, \tau\right)\right\|=o_{p}(1)
$$

where

$$
G_{2}\left(b, \gamma, \beta, \rho, \delta, \tau_{0}, \tau\right)=E\left[g_{2}\left(\xi_{i}, b, \gamma, \rho, \delta, \tau, \tau_{0}\right)\right]
$$

Consequently, it is straightforward to show that

$$
G_{2 n}(\rho)=\bar{G}_{2}(\rho)+o_{p}(1)
$$

uniformly in $\rho \in \varrho$, which, together with the Assumption A. 8 and the fact that $\bar{G}_{2}(\rho)$ is continuous and the parameter space for $\rho$ is compact, establish the consistency of $\hat{\rho}$.

We now follow the arguments of Pakes and Pollard (1989) to establish $\sqrt{n}$-consistency and asymptotic normality. First we linearize the appropriate moment equations. Similar to the arguments in the proof of

Lemma A1, we can show that

$$
\begin{aligned}
& G_{2 n}\left(\hat{\beta}(\tau, \rho), \hat{\gamma}\left(\tau_{0}\right), \rho, \delta_{n}, \tau_{0}, \tau\right) \\
= & G_{2 n}\left(\beta\left(\tau, \rho_{0}\right), \gamma\left(\tau_{0}\right), \rho_{0}, \delta_{n}, \tau_{0}, \tau\right) \\
& +\partial G_{2 \beta}\left(\tau, \tau_{0}, \rho_{0}\right)\left(\hat{\beta}\left(\tau, \rho_{0}\right)-\beta\left(\tau, \rho_{0}\right)\right) \\
& +\partial G_{2 \gamma}\left(\tau, \tau_{0}, \rho_{0}\right)\left(\hat{\gamma}\left(\tau_{0}\right)-\gamma\left(\tau_{0}\right)\right) \\
& +\left(\partial G_{2 \beta}\left(\tau, \tau_{0}, \rho_{0}\right) \frac{\partial}{\partial \rho} \beta\left(\tau, \rho_{0}\right)+\partial G_{2 \rho}\left(\tau, \tau_{0}, \rho_{0}\right)\right)\left(\rho-\rho_{0}\right)+o_{p}\left(\left(\rho-\rho_{0}\right)+n^{-1 / 2}\right) \\
= & L_{2 n}^{*}\left(\tau_{0}, \tau\right)+L_{2}\left(\tau_{0}, \tau\right)\left(\rho-\rho_{0}\right)+o_{p}\left(\left(\rho-\rho_{0}\right)+n^{-1 / 2}\right)
\end{aligned}
$$

uniformly in $\tau \in\left[\tau_{l}^{*}, \tau_{u}^{*}\right], \tau_{0} \in\left[\tau_{l}, \tau_{u}\right]$ and $\rho$ in the $o_{p}(1)$-neighborhood of $\rho_{0}$, where

$$
\begin{aligned}
\tilde{L}_{2 n}\left(\tau_{0}, \tau\right)= & G_{2 n}\left(\beta\left(\tau, \rho_{0}\right), \gamma\left(\tau_{0}\right), \rho_{0}, \tau_{0}, \tau\right) \\
& +\frac{1}{n} \sum_{i=1}^{n}\left(\partial G_{2 \beta}\left(\tau, \tau_{0}, \rho_{0}\right) \tilde{\phi}_{\beta i}\left(\tau, \rho_{0}\right)+\partial G_{2 \gamma}\left(\tau, \tau_{0}, \rho_{0}\right) \phi_{\gamma i}\left(\tau_{0}\right)\right)
\end{aligned}
$$

and

$$
L_{2}\left(\tau_{0}, \tau\right)=\partial G_{2 \beta}\left(\tau, \tau_{0}, \rho_{0}\right) \frac{\partial}{\partial \rho} \beta\left(\tau, \rho_{0}\right)+\partial G_{2 \rho}\left(\tau, \tau_{0}, \rho_{0}\right)
$$

Then, following the arguments in Pakes and Pollard (1989), we can establish the $\sqrt{n}$ consistency of $\hat{\rho}$, and furthermore

$$
\begin{aligned}
\sqrt{n}\left(\hat{\rho}-\rho_{0}\right) & =\left(\sum_{\tau_{0} \in \mathcal{J}_{0}, \tau \in \mathcal{J}_{1}} L_{2}^{\prime}\left(\tau_{0}, \tau\right) L_{2}\left(\tau_{0}, \tau\right)\right)^{-1} \sum_{\tau_{0} \in \mathcal{J}_{0}, \tau \in \mathcal{J}_{1}} L_{2}^{\prime}\left(\tau_{0}, \tau\right) \sqrt{n} \tilde{L}_{2 n}\left(\tau_{0}, \tau\right)+o_{p}(1) \\
& =\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{\rho i}+o_{p}(1)
\end{aligned}
$$

where

$$
\begin{aligned}
\sqrt{n} \tilde{L}_{2 n}\left(\tau_{0}, \tau\right)= & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_{2}\left(\xi_{i}, \beta\left(\tau, \rho_{0}\right), \gamma\left(\tau_{0}\right), \rho_{0}, 0, \tau, \tau_{0}\right) \\
& +\partial G_{2 \beta}\left(\tau, \tau_{0}, \rho_{0}\right) \tilde{\phi}_{\beta i}(\tau)+\partial G_{2 \gamma}\left(\tau, \tau_{0}, \rho_{0}\right) \phi_{\gamma i}\left(\tau_{0}\right)
\end{aligned}
$$

