QUANTILE REGRESSION WITH AN ENDOGENOUS ONE-SIDED MISCLASSIFIED BINARY REGRESSOR*

CARLOS LAMARCHE†

Abstract: Addressing misreporting of participation in social programs, which is common and has increased in all major surveys, is important to study intergenerational effects of policies. In this paper, we propose a simple two-step estimator for a quantile regression model with endogenous one-sided misreporting. The identification of the model uses a parametric first stage and information related to participation and misreporting. We show that the estimator is consistent and asymptotically normal. We also establish that a bootstrap procedure is asymptotically valid for approximating the distribution of the estimator. Simulation studies show the small sample behavior of the estimator in comparison with other methods. Finally, we illustrate the approach using U.S. survey data to estimate the intergenerational effect of mother's participation on welfare on daughter's adult income.

Keywords: Quantile regression; Misclassification; Endogenous Treatments; Survey data.
JEL classification: C21; C25; I32.

1. Introduction

A growing concern in the social sciences is the declining quality of household survey data (Meyer, Mok, and Sullivan, 2015). It is well documented that survey respondents have become less likely to answer certain questions, including whether they participate in social programs. Recent theoretical and methodological research on quantile regression have addressed important generalizations of the celebrated Koenker and Bassett's estimator, but

*This draft: February 6, 2024. First version: November 15, 2021. We are grateful to Chris Bollinger, Roger Koenker, Pierre Nguimkeu, Tom Parker, Takuya Ura, and Jim Ziliak for useful comments, as well as seminar participants at the University of Kentucky, Universidad de la República, Universidad de La Plata, the 2021 SEA conference, and the 2023 Midwest Econometrics Group conference. We would like to thank Austin Denteh for making available his code and Rob Hartley for sharing the data used in Section 5.

†Carlos Lamarche: Department of Economics, University of Kentucky, 223G Gatton College of Business & Economics, Lexington, KY 40506. Email: clamarche@uky.edu.

research on misclassified data remains sparse (Koenker, 2017). Practitioners face the limitations of classical parametric models and policy recommendations could miss important heterogeneity since they can only be based on average effects.

An exception in the literature is the recent work by Ura (2021), who develops identification results for a quantile model with a misclassified (or misreported) binary variable indicating treatment status. Misclassified regressors are common when practitioners evaluate the impact of programs using survey data with high levels of item non-response. For welfare programs, for instance, this relates to misreporting of whether a person received benefits, duration of social assistance, and amount of dollars received. Using data from Meyer, Mok, and Sullivan (2015, 2009), the left panel of Figure 1.1 shows the number of households participating in three of the most important welfare programs in the US. The programs are: Aid to Families with Dependent Children/Temporary Assistance for Needy Families (AFDC/TANF), Food Stamps, and the Supplemental Security Income. The continuous line is obtained from administrative records and the dashed lines are weighted survey estimates obtained from the Current Population Survey (CPS), the Survey of Income and Program Participation (SIPP), and the Panel Study of Income Dynamics (PSID), which is used in Section 5. The right panel of Figure 1.1 shows the reporting rate, which is defined as the ratio of survey reports and administrative cases. Reporting rates have been declining in the CPS and PSID since 1980, reaching a level as low as 55% around 2000. Because these programs target low income households, this evidence illustrates the importance of estimating quantile effects of welfare program participation while allowing misreporting to be endogenous.

Motivated by the challenges, we investigate the estimation of a quantile regression model when participation is endogenously misreported. Thus, relative to the work on partial identification of Ura (2021), we provide conditions for point identification and propose a new quantile regression estimator. Our estimation procedure adopts the partial observability model recently considered in Nguimkeu, Denteh, and Tchernis (2019), which requires a parametric first step and two exogenous measurements related to the observed (possibly endogeneous) binary regressor. In contrast to their paper, the development in the second step is critically different since we estimate heterogeneous treatment effects by using quantile regression. The estimator is simple to compute and easy to be implemented in applications when respondents report not participating in a program when in fact they did participate.

We investigate the asymptotic properties of the proposed estimator and establish three results. First, we establish the consistency of the estimator. We then derive the asymptotic

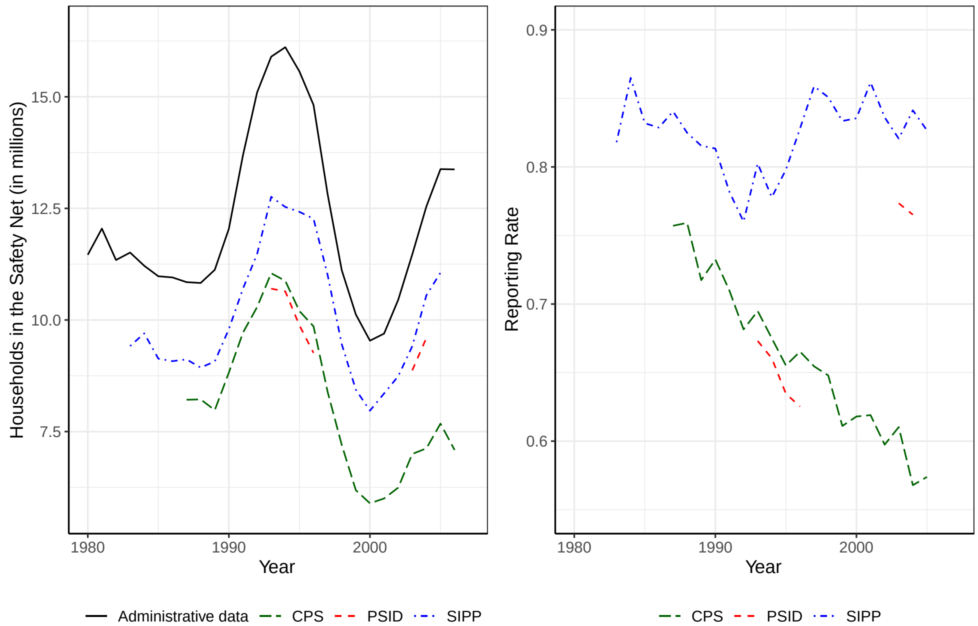

FIGURE 1.1. *Average monthly participation by year in the US safety net and estimated reporting rates.*

distribution of the proposed estimator, and obtain an asymptotic covariance matrix that could be seen as similar to the ones derived for other two-step quantile regression estimators (e.g., Ma and Koenker, 2006; Chernozhukov, Fernández-Val, and Kowalski, 2015; Chen, Galvao, and Song, 2021). Inference based on the asymptotic distribution of the proposed estimator requires estimation of nuisance parameters and first-order partial derivatives of conditional functions. Because estimation of these parameters might be difficult, it is useful to have an alternative inference procedure. Therefore, our last result is to offer a valid framework for inference. We demonstrate the validity of a bootstrap method to approximate the asymptotic distribution of the quantile estimator.

Our paper is related to research investigating mismeasured continuous regressors in a quantile regression model. Wei and Carroll (2009) develop a consistent estimator in the presence of covariate measurement error, and Wang, Stefanski, and Zhu (2012) adapt the classical quantile regression problem to Gaussian and Laplace measurement error models. He and Liang (2000) investigate estimation of quantile coefficients when errors in the outcome equation and covariates are independent and their distribution is symmetric, and Schennach (2008) considers a nonparametric measurement error model employing deconvolution methods. Chesher (2017) investigates the relationship between quantile regression functions corresponding to error-free and error-contaminated variables when the variance of the measurement error is small. Firpo, Galvao, and Song (2017) investigate measurement error in quantile regression when researchers have multiple noisy measurements of latent variables. The literature on estimation of conditional mean models is broader relative to quantile regression, and it includes Bollinger (1996), Frazis and Loewenstein (2003), Mahajan (2006), Lewbel (2007), Kreider, Pepper, Gundersen, and Jolliffe (2012), Nguimkeu, Denteh, and Tchernis (2019), DiTraglia and García-Jimeno (2019), among others.

This paper is organized as follows. The next section presents the model and the estimator, and Section 3 presents theoretical results. Section 4 investigates the small sample performance of the method, showing that the estimator has satisfactory performance under different specifications considered in the literature. Section 5 illustrates the theory and provides practical guidelines from an application of the method. Considering data from the PSID, we estimate a quantile intergenerational parameter to study how mother's participation on welfare during her daughter's childhood affects daughter's adult income. A major difficulty of estimating this intergenerational parameter is the low reporting rates in the PSID (Figure 1.1). Finally, Section 6 concludes. Mathematical proofs are offered in the Appendix.

2. Model and Methods

We consider a continuous outcome variable, $y_i \in \mathbb{R}$, a latent binary regressor $d_i^* \in \{0,1\}$ indicating true participation status, and a p-dimensional vector of exogenous independent variables, $\boldsymbol{x}_i$. It is assumed that the vector $\boldsymbol{x}_i$ includes an intercept. Instead of d_i^*, we observe a surrogate $d_i \in \{0,1\}$ for each subject $1 \leq i \leq n$, where n denotes the number of cross-sectional units. Our interest is to investigate the effect of d_i^* on the conditional distribution of the response variable y_i using the following model:

$$Q_{y_i}(\tau|\boldsymbol{x}_i, d_i^*) = \boldsymbol{x}_i'\boldsymbol{\beta}_0(\tau) + d_i^*\alpha_0(\tau), \tag{2.1}$$

where $\tau \in (0,1)$ and the function $Q_{y_i}(\tau|\boldsymbol{x}_i, d_i^*)$ is the τ-th quantile of the conditional distribution of y_i given $\boldsymbol{x}_i$ and d_i^*. Because we consider one value of τ throughout the paper, we suppress the dependence of the parameters $\alpha_0(\tau)$ and $\boldsymbol{\beta}_0(\tau)$ on τ for notational simplicity.

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \alpha)' \in \boldsymbol{\Theta} \subseteq \mathbb{R}^{p+1}$, and let $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0', \alpha_0)'$. The quantile treatment effect, α_0, is defined as,

$$\alpha_0 := Q_{y_i}(\tau|\boldsymbol{x}_i, d_i^* = 1) - Q_{y_i}(\tau|\boldsymbol{x}_i, d_i^* = 0). \tag{2.2}$$

If the participation status is known by the researcher, we might estimate $\boldsymbol{\theta}_0$ using

$$\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\alpha})' = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{x}_i'\boldsymbol{\beta} - d_i^*\alpha), \tag{2.3}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the quantile regression loss function. The problem with the estimator defined by (2.3) is that the variable d_i^* is not observed in applied practice, as in the application considered in Section 5.

2.1. **Background.** Ura (2021) investigates identification in a quantile regression model when a binary regressor indicating treatment status is not observed. He adopts the framework developed by Chernozhukov and Hansen (2005, 2006) by considering a vector of instrumental variables, denoted here by $\boldsymbol{z}_i$. The following result holds for a discrete or continuous vector of instruments, but for simplicity, we consider the case that z_i is a scalar binary variable.

Lemma 1 (Ura, 2021)**.** *Let $\mathfrak{F}_i = (y_i, z_i, \boldsymbol{x}_i')$, $\pi_0 = P(d_i = 0|d_i^* = 1, \mathfrak{F}_i)$ and $\pi_1 = P(d_i = 1|d_i^* = 0, \mathfrak{F}_i)$, and assume a monotone positive relationship between d_i^* and z_i. If $Pr(d_i \neq d_i^*|d_i^* = 0, \mathfrak{F}_i) + Pr(d_i \neq d_i^*|d_i^* = 1, \mathfrak{F}_i) < 1$, then there exists a $\kappa \in [0, 1]$ such that*

$$\tilde{\alpha}_0 := Q_{y_i}(\tau|\boldsymbol{x}_i, z_i = 1) - Q_{y_i}(\tau|\boldsymbol{x}_i, z_i = 0) = \kappa\alpha_0.$$

The result has important implications for empirical practice, as it states that the reduced-form quantile regression coefficient of y_i on z_i provides a bound to the structural treatment effect parameter (2.2). The implication is that the quantile coefficient of regressing the response variable on an instrument is biased toward zero. This result can be further investigated by extending Section 3 in Ura (2021) as follows.

Lemma 1 suggests we estimate a quantile regression model using the instrument z_i instead of the latent variable d_i^*. Define $\boldsymbol{\delta}_1 = \sqrt{n}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$, $\delta_2 = \sqrt{n}(\alpha - \alpha_0)$, and $\boldsymbol{\delta} = (\boldsymbol{\delta}_1', \delta_2)'$. In this case, after simple algebra, we can rewrite the objective function in (2.3) as:

$$\mathbb{V}_n(\boldsymbol{\delta}) = \sum_{i=1}^{n} \left\{ \rho_\tau\left(\tilde{u}_i - \frac{\boldsymbol{x}_i'\boldsymbol{\delta}_1}{\sqrt{n}} - \frac{z_i\delta_2}{\sqrt{n}}\right) - \rho_\tau(\tilde{u}_i) \right\}, \tag{2.4}$$

where $\tilde{u}_i = y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_0 - z_i\alpha_0 = u_i^* + (d_i^* - z_i)\alpha_0$ and $u_i^* = y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_0 - d_i^*\alpha_0$. Using developments in Knight (1998) and Koenker (2005), when n is large, letting $A \approx B$ mean that A is approximately distributed as B,

$$\mathbb{V}_n(\boldsymbol{\delta}) \approx -\boldsymbol{\delta}'\boldsymbol{B}_n + \frac{1}{2}\boldsymbol{\delta}'\boldsymbol{D}_n\boldsymbol{\delta}, \tag{2.5}$$

where

$$\boldsymbol{B}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{bmatrix}\boldsymbol{x}_i \\ z_i\end{bmatrix}\psi_\tau(u_i^* + (d_i^* - z_i)\alpha_0), \qquad \boldsymbol{D}_n = \frac{1}{n}\sum_{i=1}^{n} f_i(0|\boldsymbol{x}_i, z_i)\begin{bmatrix}\boldsymbol{x}_i\boldsymbol{x}_i' & z_i\boldsymbol{x}_i \\ z_i\boldsymbol{x}_i' & z_i\end{bmatrix}.$$

Using the definition of the quantile score function, we obtain:

$$\mathrm{E}\left[\psi_\tau(u_i^* + (d_i^* - z_i)\alpha_0)|\boldsymbol{x}_i, z_i\right] = \tau - \Pr(u_i^* + \alpha_0 d_i^* \leq z_i\alpha_0|\boldsymbol{x}_i, z_i).$$

Assuming that d_i^* is weakly increasing in z_i (i.e., $d_i^*(z_i = 0) \leq d_i^*(z_i = 1)$), and because $u_i^* + \alpha_0 d_i^* = y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_0$, we can write,

$$\Pr(u_i^* \leq 0|\boldsymbol{x}_i, d_i^* = z_i = 0) - \Pr(u_i^* \leq \alpha_0|\boldsymbol{x}_i, d_i^* = 0, z_i = 1) \leq 0,$$

if $\alpha_0 \geq 0$. Therefore, $\mathrm{E}\left[\boldsymbol{B}_n|\boldsymbol{x}_i, z_i\right] \leq 0$ and the solution of (2.5) is expected to be (given the approximation) downward biased if the latent variable d_i^* is exogenous.

Remark 1. Note that $\mathrm{E}\left[\psi_\tau(u_i^*)|\boldsymbol{x}_i, d_i^*\right] = 0$ does not hold if d_i^* is endogenous. Note also that d_i^* is assumed to be weakly increasing in z_i. It allows for some $1 \leq i \leq n$ but not all subjects to change d_i^* from 1 to 0 when z_i changes from 0 to 1. If the sample includes compliers instead of deniers and $\alpha_0 < 0$, the solution of (2.5) is expected to provide an upper bound for α_0.

In the next section, we propose an estimator that does not suffer from this issue. The idea is to recenter the error term in (2.4) by replacing z_i by the conditional expectation $\mathrm{E}\left[d_i^*|\boldsymbol{x}_i, z_i\right]$. For consistency of the estimator, one could require that $d_i^* - \mathrm{E}\left[d_i^*|\boldsymbol{x}_i, z_i\right]$, a Bernoulli recentered random variable, is independent of the error term of the model. However, the model presented in the next section allows for weak forms of dependence as long as the ranks of the variables are not sufficiently different. Similar conditions are introduced in the literature (Chernozhukov and Hansen, 2005; Ura, 2021) and the conditional expectation can be estimated in scenarios of incomplete data under the conditions introduced below.

2.2. **Modeling framework.** We now introduce the remaining part of the model. The latent binary variable is,

$$d_i^* = 1\{\boldsymbol{z}_i'\boldsymbol{\vartheta} + v_i \geq 0\}, \tag{2.6}$$

where $\boldsymbol{\vartheta}$ is a k_1-dimensional parameter, $\boldsymbol{z}_i$ is a vector of instruments that includes the exogenous independent variables $\boldsymbol{x}_i$, and v_i is an error term. Moreover, we model misreporting behavior by considering,

$$m_i = 1\{\boldsymbol{w}_i'\boldsymbol{\gamma} + \epsilon_i \geq 0\}, \tag{2.7}$$

where $\boldsymbol{\gamma}$ is a k_2-dimensional parameter, $\boldsymbol{w}_i$ is a vector of variables that are different than $\boldsymbol{z}_i$, and ϵ_i is an error term. Therefore, the observed binary variable is modeled by,

$$d_i = d_i^* m_i = 1\{\boldsymbol{z}_i'\boldsymbol{\vartheta} + v_i \geq 0, \boldsymbol{w}_i'\boldsymbol{\gamma} + \epsilon_i \geq 0\}. \tag{2.8}$$

The last equation follows the partial observability model introduced by Poirier (1980) and recently adopted by Nguimkeu, Denteh, and Tchernis (2019). We do not observe individual decisions d_i^* and m_i but $d_i = 1$ implies $d_i^* = 1$ and $m_i = 1$, and $d_i = 0$ implies $d_i^* = 0$ and/or $m_i = 0$. Sufficient variation in $\boldsymbol{z}_i$ and $\boldsymbol{w}_i$ allows point identification of the effect of interest under different assumptions, including non Gaussian conditions. The estimation strategy below relies on observing both $\boldsymbol{z}_i$ and $\boldsymbol{w}_i$, and at least one variable in $\boldsymbol{w}_i$ or $\boldsymbol{z}_i$ has to be continuous.

The model is completed with the following assumptions:

A1. For each $\tau \in (0,1)$, considering the quantile function in (2.1), $\Pr(y_i \leq Q_{y_i}(\tau|\boldsymbol{d}_i^*, \boldsymbol{x}_i)|\boldsymbol{x}_i, \boldsymbol{z}_i) = \tau$. Moreover, the indicator variable $d_i \in \mathbb{R}$ is generated according to equation (2.8) and the error terms (v_i, ϵ_i) are independent of $\boldsymbol{z}_i$, $\boldsymbol{w}_i$, and $\boldsymbol{x}_i$ and have unit variances.

A2. Let $F_{v,\epsilon}(v, \epsilon; \rho)$ denote the joint distribution of v and ϵ with correlation coefficient ρ, and let $F_v(v)$ denote a marginal distribution. Conditional on $(\boldsymbol{z}_i', \boldsymbol{w}_i')$, $F_{v,\epsilon}(v, \epsilon; \rho) = \Phi_{v,\epsilon}(v, \epsilon; \rho)$, the bivariate normal distribution with parameter ρ.

