Econometric Theory, **0**, 2025, 1–44.
doi:10.1017/S0266466625100121

SEMIPARAMETRIC ESTIMATION OF QUANTILE REGRESSION WITH BINARY QUANTILE SELECTION

SONGNIAN CHEN
Zhejiang University

HANGHUI ZHANG
Shanghai University of Finance and Economics

This article proposes a novel method for estimating quantile regression models that account for sample selection. Unlike the approach by Arellano and Bonhomme (2017, *Econometrica* 85(1), 1–28; hereafter referred to as AB17), which employs a parametric selection equation, our method utilizes a standard binary quantile regression model to handle the selection issue, thereby accommodating general heterogeneity in both the selection and outcome equations. We adopt a semiparametric estimation technique for the outcome quantile regression by integrating local moment conditions, resulting in $\sqrt{n}$-consistent estimators for the quantile coefficients and copula parameter. Monte Carlo simulation results demonstrate that our estimator performs well in finite samples. Additionally, we apply our method to examine the wage distribution among women using a randomly simulated sample from the US General Social Survey. Our key finding is the presence of significant positive selection among women in the US, which is notably more pronounced than the estimates produced by the AB17's model.

1. INTRODUCTION

The quantile regression framework developed by Koenker and Bassett (1978) has received a great deal of attention in theoretical as well as applied economic analysis. Through parsimonious modeling, quantile regression provides comprehensive characterization of conditional distributions, allowing for general heterogeneous covariate effects and unobserved heterogeneity. However, most quantile regression analyses have traditionally focused on single equation frameworks.

In an influential article, AB17 studied quantile regression subject to sample selection. Sample selection dates back to Heckman (1979) and arises frequently in

We are very grateful for the insightful comments and suggestions from the editor, co-editor, and two anonymous referees, which greatly help to enhance the quality of this article. Address correspondence to Hanghui Zhang, School of Economics, Shanghai University of Finance and Economics; Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, Shanghai 200433, China, e-mail: zhang.hanghui@mail.shufe.edu.cn.

practice. For example, for the study of wages and employment, AB17 noted that only the wages of employed individuals are observed, so conventional measures of wage gaps or wage inequality may be biased and wage inequality for those at work may provide a distorted picture of market-level wage inequality. To overcome such bias and recover the latent wage distribution, AB17 proposed a quantile regression model subject to a sample selection where the sample selection is modeled via a bivariate cumulative distribution function (cdf), or copula, of the errors in the outcome and the selection equation. Furthermore, in a parametric/semiparametric framework, they proposed a two-step method for the estimation of the copula parameter, and the final step for the estimation for the entire family of quantile regression coefficient process given a consistent estimator for the copula parameter. Despite this important breakthrough in extending quantile regression to a sample selection framework, AB17's approach suffers from some drawbacks. Their parametric/semiparametric framework, which is the focus of their estimation, consists of a parametric selection equation and a semiparametric quantile regression outcome equation. As a result, their model accommodates general heterogeneous effects and unobserved heterogeneity only in the outcome equation, but not in the selection equation. As Horowitz (1993) pointed out, the parametric specification of the selection equation is prone to misspecification, often excluding heteroskedasticity, which results in biased predictions of conditional selection (choice) probabilities. Such misspecification can lead to inconsistent estimates and misleading inferences about the quantile coefficients.

In this article, we overcome the drawbacks associated with the parametric/semiparametric estimation approach[1] in AB17 by relaxing the parametric specification of the selection equation. In particular, like the outcome equation, the selection equation is also modeled as a semiparametric binary quantile regression. Therefore, our framework is able to accommodate general heterogeneous effects in both equations.[2] Furthermore, by specifying both the selection equation and the outcome as quantile regression equations, respectively, we provide a more natural and coherent modeling strategy. Similar to AB17, given initial estimates for the selection equation based on existing estimators, such as, Manski (1975, 1985), Horowitz (1992), and Chen and Zhang (2015), we propose a two-step estimator for the copula parameter. Our estimation approach differs from that of AB17 in constructing moment conditions; in fact, we work with the local moment conditions. Specifically, for a pair of quantile indices that correspond to the selection equation and outcome equation, we select a subsample for which

[1] Arellano and Bonhomme (2017) provided a brief discussion on estimation of quantile regression for the outcome equation with nonparametric specification for the selection equation. One major drawback with that approach is the curse-of-dimensionality problem associated with the nonparametric structure.