Then, by (A3) of Lemma A1, we obtain

$$
\begin{aligned}
\sqrt{n}\left(\hat{\beta}(\tau, \hat{\rho})-\beta\left(\tau, \rho_{0}\right)\right) & =\sqrt{n}\left(\hat{\beta}\left(\tau, \rho_{0}\right)-\beta(\tau)\right)+\sqrt{n}\left(\beta(\tau, \hat{\rho})-\beta\left(\tau, \rho_{0}\right)\right)+o_{p}(1) \\
& =\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\phi}_{\beta i}(\tau)+\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \beta\left(\tau, \rho_{0}\right)}{\partial \rho} \phi_{\rho i}+o_{p}(1) \\
& =\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{\beta i}(\tau)+o_{p}(1)
\end{aligned}
$$

where

$$
\phi_{\beta i}(\tau)=\tilde{\phi}_{\beta i}(\tau)+\frac{\partial \beta\left(\tau, \rho_{0}\right)}{\partial \rho} \phi_{\rho i}
$$

Therefore, we have

$$
\sqrt{n}\left[\begin{array}{c}
\hat{\beta}(\tau, \hat{\rho})-\beta(\tau) \\
\hat{\rho}-\rho_{0}
\end{array}\right]=\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{\theta i}(\tau)+o_{p}(1)
$$

where $\phi_{\theta i}(\tau)=\left(\phi_{\beta i}^{\prime}(\tau), \phi_{\rho i}^{\prime}\right)^{\prime}$. Consequently, Theorem 1 follows by applying the Central Limit Theorem. Proof of Theorem 2: Following the proof of Theorem 1, we can establish the consistency of $\hat{\rho}^{*}$ and $\hat{\beta}^{*}\left(\tau, \hat{\rho}^{*}\right)$ jointly in space $P \times P_{\eta}$. Furthermore, we can also establish the asymptotic linear representations

$$
\sqrt{n}\left[\begin{array}{c}
\hat{\beta}(\tau, \hat{\rho})-\beta(\tau) \\
\hat{\rho}-\rho_{0}
\end{array}\right]=\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{\theta i}(\tau)+o_{p}(1)
$$

and

$$
\sqrt{n}\left[\begin{array}{c}
\hat{\beta}^{*}\left(\tau, \hat{\rho}^{*}\right)-\beta(\tau) \\
\hat{\rho}^{*}-\rho_{0}
\end{array}\right]=\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \eta_{i} \phi_{\theta i}(\tau)+o_{p}(1)
$$

Therefore, we have

$$
\sqrt{n}\left[\begin{array}{c}
\hat{\beta}^{*}\left(\tau, \hat{\rho}^{*}\right)-\hat{\beta}(\tau, \hat{\rho}) \\
\hat{\rho}^{*}-\hat{\rho}
\end{array}\right]=\frac{1}{\sqrt{n}} \sum_{i=1}^{n}\left(\eta_{i}-1\right) \phi_{\theta i}(\tau)+o_{p}(1)
$$

Then Theorem 2 follows from the conditional Central Limit Theorem (Th. 2.9.6, van der Vaart and Wellner, 1996).

## Appendix B. Tables for simulation

Table 2: Simulation results for homoscedastic case with Gaussian copula

| $\tau$ | $\beta(\tau)$ | $n=250$ |  |  | $n=500$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Bias | SD | RMSE | Bias | SD | RMSE |
| 0.10 | -2.282 | 0.451 | 0.173 | 0.483 | 0.299 | 0.140 | 0.330 |
|  | 1.000 | -0.001 | 0.189 | 0.188 | -0.005 | 0.146 | 0.146 |
|  | 1.000 | 0.005 | 0.177 | 0.177 | -0.005 | 0.141 | 0.141 |
| 0.25 | -1.674 | 0.092 | 0.177 | 0.199 | 0.039 | 0.134 | 0.140 |
|  | 1.000 | -0.005 | 0.157 | 0.157 | -0.005 | 0.115 | 0.115 |
|  | 1.000 | 0.002 | 0.154 | 0.154 | 0.000 | 0.112 | 0.112 |
| 0.50 | -1.000 | -0.028 | 0.161 | 0.163 | -0.037 | 0.125 | 0.130 |
|  | 1.000 | -0.002 | 0.131 | 0.131 | 0.002 | 0.098 | 0.098 |
|  | 1.000 | -0.004 | 0.131 | 0.131 | 0.009 | 0.089 | 0.090 |
| 0.75 | -0.326 | -0.031 | 0.138 | 0.141 | -0.021 | 0.108 | 0.110 |
|  | 1.000 | -0.008 | 0.125 | 0.125 | 0.001 | 0.094 | 0.093 |
|  | 1.000 | -0.005 | 0.127 | 0.127 | 0.006 | 0.089 | 0.089 |
| 0.90 | 0.282 | -0.024 | 0.135 | 0.136 | -0.015 | 0.103 | 0.104 |
|  | 1.000 | -0.007 | 0.145 | 0.145 | -0.003 | 0.103 | 0.103 |
|  | 1.000 | -0.005 | 0.142 | 0.142 | 0.000 | 0.100 | 0.099 |
| $\rho$ | $\rho_{0}=0.700$ | 0.019 | 0.100 | 0.101 | 0.017 | 0.067 | 0.069 |

