# Quantile Regression with Censored Selection and Many Controls

Kaixi Zhang* , Songnian Chen†, and Junlong Feng*

*Department of Economics, the Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong SAR, China

†Tsingshan Institute for Advanced Business Studies, Zhejiang University, Hangzhou, Zhejiang, China

Latest Version

This version: November 18, 2025

**Abstract**

This paper develops a quantile regression model with censored selection in a high-dimensional setting (HD-QRCS). We introduce a semiparametric nonlinear least squares type estimator for the copula parameters, as an alternative to the estimator using the conventional method-of-moments criterion. Our approach is based on the framework of the quantile selection model proposed by Arellano and Bonhomme (2017). This method ensures convergence to the global minimum and avoids iterative procedures. Our estimation procedures consist of three steps: the first two employ $\ell_1$-penalized quantile regression and hard-thresholding in order to mitigate estimation bias and achieve oracle rate of convergence. Lastly, we obtain copula parameters and coefficients of interest. We establish uniform convergence rates for the proposed estimators and discuss the properties of variable selection. Monte Carlo simulations show that our estimators perform well under many controls. We apply our method to American Community Survey (ACS) data from 2001 to 2023 to study wage inequality.

**Keywords:** selection, quantile regression, censoring, high-dimensional model

**JEL Codes:** C14, C31, C34

*We are grateful to [...] for helpful comments. All errors are our own. Email: kzhangbh@connect.ust.hk (K. Zhang, corresponding author), snchen2022@zju.edu.cn (S. Chen), jlfeng@ust.hk (J. Feng).

# 1 Introduction

Censoring and sample selection are common issues in economic research. For example, both arise in the study of wages and labor market participation, where income survey data are often censored due to top-coding or minimal wages, while many other variables like hours worked are naturally bound by zero from below. Additionally, because wages are only considered for employed individuals, employment status and wage level may be influenced by common unobserved variables, such as ability or skills. These problems can lead to inconsistent estimators either for mean regression or quantile regression. Quantile regression (QR) models are a valuable empirical tool for capturing the heterogeneous effects of covariates on the distribution of a response variable. In this paper, we tackle both challenges by developing high dimensional quantile regression models in the presence of sample selection (QRCS), building on the works of Arellano and Bonhomme (2017) and Chen (2018).

The basic idea for addressing sample selection is drawing on the work of Arellano and Bonhomme (2017). They model the sample selection through the bivariate cumulative distribution function, or copula, of errors in the outcome equation and the selection equation. Estimating the copula parameters is a crucial step in this framework, as it allows for the adjustment of percentile levels in standard quantile regression based on the degree of selection. Ultimately, this leads to selection-corrected estimates obtained by minimizing a rotated check function. Chen and Wang (2023) extended Arellano and Bonhomme (2017)'s quantile selection model to take into account censoring. In the existing literature, copula parameters are estimated by minimizing a method-of-moments criterion. These methods, however, have proven to be challenging to implement in practice. Specifically, when a finite number of moment conditions are used for identification, the estimation procedures may converge to a local maximum, while consistency results are applicable only to the global extremum. Special cases occur when the objective function has only one solution and the first-order condition has multiple roots (Newey and McFadden (1994)). Moreover, this method-of-moments criterion requires an iterative estimation procedures, which can also be difficult in practice.

In this paper, we use the semiparametric nonlinear least squares type estimators for copula parameters introduced by Chen et al. (2025). This approach works similarly to traditional nonlinear regression analysis, in which an infinite number of unconditional moment conditions are used for identification without losing any information. As a result, this method does not require any iterative procedures, which helps reduce computational complexity.

Furthermore, our estimators utilize Chen (2018)'s sequential censored quantile regression (CQR) approach to deal with censoring. While Powell (1986) first introduced CQR as a cornerstone, its nonlinear and non-convex loss function poses challenges for implementation. Chen (2018)'s method transforms the optimization problem into standard quantile regression by recognizing that censoring from below is unlikely to affect the top quantiles of the conditional distribution. As a result, quantile regression for the top quantiles can be estimated using standard QR techniques. This framework also provides a natural subsample selector for estimating the lower quantiles through a sequential procedure, leveraging the continuity of the quantile regression coefficient process. Furthermore, due to the widespread availability of datasets, the number of controls $p$ is often quite large, potentially exceeding the sample size $n$. This insight indicates that the previously discussed approach cannot be directly applied when $p > n$. To address this problem, we apply the $\ell_1$-penalized quantile regression (Belloni and Chernozhukov (2011)), and assume that there are only at most $s$ control variables that contribute to explaining the response variable and $s$ grows slowly than $n$, that is, $s = o(n)$.

Our proposed estimation procedure for high dimensional QRCS consists of three steps. In the first step, we use $\ell_1$ penalization with the sequential CQR method to estimate the coefficients in the selection equation of the QRCS. This step focusing on globally concerned of quantiles over a continuum of quantile indices. Globally concerned quantile regression has two strengths in high-dimensional settings: it combines all relevant information across quantiles to enhance the robustness of variable selection, and it captures global sparsity better (Zheng et al. (2015), Zheng et al. (2018)). It is important to note that using Chen (2018)'s method in this step not only avoids the nonconvex optimization problem, but also mitigates subsample selection bias because we use indicator functions as subsample selectors in the objective functions for lower quantile regression. On the other hand, we apply hard-thresholding to reduce model selection bias and select the minimal true model for use in the final step. In the second step, we estimate the coefficients in the outcome equation of the QRCS using $\ell_1$-penalized QR. Once again, we apply hard-thresholding to lessen the estimation bias and achieve the oracle rate of convergence. In the last step, by using the nonlinear least squares type estimators we discussed above, we obtain the copula parameter and all parameters of interest.

### 1.1 Literature review

There is a great deal of literature that discusses censoring and sample selection individually. Regarding CQR, since it was first introduced by Powell (1986), many following studies try to overcome the computational difficulties posed by its non-convex optimization problem. These studies aim to develop algorithms to implement Powell (1986)'s estimator, such as Buchinsky (1994), Fitzenberger (1997), Fitzenberger and Winker (2007) and Koenker (2008). Other research introduces various alternative estimators utilizing standard quantile regression on subsamples that are not affected by censoring, showing that these estimators are asymptotically equivalent to Powell (1986)'s estimator. Thus, identifying this subsample becomes crucial. Influential examples include Buchinsky and Hahn (1998), Khan and Powell (2001) and Chernozhukov and Hong (2002). Among them, Portnoy (2003) and Peng and Huang (2008) relaxed the assumptions on censoring but employed a recursive estimation scheme. In contrast, Chen (2018)'s method is less computationally intensive, especially in complex settings. On the other hand, research at the intersection of quantile methods and sample selection methods is relatively limited, with most studies concentrating on mean regression (Heckman (1974), Heckman (1979), Ahn and Powell (1993), Andrews and Schafgans (1998), Chen and Khan (2003) and Das et al. (2003)). Buchinsky and Hahn (1998), Buchinsky (2001) used the control function approach to correct sample selection bias in quantile regression, but still within a linear additive framework. Huber and Melly (2015) explored a more general, non-additive quantile model. Arellano and Bonhomme (2017) proposed widely accepted quantile selection models, where the key lies in the identification and estimation of the copula parameters. This paper adds to the literature on censored quantile selection models by offering alternative approaches for identifying and estimating the copula.

Literature on censored quantile selection models is scarce. Fernández-Val et al. (2024) developed nonseparable sample selection models that account for censoring and applied the control function approach for identifying outcome equations. Chen and Wang (2023) and Chen et al. (2024) built on the methods of Arellano and Bonhomme (2017) to estimate copula estimators, and the latter introducing a more informative selection function that goes beyond a simple binary selection.

Our paper also contributes to the framework of high-dimensional data analysis. Belloni and Chernozhukov (2011) investigated $\ell_1$-penalized quantile regression and post-$\ell_1$-penalized quantile regression to achieve the near-oracle rate of convergence. Zheng et al. (2015) and Zheng et al. (2018) employed adaptive $\ell_1$ penalties to reduce bias induced by $\ell_1$ penalties. Our paper directly uses hard-thresholding to select the minimal true model and mitigate estimation bias, obtaining the oracle rate of convergence without requiring additional steps.

### 1.2 Plan of this paper

The rest of paper is organized as follows: In section 2, we propose the censored quantile selection model and develop estimation procedures for the parameters of interest of this model. Section 3 shows the large sample properties and related theorems. Section 4 presents the Monte Carlo simulation of the finite samples performance of the proposed estimators. In Section 5, we apply the proposed method to analyze wage inequality in the United States from 2001 to 2023, using data from the American Community Survey. The proofs of theorems and part of simulation results are in the Appendix.

### 1.3 Notation

Let $\|\cdot\|$, $\|\cdot\|_1$, $\|\cdot\|_\infty$ and $\|\cdot\|_0$ denote the $\ell_2$-norm, the $\ell_1$-norm, the maximum norm and the number of nonzero components. $\|\gamma\|_{1,n} = \sum_{j=1}^{p} \hat{\sigma}_j |\gamma_j|$ denotes the $\ell_1$-norm weighted by $\hat{\sigma}_j$. For two real numbers, $a \vee b$ and $a \wedge b$ are the maximum and the minimum of $a$ and $b$, respectively. The empirical process is defined as $\mathbb{G}_n(f) = \mathbb{G}_n(f(Z_i)) \equiv n^{-1/2} \sum_{i=1}^{n} (f(Z_i) - E[f(Z_i)])$, given a random sample $Z_1, ..., Z_n$.

## 2 The estimator

### 2.1 Quantile Regression with Censored Selection (QRCS)

To better understand QRCS, we first consider the classical setting where that the number of covariates $p$ is smaller than the sample size $n$. In the next section, we will extend our model to include many controls.

Consider the following semiparametric model

$$
\begin{aligned}
&Y_1^* = X'\beta_0(U),\\
&Y_2 = \max\{Y_2^*, 0\} = \max\{X'\gamma_0(V), 0\},\\
&D = 1\{Y_2^* > 0\},\\
&Y_1 = DY_1^*,
\end{aligned}
\tag{2.1}
$$

where $Y_1^*$ and $Y_2^*$ are the latent outcome variables. $U$ and $V$ are error terms with uniform distribution on $[0,1]$, and $(U,V)$ is jointly independent of the covariates $X$. The latent outcome variables $Y_1^* = Y_1$ are observed only when $D = 1$ or $Y_2^* > 0$. The selection problem arises from the dependence between $U$ and $V$. For instance, $Y_1^*$ and $Y_2^*$ are log-wage level and hours worked. $D$ indicates labor force participation. $X$ are determinants of both levels, such as working experience, education, number of children and, etc.

Before introducing the estimation procedure, we first discuss our identification conditions for the copula parameter $\rho_0$, which measures the dependence between $U$ and $V$ and parameters $\beta_0$. Since the estimators $\hat{\gamma}$ can be obtained directly from Chen (2018)'s model, we focus solely on establishing identification for $(\beta_0, \rho_0)$.

Let $\mathcal{U} \subset (0,1)$ be a compact set of quantile indices. Similarly to the equation (5) in Arellano and Bonhomme (2017), given any pair of $u, v \in \mathcal{U}$, for observations with $X'\gamma(v) > 0$, we have

$$
\begin{aligned}
P(Y_1 < X'\beta_0(u)|Y_2^* > X'\gamma_0(v), X = x) =& P(Y_1^* < X'\beta_0(u)|X'\gamma_0(V) > X'\gamma_0(v), X = x)\\
=& P(X'\beta_0(U) < X'\beta_0(u)|X'\gamma_0(V) > X'\gamma_0(v), X = x)\\
=& P(U < u|V > v, X = x)\\
=& \frac{C(u,1,\rho_0) - C(u,v,\rho_0)}{1-v}\\
\equiv& G(u,v,\rho_0),
\end{aligned}
\tag{2.2}
$$

where $C(\cdot,\cdot,\rho)$ denotes the cumulative distribution function (c.d.f.) or copula function of $(U,V)$. Here we assume that the copula function and the function $G$ are indexed by a parameter vector $\rho_0$. $\rho_0$ measures the dependence between $U$ and $V$, which is the source of sample selection bias. For simplicity, we omit the dependence of $\rho_0$ on $x$ here. By the monotonicity of the cumulative distribution function, let

$$u = G^{-1}(\tau, v, \rho_0), \tag{2.3}$$

where $G^{-1}$ is the inverse function of $G$ with respect to the first argument. As a result, for any $\tau \in \mathcal{U}$, we have

$$G(u, v, \rho_0) = \tau,$$

which yields the following conditional moment conditions

$$E[1\{Y_1 < X'\beta_0(G^{-1}(\tau, v, \rho_0))\} - \tau | X, Y_2 > X'\gamma_0(v) > 0] = 0, \tag{2.4}$$

and unconditional moment conditions

$$E[1\{Y_2 > X'\gamma_0(v) > 0\}(1\{Y_1 < X'\beta_0(G^{-1}(\tau, v, \rho_0))\} - \tau)\phi(X)] = 0, \tag{2.5}$$

where $\phi$ is any function of $X$.

Therefore, given $\tau, v \in \mathcal{U}$, we can define $b_0(\tau, v) := \beta_0(u) = \beta_0(G^{-1}(\tau, v, \rho_0))$ as a solution to

$$E[1\{Y_2 > X'\gamma_0(v) > 0\}(1\{Y_1 < X'b\} - \tau)X] = 0, \tag{2.6}$$

which corresponds to the following minimization problem

$$\min_{b \in \mathbb{R}^p} E\rho_\tau(Y_{1i} - X_i'b)1\{Y_{2i} > X_i'\gamma_0(v) > 0\}, \tag{2.7}$$

where $\rho_\tau(u) = (\tau - 1\{u < 0\})u$ is the 'check function' in standard quantile regression. Under the full rank condition, there exists a unique solution to (2.9) and (2.15).

We have prepared all the necessary ingredients so far. We will adopt a nonlinear least square type identification strategy based on the unconditional conditions (2.9) next. Recall that in non-linear least square regression model, i.e $Y = m(X) + U$, $m(X)$ can be identified for any $x \in \mathbb{R}^n$ as $m(x) = E(Y|X = x)$. Similarly, for any $u \in \mathcal{U}$, Equation (2.2) suggests that

$$\begin{aligned}
E[1\{Y_1 < X'\beta_0(u)\}|X, Y_2 > X'\gamma_0(\tilde{v}) > 0] =& E[1\{U < G^{-1}(\tau_m, v, \rho_0)\}|X, V > \tilde{v}] \\
=& \frac{P(U < G^{-1}(\tau_m, v, \rho_0), V > \tilde{v}|X)}{P(V > \tilde{v}|X)} \\
=& \frac{C(G^{-1}(\tau_m, v, \rho_0), 1, \rho_0) - C(G^{-1}(\tau_m, v, \rho_0), \tilde{v}, \rho_0)}{1 - \tilde{v}} \\
=& G\left(G^{-1}(\tau_m, v, \rho_0), \tilde{v}, \rho_0\right)
\end{aligned}$$

with $X'\gamma(\tilde{v}) > 0$ by choosing some $\tilde{v}$ and a finite gird $\tau_1 < \tau_2 < ... < \tau_M$ on $\mathcal{U}$. To simplify our notation, define

$$G^*(\tau_m, v, \tilde{v}, \rho_0) = G\left(G^{-1}(\tau_m, v, \rho_0), \tilde{v}, \rho_0\right),$$

to be a known function of parameter $\rho_0$. Let $G_\rho^*(\tau_m, v, \tilde{v}, \cdot)$ be the derivative of $G^*$ with respect to the last argument. We make the following assumptions:

**Assumption 1.** *(i) The following equation holds*

$$G^*(\tau, v, \tilde{v}, \rho_0) = G^*(\tau, v, \tilde{v}, \rho), \quad \forall \tau, v, \tilde{v} \in (0,1)$$

*if and only if $\rho = \rho_0$.*
*(ii) The matrix*

$$\int_v \int_{\tilde{v}} \sum_{m=1}^{M} E\left[G_\rho^*(\tau_m, v, \tilde{v}, \rho_0) G_\rho^*(\tau_m, v, \tilde{v}, \rho_0) 1\{Y_{2i} > X_i' \hat{\gamma}(\tilde{v}) > 0\}\right] dv d\tilde{v}$$

*is positive definite.*

These assumptions refer to global and local identification conditions, which are quite general for copula functions specified parametrically. As a matter of fact, these are just traditional identification conditions for nonlinear least squares regression models. Under these conditions, we get $\hat{\rho}$ via

$$\min_{\rho \in \mathbb{R}} \int_{\tilde{v}} \int_v \sum_{m=1}^{M} \sum_{i=1}^{n} \Big[1\{Y_{1i} < X_i' \hat{b}(\tau_m, v)\} - G^*\left(\tau_m, v, \tilde{v}, \rho\right)\Big]^2 1\{Y_{2i} > X_i' \hat{\gamma}(\tilde{v}) > 0\} dv d\tilde{v}. \tag{2.8}$$

where we will show next the estimation procedures for $\hat{b}(\tau, v)$ and $\hat{\gamma}(v)$.

*Remark* 2.1. As opposed to Chen et al. (2024), which applies a method-of-moment criterion based on conditional moment conditions

$$E[1\{Y_2 > X'\gamma_0(v) > 0\}(1\{Y_1 < X'\beta_0(u)\} - G(u, v, \rho_0))|X] = 0, \tag{2.9}$$

and they defines $\hat{\rho}$ as an solution to

$$\min_{\rho \in \mathbb{R}} \sum_{u, v \in \mathcal{U}} \| \frac{1}{n} \sum_{i=1}^{n} 1\{Y_{2i} > X_i' \hat{\gamma}(v) > 0\}(1\{Y_1 < X' \hat{\beta}(u, \rho)\} - G(u, v, \rho))\phi(X)\|,$$

where $\phi$ is an instrumental function of $X$, our approach has the following advantages: Since $\beta_0$ no longer depends on some unknown $\rho$ at first, we do not require an iterative process. In addition, our nonlinear least square type estimators which minimize the sum of squared residuals, inherently incorporate information from every observation in the dataset. This constitutes a global optimization that ensures our estimators converge to a global minimum. The method of moments, however, is intrinsically a limited information technique. Its estimates are determined by the selected sample moments. For example, $\phi$ can be chosen as a linear combination of covariates $X$. Consequently, any information not encapsulated in the chosen moments is effectively discarded, potentially leading to estimator precision loss.

**Step 1: Estimation of $\gamma_0(v)$.** Given any $v \in \mathcal{U}$, we obtain $\hat{\gamma}(v)$ using Chen (2018)'s model. The key idea is that censoring from below affects lower tails more than right tails of the distribution. We define the grid of $\tau$-values $S_{L_n} = \{\bar{v} = v_0 > v_1 > ... > v_{L_n} = \underline{v}\}$, where $L_n$ increases with sample size. We first start with the highest value $v_0$ and denote $\hat{\gamma}(v_0)$ as the solution of the following standard quantile regression. Let $\hat{\gamma}(v_0)$ as the solution to

$$\min_{\gamma \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^{n} \rho_{v_0}(Y_{2i} - X_i'\gamma), \tag{2.10}$$

Then, we select the subsample $\{i : X_i'\hat{\gamma}(v_0) > \delta_n, i = 1, 2, ..., n\}$, where $\delta_n$ goes to zero at a slower rate than $1/L_n$, we could choose $\delta_n$ as $q_1$-th quantile of those positive $X_i'\hat{\gamma}(v_0)$. The choice of $q_1 = 5\%$ works well in practice. Because $X_i'\hat{\gamma}(v_0) > \delta_n$ implies $X_i'\gamma_0(v_1) > 0$ with large probability when sample size increases, this procedure could be written as general forms. Thus, we let $\hat{\gamma}(v_{l+1})$ as the solution to

$$\min_{\gamma \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^{n} 1\{X_i'\hat{\gamma}(v_l) > \delta_n\} \rho_{v_{l+1}}(Y_{2i} - X_i'\gamma), \tag{2.11}$$

where $\delta_n$ is $q_1$-th quantile of $X_i'\hat{\gamma}(v_l), l = 0, ..., L_n$.

**Step 2: Estimation of $b_0(\tau, v)$.** Given any $\tau \in \mathcal{U}$ and $v \in \mathcal{U}$, we proposed $\hat{b}(\tau, v)$ by solving

$$\min_{b \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(Y_{1i} - X_i'b) 1\{Y_{2i} > X_i'\hat{\gamma}(v) > 0\}, \tag{2.12}$$

**Step 3: Estimation of $\rho_0$.** According to the previous discussion, $\hat{\rho}$ can be estimated from (2.8). Finally, we obtain the effect of the variable of interest $X$ on $\tau$th quantile of conditional distribution of potential outcomes $Y_1^*$, given all other controls.

*Remark* 2.2. (The roles of copula parameters $\rho_0$) Beyond its role in identification and estimation of censored quantile selection models, the copula parameter $\rho_0$ also offers several useful applications in empirical economic research:

- **Characterizing selection mechanisms:** The sign and magnitude of $\rho_0$ reveal the direction and degree of selection. $\rho_0 > 0$ indicates positive selection, e.g., individuals with higher abilities (higher $U$) tend to work longer (higher $V$), while a negative $\rho_0$ suggests negative selection.

- **Counterfactual distribution analysis:** $\rho_0$ enables the construction of counterfactual economic scenarios. We can, for instance, examine how wage and employment distributions would change in response to different policy changes. As illustrated in Section 6 of Arellano and Bonhomme (2017), they consider a general equilibrium framework where skill prices respond to labor supply changes and use $\rho_0$ to correct the non-random selection into work, allowing them to recompute wage and employment distributions under fixed 1978 out-of-work income levels.

- **Treatment effect with selection:** This framework can be extended to models with endogeneity and self-selection into treatment, offering a flexible alternative to instrumental variable quantile regression (IVQR) methods that rely on strong rank invariance assumptions (Chernozhukov and Hansen (2005)).