[2] We appreciate the editor's comments regarding our specification of the selection equation, and in particular, enforcing monotonic non-crossing in quantile regression does not fully resolve the misspecification issue (Phillips, 2015). However, binary quantile regression is inherently more robust than other parametric models. Furthermore, in practice, when quadratic or cubic terms are present, the issue of quantile crossing may not arise. Even in the case of a linear specification in terms of the base variables, quantile crossing may not occur if the regressor has a limited range (e.g., class sizes).

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

the local moment conditions correspond to a rotated quantile function, where the subsample is constructed based on the quantile index of the selection equation. Then, we combine the local moment conditions corresponding to different pairs of quantile indices for the estimation of the copula parameter. Once an estimate for the copula parameter is available, quantile regression coefficients for the outcome equation can be estimated by combining local quantile regression estimates based on different pairs of quantile indices for the selection and outcome equation. Although estimators based on individual local moment conditions converge at nonparametric rates, aggregating over local moment conditions based on a family of quantile indices leads to the parametric rate of convergence of our quantile regression estimator for the outcome equation. AB17's estimators are known for computational simplicity, using linear programming for their rotated quantile regression for each given ρ. Similarly, our estimators retain this computational advantage. The convolution smoothing quantile regression (SQR) introduced by Fernandes, Guerre, and Horta (2019) has the desirable property of maintaining the convexity of the objective function, enabling the computation of the global optimal solution in finite samples using standard local optimization methods.

The rest of the article is organized as follows: Section 2 presents the quantile regression model subject to binary quantile selection and proposes our estimation procedure. Section 3 presents the large sample properties of our estimator. Simulation results are contained in Section 4. We apply our method to the General Social Survey (GSS) data in Section 5. Section 6 concludes. The Appendix contains proofs of the main theorems. The Supplementary Material contains some additional simulation results.

2. MODEL AND ESTIMATOR

We consider the following model:

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

$$D = 1\left\{W'\gamma(V) > 0\right\}, \tag{2}$$

where (1) is the outcome equation and (2) is the selection equation; Y^* is the latent dependent variable in the outcome equation (1), X and W are observed characteristics with $W = (X, Z)$, where Z represents the variables in the selection equation excluded from the outcome equation; U and V are univariate unobserved characteristics each with the uniform distribution on $(0,1)$ for its marginal distribution conditional on W, representing "rankings" in the latent selection and outcome equations, respectively. Following the usual set-up of quantile regressions, conditional on $W = w$, both quantile functions $\tau \to x'\beta(\tau)$ and $\tau \to w'\gamma(\tau)$ are strictly increasing in $\tau \in (0,1)$. Our observations consist of a random sample of (Y_i, D_i, W_i), $i = 1, 2, \ldots, n$, where $Y_i = D_i Y_i^*$. Our objective is the estimation of quantile regression coefficients $\beta(\tau)$, for $\tau \in (0,1)$ as well as the corresponding conditional quantile function $x'\beta(\tau)$. As in AB17 we model the selection bias through the copula function $C^*(u, v, \rho_0)$, which is the joint distribution function of (U, V) assumed to be independent of W, with the copula parameter ρ_0 capturing the dependence between U and V.

In (1) and (2), both the binary selection and outcome equations are specified through semiparametric quantile regression functions, which allows for general heterogeneous effects and provides a more coherent framework. The semiparametric model studied in Section 3 of AB17 adopted a parametric specification for the selection equation for the estimation of the quantile regression coefficients in the outcome equation. Typically, for estimating binary choice models, estimates of the parameters and conditional probabilities can be sensitive to certain types of heteroskedasticity which are generally ruled out by parametric models such as Probit or Logit model.[3] Furthermore, the parametric specification of the selection equation cannot typically be justified by economic theory. Consequently, the misspecification of the selection equation is likely to result in inconsistency and erroneous inferences for $\beta(\tau)$ as well as the conditional quantile function $x'\beta(\tau)$.[4]