Table 3: Simulation results for homoscedastic case with Frank copula

| $\tau$ | $\beta(\tau)$ | $n=250$ |  |  | $n=500$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Bias | SD | RMSE | Bias | SD | RMSE |
| 0.10 | -2.282 | 0.359 | 0.249 | 0.437 | 0.172 | 0.215 | 0.275 |
|  | 1.000 | 0.006 | 0.230 | 0.230 | 0.009 | 0.201 | 0.200 |
|  | 1.000 | -0.012 | 0.247 | 0.247 | 0.022 | 0.205 | 0.206 |
| 0.25 | -1.674 | 0.017 | 0.211 | 0.212 | -0.038 | 0.192 | 0.195 |
|  | 1.000 | 0.001 | 0.185 | 0.185 | 0.000 | 0.145 | 0.145 |
|  | 1.000 | 0.001 | 0.200 | 0.200 | 0.018 | 0.152 | 0.153 |
| 0.50 | -1.000 | -0.078 | 0.162 | 0.179 | -0.053 | 0.134 | 0.144 |
|  | 1.000 | -0.004 | 0.134 | 0.134 | 0.007 | 0.089 | 0.089 |
|  | 1.000 | 0.003 | 0.136 | 0.136 | 0.011 | 0.095 | 0.095 |
| 0.75 | -0.326 | -0.045 | 0.126 | 0.133 | -0.031 | 0.091 | 0.095 |
|  | 1.000 | 0.002 | 0.118 | 0.118 | 0.006 | 0.079 | 0.079 |
|  | 1.000 | 0.002 | 0.118 | 0.118 | -0.001 | 0.078 | 0.078 |
| 0.90 | 0.282 | -0.027 | 0.137 | 0.140 | -0.021 | 0.093 | 0.095 |
|  | 1.000 | 0.008 | 0.141 | 0.141 | 0.010 | 0.100 | 0.101 |
|  | 1.000 | 0.003 | 0.135 | 0.135 | -0.001 | 0.090 | 0.089 |
| $\rho$ | $\rho_{0}=5.628$ | 0.921 | 1.813 | 2.031 | 0.512 | 1.215 | 1.317 |
| Kendall's $\tau$ | true $=0.494$ | 0.035 | 0.083 | 0.090 | 0.021 | 0.061 | 0.065 |

Table 4: Simulation results for heteroscedastic case with Gaussian copula

|  |  | $n=250$ |  |  |  |  | $n=500$ |  |  |
| :--- | :---: | ---: | ---: | ---: | ---: | ---: | ---: | ---: | :---: |
| $\tau$ | $\beta(\tau)$ | Bias | SD | RMSE |  | Bias | SD | RMSE |  |
| 0.10 | -2.282 | 0.404 | 0.181 | 0.443 |  | 0.255 | 0.159 | 0.301 |  |
|  | 0.487 | 0.277 | 0.193 | 0.338 |  | 0.215 | 0.152 | 0.263 |  |
|  | 1.000 | -0.003 | 0.187 | 0.186 |  | -0.010 | 0.153 | 0.153 |  |
| 0.25 | -1.674 |  | 0.064 | 0.177 | 0.188 |  | 0.024 | 0.139 |  |
|  | 0.730 | 0.099 | 0.161 | 0.189 |  | 0.073 | 0.113 | 0.141 |  |
|  | 1.000 | -0.006 | 0.161 | 0.161 |  | -0.004 | 0.118 | 0.118 |  |
|  |  |  |  |  |  | -0.033 | 0.118 | 0.122 |  |
| 0.50 | -1.000 | -0.039 | 0.162 | 0.166 |  | 0.013 | 0.087 | 0.088 |  |
|  | 1.000 | 0.014 | 0.137 | 0.138 |  | 0.008 | 0.088 | 0.089 |  |
|  | 1.000 | -0.007 | 0.133 | 0.133 |  | -0.016 | 0.104 | 0.105 |  |
|  |  | -0.326 | -0.033 | 0.133 | 0.137 |  | -0.002 | 0.085 |  |
| 0.75 | 1.270 | -0.016 | 0.121 | 0.122 |  | 0.085 |  |  |  |
|  | 1.000 | -0.007 | 0.126 | 0.126 |  | 0.006 | 0.087 | 0.087 |  |
|  | 0.282 | -0.024 | 0.134 | 0.136 |  | -0.009 | 0.101 | 0.101 |  |
|  | 1.513 | -0.035 | 0.143 | 0.147 |  | -0.016 | 0.096 | 0.097 |  |
| 0.90 | 1.000 | -0.005 | 0.139 | 0.139 |  | -0.000 | 0.096 | 0.095 |  |
| $\rho$ | $\rho_{0}=0.700$ | 0.020 | 0.099 | 0.101 |  | 0.011 | 0.066 | 0.067 |  |