A3. Let $\Phi_i(\boldsymbol{\vartheta}) := F_v(\boldsymbol{z}_i'\boldsymbol{\vartheta})$, $u_i := y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_0 - \alpha_0\Phi_i(\boldsymbol{\vartheta})$, and $\xi_i := d_i^* - \Phi_i(\boldsymbol{\vartheta})$. Assume that u_i^* and ξ_i are independent, and $u_i^*|\boldsymbol{x}_i, \boldsymbol{z}_i$ and $u_i|\boldsymbol{x}_i, \boldsymbol{z}_i$ are rank invariant.

Assumption A1 is different than Assumption 1 in Nguimkeu, Denteh, and Tchernis (2019) by introducing a quantile conditional moment restriction (Chernozhukov and Hansen, 2005; Ura, 2021). Assumption A2 is similar to Assumption 3 in Nguimkeu, Denteh, and Tchernis (2019) and it allows a parametric first step. The normality assumption is convenient for the estimation of the partial observability model, but it is not needed and can be relaxed by considering other absolutely continuous distributions (e.g., bivariate logistic distributions) or semiparametric models (e.g., Cavanagh and Sherman, 1998). Assumption A3 is similar to the rank similarity condition used for identification of the IVQR model of Chernozhukov and Hansen (2005). It is weaker than assuming independence between u_i^* and v_i, because it allows for weak forms of dependence, as long as there is no systematic variation making v_i to cross the threshold (conditional on $\boldsymbol{z}_i'\boldsymbol{\vartheta}$) in equation (2.6).

2.3. **Estimation.** This section describes the proposed two-step estimator. In what follows, we extend the results in Nguimkeu, Denteh, and Tchernis (2019) and Ura (2021) for the problem studied in this paper.

Step 1. Estimate the joint distribution $F_{v,\epsilon}(v, \epsilon; \rho)$ corresponding to the variable defined in equation (2.8) by regression methods for bivariate data, and denote the estimated marginal conditional distribution of v_i as $\Phi_i(\hat{\boldsymbol{\vartheta}}) := F_v(\boldsymbol{z}_i'\hat{\boldsymbol{\vartheta}})$.

Step 2. Then $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0', \alpha_0)' \in \boldsymbol{\Theta} \subseteq \mathbb{R}^{p+1}$ can be estimated by standard quantile regression:

$$\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\alpha})' = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{x}_i'\boldsymbol{\beta} - \alpha\Phi_i(\hat{\boldsymbol{\vartheta}})), \tag{2.9}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the quantile regression loss function. We denote this estimator the quantile regression estimator for a model with endogenous misclassification (QREM).

In the first step, the variable Φ_i is estimated using a Gaussian marginal distribution and a Gaussian copula. These choices are convenient in practice and the first stage could be extended to include nonparametric alternatives (e.g., Chen, Fan, and Tsyrennikov, 2006; Han and Vytlacil, 2017; Han and Lee, 2019). Cavanagh and Sherman (1998) considered a class of semiparametric models and a rank estimator for (2.8). Although nonparametric methods are of interest, they have slow convergence rates and might not be applied to models with bivariate (partially observed) binary data. Given these considerations, the first step considers the partial observability model of Poirier (1980) and Nguimkeu, Denteh, and Tchernis (2019). Let

$$\Pr(d_i = 1|\boldsymbol{z}_i, \boldsymbol{w}_i) = \Pr(v_i \geq -\boldsymbol{z}_i'\boldsymbol{\vartheta}, \epsilon_i \geq -\boldsymbol{w}_i'\boldsymbol{\gamma}) = F_{v,\epsilon}(\boldsymbol{z}_i'\boldsymbol{\vartheta}, \boldsymbol{w}_i'\boldsymbol{\gamma}; \rho) = P_i(\boldsymbol{\vartheta}, \boldsymbol{\gamma}, \rho), \tag{2.10}$$

where ρ denotes the correlation between v_i and ϵ_i. The parameters $(\boldsymbol{\vartheta}', \boldsymbol{\gamma}', \rho)$ of the joint distribution function can be estimated by maximum likelihood (ML), considering the following log-likelihood function:

$$\ell_n(\boldsymbol{\vartheta}, \boldsymbol{\gamma}, \rho) = \frac{1}{n}\sum_{i=1}^{n}\left\{d_i \log(P_i(\boldsymbol{\vartheta}, \boldsymbol{\gamma}, \rho)) + (1-d_i)(1-\log(P_i(\boldsymbol{\vartheta}, \boldsymbol{\gamma}, \rho)))\right\}. \tag{2.11}$$

Under A2, the ML estimator defined as,

$$\{\hat{\boldsymbol{\vartheta}}, \hat{\boldsymbol{\gamma}}, \hat{\rho}\} = \underset{\boldsymbol{\vartheta},\boldsymbol{\gamma},\rho \in \boldsymbol{\Upsilon}\times\boldsymbol{\Gamma}\times\mathcal{K}}{\operatorname{argmax}} \ell_n(\boldsymbol{\vartheta}, \boldsymbol{\gamma}, \rho), \tag{2.12}$$

is consistent and asymptotically normal. In the second step, we employ quantile regression using the estimated marginal $\Phi_i(\hat{\boldsymbol{\vartheta}})$. The optimization problem formulated in (2.9) can accommodate weights, semiparametric estimation, and a penalty for high-dimensional models (Koenker, 2005). The solution can be obtained easily by computation methods developed in the R package `quantreg`.

Naturally, the procedure has advantages and disadvantages when applied to misclassification problems. Although the second step employs standard quantile regression methods, the asymptotic distribution of the QREM estimator derived in Theorem 2 below is complicated and it creates challenges for inference. Motivated by this limitation, the next section proposes a weighted bootstrap approach. Please see Section S.2 in the online appendix for an alternative estimator that allows the probability of partipation to be misspecified.

2.4. **Bootstrap Estimation.** The weighted bootstrap approach considered in this section is similar to the one previously employed in Chernozhukov, Fernández-Val, and Kowalski (2015) in cross-sectional models with control variables. The procedure, which was introduced by Jin, Ying, and Wei (2001) and further investigated by Ma and Kosorok (2005), suggests perturbing the objective function using independent draws from a non-negative distribution. It works under fairly general conditions and it can accommodate discrete regressors.

Let $\{\omega_i\}$ be a sequence of weights with mean 1 and variance 1. Using the bootstrap weights, we obtain the bootstrap estimator as follows:

$$\check{\boldsymbol{\theta}} = (\check{\boldsymbol{\beta}}', \check{\alpha})' = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n} \omega_i \rho_\tau(y_i - \boldsymbol{x}_i'\boldsymbol{\beta} - \alpha\Phi_i(\check{\boldsymbol{\vartheta}})), \tag{2.13}$$

where $\Phi_i(\check{\boldsymbol{\vartheta}})$ is estimated by the weighted bootstrap.

The procedure is implemented as follows. First, we draw a sample of weights $(\omega_1, \omega_2, \ldots, \omega_n)$. Using the multinomial version of the weighted bootstrap, and considering the sampled weights, we estimate the joint distribution and obtain the marginal distribution $\Phi_i(\check{\boldsymbol{\vartheta}})$. Second, we obtain $\check{\boldsymbol{\theta}}$ as in (2.13). We repeat the previous steps B times. Given a bootstrap sample $\{\check{\boldsymbol{\theta}}_b\}_{b=1}^B$, we obtain confidence intervals that are asymptotically valid, as demonstrated in Theorem 3 below. Let $G_j(a/2)$ and $G_j(1-a/2)$ be the $(a/2)$-th quantile and $(1-a/2)$-th quantile of the bootstrap distribution of $\sqrt{n}(\check{\theta}_j - \hat{\theta}_j)$ for $j = 1, 2, \ldots, p+1$. We obtain asymptotically valid $100(1-a)\%$ confidence intervals for $\theta_1 = \alpha_0$ by $[\hat{\alpha} - n^{-1/2}G_1(1-a/2), \hat{\alpha} - n^{-1/2}G_1(a/2)]$.

Remark 2. Standard choices for the weight distribution are the exponential and multinomial distributions. We recommend using multinomial weights in practice, because it is convenient for the parametric first step with dichotomous variables. When ω_i is a multinomial weight with probability n^{-1}, ω_i denotes the number of times that cross-sectional unit i is redrawn, and then (2.13) can be viewed as the cross-sectional pairs bootstrap estimator (Chatterjee and Bose, 2005). We adopt the recommendation in Section 5.

3. Theoretical properties

This section investigates the large sample properties of the proposed estimator. First, we establish consistency and asymptotic normality of the estimator. We then demonstrate the validity of the bootstrap estimator.

We consider the following assumptions:

A4. For each $\phi > 0$,

$$\inf_{\|\boldsymbol{\theta}\|_1 = \phi} \mathrm{E}\left[\int_0^{\boldsymbol{x}_i'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+(\alpha-\alpha_0)\Phi_i(\boldsymbol{\vartheta})} \left(F_u(s|\boldsymbol{x}_i, \boldsymbol{z}_i) - \tau\right) \mathrm{d}s\right] = \epsilon_\phi > 0,$$

where $F_u := F_{u_i|\boldsymbol{x}_i,\boldsymbol{z}_i}$ is the cumulative distribution function of $u_i = y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_0 - \alpha_0\Phi_i(\boldsymbol{\vartheta}_0)$ conditional on $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$, and $\Phi_i(\boldsymbol{\vartheta}) := \Phi(\boldsymbol{z}_i'\boldsymbol{\vartheta})$.

A5. The vector $\boldsymbol{h}_i = (\boldsymbol{x}_i', \boldsymbol{z}_i', \boldsymbol{w}_i')'$ satisfies $\max_{1 \le i \le n} \|\boldsymbol{h}_i\| < M < \infty$ a.s.

A6. Let $\psi(d_i, \boldsymbol{z}_i, \boldsymbol{w}_i) =: \psi(\boldsymbol{L}_i)$ be an influence function corresponding to (2.11), with $\mathrm{E}[\psi(\boldsymbol{L}_i)] = 0$ and $\mathrm{E}[\psi(\boldsymbol{L}_i)\psi(\boldsymbol{L}_i)'] \le K_\psi < \infty$. Then (a) $\hat{\boldsymbol{\vartheta}} \xrightarrow{p} \boldsymbol{\vartheta}_0$ and (b) the estimator $\hat{\boldsymbol{\vartheta}}$ admits an asymptotically linear representation,

$$\sqrt{n}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(\boldsymbol{z}_i) + o_p(1),$$

where $\mathrm{E}[\psi(\boldsymbol{z}_i)] = 0$ and $\mathrm{E}[\psi(\boldsymbol{z}_i)\psi(\boldsymbol{z}_i)'] = \boldsymbol{\Omega}$, a positive definite matrix.

The first two conditions are similar to the ones used in the quantile regression literature (e.g., Koenker, 2005). Assumption A4 is an identification condition and it is sufficient for consistency, without requiring F_u to have a derivative. The condition guarantees the convexity of the objective function, and thus, the uniqueness of $\boldsymbol{\theta}_0$ as $n \to \infty$. Assumption A5 is common to impose appropriate moment conditions on the covariates in the first and second stages, and it can be relaxed by imposing stochastic bounds based on moment conditions. The last condition A6 is used for the expansion of estimators with sufficiently smooth objective functions (Newey and McFadden, 1994) and quantile based estimators (Chernozhukov, Fernández-Val, and Kowalski, 2015), although here the first stage does not require a smoothness condition. The first part of A6 is needed for the consistency of the estimator of $\mathrm{E}\left[d_i^* | \boldsymbol{x}_i, \boldsymbol{z}_i\right]$, obtained from (2.11) under Assumption A2. The first stage is similar to the method employed by Nguimkeu, Denteh, and Tchernis (2019), and therefore, the result below uses a similar condition. The second part of the condition is similar to Assumption 4 in Chernozhukov, Fernández-Val, and Kowalski (2015). The condition has to be verified for the estimators employed in the first stage, and it is satisfied under condition A2.

The following result establishes the consistency of the two-step estimator:

Theorem 1. *Under Assumptions A1-A6.a, as $n \to \infty$, the estimator $\hat{\boldsymbol{\theta}}$ defined in equation* (2.9) *is a consistent estimator of $\boldsymbol{\theta}_0$.*

We consider the following additional conditions, which are similar to A1 and A2 in Ma and Koenker (2006) and A1 and A5 in Chen, Galvao, and Song (2021):

A7. The conditional cumulative distribution function of the error term u_i, $F_{u_i|\boldsymbol{x}_i,\boldsymbol{z}_i}$, has continuous derivative $f_{u_i|\boldsymbol{x}_i,\boldsymbol{z}_i}$ that is uniformly bounded away from 0 and ∞ at the τ-th conditional quantile, for $1 \leq i \leq n$.

A8. Let $\boldsymbol{X}_i(\boldsymbol{\vartheta}) := (\boldsymbol{x}_i', \Phi_i(\boldsymbol{\vartheta}))'$ and $\dot{\boldsymbol{X}}_i(\boldsymbol{\vartheta}) := \partial_\Phi \boldsymbol{X}_i(\boldsymbol{\vartheta})$. There exist positive definite matrices $\boldsymbol{D}_j$ for $j \in \{0,1,2,3\}$ defined as,

$$\boldsymbol{D}_0 = \tau(1-\tau) \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i(\boldsymbol{\vartheta}) \boldsymbol{X}_i(\boldsymbol{\vartheta})', \quad \boldsymbol{D}_1 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f_u(0|\boldsymbol{x}_i, \boldsymbol{z}_i) \boldsymbol{X}_i(\boldsymbol{\vartheta}) \boldsymbol{X}_i(\boldsymbol{\vartheta})',$$

$$\boldsymbol{D}_2 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f_u(0|\boldsymbol{x}_i, \boldsymbol{z}_i) \boldsymbol{X}_i(\boldsymbol{\vartheta}) \dot{\boldsymbol{X}}_i(\boldsymbol{\vartheta})' \boldsymbol{\theta} \nabla_{\boldsymbol{\vartheta}} \Phi_i(\boldsymbol{\vartheta}), \quad \boldsymbol{D}_3 = \lim_{n\to\infty} n \mathrm{E}\left[(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}})(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' \right],$$

where $\hat{\boldsymbol{\theta}}$ is the estimator in (2.9) and $\tilde{\boldsymbol{\theta}}$ is an estimator that uses $\Phi_i(\boldsymbol{\vartheta})$ instead of $\Phi_i(\hat{\boldsymbol{\vartheta}})$ in (2.9).

Under these conditions, we obtain the asymptotic distribution of the estimator:

Theorem 2. *Under Assumptions A1-A8, as $n \to \infty$,*

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{D}_1^{-1}[\boldsymbol{D}_0 + \boldsymbol{D}_2 \boldsymbol{\Omega} \boldsymbol{D}_2' - \boldsymbol{D}_3]\boldsymbol{D}_1^{-1}).$$

The asymptotic variance of the proposed estimator is given in Theorem 2. Although the literature on quantile regression offers a variety of inference methods, the complicated form of the asymptotic covariance matrix represents a major challenge for carrying out analytical inference. The asymptotic variance requires estimation of nuisance parameters by non-parametric methods, which is standard in quantile regression although it can introduce size distortions (He, 2018). In the case of misclassification, we also need to approximate and estimate partial derivatives of distribution functions. Motivated by these limitations, we propose a weighted bootstrap procedure in (2.13) and we establish its validity in Theorem 3.

Consider the following assumptions:

B1. Let $(\omega_1, \omega_2, \ldots, \omega_n)$ be a vector of independent and identically distributed weights from a non-negative distribution with $\mathrm{E}[\omega_i] = 1$ and $\mathrm{Var}[\omega_i] = 1$. The weights are independent of the variables $\{(y_i, d_i, \boldsymbol{x}_i', \boldsymbol{z}_i', \boldsymbol{w}_i')\}$ for $1 \le i \le n$.

B2. Let $o_{p^\star}$ denote an stochastic order interpreted as conditional on the observed sample. Assume that $\|\check{\boldsymbol{\vartheta}} - \hat{\boldsymbol{\vartheta}}\| = o_{p^\star}(1)$ and the bootstrap estimator $\check{\boldsymbol{\vartheta}}$ admits an asymptotically linear representation,

$$\sqrt{n}(\check{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \omega_i \psi(\boldsymbol{z}_i) + o_p(1),$$

where $(\omega_1, \omega_2, \ldots, \omega_n)$ satisfies the conditions in B1, $\mathrm{E}[\psi(\boldsymbol{z}_i)] = 0$, and $\mathrm{E}[\psi(\boldsymbol{z}_i)\psi(\boldsymbol{z}_i)'] = \boldsymbol{\Omega}$. Define $\Phi_i(\check{\boldsymbol{\vartheta}}) := \Phi(\boldsymbol{z}_i'\check{\boldsymbol{\vartheta}})$ as the bootstrap estimator for the conditional marginal distribution.

Assumption B1 is standard and it describes the weights used to perturb the objective function. Exponential weights have been adopted in a variety of models (e.g., Peng and Huang, 2008; Chernozhukov, Fernández-Val, and Kowalski, 2015, among others), but we consider ω_i to be multinomial drawn with probability n^{-1}, as discussed in Remark 2. Assumption B2 is a high level assumption that requires the consistency and asymptotic normality of the first stage estimator. This is an important assumption, since in some situations, the consistency of the bootstrap estimator does not imply the consistency of the bootstrap estimator of second moments (Hahn and Liao, 2021).

Theorem 3. *Under the conditions of Theorem 2 and Assumptions B1 and B2, the bootstrap estimator defined in* (2.13)*, as* $n \to \infty$ *and conditional on the observed sample,*

$$\sqrt{n}(\check{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) \Rightarrow B^{\star}(\tau),$$

where $B^{\star}(\tau)$ *is a Gaussian process with mean zero and covariance matrix* $\boldsymbol{D}_1^{-1}[\boldsymbol{D}_0 + \boldsymbol{D}_2 \boldsymbol{\Omega} \boldsymbol{D}_2' - \boldsymbol{D}_3]\boldsymbol{D}_1^{-1}$ *as in Theorem 2.*

The result of Theorem 3 implies that the weighted bootstrap procedure is consistent as an estimator of the asymptotic distribution of the two-step estimator (2.9).

4. Simulation Study

This section presents the results of a simulation study designed to evaluate the performance of the proposed estimator in finite samples. We consider a data generating process that follows closely the designs in Nguimkeu, Denteh, and Tchernis (2019) and Ura (2021). Observations for the dependent variable y_i for $i = 1, 2, \ldots, n$ are generated according to the following model:

$$y_i = \beta_0 + \beta_1 x_i + \left[\exp\left(F_u(u_i) - 0.5\right) - 1.2\right] d_i^* + u_i, \tag{4.1}$$

where F_u is the cumulative distribution of u_i and x_i is an i.i.d. random variable distributed as $\mathcal{U}[0, 1]$. The true value of $\alpha_0 = \exp(\tau - 0.5) - 1.2$ changes across quantiles as in Ura (2021). When $\tau = 0.5$, the parameter of interest $\alpha_0 = -0.2$ as in Nguimkeu, Denteh, and Tchernis (2019).