For the estimation of the quantile regression coefficients in the outcome equation, similar to AB17, we adopt a moment based approach; but unlike AB17, we work with local moment conditions. Specifically, for a given $v \in (0,1)$, we consider observations for which $W'\gamma(v) = 0$ holds, then we can write the binary selection indicator as

$$\begin{aligned} D &= 1\left\{W'\gamma(V) > 0\right\} \\ &= 1\left\{W'\gamma(V) > W'\gamma(v)\right\} \\ &= 1\{V > v\}, \end{aligned} \tag{3}$$

and consequently, we have

$$\begin{aligned} &E\left[D1\left\{Y \le X'\beta(\tau)\right\}|X, W'\gamma(v) = 0\right] \\ &\quad = E\left[1\left\{X'\beta(U) \le X'\beta(\tau), W'\gamma(V) > 0\right\}|X, W'\gamma(v) = 0\right] \\ &\quad = E\left[1\{U \le \tau, V \ge v\}\right]. \end{aligned}$$

The above equations yield the following conditional moment equation:

$$E\left[D\left(1\left\{Y \le X'\beta(\tau)\right\} - C(\tau, v, \rho_0)\right)|X, W'\gamma(v) = 0\right] = 0, \tag{4}$$

where $C(\cdot,\cdot,\rho)$ denotes the conditional copula,

$$C(\tau, v, \rho_0) = \frac{E(1\{U \le \tau, V > v\})}{E(1\{V > v\})} = \frac{C^*(\tau, 1, \rho_0) - C^*(\tau, v, \rho_0)}{1 - v},$$

[3] Typically, parametric binary choice models follow a specification,

$$D = 1\{W'\gamma_0 + \gamma_c(V) > 0\},$$

where, for example, $\gamma_c(V) = \Phi^{-1}(V)$ for the probit model.

[4] Arellano and Bonhomme (2017) also discussed the possibility of a nonparametrically selection equation; however, complete nonparametric specification will lead to the curse of dimensionality.

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

with $C^{*}(u,v,\rho_0)$ denoting the copula function, namely, the joint distribution function of (U,V). From the conditional moment equation (4), we can construct the following unconditional moment equations for $(\beta(\tau),\rho_0),\tau\in(0,1)$,

$$E\left[D\left(1\left\{Y\leq X^{'}\beta(\tau)\right\}-C(\tau,v,\rho_0)\right)X|W'\gamma(v)=0\right]=0, \tag{5}$$

and

$$E\left[D\left(1\left\{Y\leq X^{'}\beta(\tau)\right\}-C(\tau,v,\rho_0)\right)Z|W'\gamma(v)=0\right]=0. \tag{6}$$

Our identification strategy, in spirit, follows the common identification strategy used for nonlinear regression models. Essentially, we view $C^{*}(\tau,v,\rho)$ as a function of v indexed by (τ,ρ). $C^{*}(\tau,v,\rho_0)$ is distinguishable in the sense that for any given $\tau,\tilde{\tau}\in(0,1)$ and $\rho\in(-1,1)$, it is basically required that $C^{*}(\tau,v,\rho_0)=C^{*}(\tilde{\tau},v,\rho)$ for all $v\in(0,1)$, if and only if $\tau=\tilde{\tau}$ and $\rho=\rho_0$. The Proposition 1 in the Appendix shows the identification result.

Now we turn to the estimation strategy of the parameters of interest. Let $[v_l,v_u]$ denote a range of v in the support of the conditional choice probabilities $P(W)=E(D|W)$ for which we can estimate $\gamma(v)$ reasonably precisely. Let $\hat{\gamma}(v)$ denote an estimator for $\gamma(v)$ for $v\in[v_l,v_u]$.[5] Existing estimators for $\gamma(v)$ include the maximum score estimator (Manski, 1975, 1985, MSC estimator), the smoothed MSC estimator (Horowitz, 1992, see also, Kordas, 2006 and Volgushev, 2020, SMSC estimator), the local polynomial smoothed version of the MSC estimator by Chen and Zhang (2015), and more recently, the two-step version Gao et al (2022), among others. Note that given ρ_0 and the SMSC estimator $\hat{\gamma}(v)$ for the binary quantile selection equation, $\beta(\tau)$ can be estimated by a solution to

$$\frac{1}{n}\sum_{i=1}^{n}D_i\left[\left(1\left\{Y_i-X_i'b\leq 0\right\}-C(\tau,v,\rho_0)\right)X_i\right]\frac{1}{h_1}k_1\left(\frac{W_i'\hat{\gamma}(v)}{h_1}\right)=0, \tag{7}$$

which is the local estimating equation corresponding to the following local quantile regression:

$$\min_{b\in B}\sum_{i=1}^{n}D_i\rho_{C(\tau,v,\rho_0)}(Y_i-X_i'b)\frac{1}{h_1}k_1\left(\frac{W_i'\hat{\gamma}(v)}{h_1}\right), \tag{8}$$

where $\rho_\tau(u)=u(\tau-1\{u<0\})$ is the usual check function in quantile regression, $k_1(\cdot)$ is a kernel density function used to provide a smooth weighting, which gives more weight to observations with $W_i'\hat{\gamma}(v)$ close to zero.