Table 5: Simulation results for heteroscedastic case with Frank copula

| $\tau$ | $\beta(\tau)$ | $n=250$ |  |  | $n=500$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Bias | SD | RMSE | Bias | SD | RMSE |
| 0.10 | -2.282 | 0.298 | 0.276 | 0.406 | 0.124 | 0.232 | 0.263 |
|  | 0.487 | 0.288 | 0.245 | 0.378 | 0.219 | 0.207 | 0.301 |
|  | 1.000 | -0.003 | 0.268 | 0.268 | 0.028 | 0.212 | 0.213 |
| 0.25 | -1.674 | -0.018 | 0.231 | 0.231 | -0.049 | 0.200 | 0.205 |
|  | 0.730 | 0.101 | 0.197 | 0.221 | 0.073 | 0.140 | 0.157 |
|  | 1.000 | 0.003 | 0.205 | 0.205 | 0.016 | 0.152 | 0.153 |
| 0.50 | -1.000 | -0.090 | 0.172 | 0.194 | -0.050 | 0.135 | 0.143 |
|  | 1.000 | 0.006 | 0.136 | 0.136 | 0.011 | 0.088 | 0.089 |
|  | 1.000 | 0.004 | 0.136 | 0.136 | 0.007 | 0.093 | 0.093 |
| 0.75 | -0.326 | -0.044 | 0.123 | 0.131 | -0.024 | 0.092 | 0.095 |
|  | 1.270 | -0.012 | 0.119 | 0.119 | -0.003 | 0.076 | 0.076 |
|  | 1.000 | 0.001 | 0.118 | 0.118 | -0.002 | 0.076 | 0.076 |
| 0.90 | 0.282 | -0.019 | 0.136 | 0.138 | -0.010 | 0.096 | 0.097 |
|  | 1.513 | -0.017 | 0.137 | 0.138 | -0.011 | 0.093 | 0.094 |
|  | 1.000 | 0.005 | 0.132 | 0.132 | -0.001 | 0.090 | 0.090 |
| $\rho$ | $\rho_{0}=5.628$ | 0.911 | 1.876 | 2.083 | 0.422 | 1.202 | 1.273 |
| Kendall's $\tau$ | true $=0.494$ | 0.034 | 0.085 | 0.091 | 0.016 | 0.062 | 0.064 |

Table 6: Simulation results for homoscedastic Gaussian case without excluded covariates

| $\tau$ | $\beta(\tau)$ | $n=250$ |  |  | $n=500$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Bias | SD | RMSE | Bias | SD | RMSE |
| 0.10 | -2.282 | 0.430 | 0.191 | 0.470 | 0.276 | 0.145 | 0.312 |
|  | 1.000 | 0.010 | 0.200 | 0.200 | -0.008 | 0.152 | 0.152 |
|  | 1.000 | 0.020 | 0.194 | 0.195 | 0.012 | 0.150 | 0.150 |
| 0.25 | -1.674 | 0.098 | 0.183 | 0.207 | 0.035 | 0.142 | 0.146 |
|  | 1.000 | 0.003 | 0.162 | 0.162 | -0.002 | 0.123 | 0.123 |
|  | 1.000 | 0.012 | 0.159 | 0.160 | -0.001 | 0.113 | 0.113 |
| 0.50 | -1.000 | -0.018 | 0.174 | 0.175 | -0.034 | 0.121 | 0.125 |
|  | 1.000 | 0.005 | 0.129 | 0.129 | 0.006 | 0.095 | 0.095 |
|  | 1.000 | 0.005 | 0.137 | 0.137 | 0.007 | 0.092 | 0.092 |
| 0.75 | -0.326 | -0.031 | 0.143 | 0.146 | -0.019 | 0.101 | 0.102 |
|  | 1.000 | 0.014 | 0.127 | 0.127 | 0.004 | 0.097 | 0.096 |
|  | 1.000 | 0.003 | 0.123 | 0.123 | 0.005 | 0.094 | 0.094 |
| 0.90 | 0.282 | -0.026 | 0.135 | 0.137 | -0.013 | 0.101 | 0.102 |
|  | 1.000 | 0.020 | 0.149 | 0.150 | 0.003 | 0.105 | 0.105 |
|  | 1.000 | 0.004 | 0.135 | 0.135 | 0.001 | 0.104 | 0.104 |
| $\rho$ | $\rho_{0}=0.700$ | 0.015 | 0.103 | 0.104 | 0.017 | 0.070 | 0.072 |