### 2.2 High-dimensional CRCS

We now consider the case in which the number of covariates $p$ is large, possibly larger than the number of samples $n$. It is easy to show that the full rank conditions for low dimensional QR are not satisfied when $p \gg n$. We use high dimensional sparse models to solve this problem. Following the $\ell_1$-penalized quantile regression proposed by Belloni and Chernozhukov (2011), we define the globally concerned sparse supports for the true parameters $\gamma_0(v)$ and $b_0(\tau, v)$ as

$$\mathbb{S}_\gamma(v) = \text{support}(\gamma_0(v)) = \{j \in \{1, ..., p\} : |\gamma_{0j}(v)| > 0\},$$

and

$$\mathbb{S}_b(\tau, v) = \text{support}(b(\tau, v)) = \{j \in \{1, ..., p\} : |b_{0j}(\tau, v)| > 0\}$$

having at most $s^\gamma, s^b \leq n/\log(n(p+1))$ nonzero components respectively for all $\tau, v \in \mathcal{U}$[1].

Let $\hat{\sigma}_j^2 = \frac{1}{n}\sum_{i=1}^n x_{ij}^2$. Define $\tilde{\gamma}(v_0)$ as the solution to

$$\min_{\gamma \in \mathbb{R}^p} \frac{1}{n}\sum_{i=1}^n \rho_{v_0}(Y_{2i} - X_i'\gamma) + \frac{\lambda}{n}\|\gamma\|_{1,n}, \tag{2.13}$$

and $\tilde{\gamma}(v_{l+1})$ as the solution to

$$\min_{\gamma \in \mathbb{R}^p} \frac{1}{n}\sum_{i=1}^n 1\{X_i'\tilde{\gamma}(v_l) > \delta_n\}\rho_{v_{l+1}}(Y_{2i} - X_i'\gamma) + \frac{\lambda}{n}\|\gamma\|_{1,n}, \tag{2.14}$$

for $v_{L_n} < ... < v_1 < v_0 \in \mathcal{U}$.

Similarly, we proposed $\tilde{b}(\tau, v)$ by solving

$$\min_{b \in \mathbb{R}^p} \frac{1}{n}\sum_{i=1}^n \rho_\tau(Y_{1i} - X_i'b)1\{Y_{2i} > X_i'\tilde{\gamma}(v) > 0\} + \frac{\lambda}{n}\|b\|_{1,n}, \tag{2.15}$$

for any $\tau, v \in \mathcal{U}$.

There are two types of parameters in (2.8) with many controls: $b_0(\tau, v)$ and $\gamma_0(v)$ are high-dimensional nuisance parameters, whereas $\rho_0$ is a low-dimensional copula parameter that describes the relationship between $U$ and $V$. To reduce the bias caused by model selection, we employ hard-thresholded QR estimators in the presence of selection and censoring. Define the hard-thresholded

[1] That is, $\sup_{v \in \mathcal{U}} |\mathbb{S}_\gamma(v)| \leq s^\gamma$ and $\sup_{\tau, v \in \mathcal{U}} |\mathbb{S}_b(\tau, v)| \leq s^b$.

estimators as

$$\hat{b}_j(\tau, v) = \tilde{b}_j(\tau, v)1\{|\tilde{b}_j(\tau, v)| > \varepsilon^b\} \quad \text{and} \quad \hat{\gamma}_j(v) = \tilde{\gamma}_j(v)1\{|\tilde{\gamma}_j(v)| > \varepsilon^\gamma\},$$

for some $\varepsilon^b \geq 0$ and $\varepsilon^\gamma \geq 0$, where $j = 1, .., p$ and $\tau, v \in \mathcal{U}$. The next section will provide details on how to choose $\varepsilon^b$ and $\varepsilon^\gamma$.

*Remark* 2.3. While Belloni and Chernozhukov (2011) mention that perfect model selection through hard-thresholding requires knowledge of the unknown quantity, for example, $\min_{j\in\mathbb{S}_\gamma(v)} |\gamma_j(v)|$, which is typically unavailable in practice, our approach adopts a hard-thresholding procedure under a mild assumption. Specifically, we assume that

$$\min_{j\in\mathbb{S}_\gamma(v)} |\gamma_j(v)| \geq c_n^\gamma,$$

where $c_n^\gamma$ is a sequence of positive constants converging to zero at an appropriate rate. This condition is weaker than requiring exact knowledge of the minimal signal strength and is often reasonable in high-dimensional settings. Consequently, our method retains some benefits of hard-thresholding, such as computational tractability and bias reduction, while remaining feasible in empirical applications.

## 3 Asymptotic Results

In this section, we present the asymptotic properties of our estimators. We first demonstrate that estimators $\tilde{\gamma}(v)$ and $\tilde{b}(\tau, v)$ for all $\tau, v \in \mathcal{U}$ fall into restricted sets that is smaller than the full parameter space.

**Assumption 2.** *$\{Y_{1i}, X_i, Y_{2i}, D_i\}_{i=1}^n$ is an i.i.d sample generated from model (2.1).*

**Assumption 3.** *The copula function $C(u, v, \rho)$ is strictly increasing in $u$ and $v$ and continuously differentiable with respect to $(u, v, \rho)$.*

Assumption 2 describes the data generation process, while Assumption 3 specifies the properties of the copula function. The latter holds for several widely used copulas, such as the Gaussian, Frank, and Gumbel copulas. Under these assumptions, the function

$$G(u, v, \rho_0) = \frac{C(u, 1, \rho_0) - C(u, v, \rho_0)}{1 - v}$$

is continuously differentiable with respect to $(u, v, \rho)$.

**Assumption 4.** *Let $\epsilon(v) = Y_2^* - X'\gamma_0(v)$ and $\eta(u) = Y_1^* - X'\beta_0(u)$.*

*(i) For all $v, u \in \mathcal{U}$, the elements of $\epsilon(v)$ and $\eta(u)$ are almost surely strictly decreasing in $v$ and $u$, respectively.*

*(ii) The conditional density of* $f_{\epsilon_i(v)|X_i}(\cdot|x_i)$ *and* $g_{\eta_i(u)|X_i}(\cdot|x_i)$ *for all* $v, u \in \mathcal{U}$ *satisfies*

$$0 < \underline{f} = \inf_{\substack{v\in\mathcal{U},1\le i\le n\\ s\in\mathbb{R}}} f_{\epsilon_i(v)|X_i}(s) \le \sup_{\substack{v\in\mathcal{U},1\le i\le n\\ s\in\mathbb{R}}} f_{\epsilon_i(v)|X_i}(s) = \bar{f} < \infty \quad a.s.$$

*and*

$$0 < \underline{g} = \inf_{\substack{u\in\mathcal{U},1\le i\le n\\ s\in\mathbb{R}}} g_{\eta_i(u)|X_i}(s) \le \sup_{\substack{u\in\mathcal{U},1\le i\le n\\ s\in\mathbb{R}}} g_{\eta_i(u)|X_i}(s) = \bar{g} < \infty \quad a.s.$$

*(iii) Suppose* $\max_{1\le i\le n}\|X_i\| \le C_X$ *for a nonnegative constant* $C_X$*.*

Assumption 4 ensures the smoothness and boundedness of the underlying distributions. Elements in $\epsilon(v)$ and $\eta(u)$ being strictly decreasing in $v$ and $u$ almost surely implies that for almost all realizations of the $X$, $Y_1^*$, $Y_2^*$ are continuously distributed. $(ii)$ imposes mild smoothness conditions on the conditional density of the response variable given covariates $X$, and does not require normality or homoscedasticity assumptions. The requirement that the conditional density is bounded away from zero at the conditional quantile is standard. The uniform boundedness condition on covariates $X$ serves multiple roles: first, it ensures $1\{X_i'\hat{\gamma}(v_0) > \delta_n\}$ implies $1\{X_i'\gamma_0(v_1) > 0\}$ with probability approaching 1. Then, it enables the application of concentration inequalities and controls the empirical process terms. However, $(ii)$ and $(iii)$ can be relaxed to a slightly more general condition, as discussed in see Belloni and Chernozhukov (2011), Zheng et al. (2015) and Zheng et al. (2018).

**Assumption 5.** *The coefficients* $\gamma_0(v)$ *and* $\beta_0(u)$ *are sparse*

$$\sup_{v\in\mathcal{U}}\|\gamma(v)\|_0 \le s^\gamma \quad and \quad \sup_{u\in\mathcal{U}}\|\beta(u)\|_0 \le s^b,$$

$\gamma_0(v)$ *and* $\beta_0(u)$ *are Lipschitz in* $v$ *and* $u$ *respectively, that is, there exists a constant* $C_L = O(\sqrt{p})$ *such that*

$$\|\gamma_0(v') - \gamma_0(v'')\| \le C_L|v' - v''|, \quad \forall v', v'' \in \mathcal{U}.$$

*and*

$$\|\beta_0(u') - \beta_0(u'')\| \le C_L|u' - u''|, \quad \forall u', u'' \in \mathcal{U}.$$

Assumption 5 ensures the sparsity and smoothness of the coefficients $\gamma_0(v)$ and $\beta_0(u)$.

**Assumption 6.** *The length of the grid* $|v_{i-1} - v_i| \le \epsilon_v, i = 1, ..., L_n,$ *where* $\epsilon_v = o(\delta_n)$ *and* $\epsilon_v \le \frac{\delta_n}{3C_LC_X}$*.* $\delta_n \to 0$*,* $L_n \to \infty$*,* $L_n = o(n^{1/2})$ *and* $L_n\delta_n \to \infty$ *as* $n \to \infty$*.*

This assumption is the same as Chen (2018), it requires the cutoff sequence $\delta_n$ goes to zero slowly while grid points increase faster than $1/\delta_n$. Following lemma establishes a key property that guarantees the validity of the sequential selection mechanism in the quantile regression model with censored selections.

**Lemma 3.1.** *Under Assumption 2 - 6, if* $\hat{\gamma}(v_0)$ *is a consistent estimator of* $\gamma_0(v_0)$*,* $1\{X_i'\hat{\gamma}(v_0) > \delta_n\}$ *implies* $1\{X_i'\gamma_0(v_1) > 0\}$ *w.p.a.1.*

*Proof.* Since

$$X_i'\gamma_0(v_1) = X_i'\hat{\gamma}(v_0) + X_i'(\gamma_0(v_1) - \gamma_0(v_0)) - X_i'(\hat{\gamma}(v_0) - \gamma_0(v_0)),$$

we have

$$\begin{aligned}
&P\left(X_i'\hat{\gamma}(v_0) > \delta_n\right) \\
=&P\left(X_i'\gamma_0(v_1) + X_i'(\gamma_0(v_0) - \gamma_0(v_1)) + X_i'(\hat{\gamma}(v_0) - \gamma_0(v_0)) > \delta_n\right) \\
\leq&P\left(X_i'\gamma_0(v_1) > \tfrac{\delta_n}{3}\right) + P\left(X_i'(\gamma_0(v_0) - \gamma_0(v_1)) > \tfrac{\delta_n}{3}\right) + P\left(X_i'(\hat{\gamma}(v_0) - \gamma_0(v_0)) > \tfrac{\delta_n}{3}\right) \\
:=&(I) + (II) + (III).
\end{aligned}$$

For $(II)$:

$$\begin{aligned}
(II) &\leq P\left(|X_i'(\gamma_0(v_0) - \gamma_0(v_1))| > \tfrac{\delta_n}{3}\right) \\
&\leq P\left(\|X_i\|\|\gamma_0(v_0) - \gamma_0(v_1)\| > \tfrac{\delta_n}{3}\right) \\
&\leq P\left(C_X C_L |v_0 - v_1| > \tfrac{\delta_n}{3}\right) \\
&= P\left(|v_0 - v_1| > \tfrac{\delta_n}{3C_X C_L}\right) = 0,
\end{aligned}$$

by Cauchy-Schwarz inequality, assumptions 5 and 6.

For $(III)$:

$$\begin{aligned}
(III) &\leq P\left(\|X_i\|\|\hat{\gamma}(v_0) - \gamma_0(v_0)\| > \tfrac{\delta_n}{3}\right) \\
&\leq P\left(\|\hat{\gamma}(v_0) - \gamma_0(v_0)\| > \tfrac{\delta_n}{3C_X}\right) = 0,
\end{aligned}$$

by consistency of $\hat{\gamma}(v_0)$. □

**Assumption 7.** *Covariates are normalized such that $\sigma_j^2 = Ex_{ij}^2 = 1$, for all $j = 1, ..., p$ and $\hat{\sigma}_j^2 = \frac{1}{n}\sum_{i=1}^n x_{ij}^2$ obeys $\max_{1\leq j\leq p}|\hat{\sigma}_j - 1| \leq \mathcal{E}_X$ with probability approaching to 1, where $\mathcal{E}_X$ is an arbitrary number in $(0, 1)$.*

Assumption 7 establishes normalization conditions required for the analysis of LASSO-type estimators. Since heterogeneous variances among covariates will result in variable-dependent coefficient shrinkage: variables with larger scales will be penalized disproportionately, whereas those with smaller scales will be under-regularized. This can lead to inconsistent model selection.

We will then establish the identification of the parameters $(\gamma_0(v), b_0(\tau, v))$ uniformly over all quantile indices $v, \tau \in \mathcal{U}$, within restricted sets. These sets constrain the estimators $\hat{\gamma}(v) - \gamma_0(v)$ and $\hat{b}(\tau, v) - b_0(\tau, v)$ to cones defined below.

**Lemma 3.2** (Restricted sets)**.** *Define $\Delta_\gamma(v) := \gamma(v) - \gamma_0(v)$ for all $v \in \mathcal{U}$, and cone $\mathcal{A}_v$*

$$\mathcal{A}_v = \{\Delta_\gamma \in \mathbb{R}^P : \|\Delta_{\mathbb{S}_\gamma^c(v)}\|_1 \leq C_{\mathcal{E}}\|\Delta_{\mathbb{S}_\gamma(v)}\|_1\},$$

*where* $C_{\mathcal{E}} = \frac{\mathcal{E}_X+1}{\mathcal{E}_X(1-\mathcal{E}_X)}$. *Under Assumptions 2 - 7, if*

$$\lambda \geq C_\lambda \sqrt{n \log((p+1)n)},$$

*where* $\lambda = O(\sqrt{n \log((p+1)n)})$ *and* $C_\lambda = 2\sqrt{2}(\frac{1+\mathcal{E}_X}{1-\mathcal{E}_X})^2$, *we have (1)*

$$(1-\mathcal{E}_X)\|\Delta_\gamma\|_1 \leq \|\Delta_\gamma\|_{1,n} \leq (1+\mathcal{E}_X)\|\Delta_\gamma\|_1,$$

*for all* $\Delta_\gamma \in \mathbb{R}^p$ *and (2)*

$$\hat{\gamma}(v) \in \Gamma_0 = \{\gamma(v) \in \mathbb{R}^p : \gamma(v) - \gamma_0(v) \in \mathcal{A}_v \cap \mathcal{R}_v\},$$

*where* $\mathcal{R}_v = \{\Delta_\gamma \in \mathbb{R}^p : \|\Delta_{\mathbb{S}^c_\gamma(v)}\|_0 \leq n\}$.

*Similarly, given* $v \in \mathcal{U}$, *define* $\Delta_b(\tau) := b(\tau, v) - b_0(\tau, v)$ *for all* $\tau \in \mathcal{U}$, *and cone* $\mathcal{B}_\tau$

$$\mathcal{B}_\tau = \{\Delta_b \in \mathbb{R}^P : \|\Delta_{\mathbb{S}^c_b(\tau,v)}\|_1 \leq C_{\mathcal{E}}\|\Delta_{\mathbb{S}_b(\tau,v)}\|_1\}.$$

*Under the same conditions, the estimators* $\hat{b}(\tau, v)$ *lies in the restricted set*

$$\hat{b}(\tau, v) \in \{b(\tau, v) \in \mathbb{R}^p : b(\tau, v) - b_0(\tau, v) \in \mathcal{B}_v \cap \mathcal{R'}_v\},$$

*where* $\mathcal{R}'_\tau = \{\Delta_b \in \mathbb{R}^p : \|\Delta_{\mathbb{S}^c_b(\tau,v)}\|_0 \leq n\}$.

*Proof.* See Section A.1 in Appendix A. □

**Assumption 8.** *The restricted eigenvalue of the population design matrices satisfy*

$$\inf_{v \in \mathcal{U}} \inf_{\substack{\Delta_\gamma \neq 0 \\ \Delta_\gamma \in \mathcal{A}_v}} \frac{\Delta'_\gamma E\left(1\{X_i\gamma_0(v) > \delta_0\}X_iX'_i\right)\Delta_\gamma}{\|\Delta_\gamma\|^2} = \sigma^2_{\min} > 0$$

*and*

$$\inf_{v \in \mathcal{U}} \inf_{\substack{\Delta_b \neq 0 \\ \Delta_b \in \mathcal{B}_\tau}} \frac{\Delta'_b E\left(1\{Y_{2i} > X'_i\gamma_0(v) > 0\}X_iX'_i\right)\Delta_b}{\|\Delta_b\|^2} = \delta^2_{\min} > 0$$

*for all* $i$.

Assumption 8 is a common condition in the high dimensional data analysis. Since the minimum eigenvalue of the population design matrix inevitably tends to zero when the dimension $p$ grows with or exceeds the sample size $n$, these restricted eigenvalue conditions ensure that the covariance matrix remains positive definite over the subset of the parameter space. Under this assumption, Theorem 3.1 prove that the parameters can be identified over the restricted set. Further, Theorem 3.2 shows that the estimators attain a near optimal rate of convergence.

**Theorem 3.1** (Identification over restricted sets). *If Assumptions 2 - 8 hold, $\hat{\gamma}(v_l)$ can be identified over set $\Gamma_0$, $l = 0, ..., L_n$.*

*Proof.* See Section A.2 in Appendix A. $\square$

**Theorem 3.2** (Convergence rate). *Let*

$$C_\lambda \sqrt{n \log((p+1)n)} \le \lambda \le C'_\lambda \sqrt{n \log((p+1)n)},$$

*where $C'_\lambda \ge C_\lambda = 2\sqrt{2}(\frac{1+\mathcal{E}_X}{1-\mathcal{E}_X})^2$. If Assumption 2 - 8 hold, there exists a constant $C_{\sup} > 0$ such that the following inequality holds w.p.a.1:*

$$\sup_{v_l \in \mathcal{U}} \|\hat{\gamma}(v_l) - \gamma_0(v_l)\|^2 \le t^2 := \left( \frac{3(C_X \vee 1)^2}{\sigma^2_{\min} \underline{f}} \right)^2 (C_{\sup} + (1 + \mathcal{E}_X) C'_\lambda)^2 \frac{s^\gamma \log((p+1)n)}{n}$$

*for $l = 0, .., L_n$.*

*Proof.* See Section A.3 in Appendix A. $\square$

We now examine the model selection properties of the high-dimensional CRCS. In contrast to the existing literature, we adopt a hard thresholding procedure. This method effectively removes covariates with coefficients below a data-driven threshold.

**Theorem 3.3** (Hard-thresholding). *Let $r^\gamma = \sup_{v \in \mathcal{U}} \|\tilde{\gamma}(v) - \gamma(v)\|$. Assume $\inf_{v \in \mathcal{U}} \min_{j \in \mathbb{S}_\gamma(v)} |\gamma_j(v)| > c^\gamma$, where $r^\gamma < c^\gamma$, then*

$$\mathbb{S}_\gamma(v) := \mathrm{support}(\gamma(v)) \subseteq \tilde{\mathbb{S}}_\gamma(v) := \mathrm{support}(\tilde{\gamma}(v)) \quad \textit{for all } v \in \mathcal{U}. \tag{3.1}$$

*Moreover, the hard-thresholded estimator $\hat{\gamma}(v)$, defined for any $\varepsilon^\gamma \ge 0$ by*

$$\hat{\gamma}_j(v) = \tilde{\gamma}_j(v) 1\{|\tilde{\gamma}_j(v)| > \varepsilon^\gamma\}, \quad v \in \mathcal{U}, j = 1, \ldots, p, \tag{3.2}$$

*provided that $\varepsilon^\gamma$ is chosen such that $r^\gamma < \varepsilon^\gamma < c^\gamma - r^\gamma$, satisfies*

$$\mathrm{support}(\hat{\gamma}(v)) = \mathbb{S}_\gamma(v) \quad \textit{for all } v \in \mathcal{U}.$$

*Proof.* By assumption $\sup_{v \in \mathcal{U}} \|\tilde{\gamma}_j(v) - \gamma(v)\|_\infty \le \sup_{v \in \mathcal{U}} \|\tilde{\gamma}_j(v) - \gamma(v)\| \le r^\gamma < c^\gamma$, which immediately implies

$$\mathbb{S}_\gamma(v) := \mathrm{support}(\gamma(v)) \subseteq \tilde{\mathbb{S}}_\gamma(v) := \mathrm{support}(\tilde{\gamma}(v)) \quad \text{for all } v \in \mathcal{U}.$$

since $j \in \mathbb{S}_\gamma(v)$ implies $j \in \tilde{\mathbb{S}}_\gamma(v)$.

Consider the hard-thresholded estimator next. To establish the inclusion, by triangle inequality

$$\begin{aligned}|\tilde{\gamma}_j(v)| &\geq |\gamma_j(v)| - |\tilde{\gamma}_j(v) - \gamma_j(v)| \\ &\geq \min_{j\in\mathbb{S}_\gamma(v)} |\gamma_j(v)| - \|\tilde{\gamma}_j(v) - \gamma(v)\|_\infty > 0\end{aligned}$$

Therefore, if $j \in \mathbb{S}_\gamma(v)$, then $j \in \tilde{\mathbb{S}}_\gamma(v)$ and $\tilde{\mathbb{S}}_\gamma(v) \subseteq \mathbb{S}_\gamma(v)$. □

As demonstrated in the theorem above, both $c^\gamma$ and $r^\gamma$ converge to zero as the sample size $n$ increases. This asymptotic behavior suggest that, in practice, one could select an arbitrarily small threshold $\varepsilon^\gamma$, or using cross-validation to ensure a robust model selection performance.

The theorem and proof for $\hat{b}(\tau, v)$ are omitted for brevity, as a similar argument applies with minor changes in notation. Following theorem establishes the asymptotic properties of the final step estimator $\hat{\rho}$.