Due to local weighting, the estimator in (8) will converge at a nonparametric rate. To improve the rate of convergence, we integrate the objective function in (8) over a continuum of vs by working with the objective function

$$\int_{v_l}^{v_u}\sum_{i=1}^{n}D_i\rho_{C(\tau,v,\rho_0)}(Y_i-X_i'b)\frac{1}{h_1}k_1\left(\frac{W_i'\hat{\gamma}(v)}{h_1}\right)\omega(v)\,dv, \tag{9}$$

[5] We discuss the range of τ_0 and the support of $P(W)$ in more detail in a remark below.

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

where $\omega(\cdot)$ is some nonnegative and integrable weight function required to be non-degenerated such that it only takes positive values on some interval $[v_l, v_u] \subset \mathcal{S}_{p(W)}$, where $\mathcal{S}_{p(W)}$ is the support of the propensity score. Furthermore, in light of the favorable asymptotic properties and finite sample performance associated with smoothed quantile regression proposed by Fernandes et al. (2019), we consider the convolution-type smoothing objective function of (9).

We now describe the details of our estimation algorithm for $\beta(\tau)$, $\tau \in \left[\tau_l^*, \tau_u^*\right]$ in $(0,1)$ and ρ_0: both parameters $(\beta(\tau), \rho_0)$ are solved in a profiled manner, similar to AB17. Suppose $\hat{\gamma}(v)$s are obtained by Horowitz's SMSC methods for $v \in [v_l, v_u]$.

Step 1: For a given ρ, $v \in [v_l, v_u]$ and $\tau \in \mathcal{J} \subset \left[\tau_l^*, \tau_u^*\right]$, where $\mathcal{J}$ is a set of quantile indices for the purpose of estimating ρ_0, define $\hat{\beta}(\tau, \rho)$ as a solution to

$$\min_{b \in B} \int_{v_l}^{v_u} M_{n1}\left(b, \hat{\gamma}(v), \rho, \tau, v\right) \omega(v)\, dv, \tag{10}$$

where M_{n1} represents a convolution-type smoothed objective function,

$$M_{n1}(b, \gamma, \rho, \tau, v) = \int_{-\infty}^{\infty} \rho_{C(\tau, v, \rho)}(t) f_n(t, b, \gamma)\, dt, \tag{11}$$

with

$$f_n(t, b, \gamma) = \frac{1}{nh_1h_2} \sum_{i=1}^{n} D_i k_2 \left(\frac{t - \left(Y_i - X_i'b\right)}{h_2}\right) k_1\left(\frac{W_i'\gamma}{h_1}\right),$$

$k_2(\cdot)$ is a kernel density function and h_2 is the corresponding bandwidth. For several commonly used kernel functions, we can compute (11) explicitly, for which details are provided below in the simulation section.

Step 2: Based on the first step estimator $\hat{\beta}(\tau, \rho)$, for any fixed ρ and $\tau \in \mathcal{J}$, define our estimator for ρ_0, $\hat{\rho}$, which solves

$$\min_{\rho \in \varrho} \bar{G}_{2n}(\rho),$$

where

$$\bar{G}_{2n}(\rho) = \sum_{\tau_j \in \mathcal{J}} \int_{v_l}^{v_u} \left\| G_{2n}\left(\hat{\beta}\left(\tau_j, \rho\right), \hat{\gamma}(v), \rho, v, \tau_j\right) \right\|^2 \omega(v)\, dv,$$

with

$$G_{2n}(b, \gamma, \rho, v, \tau) = \frac{1}{nh_1} \sum_{i=1}^{n} D_i \left(K_2\left(\frac{X_i'b - Y_i}{h_2}\right) - C(\tau, v, \rho)\right) k_1\left(\frac{W'\gamma}{h_1}\right) Z_i,$$

where $\|\cdot\|$ in $\bar{G}_{2n}(\cdot)$ is the Euclidean norm; $K_2(t) = \int_{-\infty}^{t} k_2(v)\,dv$ and $K_2(-t/h_2)$ represents a smoothed version of the indicator function $1\{t < 0\}$.