Table 7: Simulation results for homoscedastic Frank case without excluded covariates

| $\tau$ | $\beta(\tau)$ | $n=250$ |  |  | $n=500$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Bias | SD | RMSE | Bias | SD | RMSE |
| 0.10 | -2.282 | 0.366 | 0.259 | 0.448 | 0.158 | 0.230 | 0.279 |
|  | 1.000 | -0.012 | 0.234 | 0.234 | -0.002 | 0.228 | 0.228 |
|  | 1.000 | -0.001 | 0.246 | 0.246 | 0.011 | 0.200 | 0.200 |
| 0.25 | -1.674 | 0.038 | 0.229 | 0.232 | -0.035 | 0.193 | 0.196 |
|  | 1.000 | -0.007 | 0.184 | 0.184 | -0.007 | 0.163 | 0.163 |
|  | 1.000 | -0.009 | 0.199 | 0.198 | 0.004 | 0.144 | 0.143 |
| 0.50 | -1.000 | -0.063 | 0.180 | 0.190 | -0.054 | 0.142 | 0.152 |
|  | 1.000 | -0.001 | 0.119 | 0.118 | -0.003 | 0.100 | 0.100 |
|  | 1.000 | 0.002 | 0.134 | 0.134 | -0.001 | 0.093 | 0.093 |
| 0.75 | -0.326 | -0.039 | 0.129 | 0.135 | -0.033 | 0.097 | 0.102 |
|  | 1.000 | 0.002 | 0.106 | 0.105 | 0.001 | 0.083 | 0.083 |
|  | 1.000 | 0.002 | 0.115 | 0.115 | 0.003 | 0.081 | 0.081 |
| 0.90 | 0.282 | -0.031 | 0.134 | 0.137 | -0.016 | 0.095 | 0.097 |
|  | 1.000 | 0.001 | 0.127 | 0.127 | 0.001 | 0.096 | 0.096 |
|  | 1.000 | -0.005 | 0.130 | 0.130 | 0.005 | 0.098 | 0.098 |
| $\rho$ | $\rho_{0}=5.628$ | 0.843 | 1.798 | 1.984 | 0.563 | 1.330 | 1.443 |
| Kendall's $\tau$ | true $=0.494$ | 0.031 | 0.088 | 0.093 | 0.022 | 0.066 | 0.070 |

Table 8: Simulation results for heteroscedastic Gaussian case without excluded covariates

| $\tau$ | $\beta(\tau)$ | $n=250$ |  |  | $n=500$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Bias | SD | RMSE | Bias | SD | RMSE |
| 0.10 | -2.282 | 0.395 | 0.198 | 0.442 | 0.245 | 0.156 | 0.290 |
|  | 0.487 | 0.276 | 0.227 | 0.358 | 0.185 | 0.168 | 0.250 |
|  | 1.000 | 0.008 | 0.215 | 0.215 | -0.005 | 0.173 | 0.173 |
| 0.25 | -1.674 | 0.076 | 0.190 | 0.204 | 0.019 | 0.146 | 0.147 |
|  | 0.730 | 0.102 | 0.178 | 0.205 | 0.062 | 0.130 | 0.144 |
|  | 1.000 | 0.001 | 0.176 | 0.176 | -0.008 | 0.127 | 0.127 |
| 0.50 | -1.000 | -0.030 | 0.175 | 0.177 | -0.039 | 0.122 | 0.128 |
|  | 1.000 | 0.023 | 0.132 | 0.134 | 0.012 | 0.105 | 0.105 |
|  | 1.000 | 0.002 | 0.147 | 0.147 | 0.004 | 0.101 | 0.101 |
| 0.75 | -0.326 | -0.032 | 0.142 | 0.145 | -0.019 | 0.103 | 0.104 |
|  | 1.270 | 0.010 | 0.133 | 0.133 | 0.001 | 0.099 | 0.099 |
|  | 1.000 | 0.007 | 0.133 | 0.133 | 0.007 | 0.102 | 0.102 |
| 0.90 | 0.282 | -0.018 | 0.131 | 0.132 | -0.011 | 0.100 | 0.100 |
|  | 1.513 | 0.004 | 0.157 | 0.157 | -0.009 | 0.108 | 0.108 |
|  | 1.000 | 0.007 | 0.148 | 0.148 | 0.003 | 0.114 | 0.114 |
| $\rho$ | $\rho_{0}=0.700$ | 0.017 | 0.100 | 0.102 | 0.017 | 0.072 | 0.073 |