The latent indicator variable is $d_i^* = 1\{\vartheta_0 + \vartheta_1 z_i + v_i \geq 0\}$, and the misclassification indicator is $m_i = 1\{\gamma_0 + \gamma_1 w_i + \epsilon_i \geq c\}$, where the parameter c is set to control the proportion of false negatives, that is, observations with $d_i^* = 1$ and $d_i = 0$. The instrument z_i and additional regressor w_i are i.i.d. random variables distributed as $\mathcal{N}(0, 1)$ as in Nguimkeu, Denteh, and Tchernis (2019). The observed binary regressor variable is generated as follows:

$$d_i = d_i^* \, m_i = 1\{\vartheta_0 + \vartheta_1 z_i + v_i \geq 0, \gamma_0 + \gamma_1 w_i + \epsilon_i \geq c\}. \tag{4.2}$$

In all variants of the models, $\vartheta_0 = 0.1$, $\beta_0 = \beta_1 = \vartheta_1 = 1$, $\gamma_1 = 2$, and $\gamma_0 = 0.01$.

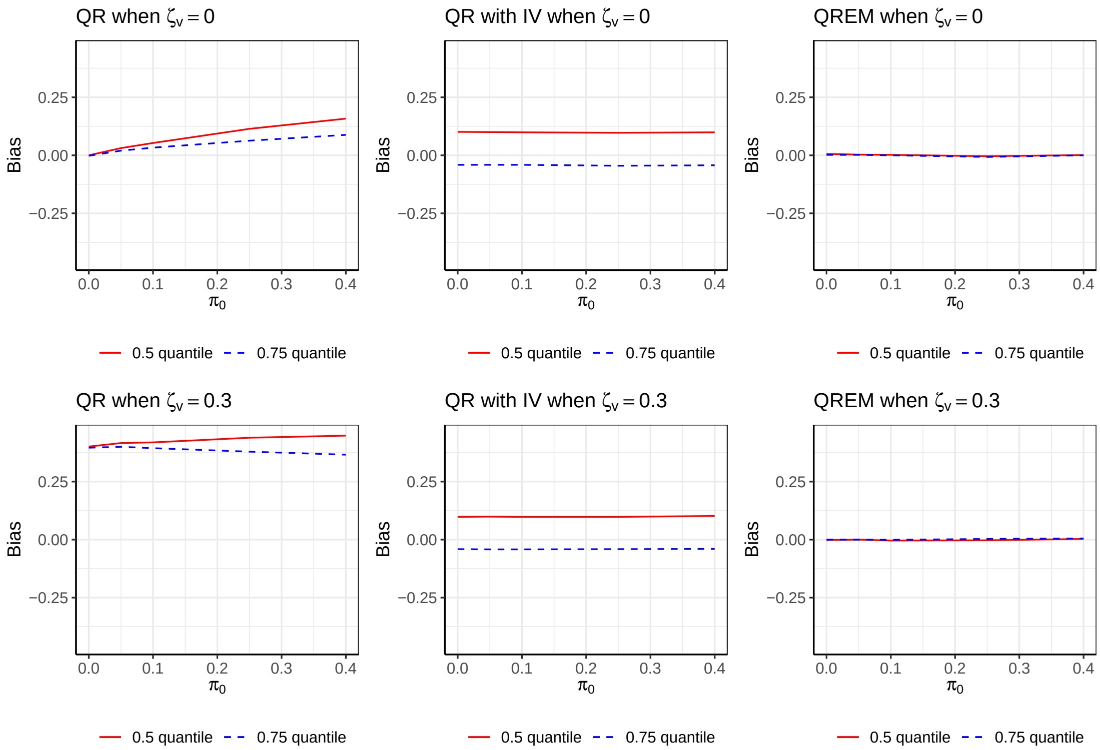

FIGURE 4.1. *Bias for α_0 when the marginal distribution is Normal. QR denotes quantile regression, QR with IV uses an instrumental variable as regressor, and QREM is the two-step estimator.*

Let $\pi_0 = \Pr(d_i = 0|d_i^* = 1)$ and $\pi_1 = \Pr(d_i = 1|d_i^* = 0)$. We set c in (4.2) to generate 0%, 5%, 10%, 25%, and 40% of false negatives in Figure 4.1 and Figure 4.2 with one-sided misreporting. We consider the case of both $\pi_0 > 0$ and $\pi_1 > 0$ in the online appendix.

The errors in equations (4.1) and (4.2) are i.i.d. random variables distributed from a trivariate distribution $F_{u,v,\epsilon}$ with mean zero and covariance:

$$\Sigma = \begin{pmatrix} \sigma^2 & \sigma\zeta_v & \sigma\zeta_\epsilon \\ \sigma\zeta_v & 1 & \rho \\ \sigma\zeta_\epsilon & \rho & 1 \end{pmatrix} = \begin{pmatrix} 1 & \zeta_v & \zeta_\epsilon \\ \zeta_v & 1 & 0.3 \\ \zeta_\epsilon & 0.3 & 1 \end{pmatrix}, \tag{4.3}$$

where $\sigma = 1$ and $\rho = 0.3$ in all variants of the experiments considered in this section. Note that the case of exogenous misreporting and exogenous participation is obtained by setting $\zeta_\epsilon = \zeta_v = 0$. In this case, the quantile regression (QR) estimator is consistent for α_0, and it expected to perform well in finite samples. To examine the performance of quantile

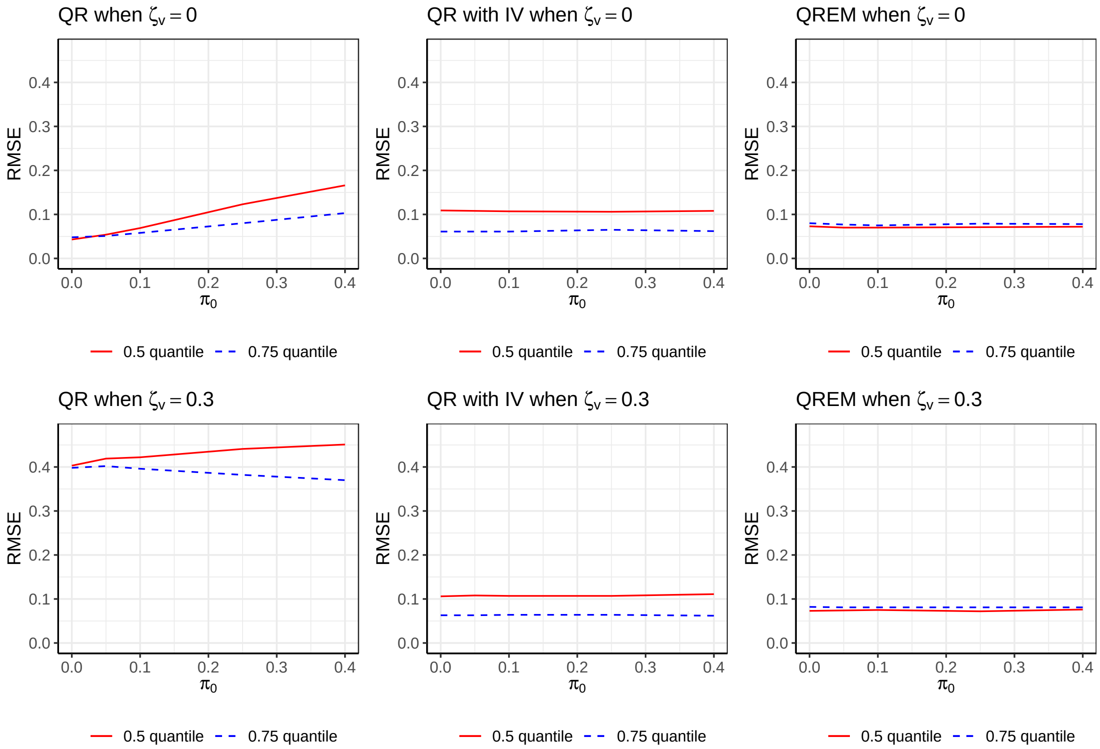

FIGURE 4.2. *Root mean square error (RMSE) for α_0 when the marginal distribution is Normal. QR denotes quantile regression, QR with IV uses an instrumental variable as regressor, and QREM is the two-step estimator.*

regression estimators in the case of endogenous misreporting and participation, we follow Nguimkeu, Denteh, and Tchernis (2019) and set $\zeta_v = 0.3$ and $\zeta_\epsilon = 0.2$. Finally, we consider different trivariate distributions. We first consider joint normality, and then, we report results obtained by assuming that $F_{u,v,\epsilon}$ is t_3 (i.e., t-student distribution with 3 degrees of freedom), and χ^2_3 (i.e., χ^2 distribution with 3 degrees of freedom) centered to have zero mean.

Figures 4.1 and 4.2 present bias and root mean square error (RMSE) of the proposed estimators in comparison with QR and the estimator that uses an IV. Following Lemma 1, we consider QR using $1\{z_i \geq 0\}$ instead of d_i as a regressor (IV). It is important to bear in mind that these estimators do not allow for endogenous misclassification and could be biased in finite samples. Lastly, the figures show results obtained by the QREM estimator introduced in Section 2.3. The bias and RMSE are for the parameter of interest α_0 at $\tau \in \{0.5, 0.75\}$, obtained from 1000 samples of size 5000.

The top panels of the figures present the case of no endogenous participation, while the lower panels present results from allowing participation to be endogenous. The left panel in Figure 4.1 provides evidence of the bias of QR under exogenous misclassification and participation. As expected, the QR is unbiased if the proportion of false negatives is zero (i.e., $\pi_0 = 0$). The performance of QR and the IV estimator deteriorate when $\pi_0 > 0$, reaching bias that range from 20% to 50% at $\pi_0 = 0.4$ and $\tau = 0.5$. In the case of endogenous participation and misclassification, the proposed estimator have almost zero bias and the lowest MSE. The results across quantiles lead to similar conclusions.

5. An Empirical Application

Identifying the effect of public assistance on different economic outcomes is plagued with challenges ranging from selection to endogenous misclassification (Kreider, Pepper, Gundersen, and Jolliffe, 2012; Dahl, Kostøl, and Mogstad, 2014; Hartley, Lamarche, and Ziliak, 2022). In this section, we consider data from the Panel Study of Income Dynamics (PSID) to study the effect of Aid to Families with Dependent Children (AFDC), a major welfare program implemented before 1996. We apply our quantile regression approach to estimate the effect of welfare participation as a child on family income as an adult. Our findings are consistent with theory as they show that the estimator that uses instrumental variables provides an upper bound to results obtained by the proposed estimator. When we address the possibility of endogenous misclassification, maternal participation in the program has a negative effect across the conditional income distribution of adult daughters. Our findings reveal that standard approaches underestimate the effect of welfare participation, and that the intergenerational impact is significantly larger among low-income daughters.

5.1. **Data.** We use data from the PSID, which is an annual survey often used in intergenerational studies in the U.S. The sample includes 2038 daughters who were between the ages of 12 and 18 when their mother received cash welfare. These daughters have been followed into adulthood and the PSID provides information on adult family income and other characteristics. We focus on linked mother-daughter pairs over PSID survey years from 1968 to 1996, before a major reform replaced Aid to Families with Dependent Children (AFDC) with a new program. Our analysis is then restricted to the years before welfare reform. The sample includes families from the nationally representative Survey Research Center (SRC)

	Equations:	
	Participation	Misclassification
AFDC benefit standard, average	1.864	
	(0.518)	
AFDC benefit standard, maximum	-1.630	
	(0.472)	
PSID Reporting Rates		3.785
		(2.004)
Earnings below 130% FPL	1.268	-0.254
	(0.427)	(0.893)

TABLE 5.1. *Partial observability probit estimates. Column (1) shows results for the probability of participation and column (2) shows results for the probability of true reporting. Standard errors are in parenthesis. FPL denotes federal poverty line.*

subsample and the Survey of Economic Opportunity (SEO) subsample, which oversamples low-income and minority families.

Family income of the adult daughter is the focus of our investigation. Daughter's adult income includes total family taxable income and welfare cash transfers of the head, spouse and other family members. A daughter is considered an adult at first childbirth or when establishing a new family unit if she is older than 14 years old. We construct the average of annual income (in 2016 dollars) across the daughter's adult years from the age of 19 through age 27. We measure maternal participation during the child's critical exposure period of 12-18 years of age, and our data reveals that approximately 30 percent of children grow up under welfare.

5.2. **Empirical results.** We estimate the intergenerational effect of mother's participation on AFDC on daughter's adult income, say α_0 as defined in equation (2.2), considering the quantile regression estimator (2.9). We proceed in two steps. We first estimate equation (2.8) considering a partial observability probit model. Then, we estimate the parameter of interest by quantile regression considering the predicted probabilities obtained in the first step and a vector of independent variables $\boldsymbol{x}_i$ that includes mother's education, daughter's marital status, an indicator variable for race of the daughter, and the logarithm of mother's financial income.

Table 5.1 presents the results of the first step. To address the possible endogeneity of welfare participation, we use instrumental variables that are constructed following closely

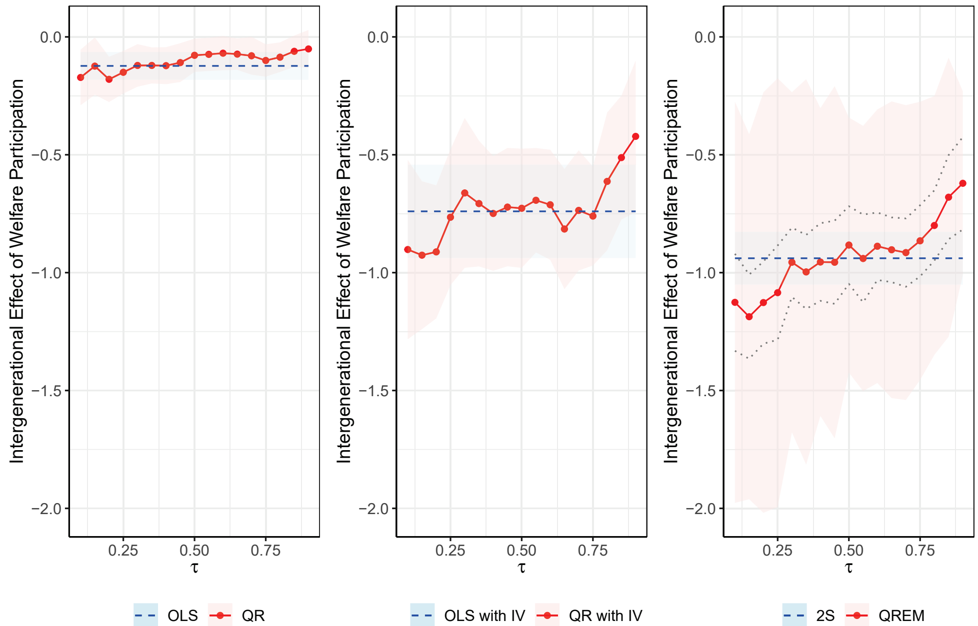

FIGURE 5.1. *The effect of mother's welfare participation on daughter's income. OLS denotes ordinary least squares, QR denotes quantile regression, 2S is a two-step estimator for the conditional mean model, and EM is the proposed two-step estimator for a model with endogeneous misclassification. The dotted intervals represent uncorrected point-wise confidence intervals.*

Hartley, Lamarche, and Ziliak (2022). The instruments are based on the state-level AFDC maximum benefit guarantee in the years when the child is 12-18 years of age. We use the average and maximum values. This policy is determined by state legislatures and affects the participation decision of the mother via her welfare status, as opposed to her poverty status and adult daughter poverty status. To address endogenous misclassification, we use average PSID reporting rates for dollar amount in transfers and number of cases for AFDC corresponding to the years when the daughter was between 12 and 18 years old. We also include an indicator variable for poverty status. All coefficients have the expected signs and are statistically significant (with the exception of earnings below the federal poverty line in the misclassification equation).

Figure 5.1 presents the main results. The figure reports the intergenerational effect of AFDC participation on daughter's income estimated by different methods. The horizontal dashed lines represent results obtained by OLS and the continuous lines with dots show results obtained by QR. The panel on the left show OLS and QR results considering an indicator variable for reported participation in AFDC (i.e., d_i) as a regressor. Following Lemma 1, the panel in the middle uses the AFDC maximum benefit (i.e., $z_{2,i}$) as a regressor. The right panel of Figure 5.1 presents results using the estimator proposed by Nguimkeu, Denteh, and Tchernis (2019) (denoted by 2S) and the proposed QREM estimator in (2.9). The figures present 95% point-wise confidence intervals. The last panel presents bootstrap confidence intervals (gray area) as well as a confidence interval obtained by estimating the asymptotic variance using kernel methods (dashed lines) without adjusting for the estimation error in the first step.

The findings presented in Figure 5.1 illustrate the importance of considering mother's possible endogenous misreporting. QR results are expected to be small and biased towards zero. On the other hand, QR with instrumental variables as in Lemma 1 provides an upper bound across τ, and consistent with expectations, the QREM estimator suggests a smaller intergenerational coefficient than the estimator that includes an IV as regressor. Welfare participation of the mother during her daughter's childhood is associated with a 68% reduction on daughter's adult income at the 0.1 quantile and a 46% reduction at the 0.9 quantile. These effects are underestimated by 8% to 53% if the intergenerational parameter α_0 is estimated by existing methods.

6. Conclusion and discussion

This paper investigates the estimation of a quantile regression model with a misclassified binary regressor. We propose a two-step approach and show that the estimator is consistent and asymptotically normal. The identification of the model relies on a parametric first stage and the use of additional measurements including instrumental variables. We also propose a bootstrap approach and establish the validity of the estimator. Considering data from the PSID, we estimate a quantile intergenerational parameter to study how mother's participation on welfare during her daughter's childhood affects daughter's adult income. We find that existing methods underestimate the intergenerational impact of welfare participation.

A number of recent papers have contributed to a deeper understanding of the challenges of employing household survey data (e.g., Meyer, Mok, and Sullivan, 2015; Bollinger, Hirsch, Hokayem, and Ziliak, 2019). While the use of administrative data has been an important tool to assess issues of item non-response and misreporting, the availability of administrative data among social scientists is not widespread. Although the paper examines the sensitivity of the proposed estimator to departures of functional form assumptions, developing more flexible approaches is naturally critical. To the best of our knowledge, this paper is the first attempt to investigate point estimation of heterogeneous effects without assuming exogenous or random misreporting. The estimator is simple to compute, and it provides scientists using survey data the possibility of contributing to the critical policy debate on the effects of social programs.

Appendix A. Proof of main results

Remarks on notation and definitions: Throughout the appendix, we define $\boldsymbol{\theta} = (\boldsymbol{\beta}', \alpha)'$, but we suppress the dependency on τ for notational simplicity. The proofs below refer to Knight's (1998) identity: $\rho_\tau(u-v) - \rho_\tau(u) = -v\psi_\tau(u) + \int_0^v (I(u \leq s) - I(u \leq 0))\mathrm{d}s$, where $\rho_\tau = u(\tau - I(u < 0))$ is the quantile regression check function and $\psi_\tau(u) = \tau - I(u < 0)$ is the associated score function.

It is convenient to introduce additional notation. Let $\Phi_i(\boldsymbol{\vartheta}) := \Phi_i(\boldsymbol{z}_i'\boldsymbol{\vartheta})$, $\boldsymbol{X}_i(\boldsymbol{\vartheta}) := (\boldsymbol{x}_i', \Phi_i(\boldsymbol{\vartheta}))'$, and $\dot{\boldsymbol{X}}_i(\boldsymbol{\vartheta}) := \partial_\Phi \boldsymbol{X}_i(\boldsymbol{\vartheta})$. Moreover, let $\boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})$ denote the vector $\boldsymbol{X}_i(\boldsymbol{\vartheta})$ evaluated at the estimated values. Consider the unfeasible estimator,

$$\tilde{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmin}} \left\{ \frac{1}{n} \sum_{i=1}^n \rho_\tau(y_i - \boldsymbol{\theta}'\boldsymbol{X}_i(\boldsymbol{\vartheta}_0)) \right\},$$

and the feasible version of $\tilde{\boldsymbol{\theta}}$ as in (2.9):

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmin}} \left\{ \frac{1}{n} \sum_{i=1}^n \rho_\tau(y_i - \boldsymbol{\theta}'\boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})) \right\}.$$

Proof of Theorem 1. Write $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = (\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}) + (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$. Consistency is established in two steps. We first show that $\tilde{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$, and in the second part of the proof, we demonstrate that $\hat{\boldsymbol{\theta}} \xrightarrow{p} \tilde{\boldsymbol{\theta}}$.