**Theorem 3.4.** *If Assumptions 1-8 hold, then $\hat{\rho}$ is consistent for $\rho_0$, and is asymptotically normal:*

$$\sqrt{n}(\hat{\rho} - \rho_0) \to^d N(0, \Sigma_\rho).$$

*Proof.* See Section A.4 in Appendix A for details. □

# 4 Simulation

This section first presents Monte Carlo simulations to evaluate the finite-sample performance of the proposed estimator $\hat{\rho}$. Our analysis begins with the classical setting where the number of covariates $p$ is fixed and smaller than the sample size $n$. The data are generated from the following model

$$\begin{aligned}Y_1^* &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \sigma(X) \cdot \Phi^{-1}(U) \\ Y_2^* &= \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \Phi^{-1}(V) \\ D &= 1\{Y_2^* > 0\} \\ Y_2 &= \max(Y_2^*, 0) \\ Y_1 &= DY_1^*\end{aligned}$$

where $X_1 \sim 0.5 \cdot N_1$, $X_2 \sim U[0,2]$ and $N_1$ are independent standard normal variables; we set $\beta_0 = \beta_1 = \beta_2 = 1, \gamma_0 = 0.2, \gamma_1 = \gamma_2 = 1$; $U$ and $V$ are uniform on $(0,1)$, with their dependence modelled through different copulas, and both are independent of $X$. $\sigma(X) = 1$ for the homoscedastic case and $\sigma(X) = 1 + 0.5 \cdot X_1$ for the heteroscedastic case. The gird of $\tau$-value to obtain $\hat{\gamma}$ is given by $\{\bar{v} = 0.99 > 0.94 > ... > 0.34 = \underline{v}\}$ (see Chen et al. (2024)) and the quantile indices to estimate $\hat{b}$ is given by $\{0.3, 0.5, 0.7, 0.9\}$. Furthermore, to assess the robustness of the estimator $\hat{\rho}$ to different dependency structures, we consider the following two copulas in our simulation design:

**Frank copula:**

$$C(u, v, \eta) = \frac{1}{\ln(1-\eta)} \ln[1 - \frac{1}{\eta}[1 - \exp(u \ln(1-\eta))][1 - \exp(v \ln(1-\eta))]],$$

where $\eta \in \mathbb{R} \backslash \{0\}$.

**Gaussian copula:**

$$C(u, v, \rho) = \Phi_\rho(\Phi(u), \Phi(v); \rho),$$

where $\Phi(\cdot)$ is CDF of standard univariate normal and $\Phi_\rho(\cdot)$ denotes the joint CDF of normal distribution with correlation $\rho \in (-1, 1)$.

Tables 1 and 2 provide the simulation results for the Frank and Gaussian copulas under the homoscedastic case, respectively. The simulation is repeated 400 times for sample sizes of $n = 300$ and 600. To ensures a fair comparison of the degree of dependence, the true copula parameters are set such that Kendall's $\tau$ takes values of $\pm 0.7$ for the Frank copula, and the correlation coefficient $\rho$ is $\pm 0.7$ for the Gaussian copula. The censoring percentages are about 17%. We report the bias, root mean square error (RMSE), and standard deviation (SD) of the estimator. As the sample size increases from 300 to 600, both the RMSE and SD decrease at a rate roughly proportional to $1/\sqrt{n}$.

Table 1: Performance for Frank copula (homoscedastic case)

| Parameter | True | Bias | RMSE | SD |
|---|---|---|---|---|
| Censoring percentages $\approx$ 17% | | | | |
| | | $n = 300$ | | |
| $\eta$ | 11.4115 | -0.6788 | 1.9436 | 1.8235 |
| Kendall's $\tau$ | 0.7000 | -0.0212 | 0.0480 | 0.0430 |
| | | $n = 600$ | | |
| $\eta$ | 11.4115 | -0.4921 | 1.3541 | 1.2631 |
| Kendall's $\tau$ | 0.7000 | -0.0138 | 0.0332 | 0.0296 |
| | | $n = 300$ | | |
| $\eta$ | -11.4115 | 0.4961 | 1.7969 | 1.7292 |
| Kendall's $\tau$ | -0.7000 | 0.0163 | 0.0447 | 0.0420 |
| | | $n = 600$ | | |
| $\eta$ | -11.4115 | 0.4222 | 1.3555 | 1.2897 |
| Kendall's $\tau$ | -0.7000 | 0.0123 | 0.0332 | 0.0302 |

The simulation results for the heteroscedastic design are shown in Tables 3 and 4. Comparing these results with those in Tables 1 and 2 reveals that the performance of our estimator remains robust. We further explore the robustness of our estimator in the Appendix C. Based on results from scenarios with varying dependence strengths (both weaker and stronger) and a more severe censoring level (the censoring percentages are about 35%), the performance remains stable.

Table 2: Performance for Gaussian copula (homoscedastic case)

| Parameter | True | Bias | RMSE | SD |
|---|---|---|---|---|
| Censoring percentages $\approx 17\%$ | | | | |
| | $n = 300$ | | | |
| $\rho$ | 0.7000 | -0.0291 | 0.0768 | 0.0710 |
| | $n = 600$ | | | |
| $\rho$ | 0.7000 | -0.0198 | 0.0510 | 0.0466 |
| | | | | |
| | $n = 300$ | | | |
| $\rho$ | -0.7000 | 0.0088 | 0.0707 | 0.0702 |
| | $n = 600$ | | | |
| $\rho$ | -0.7000 | 0.0101 | 0.0469 | 0.0462 |

Table 3: Performance for Frank copula (heteroscedastic case)

| Parameter | True | Bias | RMSE | SD |
|---|---|---|---|---|
| Censoring percentages $\approx 17\%$ | | | | |
| | $n = 300$ | | | |
| $\eta$ | 11.4115 | -0.7331 | 1.9557 | 1.8154 |
| Kendall's $\tau$ | 0.7000 | -0.0225 | 0.0480 | 0.0429 |
| | $n = 600$ | | | |
| $\eta$ | 11.4115 | -0.5150 | 1.3821 | 1.2842 |
| Kendall's $\tau$ | 0.7000 | -0.0144 | 0.0332 | 0.0303 |
| | | | | |
| | $n = 300$ | | | |
| $\eta$ | -11.4115 | 0.5627 | 1.8415 | 1.7556 |
| Kendall's $\tau$ | -0.7000 | 0.0181 | 0.0458 | 0.0427 |
| | $n = 600$ | | | |
| $\eta$ | -11.4115 | 0.4435 | 1.3767 | 1.3050 |
| Kendall's $\tau$ | -0.7000 | 0.0129 | 0.0332 | 0.0307 |

Table 4: Performance for Gaussian copula (heteroscedastic case)

| Parameter | True | Bias | RMSE | SD |
|---|---|---|---|---|
| Censoring percentages $\approx 17\%$ | | | | |
| | $n = 300$ | | | |
| $\rho$ | 0.7000 | -0.0287 | 0.0768 | 0.0712 |
| | $n = 600$ | | | |
| $\rho$ | 0.7000 | -0.0211 | 0.0520 | 0.0474 |
| | | | | |
| | $n = 300$ | | | |
| $\rho$ | -0.7000 | 0.0096 | 0.0721 | 0.0718 |
| | $n = 600$ | | | |
| $\rho$ | -0.7000 | 0.0109 | 0.0479 | 0.0463 |

We now investigate the high-dimensional setting where the number of covariates is set to $p = 500$. We consider both scenarios where the sample size $n = 1000$ and $n = 2000$. The data generating process (DGP) is specified as follows:

$$
\begin{aligned}
Y_1^* &= 1 + X_1 + X_2 + X_3 + \sigma(X) \cdot \Phi^{-1}(U) \\
Y_2^* &= 2.5 + X_1 + X_2 + X_3 + Z_1 + \Phi^{-1}(V) \\
D &= 1\{Y_2^* > 0\} \\
Y_2 &= \max(Y_2^*, 0) \\
Y_1 &= DY_1^*
\end{aligned}
$$

where the covariates $(X_1, X_2, X_3)$ and $(Z_1, ..., Z_{496})$ are independently drawn from a standard normal distribution. The gird of $\tau$-value to obtain $\hat{\gamma}$ is given by $\{\bar{v} = 0.95 > 0.9 > ... > 0.05 = \underline{v}\}$ and other settings remain unchanged from the previous design. The censoring percentages are about 20%. For the penalty level $\lambda$, we employ the criterion proposed by Belloni and Chernozhukov (2011) (BC):

$$
\lambda = 2\Lambda(1 - \alpha|X),
$$

where $\Lambda(1 - \alpha|X) := (1 - \alpha)$-quantile of $\Lambda$ conditional on X. The random variable

$$
\Lambda = n \sup_{u \in \mathcal{U}} \max_{1 \leq j \leq dim(\gamma)} |\frac{1}{n} \sum_{i=1}^{n} [\frac{x_{ij}(u - 1\{u_i \leq u\})}{\hat{\sigma_j}\sqrt{u(1-u)}}]|,
$$

where $u_1, ..., u_n$ are *i.i.d.* uniform $(0, 1)$ random variables that are independently distributed from the controls $X_1, .., X_n$. $\alpha$ is set to 0.05 and the threshold to $\varepsilon^{\gamma} = 0.001$. To implement our estimation procedure, we choose the Frank copula,and we obtain very similar results (not reported) by adopting the Gaussian copula.

Table 5: Performance for high-dimensional case with Frank copula (Kendall's $\tau = 0.7$, homoscedastic case)

| | | | $n = 1000$ | | | $n = 2000$ | | |
|---|---|---|---|---|---|---|---|---|
| $\tau$ | Parameter | True | Bias | SD | RMSE | Bias | SD | RMSE |
| Censoring percentages $\approx$ 20% | | | | | | | | |
| 0.10 | $\gamma_0$ | 1.2184 | 0.0307 | 0.0737 | 0.0798 | 0.0342 | 0.0464 | 0.0576 |
| | $\gamma_1$ | 1.0000 | -0.0972 | 0.0767 | 0.1237 | -0.0577 | 0.0572 | 0.0812 |
| | $\gamma_2$ | 1.0000 | -0.0541 | 0.0886 | 0.1037 | -0.0473 | 0.0618 | 0.0778 |
| | $\gamma_3$ | 1.0000 | -0.0569 | 0.0870 | 0.1039 | -0.0390 | 0.0627 | 0.0738 |
| | $\gamma_4$ | 1.0000 | -0.0875 | 0.0738 | 0.1144 | -0.0683 | 0.0566 | 0.0886 |
| 0.15 | $\gamma_0$ | 1.4636 | 0.0139 | 0.0668 | 0.0681 | 0.0197 | 0.0407 | 0.0451 |

Continued on next page

Table 5: Performance for high-dimensional case with Frank copula (Kendall's $\tau = 0.7$, homoscedastic case) (continued)

| | | | $n = 1000$ | | | $n = 2000$ | | |
|---|---|---|---|---|---|---|---|---|
| $\tau$ | Parameter | True | Bias | SD | RMSE | Bias | SD | RMSE |
| | $\gamma_1$ | 1.0000 | -0.0801 | 0.0686 | 0.1054 | -0.0489 | 0.0472 | 0.0679 |
| | $\gamma_2$ | 1.0000 | -0.0448 | 0.0798 | 0.0914 | -0.0376 | 0.0541 | 0.0658 |
| | $\gamma_3$ | 1.0000 | -0.0430 | 0.0769 | 0.0880 | -0.0315 | 0.0565 | 0.0646 |
| | $\gamma_4$ | 1.0000 | -0.0708 | 0.0671 | 0.0975 | -0.0577 | 0.0490 | 0.0756 |
| 0.20 | $\gamma_0$ | 1.6584 | 0.0027 | 0.0617 | 0.0617 | 0.0113 | 0.0389 | 0.0404 |
| | $\gamma_1$ | 1.0000 | -0.0667 | 0.0618 | 0.0908 | -0.0438 | 0.0449 | 0.0626 |
| | $\gamma_2$ | 1.0000 | -0.0401 | 0.0713 | 0.0817 | -0.0322 | 0.0513 | 0.0605 |
| | $\gamma_3$ | 1.0000 | -0.0359 | 0.0708 | 0.0793 | -0.0271 | 0.0519 | 0.0585 |
| | $\gamma_4$ | 1.0000 | -0.0607 | 0.0621 | 0.0868 | -0.0495 | 0.0445 | 0.0665 |
| 0.25 | $\gamma_0$ | 1.8255 | -0.0026 | 0.0557 | 0.0557 | 0.0050 | 0.0369 | 0.0372 |
| | $\gamma_1$ | 1.0000 | -0.0587 | 0.0579 | 0.0824 | -0.0409 | 0.0417 | 0.0583 |
| | $\gamma_2$ | 1.0000 | -0.0373 | 0.0623 | 0.0726 | -0.0286 | 0.0462 | 0.0543 |
| | $\gamma_3$ | 1.0000 | -0.0297 | 0.0660 | 0.0722 | -0.0236 | 0.0475 | 0.0529 |
| | $\gamma_4$ | 1.0000 | -0.0574 | 0.0592 | 0.0824 | -0.0447 | 0.0434 | 0.0623 |
| 0.30 | $\gamma_0$ | 1.9756 | -0.0055 | 0.0510 | 0.0512 | 0.0002 | 0.0353 | 0.0352 |
| | $\gamma_1$ | 1.0000 | -0.0544 | 0.0537 | 0.0764 | -0.0383 | 0.0413 | 0.0563 |
| | $\gamma_2$ | 1.0000 | -0.0322 | 0.0617 | 0.0696 | -0.0265 | 0.0430 | 0.0504 |
| | $\gamma_3$ | 1.0000 | -0.0308 | 0.0625 | 0.0696 | -0.0211 | 0.0470 | 0.0514 |
| | $\gamma_4$ | 1.0000 | -0.0544 | 0.0562 | 0.0782 | -0.0419 | 0.0411 | 0.0586 |
| 0.35 | $\gamma_0$ | 2.1147 | -0.0095 | 0.0494 | 0.0503 | -0.0025 | 0.0323 | 0.0324 |
| | $\gamma_1$ | 1.0000 | -0.0529 | 0.0539 | 0.0755 | -0.0367 | 0.0394 | 0.0538 |
| | $\gamma_2$ | 1.0000 | -0.0287 | 0.0596 | 0.0661 | -0.0242 | 0.0418 | 0.0483 |
| | $\gamma_3$ | 1.0000 | -0.0274 | 0.0601 | 0.0660 | -0.0200 | 0.0442 | 0.0485 |
| | $\gamma_4$ | 1.0000 | -0.0537 | 0.0532 | 0.0756 | -0.0402 | 0.0404 | 0.0569 |
| 0.40 | $\gamma_0$ | 2.2467 | -0.0124 | 0.0484 | 0.0500 | -0.0041 | 0.0311 | 0.0313 |
| | $\gamma_1$ | 1.0000 | -0.0522 | 0.0522 | 0.0738 | -0.0341 | 0.0374 | 0.0506 |
| | $\gamma_2$ | 1.0000 | -0.0279 | 0.0572 | 0.0636 | -0.0241 | 0.0413 | 0.0477 |
| | $\gamma_3$ | 1.0000 | -0.0257 | 0.0581 | 0.0635 | -0.0201 | 0.0415 | 0.0461 |
| | $\gamma_4$ | 1.0000 | -0.0510 | 0.0528 | 0.0734 | -0.0383 | 0.0390 | 0.0546 |
| 0.45 | $\gamma_0$ | 2.3743 | -0.0161 | 0.0467 | 0.0493 | -0.0070 | 0.0316 | 0.0323 |
| | $\gamma_1$ | 1.0000 | -0.0484 | 0.0524 | 0.0713 | -0.0346 | 0.0364 | 0.0502 |

Continued on next page

Table 5: Performance for high-dimensional case with Frank copula (Kendall's $\tau = 0.7$, homoscedastic case) (continued)

| | | | $n = 1000$ | | | $n = 2000$ | | |
|---|---|---|---|---|---|---|---|---|
| $\tau$ | Parameter | True | Bias | SD | RMSE | Bias | SD | RMSE |
| | $\gamma_2$ | 1.0000 | -0.0269 | 0.0562 | 0.0623 | -0.0213 | 0.0409 | 0.0461 |
| | $\gamma_3$ | 1.0000 | -0.0249 | 0.0575 | 0.0626 | -0.0198 | 0.0408 | 0.0453 |
| | $\gamma_4$ | 1.0000 | -0.0489 | 0.0522 | 0.0715 | -0.0369 | 0.0390 | 0.0537 |
| 0.50 | $\gamma_0$ | 2.5000 | -0.0184 | 0.0458 | 0.0493 | -0.0083 | 0.0314 | 0.0325 |
| | $\gamma_1$ | 1.0000 | -0.0481 | 0.0496 | 0.0691 | -0.0350 | 0.0353 | 0.0497 |
| | $\gamma_2$ | 1.0000 | -0.0267 | 0.0574 | 0.0632 | -0.0205 | 0.0399 | 0.0448 |
| | $\gamma_3$ | 1.0000 | -0.0234 | 0.0577 | 0.0622 | -0.0188 | 0.0408 | 0.0449 |
| | $\gamma_4$ | 1.0000 | -0.0482 | 0.0513 | 0.0704 | -0.0374 | 0.0404 | 0.0550 |
| 0.55 | $\gamma_0$ | 2.6257 | -0.0185 | 0.0454 | 0.0490 | -0.0099 | 0.0309 | 0.0324 |
| | $\gamma_1$ | 1.0000 | -0.0499 | 0.0511 | 0.0714 | -0.0330 | 0.0358 | 0.0487 |
| | $\gamma_2$ | 1.0000 | -0.0270 | 0.0572 | 0.0632 | -0.0215 | 0.0393 | 0.0447 |
| | $\gamma_3$ | 1.0000 | -0.0242 | 0.0586 | 0.0633 | -0.0190 | 0.0406 | 0.0448 |
| | $\gamma_4$ | 1.0000 | -0.0494 | 0.0524 | 0.0720 | -0.0368 | 0.0405 | 0.0547 |
| 0.60 | $\gamma_0$ | 2.7533 | -0.0194 | 0.0444 | 0.0485 | -0.0101 | 0.0303 | 0.0319 |
| | $\gamma_1$ | 1.0000 | -0.0502 | 0.0535 | 0.0733 | -0.0336 | 0.0348 | 0.0484 |
| | $\gamma_2$ | 1.0000 | -0.0268 | 0.0579 | 0.0637 | -0.0204 | 0.0389 | 0.0439 |
| | $\gamma_3$ | 1.0000 | -0.0273 | 0.0570 | 0.0632 | -0.0197 | 0.0402 | 0.0447 |
| | $\gamma_4$ | 1.0000 | -0.0486 | 0.0520 | 0.0711 | -0.0364 | 0.0393 | 0.0535 |
| 0.65 | $\gamma_0$ | 2.8853 | -0.0189 | 0.0437 | 0.0476 | -0.0071 | 0.0308 | 0.0316 |
| | $\gamma_1$ | 1.0000 | -0.0515 | 0.0539 | 0.0745 | -0.0361 | 0.0341 | 0.0497 |
| | $\gamma_2$ | 1.0000 | -0.0279 | 0.0577 | 0.0640 | -0.0219 | 0.0401 | 0.0456 |
| | $\gamma_3$ | 1.0000 | -0.0288 | 0.0582 | 0.0648 | -0.0217 | 0.0415 | 0.0467 |
| | $\gamma_4$ | 1.0000 | -0.0517 | 0.0519 | 0.0732 | -0.0372 | 0.0387 | 0.0536 |
| 0.70 | $\gamma_0$ | 3.0244 | -0.0130 | 0.0438 | 0.0456 | -0.0007 | 0.0308 | 0.0307 |
| | $\gamma_1$ | 1.0000 | -0.0537 | 0.0540 | 0.0762 | -0.0408 | 0.0355 | 0.0540 |
| | $\gamma_2$ | 1.0000 | -0.0346 | 0.0579 | 0.0674 | -0.0261 | 0.0393 | 0.0471 |
| | $\gamma_3$ | 1.0000 | -0.0310 | 0.0567 | 0.0646 | -0.0271 | 0.0400 | 0.0483 |
| | $\gamma_4$ | 1.0000 | -0.0560 | 0.0523 | 0.0766 | -0.0417 | 0.0379 | 0.0563 |
| 0.75 | $\gamma_0$ | 3.1745 | 0.0016 | 0.0439 | 0.0439 | 0.0138 | 0.0287 | 0.0318 |
| | $\gamma_1$ | 1.0000 | -0.0613 | 0.0539 | 0.0816 | -0.0506 | 0.0359 | 0.0620 |
| | $\gamma_2$ | 1.0000 | -0.0457 | 0.0600 | 0.0754 | -0.0323 | 0.0399 | 0.0513 |

Continued on next page

Table 5: Performance for high-dimensional case with Frank copula (Kendall's $\tau = 0.7$, homoscedastic case) (continued)

| $\tau$ | Parameter | True | $n = 1000$ | | | $n = 2000$ | | |
|---|---|---|---|---|---|---|---|---|
| | | | Bias | SD | RMSE | Bias | SD | RMSE |
| | $\gamma_3$ | 1.0000 | -0.0389 | 0.0587 | 0.0704 | -0.0364 | 0.0408 | 0.0547 |
| | $\gamma_4$ | 1.0000 | -0.0646 | 0.0540 | 0.0842 | -0.0515 | 0.0381 | 0.0641 |
| 0.80 | $\gamma_0$ | 3.3416 | 0.0303 | 0.0457 | 0.0548 | 0.0454 | 0.0301 | 0.0544 |
| | $\gamma_1$ | 1.0000 | -0.0780 | 0.0548 | 0.0953 | -0.0700 | 0.0379 | 0.0796 |
| | $\gamma_2$ | 1.0000 | -0.0655 | 0.0610 | 0.0894 | -0.0497 | 0.0424 | 0.0653 |
| | $\gamma_3$ | 1.0000 | -0.0599 | 0.0582 | 0.0835 | -0.0537 | 0.0423 | 0.0683 |
| | $\gamma_4$ | 1.0000 | -0.0818 | 0.0585 | 0.1005 | -0.0710 | 0.0391 | 0.0810 |
| 0.85 | $\gamma_0$ | 3.5364 | 0.0894 | 0.0514 | 0.1031 | 0.1046 | 0.0317 | 0.1093 |
| | $\gamma_1$ | 1.0000 | -0.1162 | 0.0578 | 0.1298 | -0.1067 | 0.0406 | 0.1142 |
| | $\gamma_2$ | 1.0000 | -0.0982 | 0.0668 | 0.1187 | -0.0857 | 0.0458 | 0.0971 |
| | $\gamma_3$ | 1.0000 | -0.0943 | 0.0629 | 0.1133 | -0.0851 | 0.0454 | 0.0964 |
| | $\gamma_4$ | 1.0000 | -0.1197 | 0.0627 | 0.1351 | -0.1066 | 0.0429 | 0.1149 |
| 0.90 | $\gamma_0$ | 3.7816 | 0.1834 | 0.0587 | 0.1925 | 0.2133 | 0.0377 | 0.2166 |
| | $\gamma_1$ | 1.0000 | -0.1739 | 0.0695 | 0.1872 | -0.1669 | 0.0495 | 0.1741 |
| | $\gamma_2$ | 1.0000 | -0.1545 | 0.0759 | 0.1721 | -0.1407 | 0.0553 | 0.1511 |
| | $\gamma_3$ | 1.0000 | -0.1463 | 0.0713 | 0.1627 | -0.1443 | 0.0543 | 0.1541 |
| | $\gamma_4$ | 1.0000 | -0.1826 | 0.0686 | 0.1950 | -0.1681 | 0.0484 | 0.1749 |

Given that our setup is an additive location model when $\sigma(X) = 1$, the coefficients of the outcome equation $\hat{\beta}(\tau)$ can be directly obtained from

$$\hat{\beta}(\tau) = \frac{1}{L_n} \sum_{l=1}^{L_n} \hat{b}(\tau, v_l)$$

without estimating $\hat{\rho}$. Furthermore, due to the relatively small sample sizes considered in this setting, the second step estimation does not employ high-dimensional techniques. In empirical work, we expand our analysis to larger sample sizes in order to fully accommodate a high-dimensional setting in all estimation steps. Table 5 and 6 summarize the estimation results for the first-stage and second-stage coefficients under the homoscedastic case, demonstrating the robust performance of our estimator in both steps. Table 7 presents the estimation results for both the copula parameters and the selection-corrected quantile regression coefficients. Table 8 presents the analogous set of results under the heteroscedastic case. A comparison with Table 7 reveals that the estimates for both the copula parameters and the selection-corrected quantile regression coefficients remain remarkably stable.