Table 9: Simulation results for heteroscedastic Frank case without excluded covariates

| $\tau$ | $\beta(\tau)$ | $n=250$ |  |  | $n=500$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Bias | SD | RMSE | Bias | SD | RMSE |
| 0.10 | -2.282 | 0.313 | 0.286 | 0.423 | 0.124 | 0.247 | 0.276 |
|  | 0.487 | 0.257 | 0.277 | 0.378 | 0.174 | 0.237 | 0.294 |
|  | 1.000 | -0.016 | 0.288 | 0.288 | -0.010 | 0.240 | 0.240 |
| 0.25 | -1.674 | 0.007 | 0.247 | 0.247 | -0.064 | 0.208 | 0.218 |
|  | 0.730 | 0.090 | 0.210 | 0.228 | 0.050 | 0.165 | 0.172 |
|  | 1.000 | -0.018 | 0.220 | 0.221 | -0.003 | 0.169 | 0.169 |
| 0.50 | -1.000 | -0.074 | 0.187 | 0.201 | -0.061 | 0.145 | 0.157 |
|  | 1.000 | 0.009 | 0.136 | 0.137 | -0.004 | 0.105 | 0.105 |
|  | 1.000 | 0.001 | 0.144 | 0.144 | -0.001 | 0.103 | 0.103 |
| 0.75 | -0.326 | -0.044 | 0.132 | 0.138 | -0.033 | 0.097 | 0.103 |
|  | 1.270 | -0.010 | 0.121 | 0.121 | -0.008 | 0.085 | 0.085 |
|  | 1.000 | 0.003 | 0.126 | 0.126 | 0.005 | 0.088 | 0.088 |
| 0.90 | 0.282 | -0.030 | 0.135 | 0.138 | -0.015 | 0.098 | 0.099 |
|  | 1.513 | -0.024 | 0.144 | 0.146 | -0.011 | 0.102 | 0.102 |
|  | 1.000 | -0.001 | 0.142 | 0.142 | 0.005 | 0.103 | 0.103 |
| $\rho$ | $\rho_{0}=5.628$ | 0.903 | 1.896 | 2.098 | 0.568 | 1.350 | 1.463 |
| Kendall's $\tau$ | true $=0.494$ | 0.033 | 0.089 | 0.095 | 0.022 | 0.067 | 0.071 |