Let $\hat{\boldsymbol{\theta}}$ be the minimizer of the normalized objective function

$$\mathbb{M}_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \rho_\tau\left(y_i - \boldsymbol{\theta}'\boldsymbol{X}_i(\boldsymbol{\vartheta}_0)\right) = \frac{1}{n} \sum_{i=1}^n \rho_\tau\left(u_i - (\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\boldsymbol{X}_i(\boldsymbol{\vartheta}_0)\right),$$

where $u_i = y_i - \boldsymbol{\theta}_0'\boldsymbol{X}_i(\boldsymbol{\vartheta}_0)$, and $(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\boldsymbol{X}_i(\boldsymbol{\vartheta}_0) = \boldsymbol{x}_i'(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \Phi_i(\boldsymbol{\vartheta}_0)(\alpha - \alpha_0)$. Let $\Delta_n(\boldsymbol{\theta}) = \mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n(\boldsymbol{\theta}_0)$, that is,

$$\Delta_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \left\{ \rho_\tau\left(u_i - (\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\boldsymbol{X}_i(\boldsymbol{\vartheta}_0)\right) - \rho_\tau\left(u_i\right) \right\}.$$

By Knight's (1998) identity, $\Delta_n(\boldsymbol{\theta}) = \mathbb{V}_n^{(1)} + \mathbb{V}_n^{(2)}$, where,

$$\mathbb{V}_n^{(1)}(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{i=1}^{n}\left\{\boldsymbol{x}_i'(\boldsymbol{\beta}-\boldsymbol{\beta}_0) + \Phi_i(\boldsymbol{\vartheta}_0)(\alpha-\alpha_0)\right\}\psi_\tau(u_i),$$

$$\mathbb{V}_n^{(2)}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\boldsymbol{x}_i'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+\Phi_i(\boldsymbol{\vartheta}_0)(\alpha-\alpha_0)} \left(I(u_i \le s) - I(u_i \le 0)\right) \mathrm{d}s,$$

Note that under Assumptions A1 and A3, $\mathrm{E}\left[\mathbb{V}_n^{(1)}(\boldsymbol{\theta})\right] = 0$, because the quantile of u_i conditional on $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ is equal to the conditional quantile of u_i^* conditional on $\boldsymbol{x}_i$ and d_i^*.

We first show the consistency of $\tilde{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}_0$. For each $\phi > 0$, define the ball $\mathcal{B}(\phi) := \{\boldsymbol{\theta} : \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_1 \le \phi\}$ and the boundary $\partial\mathcal{B}(\phi) := \{\boldsymbol{\theta} : \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_1 = \phi\}$. For each $\boldsymbol{\theta} \notin \mathcal{B}(\phi)$, define $\bar{\boldsymbol{\theta}} = r\boldsymbol{\theta} + (1-r)\boldsymbol{\theta}_0$ where $r = \phi/\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_1$. By construction, $r \in (0,1)$, and $\bar{\boldsymbol{\theta}}$ is in the boundary set of $\mathcal{B}(\phi)$.

By the convexity of $\mathbb{M}_n(\boldsymbol{\theta})$,

$$r\mathbb{M}_n(\boldsymbol{\theta}) + (1-r)\mathbb{M}_n(\boldsymbol{\theta}_0) \ge \mathbb{M}_n(\bar{\boldsymbol{\theta}}),$$

or, $r\big(\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n(\boldsymbol{\theta}_0)\big) \ge \mathbb{M}_n(\bar{\boldsymbol{\theta}}) - \mathbb{M}_n(\boldsymbol{\theta}_0) = \mathrm{E}\left[\Delta_n(\bar{\boldsymbol{\theta}})\right] + \big(\Delta_n(\bar{\boldsymbol{\theta}}) - \mathrm{E}\left[\Delta_n(\bar{\boldsymbol{\theta}})\right]\big)$. Under Assumption A4, we obtain,

$$\mathrm{E}\left[\Delta_n(\boldsymbol{\theta})\right] = \mathrm{E}\left[\int_0^{\boldsymbol{x}_i'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+\Phi_i(\boldsymbol{\vartheta})(\alpha-\alpha_0)} (F_u(s|\boldsymbol{x}_i,\boldsymbol{z}_i) - \tau)\mathrm{d}s\right] \ge \epsilon_\phi,$$

for some $\epsilon_\phi > 0$. Then,

$$r(\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n(\boldsymbol{\theta}_0)) = r\Delta_n(\boldsymbol{\theta}) \ge \epsilon_\phi + \left(\Delta_n(\bar{\boldsymbol{\theta}}) - \mathrm{E}\left[\Delta_n(\bar{\boldsymbol{\theta}})\right]\right).$$

By definition of $\tilde{\boldsymbol{\theta}}$ as the minimizer of $\frac{1}{n}\sum_{i=1}^{n}\mathbb{M}_n(\boldsymbol{\theta})$, we obtain the following inclusion relationships:

$$\left\{\|\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|_1 > \phi\right\} \subseteq \left\{\tilde{\boldsymbol{\theta}} \notin \mathcal{B}(\phi) \text{ and } \mathbb{M}_n(\tilde{\boldsymbol{\theta}}) \le \mathbb{M}_n(\boldsymbol{\theta}_0)\right\} \subseteq \left\{\sup_{\boldsymbol{\theta}\in\mathcal{B}(\phi)}\Big|\Delta_n(\boldsymbol{\theta}) - \mathrm{E}\left[\Delta_n(\boldsymbol{\theta})\right]\Big| \ge \epsilon_\phi\right\}.$$

Using standard arguments (i.e., derivations analogous to those in Theorem 1 in Lamarche and Parker, 2023), for any $\epsilon > 0$ and constants $D > 0$ and $C > 0$,

$$\mathrm{P}\left\{\sup_{\boldsymbol{\theta}\in\mathcal{B}(\phi)}|\Delta_n(\boldsymbol{\theta}) - \mathrm{E}\left[\Delta_n(\boldsymbol{\theta})\right]| > \epsilon\right\} \le 2C\exp\{-Dn\}.$$

Therefore, with probability tending to zero, $\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1 > \phi$, and thus, $\tilde{\boldsymbol{\theta}}$ is consistent for $\boldsymbol{\theta}_0$.

We now demonstrate the consistency of the feasible estimator $\hat{\boldsymbol{\theta}}$ for the unfeasible $\tilde{\boldsymbol{\theta}}$. Write the objective function of $\hat{\boldsymbol{\theta}}$ as follows:

$$\rho_\tau(y_i - \boldsymbol{\theta}'\boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})) = \rho_\tau(y_i - \boldsymbol{x}_i'\boldsymbol{\beta} - \Phi_i(\hat{\boldsymbol{\vartheta}})\alpha) = \rho_\tau(y_i - \boldsymbol{x}_i'\boldsymbol{\beta} - \Phi_i(\boldsymbol{\vartheta}_0)\alpha - (\Phi_i(\hat{\boldsymbol{\vartheta}}) - \Phi_i(\boldsymbol{\vartheta}_0))\alpha).$$

Using a version of Knight's identity, $|\rho_\tau(u - v) - \rho_\tau(u)| \le 3|v|$, we write,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \Big|\rho_\tau(y_i - \boldsymbol{\theta}'\boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})) - \rho_\tau(y_i - \boldsymbol{\theta}'\boldsymbol{X}_i(\boldsymbol{\vartheta}_0))\Big| &\le 3\frac{1}{n}\sum_{i=1}^n |(\Phi_i(\hat{\boldsymbol{\vartheta}}) - \Phi_i(\boldsymbol{\vartheta}_0))\alpha| \\
&\le 3|\alpha|\frac{1}{n}\sum_{i=1}^n |\Phi_i(\hat{\boldsymbol{\vartheta}}) - \Phi_i(\boldsymbol{\vartheta}_0)|.
\end{aligned}$$

Under Assumptions A2 and A6, expanding $\Phi(\hat{\boldsymbol{\vartheta}})$ close enough to $\boldsymbol{\vartheta}_0$, we obtain

$$\Phi(\hat{\boldsymbol{\vartheta}}) = \Phi(\boldsymbol{\vartheta}_0) + \Psi(\boldsymbol{\vartheta})(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) + e(\boldsymbol{z}, \hat{\gamma}),$$

where $\Psi(\boldsymbol{\vartheta}) = \nabla_{\boldsymbol{\vartheta}}\Phi(\boldsymbol{\vartheta}) = \varphi(\boldsymbol{z}'\boldsymbol{\vartheta})\boldsymbol{z}'$ and $e(\boldsymbol{z}, \hat{\gamma}) = (\Psi(\bar{\boldsymbol{\vartheta}}) - \Psi(\boldsymbol{\vartheta}_0))'(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0)$ for a mean value $\bar{\gamma}$.

It follows that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n |\Phi_i(\hat{\boldsymbol{\vartheta}}) - \Phi_i(\boldsymbol{\vartheta}_0)| &\le \frac{1}{n}\sum_{i=1}^n \Big(|\Psi_i(\boldsymbol{\vartheta})(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0)| + |e(\boldsymbol{z}_i, \hat{\gamma})|\Big) \\
&\le \|\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\| \ \frac{1}{n}\sum_{i=1}^n \|\Psi_i(\boldsymbol{\vartheta})\| + \frac{1}{n}\sum_{i=1}^n |e(\boldsymbol{z}_i, \hat{\gamma})| \\
&\le \|\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\| \ \frac{1}{n}\sum_{i=1}^n \|\Psi_i(\boldsymbol{\vartheta})\| + \|\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\| \ \frac{1}{n}\sum_{i=1}^n \|\Psi_i(\bar{\boldsymbol{\vartheta}}) - \Psi_i(\boldsymbol{\vartheta}_0)\| \\
&\le \|\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\| \ \max_{1 \le i \le n} \|\boldsymbol{z}_i\| + o_p(1),
\end{aligned}$$

by continuity of Ψ_i implied by A2. Under A5, it can be established that,

$$\frac{1}{n}\sum_{i=1}^n \Big|\rho_\tau(y_i - \boldsymbol{\theta}'\boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})) - \rho_\tau(y_i - \boldsymbol{\theta}'\boldsymbol{X}_i(\boldsymbol{\vartheta}_0)))\Big| \le M \ \|\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\| + o_p(1).$$

Therefore, as $n \to \infty$, $\hat{\boldsymbol{\theta}}$ converges in probability to $\tilde{\boldsymbol{\theta}}$ under Assumption A6.a.

Combining the two previous results completes the proof. $\square$

The following definitions are used in Lemma 2 and Theorem 2. Consider,

$$\begin{aligned}
\mathbb{H}_n(\boldsymbol{\theta},\boldsymbol{\vartheta}) &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i(\boldsymbol{\vartheta})\psi_\tau\left(y_i-\boldsymbol{\theta}'\boldsymbol{X}_i(\boldsymbol{\vartheta})\right),\\
H_n(\boldsymbol{\theta},\boldsymbol{\vartheta}) &= \mathrm{E}\left[\mathbb{H}_n(\boldsymbol{\theta},\boldsymbol{\vartheta})\right] = \mathrm{E}\left[\boldsymbol{X}_i(\boldsymbol{\vartheta})(\tau - F_u((\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\boldsymbol{X}_i(\boldsymbol{\vartheta})|\boldsymbol{x}_i,\boldsymbol{z}_i)\right],\\
\mathbb{K}_n(\boldsymbol{\theta},\boldsymbol{\vartheta}) &= \frac{1}{n}\sum_{i=1}^{n} f_u(0|\boldsymbol{x}_i,\boldsymbol{z}_i)\boldsymbol{X}_i(\boldsymbol{\vartheta})\dot{\boldsymbol{X}}_i(\boldsymbol{\vartheta})'\boldsymbol{\theta}\Psi_i(\boldsymbol{\vartheta}).
\end{aligned}$$

Lemma 2. *Under Assumptions A1-A8, if* $\sqrt{n}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)=O_p(1)$ *and*

$$\sqrt{n}(\Phi_i(\hat{\boldsymbol{\vartheta}})-\Phi_i(\boldsymbol{\vartheta}_0)) = \Psi_i(\boldsymbol{\vartheta})\sqrt{n}(\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0)+o_p(1),$$

for $\Psi_i(\boldsymbol{\vartheta}) = \nabla_{\boldsymbol{\vartheta}}\Phi_i(\boldsymbol{\vartheta}) = \varphi(\boldsymbol{z}_i'\boldsymbol{\vartheta})\boldsymbol{z}_i'$, *then*

$$\sqrt{n}H_n(\tilde{\boldsymbol{\theta}},\hat{\boldsymbol{\vartheta}}) = \boldsymbol{D}_1\sqrt{n}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+\boldsymbol{D}_2\sqrt{n}(\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0).$$

Proof. The proof follows closely Lemma 2 in Chernozhukov, Fernández-Val, and Kowalski (2015), although it is significantly simpler because we do not estimate the arguments of an indicator variable to define subsets of observations.

Let $\bar{\boldsymbol{\vartheta}}$ be on the line connecting $\boldsymbol{\vartheta}_0$ and $\hat{\boldsymbol{\vartheta}}$ and $\bar{\boldsymbol{\theta}}$ be on the line connecting $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}$. Using the mean value theorem on the expected value of the quantile score:

$$\begin{aligned}
\mathrm{E}\left[\psi_\tau\left(y_i-\tilde{\boldsymbol{\theta}}'\boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})\right)\right] &= f_u((\bar{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0)'\boldsymbol{X}_i(\bar{\boldsymbol{\vartheta}})|\boldsymbol{x}_i,\boldsymbol{z}_i)\left[\boldsymbol{X}_i(\bar{\boldsymbol{\vartheta}})'(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+\dot{\boldsymbol{X}}_i(\bar{\boldsymbol{\vartheta}})'\bar{\boldsymbol{\theta}}(\Phi_i(\hat{\boldsymbol{\vartheta}})-\Phi_i(\boldsymbol{\vartheta}_0))\right],\\
&= f_u((\bar{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0)'\boldsymbol{X}_i(\bar{\boldsymbol{\vartheta}})|\boldsymbol{x}_i,\boldsymbol{z}_i)\left[\boldsymbol{X}_i(\bar{\boldsymbol{\vartheta}})'(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+\dot{\boldsymbol{X}}_i(\bar{\boldsymbol{\vartheta}})'\bar{\boldsymbol{\theta}}\Psi_i(\boldsymbol{\vartheta})(\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0)+o_p(1)\right],
\end{aligned}$$

where the first equality follows by mean value expansion and the continuity of $\boldsymbol{X}_i(\boldsymbol{\vartheta})$, and the last expression follows by expanding the marginal CDF Φ_i under Assumption A2. Using the result in Theorem 1 and multiplying by $\sqrt{n}$, we obtain,

$$\mathrm{E}\left[\sqrt{n}\psi_\tau\left(y_i-\boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})'\tilde{\boldsymbol{\theta}}\right)\right] = f_u(0|\boldsymbol{x}_i,\boldsymbol{z}_i)\left[\boldsymbol{X}_i(\boldsymbol{\vartheta}_0)'\sqrt{n}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+\dot{\boldsymbol{X}}_i(\boldsymbol{\vartheta}_0)'\boldsymbol{\theta}_0\Psi_i(\boldsymbol{\vartheta})\sqrt{n}(\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0)\right]+o_p(1),$$

by Assumption A6.a on the consistency of $\hat{\boldsymbol{\vartheta}}$ for $\boldsymbol{\vartheta}_0$.

Multiplying the last expression by $\boldsymbol{X}_i(\boldsymbol{\vartheta}_0)$ gives the desired result. $\square$

Proof of Theorem 2. By the computational property of the quantile regression estimator, $\|\mathbb{H}_n(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\vartheta}})\| = O_p(n^{-1})$. To see this,

$$\begin{aligned}
\left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}}) \psi_\tau(y_i - \hat{\boldsymbol{\theta}}' \boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})) \right\| &= \left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}}) I(y_i = \hat{\boldsymbol{\theta}}' \boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})) \right\| \\
&\leq \left\| \frac{1}{n} \sum_{i=1}^{n} \max_{1 \leq i \leq n} \{\boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})\} I(y_i = \hat{\boldsymbol{\theta}}' \boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})) \right\| \\
&\leq \left| \sum_{i=1}^{n} I(y_i = \hat{\boldsymbol{\theta}}' \boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})) \right| \frac{1}{n} \max_{1 \leq i \leq n} \left\| \boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}}) \right\| = \frac{p}{n} \max_{1 \leq i \leq n} \left\| \boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}}) \right\|.
\end{aligned}$$

It follows that,

$$\begin{aligned}
o_p(1) = \sqrt{n} \mathbb{H}_n(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\vartheta}}) &= \sqrt{n} \left(\mathbb{H}_n(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\vartheta}}) - H_n(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\vartheta}}) + H_n(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\vartheta}}) \right) \\
&= \sqrt{n}(\mathbb{H}_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0) - H_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0) + o_p(1)) + \sqrt{n} H_n(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\vartheta}}) \\
&= \sqrt{n}(\mathbb{H}_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0) + o_p(1)) + \boldsymbol{D}_1 \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + (\mathbb{K}_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0) \sqrt{n}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) + o_p(1)),
\end{aligned}$$

where the last equality follows by Lemma 2. By the invertibility of $\boldsymbol{D}_1$ implied by Assumption A8, we solve for $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ and obtain the Bahadur representation of the two-step estimator:

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\boldsymbol{D}_1^{-1} \sqrt{n} \mathbb{H}_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0) - \boldsymbol{D}_1^{-1} \mathbb{K}_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0) \sqrt{n}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) + o_p(1). \tag{A.1}$$

It is clear that the first term corresponds to the Bahadur representation of the quantile regression estimator for known $\boldsymbol{\vartheta}_0$. Thus, the first term is $\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_p(1)$, and then, it must be the case that the second term is $\sqrt{n}(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}) + o_p(1)$, because

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \sqrt{n}(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}).$$

The Bahadur representation (A.1) is similar to the Bahadur representation of other two-step quantile regression estimator (see, e.g., Theorem 2 in Ma and Koenker, 2006).