Table 6: Performance for high-dimensional case with Frank copula (Kendall's $\tau = 0.7$, homoscedastic case)

| $\tau$ | Parameter | True | $n = 1000$ Bias | SD | RMSE | $n = 2000$ Bias | SD | RMSE |
|---|---|---|---|---|---|---|---|---|
| Censoring percentages $\approx$ 20% | | | | | | | | |
| 0.3 | $\beta_1$ | 1.00 | -0.0645 | 0.0783 | 0.1013 | -0.0711 | 0.0551 | 0.0899 |
| | $\beta_2$ | 1.00 | -0.0727 | 0.0818 | 0.1094 | -0.0601 | 0.0589 | 0.0841 |
| | $\beta_3$ | 1.00 | -0.0985 | 0.0793 | 0.1264 | -0.1044 | 0.0564 | 0.1186 |
| 0.5 | $\beta_1$ | 1.00 | -0.0587 | 0.0756 | 0.0956 | -0.0626 | 0.0511 | 0.0808 |
| | $\beta_2$ | 1.00 | -0.0659 | 0.0766 | 0.1010 | -0.0564 | 0.0568 | 0.0800 |
| | $\beta_3$ | 1.00 | -0.0903 | 0.0689 | 0.1135 | -0.0938 | 0.0524 | 0.1074 |
| 0.7 | $\beta_1$ | 1.00 | -0.0550 | 0.0813 | 0.0981 | -0.0541 | 0.0546 | 0.0768 |
| | $\beta_2$ | 1.00 | -0.0537 | 0.0826 | 0.0984 | -0.0511 | 0.0595 | 0.0784 |
| | $\beta_3$ | 1.00 | -0.0823 | 0.0717 | 0.1091 | -0.0816 | 0.0544 | 0.0980 |
| 0.9 | $\beta_1$ | 1.00 | -0.0472 | 0.1060 | 0.1160 | -0.0447 | 0.0719 | 0.0845 |
| | $\beta_2$ | 1.00 | -0.0433 | 0.1082 | 0.1164 | -0.0420 | 0.0779 | 0.0884 |
| | $\beta_3$ | 1.00 | -0.0728 | 0.0941 | 0.1189 | -0.0675 | 0.0708 | 0.0977 |

Tables 6 summarize the estimation results for the first-stage and second-stage coefficients under the high-dimensional setting, demonstrating the robust performance of our estimator in both steps.

Table 7: Performance for high-dimensional case with Frank copula (Kendall's $\tau = 0.7$, homoscedastic case)

| $\tau$ | Parameter | True | $n = 1000$ Bias | SD | RMSE | $n = 2000$ Bias | SD | RMSE |
|---|---|---|---|---|---|---|---|---|
| Censoring percentages $\approx$ 20% | | | | | | | | |
| | Kendall's $\tau$ | 0.7000 | -0.0351 | 0.0412 | 0.0212 | -0.0200 | 0.02646 | 0.0163 |
| 0.3 | $\beta_0$ | 0.4756 | 0.1974 | 0.0767 | 0.2118 | 0.1501 | 0.0667 | 0.1642 |
| | $\beta_1$ | 1.0000 | -0.0421 | 0.0743 | 0.0853 | -0.0310 | 0.0598 | 0.0673 |
| | $\beta_2$ | 1.0000 | -0.0368 | 0.0787 | 0.0868 | -0.0278 | 0.0660 | 0.0715 |
| | $\beta_3$ | 1.0000 | -0.0455 | 0.0743 | 0.0871 | -0.0366 | 0.0596 | 0.0699 |
| 0.5 | $\beta_0$ | 1.0000 | 0.0702 | 0.0606 | 0.0927 | 0.0477 | 0.0483 | 0.0679 |
| | $\beta_1$ | 1.0000 | -0.0385 | 0.0580 | 0.0695 | -0.0286 | 0.0462 | 0.0542 |
| | $\beta_2$ | 1.0000 | -0.0287 | 0.0625 | 0.0687 | -0.0219 | 0.0491 | 0.0537 |

Continued on next page

Table 7: Performance for high-dimensional case with Frank copula (Kendall's $\tau = 0.7$, homoscedastic case) (continued)

| | | | $n = 1000$ | | | $n = 2000$ | | |
|---|---|---|---|---|---|---|---|---|
| $\tau$ | Parameter | True | Bias | SD | RMSE | Bias | SD | RMSE |
| | $\beta_3$ | 1.0000 | -0.0404 | 0.0552 | 0.0684 | -0.0355 | 0.0444 | 0.0568 |
| 0.7 | $\beta_0$ | 1.5244 | 0.0232 | 0.0534 | 0.0582 | 0.0195 | 0.0389 | 0.0435 |
| | $\beta_1$ | 1.0000 | -0.0326 | 0.0545 | 0.0635 | -0.0267 | 0.0389 | 0.0471 |
| | $\beta_2$ | 1.0000 | -0.0217 | 0.0609 | 0.0646 | -0.0166 | 0.0430 | 0.0460 |
| | $\beta_3$ | 1.0000 | -0.0412 | 0.0523 | 0.0665 | -0.0327 | 0.0386 | 0.0505 |
| 0.9 | $\beta_0$ | 2.2816 | 0.0128 | 0.0606 | 0.0618 | 0.0128 | 0.0422 | 0.0440 |
| | $\beta_1$ | 1.0000 | -0.0259 | 0.0687 | 0.0733 | -0.0244 | 0.0481 | 0.0538 |
| | $\beta_2$ | 1.0000 | -0.0253 | 0.0792 | 0.0830 | -0.0143 | 0.0524 | 0.0543 |
| | $\beta_3$ | 1.0000 | -0.0315 | 0.0663 | 0.0733 | -0.0268 | 0.0492 | 0.0559 |

Table 8: Performance for high-dimensional case with Frank copula (Kendall's $\tau = 0.7$, heteroscedastic case)

| | | | $n = 1000$ | | | $n = 2000$ | | |
|---|---|---|---|---|---|---|---|---|
| $\tau$ | Parameter | True | Bias | SD | RMSE | Bias | SD | RMSE |
| Censoring percentages $\approx 20\%$ | | | | | | | | |
| | Kendall's $\tau$ | 0.7000 | -0.0482 | 0.0529 | 0.0223 | -0.0316 | 0.0361 | 0.0173 |
| 0.3 | $\beta_0$ | 0.4756 | 0.1489 | 0.0903 | 0.1741 | 0.1020 | 0.0833 | 0.1316 |
| | $\beta_1$ | 0.7378 | 0.1323 | 0.0740 | 0.1515 | 0.1270 | 0.0615 | 0.1411 |
| | $\beta_2$ | 1.0000 | -0.0531 | 0.0896 | 0.1041 | -0.0440 | 0.0774 | 0.0889 |
| | $\beta_3$ | 1.0000 | -0.0664 | 0.0840 | 0.1070 | -0.0571 | 0.0707 | 0.0908 |
| 0.5 | $\beta_0$ | 1.0000 | 0.0410 | 0.0666 | 0.0781 | 0.0233 | 0.0545 | 0.0592 |
| | $\beta_1$ | 1.0000 | 0.0281 | 0.0600 | 0.0661 | 0.0293 | 0.0466 | 0.0549 |
| | $\beta_2$ | 1.0000 | -0.0385 | 0.0686 | 0.0786 | -0.0312 | 0.0533 | 0.0617 |
| | $\beta_3$ | 1.0000 | -0.0517 | 0.0623 | 0.0809 | -0.0458 | 0.0462 | 0.0650 |
| 0.7 | $\beta_0$ | 1.5244 | 0.0089 | 0.0553 | 0.0560 | 0.0087 | 0.0401 | 0.0409 |
| | $\beta_1$ | 1.2622 | -0.0212 | 0.0560 | 0.0598 | -0.0145 | 0.0368 | 0.0395 |
| | $\beta_2$ | 1.0000 | -0.0271 | 0.0616 | 0.0672 | -0.0202 | 0.0415 | 0.0461 |
| | $\beta_3$ | 1.0000 | -0.0459 | 0.0542 | 0.0710 | -0.0378 | 0.0383 | 0.0538 |
| 0.9 | $\beta_0$ | 2.2816 | 0.0097 | 0.0609 | 0.0616 | 0.0091 | 0.0439 | 0.0447 |

Continued on next page

Table 8: Performance for high-dimensional case with Frank copula (Kendall's $\tau = 0.7$, heteroscedastic case) (continued)

| | | | $n = 1000$ | | | $n = 2000$ | | |
|---|---|---|---|---|---|---|---|---|
| $\tau$ | Parameter | True | Bias | SD | RMSE | Bias | SD | RMSE |
| | $\beta_1$ | 1.6408 | -0.0476 | 0.0631 | 0.0790 | -0.0401 | 0.0433 | 0.0590 |
| | $\beta_2$ | 1.0000 | -0.0250 | 0.0755 | 0.0794 | -0.0150 | 0.0487 | 0.0509 |
| | $\beta_3$ | 1.0000 | -0.0336 | 0.0675 | 0.0753 | -0.0283 | 0.0474 | 0.0551 |

# 5 An application: Wage Inequality in the United States, 2001–2023

This section applies our high-dimensional quantile regression models with censored selection to analyze the evolution of wage inequality and gender wage gaps in the United States using data from the American Community Survey (ACS) from 2001 to 2023. The primary goal is to recover the latent wage distribution for both male and female, free from the distortions of sample selection, to estimate the gender wage gap at given quantile levels.

## 5.1 Data description

The analysis employs the public-use files of the ACS for the years 2001 through 2023. The sample is restricted to individuals aged between 16 and 64 who reported working in the year preceding the survey. Residents of institutional group quarters, such as prisons, are excluded, as are unpaid family workers. Self-employed workers are excluded. To ensure computational feasibility, a 10% stratified random sample is drawn, where stratification is based on state, race, and education level. The final sample consists of 2,706,867 observations across the 2001-2023 period. Annual sample sizes range from 49,948 in 2002 to 142,493 in 2023. The sample includes 1,350,145 males (49.9%) and 1,356,722 females (50.1%). Table 9 reports the detailed sample composition by year and gender.

Table 9: Descriptive statistics

| | Total | | Male | | Female | |
|---|---|---|---|---|---|---|
| Year | $n$ | % | $n$ | % | $n$ | % |
| 2001 | 55,778 | 2.1 | 28,128 | 50.4 | 27,650 | 49.6 |
| 2002 | 49,948 | 1.8 | 25,138 | 50.3 | 24,810 | 49.7 |
| 2003 | 54,752 | 2.0 | 27,567 | 50.4 | 27,185 | 49.6 |

Continued on next page

Table 9: Descriptive statistics (continued)

| | Total | | Male | | Female | |
|---|---|---|---|---|---|---|
| Year | $n$ | % | $n$ | % | $n$ | % |
| 2004 | 54,391 | 2.0 | 27,325 | 50.2 | 27,066 | 49.8 |
| 2005 | 130,392 | 4.8 | 65,579 | 50.3 | 64,813 | 49.7 |
| 2006 | 133,776 | 4.9 | 67,331 | 50.3 | 66,445 | 49.7 |
| 2007 | 134,685 | 5.0 | 67,620 | 50.2 | 67,065 | 49.8 |
| 2008 | 135,135 | 5.0 | 67,821 | 50.2 | 67,314 | 49.8 |
| 2009 | 133,429 | 4.9 | 66,615 | 49.9 | 66,814 | 50.1 |
| 2010 | 131,191 | 4.8 | 65,502 | 49.9 | 65,689 | 50.1 |
| 2011 | 129,839 | 4.8 | 65,077 | 50.1 | 64,762 | 49.9 |
| 2012 | 130,370 | 4.8 | 65,432 | 50.2 | 64,938 | 49.8 |
| 2013 | 132,769 | 4.9 | 67,099 | 50.5 | 65,670 | 49.5 |
| 2014 | 132,513 | 4.9 | 66,576 | 50.2 | 65,937 | 49.8 |
| 2015 | 133,547 | 4.9 | 67,402 | 50.5 | 66,145 | 49.5 |
| 2016 | 134,306 | 5.0 | 67,786 | 50.5 | 66,520 | 49.5 |
| 2017 | 136,491 | 5.0 | 68,847 | 50.4 | 67,644 | 49.6 |
| 2018 | 137,547 | 5.1 | 69,436 | 50.5 | 68,111 | 49.5 |
| 2019 | 137,867 | 5.1 | 70,123 | 50.9 | 67,744 | 49.1 |
| 2020 | 110,549 | 4.1 | 55,890 | 50.6 | 54,659 | 49.4 |
| 2021 | 133,627 | 4.9 | 68,026 | 50.9 | 65,601 | 49.1 |
| 2022 | 140,583 | 5.2 | 71,600 | 50.9 | 68,983 | 49.1 |
| 2023 | 142,493 | 5.3 | 72,516 | 50.9 | 69,977 | 49.1 |
| Total | 2,706,867 | 100.0 | 1,350,145 | 49.9 | 1,356,722 | 50.1 |

In our model, $Y_1$ represents the logarithm of the hourly wage, while $Y_2$ denotes the average daily work hours. The hourly wage is constructed as annual wage and salary income divided by the product of weeks worked and usual weekly hours. Top-coded yearly wages are multiplied by a factor of 1.5 and hourly wages are set not to exceed this value divided by 1,750 hours (50 weeks $\times$ 35 hours) (Autor et al. (2013)). The average daily work hours are obtained by dividing usual hours worked per week by seven. All wages are inflated to 2023 dollars using the Personal Consumption Expenditure (PCE) Index. The set of control variables $X$ includes age, age squared, age cubed, three educational attainment dummies, 50 state dummies, a marital status dummy, 8 race category dummies, 14 occupation dummies, and 18 industry dummies.

## 5.2 Estimation procedures

We implement our estimation procedure using the Frank copula to capture the dependence between $U$ and $V$, the unobserved variables in the wage and hours equations, respectively. Results using the Gaussian copula are very similar and are not reported, in particular, the estimated rank correlation is nearly identical to that obtained with the Frank copula. This suggests that our empirical findings are not sensitive to the choice of copula.

The model is estimated separately for males and females for each year from 2001 to 2023. We let the copula parameter be gender and year specific. First, $\hat{\gamma}(v)$ is obtained from the three-step estimation procedure of Chen (2018). The grid of $v$-values used to compute $\hat{\gamma}$ is $\{0.95, 0.9, \ldots, 0.05\}$. The subsample selector $\delta_n$ is set as the 1% quantile of the positive values of $X_i'\hat{\gamma}(\cdot)$ in all steps. The threshold is set to 0.0001 for both $\hat{\gamma}(v)$ and $\hat{b}(\tau, v)$. To estimate $\hat{\rho}$, we use $\tau \in \{0.3, 0.5, 0.7, 0.9\}$ and 703 grid points. Finally, the rotated quantile regression coefficients $\hat{\beta}(u)$ are estimated over the quantile indices $\{0.02, 0.04, 0.06, \ldots, 0.98\}$.

## 5.3 Results

Table 10 reports the estimated Spearman (or rank) correlation coefficients between the errors in the wage and selection equations for males and females from 2001 to 2023. The negative coefficients indicate a negative selection into employment. For males, the rank correlation ranges from -0.164 to -0.075, suggesting a persistent selection effect throughout the sample period. In contrast, the selection pattern for females is generally weaker than those for males, with coefficients ranging from -0.105 to nearly zero.

Table 10: Spearman's $\rho$ by gender and year

| Year | Male | Female |
|---|---|---|
| 2001 | -0.164 | -0.045 |
| 2002 | -0.150 | -0.060 |
| 2003 | -0.150 | -0.030 |
| 2004 | -0.135 | -0.105 |
| 2005 | -0.120 | -0.030 |
| 2006 | -0.120 | -0.045 |
| 2007 | -0.105 | 0.000 |
| 2008 | -0.105 | -0.015 |
| 2009 | -0.075 | -0.015 |
| 2010 | -0.120 | -0.015 |
| 2011 | -0.075 | -0.015 |

Continued on next page

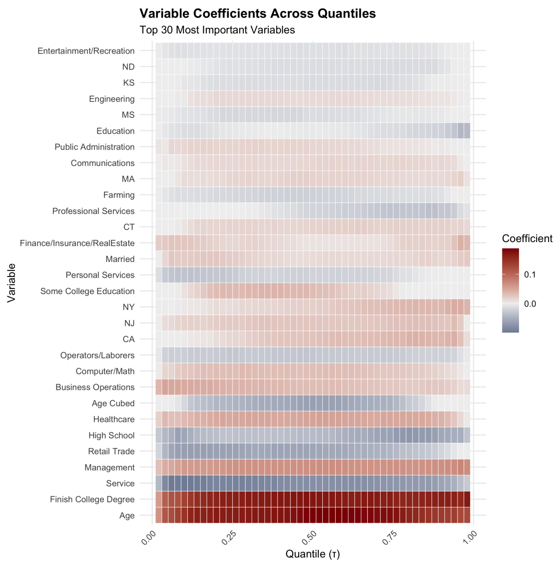
Table 10: Spearman's $\rho$ by gender and year (continued)

| Year | Male | Female |
|---|---|---|
| 2012 | -0.105 | 0.000 |
| 2013 | -0.105 | -0.015 |
| 2014 | -0.090 | 0.000 |
| 2015 | -0.075 | -0.015 |
| 2016 | -0.090 | -0.030 |
| 2017 | -0.090 | -0.030 |
| 2018 | -0.075 | -0.030 |
| 2019 | -0.090 | -0.015 |
| 2020 | -0.090 | -0.045 |
| 2021 | -0.090 | -0.075 |
| 2022 | -0.090 | -0.060 |
| 2023 | -0.075 | -0.045 |

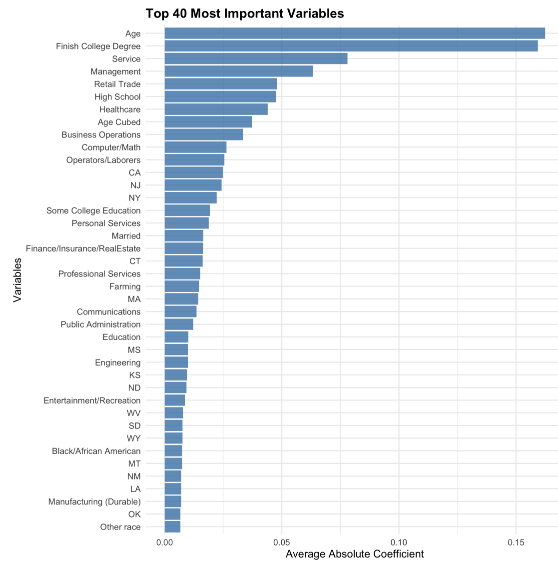
Figures 1 to 3 present the selection-corrected quantile regression coefficients for the years 2001, 2011, and 2021, separately by gender. Specifically, Figure 1 displays the results for 2001. In these figures, the heatmaps (subfigures (a) and (c)) illustrate the variation of each coefficient across quantiles for females and males, respectively, with color intensity corresponding to the magnitude of the effect. Meanwhile, the bar charts (subfigures (b) and (d)) rank the top 40 variables by their average absolute coefficients. Similar results are observed in 2011 (Figure 2) and 2021 (Figure 3).