## References

[1] Amemiya, T. (1978), The estimation of a simultaneous equation generalized probit model, Econometrica, 46, 1193-1205.
[2] Amemiya, T. (1979), The estimation of a simultaneous tobit model, International Economic Review, 20, 169-81.
[3] Amemiya, T. (1985), Advanced Econometrics, Harvard University Press.
[4] Arellano, M. and S. Bonhomme (2017), Quantile selection models: with an application to understanding changes in wage inequality, Econometrica, 85(1), 1-28.
[5] Arellano, M. and S. Bonhomme (2017), Supplemental material to "quantile selection models with an application to understanding changes in wage inequality", Econometrica, 85(1), 1-28.
[6] Blau, F. D., and L. M. Kahn (2017), The gender wage gap: Extent, trends, and explanations, Journal of economic literature, 55(3), 789-865.
[7] 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(2), 323-363.
[8] Blundell, R., T. MaCurdy, and C. Meghir (2007), Labor supply models: Unobserved heterogeneity, nonparticipation and dynamics, Handbook of Econometrics, 6, 4667-4775.
[9] 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.
[10] Buchinsky, M. (1994), Changes in the US wage structure 1963-1987: Application of quantile regression, Econometrica, 405-458.
[11] 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), 1-30.
[12] Buchinsky, M. (2001). Quantile regression with sample selection: Estimating women's return to education in the US. Empirical Economics, 26(1), 87-113.
[13] Carroll, R. J., J. Fan, I. Gijbels, and M.P. Wand (1997), Generalized partially linear single-index models, Journal of the American Statistical Association, 92, 438, 477-489.
[14] Chamberlain, G. (1994), Quantile regression, censoring, and the structure of wages, Advances in econometrics: sixth world congress, Vol. 2, pp. 171-209.
[15] Chen, S. (1997), Semiparametric estimation of the Type-3 Tobit model, Journal of Econometrics, 80, 1-34.
[16] Chen, S. (2018), Sequential estimation of censored quantile regression models, Journal of Econometrics, 207, 1, 30-52.
[17] Chen, X., O. Linton, and I. Van Keilegom (2003), Estimation of semiparametric models when the criterion function is not smooth, Econometrica, 71(5), 1591-1608.
[18] Chernozhukov, V., I. Fernández-Val, and A. E. Kowalski (2015), Quantile regression with censoring and endogeneity, Journal of Econometrics, 186(1), 201-221.
[19] Chernozhukov, V., I. Fernández-Val, and S. Luo (2022), Distribution regression with sample selection, with an application to wage decompositions in the UK, Working paper.
[20] Chernozhukov, V and C. Hansen (2006), Instrumental quantile regression inference for structural and treatment effect models, Journal of Econometrics, 132, 2, 491-525.
[21] Chernozhukov, V., and C. Hansen (2008), Instrumental variable quantile regression: A robust inference approach, Journal of Econometrics, 142(1), 379-398.
[22] Chernozhukov, V and H. Hong (2002), Three-step censored quantile regression and extramarital affairs, Journal of the American Statistical Association, 97, 459, 872-882.
[23] Christofides, L. N., Q. Li, Z. Liu, and I. Min (2003), Recent two-stage sample selection procedures with an application to the gender wage gap, Journal of Business \& Economic Statistics, 21(3), 396-405.
[24] Dolton, P. J. and G. H. Makepeace (1987), Marital status, child rearing and earnings differentials in the graduate labour market, The Economic Journal, 97(388), 897-922.
[25] Ermisch, J. F. and R. E. Wright (1994), Interpretation of negative sample selection effects in wage offer equations, Applied Economics Letters, 1(11), 187-189.
[26] Fernández-Val, I., F. Peracchi, A. van Vuuren, and F. Vella (2019), Decomposing changes in the distribution of real hourly wages in the US, Cemmap Working Paper.
[27] Fernández-Val, I., A. van Vuuren, and F. Vella (2021). Nonseparable sample selection models with censored selection rules. Journal of Econometrics.
[28] Gosling, A., S. Machin, and C. Meghir (2000), The changing distribution of male wages in the UK, The Review of Economic Studies, 67(4), 635-666.
[29] Gronau, R. (1974), Wage comparisons-A selectivity bias, Journal of Political Economy, 82(6), 1119-1143.
[30] Heckman, J. J. (1974), Shadow prices, market wages, and labor supply, Econometrica, 679-694.
[31] Heckman, J. J. (1979), Sample selection bias as a specification error, Econometrica, 47, 1, 153-161.
[32] Heckman, J. J. and G. L. Sedlacek (1990), Self-selection and the distribution of hourly wages, Journal of Labor Economics, 8(1, Part 2), S329-S363.
[33] Honoré, B. E. (1992), Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects, Econometrica, 60, 3, 533-565.
[34] Honoré, B., E. Kyriazidou and C. Udry (1997), Estimation of type 3 tobit models using symmetric trimming and pairwise comparisons, Journal of Econometrics, 76, 107-28.
[35] Huber, M. and B. Melly (2015), A test of the conditional independence assumption in sample selection models, Journal of Applied Econometrics, 30(7), 1144-1168.
[36] Juhn, C., K. M. Murphy, and B. Pierce (1993), Wage inequality and the rise in returns to skill, Journal of Political Economy, 101(3), 410-442.
[37] Killingsworth, M. R. and J. J. Heckman (1986), Female labor supply: A survey, Handbook of Labor Economics, 1, 103-204.
[38] Koenker, R and G. Bassett (1978), Regression quantiles, Econometrica, 46, 1, 33-50.
[39] Krueger, A. B. and D. M. Whitmore (2001), The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR, The Economic Journal, 111(468), 1-28.
[40] Lee, D. S. (2009), Training, wages, and sample selection: Estimating sharp bounds on treatment effects, The Review of Economic Studies, 76(3), 1071-1102.
[41] Lee, L. F. (1994), Semiparametric two-stage estimation of sample selection models subject to Tobit-type selection rules, Journal of Econometrics, 61(2), 305-344.
[42] Lee, M.J. and F. Vella (2006), A semi-parametric estimator for censored selection models with endogeneity, Journal of Econometrics, 130, 235-52.
[43] Maasoumi, E. and L. Wang (2019), The gender gap between earnings distributions. Journal of Political Economy, 127(5), 2438-2504.
[44] Machado, J. A. and J. Mata (2005), Counterfactual decomposition of changes in wage distributions using quantile regression, Journal of Applied Econometrics, 20(4), 445-465.
[45] Mocan, N. and B. Unel (2017), Skill-Biased Technological Change, Earnings of Unskilled Workers, and Crime, Review of Law E Economics, 13(3).
[46] Mulligan, C. B. and Y. Rubinstein (2008). Selection, investment, and women's relative wages over time. The Quarterly Journal of Economics, 123(3), 1061-1110.
[47] Newey, W. K. and D. McFadden (1994), Large sample estimation and hypothesis testing, Handbook of Econometrics, 4, 2111-2245.
[48] Pakes, A. and D. Pollard (1989), Simulation and the asymptotics of optimization estimators, Econometrica, 57, 5, 1027-1057.
[49] Powell, J. (1984), Least absolute deviations estimation for the censored regression model, Journal of Econometrics, 25, 3, 303-325.
[50] Powell, J. (1986), Censored regression quantiles, Journal of Econometrics, 32, 1, 143-155.
[51] Steinberg, D. (1989), Induced work participation and the returns to experience for welfare women: Evidence from a social experiment, Journal of Econometrics, 41(3), 321-340.
[52] van der Vaart, A. W., and J. Wellner (1996), Weak convergence and empirical processes: with applications to statistics, Springer.
[53] Vella, F. (1993), A simple estimator for simultaneous models with censored endogenous regressors, International Economic Review, 34, 441-57
[54] Wooldridge, J. M. (1998), Selection corrections with a censored selection variable, mimeo, Michigan State University Department of Economics.
[55] Wright, R. E. and J. F. Ermisch (1991), Gender discrimination in the British labour market: a reassessment, The Economic Journal, 101(406), 508-522.