Considering the first term in equation (A.1), we obtain the following result based on the Bahadur representation of the estimator $\tilde{\boldsymbol{\theta}}$:

$$\boldsymbol{D}_1^{-1} \sqrt{n} \mathbb{H}_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0) = \boldsymbol{D}_1^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{X}_i(\boldsymbol{\vartheta}_0) \psi_\tau(y_i - \boldsymbol{\theta}_0' \boldsymbol{X}_i(\boldsymbol{\vartheta}_0)) + o_p(1) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{D}_1^{-1} \boldsymbol{D}_0 \boldsymbol{D}_1^{-1}).$$

Considering the second term in equation (A.1), by an application of the CLT,

$$\boldsymbol{D}_1^{-1}\mathbb{K}_n(\boldsymbol{\theta}_0,\boldsymbol{\vartheta}_0)\sqrt{n}(\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0)=\boldsymbol{D}_1^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i(\boldsymbol{\vartheta})f_u(0|\boldsymbol{x}_i,\boldsymbol{z}_i)\dot{\boldsymbol{X}}_i(\boldsymbol{\vartheta}_0)'\boldsymbol{\theta}_0\Psi_i(\boldsymbol{\vartheta}_0)\sqrt{n}(\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0)+o_p(1)$$
$$\xrightarrow{d}\mathcal{N}(\mathbf{0},\boldsymbol{D}_1^{-1}\boldsymbol{D}_2\boldsymbol{\Omega}\boldsymbol{D}_2'\boldsymbol{D}_1^{-1}).$$

The proof is completed after combining the results on the limiting distributions of $\sqrt{n}(\hat{\boldsymbol{\theta}}-\tilde{\boldsymbol{\theta}})$ and $\sqrt{n}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$, and recognizing that the two terms in (A.1) are not independent. $\square$

Proof of Theorem 3. The proof is organized in three steps. We begin verifying conditions for stochastic equicontinuity to apply the arguments of Theorem 2 in Chernozhukov, Fernández-Val, and Kowalski (2015). Second, we obtain the Bahadur representation of the bootstrap estimator $\check{\boldsymbol{\theta}}$. Finally, we employ the multiplier CLT to obtain the desired result. In this proof, we let $\mathrm{P}^\star\{\cdot\}=\mathrm{P}\{\cdot|\boldsymbol{S}\}$ denote the probability calculated conditional on the observed sample $\boldsymbol{S}$ and the stochastic orders $O_{p^\star}(\cdot)$ and $o_{p^\star}(\cdot)$ are interpreted conditional on the sample $\boldsymbol{S}$.

Recall that $\psi_\tau(u)=\tau-I(u\leq 0)$ and note that $I(y_i-\boldsymbol{\theta}'\boldsymbol{X}_i(\boldsymbol{\vartheta})<0)$ belongs to the type I class of functions of Andrews (1994). The function

$$\left\{\frac{1}{n}\sum_{i=1}^{n}\psi_\tau(y_i-\boldsymbol{\theta}'\boldsymbol{X}_i(\boldsymbol{\vartheta}))\boldsymbol{X}_i(\boldsymbol{\vartheta})|\boldsymbol{\theta},\boldsymbol{\vartheta}\in\boldsymbol{\Theta}\times\boldsymbol{\Gamma}\right\},$$

is a Donsker class and a Glivenko-Cantelli class. By Theorem 2.9.2 in van der Vaart and Wellner (1996), it can be verified that the product of a random variable ω_i and the previous function,

$$\left\{\frac{1}{n}\sum_{i=1}^{n}\omega_i\psi_\tau(y_i-\boldsymbol{\theta}'\boldsymbol{X}_i(\boldsymbol{\vartheta}))\boldsymbol{X}_i(\boldsymbol{\vartheta})|\boldsymbol{\theta},\boldsymbol{\vartheta}\in\boldsymbol{\Theta}\times\boldsymbol{\Gamma}\right\},$$

also belong to the Donsker class. Then, the conditions of Lemma 1 (a and b) in Chernozhukov, Fernández-Val, and Kowalski (2015) are satisfied under Assumptions B1 and B2. By Lemma 2 and Step 1 in Theorem 2 in Chernozhukov, Fernández-Val, and Kowalski (2015), $\sqrt{n}(\check{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)=O_p(1)$, unconditional on the data.

Note that ω_i denotes the number of times unit i is redrawn from the original sample. Thus, the asymptotic distribution of $(\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)',\sqrt{n}(\hat{\alpha}-\alpha_0))'$ is approximated by $(\sqrt{n}(\check{\boldsymbol{\beta}}-$

$\hat{\boldsymbol{\beta}})', \sqrt{n}(\check{\alpha}-\hat{\alpha}))'$ where

$$\begin{aligned}
\check{\boldsymbol{\theta}} = \left(\check{\boldsymbol{\beta}}', \check{\alpha}\right)' &= \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \left\{ \frac{1}{n}\sum_{i=1}^{n} \omega_i \rho_\tau \left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta} - \alpha\Phi_i(\check{\boldsymbol{\vartheta}})\right) \right\}, \\
&= \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \left\{ \frac{1}{n}\sum_{i=1}^{n} \omega_i \rho_\tau \left(\hat{u}_i - \boldsymbol{x}_i'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) - \Phi_i(\check{\boldsymbol{\vartheta}})(\alpha - \hat{\alpha})\right) - \rho_\tau\left(\hat{u}_i\right) \right\},
\end{aligned}$$

where $\hat{u}_i = y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}} - \hat{\alpha}\Phi_i(\check{\boldsymbol{\vartheta}})$. Since ω_i is a multinomial weight with probability $1/n$, it is straightforward to calculate that the expected value of the objective function with respect to the bootstrap weights conditional on $\boldsymbol{S}$ is minimized at $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\alpha})$.

Similarly to the proof of Theorem 2,

$$\begin{aligned}
\left\| \frac{1}{n}\sum_{i=1}^{n} \omega_i \boldsymbol{X}_i(\check{\boldsymbol{\vartheta}})\psi_\tau(y_i - \check{\boldsymbol{\theta}}'\boldsymbol{X}_i(\check{\boldsymbol{\vartheta}}))\right\| &= \left\| \frac{1}{n}\sum_{i=1}^{n} \omega_i \boldsymbol{X}_i(\check{\boldsymbol{\vartheta}}) I(y_i = \check{\boldsymbol{\theta}}'\boldsymbol{X}_i(\check{\boldsymbol{\vartheta}}))\right\| \\
&\le \left\| \frac{1}{n}\sum_{i=1}^{n} \max_{1\le i\le n}\{\omega_i\} \max_{1\le i\le n}\{\boldsymbol{X}_i(\check{\boldsymbol{\vartheta}})\} I(y_i = \check{\boldsymbol{\theta}}'\boldsymbol{X}_i(\check{\boldsymbol{\vartheta}}))\right\| \\
&\le \left| \sum_{i=1}^{n} I(y_i = \check{\boldsymbol{\theta}}'\boldsymbol{X}_i(\check{\boldsymbol{\vartheta}}))\right| \frac{1}{n} \max_{1\le i\le n}\{\omega_i\} \max_{1\le i\le n} \left\|\boldsymbol{X}_i(\check{\boldsymbol{\vartheta}})\right\| \\
&= \frac{p}{n} \max_{1\le i\le n}\{\omega_i\} \max_{1\le i\le n} \left\|\boldsymbol{X}_i(\check{\boldsymbol{\vartheta}})\right\|.
\end{aligned}$$

Therefore,

$$\begin{aligned}
o_p(1) = \sqrt{n}\mathbb{H}_n(\check{\boldsymbol{\theta}}, \check{\boldsymbol{\vartheta}}) &= \sqrt{n}\left(\mathbb{H}_n(\check{\boldsymbol{\theta}}, \check{\boldsymbol{\vartheta}}) - H_n(\check{\boldsymbol{\theta}}, \check{\boldsymbol{\vartheta}}) + H_n(\check{\boldsymbol{\theta}}, \check{\boldsymbol{\vartheta}})\right) \\
&= \sqrt{n}(\mathbb{H}_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0) - H_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0) + o_p(1)) + \sqrt{n}H_n(\check{\boldsymbol{\theta}}, \check{\boldsymbol{\vartheta}}) \\
&= \sqrt{n}(\mathbb{H}_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0) + o_p(1)) + \boldsymbol{D}_1\sqrt{n}(\check{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + (\mathbb{K}_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0)\sqrt{n}(\check{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) + o_p(1)),
\end{aligned}$$

where the last equality follows by Lemma 2. We solve for $\sqrt{n}(\check{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ to obtain the Bahadur representation of the bootstrap estimator:

$$\begin{aligned}
\sqrt{n}(\check{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = &- \boldsymbol{D}_1^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \omega_i \boldsymbol{X}_i(\boldsymbol{\vartheta}_0)\psi_\tau(y_i - \boldsymbol{X}_i(\boldsymbol{\vartheta})'\boldsymbol{\theta}_0) \\
&- \boldsymbol{D}_1^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \omega_i \boldsymbol{X}_i(\boldsymbol{\vartheta}_0) f_u(0|\boldsymbol{x}_i, \boldsymbol{z}_i)\dot{\boldsymbol{X}}_i(\boldsymbol{\vartheta}_0)'\boldsymbol{\theta}_0\Psi_i(\boldsymbol{\vartheta}_0)\sqrt{n}(\check{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) + o_p(1).
\end{aligned}$$

We now obtain the Bahadur representation of the bootstrap estimator conditional on the data. Noting that $\sqrt{n}(\check{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) = \sqrt{n}(\check{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, using the result in Theorem 2, we obtain,

$$\begin{aligned}
\sqrt{n}(\check{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) = &- \boldsymbol{D}_1^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\omega_i - 1) \boldsymbol{X}_i(\boldsymbol{\vartheta}_0) \psi_\tau (y_i - \boldsymbol{X}_i(\boldsymbol{\vartheta}_0)' \boldsymbol{\theta}_0) - \\
&- \boldsymbol{D}_1^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \omega_i f_u(0|\boldsymbol{x}_i, \boldsymbol{z}_i) \dot{\boldsymbol{X}}_i(\boldsymbol{\vartheta}_0)' \boldsymbol{\theta} \Psi_i(\boldsymbol{\vartheta}_0) \sqrt{n}((\hat{\boldsymbol{\vartheta}} + o_{p^\star}(1)) - \boldsymbol{\vartheta}_0) \\
&+ \boldsymbol{D}_1^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f_u(0|\boldsymbol{x}_i, \boldsymbol{z}_i) \dot{\boldsymbol{X}}_i(\boldsymbol{\vartheta}_0)' \boldsymbol{\theta}_0 \Psi_i(\boldsymbol{\vartheta}_0) \sqrt{n}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) + o_p(1) \\
= &- \boldsymbol{D}_1^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\omega_i - 1) \boldsymbol{X}_i(\boldsymbol{\vartheta}_0) \Big[\psi_\tau (y_i - \boldsymbol{X}_i(\boldsymbol{\vartheta}_0)' \boldsymbol{\theta}_0) + \\
&\qquad f_u(0|\boldsymbol{x}_i, \boldsymbol{z}_i) \dot{\boldsymbol{X}}_i(\boldsymbol{\vartheta}_0)' \boldsymbol{\theta}_0 \Psi_i(\boldsymbol{\vartheta}_0) \sqrt{n}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) \Big] + o_p(1) + o_{p^\star}(1),
\end{aligned}$$

where the first equality holds by Assumption B2.

By the Conditional Multiplier CLT as in Lemma 2.9.5 in van der Vaart and Wellner (1996), we have that conditional on the data,

$$\sqrt{n}(\check{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{D}_1^{-1}[\boldsymbol{D}_0 + \boldsymbol{D}_2 \boldsymbol{\Omega} \boldsymbol{D}_2' - \boldsymbol{D}_3] \boldsymbol{D}_1^{-1}).$$

The last statement means that, for each $t \in \mathbb{R}^{p+1}$,

$$\mathrm{P}^\star \left\{ \sqrt{n}(\check{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) \le t \right\} - \mathrm{P} \left\{ \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \le t \right\} \to 0.$$

$\square$

References

Andrews, D. W. (1994): "Chapter 37 Empirical process methods in econometrics," vol. 4 of *Handbook of Econometrics*, pp. 2247–2294. Elsevier.

Bollinger, C. R. (1996): "Bounding mean regressions when a binary regressor is mismeasured," *Journal of Econometrics*, 73(2), 387–399.

Bollinger, C. R., B. T. Hirsch, C. M. Hokayem, and J. P. Ziliak (2019): "Trouble in the Tails? What We Know about Earnings Nonresponse 30 Years after Lillard, Smith, and Welch," *Journal of Political Economy*, 127(5), 2143–2185.

Cavanagh, C., and R. P. Sherman (1998): "Rank estimators for monotonic index models," *Journal of Econometrics*, 84(2), 351–381.

Chatterjee, S., and A. Bose (2005): "Generalized bootstrap for estimating equations," *Annals of Statistics*, 33(1), 414–436.

Chen, L., A. F. Galvao, and S. Song (2021): "Quantile Regression with Generated Regressors," *Econometrics*, 9(2), 1–35.

Chen, X., Y. Fan, and V. Tsyrennikov (2006): "Efficient Estimation of Semiparametric Multivariate Copula Models," *Journal of the American Statistical Association*, 101(475), 1228–1240.

Chernozhukov, V., I. Fernández-Val, and A. Kowalski (2015): "Quantile regression with censoring and endogeneity," *Journal of Econometrics*, 186(1), 201–221.

Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73(1), 245–262.

——— (2006): "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics*, 132(2), 491 – 525.

Chesher, A. (2017): "Understanding the effect of measurement error on quantile regressions," *Journal of Econometrics*, 200(2), 223–237.

Dahl, G. B., A. R. Kostøl, and M. Mogstad (2014): "Family Welfare Cultures," *The Quarterly Journal of Economics*, 129(4), 1711–1752.

DiTraglia, F. J., and C. García-Jimeno (2019): "Identifying the effect of a mis-classified, binary, endogenous regressor," *Journal of Econometrics*, 209(2), 376–390.

Firpo, S., A. F. Galvao, and S. Song (2017): "Measurement errors in quantile regression models," *Journal of Econometrics*, 198(1), 146–164.

Frazis, H., and M. A. Loewenstein (2003): "Estimating linear regressions with mismeasured, possibly endogenous, binary explanatory variables," *Journal of Econometrics*, 117(1), 151–178.

Hahn, J., and Z. Liao (2021): "Bootstrap Standard Error Estimates and Inference," *Econometrica*, 89(4), 1963–1977.

Han, S., and S. Lee (2019): "Estimation in a generalization of bivariate probit models with dummy endogenous regressors," *Journal of Applied Econometrics*, 34(6), 994–1015.

Han, S., and E. J. Vytlacil (2017): "Identification in a generalization of bivariate probit models with dummy endogenous regressors," *Journal of Econometrics*, 199(1), 63–73.

Hartley, R. P., C. Lamarche, and J. P. Ziliak (2022): "Welfare Reform and the Intergenerational Transmission of Dependence," *Journal of Political Economy*, 130(3), 523–565.

He, X. (2018): "Resampling Methods," in *Handbook of Quantile Regression*, pp. 7–17. CRC Press.

He, X., and H. Liang (2000): "Quantile Regression Estimates for a Class of Linear and Partially Linear Errors-in-Variables Models," *Statistica Sinica*, 10(1), 129–140.

Jin, Z., Z. Ying, and L. J. Wei (2001): "A simple resampling method by perturbing the minimand," *Biometrika*, 88(2), 381–390.

Knight, K. (1998): "Limiting Distributions for L_1 Regression Estimators Under General Conditions," *Annals of Statistics*, 26, 755–770.

Koenker, R. (2005): *Quantile Regression*. Cambridge University Press.

KOENKER, R. (2017): "Quantile Regression: 40 Years On," *Annual Review of Economics*, 9(1), 155–176.

KREIDER, B., J. V. PEPPER, C. GUNDERSEN, AND D. JOLLIFFE (2012): "Identifying the Effects of SNAP (Food Stamps) on Child Health Outcomes When Participation Is Endogenous and Misreported," *Journal of the American Statistical Association*, 107(499), 958–975.

LAMARCHE, C., AND T. PARKER (2023): "Wild Bootstrap Inference for Penalized Quantile Regression for Panel Data," Journal of Econometrics, forthcoming.

LEWBEL, A. (2007): "Estimation of Average Treatment Effects with Misclassification," *Econometrica*, 75(2), 537–551.

MA, L., AND R. KOENKER (2006): "Quantile regression methods for recursive structural equation models," *Journal of Econometrics*, 134(2), 471–506.

MA, S., AND M. R. KOSOROK (2005): "Robust semiparametric M-estimation and the weighted bootstrap," *Journal of Multivariate Analysis*, 96(1), 190 – 217.

MAHAJAN, A. (2006): "Identification and Estimation of Regression Models with Misclassification," *Econometrica*, 74(3), 631–665.

MEYER, B. D., W. K. C. MOK, AND J. X. SULLIVAN (2009): "The Under-Reporting of Transfers in Household Surveys: Its Nature and Consequences," Working Paper 15181, National Bureau of Economic Research.

MEYER, B. D., W. K. C. MOK, AND J. X. SULLIVAN (2015): "Household Surveys in Crisis," *Journal of Economic Perspectives*, 29(4), 199–226.

NEWEY, W. K., AND D. MCFADDEN (1994): "Chapter 36 Large sample estimation and hypothesis testing," vol. 4 of *Handbook of Econometrics*, pp. 2111–2245. Elsevier.

NGUIMKEU, P., A. DENTEH, AND R. TCHERNIS (2019): "On the estimation of treatment effects with endogenous misreporting," *Journal of Econometrics*, 208(2), 487–506.

PENG, L., AND Y. HUANG (2008): "Survival Analysis With Quantile Regression Models," *Journal of the American Statistical Association*, 103, 637–649.

POIRIER, D. J. (1980): "Partial observability in bivariate probit models," *Journal of Econometrics*, 12(2), 209–217.

SCHENNACH, S. M. (2008): "Quantile Regression with Mismeasured Covariates," *Econometric Theory*, 24(4), 1010–1043.

URA, T. (2021): "Instrumental Variable Quantile Regression with Misclassification," *Econometric Theory*, 37(1), 169–204.

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.

WANG, H. J., L. A. STEFANSKI, AND Z. ZHU (2012): "Corrected-loss estimation for quantile regression with covariate measurement errors," *Biometrika*, 99(2), 405–421.

Wei, Y., and R. J. Carroll (2009): "Quantile Regression With Measurement Error," *Journal of the American Statistical Association*, 104(487), 1129–1143.

SUPPLEMENTARY APPENDIX TO "QUANTILE REGRESSION WITH AN ENDOGENOUS ONE-SIDED MISCLASSIFIED BINARY REGRESSOR"*

CARLOS LAMARCHE†

Remarks on notation and definitions: Throughout the appendix, we define $\boldsymbol{\theta} = (\boldsymbol{\beta}', \alpha)'$, but we suppress the dependency on τ for notational simplicity. The proofs below refer to Knight's (1998) identity: $\rho_\tau(u-v) - \rho_\tau(u) = -v\psi_\tau(u) + \int_0^v (I(u \leq s) - I(u \leq 0))\mathrm{d}s$, where $\rho_\tau = u(\tau - I(u < 0))$ is the quantile regression check function and $\psi_\tau(u) = \tau - I(u < 0)$ is the associated score function.