Turning to the wage distribution, Figure 4 shows the trends in log wages from 2001 to 2023 across different quantile indices. The counterfactual distributions are constructed following the method of Machado and Mata (2005). Each panel corresponds to a specific quantile of the wage distribution. In particular, solid lines represent observed wages conditional on employment, while dashed lines show selection-corrected (simulated) wages, with blue lines indicating male wages and red lines indicating female wages. Furthermore, Figure 5 provides the path of gender wage gaps across the distribution between 2001 and 2023. It reveals that the difference between observed and selection-corrected gaps is larger at the higher quantiles.

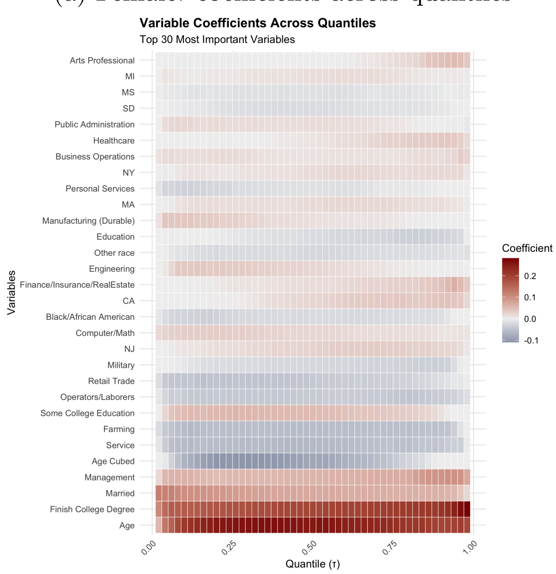
In terms of within-gender wage dispersion, Table 11 reports the 90/10 percentile ratios for male and female wages separately. The observed ratio for males rose from 1.801 in 2001 to a peak of 1.940 in 2011, then fell to 1.858 in 2023. For females, the observed ratio increased from 1.810 to a peak of 1.954 in 2014 before declining to 1.873 in 2023. The post-2020 decrease may reflect labor market disruptions during the COVID-19 pandemic. Across both genders, the corrected ratios are lower than the observed values.



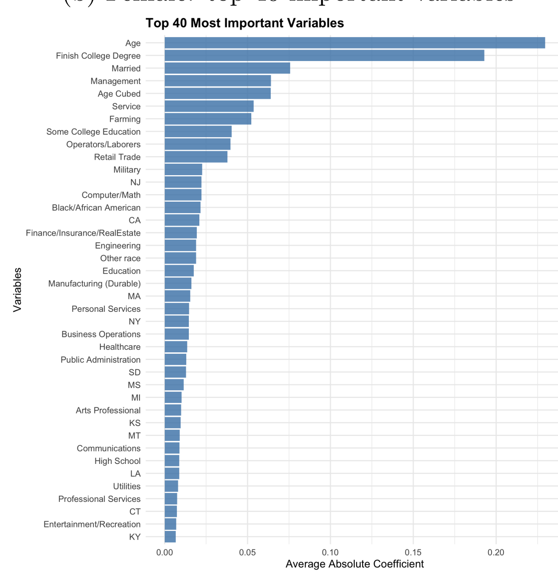

(a) Female: coefficients across quantiles




(b) Female: top 40 important variables



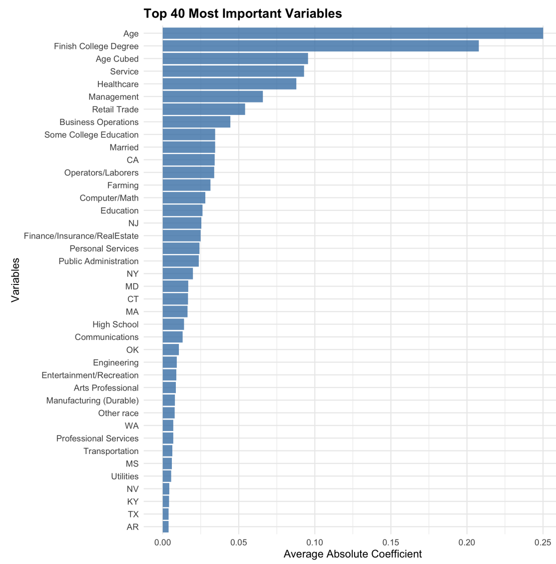

(c) Male: coefficients across quantiles



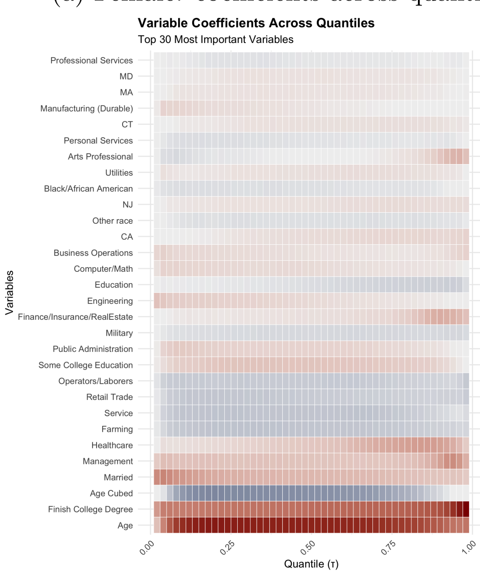

(d) Male: top 40 important variables

Figure 1: Selection-corrected coefficients for the year 2001. Subfigures (a) and (c) visualize how coefficients vary across quantiles for females and males respectively. Subfigures (b) and (d) rank the top 40 variables by their average absolute coefficient values, measuring overall influence on log wages.

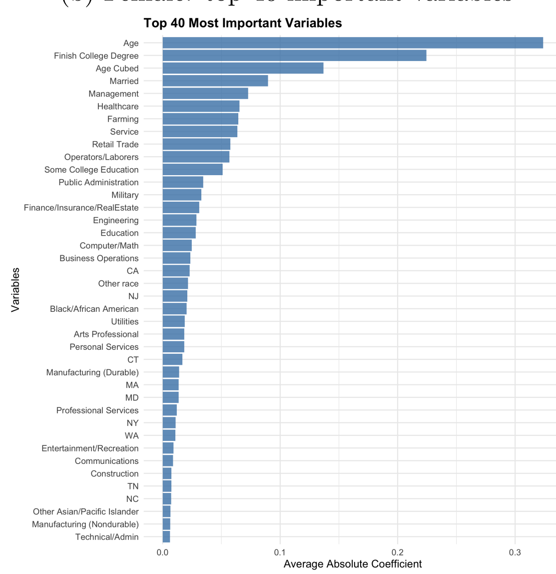

(a) Female: coefficients across quantiles




(b) Female: top 40 important variables

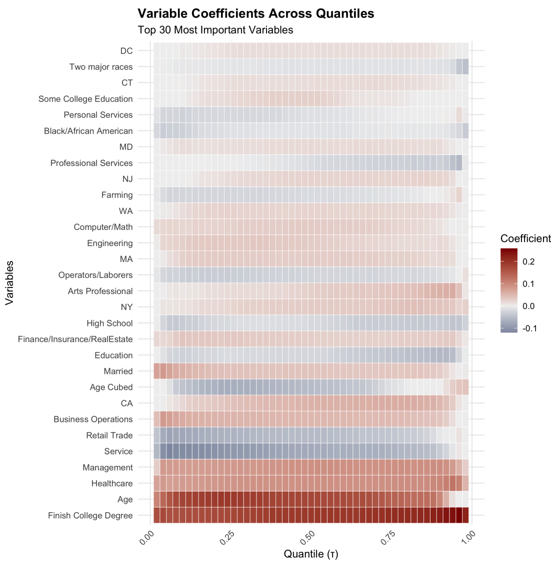



(c) Male: coefficients across quantiles

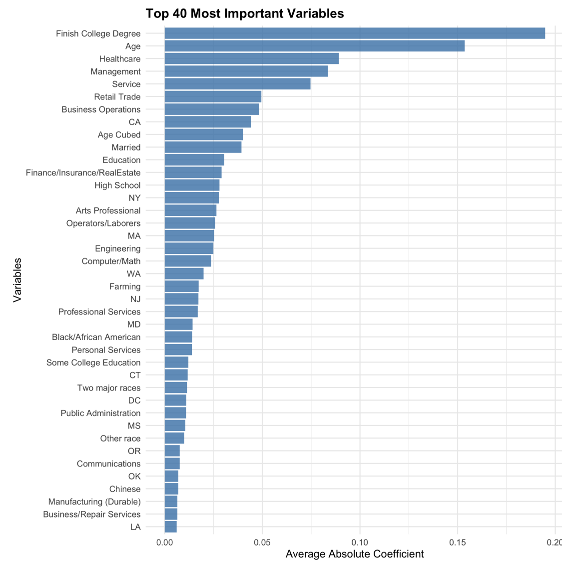



(d) Male: top 40 important variables

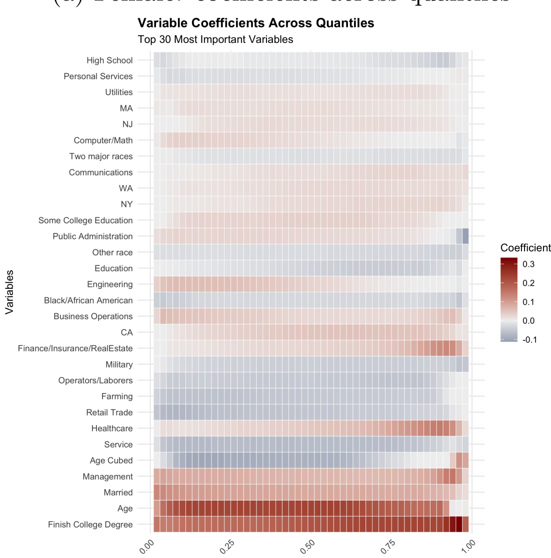
Figure 2: Selection-corrected coefficients for the year 2011. Subfigures (a) and (c) visualize how coefficients vary across quantiles for females and males respectively. Subfigures (b) and (d) rank the top 40 variables by their average absolute coefficient values, measuring overall influence on log wages.



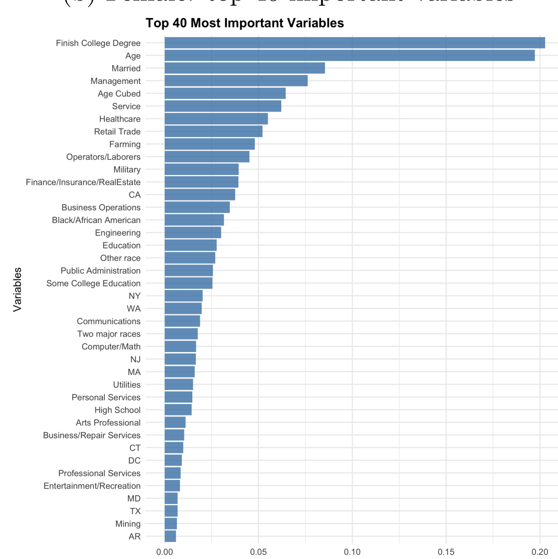

(a) Female: coefficients across quantiles




(b) Female: top 40 important variables




(c) Male: coefficients across quantiles




(d) Male: top 40 important variables

Figure 3: Selection-corrected coefficients for the year 2021. Subfigures (a) and (c) visualize how coefficients vary across quantiles for females and males respectively. Subfigures (b) and (d) rank the top 40 variables by their average absolute coefficient values, measuring overall influence on log wages.

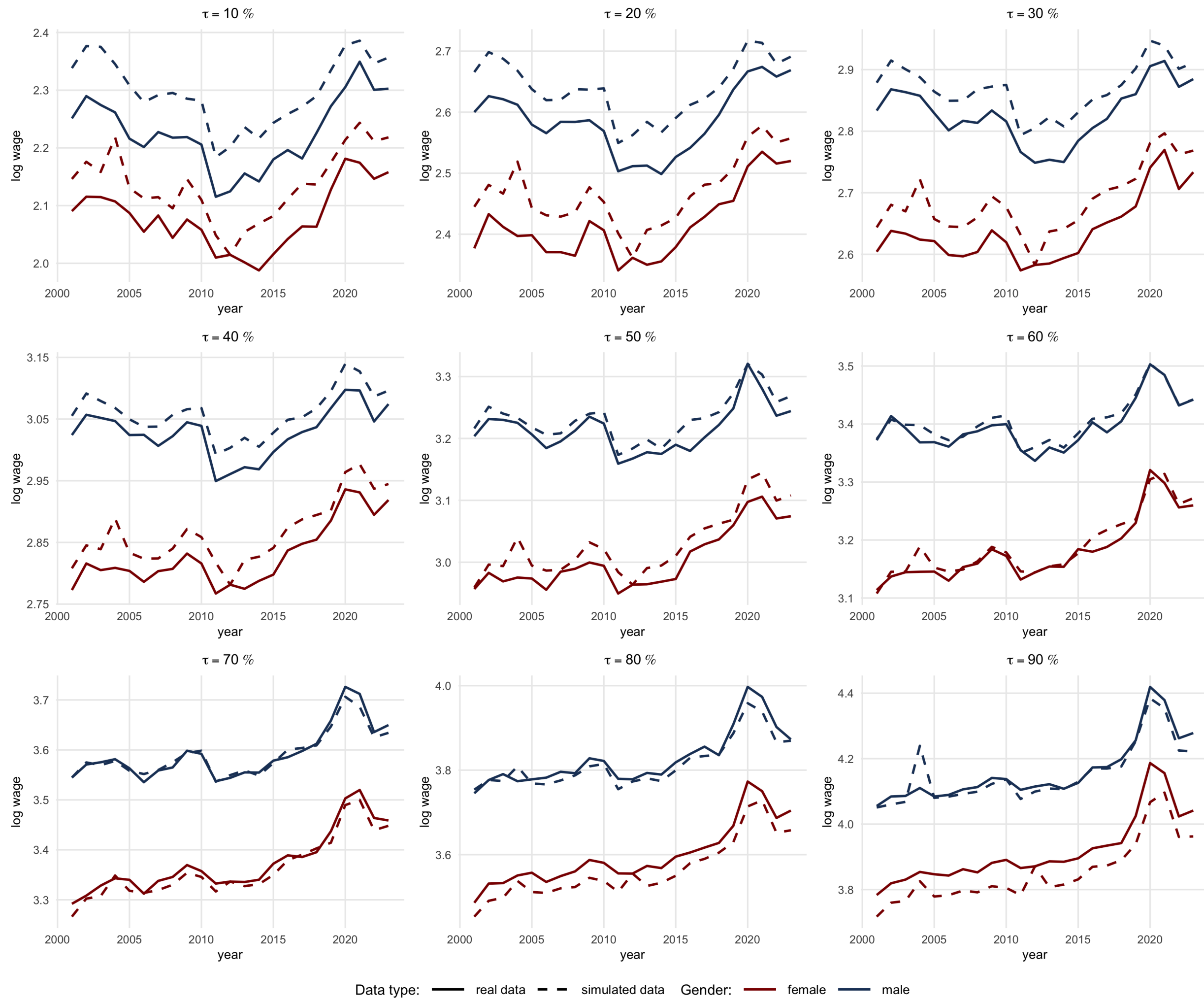



Figure 4: Wage quantiles over time

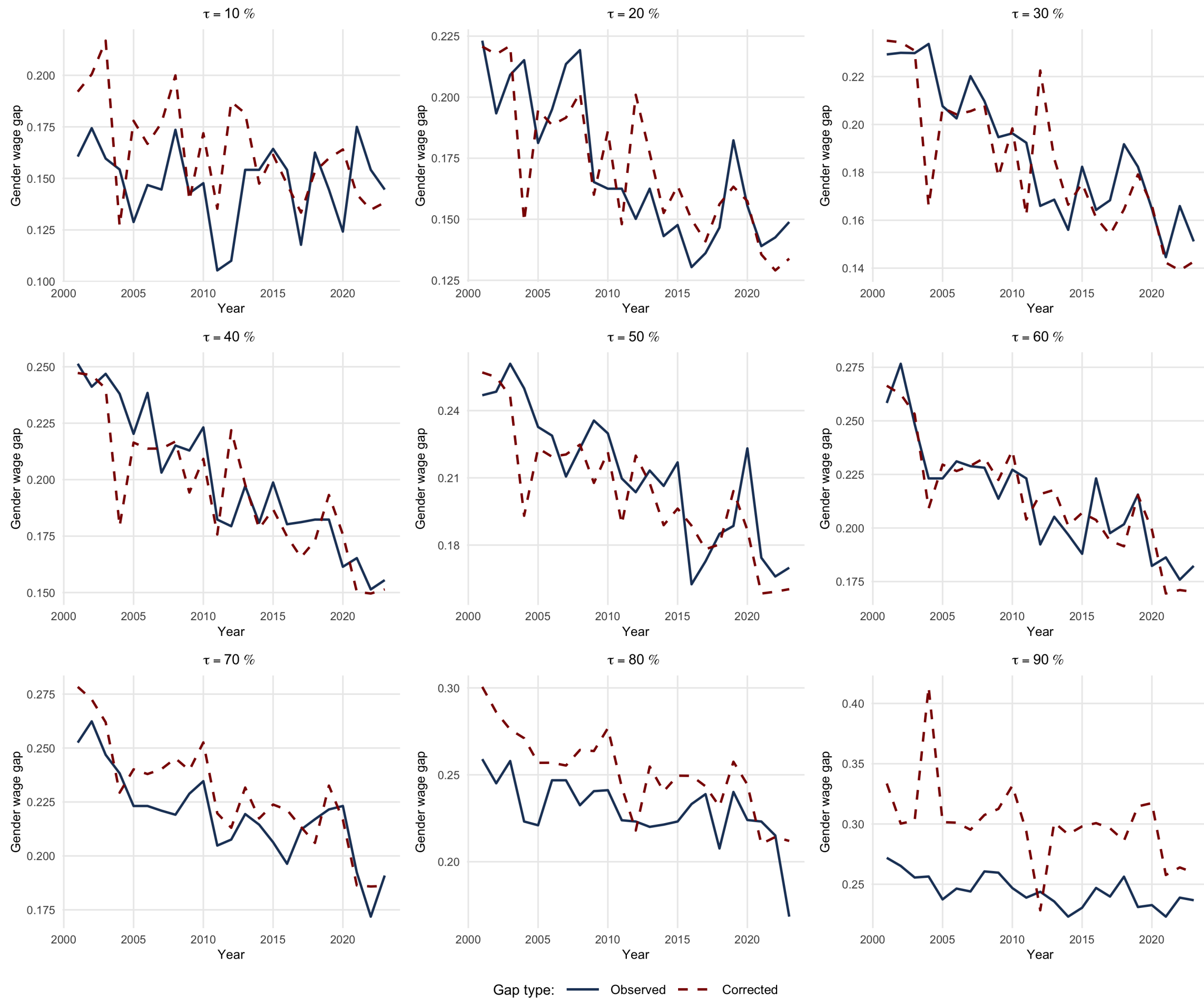



Figure 5: Gender wage gaps

Table 11: 90/10 percentile ratios

| | Male | | Female | |
|---|---|---|---|---|
| Year | Observed | Corrected | Observed | Corrected |
| 2001 | 1.801 | 1.732 | 1.810 | 1.732 |
| 2002 | 1.784 | 1.709 | 1.805 | 1.728 |
| 2003 | 1.797 | 1.713 | 1.811 | 1.744 |
| 2004 | 1.817 | 1.808 | 1.829 | 1.724 |
| 2005 | 1.843 | 1.768 | 1.843 | 1.774 |
| 2006 | 1.857 | 1.792 | 1.870 | 1.791 |
| 2007 | 1.843 | 1.786 | 1.854 | 1.796 |
| 2008 | 1.855 | 1.786 | 1.885 | 1.810 |
| 2009 | 1.866 | 1.804 | 1.870 | 1.776 |
| 2010 | 1.876 | 1.813 | 1.890 | 1.803 |
| 2011 | 1.940 | 1.867 | 1.923 | 1.847 |
| 2012 | 1.937 | 1.862 | 1.921 | 1.921 |
| 2013 | 1.912 | 1.837 | 1.942 | 1.853 |
| 2014 | 1.918 | 1.853 | 1.954 | 1.844 |
| 2015 | 1.893 | 1.841 | 1.932 | 1.840 |
| 2016 | 1.900 | 1.846 | 1.923 | 1.833 |
| 2017 | 1.913 | 1.836 | 1.906 | 1.812 |
| 2018 | 1.886 | 1.824 | 1.910 | 1.821 |
| 2019 | 1.873 | 1.822 | 1.892 | 1.811 |
| 2020 | 1.917 | 1.844 | 1.919 | 1.837 |
| 2021 | 1.864 | 1.825 | 1.911 | 1.826 |
| 2022 | 1.853 | 1.801 | 1.874 | 1.791 |
| 2023 | 1.858 | 1.792 | 1.873 | 1.786 |

# Conclusions

In this paper, we propose a semiparametric nonlinear least squares type estimator for copula parameters to correct sample selection. The estimation procedure employs $\ell_1$-penalized quantile regression combined with hard-thresholding in its first two stages to mitigate estimation bias and achieve the oracle rate of convergence, before recovering the copula parameters and coefficients of interest in the final step. We establish uniform asymptotic properties for the proposed estimators, including convergence rates and variable selection consistency.

# References

Ahn, H. and J. L. Powell (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics 58*(1-2), 3–29.

Andrews, D. W. and M. M. Schafgans (1998). Semiparametric estimation of the intercept of a sample selection model. *The Review of Economic Studies 65*(3), 497–517.

Arellano, M. and S. Bonhomme (2017). Quantile selection models with an application to understanding changes in wage inequality. *Econometrica 85*(1), 1–28.

Autor, D. H., D. Dorn, and G. H. Hanson (2013). The china syndrome: Local labor market effects of import competition in the united states. *American economic review 103*(6), 2121–2168.

Belloni, A. and V. Chernozhukov (2011). $l_1$-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics 39*(1), 82–130.

Buchinsky, M. (1994). Changes in the us wage structure 1963-1987: Application of quantile regression. *Econometrica: Journal of the Econometric Society*, 405–458.

Buchinsky, M. (2001). Quantile regression with sample selection: Estimating women's return to education in the us. *Empirical Economics 26*(1), 87–113.