[^0]:    *Tsingshan Institute for Advanced Business Study and School of Economics, Zhejiang University, Hangzhou, China.
    ${ }^{\dagger}$ School of Economics, The Shanghai University of Finance and Economics (SUFE), Shanghai, China.
    $\ddagger$ School of Economics and Dishui Lake Advanced Finance institute, The Shanghai University of Finance and Economics (SUFE), Shanghai, China.

[^1]:    ${ }^{1}$ Arellano and Bonhomme (2017) also considered identification and estimation with nonparametric specification for the selection equation. However, they mainly focused on the estimation procedure which requires the probit specification for the selection equation. They discussed various extensions with nonparametric selection equation, but these extensions are generally not very practical, due to curse of dimensionality. In addition, the probit specification imposes a strong homoscedasticity assumption, which typically does not hold in practice.

[^2]:    ${ }^{2}$ We thank one referee for pointing this out.
    ${ }^{3}$ The data set is publicly available through https://www.econometricsociety.org/content/ supplement-quantile-selection-models-application-understanding-changes-wage-inequality-0.

[^3]:    ${ }^{4}$ Our estimator has been implemented by an $R$ package available through https://drive.google.com/drive/folders/ 1G6SSQqVDYYakUn11JGX9fVBkZ8Sv3hoc?usp=sharing.

[^4]:    ${ }^{5}$ As in Arellano and Bonhomme (2017), we can relax this assumption by allowing the copula function to depend on the regressors with a parametric structure.

[^5]:    ${ }^{6}$ Both the Gaussian copula of $C_{G}^{*}(\cdot, \cdot, 0.7)$ and Frank copula of $C_{F}^{*}(\cdot, \cdot, 5.628)$ have a Kendall's tau coefficient of about 0.494 .
    ${ }^{7}$ We follow the same strategy in the empirical application. Also note that the semiparametric estimation procedure in Arellano and Bonhomme (2017) does not involve choosing a tuning parameter.

[^6]:    ${ }^{8}$ Note that the parameter $\rho$ of Gaussian copula is in $[-1,1]$, but the one of Frank copula is in $\mathbb{R} /\{0\}$. We report the statistics for both $\rho$ and Kendall's $\tau$, which is in $[-1,1]$, in the Frank copula case.

[^7]:    ${ }^{9}$ The raw data of UK Family Expenditure Survey can be accessed in the UK Data Service through https://www. ukdataservice. ac.uk.
    ${ }^{10}$ This drops about $2.95 \%$ of the observations that remained in the sample after previous cleaning procedure.

[^8]:    ${ }^{11}$ Blundell et al. (2007) and Arellano and Bonhomme (2017) also use this constructed measure of out-of-income to study the changes in wage distributions of males and females in UK.
    ${ }^{12}$ Interdecile rage (IDR) is defined as the difference between the 9 th decile and 1st decile.

[^9]:    ${ }^{13} \mathrm{We}$ also tried $\varphi\left(X, \tau_{0}\right)=\left(1, X,\left[X^{\prime} \gamma\left(\tau_{0}\right)\right]^{2}\right)$ and obtained very similar estimation results.
    ${ }^{14} \otimes$ denotes the Kronecker tensor product.
    ${ }^{15}$ We also applied the Type 3 Tobit model in empirical analysis. It turns out the latent average wage curve after the Tobit correction is very similar to the latent median wage curve based on our estimation procedure. However, the strong homogeneity associated with the Tobit 3 model implies parallel latent quantile wage curves, in contrast to our selection corrected heterogeneous decile curves.
    ${ }^{16}$ Our standard errors for the copula parameters are smaller than Arellano and Bonhomme (2017), since our model is more informative with a censored selection (based on working hours) whereas they are working with a binary selection which codes the working hours into binary employment status.
    ${ }^{17}$ A negative copula coefficient of Arellano and Bonhomme(2017) (resp. ours) means positive (resp. negative) selection into employment and vice versa, since Arellano and Bonhomme (2017) adopted a different sign normalization from ours.

[^10]:    ${ }^{18}$ We thank an anonymous referee for suggesting this explanation.

[^11]:    ${ }^{19}$ To further explore possible causes behind the discrepancy between our findings and those of Arellano and Bonhomme (2017), we have conducted some further analysis. In particular, we implement our procedure with the exclusion restriction imposed; it turns out that the results are not significantly different, even though the out-of-work income is statistically significant in the wage equation. Another angle worth exploring is the validity of the probit specification for participation adopted in Arellano and Bonhomme (2017), which rules out heteroscedasticity commonly present in practice. Indeed, some specification analysis indicates that the probit participation for the current application is seriously misspecified.