Step 3: Finally, we estimate $\beta(\tau)$ by $\hat{\beta}(\tau, \hat{\rho})$ for $\tau \in [\tau_l^*, \tau_u^*]$.

Remark 1 (Range of v and weight function). Note that the binary quantile selection equation (3) implies

$$P(W) \gtrless 1 - v \Leftrightarrow W'\gamma(v) \gtrless 0, \quad \textbf{(12)}$$
$$P(W) = 1 - v \Leftrightarrow W'\gamma(v) = 0.$$

Hence, the range of v, $[v_l, v_u]$ is chosen to be within the support of $P(W)$ and a natural choice of $\omega(v)$ could be simply the uniform weight contained in the support of $P(W)$, which is used throughout our simulation studies, namely, $\omega(v) = 1\left\{1 - \bar{p} + \varepsilon \le v \le 1 - \underline{p} - \varepsilon\right\}$, where $\underline{p} = \min_w P(w)$ and $\bar{p} = \max_w P(w)$ and ε is some small positive value (e.g., upper/lower 5–10% percentiles of $P(W)$). In practice, $\underline{p}$ and $\bar{p}$ can be based on some preliminary estimates.

Remark 2 ($\beta(\tau, \rho)$ for a given ρ). Note $\hat{\beta}(\tau, \rho)$ is an estimator for $\beta(\tau, \rho)$ which solves

$$\min_{b \in B} \bar{M}_1(b, \rho, \tau),$$

where

$$\bar{M}_1(b, \rho, \tau) = \int_{v_l}^{v_u} M_1(b, \gamma(v), \rho, \tau, v)\,\omega(v)dv,$$

with

$$M_1(b, \gamma(v), \rho, \tau, v) = E\left[1\{V > v\}\,\rho_{C(\tau, v, \rho)}\left(Y_i - X_i'b\right)|W'\gamma(v) = 0\right]p_{W'\gamma(v)}(0),$$

where $p_{W'\gamma(v)}(t)$ denotes the density of $W'\gamma(v)$ at t. Similar to the quantile regression analysis under misspecification (Angrist, Chernozhukov, and Fernandez-Val, 2006), $\beta(\tau, \rho)$ satisfies

$$\begin{aligned}
&\int_{v_l}^{v_u} E\left[D\left(1\left\{Y \le X'\beta(\tau, \rho)\right\} - C(\tau, v, \rho)\right)X|W'\gamma(v) = 0\right]p_{W'\gamma(v)}(0)\omega(v)\,dv \\
&\quad = E\Big[1\{V > 1 - P(W)\}\left(1\left\{Y \le X'\beta(\tau, \rho)\right\} - C(\tau, 1 - P(W), \rho)\right) \\
&\qquad \times X f_{Y_2^*}(0, W)\,\omega(1 - P(W))\Big] \\
&\quad = 0,
\end{aligned}$$

where $f_{Y^*}(\cdot, w)$ is the conditional density of Y^* given $W = w$. Thus under a full rank condition (Assumption 6) and some smoothness condition, $\beta(\tau, \rho)$ is unique and continuous and differentiable with respect to (τ, ρ).

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

Remark 3 (Computation of $\hat{\beta}(\tau,\rho)$). AB17's estimators are known for their computational simplicity. Given ρ, their rotated quantile regression can still be implemented using linear programming. Similarly, our estimators offer the same computational advantage. The convolution SQR introduced by Fernandes et al. (2019) has the desirable property of maintaining the convexity of the objective function, enabling the computation of the global optimal solution in finite samples using standard local optimization methods. Moreover, the SQR estimator outperforms standard quantile regression in terms of asymptotic mean squared error (AMSE). He et al. (2023) demonstrates that SQR provides significant computational advantages over standard quantile regression, especially for very large samples and/or a large number of regressors, along with slightly superior performance. Other versions of smoothed quantile regression have also been explored by Horowitz (1998) (non-convex smoothing). Horowitz's smoothing approach has been widely used for various QR-related problems, such as Galvao and Kato (2016), Kaplan and Sun (2017), and among others.

3. LARGE SAMPLE PROPERTIES

We now describe the large sample properties of our estimator. We make the following assumptions:

Assumption 1. $\{X_i, W_i, D_i, Y_i\}_{i=1}^n$ is a random sample generated from (1) and (2).