Buchinsky, M. and J. Hahn (1998). An alternative estimator for the censored quantile regression model. *Econometrica*, 653–671.

Chen, S. (2018). Sequential estimation of censored quantile regression models. *Journal of Econometrics 207*(1), 30–52.

Chen, S., J. Feng, and K. Zhang (2025). Semiparatric estimation of the copula parameter in a quantile selection model. *Working Paper*.

Chen, S. and S. Khan (2003). Semiparametric estimation of a heteroskedastic sample selection model. *Econometric Theory 19*(6), 1040–1064.

Chen, S., N. Liu, H. Zhang, and Y. Zhou (2024). Estimation of wage inequality in the uk by quantile regression with censored selection. *Journal of Econometrics*, 105733.

Chen, S. and Q. Wang (2023). Quantile regression with censoring and sample selection. *Journal of Econometrics 234*(1), 205–226.

Chernozhukov, V. and C. Hansen (2005). An iv model of quantile treatment effects. *Econometrica 73*(1), 245–261.

Chernozhukov, V. and H. Hong (2002). Three-step censored quantile regression and extramarital affairs. *Journal of the American Statistical Association 97*(459), 872–882.

Das, M., W. K. Newey, and F. Vella (2003). Nonparametric estimation of sample selection models. *The Review of Economic Studies 70*(1), 33–58.

Feng, J. (2024). Nuclear norm regularized quantile regression with interactive fixed effects. *Econometric Theory 40*(6), 1391–1421.

Fernández-Val, I., A. van Vuuren, and F. Vella (2024). Nonseparable sample selection models with censored selection rules. *Journal of Econometrics 240*(2), 105088.

Fitzenberger, B. (1997). 15 a guide to censored quantile regressions. *Handbook of statistics 15*, 405–437.

Fitzenberger, B. and P. Winker (2007). Improving the computation of censored quantile regressions. *Computational Statistics & Data Analysis 52*(1), 88–108.

Heckman, J. (1974). Shadow prices, market wages, and labor supply. *Econometrica: journal of the econometric society*, 679–694.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 153–161.

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.

Khan, S. and J. L. Powell (2001). Two-step estimation of semiparametric censored regression models. *Journal of Econometrics 103*(1-2), 73–110.

Knight, K. (1998). Limiting distributions for $l_1$ regression estimators under general conditions. *Annals of Statistics*, 755–770.

Koenker, R. (2008). Censored quantile regression redux. *Journal of Statistical Software 27*, 1–25.

Ledoux, M. and M. Talagrand (2013). *Probability in Banach Spaces: Isoperimetry and Processes.* Springer Science & Business Media.

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.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics 4*, 2111–2245.

Peng, L. and Y. Huang (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association 103*(482), 637–649.

Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association 98*(464), 1001–1012.

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

Van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes.* Springer.

Zheng, Q., L. Peng, and X. He (2015). Globally adaptive quantile regression with ultra-high dimensional data. *Annals of statistics 43*(5), 2225.

Zheng, Q., L. Peng, and X. He (2018). High dimensional censored quantile regression. *Annals of statistics 46*(1), 308.

## Appendix A   Proofs of Some Lemmas and Theorems

### A.1   Proof of Lemma 3.2

*Proof.* According to Assumption 7, we have $1-\mathcal{E}_X \leq \hat{\sigma}_j \leq 1+\mathcal{E}_X$ for all $j=1,...,p$. As a result, (1) is obtained

$$\sum_{j=1}^{p}(1-\mathcal{E}_X)\left|\Delta_{\gamma_j}(v)\right| \leq \|\Delta_\gamma(v)\|_{1,n} = \sum_{j=1}^{p}\hat{\sigma}_j\left|\Delta_{\gamma_j}(v)\right| \leq \sum_{j=1}^{p}(1+\mathcal{E}_X)\left|\Delta_{\gamma_j}(v)\right|.$$

Since $\hat{\gamma}(v_0)$ has the same properties in standard $l_1$-penalized quartile regression, $\hat{\gamma}(v_0) \in \Gamma_0$, as discussed in Lemma 3 of Belloni and Chernozhukov (2011). Our discussion will focus on the restricted sets of $\hat{\gamma}(v_l), l=1,\cdots,L_n$. Without loss of generality, we choose $l=1$ and $v=v_1$. By Lemma 3.1, $V_i(v)=Y_{2i}^*-X_i'\gamma_0(v)=Y_{2i}-X_i'\gamma_0(v)$ as follows.

Let

$$\nabla\rho_v(V_i(v)) = v1\{V_i(v)\geq 0\}+(v-1)1\{V_i(v)<0\}$$

be the subgradient of $\rho_v(\cdot)$ evaluated at $V_i(v)$ for all $i=1,\cdots,n$. We will use the following lemma in this proof.

**Lemma A.1.** *Under Assumption 7, the following inequality holds w.p.a.1*

$$\sup_{v\in\mathcal{U}}\max_{1\leq j\leq p}|\langle\nabla\rho_v(V(v)),X_j\rangle| \leq 2\sqrt{2}\,(1+\mathcal{E}_X)\sqrt{n\log((p+1)n)},$$

*where $\mathcal{E}_X$ is defined in Assumption 7.*

*Proof.* The proof see Appendix B. □

By the definition of $\hat{\gamma}(v)$, we have

$$\inf_{v\in\mathcal{U}}\frac{1}{n}\sum_{i=1}^{n}\left[1\{X_i'\hat{\gamma}(v_0)>\delta_n\}\left(\rho_v\left(V_i(v)\right)-\rho_v(V_i(v)-X_i'\hat{\Delta}_\gamma(v))\right)\right]+\frac{\lambda}{n}\left[\sum_{j=1}^{p}\hat{\sigma}_j\left(|\gamma_{0j}(v)|-|\hat{\gamma}_j(v)|\right)\right]\geq 0.$$

For all $v\in\mathcal{U}$, since the objective function of quantile regression is a convex function, the first part can be bounded by

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\left[1\{X_i'\hat{\gamma}(v_0)>\delta_n\}\left(\rho_v\left(V_i(v)\right)-\rho_v(V_i(v)-X_i'\hat{\Delta}_\gamma(v))\right)\right]\\ \leq&\frac{1}{n}|\sum_{i=1}^{n}\left[1\{X_i'\hat{\gamma}(v_0)>\delta_n\}\nabla\rho_v(V_i(v))X_i'\hat{\Delta}_\gamma(v)\right]|\\ \leq&\frac{1}{n}|\sum_{i=1}^{n}\left[\nabla\rho_v(V_i(v))X_i'\hat{\Delta}_\gamma(v)\right]|\end{aligned}$$

$$
\begin{aligned}
&=\frac{1}{n}\left|\sum_{i=1}^{n}\left[\nabla\rho_v(V_i(v))X_i'\hat{\Sigma}^{-1}\hat{\Sigma}\hat{\Delta}_\gamma(v)\right]\right| \\
&\leq\frac{1}{n}\|\hat{\Sigma}\hat{\Delta}_\gamma(v)\|_1\|\sum_{i=1}^{n}\nabla\rho_v(V_i(v))X_i'\hat{\Sigma}^{-1}\|_\infty \\
&\leq\frac{1}{n}\|\hat{\Delta}_\gamma(v)\|_{1,n}\max_{1\leq j\leq p}\left|\sum_{i=1}^{n}\nabla\rho_v(V_i(v))x_{ij}\hat{\sigma}_j^{-1}\right| \\
&\leq\frac{1}{n}\|\hat{\Delta}_\gamma(v)\|_{1,n}\max_{1\leq j\leq p}|\hat{\sigma}_j^{-1}|\max_{1\leq j\leq p}\left|\sum_{i=1}^{n}\nabla\rho_v(V_i(v))x_{ij}\right| \\
&\leq\frac{1}{n}\|\hat{\Delta}_\gamma(v)\|_{1,n}\frac{1}{1-\mathcal{E}_X}2\sqrt{2}(1+\mathcal{E}_X)\sqrt{n\log((p+1)n)} \\
&=2\sqrt{2}\frac{1+\mathcal{E}_X}{1-\mathcal{E}_X}\sqrt{\frac{\log((p+1)n)}{n}}\|\hat{\Delta}_\gamma(v)\|_{1,n},
\end{aligned}
$$

where $\hat{\Sigma}=\operatorname{diag}(\hat{\sigma}_1,\cdots,\hat{\sigma}_p)$. The first inequality is by the definition of subgradient and the last inequality follows Lemma A.1. Therefore,

$$
\begin{aligned}
0 &\leq \frac{1}{n}\sum_{i=1}^{n}\left[1\{X_i'\hat{\gamma}(v_0)>\delta_n\}\left(\rho_v\left(V_i(v)\right)-\rho_v(V_i(v)-X_i'\hat{\Delta}_\gamma(v))\right)\right]+\frac{\lambda}{n}\left[\sum_{j=1}^{p}\hat{\sigma}_j\left(|\gamma_{0j}(v)|-|\hat{\gamma}_j(v)|\right)\right] \\
&\leq 2\sqrt{2}\frac{1+\mathcal{E}_X}{1-\mathcal{E}_X}\sqrt{\frac{\log((p+1)n)}{n}}\sum_{j=1}^{p}\hat{\sigma}_j\left|\hat{\gamma}_j(v)-\gamma_{0j}(v)\right|+\frac{\lambda}{n}\left[\sum_{j=1}^{p}\hat{\sigma}_j\left(|\gamma_{0j}(v)|-|\hat{\gamma}_j(v)|\right)\right] \\
&\leq \lambda\sum_{j=1}^{p}\frac{1-\mathcal{E}_X}{1+\mathcal{E}_X}\hat{\sigma}_j\left|\hat{\gamma}_j(v)-\gamma_{0j}(v)\right|+\lambda\left[\sum_{j=1}^{p}\hat{\sigma}_j\left(|\gamma_{0j}(v)|-|\hat{\gamma}_j(v)|\right)\right].
\end{aligned}
$$

After canceling $\lambda$, we obtain

$$
\frac{2\mathcal{E}_X}{1+\mathcal{E}_X}\sum_{j=1}^{p}\hat{\sigma}_j\left|\hat{\gamma}_j(v)-\gamma_{0j}(v)\right|\leq\sum_{j=1}^{p}\hat{\sigma}_j\left(|\hat{\gamma}_j(v)-\gamma_{0j}(v)|+|\gamma_{0j}(v)|-|\hat{\gamma}_j(v)|\right). \tag{A.1}
$$

Notice that, when $j\in\mathbb{S}_\gamma^c(v)=\{j\in\{1,...,p\}:|\gamma_{0j}(v)|=0\}$,

$$
\hat{\sigma}_j\left(|\hat{\gamma}_j(v)-\gamma_{0j}(v)|+|\gamma_{0j}(v)|-|\hat{\gamma}_j(v)|\right)=0,
$$

which implies

$$
\begin{aligned}
\sum_{j=1}^{p}\hat{\sigma}_j\left(|\hat{\gamma}_j(v)-\gamma_{0j}(v)|+|\gamma_{0j}(v)|-|\hat{\gamma}_j(v)|\right) &= \sum_{j\in\mathbb{S}_\gamma(v)}\hat{\sigma}_j\left(|\hat{\gamma}_j(v)-\gamma_{0j}(v)|+|\gamma_{0j}(v)|-|\hat{\gamma}_j(v)|\right) \\
&\leq 2\sum_{j\in\mathbb{S}_\gamma(v)}\hat{\sigma}_j\left|\hat{\gamma}_j(v)-\gamma_{0j}(v)\right|=2\|\hat{\gamma}_{\mathbb{S}_\gamma}(v)-\gamma_0(v)\|_{1,n}.
\end{aligned} \tag{A.2}
$$

Combine inequalities (A.1), (A.2) and part (1), we have

$$\|\hat{\gamma}_{\mathbb{S}^c_\gamma}(v)\|_{1,n} \leq \frac{1}{\mathcal{E}_X}\|\hat{\gamma}_{\mathbb{S}_\gamma}(v) - \gamma_0(v)\|_{1,n},$$

as a result

$$\|\hat{\Delta}_{\mathbb{S}^c_\gamma(v)}\|_1 \leq \frac{\mathcal{E}_X + 1}{\mathcal{E}_X(1-\mathcal{E}_X)}\|\hat{\Delta}_{\mathbb{S}_\gamma(v)}\|_1.$$

By Lemma 9 in Belloni and Chernozhukov (2011), $\|\hat{\Delta}_{\mathbb{S}^c_\gamma(v)}\|_0 \leq n$ w.p.a.1 uniformly in $v \in \mathcal{U}$.

Following a similar argument as in the proof for $\hat{\gamma}(v)$, it can be shown that $\hat{b}(\tau, v)$ also lies in the restricted set

$$\hat{b}(\tau, v) \in \{b(\tau, v) \in \mathbb{R}^p : b(\tau, v) - b_0(\tau, v) \in \mathcal{B}_v \cap \mathcal{R}'_v\},$$

under identical conditions. □

## A.2 Proof of Theorem 3.1

*Proof.* Given $\hat{\gamma}(v_0)$ can be identified over $\Gamma_0$ and $\hat{\gamma}\,(v_0)$ is a consistent estimator of $\gamma_0(v_0)$, we choose $l = 1$ and $v = v_1$ based on what has already been discussed. Since

$$\begin{aligned}
&E\left[1\{X_i'\hat{\gamma}(v_0) > \delta_n\}\left(\rho_v\left(\max\{Y_{2i}^*, 0\} - X_i'\gamma(v)\right) - \rho_v\left(\max\{Y_{2i}^*, 0\} - X_i'\gamma_0(v)\right)\right)\right] \\
=&E\left[1\{X_i'\hat{\gamma}(v_0) > \delta_n\}\left(\rho_v\left(\max\{V_i(v), -X_i'\gamma_0(v)\} - X_i'\Delta_\gamma(v)\right) - \rho_v\left(\max\{V_i(v), -X_i'\gamma_0(v)\}\right)\right)\right]
\end{aligned}$$

is nonconvex, first we prove that $1\{X_i'\hat{\gamma}(v_0) > \delta_n\}$ is lower bounded by $1\{X_i'\gamma_0(v) > \delta_0\}$, where $\delta_0 > 3\delta_n > 0$. As we know from Lemma 3.1, $X_i'\gamma_0(v_1) > \frac{\delta_n}{3}$ when $X_i\hat{\gamma}(v_0) > \delta_n$, it also implies that $X_i'\gamma_0(v_1) > \delta_0 > 3\delta_n$ when $X_i\hat{\gamma}(v_0) > \delta_n$. Thus,

$$\begin{aligned}
&1\{X_i'\hat{\gamma}(v_0) > \delta_n\} \\
=&1\{X_i'\gamma_0(v) > \delta_0, X_i'\hat{\gamma}(v_0) > \delta_n\} + 1\{X_i'\gamma_0(v) \leq \delta_0, X_i'\hat{\gamma}(v_0) > \delta_n\} \\
=&1\{X_i'\gamma_0(v) > \delta_0\} + 1\{\frac{\delta_n}{3} < X_i'\gamma_0(v) \leq \delta_0, X_i'\hat{\gamma}(v_0) > \delta_n\}.
\end{aligned}$$

Next, from Knight's identity (Knight (1998)), for any two scalars $u$ and $v$ we have

$$\rho_v(u - v) - \rho_v(u) = -v(v - 1\{u < 0\}) + \int_0^v \left(1\{u \leq s\} - 1\{u \leq 0\}\right) ds.$$

Let $u = V_i(v)$ and $v = X_i'\Delta_\gamma(v)$, using definition of $V_i(v)$ and the law of iterated expectation

$$-E\left[1\{X_i'\gamma_0(v) > \delta_0\}v(v - 1\{u < 0\})\right] = -E\left[1\{X_i'\gamma_0(v) > \delta_0\}vE\left[(v - 1\{u < 0\})|X_i\right]\right] = 0.$$

When we combine these two arguments, we get

$$E\left[1\{X_i'\hat{\gamma}(v_0) > \delta_n\}\left(\rho_v\left(\max\{V_i(v), -X_i'\gamma_0(v)\} - X_i'\Delta_\gamma(v)\right) - \rho_v\left(\max\{V_i(v), -X_i'\gamma_0(v)\}\right)\right)\right]$$

$$
\begin{aligned}
&\geq E\left[1\{X_i'\gamma_0(v)>\delta_0\}\left(\rho_v\left(\max\{V_i(v),-X_i'\gamma_0(v)\}-X_i'\Delta_\gamma(v)\right)-\rho_v\left(\max\{V_i(v),-X_i'\gamma_0(v)\}\right)\right)\right]\\
&=E\left[1\{X_i'\gamma_0(v)>\delta_0\}\left(\rho_v\left(V_i(v)-X_i'\Delta_\gamma(v)\right)-\rho_v\left(V_i(v)\right)\right)\right]\\
&=E\left[1\{X_i'\gamma_0(v)>\delta_0\}\int_0^{X_i'\Delta_\gamma(v)}\left(1\{V_i(v)\leq s\}-1\{V_i(v)\leq 0\}\right)ds\right]\\
&=E\left[1\{X_i'\gamma_0(v)>\delta_0\}E\left[\int_0^{X_i'\Delta_\gamma(v)}\left(1\{V_i(v)\leq s\}-1\{V_i(v)\leq 0\}\right)ds|X_i\right]\right]\\
&=E\left[1\{X_i'\gamma_0(v)>\delta_0\}\int_0^{X_i'\Delta_\gamma(v)}\left(F_{V_i(v)|X_i}(s)-F_{V_i(v)|X_i}(0)\right)ds\right]\\
&=E\left[1\{X_i'\gamma_0(v)>\delta_0\}\int_0^{X_i'\Delta_\gamma(v)}sf_{V_i(v)|X_i}(\tilde{s})ds\right],
\end{aligned}
$$

where the last equation is by mean value theorem where $\tilde{s}$ is the mean value. Follow Lemma S.B.6 in Feng (2024), define $0<\kappa_i=\frac{1}{(C_X\|\Delta_\gamma\|/t)\vee 1}\leq 1$ for any $t>0$, we could further lower the bound

$$
\begin{aligned}
&E\left[1\{X_i'\gamma_0(v)>\delta_0\}\int_0^{X_i'\Delta_\gamma(v)}sf_{V_i(v)|X_i}(\tilde{s})ds\right]\\
\geq&E\left[1\{X_i'\gamma_0(v)>\delta_0\}\int_0^{\kappa_iX_i'\Delta_\gamma(v)}sf_{V_i(v)|X_i}(\tilde{s})ds\right]\\
\geq&\frac{1}{2}\underline{f}E\left[1\{X_i'\gamma_0(v)>\delta_0\}\frac{\|X_i'\Delta_\gamma(v)\|^2}{\left((C_X\|\Delta_\gamma\|/t)\vee 1\right)^2}\right]\\
=&\underline{f}\Delta_\gamma(v)'\frac{E\left[1\{X_i'\gamma_0(v)>\delta_0\}X_iX_i'\right]}{2\left((C_X\|\Delta_\gamma\|/t)\vee 1\right)^2}\Delta_\gamma(v)\\
\geq&\frac{\sigma^2_{\min}\underline{f}}{2\left((C_X\|\Delta_\gamma\|/t)\vee 1\right)^2}\|\Delta_\gamma(v)\|^2\\
\geq&\frac{\sigma^2_{\min}\underline{f}}{2\left((C_X\|\Delta_\gamma\|/t)\vee 1\right)^2}\|\Delta_{\mathbb{S}_\gamma}(v)\|^2\\
\geq&\frac{\sigma^2_{\min}\underline{f}}{2s^\gamma\left((C_X\|\Delta_\gamma\|/t)\vee 1\right)^2}\|\Delta_{\mathbb{S}_\gamma}(v)\|_1^2\\
\geq&\frac{\sigma^2_{\min}\underline{f}}{2s^\gamma\left((C_X\|\Delta_\gamma\|/t)\vee 1\right)^2(1+C_{\mathcal{E}})^2}\|\Delta_\gamma(v)\|_1^2\geq 0,
\end{aligned}
$$

where the last two equations are due to $\gamma(v)\in\Gamma_0$ and $\|\Delta_{\mathbb{S}_\gamma}(v)\|_1\leq\sqrt{s^\gamma}\|\Delta_{\mathbb{S}_\gamma}(v)\|$. Finally, as $\hat{\gamma}(v_0)$ obtained from the standard $\ell_1$-penalized QR, $\hat{\gamma}(v_0)$ can be identified over $\Gamma_0$ and $\hat{\gamma}(v_0)$ is a consistent estimator of $\gamma_0(v_0)$ following Lemma 3 and Theorem 2 of Belloni and Chernozhukov (2011). □

### A.3 Proof of Theorem 3.2

*Proof.* In line with Lemma 5 in Belloni and Chernozhukov (2011), the control of empirical error is proved first. W.L.O.G, $v = v_1 \in \mathcal{U}$ hereafter. The proof of the following lemma is in Section B.2 of Appendix B.

**Lemma A.2.** *Under Assumption 2 - 8, for any $t > 0$, let $\mathcal{R}(t)$ be the emprical process*

$$\mathcal{R}(t) = \sup_{\substack{v_{l+1}\in\mathcal{U} \\ \gamma-\gamma_0\in\mathcal{A}_v \\ \|\gamma-\gamma_0\|^2\le t^2}} \left|\mathbb{G}_n\left(1\{X_i'\hat{\gamma}(v_l) > \delta_n\}\left(\rho_{v_{l+1}}\left(\max\{Y_{2i}^*, 0\} - X_i'\gamma(v_{l+1})\right) - \rho_{v_{l+1}}\left(\max\{Y_{2i}^*, 0\} - X_i'\gamma_0(v_{l+1})\right)\right)\right)\right|,$$

*for $l = 0, 1, .., L_n$. Then, there exist a constant $C_{\sup} > 0$ such that*

$$\mathcal{R}(t) \le C_{\sup}\sqrt{s^\gamma \log((p+1)n)}t$$

*with probability approaching to 1.*

Next, define the following events:
$\Omega_0$ := the event that $\hat{\gamma}(v) - \gamma_0(v) \in \mathcal{A}_v$ uniformly in $v \in \mathcal{U}$.
$\Omega_2$ := the event that the bound on the empirical error process $\mathbb{G}_n$ in Lemma A.2 holds.
Note that both events occur w.p.a.1, it is then sufficient to show that the following event $\Omega_I$:

$$\exists v \in \mathcal{U} : \|\Delta_\gamma(v)\|^2 \ge t^2$$

where $t = \frac{3(C_X\vee 1)^2}{\sigma^2_{\min}\underline{f}}(C_{\sup} + (1+\mathcal{E}_X)C_\lambda)\sqrt{\frac{s^\gamma\log((p+1)n)}{n}}$ is impossible.