It is convenient to introduce additional notation. Let $\Phi_i(\boldsymbol{\vartheta}) := \Phi_i(\boldsymbol{z}_i'\boldsymbol{\vartheta})$, $\boldsymbol{X}_i(\boldsymbol{\vartheta}) := (\boldsymbol{x}_i', \Phi_i(\boldsymbol{\vartheta}))'$, and $\dot{\boldsymbol{X}}_i(\boldsymbol{\vartheta}) := \partial_\Phi \boldsymbol{X}_i(\boldsymbol{\vartheta})$. Moreover, let $\boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})$ denote the vector $\boldsymbol{X}_i(\boldsymbol{\vartheta})$ evaluated at the estimated values. Consider the unfeasible estimator,

$$\tilde{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \left\{ \frac{1}{n} \sum_{i=1}^n \rho_\tau(y_i - \boldsymbol{\theta}'\boldsymbol{X}_i(\boldsymbol{\vartheta}_0)) \right\},$$

and the feasible version of $\tilde{\boldsymbol{\theta}}$ as in (2.9):

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \left\{ \frac{1}{n} \sum_{i=1}^n \rho_\tau(y_i - \boldsymbol{\theta}'\boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})) \right\}.$$

S.1. Proof of main results

Proof of Theorem 1. Write $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = (\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}) + (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$. Consistency is established in two steps. We first show that $\tilde{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$, and in the second part of the proof, we demonstrate that $\hat{\boldsymbol{\theta}} \xrightarrow{p} \tilde{\boldsymbol{\theta}}$.

Let $\hat{\boldsymbol{\theta}}$ be the minimizer of the normalized objective function

$$\mathbb{M}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^n \rho_\tau\left(y_i - \boldsymbol{\theta}'\boldsymbol{X}_i(\boldsymbol{\vartheta}_0)\right) = \frac{1}{n}\sum_{i=1}^n \rho_\tau\left(u_i - (\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\boldsymbol{X}_i(\boldsymbol{\vartheta}_0)\right),$$

where $u_i = y_i - \boldsymbol{\theta}_0'\boldsymbol{X}_i(\boldsymbol{\vartheta}_0)$, and $(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\boldsymbol{X}_i(\boldsymbol{\vartheta}_0) = \boldsymbol{x}_i'(\boldsymbol{\beta}-\boldsymbol{\beta}_0) + \Phi_i(\boldsymbol{\vartheta}_0)(\alpha - \alpha_0)$. Let $\Delta_n(\boldsymbol{\theta}) = \mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n(\boldsymbol{\theta}_0)$, that is,

$$\Delta_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^n \left\{\rho_\tau\left(u_i - (\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\boldsymbol{X}_i(\boldsymbol{\vartheta}_0)\right) - \rho_\tau\left(u_i\right)\right\}.$$

*This draft: February 6, 2024.

†Carlos Lamarche: Department of Economics, University of Kentucky, 223G Gatton College of Business & Economics, Lexington, KY 40506. Email: clamarche@uky.edu.

By Knight's (1998) identity, $\Delta_n(\boldsymbol{\theta}) = \mathbb{V}_n^{(1)} + \mathbb{V}_n^{(2)}$, where,

$$\mathbb{V}_n^{(1)}(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{i=1}^{n}\left\{\boldsymbol{x}_i'(\boldsymbol{\beta}-\boldsymbol{\beta}_0) + \Phi_i(\boldsymbol{\vartheta}_0)(\alpha-\alpha_0)\right\}\psi_\tau(u_i),$$

$$\mathbb{V}_n^{(2)}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\boldsymbol{x}_i'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+\Phi_i(\boldsymbol{\vartheta}_0)(\alpha-\alpha_0)}\left(I(u_i \le s) - I(u_i \le 0)\right)\mathrm{d}s,$$

Note that under Assumptions A1 and A3, $\mathrm{E}\left[\mathbb{V}_n^{(1)}(\boldsymbol{\theta})\right] = 0$, because the quantile of u_i conditional on $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ is equal to the conditional quantile of u_i^* conditional on $\boldsymbol{x}_i$ and d_i^*.

We first show the consistency of $\tilde{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}_0$. For each $\phi > 0$, define the ball $\mathcal{B}(\phi) := \{\boldsymbol{\theta} : \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_1 \le \phi\}$ and the boundary $\partial\mathcal{B}(\phi) := \{\boldsymbol{\theta} : \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_1 = \phi\}$. For each $\boldsymbol{\theta} \notin \mathcal{B}(\phi)$, define $\bar{\boldsymbol{\theta}} = r\boldsymbol{\theta} + (1-r)\boldsymbol{\theta}_0$ where $r = \phi/\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|_1$. By construction, $r \in (0,1)$, and $\bar{\boldsymbol{\theta}}$ is in the boundary set of $\mathcal{B}(\phi)$.

By the convexity of $\mathbb{M}_n(\boldsymbol{\theta})$,

$$r\mathbb{M}_n(\boldsymbol{\theta}) + (1-r)\mathbb{M}_n(\boldsymbol{\theta}_0) \ge \mathbb{M}_n(\bar{\boldsymbol{\theta}}),$$

or, $r\big(\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n(\boldsymbol{\theta}_0)\big) \ge \mathbb{M}_n(\bar{\boldsymbol{\theta}}) - \mathbb{M}_n(\boldsymbol{\theta}_0) = \mathrm{E}\left[\Delta_n(\bar{\boldsymbol{\theta}})\right] + \big(\Delta_n(\bar{\boldsymbol{\theta}}) - \mathrm{E}\left[\Delta_n(\bar{\boldsymbol{\theta}})\right]\big)$. Under Assumption A4, we obtain,

$$\mathrm{E}\left[\Delta_n(\boldsymbol{\theta})\right] = \mathrm{E}\left[\int_0^{\boldsymbol{x}_i'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+\Phi_i(\boldsymbol{\vartheta})(\alpha-\alpha_0)}(F_u(s|\boldsymbol{x}_i,\boldsymbol{z}_i) - \tau)\mathrm{d}s\right] \ge \epsilon_\phi,$$

for some $\epsilon_\phi > 0$. Then,

$$r(\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n(\boldsymbol{\theta}_0)) = r\Delta_n(\boldsymbol{\theta}) \ge \epsilon_\phi + \left(\Delta_n(\bar{\boldsymbol{\theta}}) - \mathrm{E}\left[\Delta_n(\bar{\boldsymbol{\theta}})\right]\right).$$

By definition of $\tilde{\boldsymbol{\theta}}$ as the minimizer of $\frac{1}{n}\sum_{i=1}^{n}\mathbb{M}_n(\boldsymbol{\theta})$, we obtain the following inclusion relationships:

$$\left\{\|\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|_1 > \phi\right\} \subseteq \left\{\tilde{\boldsymbol{\theta}} \notin \mathcal{B}(\phi) \text{ and } \mathbb{M}_n(\tilde{\boldsymbol{\theta}}) \le \mathbb{M}_n(\boldsymbol{\theta}_0)\right\} \subseteq \left\{\sup_{\boldsymbol{\theta}\in\mathcal{B}(\phi)}\Big|\Delta_n(\boldsymbol{\theta}) - \mathrm{E}\left[\Delta_n(\boldsymbol{\theta})\right]\Big| \ge \epsilon_\phi\right\}.$$

Using standard arguments (i.e., derivations analogous to those in Theorem 1 in Lamarche and Parker, 2023), for any $\epsilon > 0$ and constants $D > 0$ and $C > 0$,

$$\mathrm{P}\left\{\sup_{\boldsymbol{\theta}\in\mathcal{B}(\phi)}\left|\Delta_n(\boldsymbol{\theta}) - \mathrm{E}\left[\Delta_n(\boldsymbol{\theta})\right]\right| > \epsilon\right\} \le 2C\exp\{-Dn\}.$$

Therefore, with probability tending to zero, $\|\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|_1 > \phi$, and thus, $\tilde{\boldsymbol{\theta}}$ is consistent for $\boldsymbol{\theta}_0$.

We now demonstrate the consistency of the feasible estimator $\hat{\boldsymbol{\theta}}$ for the unfeasible $\tilde{\boldsymbol{\theta}}$. Write the objective function of $\hat{\boldsymbol{\theta}}$ as follows:

$$\rho_\tau(y_i - \boldsymbol{\theta}'\boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})) = \rho_\tau(y_i - \boldsymbol{x}_i'\boldsymbol{\beta} - \Phi_i(\hat{\boldsymbol{\vartheta}})\alpha) = \rho_\tau(y_i - \boldsymbol{x}_i'\boldsymbol{\beta} - \Phi_i(\boldsymbol{\vartheta}_0)\alpha - (\Phi_i(\hat{\boldsymbol{\vartheta}}) - \Phi_i(\boldsymbol{\vartheta}_0))\alpha).$$

Using a version of Knight's identity, $|\rho_\tau(u-v)-\rho_\tau(u)| \le 3|v|$, we write,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left|\rho_\tau(y_i-\boldsymbol{\theta}'\boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}}))-\rho_\tau(y_i-\boldsymbol{\theta}'\boldsymbol{X}_i(\boldsymbol{\vartheta}_0))\right| &\le 3\frac{1}{n}\sum_{i=1}^{n}|(\Phi_i(\hat{\boldsymbol{\vartheta}})-\Phi_i(\boldsymbol{\vartheta}_0))\alpha| \\ &\le 3|\alpha|\frac{1}{n}\sum_{i=1}^{n}|\Phi_i(\hat{\boldsymbol{\vartheta}})-\Phi_i(\boldsymbol{\vartheta}_0)|.\end{aligned}$$

Under Assumptions A2 and A6, expanding $\Phi(\hat{\boldsymbol{\vartheta}})$ close enough to $\boldsymbol{\vartheta}_0$, we obtain

$$\Phi(\hat{\boldsymbol{\vartheta}})=\Phi(\boldsymbol{\vartheta}_0)+\Psi(\boldsymbol{\vartheta})(\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0)+e(\boldsymbol{z},\hat{\gamma}),$$

where $\Psi(\boldsymbol{\vartheta})=\nabla_{\boldsymbol{\vartheta}}\Phi(\boldsymbol{\vartheta})=\varphi(\boldsymbol{z}'\boldsymbol{\vartheta})\boldsymbol{z}'$ and $e(\boldsymbol{z},\hat{\gamma})=(\Psi(\bar{\boldsymbol{\vartheta}})-\Psi(\boldsymbol{\vartheta}_0))'(\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0)$ for a mean value $\bar{\gamma}$.

It follows that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}|\Phi_i(\hat{\boldsymbol{\vartheta}})-\Phi_i(\boldsymbol{\vartheta}_0)| &\le \frac{1}{n}\sum_{i=1}^{n}\Big(|\Psi_i(\boldsymbol{\vartheta})(\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0)|+|e(\boldsymbol{z}_i,\hat{\gamma})|\Big) \\ &\le \|\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\|\ \frac{1}{n}\sum_{i=1}^{n}\|\Psi_i(\boldsymbol{\vartheta})\|+\frac{1}{n}\sum_{i=1}^{n}|e(\boldsymbol{z}_i,\hat{\gamma})| \\ &\le \|\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\|\ \frac{1}{n}\sum_{i=1}^{n}\|\Psi_i(\boldsymbol{\vartheta})\|+\|\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\|\ \frac{1}{n}\sum_{i=1}^{n}\|\Psi_i(\bar{\boldsymbol{\vartheta}})-\Psi_i(\boldsymbol{\vartheta}_0)\| \\ &\le \|\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\|\ \max_{1\le i\le n}\|\boldsymbol{z}_i\|+o_p(1),\end{aligned}$$

by continuity of Ψ_i implied by A2. Under A5, it can be established that,

$$\frac{1}{n}\sum_{i=1}^{n}\left|\rho_\tau(y_i-\boldsymbol{\theta}'\boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}}))-\rho_\tau(y_i-\boldsymbol{\theta}'\boldsymbol{X}_i(\boldsymbol{\vartheta}_0)))\right|\le M\ \|\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\|+o_p(1).$$

Therefore, as $n\to\infty$, $\hat{\boldsymbol{\theta}}$ converges in probability to $\tilde{\boldsymbol{\theta}}$ under Assumption A6.a.

Combining the two previous results completes the proof. $\square$

The following definitions are used in Lemma 1 and Theorem 2. Consider,

$$\begin{aligned}\mathbb{H}_n(\boldsymbol{\theta},\boldsymbol{\vartheta}) &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i(\boldsymbol{\vartheta})\psi_\tau\left(y_i-\boldsymbol{\theta}'\boldsymbol{X}_i(\boldsymbol{\vartheta})\right), \\ H_n(\boldsymbol{\theta},\boldsymbol{\vartheta}) &= \mathrm{E}\left[\mathbb{H}_n(\boldsymbol{\theta},\boldsymbol{\vartheta})\right]=\mathrm{E}\left[\boldsymbol{X}_i(\boldsymbol{\vartheta})(\tau-F_u((\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\boldsymbol{X}_i(\boldsymbol{\vartheta})|\boldsymbol{x}_i,\boldsymbol{z}_i)\right], \\ \mathbb{K}_n(\boldsymbol{\theta},\boldsymbol{\vartheta}) &= \frac{1}{n}\sum_{i=1}^{n}f_u(0|\boldsymbol{x}_i,\boldsymbol{z}_i)\boldsymbol{X}_i(\boldsymbol{\vartheta})\dot{\boldsymbol{X}}_i(\boldsymbol{\vartheta})'\boldsymbol{\theta}\Psi_i(\boldsymbol{\vartheta}).\end{aligned}$$

Lemma 1. *Under Assumptions A1-A8, if* $\sqrt{n}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)=O_p(1)$ *and*

$$\sqrt{n}(\Phi_i(\hat{\boldsymbol{\vartheta}})-\Phi_i(\boldsymbol{\vartheta}_0))=\Psi_i(\boldsymbol{\vartheta})\sqrt{n}(\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0)+o_p(1),$$

for $\Psi_i(\boldsymbol{\vartheta})=\nabla_{\boldsymbol{\vartheta}}\Phi_i(\boldsymbol{\vartheta})=\varphi(\boldsymbol{z}_i'\boldsymbol{\vartheta})\boldsymbol{z}_i'$*, then*

$$\sqrt{n}H_n(\tilde{\boldsymbol{\theta}},\hat{\boldsymbol{\vartheta}})=\boldsymbol{D}_1\sqrt{n}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+\boldsymbol{D}_2\sqrt{n}(\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0).$$

Proof. The proof follows closely Lemma 2 in Chernozhukov, Fernández-Val, and Kowalski (2015), although it is significantly simpler because we do not estimate the arguments of an indicator variable to define subsets of observations.

Let $\bar{\boldsymbol{\vartheta}}$ be on the line connecting $\boldsymbol{\vartheta}_0$ and $\hat{\boldsymbol{\vartheta}}$ and $\bar{\boldsymbol{\theta}}$ be on the line connecting $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}$. Using the mean value theorem on the expected value of the quantile score:

$$\begin{aligned}\mathrm{E}\left[\psi_\tau\left(y_i - \tilde{\boldsymbol{\theta}}'\boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})\right)\right] &= f_u((\bar{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0)'\boldsymbol{X}_i(\bar{\boldsymbol{\vartheta}})|\boldsymbol{x}_i, \boldsymbol{z}_i)\left[\boldsymbol{X}_i(\bar{\boldsymbol{\vartheta}})'(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \dot{\boldsymbol{X}}_i(\bar{\boldsymbol{\vartheta}})'\bar{\boldsymbol{\theta}}(\Phi_i(\hat{\boldsymbol{\vartheta}}) - \Phi_i(\boldsymbol{\vartheta}_0))\right], \\ &= f_u((\bar{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0)'\boldsymbol{X}_i(\bar{\boldsymbol{\vartheta}})|\boldsymbol{x}_i, \boldsymbol{z}_i)\left[\boldsymbol{X}_i(\bar{\boldsymbol{\vartheta}})'(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \dot{\boldsymbol{X}}_i(\bar{\boldsymbol{\vartheta}})'\bar{\boldsymbol{\theta}}\Psi_i(\boldsymbol{\vartheta})(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) + o_p(1)\right],\end{aligned}$$

where the first equality follows by mean value expansion and the continuity of $\boldsymbol{X}_i(\boldsymbol{\vartheta})$, and the last expression follows by expanding the marginal CDF Φ_i under Assumption A2. Using the result in Theorem 1 and multiplying by $\sqrt{n}$, we obtain,

$$\mathrm{E}\left[\sqrt{n}\psi_\tau\left(y_i - \boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})'\tilde{\boldsymbol{\theta}}\right)\right] = f_u(0|\boldsymbol{x}_i, \boldsymbol{z}_i)\left[\boldsymbol{X}_i(\boldsymbol{\vartheta}_0)'\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \dot{\boldsymbol{X}}_i(\boldsymbol{\vartheta}_0)'\boldsymbol{\theta}_0\Psi_i(\boldsymbol{\vartheta})\sqrt{n}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0)\right] + o_p(1),$$

by Assumption A6.a on the consistency of $\hat{\boldsymbol{\vartheta}}$ for $\boldsymbol{\vartheta}_0$.

Multiplying the last expression by $\boldsymbol{X}_i(\boldsymbol{\vartheta}_0)$ gives the desired result. □

Proof of Theorem 2. By the computational property of the quantile regression estimator, $\|\mathbb{H}_n(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\vartheta}})\| = O_p(n^{-1})$. To see this,

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})\psi_\tau(y_i - \hat{\boldsymbol{\theta}}'\boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}}))\right\| &= \left\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})I(y_i = \hat{\boldsymbol{\theta}}'\boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}}))\right\| \\ &\leq \left\|\frac{1}{n}\sum_{i=1}^n \max_{1\leq i\leq n}\{\boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})\}I(y_i = \hat{\boldsymbol{\theta}}'\boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}}))\right\| \\ &\leq \left|\sum_{i=1}^n I(y_i = \hat{\boldsymbol{\theta}}'\boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}}))\right| \frac{1}{n}\max_{1\leq i\leq n}\left\|\boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})\right\| = \frac{p}{n}\max_{1\leq i\leq n}\left\|\boldsymbol{X}_i(\hat{\boldsymbol{\vartheta}})\right\|.\end{aligned}$$

It follows that,

$$\begin{aligned}o_p(1) = \sqrt{n}\mathbb{H}_n(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\vartheta}}) &= \sqrt{n}\left(\mathbb{H}_n(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\vartheta}}) - H_n(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\vartheta}}) + H_n(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\vartheta}})\right) \\ &= \sqrt{n}(\mathbb{H}_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0) - H_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0) + o_p(1)) + \sqrt{n}H_n(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\vartheta}}) \\ &= \sqrt{n}(\mathbb{H}_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0) + o_p(1)) + \boldsymbol{D}_1\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + (\mathbb{K}_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0)\sqrt{n}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) + o_p(1)),\end{aligned}$$

where the last equality follows by Lemma 1. By the invertibility of $\boldsymbol{D}_1$ implied by Assumption A8, we solve for $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ and obtain the Bahadur representation of the two-step estimator:

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\boldsymbol{D}_1^{-1}\sqrt{n}\mathbb{H}_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0) - \boldsymbol{D}_1^{-1}\mathbb{K}_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0)\sqrt{n}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) + o_p(1). \tag{S.1}$$

It is clear that the first term corresponds to the Bahadur representation of the quantile regression estimator for known $\boldsymbol{\vartheta}_0$. Thus, the first term is $\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_p(1)$, and then, it must be the case that the second term is $\sqrt{n}(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}) + o_p(1)$, because

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \sqrt{n}(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}).$$

The Bahadur representation (S.1) is similar to the Bahadur representation of other two-step quantile regression estimator (see, e.g., Theorem 2 in Ma and Koenker, 2006).