Assumption 2. (U,V) is jointly statistically independent of W and the bivariate distribution of (U,V) is absolutely continuous with respect to the Lebesgue measure with standard uniform marginals and rectangular support. We denote its c.d.f as $C^*(u,v,\rho_0)$. The copula function $C^*(u,v,\rho)$ is twice continuously differentiable with respect to ρ. X has a bounded support.[6]

Write $Y_2^* = W'\gamma(V)$ and let $p(w) = p(w_1,\tilde{w})$ denote the density of W and also define the conditional density of $\left(Y^*, Y_2^*\right)$ given $W = w$, by $g(y_1,y_2|w) = g(y_1,y_2|w_1,\tilde{w})$. And let $f_{Y^*}(y,w)$ denote the marginal density of Y^* given $W = w$.

Assumption 3. $g(y_1,y_2|w)$ is continuously differentiable in (y_1,y_2) with positive density on $\mathbb{R}^2$. In addition, $g(y_1,y_2|w_1,\tilde{w})$ and $p(w_1,\tilde{w})$ are (s_1+3)th order continuously differentiable with respect to w_1 and $g(y_1,y_2|w)$ is (s_2+1)th $((s_1+1)$ th$)$ order continuously differentiable with respect to $y_1\,(y_2)$. The weight function, $\omega(v)$, is also (s_1+3)th order continuously differentiable in v and $\omega(v) > 0$ for $v \in [v_l, v_u] \subseteq \mathcal{S}_{p(W)}$.

Assumption 4. (i) $\beta(\tau)$, $\gamma(v)$ and ρ_0 are interior points of compact set $B \times \Gamma \times \varrho$, respectively, for any $v \in [v_l, v_u]$ and $\tau \in [\tau_l^*, \tau_u^*]$; (ii) $\gamma(v) = (\gamma_1(v), \tilde{\gamma}'(v))'$ such that $\inf_{v\in[v_l,v_u]} |\gamma_1(v)| > 0$.

Following Volgushev (2020), with a slight abuse of notation, let $\gamma(v)$ denote the rescaled parameter $\gamma(v)/|\gamma_1(v)|$.

[6]More generally, we can assume $EX^p < \infty, p \geq 2$ without requiring that X has bounded support.

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

Assumption 5. The first stage SMSC $\hat{\gamma}(\cdot)$ satisfies the following asymptotic linear representation:

$$\left(\hat{\gamma}(v)-\gamma(v)\right)=-Q_v^{-1}\frac{1}{n}\sum_{i=1}^{n}(D_i-(1-v))\frac{1}{h}k\left(\frac{W_i'\gamma(v)}{h}\right)\tilde{W}_i+o_p\left(n^{-1/2}\right),$$

$$\sup_{v\in[v_l,v_u]}\left|\hat{\gamma}(v)-\gamma(v)\right|=O_p\left(\delta_{n\gamma}\right),$$

uniformly over $v\in[v_l,v_u]$, where $\delta_{n\gamma}=(nh)^{-1/2}\ln n+h^s$, k is sth order kernel such that $\int u^j k(u)du=0$ for $j<s$ and $\int u^j k(u)du$ is finite and nonzero for $j=s$. Q_v is defined in Horowitz (1992).

Assumption 6. $E\left[1\{V>v\}f_{Y^*}(X'b,W)XX'|W'\gamma(v)=0\right]$ is uniformly nonsingular over $(b,v)\in B\times[v_l,v_u]$ in that its minimum eigenvalue is bounded away from some $c_0>0$.

Assumption 7. The kernel functions k_1 and k_2 are three times differentiable with bounded support; k_1,k_2 and their derivative functions are uniformly bounded; k_1 and k_2 are s_1th and s_2th order kernel functions: for some s_1 and $s_2\geq 2$ and each integer $1\leq j_1\leq s_1$ and $1\leq j_2\leq s_2$,

$$\int u^{j_i}k_i(u)\,du=\begin{cases}=0, \text{ if } j_i<s_{i,}\\ \text{is finite, if } j_i=s_{i,}\end{cases}\quad i=1,2.$$

Define $\delta_{n1}=(nh_1)^{-1/2}\ln n+h_1^{s_1}+h_2^{s_2}$, $\delta_{n2}=\left(nh_1^3\right)^{-1/2}\ln n+h_1^{s_1}+h_2^{s_2},\delta_{n3}=(nh_1h_2)^{-1/2}\ln n+h_1^{s_1}+h_2^{s_2},\delta_{n4}=\left(nh_1h_2^3\right)^{-1/2}\ln n+h_1^{s_1}+h_2^{s_2}$.