By the fact that $\mathcal{A}_v$ is a cone, we can replace $\|\Delta_\gamma\| \ge t$ by $\|\Delta_\gamma\| = t$ and by the convexity of the objective function, the event $\Omega_I$ implies that for some $v \in \mathcal{U}$,

$$\begin{aligned}
0 &\ge \inf_{\substack{\Delta_\gamma\in\mathcal{A}_v \\ \|\Delta_\gamma\|^2=t^2}} \frac{1}{n}\sum_{i=1}^n 1\{X_i'\hat{\gamma}(v_0) > \delta_n\}\left(\rho_v\left(V_i(v) - X_i'\Delta_\gamma(v)\right) - \rho_v\left(V_i(v)\right)\right) + \frac{\lambda}{n}\left[\|\gamma_0(v)+\Delta_\gamma(v)\|_{1,n} - \|\gamma_0(v)\|_{1,n}\right] \\
&\ge \inf_{\substack{\Delta_\gamma\in\mathcal{A}_v \\ \|\Delta_\gamma\|^2=t^2}} \frac{1}{n}\sum_{i=1}^n 1\{X_i'\gamma_0(v) > \delta_0\}\left(\rho_v\left(V_i(v) - X_i'\Delta_\gamma(v)\right) - \rho_v\left(V_i(v)\right)\right) + \frac{\lambda}{n}\left[\|\gamma_0(v)+\Delta_\gamma(v)\|_{1,n} - \|\gamma_0(v)\|_{1,n}\right] \\
&= \inf_{\substack{\Delta_\gamma\in\mathcal{A}_v \\ \|\Delta_\gamma\|^2=t^2}} \underbrace{E\left(1\{X_i'\gamma_0(v) > \delta_0\}\left(\rho_v\left(V_i(v) - X_i'\Delta_\gamma(v)\right) - \rho_v\left(V_i(v)\right)\right)\right)}_{\text{Expectation}} \\
&\quad + \frac{1}{\sqrt{n}}\underbrace{\mathbb{G}_n\left(1\{X_i'\gamma_0(v) > \delta_0\}\left(\rho_v\left(V_i(v) - X_i'\Delta_\gamma(v)\right) - \rho_v\left(V_i(v)\right)\right)\right)}_{\text{Empirical error}} \\
&\quad + \frac{\lambda}{n}\underbrace{\left[\|\gamma_0(v)+\Delta_\gamma(v)\|_{1,n} - \|\gamma_0(v)\|_{1,n}\right]}_{\text{Penalty}}.
\end{aligned}$$

For the expectation part, Theorem 3.1 implies

$$\begin{aligned} E\left(1\{X_i'\gamma_0(v)>\delta_0\}\left(\rho_v\left(V_i(v)-X_i'\Delta_\gamma(v)\right)-\rho_v\left(V_i(v)\right)\right)\right) &\geq \frac{\sigma_{\min}^2\underline{f}}{2\left((C_X\|\Delta_\gamma\|/t)\vee 1\right)^2}\|\Delta_\gamma(v)\|^2 \\ &= \frac{\sigma_{\min}^2\underline{f}}{2\left(C_X\vee 1\right)^2}t^2. \end{aligned}$$

Next, following Lemma A.2, the empirical error process was bounded by

$$\frac{1}{\sqrt{n}}\sup_{\substack{v\in\mathcal{U}\\ \Delta_\gamma\in\mathcal{A}_v\\ \|\Delta_\gamma\|^2=t^2}}\left|\mathbb{G}_n\left(1\{X_i'\gamma_0(v)>\delta_0\}\left(\rho_v\left(V_i(v)-X_i'\Delta_\gamma(v)\right)-\rho_v\left(V_i(v)\right)\right)\right)\right|\leq C_{\sup}\sqrt{\frac{s^\gamma\log((p+1)n)}{n}}t.$$

Lastly,

$$\begin{aligned} \sup_{\substack{v\in\mathcal{U}\\ \Delta_\gamma\in\mathcal{A}_v\\ \|\Delta_\gamma\|^2=t^2}}\frac{\lambda}{n}\left|\|\gamma_0(v)+\Delta_\gamma(v)\|_{1,n}-\|\gamma_0(v)\|_{1,n}\right| &= \sup_{\substack{v\in\mathcal{U}\\ \Delta_\gamma\in\mathcal{A}_v\\ \|\Delta_\gamma\|^2=t^2}}\frac{\lambda}{n}\left|\sum_{j=1}^p\hat{\sigma}_j\left(|\gamma_j(v)|-|\hat{\gamma}_{0j}(v)|\right)\right| \\ &\leq \sup_{\substack{v\in\mathcal{U}\\ \Delta_\gamma\in\mathcal{A}_v\\ \|\Delta_\gamma\|^2=t^2}}\frac{\lambda}{n}\left|\left\|\Delta_{\mathbb{S}_\gamma(v)}\right\|_{1,n}\right| \\ &\leq \sup_{\substack{v\in\mathcal{U}\\ \Delta_\gamma\in\mathcal{A}_v\\ \|\Delta_\gamma\|^2=t^2}}\frac{\lambda}{n}\left|(1+\mathcal{E}_X)\left\|\Delta_{\mathbb{S}_\gamma(v)}\right\|_1\right| \\ &\leq \sup_{\substack{v\in\mathcal{U}\\ \Delta_\gamma\in\mathcal{A}_v\\ \|\Delta_\gamma\|^2=t^2}}\frac{\lambda}{n}(1+\mathcal{E}_X)\left\|\Delta_\gamma(v)\right\|_1 \\ &\leq \frac{\lambda}{n}(1+\mathcal{E}_X)\sqrt{s^\gamma}t \\ &= (1+\mathcal{E}_X)C_\lambda'\sqrt{\frac{s^\gamma\log\left((p+1)n\right)}{n}}t, \end{aligned}$$

where the first inequality follows Equation A.2 and the last equation is let $\lambda\leq C_\lambda'\sqrt{n\log\left((p+1)n\right)}$. Therefore, let $t=\frac{3(C_X\vee 1)^2}{\sigma_{\min}^2\underline{f}}(C_{\sup}+(1+\mathcal{E}_X)C_\lambda')\sqrt{\frac{s^\gamma\log((p+1)n)}{n}}$ such that

$$\frac{\sigma_{\min}^2\underline{f}}{2\left(C_X\vee 1\right)^2}t^2-C_{\sup}\sqrt{\frac{s^\gamma\log((p+1)n)}{n}}t-(1+\mathcal{E}_X)C_\lambda'\sqrt{\frac{s^\gamma\log\left((p+1)n\right)}{n}}t>0$$

with probability one. This yields the desired results. □

### A.4 Proofs of Theorem 3.4

*Proof.* The proof proceeds in two steps: first, we establish consistency, and then derive the asymptotic distribution.

**Consistency:** Under Assumptions 1–8, $\hat{\rho} \xrightarrow{p} \rho_0$.

**Asymptotic Normality:** The estimator $\hat{\rho}$ satisfies the first-order condition:

$$\frac{1}{n}\sum_{i=1}^{n} \psi(w_i; \hat{\rho}) = 0,$$

where $\psi(w_i; \rho)$ is the score function. Applying a Taylor expansion around $\rho_0$ yields:

$$0 = \frac{1}{n}\sum_{i=1}^{n} \psi(w_i; \rho_0) + \left[\frac{1}{n}\sum_{i=1}^{n} \nabla_\rho \psi(w_i; \bar{\rho})\right] (\hat{\rho} - \rho_0),$$

for some $\bar{\rho}$ between $\hat{\rho}$ and $\rho_0$. Rearranging terms gives:

$$\sqrt{n}(\hat{\rho} - \rho_0) = -\left[\frac{1}{n}\sum_{i=1}^{n} \nabla_\rho \psi(w_i; \bar{\rho})\right]^{-1} \cdot \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi(w_i; \rho_0).$$

By the uniform law of large numbers and the consistency of $\hat{\rho}$, the Hessian term converges in probability:

$$\frac{1}{n}\sum_{i=1}^{n} \nabla_\rho \psi(w_i; \bar{\rho}) \xrightarrow{p} H(\rho_0),$$

where $H(\rho_0)$ is nonsingular by Assumption 1. By the central limit theorem, the score term satisfies:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi(w_i; \rho_0) \xrightarrow{d} N(0, \Omega).$$

Therefore, by applying Slutsky's theorem and the Delta method, we conclude:

$$\sqrt{n}(\hat{\rho} - \rho_0) \xrightarrow{d} N\left(0, H(\rho_0)^{-1}\Omega H(\rho_0)^{-1}\right).$$

□

## Appendix B Proofs of Lemmas

### B.1 Proof of Lemma A.1

*Proof.* Define the event:

$\Omega_1 :=$ the event that $\max_{1 \le j \le p} \|X_j\|^2 \le n(\mathcal{E}_X + 1)^2$.

Under Assumption 7, $P(\Omega_1) \to 1$ and $P(\Omega_1^c) \to 0$.

Let $B = 2\sqrt{2}\left(1+\mathcal{E}_X\right)\sqrt{n\log((p+1)n)}$. By law of total probability, we have

$$P\left(\sup_{v\in\mathcal{U}}\max_{1\le j\le p}|\langle\nabla\rho_v(V(v)),X_j\rangle|>B\right)\le P\left(\sup_{v\in\mathcal{U}}\max_{1\le j\le p}|\langle\nabla\rho_v(V(v)),X_j\rangle|>B|\Omega_1\right)P\left(\Omega_1\right)+P\left(\Omega_1^c\right).$$

Thus, it is sufficient to show $P\left(\sup_{v\in\mathcal{U}}\max_{1\le j\le p}|\langle\nabla\rho_v(V(v)),X_j\rangle|>B|\Omega_1\right)$ converges to 0.

Let us define a $\epsilon$-grid of the interval $[0,1]$, i.e., $v^1<v^2<...<v^M\in\mathcal{U}_M$. Let $\epsilon=\frac{1}{\sqrt{n}}$ and $M\epsilon\le 1$. Following triangle inequality, we obtain

$$\begin{aligned}
&\sup_{\substack{v\in\mathcal{U}\\1\le j\le p}}|\langle\nabla\rho_v(V(v)),X_j\rangle|\\
\le&\max_{\substack{v^m\in\mathcal{U}_M\\1\le j\le p}}|\langle\nabla\rho_{v^m}(V(v^m)),X_j\rangle|+\sup_{\substack{|v-v^m|\le\epsilon,v^m\in\mathcal{U}_M\\1\le j\le p}}|\langle\nabla\rho_v(V(v))-\nabla\rho_{v^m}(V(v^m)),X_j\rangle|\\
\equiv&\mathcal{A}_1+\mathcal{A}_2.
\end{aligned}$$

Then,

$$P\left(\sup_{\substack{v\in\mathcal{U}\\1\le j\le p}}|\langle\nabla\rho_v(V(v)),X_j\rangle|>B|\Omega_1\right)\le 2\max\left\{P\left(\mathcal{A}_1>\frac{B}{2}|\Omega_1\right)+P\left(\mathcal{A}_2>\frac{B}{2}|\Omega_1\right)\right\}$$

by the union bound.

**Bound on $\mathcal{A}_1$**:

$$\begin{aligned}
P\left(\mathcal{A}_1>\frac{B}{2}|\Omega_1\right)&=P\left(\bigcup_{v^m\in\mathcal{U}_M}\bigcup_{1\le j\le p}|\langle\nabla\rho_{v^m}(V(v^m)),X_j\rangle|>\frac{B}{2}|\Omega_1\right)\\
&\le\sum_{v^m\in\mathcal{U}_M}\sum_{j=1}^{p}P\left(|\langle\nabla\rho_{v^m}(V(v^m)),X_j\rangle|>\frac{B}{2}|\Omega_1\right)\\
&\le\sum_{v^m\in\mathcal{U}_M}\sum_{j=1}^{p}\max_{\substack{v^m\in\mathcal{U}_M\\1\le j\le p}}P\left(|\langle\nabla\rho_{v^m}(V(v^m)),X_j\rangle|>\frac{B}{2}|\Omega_1\right)\\
&\le Mp\max_{\substack{v^m\in\mathcal{U}_M\\1\le j\le p}}P\left(|\langle\nabla\rho_{v^m}(V(v^m)),X_j\rangle|>\frac{B}{2}|\Omega_1\right)\\
&\le\frac{p}{\epsilon}\max_{\substack{v^m\in\mathcal{U}_M\\1\le j\le p}}P\left(|\langle\nabla\rho_{v^m}(V(v^m)),X_j\rangle|>\frac{B}{2}|\Omega_1\right)\\
&=\frac{p}{\epsilon}\max_{\substack{v^m\in\mathcal{U}_M\\1\le j\le p}}E\left[P\left(|\langle\nabla\rho_{v^m}(V(v^m)),X_j\rangle|>\frac{B}{2}|X,\Omega_1\right)|\Omega_1\right],
\end{aligned}$$

where the second inequality is by the union bound and the following inequality is due to $M\le\frac{1}{\epsilon}$. The last equation is by the law of iterative expectation. Since $\nabla\rho_v(V(v))$ are independent with zero mean conditional on $X$ and are uniformly bounded within $[-1,1]$, following Hoeffding's inequality, the previous equation can be further bounded by

$$\frac{2p}{\epsilon}\max_{1\leq j\leq p}\exp\left(-\frac{\frac{1}{4}B^2}{2\|X_j\|^2}\right)\leq 2p\sqrt{n}\exp\left(-\frac{\frac{1}{4}B^2}{2n\left(\mathcal{E}_X+1\right)^2}\right)=\frac{2p}{(p+1)\sqrt{n}}\to 0.$$

**Bound on $\mathcal{A}_2$**: Since

$$\begin{aligned}
&\left(\nabla\rho_v(V_i(v))-\nabla\rho_{v^m}(V_i(v^m))\right)_i\\
=&v1\left\{V_i(v)\geq 0\right\}+(v-1)1\left\{V_i(v)<0\right\}-v^m1\left\{V_i(v^m)\geq 0\right\}-(v^m-1)1\left\{V_i(v^m)<0\right\}\\
=&\underbrace{v-v^m}_{(I)_i}+\underbrace{1\left\{V_i(v^m)<0\right\}-1\left\{V_i(v)<0\right\}}_{(II)_i},
\end{aligned}$$

we can bound $P\left(\mathcal{A}_2>\frac{B}{2}|\Omega_1\right)$ into two parts:

$$P\left(\mathcal{A}_2>\frac{B}{2}|\Omega_1\right)\leq 2\max\left\{P\left(\sup_{\substack{|v-v^m|\leq\epsilon,v^m\in\mathcal{U}_M\\1\leq j\leq p}}|\langle I,X_j\rangle|>\frac{B}{4}|\Omega_1\right),P\left(\sup_{\substack{|v-v^m|\leq\epsilon,v^m\in\mathcal{U}_M\\1\leq j\leq p}}|\langle II,X_j\rangle|>\frac{B}{4}|\Omega_1\right)\right\}$$

By Cauthy-Schwarz inequality,

$$\begin{aligned}
P\left(\sup_{\substack{|v-v^m|\leq\epsilon,v^m\in\mathcal{U}_M\\1\leq j\leq p}}|\langle I,X_j\rangle|>\frac{B}{4}|\Omega_1\right)&\leq P\left(\sup_{\substack{|v-v^m|\leq\epsilon,v^m\in\mathcal{U}_M\\1\leq j\leq p}}\|I\|\cdot\|X_j\|>\frac{B}{4}|\Omega_1\right)\\
&\leq P\left(\sqrt{C_Xn}>\frac{B}{4}|\Omega_1\right)\\
&\leq P\left(\sqrt{C_Xn}>\frac{\sqrt{2}}{2}\left(1+\mathcal{E}_X\right)\sqrt{n\log((p+1)n)}|\Omega_1\right)=0
\end{aligned}$$

for large enough $n$. Furthermore, we bound

$$\begin{aligned}
&P\left(\sup_{\substack{|v-v^m|\leq\epsilon,v^m\in\mathcal{U}_M\\1\leq j\leq p}}|\langle II,X_j\rangle|>\frac{B}{4}|\Omega_1\right)\\
\leq&P\left(\sup_{\substack{v^m-\epsilon\leq v\leq v^m\\v^m\in\mathcal{U}_M\\1\leq j\leq p}}|\langle II,X_j\rangle|>\frac{B}{8}|\Omega_1\right)\\
&+P\left(\sup_{\substack{v^m\leq v\leq v^m+\epsilon\\v^m\in\mathcal{U}_M\\1\leq j\leq p}}|\langle II,X_j\rangle|>\frac{B}{8}|\Omega_1\right)
\end{aligned}$$

$$
\begin{aligned}
\leq & P\left(\max_{\substack{v^m\in\mathcal{U}_M\\1\leq j\leq p}}\left|\langle 1\left\{V(v^m)<0\right\}-1\left\{V(v^m-\epsilon)<0\right\},|X_j|\rangle\right|>\frac{B}{8}|\Omega_1\right)\\
&+P\left(\max_{\substack{v^m\in\mathcal{U}_M\\1\leq j\leq p}}\left|\langle 1\left\{V(v^m)<0\right\}-1\left\{V(v^m+\epsilon)<0\right\},|X_j|\rangle\right|>\frac{B}{8}|\Omega_1\right)\\
\leq & P\left(\max_{\substack{v^m\in\mathcal{U}_M\\1\leq j\leq p}}\left|\langle 1\left\{V(v^m)<0\right\}-1\left\{V(v^m-\epsilon)<0\right\}-\epsilon e_n,|X_j|\rangle\right|>\frac{B}{8}-\sqrt{C_Xn}|\Omega_1\right)\\
&+P\left(\max_{\substack{v^m\in\mathcal{U}_M\\1\leq j\leq p}}\left|\langle 1\left\{V(v^m)<0\right\}-1\left\{V(v^m+\epsilon)<0\right\}-\epsilon e_n,|X_j|\rangle\right|>\frac{B}{8}-\sqrt{C_Xn}|\Omega_1\right)\\
\leq & 4p\sqrt{n}\exp\left(-\frac{(B/8-\sqrt{C_Xn})^2}{2C_Xn}\right)\to 0,
\end{aligned}
$$

where $e_n=(1,...,1)'$. For details, see Lemma S.B.1 in Feng (2024). Consequently, we achieve the desired results. □

### B.2 Proof of Lemma A.2

*Proof.* Let $M=C_{\sup}\sqrt{s^\gamma\log((p+1)n)}t$. For any fixed $\Delta_\gamma$ with $\|\Delta_\gamma\|^2\leq t^2$, we have

$$
\begin{aligned}
&\mathrm{Var}\left(\mathbb{G}_n\left(1\{X_i'\hat{\gamma}(v_0)>\delta_n\}\left(\rho_v\left(\max\{Y_{2i}^*,0\}-X_i'\gamma(v)\right)-\rho_v\left(\max\{Y_{2i}^*,0\}-X_i'\gamma_0(v)\right)\right)\right)\right)\\
=&\mathrm{Var}\left(1\{X_i'\hat{\gamma}(v_0)>\delta_n\}\left(\rho_v\left(\max\{Y_{2i}^*,0\}-X_i'\gamma(v)\right)-\rho_v\left(\max\{Y_{2i}^*,0\}-X_i'\gamma_0(v)\right)\right)\right)\\
=&E\left(1\{X_i'\hat{\gamma}(v_0)>\delta_n\}\left(\rho_v\left(\max\{Y_{2i}^*,0\}-X_i'\gamma(v)\right)-\rho_v\left(\max\{Y_{2i}^*,0\}-X_i'\gamma_0(v)\right)\right)\right)^2\\
\leq&E\left(1\{X_i'\gamma_0(v)>\frac{\delta_n}{3}\}\left(\rho_v\left(Y_{2i}^*-X_i'\gamma(v)\right)-\rho_v\left(Y_{2i}^*-X_i'\gamma_0(v)\right)\right)\right)^2\\
=&E\left(1\{X_i'\gamma_0(v)>\frac{\delta_n}{3}\}\left(\rho_v\left(V_i(v)-X_i'\Delta_\gamma\right)-\rho_v\left(V_i(v)\right)\right)\right)^2\\
\leq&E\left(\rho_v\left(V_i(v)-X_i'\Delta_\gamma\right)-\rho_v\left(V_i(v)\right)\right)^2\\
<&E(X_i'\Delta_\gamma)^2=E\left(\sum_{j=1}^p x_{ij}\Delta_{\gamma_j}\right)^2\\
\leq&E\left(\left(\sum_{j=1}^p x_{ij}^2\right)\left(\sum_{j=1}^p\Delta_{\gamma_j}^2\right)\right)\\
\leq&E\|X_i\|^2\|\Delta_\gamma\|^2\leq C_X^2t^2,
\end{aligned}
$$

where the last third inequality follows Lemma B.1 and following inequality is due to Cauchy–Schwarz inequality.