Considering the first term in equation (S.1), we obtain the following result based on the Bahadur representation of the estimator $\tilde{\boldsymbol{\theta}}$:

$$\boldsymbol{D}_1^{-1}\sqrt{n}\mathbb{H}_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0) = \boldsymbol{D}_1^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{X}_i(\boldsymbol{\vartheta}_0)\psi_\tau(y_i - \boldsymbol{\theta}_0'\boldsymbol{X}_i(\boldsymbol{\vartheta}_0)) + o_p(1) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{D}_1^{-1}\boldsymbol{D}_0\boldsymbol{D}_1^{-1}).$$

Considering the second term in equation (S.1), by an application of the CLT,

$$\begin{aligned}\boldsymbol{D}_1^{-1}\mathbb{K}_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0)\sqrt{n}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) &= \boldsymbol{D}_1^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i(\boldsymbol{\vartheta})f_u(0|\boldsymbol{x}_i, \boldsymbol{z}_i)\dot{\boldsymbol{X}}_i(\boldsymbol{\vartheta}_0)'\boldsymbol{\theta}_0\Psi_i(\boldsymbol{\vartheta}_0)\sqrt{n}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) + o_p(1)\\ &\xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{D}_1^{-1}\boldsymbol{D}_2\boldsymbol{\Omega}\boldsymbol{D}_2'\boldsymbol{D}_1^{-1}).\end{aligned}$$

The proof is completed after combining the results on the limiting distributions of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}})$ and $\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, and recognizing that the two terms in (S.1) are not independent. $\square$

Proof of Theorem 3. The proof is organized in three steps. We begin verifying conditions for stochastic equicontinuity to apply the arguments of Theorem 2 in Chernozhukov, Fernández-Val, and Kowalski (2015). Second, we obtain the Bahadur representation of the bootstrap estimator $\check{\boldsymbol{\theta}}$. Finally, we employ the multiplier CLT to obtain the desired result. In this proof, we let $\mathrm{P}^\star\{\cdot\} = \mathrm{P}\{\cdot|\boldsymbol{S}\}$ denote the probability calculated conditional on the observed sample $\boldsymbol{S}$ and the stochastic orders $O_{p^\star}(\cdot)$ and $o_{p^\star}(\cdot)$ are interpreted conditional on the sample $\boldsymbol{S}$.

Recall that $\psi_\tau(u) = \tau - I(u \leq 0)$ and note that $I(y_i - \boldsymbol{\theta}'\boldsymbol{X}_i(\boldsymbol{\vartheta}) < 0)$ belongs to the type I class of functions of Andrews (1994). The function

$$\left\{\frac{1}{n}\sum_{i=1}^{n}\psi_\tau(y_i - \boldsymbol{\theta}'\boldsymbol{X}_i(\boldsymbol{\vartheta}))\boldsymbol{X}_i(\boldsymbol{\vartheta})|\boldsymbol{\theta}, \boldsymbol{\vartheta} \in \boldsymbol{\Theta} \times \boldsymbol{\Gamma}\right\},$$

is a Donsker class and a Glivenko-Cantelli class. By Theorem 2.9.2 in van der Vaart and Wellner (1996), it can be verified that the product of a random variable ω_i and the previous function,

$$\left\{\frac{1}{n}\sum_{i=1}^{n}\omega_i\psi_\tau(y_i - \boldsymbol{\theta}'\boldsymbol{X}_i(\boldsymbol{\vartheta}))\boldsymbol{X}_i(\boldsymbol{\vartheta})|\boldsymbol{\theta}, \boldsymbol{\vartheta} \in \boldsymbol{\Theta} \times \boldsymbol{\Gamma}\right\},$$

also belong to the Donsker class. Then, the conditions of Lemma 1 (a and b) in Chernozhukov, Fernández-Val, and Kowalski (2015) are satisfied under Assumptions B1 and B2. By Lemma 2 and Step 1 in Theorem 2 in Chernozhukov, Fernández-Val, and Kowalski (2015), $\sqrt{n}(\check{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_p(1)$, unconditional on the data.

Note that ω_i denotes the number of times unit i is redrawn from the original sample. Thus, the asymptotic distribution of $(\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)', \sqrt{n}(\hat{\alpha}-\alpha_0))'$ is approximated by $(\sqrt{n}(\check{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}})', \sqrt{n}(\check{\alpha}-\hat{\alpha}))'$ where

$$\begin{aligned}\check{\boldsymbol{\theta}} = \left(\check{\boldsymbol{\beta}}', \check{\alpha}\right)' &= \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \left\{ \frac{1}{n}\sum_{i=1}^{n} \omega_i \rho_\tau\left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta} - \alpha\Phi_i(\check{\boldsymbol{\vartheta}})\right)\right\}, \\ &= \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}} \left\{ \frac{1}{n}\sum_{i=1}^{n} \omega_i \rho_\tau\left(\hat{u}_i - \boldsymbol{x}_i'(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}) - \Phi_i(\check{\boldsymbol{\vartheta}})(\alpha-\hat{\alpha})\right) - \rho_\tau\left(\hat{u}_i\right)\right\},\end{aligned}$$

where $\hat{u}_i = y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}} - \hat{\alpha}\Phi_i(\check{\boldsymbol{\vartheta}})$. Since ω_i is a multinomial weight with probability $1/n$, it is straightforward to calculate that the expected value of the objective function with respect to the bootstrap weights conditional on $\boldsymbol{S}$ is minimized at $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\alpha})$.

Similarly to the proof of Theorem 2,

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^{n}\omega_i \boldsymbol{X}_i(\check{\boldsymbol{\vartheta}})\psi_\tau(y_i - \check{\boldsymbol{\theta}}'\boldsymbol{X}_i(\check{\boldsymbol{\vartheta}}))\right\| &= \left\|\frac{1}{n}\sum_{i=1}^{n}\omega_i \boldsymbol{X}_i(\check{\boldsymbol{\vartheta}}) I(y_i = \check{\boldsymbol{\theta}}'\boldsymbol{X}_i(\check{\boldsymbol{\vartheta}}))\right\| \\ &\le \left\|\frac{1}{n}\sum_{i=1}^{n}\max_{1\le i\le n}\{\omega_i\}\max_{1\le i\le n}\{\boldsymbol{X}_i(\check{\boldsymbol{\vartheta}})\} I(y_i = \check{\boldsymbol{\theta}}'\boldsymbol{X}_i(\check{\boldsymbol{\vartheta}}))\right\| \\ &\le \left|\sum_{i=1}^{n} I(y_i = \check{\boldsymbol{\theta}}'\boldsymbol{X}_i(\check{\boldsymbol{\vartheta}}))\right| \frac{1}{n}\max_{1\le i\le n}\{\omega_i\}\max_{1\le i\le n}\left\|\boldsymbol{X}_i(\check{\boldsymbol{\vartheta}})\right\| \\ &= \frac{p}{n}\max_{1\le i\le n}\{\omega_i\}\max_{1\le i\le n}\left\|\boldsymbol{X}_i(\check{\boldsymbol{\vartheta}})\right\|.\end{aligned}$$

Therefore,

$$\begin{aligned}o_p(1) = \sqrt{n}\mathbb{H}_n(\check{\boldsymbol{\theta}}, \check{\boldsymbol{\vartheta}}) &= \sqrt{n}\left(\mathbb{H}_n(\check{\boldsymbol{\theta}}, \check{\boldsymbol{\vartheta}}) - H_n(\check{\boldsymbol{\theta}}, \check{\boldsymbol{\vartheta}}) + H_n(\check{\boldsymbol{\theta}}, \check{\boldsymbol{\vartheta}})\right) \\ &= \sqrt{n}(\mathbb{H}_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0) - H_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0) + o_p(1)) + \sqrt{n}H_n(\check{\boldsymbol{\theta}}, \check{\boldsymbol{\vartheta}}) \\ &= \sqrt{n}(\mathbb{H}_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0) + o_p(1)) + \boldsymbol{D}_1\sqrt{n}(\check{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + (\mathbb{K}_n(\boldsymbol{\theta}_0, \boldsymbol{\vartheta}_0)\sqrt{n}(\check{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) + o_p(1)),\end{aligned}$$

where the last equality follows by Lemma 1. We solve for $\sqrt{n}(\check{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ to obtain the Bahadur representation of the bootstrap estimator:

$$\begin{aligned}\sqrt{n}(\check{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = &-\boldsymbol{D}_1^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\omega_i \boldsymbol{X}_i(\boldsymbol{\vartheta}_0)\psi_\tau(y_i - \boldsymbol{X}_i(\boldsymbol{\vartheta})'\boldsymbol{\theta}_0) \\ &-\boldsymbol{D}_1^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\omega_i \boldsymbol{X}_i(\boldsymbol{\vartheta}_0) f_u(0|\boldsymbol{x}_i, \boldsymbol{z}_i)\dot{\boldsymbol{X}}_i(\boldsymbol{\vartheta}_0)'\boldsymbol{\theta}_0\Psi_i(\boldsymbol{\vartheta}_0)\sqrt{n}(\check{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) + o_p(1).\end{aligned}$$

We now obtain the Bahadur representation of the bootstrap estimator conditional on the data. Noting that $\sqrt{n}(\check{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) = \sqrt{n}(\check{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, using the result in Theorem 2, we obtain,

$$
\begin{aligned}
\sqrt{n}(\check{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) = & - \boldsymbol{D}_1^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\omega_i - 1) \boldsymbol{X}_i(\boldsymbol{\vartheta}_0) \psi_\tau(y_i - \boldsymbol{X}_i(\boldsymbol{\vartheta}_0)' \boldsymbol{\theta}_0) - \\
& - \boldsymbol{D}_1^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \omega_i f_u(0|\boldsymbol{x}_i, \boldsymbol{z}_i) \dot{\boldsymbol{X}}_i(\boldsymbol{\vartheta}_0)' \boldsymbol{\theta} \Psi_i(\boldsymbol{\vartheta}_0) \sqrt{n}((\hat{\boldsymbol{\vartheta}} + o_{p^\star}(1)) - \boldsymbol{\vartheta}_0) \\
& + \boldsymbol{D}_1^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f_u(0|\boldsymbol{x}_i, \boldsymbol{z}_i) \dot{\boldsymbol{X}}_i(\boldsymbol{\vartheta}_0)' \boldsymbol{\theta}_0 \Psi_i(\boldsymbol{\vartheta}_0) \sqrt{n}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) + o_p(1) \\
= & - \boldsymbol{D}_1^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\omega_i - 1) \boldsymbol{X}_i(\boldsymbol{\vartheta}_0) \Big[\psi_\tau(y_i - \boldsymbol{X}_i(\boldsymbol{\vartheta}_0)' \boldsymbol{\theta}_0) + \\
& \qquad f_u(0|\boldsymbol{x}_i, \boldsymbol{z}_i) \dot{\boldsymbol{X}}_i(\boldsymbol{\vartheta}_0)' \boldsymbol{\theta}_0 \Psi_i(\boldsymbol{\vartheta}_0) \sqrt{n}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) \Big] + o_p(1) + o_{p^\star}(1),
\end{aligned}
$$

where the first equality holds by Assumption B2.

By the Conditional Multiplier CLT as in Lemma 2.9.5 in van der Vaart and Wellner (1996), we have that conditional on the data,

$$
\sqrt{n}(\check{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{D}_1^{-1}[\boldsymbol{D}_0 + \boldsymbol{D}_2 \boldsymbol{\Omega} \boldsymbol{D}_2' - \boldsymbol{D}_3] \boldsymbol{D}_1^{-1}).
$$

The last statement means that, for each $t \in \mathbb{R}^{p+1}$,

$$
\mathrm{P}^\star \left\{ \sqrt{n}(\check{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) \leq t \right\} - \mathrm{P} \left\{ \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \leq t \right\} \to 0.
$$

□

S.2. Extension: A parametrically guided estimator

Following A2, the estimator in (2.9) is obtained after estimating a Gaussian copula in the first step. The procedure allows for misspecification of the joint distribution, but it requires the marginal distribution $F_v(v)$ to be correctly specified. This might be viewed as a stronger condition relative to the ones employed in other quantile regression models. For instance, in models with endogenous regressors and no misclassification, control functions are correctly specified for consistent estimation (Ma and Koenker, 2006; Chernozhukov, Fernández-Val, and Kowalski, 2015).

A number of studies develop approximations to bivariate distributions for binary data using the bivariate normal distribution. One could consider,

$$
P_i(\boldsymbol{\vartheta}, \boldsymbol{\gamma}, \rho) = \Phi_i(\boldsymbol{\vartheta}, \boldsymbol{\gamma}, \rho) \left(1 + \sum_j \sum_k \frac{K_{ik}}{j!k!} \mathrm{E}\left[H_{jk}(v, \epsilon, \rho) \right] \right),
$$

where $H_{jk}(v, \epsilon, \rho)$ are bivariate Hermite polynomials and K_{ik} are cumulants (Murphy, 2007). The function $H_{jk} := ((-1)^{j+k} D_1^j D_1^k \varphi(v, \epsilon, \rho))/\varphi(v, \epsilon, \rho)$, where D_1 is an operator for differentiation and $\phi(v, \epsilon, \rho)$ is the density function of the normal distribution. Another alternative is to consider parametrically guided methods, which can improve the performance of estimators that employ a parametric first step (Fan, Wu, and Feng, 2009). Assume that the function F_v is known up to a small approximation error following the identify: $F_v = \Phi_v + r_v\Phi_v$, where $r_v = (F_v - \Phi_v)/\Phi_v$ and Φ_v denotes the Gaussian cumulative distribution function. Because the approximation error is a function of $\boldsymbol{z}_i$, this suggests that we can augment the model by including r_v as a control variate. This strategy requires a small adaptation of the IVQR method proposed in Chernozhukov and Hansen (2006).

The estimation procedure below uses a third step to produce an estimator of α_0 that corresponds to the smallest distance between F_v and Φ_v. Define $C_i(\tau, \alpha, \boldsymbol{\beta}, \eta) = \rho_\tau\left(y_i - \alpha\Phi_i(\hat{\boldsymbol{\vartheta}}) - \boldsymbol{x}_i'\boldsymbol{\beta} - \eta\hat{r}_i\right)$, where $\hat{r}_i$ is a nonparametric estimator of $F_v(\boldsymbol{z}_i) - \Phi(\boldsymbol{z}_i'\hat{\boldsymbol{\vartheta}})$, where $\Phi(\boldsymbol{z}_i'\hat{\boldsymbol{\vartheta}})$ is obtained as in Step 1. We first minimize the objective function $C_i(\tau, \alpha, \boldsymbol{\beta}, \eta)$ for $\boldsymbol{\beta}$, η as functions of τ and α,

$$\{\hat{\boldsymbol{\beta}}_{3S}(\tau, \alpha), \hat{\eta}_{3S}(\tau, \alpha)\} = \underset{\boldsymbol{\beta}, \eta \in \mathcal{B}\times\mathcal{G}}{\operatorname{argmin}} \sum_{i=1}^{n} C_i(\tau, \alpha; \boldsymbol{\beta}, \eta). \tag{S.1}$$

Then we estimate the parameter of interest by finding the value of α which minimizes a globally convex function defined on η:

$$\hat{\alpha}_{3S} = \underset{\alpha\in\mathcal{A}}{\operatorname{argmin}} \ \left\{\hat{\eta}_{3S}(\tau, \alpha)^2\right\}. \tag{S.2}$$

The three-step quantile regression estimator (3SQR) for a model with endogenous misclassification is defined as $\hat{\boldsymbol{\theta}}_{3S} = (\hat{\boldsymbol{\beta}}_{3S}(\hat{\alpha}_{3S})', \hat{\alpha}_{3S})'$.

Remark S.1. A non-parametric estimator of $F_v(\boldsymbol{z}_i)$ can be obtained by considering $d_i^* - \Phi_i(\hat{\boldsymbol{\vartheta}})|d_i = 1$ for $1 \leq i \leq n$ if $\Pr(d_i = 1|d_i^* = 0, \boldsymbol{x}_i) = 0$. Naturally, we observe d_i^* only if $m_i = 1$, and therefore, the 3SQR is in principle unfeasible. Because the small approximation error depends on $\boldsymbol{z}_i$, we implement the estimator considering $r_i = F_v(\boldsymbol{z}_i) - \Phi(\boldsymbol{z}_i'\hat{\boldsymbol{\vartheta}}) = g(\mu, \boldsymbol{z}_i)$, which is estimated as a linear function of the instruments. We evaluate the performance of such estimator in Section S.3.

S.3. Additional simulation results

This section presents additional monte carlo results considering the data generating process introduced in S.3.

Table S.1 presents bias and root mean square error (RMSE) of the proposed estimators in comparison with QR and two IV estimators. Following Lemma 1, we consider QR using $1\{z_i \geq 0\}$ instead of d_i as a regressor (IV). Moreover, we consider the IV quantile regression estimator proposed by Chernozhukov and Hansen (2006) (CH). The CH estimator uses the observed variable d_i as a regressor, but identification relies on the instrumental variable z_i. It is important to bear in mind that the QR, IV, and CH estimators do not allow for endogenous misclassification and could be biased in finite samples. Lastly, the table shows results obtained by the QREM and 3SQR estimators introduced in Sections 2.3 and S.2. The bias and RMSE are for the parameter of interest α_0 at $\tau \in \{0.5, 0.75\}$, obtained from 1000 samples of size 5000.

The upper panel of Table S.1 provides evidence of the biases present in the application of QR methods under exogenous misclassification and participation. As expected, the QR, IV and CH estimators are unbiased if the proportion of false negatives is zero (i.e., $\pi_0 = 0$). The performance of these estimators deteriorate when $\pi_0 > 0$, reaching bias that range from 20% to 50% at $\pi_0 = 0.4$ and $\tau = 0.5$. In the case of endogenous participation and misclassification, the CH estimator offers, in general, lower bias relative to QR and IV. The proposed estimators have almost zero bias and the lowest MSE in models with endogenous misclassification. The results across quantiles lead to similar conclusions.

Tables S.2 and S.3 show the bias and RMSE when the error distributions are t_3 and χ^2_3. We continue to see that the performance of the QREM and 3SQR estimators are satisfactory and they offer substantially better MSE performance relative to existing methods. In Table S.3, the 3SQR estimator has slightly smaller bias and better MSE performance.

Figure S.1 further investigates the difference in performance between QREM and 3SQR. To illustrate the impact of deviations from Gaussian conditions, we assume that the distribution of errors is χ^2_3. The figure shows the bias (left panel) and RMSE (right panel) of QREM and 3SQR estimators as the sample size increases from $n = 1000$ to $n = 5000$. We present results for α_0 at $\tau \in \{0.5, 0.75\}$. Consistent again with expectations, we observe that when we deviate from joint normality, the 3SQR estimator offers the best finite sample performance. The differences between estimators tend to disappear as n increases, and the 3SQR estimator offers significant gains in terms of MSE when the sample size is $n \leq 2000$.

Lastly, we also investigate the performance of the estimators when the proportion of false positives is different than zero. We use the same data generating process described in equation (4.1) but we generate the observed binary regressor as follows:

$$d_i = d_i^* 1\{\gamma_0 + \gamma_1 w_i + \epsilon_i \geq c\} + (1 - d_i^*) 1\{\kappa_i < b\}, \tag{S.1}$$

where $b \in [0, 1]$ is a parameter and κ_i is an i.i.d. random variable distributed as $U[0, 1]$. Naturally, $b = 0$ implies that $\pi_1 = 0$, while when $b > 0$, the sample includes $100 \times b\%$ of false positives. Following Nguimkeu, Denteh, and Tchernis (2019), we set $b = 0.10$. Relative to Table S.1, the finite sample performance of the estimators in Table S.4 deteriorate, although the proposed approaches continue to perform quite well in comparison with existing methods.