**Lemma B.1.** *The check function*

$$\rho_\tau(u) = \begin{cases} u\tau & \text{if } u \geq 0, \\ u(\tau - 1) & \text{if } u < 0 \end{cases}$$

*is a contraction mapping for* $0 < \tau < 1$*.*

**Proof:** See Section B.3 in Appendix B.

By the symmetrization Lemma for probabilities, Lemma 2.3.7 in Van der Vaart and Wellner (1996), we obtain

$$\begin{aligned} P(\mathcal{R}(t) > M) &\leq \frac{2P\left(\mathcal{R}^o(t) > \frac{M}{4}\right)}{1 - \frac{4C_X^2 t^2}{M^2}} \\ &\leq \frac{2P\left(\mathcal{R}^o(t) > \frac{M}{4}|\Omega_1\right) P(\Omega_1) + 2P\left(\Omega_1^c\right)}{1 - \frac{4C_X^2 t^2}{M^2}} \\ &\leq \frac{2P\left(\mathcal{R}^o(t) > \frac{M}{4}|\Omega_1\right)}{1 - \frac{4C_X^2 t^2}{M^2}} \end{aligned}$$

where $\mathcal{R}^o(t)$ is the symmetrized version of $\mathcal{R}(t)$, constructed by replacing the empirical process $\mathbb{G}_n$ with its symmetrized version $\mathbb{G}_n^o$. The first inequality follows Lemma 2.3.7 of Van der Vaart and Wellner (1996), and the last inequality is due to $P(\Omega_1) \to 1$, which is defined in Lemma A.1. Since $\frac{C_X t}{M} \to 0$, it is sufficient to show $P\left(\mathcal{R}^o(t) > \frac{M}{4}|\Omega_1\right) \to 0$ in next section.

We divide $\mathcal{R}^o(t)$ into two parts. Notice that

$$\begin{aligned} &\rho_v\left(V_i(v) - X_i'\Delta_\gamma\right) - \rho_v\left(V_i(v)\right) \\ =&\left(v - 1\{V_i(v) - X_i'\Delta_\gamma < 0\}\right)\left(V_i(v) - X_i'\Delta_\gamma\right) - \left(v - 1\{V_i(v) < 0\}\right)V_i(v) \\ =&v(V_i(v) - X_i'\Delta_\gamma) - vV_i(v) - 1\{V_i(v) - X_i'\Delta_\gamma < 0\}(V_i(v) - X_i'\Delta_\gamma) + 1\{V_i(v) < 0\}V_i(v) \\ =& - vX_i'\Delta_\gamma + \delta_i(X_i'\Delta_\gamma, v), \end{aligned}$$

where $\delta_i(X_i'\Delta_\gamma, v) = (V_i(v) - X_i'\Delta_\gamma)_- - (V_i(v))_-$. Define

$$\begin{aligned} \mathcal{B}^o(t) &= \sup_{\substack{v\in\mathcal{U} \\ \Delta_\gamma\in\mathcal{A}_v \\ \|\Delta_\gamma\|^2\leq t^2}} |\mathbb{G}_n^o(X_i'\Delta_\gamma)|, \\ \mathcal{C}^o(t) &= \sup_{\substack{v\in\mathcal{U} \\ \Delta_\gamma\in\mathcal{A}_v \\ \|\Delta_\gamma\|^2\leq t^2}} |\mathbb{G}_n^o(\delta_i(X_i'\Delta_\gamma, v))|, \end{aligned}$$

then

$$\begin{aligned}
\mathcal{R}(t) \leq & \sup_{\substack{v\in\mathcal{U}\\ \Delta_\gamma\in\mathcal{A}_v\\ \|\Delta_\gamma\|^2\leq t^2}} |\mathbb{G}_n^o\left(1\{X_i'\gamma_0(v) > \frac{\delta_n}{3}\}\left(\rho_v\left(V_i(v) - X_i'\Delta_\gamma\right) - \rho_v\left(V_i(v)\right)\right)\right)| \\
\leq & \sup_{\substack{v\in\mathcal{U}\\ \Delta_\gamma\in\mathcal{A}_v\\ \|\Delta_\gamma\|^2\leq t^2}} |\mathbb{G}_n^o\left(\rho_v\left(V_i(v) - X_i'\Delta_\gamma\right) - \rho_v\left(V_i(v)\right)\right)| \\
\leq & \mathcal{B}^o(t) + \mathcal{C}^o(t).
\end{aligned}$$

Thus,

$$\begin{aligned}
& P\left(\mathcal{R}^o(t) > \frac{M}{4}|\Omega_1\right) \\
\leq & P\left(\mathcal{B}^o(t) + \mathcal{C}^o(t) > \frac{M}{4}|\Omega_1\right) \\
\leq & 2\max\left\{P\left(\mathcal{B}^o(t) > \frac{M}{8}|\Omega_1\right), P\left(\mathcal{C}^o(t) > \frac{M}{8}|\Omega_1\right)\right\}.
\end{aligned}$$

w.p.a.1.

**Bound on $\mathcal{B}^o(t)$**: Let $M_1 = \sqrt{2}(\mathcal{E}_X + 1)\sqrt{s^\gamma \log((p+1)n)}t$. There exists a $\lambda \geq 0$,

$$\begin{aligned}
P\left(\mathcal{B}^o(t) > M_1|X,\Omega_1\right) = & P\left(\exp\left(\lambda\mathcal{B}^o(t)\right) > \exp\left(\lambda M_1\right)|X,\Omega_1\right) \\
\leq & \min_{\lambda\geq 0} \frac{E\left(\exp\left(\lambda\mathcal{B}^o(t)\right)|X,\Omega_1\right)}{\exp\left(\lambda M_1\right)} \\
\leq & \min_{\lambda\geq 0} e^{-\lambda M_1} E\left(\exp\left(\lambda \sup_{\|\Delta_\gamma\|^2\leq t^2} \|\Delta_\gamma\|_1 \max_{1\leq j\leq p} |\mathbb{G}_n^o(x_{ij})|\right)|X,\Omega_1\right) \\
\leq & 2p\min_{\lambda\geq 0} e^{-\lambda M_1} \max_{1\leq j\leq p} E\left(\exp\left(\lambda\sqrt{s^\gamma}t\mathbb{G}_n^o(x_{ij})\right)|X,\Omega_1\right),
\end{aligned}$$

the first inequality is by Markov's inequality and the last equation follows from the bound for the symmetric random variable $Z_j$

$$E\left(\max_{1\leq j\leq p} e^{|Z_j|}\right) \leq p\max_{1\leq j\leq p} E\left(e^{|Z_j|}\right) \leq p\max_{1\leq j\leq p} E\left(e^{Z_j} + e^{-Z_j}\right) \leq 2p\max_{1\leq j\leq p} E\left(e^{Z_j}\right)$$

and $\sup_{v\in\mathcal{U}} |\mathbb{S}_\gamma(v)| \leq s^\gamma$. Since for any $\lambda \geq 0$ and a Rademacher variable $\xi$, one has

$$E\left(e^{\lambda\xi}\right) = \frac{1}{2}e^\lambda + \frac{1}{2}e^{-\lambda} \leq e^{\frac{\lambda^2}{2}},$$

$P\left(\mathcal{B}^o(t) > M_1|X,\Omega_1\right)$ can be further bounded by

$$P\left(\mathcal{B}^o(t) > M_1|X,\Omega_1\right) \leq 2p\min_{\lambda\geq 0} e^{-\lambda M_1} \max_{1\leq j\leq p} \exp\left(\frac{1}{2}\lambda^2 s^\gamma t^2 \frac{1}{n}\|X_j\|^2\right)$$

$$
\begin{aligned}
&\leq 2p \min_{\lambda \geq 0} e^{-\lambda M_1} \exp\left(\frac{\lambda^2 s^\gamma t^2 (\mathcal{E}_X + 1)^2}{2}\right) \\
&\leq 2p \min_{\lambda \geq 0} \exp\left(-\lambda M_1 + \frac{\lambda^2 s^\gamma t^2 (\mathcal{E}_X + 1)^2}{2}\right) \\
&\leq 2p \exp\left(-\frac{M_1^2}{2 s^\gamma t^2 (\mathcal{E}_X + 1)^2}\right) \\
&= \frac{2p}{(p+1)n} \to 0,
\end{aligned}
$$

where the minimum of the third inequality as a function of $\lambda$, we cound define $g : \mathbb{R}^+ \to \mathbb{R}$ such that $g(\lambda) = -\lambda M_1 + \frac{s^\gamma t^2 (\mathcal{E}_X+1)^2}{2} \lambda^2$, then $g(\lambda)$ achieves its minimum at $\lambda = \frac{M_1}{s^\gamma t^2 (\mathcal{E}_X+1)^2}$. Plugging in this value of $\lambda$ and $M_1$ in the last two equalities we obtain the desired result. Therefore,

$$
\begin{aligned}
P\left(\mathcal{B}^o(t) > M_1 | X, \Omega_1\right) &= E\left(1\{\mathcal{B}^o(t) > M_1\} | \Omega_1\right) \\
&= E\left(E\left(1\{\mathcal{B}^o(t) > M_1\} | X, \Omega_1\right) | \Omega_1\right) \\
&= E\left(P\left(\mathcal{B}^o(t) > M_1 | X, \Omega_1\right) | \Omega_1\right) \\
&\leq E\left(\frac{2p}{(p+1)n}\right) \leq E\left(\frac{2}{n}\right) \to 0.
\end{aligned}
$$

**Bound on $\mathcal{C}^o(t)$:** Let $M_2 = 4(\mathcal{E}_X + 1)\sqrt{s^\gamma C_\xi \log\left((p+1)n\right)} t$. For any $\hat{v} \in \mathcal{U}$, we have

$$
\begin{aligned}
\delta_i(X_i' \Delta_\gamma, v) &= (V_i(v) - X_i' \Delta_\gamma)_- - (V_i(v))_- \\
&= (V_i(\hat{v}) - X_i'(\Delta_\gamma + \gamma_0(v) - \gamma_0(\hat{v})))_- - (V_i(\hat{v}))_- - (V_i(\hat{v}) - X_i'(\gamma_0(v) - \gamma_0(\hat{v})))_- + (V_i(\hat{v}))_- \\
&= \delta_i(X_i'(\Delta_\gamma + \gamma_0(v) - \gamma_0(\hat{v})), \hat{v}) - \delta_i(X_i'(\gamma_0(v) - \gamma_0(\hat{v})), \hat{v}).
\end{aligned}
$$

By triangle inequality, we have

$$
\begin{aligned}
\mathcal{C}^o(t) &\leq \sup_{\substack{v \in \mathcal{U}, |v - \hat{v}| < \epsilon \\ \hat{v} \in \mathcal{U}_k \\ \Delta_\gamma \in \mathcal{A}_v, \|\Delta_\gamma\|^2 \leq t^2}} |\mathbb{G}_n^o(\delta_i(X_i'(\Delta_\gamma + \gamma_0(v) - \gamma_0(\hat{v})), \hat{v}))| + \sup_{\substack{v \in \mathcal{U}, |v - \hat{v}| < \epsilon \\ \hat{v} \in \mathcal{U}_k}} |\mathbb{G}_n^o(\delta_i(X_i'(\gamma_0(v) - \gamma_0(\hat{v})), \hat{v}))| \\
&\leq 2 \sup_{\substack{\hat{v} \in \mathcal{U}_k \\ \|\Delta_\gamma\|^2 \leq 2C_\xi t^2}} |\mathbb{G}_n^o(\delta_i(X_i' \Delta_\gamma, \hat{v}))| := 2\mathcal{D}^o(t),
\end{aligned}
$$

where the first inequality is by $\epsilon$-net of $\mathcal{U}$, where $\epsilon = \frac{t}{\sqrt{p+1}}$ and $\epsilon k \leq 1$. The last inequality is by treating $\Delta_\gamma + \gamma_0(v) - \gamma_0(\hat{v})$ and $\gamma_0(v) - \gamma_0(\hat{v})$ as new $\Delta_\gamma$. By Assumption 5, we have

$$
\|\gamma_0(v) - \gamma_0(\hat{v})\|^2 \leq C_L^2 (v - \hat{v})^2 \leq \frac{C_L^2}{p+1} t^2,
$$

then,

$$
\|\Delta_\gamma + \gamma_0(v) - \gamma_0(\hat{v})\|^2 \leq \left(\|\Delta_\gamma\| + \|\gamma_0(v) - \gamma_0(\hat{v})\|\right)^2
$$

$$
\begin{aligned}
&\leq \left(t + \frac{C_L}{\sqrt{p+1}} t\right)^2 \\
&\leq 2\left(1 + \frac{C_L^2}{p+1}\right) t^2 \leq 2C_\xi t^2
\end{aligned}
$$

where $C_\xi \geq 1 + \frac{C_L^2}{p+1}$ as $C_L = O(\sqrt{p})$. The third inequality is because $2\left(a^2+b^2\right) \geq (a+b)^2$ for $a, b \in \mathbb{R}$.

As we discussed previously, we bound

$$
\begin{aligned}
P\left(\mathcal{C}^o(t) > 2M_2 | X, \Omega_1\right) \leq & P\left(\mathcal{D}^o(t) > M_2 | X, \Omega_1\right) \\
= & P\left(\exp\left(\lambda' \mathcal{D}^o(t)\right) > \exp\left(\lambda' M_2\right) | X, \Omega_1\right) \\
\leq & \min_{\lambda' \geq 0} \frac{E\left(\exp\left(\lambda' \mathcal{D}^o(t)\right) | X, \Omega_1\right)}{\exp\left(\lambda' M_2\right)} \\
\leq & \frac{1}{\epsilon} \min_{\lambda' \geq 0} e^{-\lambda' M_2} \max_{\hat{v} \in \mathcal{U}_k} E\left(\exp\left(\lambda' \sup_{\|\Delta_\gamma\|^2 \leq 2C_\xi t^2} \left|\mathbb{G}_n^o\left(\delta_i\left(X_i' \Delta_\gamma, \hat{v}\right)\right)\right|\right) | X, \Omega_1\right) \\
\leq & \frac{\sqrt{p+1}}{t} \min_{\lambda' \geq 0} e^{-\lambda' M_2} \max_{\hat{v} \in \mathcal{U}_k} E\left(\exp\left(2\lambda' \sup_{\|\Delta_\gamma\|^2 \leq 2C_\xi t^2} \left|\mathbb{G}_n^o\left(X_i' \Delta_\gamma\right)\right|\right) | X, \Omega_1\right) \\
\leq & \frac{\sqrt{p+1}}{t} \min_{\lambda' \geq 0} e^{-\lambda' M_2} E\left(\exp\left(2\lambda' \sup_{\|\Delta_\gamma\|^2 \leq 2C_\xi t^2} \|\Delta_\gamma\|_1 \max_{1 \leq j \leq p} |\mathbb{G}_n^o(x_{ij})|\right) | X, \Omega_1\right) \\
\leq & \frac{2p\sqrt{p+1}}{t} \min_{\lambda' \geq 0} e^{-\lambda' M_2} \max_{1 \leq j \leq p} E\left(\exp\left(2\lambda' \sqrt{s^\gamma} \sqrt{2C_\xi} t \mathbb{G}_n^o(x_{ij})\right) | X, \Omega_1\right) \\
\leq & \frac{2p\sqrt{p+1}}{t} \min_{\lambda' \geq 0} e^{-\lambda' M_2} \max_{1 \leq j \leq p} \exp\left(4\lambda'^2 s^\gamma C_\xi t^2 \frac{1}{n} \|X_j\|^2\right) \\
\leq & \frac{2p\sqrt{p+1}}{t} \min_{\lambda' \geq 0} \exp\left(-\lambda' M_2 + 4\lambda'^2 s^\gamma C_\xi t^2 (\mathcal{E}_X + 1)^2\right) \\
\leq & \frac{2p\sqrt{p+1}}{t} \min_{\lambda' \geq 0} \exp\left(-\frac{M_2^2}{16 s^\gamma C_\xi t^2 (\mathcal{E}_X + 1)^2}\right) \\
= & \frac{2p\sqrt{p+1}}{(p+1)nt} \leq \frac{2(p+1)^{\frac{1}{2}}}{nt} \to 0,
\end{aligned}
$$

where the fourth inequality is by Theorem 4.12 of Ledoux and Talagrand (2013), and by contractivity of $\delta_i(\cdot)$, i.e., $|\delta_i(a, \hat{v}) - \delta_i(a, \hat{v})| \leq |a - b|$, with $\delta_i(0) = 0$. Thus,

$$
\begin{aligned}
P(\mathcal{R}^o(t) > \frac{M}{4} | \Omega_1) &\leq 2 \max\left\{P\left(\mathcal{B}^o(t) > \frac{M}{8} | \Omega_1\right), P\left(\mathcal{C}^o(t) > \frac{M}{8} | \Omega_1\right)\right\} \\
&\leq 2 \max\left\{P(\mathcal{B}^o(t) > M_1 | \Omega_1), P(\mathcal{C}^o(t) > 2M_2 | \Omega_1)\right\} \\
&\to 0,
\end{aligned}
$$

then, we choose $M = C_{\sup} \sqrt{s^\gamma \log((p+1)n)} t \geq 8 \max\{M_1, 2M_2\}$, where $C_{\sup} = 64\sqrt{C_\xi}(\mathcal{E}_X + 1)$, and finally obtain $P(\mathcal{R}(t) > M) \to 0$ w.p.a.1. $\square$

### B.3 Proof of Lemma B.1

*Proof.* **Case 1**: $u_1 \geq 0$ and $u_2 \geq 0$.

$$|\rho_\tau(u_1) - \rho_\tau(u_2)| = |\tau u_1 - \tau u_2| = \tau |u_1 - u_2|.$$

**Case 2**: $u_1 < 0$ and $u_2 < 0$.

$$|\rho_\tau(u_1) - \rho_\tau(u_2)| = |(\tau - 1)u_1 - (\tau - 1)u_2| = (\tau - 1)|u_1 - u_2|.$$

**Case 3**: $u_1 \geq 0$ and $u_2 < 0$.

$$|\rho_\tau(u_1) - \rho_\tau(u_2)| = |\tau u_1 - (\tau - 1)u_2| \leq \max\{\tau - 1, \tau\}|u_1 - u_2|.$$

**Case 4**: $u_1 < 0$ and $u_2 \geq 0$.

$$|\rho_\tau(u_1) - \rho_\tau(u_2)| = |(\tau - 1)u_1 - \tau u_2| \leq \max\{\tau - 1, \tau\}|u_1 - u_2|$$

Therefore, there exists a $0 \leq k < 1$ such that $|\rho_\tau(u_1) - \rho_\tau(u_2)| \leq k|u_1 - u_2| < |u_1 - u_2|$. $\square$

# Appendix C Additional Simulation Results

Table 12: Performance for Frank copula under different dependence strengths

| Parameter | True | Bias | RMSE | SD |
|---|---|---|---|---|
| Censoring Rate ≈ 17% | | | | |
| | | $n = 300$ | | |
| $\eta$ | 2.9174 | -0.1776 | 0.9907 | 0.9759 |
| Kendall's $\tau$ | 0.3000 | -0.0215 | 0.0900 | 0.0875 |
| | | $n = 600$ | | |
| $\eta$ | 2.9174 | -0.0815 | 0.7685 | 0.7651 |
| Kendall's $\tau$ | 0.3000 | -0.0108 | 0.0686 | 0.0679 |
| | | $n = 300$ | | |
| $\eta$ | 5.7363 | -0.2410 | 1.1815 | 1.1581 |
| Kendall's $\tau$ | 0.5000 | -0.0202 | 0.0678 | 0.0651 |
| | | $n = 600$ | | |
| $\eta$ | 5.7363 | -0.2359 | 0.8321 | 0.7990 |
| Kendall's $\tau$ | 0.5000 | -0.0165 | 0.0490 | 0.0461 |
| | | $n = 300$ | | |
| $\eta$ | 11.4115 | -0.6788 | 1.9436 | 1.8235 |
| Kendall's $\tau$ | 0.7000 | -0.0212 | 0.0480 | 0.0430 |
| | | $n = 600$ | | |
| $\eta$ | 11.4115 | -0.4921 | 1.3541 | 1.2631 |
| Kendall's $\tau$ | 0.7000 | -0.0138 | 0.0332 | 0.0296 |
| | | $n = 300$ | | |
| $\eta$ | 38.2812 | -6.5997 | 7.5814 | 3.7357 |
| Kendall's $\tau$ | 0.9000 | -0.0214 | 0.0265 | 0.0160 |
| | | $n = 600$ | | |
| $\eta$ | 38.2812 | -5.1848 | 5.7064 | 2.3864 |
| Kendall's $\tau$ | 0.9000 | -0.0154 | 0.0173 | 0.0088 |

Table 13: Performance for Frank copula under high censoring rate

| Parameter | True | Bias | RMSE | SD |
|---|---|---|---|---|
| Censoring Rate $\approx 35\%$ | | | | |
| | $n = 300$ | | | |
| $\eta$ | 11.4115 | -1.3657 | 2.3538 | 1.9196 |
| Kendall's $\tau$ | 0.7000 | -0.0401 | 0.0648 | 0.0506 |
| | $n = 600$ | | | |
| $\eta$ | 11.4115 | -1.2469 | 1.7992 | 1.2987 |
| Kendall's $\tau$ | 0.7000 | -0.0331 | 0.0469 | 0.0339 |
| | | | | |
| | $n = 300$ | | | |
| $\eta$ | -11.4115 | 1.0744 | 2.0066 | 1.6968 |
| Kendall's $\tau$ | -0.7000 | 0.0310 | 0.0548 | 0.0456 |
| | $n = 600$ | | | |
| $\eta$ | -11.4115 | 1.0229 | 1.6953 | 1.3537 |
| Kendall's $\tau$ | -0.7000 | 0.0275 | 0.0447 | 0.0350 |

Notes: Due to the change in censoring rate, $\gamma_0$ is set to $-0.5$ in this DGP.

Table 14: Performance for Gaussian copula under high censoring rate

| Parameter | True | Bias | RMSE | SD |
|---|---|---|---|---|
| Censoring Rate $\approx 35\%$ | | | | |
| | $n = 300$ | | | |
| $\rho$ | 0.7000 | -0.0448 | 0.0922 | 0.0806 |
| | $n = 600$ | | | |
| $\rho$ | 0.7000 | -0.0359 | 0.0632 | 0.0523 |
| | | | | |
| | $n = 300$ | | | |
| $\rho$ | -0.7000 | 0.0223 | 0.0837 | 0.0810 |
| | $n = 600$ | | | |
| $\rho$ | -0.7000 | 0.0226 | 0.0566 | 0.0522 |

Notes: Due to the change in censoring rate, $\gamma_0$ is set to $-0.5$ in this DGP.