References

Andrews, D. W. (1994): "Chapter 37 Empirical process methods in econometrics," vol. 4 of *Handbook of Econometrics*, pp. 2247–2294. Elsevier.

Chernozhukov, V., I. Fernández-Val, and A. Kowalski (2015): "Quantile regression with censoring and endogeneity," *Journal of Econometrics*, 186(1), 201–221.

Chernozhukov, V., and C. Hansen (2006): "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics*, 132(2), 491 – 525.

				$\tau = 0.50$ Quantile					$\tau = 0.75$ Quantile				
	π_0	ζ_v	ζ_ϵ	QR	IV	CH	EM	3S	QR	IV	CH	EM	3S
Bias	0.00	0.00	0.00	0.000	0.100	-0.001	0.002	0.001	0.000	-0.042	-0.001	-0.001	-0.002
RMSE	0.00	0.00	0.00	0.042	0.109	0.074	0.072	0.074	0.047	0.062	0.082	0.079	0.082
Bias	0.25	0.00	0.00	0.033	0.103	-0.056	0.009	0.010	-0.015	-0.040	0.039	0.001	0.002
RMSE	0.25	0.00	0.00	0.057	0.112	0.108	0.074	0.076	0.053	0.062	0.126	0.083	0.085
Bias	0.40	0.00	0.00	0.046	0.098	-0.103	0.000	-0.001	-0.022	-0.043	0.061	-0.001	-0.001
RMSE	0.40	0.00	0.00	0.066	0.107	0.130	0.074	0.076	0.059	0.062	0.151	0.079	0.081
Bias	0.00	0.30	0.00	0.402	0.099	0.000	-0.001	-0.001	0.396	-0.041	-0.001	0.000	0.001
RMSE	0.00	0.30	0.00	0.405	0.109	0.076	0.075	0.077	0.399	0.063	0.083	0.082	0.085
Bias	0.25	0.30	0.00	0.357	0.096	-0.069	-0.005	-0.005	0.297	-0.041	0.032	0.002	0.001
RMSE	0.25	0.30	0.00	0.360	0.105	0.112	0.074	0.076	0.301	0.062	0.119	0.082	0.084
Bias	0.40	0.30	0.00	0.339	0.098	-0.105	-0.005	-0.005	0.255	-0.042	0.061	0.000	0.000
RMSE	0.40	0.30	0.00	0.343	0.106	0.129	0.071	0.073	0.260	0.061	0.145	0.077	0.079
Bias	0.00	0.00	0.20	0.000	0.101	0.003	0.006	0.005	-0.002	-0.041	0.003	0.002	0.003
RMSE	0.00	0.00	0.20	0.043	0.109	0.076	0.073	0.075	0.048	0.061	0.083	0.080	0.082
Bias	0.25	0.00	0.20	0.114	0.097	-0.072	-0.004	-0.004	0.063	-0.045	0.022	-0.007	-0.007
RMSE	0.25	0.00	0.20	0.123	0.106	0.113	0.071	0.072	0.080	0.065	0.110	0.079	0.080
Bias	0.40	0.00	0.20	0.158	0.099	-0.102	0.001	0.001	0.088	-0.043	0.058	0.000	0.000
RMSE	0.40	0.00	0.20	0.166	0.108	0.131	0.072	0.074	0.103	0.062	0.145	0.078	0.080
Bias	0.00	0.30	0.20	0.401	0.098	-0.001	-0.002	-0.002	0.396	-0.041	-0.002	0.000	-0.001
RMSE	0.00	0.30	0.20	0.403	0.106	0.074	0.073	0.074	0.398	0.063	0.082	0.082	0.084
Bias	0.25	0.30	0.20	0.439	0.098	-0.066	-0.003	-0.002	0.379	-0.041	0.029	0.003	0.002
RMSE	0.25	0.30	0.20	0.441	0.107	0.108	0.072	0.073	0.382	0.064	0.112	0.081	0.081
Bias	0.40	0.30	0.20	0.448	0.102	-0.097	0.003	0.003	0.366	-0.040	0.062	0.005	0.004
RMSE	0.40	0.30	0.20	0.451	0.111	0.129	0.076	0.078	0.370	0.062	0.149	0.081	0.084

TABLE S.1. *Bias and root mean square error (RMSE) of quantile regression estimators for α_0 when the distribution of errors is Normal. QR denotes quantile regression, IV denotes QR with instrumental variables, CH denotes the IVQR estimator, EM the proposed two-step estimator, and 3S is the proposed three-step estimator.*

FAN, J., Y. WU, AND Y. FENG (2009): "Local Quasi-Likelihood with a Parametric Guide," *The Annals of Statistics*, 37(6B), 4153–4183.

KNIGHT, K. (1998): "Limiting Distributions for L_1 Regression Estimators Under General Conditions," *Annals of Statistics*, 26, 755–770.

LAMARCHE, C., AND T. PARKER (2023): "Wild Bootstrap Inference for Penalized Quantile Regression for Panel Data," Journal of Econometrics, forthcoming.

MA, L., AND R. KOENKER (2006): "Quantile regression methods for recursive structural equation models," *Journal of Econometrics*, 134(2), 471–506.

MURPHY, A. (2007): "Score tests of normality in bivariate probit models," *Economics Letters*, 95(3), 374–379.

NGUIMKEU, P., A. DENTEH, AND R. TCHERNIS (2019): "On the estimation of treatment effects with endogenous misreporting," *Journal of Econometrics*, 208(2), 487–506.

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.

	π_0	ζ_v	ζ_ϵ	$\tau = 0.50$ Quantile QR	IV	CH	EM	3S	$\tau = 0.75$ Quantile QR	IV	CH	EM	3S
Bias	0.00	0.00	0.00	0.001	0.099	-0.001	-0.018	-0.017	-0.010	-0.044	-0.001	0.002	0.002
RMSE	0.00	0.00	0.00	0.045	0.109	0.081	0.087	0.089	0.056	0.069	0.100	0.101	0.104
Bias	0.25	0.00	0.00	0.034	0.100	-0.075	-0.020	-0.018	-0.037	-0.044	0.033	0.003	0.003
RMSE	0.25	0.00	0.00	0.059	0.110	0.135	0.090	0.090	0.069	0.071	0.149	0.107	0.107
Bias	0.40	0.00	0.00	0.048	0.100	-0.139	-0.017	-0.015	-0.022	-0.042	0.068	0.005	0.004
RMSE	0.40	0.00	0.00	0.073	0.111	0.194	0.092	0.093	0.066	0.071	0.202	0.108	0.110
Bias	0.00	0.30	0.00	0.457	0.102	0.008	-0.011	-0.009	0.487	-0.041	0.004	0.010	0.009
RMSE	0.00	0.30	0.00	0.459	0.112	0.080	0.087	0.088	0.490	0.070	0.103	0.107	0.109
Bias	0.25	0.30	0.00	0.393	0.098	-0.074	-0.022	-0.019	0.349	-0.041	0.043	0.012	0.011
RMSE	0.25	0.30	0.00	0.396	0.107	0.130	0.088	0.090	0.355	0.070	0.148	0.106	0.109
Bias	0.40	0.30	0.00	0.371	0.097	-0.146	-0.024	-0.021	0.319	-0.045	0.060	0.001	0.000
RMSE	0.40	0.30	0.00	0.375	0.108	0.198	0.092	0.093	0.326	0.074	0.204	0.115	0.117
Bias	0.00	0.00	0.20	-0.001	0.099	-0.005	-0.021	-0.020	-0.011	-0.044	-0.005	-0.001	-0.003
RMSE	0.00	0.00	0.20	0.047	0.109	0.085	0.091	0.094	0.056	0.070	0.108	0.109	0.111
Bias	0.25	0.00	0.20	0.135	0.099	-0.075	-0.021	-0.018	0.082	-0.043	0.036	0.009	0.010
RMSE	0.25	0.00	0.20	0.143	0.109	0.135	0.091	0.093	0.101	0.071	0.149	0.109	0.111
Bias	0.40	0.00	0.20	0.186	0.099	-0.145	-0.020	-0.018	0.132	-0.041	0.069	0.010	0.009
RMSE	0.40	0.00	0.20	0.194	0.110	0.196	0.089	0.090	0.147	0.067	0.188	0.104	0.106
Bias	0.00	0.30	0.20	0.459	0.099	0.002	-0.018	-0.016	0.489	-0.046	-0.004	0.001	0.001
RMSE	0.00	0.30	0.20	0.462	0.109	0.083	0.091	0.092	0.493	0.075	0.106	0.112	0.113
Bias	0.25	0.30	0.20	0.495	0.101	-0.073	-0.020	-0.019	0.468	-0.042	0.034	0.009	0.008
RMSE	0.25	0.30	0.20	0.497	0.110	0.130	0.088	0.089	0.473	0.070	0.147	0.109	0.111
Bias	0.40	0.30	0.20	0.511	0.099	-0.136	-0.020	-0.017	0.481	-0.041	0.066	0.008	0.007
RMSE	0.40	0.30	0.20	0.514	0.109	0.188	0.089	0.091	0.485	0.071	0.189	0.109	0.110

TABLE S.2. *Bias and root mean square error (RMSE) of quantile regression estimators for α_0 when the distribution of errors is t_3. QR denotes quantile regression, IV denotes QR with instrumental variables, CH denotes the IVQR estimator, EM the proposed two-step estimator, and 3S is the proposed three-step estimator.*

				$\tau = 0.50$ Quantile					$\tau = 0.75$ Quantile				
	π_0	ζ_v	ζ_ϵ	QR	IV	CH	EM	3S	QR	IV	CH	EM	3S
Bias	0.00	0.00	0.00	-0.001	0.151	-0.003	0.000	0.002	-0.001	-0.061	-0.008	-0.007	-0.012
RMSE	0.00	0.00	0.00	0.083	0.172	0.254	0.255	0.251	0.128	0.137	0.384	0.395	0.382
Bias	0.25	0.00	0.00	0.031	0.157	-0.037	0.033	0.034	-0.014	-0.058	0.041	0.011	0.009
RMSE	0.25	0.00	0.00	0.097	0.177	0.359	0.257	0.257	0.138	0.141	0.511	0.410	0.397
Bias	0.40	0.00	0.00	0.036	0.147	-0.103	0.000	0.001	-0.027	-0.066	0.027	-0.003	-0.006
RMSE	0.40	0.00	0.00	0.103	0.168	0.390	0.261	0.258	0.154	0.139	0.531	0.394	0.384
Bias	0.00	0.30	0.00	1.044	0.151	0.001	-0.005	-0.002	1.431	-0.061	-0.006	-0.005	-0.006
RMSE	0.00	0.30	0.00	1.048	0.172	0.257	0.264	0.262	1.437	0.144	0.372	0.401	0.389
Bias	0.25	0.30	0.00	0.865	0.146	-0.089	-0.013	-0.011	1.084	-0.066	0.024	-0.007	-0.009
RMSE	0.25	0.30	0.00	0.871	0.167	0.358	0.267	0.262	1.094	0.142	0.490	0.403	0.389
Bias	0.40	0.30	0.00	0.803	0.146	-0.117	-0.012	-0.010	0.975	-0.066	0.018	-0.015	-0.014
RMSE	0.40	0.30	0.00	0.810	0.165	0.375	0.253	0.250	0.987	0.135	0.517	0.371	0.363
Bias	0.00	0.00	0.20	0.002	0.152	0.013	0.015	0.017	-0.001	-0.057	0.025	0.024	0.020
RMSE	0.00	0.00	0.20	0.085	0.173	0.261	0.261	0.257	0.132	0.134	0.379	0.390	0.378
Bias	0.25	0.00	0.20	0.273	0.145	-0.097	-0.014	-0.012	0.305	-0.069	-0.011	-0.023	-0.024
RMSE	0.25	0.00	0.20	0.291	0.166	0.348	0.252	0.250	0.335	0.142	0.476	0.383	0.374
Bias	0.40	0.00	0.20	0.373	0.150	-0.109	0.006	0.007	0.430	-0.065	0.024	0.001	-0.001
RMSE	0.40	0.00	0.20	0.387	0.170	0.395	0.252	0.250	0.457	0.139	0.533	0.390	0.384
Bias	0.00	0.30	0.20	1.043	0.145	-0.004	-0.009	-0.007	1.432	-0.064	-0.012	-0.006	-0.008
RMSE	0.00	0.30	0.20	1.047	0.167	0.251	0.258	0.256	1.437	0.144	0.372	0.397	0.386
Bias	0.25	0.30	0.20	1.125	0.147	-0.078	-0.005	-0.003	1.446	-0.062	0.022	-0.001	-0.005
RMSE	0.25	0.30	0.20	1.130	0.167	0.339	0.254	0.249	1.453	0.143	0.477	0.388	0.380
Bias	0.40	0.30	0.20	1.162	0.154	-0.073	0.012	0.013	1.460	-0.060	0.035	0.012	0.008
RMSE	0.40	0.30	0.20	1.168	0.174	0.396	0.268	0.265	1.469	0.139	0.526	0.401	0.392

TABLE S.3. *Bias and root mean square error (RMSE) of quantile regression estimators for α_0 when the distribution of errors is χ^2_3. QR denotes quantile regression, IV denotes QR with instrumental variables, CH denotes the IVQR estimator, EM the proposed two-step estimator, and 3S is the proposed three-step estimator.*

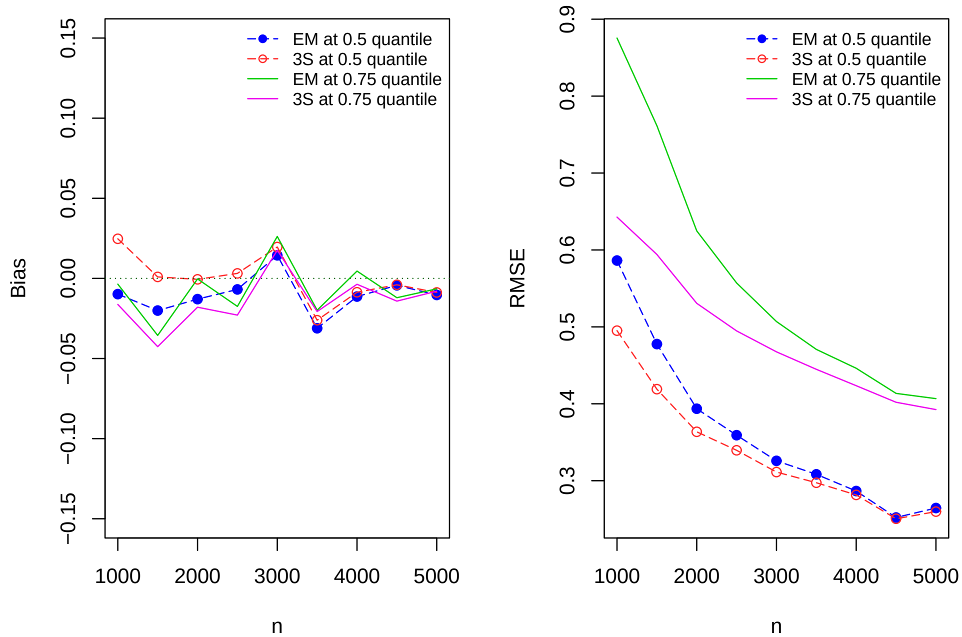

FIGURE S.1. *Bias and root mean square error (RMSE) for* α_0 *when the marginal distribution is* χ^2_3*. EM is the two-step estimator and 3S is the three-step estimator. The sample size is denoted by* n*.*

	π_0	ζ_v	ζ_ϵ	$\tau = 0.50$ Quantile QR	IV	CH	EM	3S	$\tau = 0.75$ Quantile QR	IV	CH	EM	3S
Bias	0.00	0.00	0.00	0.024	0.100	-0.021	-0.020	-0.019	-0.010	-0.044	0.007	0.006	0.007
RMSE	0.00	0.00	0.00	0.048	0.108	0.087	0.085	0.087	0.046	0.064	0.091	0.087	0.091
Bias	0.25	0.00	0.00	0.065	0.098	-0.094	-0.027	-0.027	-0.028	-0.041	0.054	0.012	0.013
RMSE	0.25	0.00	0.00	0.079	0.108	0.128	0.088	0.089	0.053	0.062	0.136	0.089	0.090
Bias	0.40	0.00	0.00	0.087	0.101	-0.116	-0.021	-0.021	-0.036	-0.044	0.080	0.006	0.007
RMSE	0.40	0.00	0.00	0.098	0.110	0.138	0.087	0.088	0.062	0.064	0.174	0.090	0.091
Bias	0.00	0.30	0.00	0.380	0.099	-0.020	-0.022	-0.022	0.350	-0.041	0.007	0.010	0.009
RMSE	0.00	0.30	0.00	0.383	0.108	0.083	0.084	0.085	0.353	0.062	0.090	0.090	0.092
Bias	0.25	0.30	0.00	0.327	0.097	-0.096	-0.031	-0.030	0.232	-0.042	0.042	0.008	0.007
RMSE	0.25	0.30	0.00	0.330	0.106	0.125	0.088	0.087	0.237	0.063	0.133	0.091	0.091
Bias	0.40	0.30	0.00	0.302	0.101	-0.113	-0.022	-0.022	0.173	-0.039	0.087	0.015	0.015
RMSE	0.40	0.30	0.00	0.305	0.110	0.136	0.086	0.086	0.181	0.063	0.176	0.095	0.096
Bias	0.00	0.00	0.20	0.023	0.101	-0.022	-0.018	-0.020	-0.008	-0.042	0.011	0.012	0.011
RMSE	0.00	0.00	0.20	0.048	0.110	0.084	0.082	0.084	0.047	0.062	0.093	0.089	0.092
Bias	0.25	0.00	0.20	0.142	0.101	-0.088	-0.021	-0.021	0.049	-0.043	0.046	0.008	0.009
RMSE	0.25	0.00	0.20	0.149	0.110	0.124	0.088	0.088	0.069	0.064	0.135	0.092	0.093
Bias	0.40	0.00	0.20	0.185	0.100	-0.117	-0.024	-0.023	0.063	-0.043	0.075	0.009	0.009
RMSE	0.40	0.00	0.20	0.191	0.109	0.139	0.089	0.089	0.080	0.064	0.170	0.091	0.093
Bias	0.00	0.30	0.20	0.381	0.099	-0.024	-0.023	-0.024	0.353	-0.041	0.005	0.010	0.008
RMSE	0.00	0.30	0.20	0.383	0.108	0.086	0.084	0.086	0.355	0.062	0.087	0.089	0.091
Bias	0.25	0.30	0.20	0.402	0.098	-0.090	-0.027	-0.026	0.312	-0.041	0.045	0.011	0.011
RMSE	0.25	0.30	0.20	0.405	0.106	0.122	0.085	0.085	0.316	0.062	0.131	0.090	0.093
Bias	0.40	0.30	0.20	0.403	0.098	-0.117	-0.025	-0.025	0.280	-0.040	0.080	0.016	0.016
RMSE	0.40	0.30	0.20	0.405	0.106	0.137	0.084	0.085	0.285	0.063	0.168	0.092	0.094

TABLE S.4. *Bias and root mean square error (RMSE) of quantile regression estimators for α_0 when the distribution of errors is Normal and the probability of false positives is 10%. QR denotes quantile regression, IV denotes QR with instrumental variables, CH denotes the IVQR estimator, EM the proposed two-step estimator, and 3S is the proposed three-step estimator.*