Assumption 8. Bandwidth sequences satisfy (i) $h^s=o\left(n^{-1/2}\right)$, and $h_1^{s_1}+h_2^{s_2}=o\left(n^{-1/2}\right)$; (ii) $\delta_{n\gamma},\delta_{nj}=o(1)$ for $j=1,2,3$ satisfy $\delta_{n\gamma}\delta_{n2}=o\left(n^{-1/2}\right)$, $\delta_{n\gamma}h_1^{-1}=o(1)$, and $\left(\delta_{n1}+\delta_{n\gamma}\right)\delta_{n4}=o(1)$.

Let

$$\bar{G}_2(\rho)=\sum_{\tau_j\in\mathcal{J}}\int_{v_l}^{v_u}\left\|E\left[D\left[1\left\{Y\leq X'\beta\left(\tau_j,\rho\right)\right\}-C\left(\tau_j,v,\rho\right)\right]Z|W'\gamma(v)=0\right]\right.$$
$$\left.\times p_{W'\gamma(v)}(0)\right\|^2\omega(v)\,dv.$$

Assumption 9. ρ_0 is the unique minimizer of $\bar{G}_2(\rho)$.

Assumption 10. $\frac{d^2}{d\rho^2}\bar{G}_2(\rho_0)>0$.

Assumptions 1 and 2 describe the data generating mechanism. Assumption 3 provides some smoothness and boundedness conditions on the joint conditional distribution of Y^* and Y_2^* given the exogenous variables. The compactness condition in Assumption 4 is standard for extremum estimators. Regarding the binary selection equation, it is well known that some normalization is required for model identification (see, for example, Manski, 1975, 1985; Horowitz, 1992; Kordas, 2006; Volgushev, 2020). Here, for notational convenience, we adopt the

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

approach of Horowitz (1992), Kordas (2006), Manski (1975, 1985), and Volgushev (2020). Assumption 5 states that the asymptotic linear representation properties are satisfied by $\hat{\gamma}(v)$. For the SMSC estimator $\hat{\gamma}(v)$ proposed by Horowitz (1992), Theorem 2.5 of Volgushev (2020) provides a set of sufficient conditions for Assumption 5. Assumption 6 is the conditional full rank condition uniformly over (b, v) to ensure that $\hat{\beta}(\tau, \rho)$ satisfies a uniformly asymptotic linear representation. Assumptions 7 and 8 place some restrictions on the kernel function and the bandwidth sequences. We consider a Gaussian kernel for k_1 and set $s = 4, s_1 = s_2 = 2$, and $h = n^{-\nu}, h_1 = n^{-\nu_1}, h_2 = n^{-\nu_2}$, then, for example, $\nu = 2/15, 1/4 < \nu_1 < 13/45, \nu_2 = 1/3$ will be some suitable choices of bandwidths. Assumption 9 states the global identification condition for our moment-based estimator, similar to that of AB17. The proposition in the first appendix essentially provides conditions for global identification based on an infinite number of moment conditions. The nonzero nature of $\frac{d^2}{d\rho^2}\bar{G}_2(\rho_0)$ (full rankness) in Assumption 10 is the local identification condition for ρ_0 which is analogous to the Hessian form of the information matrix in the maximum likelihood estimation and other extremum estimators. In fact, Assumption 10 generally holds true since in the proofs of large sample properties, it is shown that

$$\frac{d^2}{d\rho^2}\bar{G}_2(\rho_0) = \sum_{\tau_j \in \mathcal{J}} \int_{v_l}^{v_u} \left[G_{2\rho}^2\left(\tau_j, v\right)\omega(v)\,dv\right],$$

where

$$G_{2\rho}(\tau, v) = \frac{d}{d\rho}G_2(\beta(\tau, \rho_0), \gamma(v), \rho_0, v, \tau),$$

with

$$\begin{aligned}
&G_2(\beta(\tau, \rho), \gamma(v), \rho, v, \tau) \\
&\quad = E\left[D_i\left[1\left\{Y_i \le X_i'\beta(\tau, \rho)\right\} - C(\tau, v, \rho)\right]Z_i | W_i'\gamma(v) = 0\right]p_{W'\gamma(v)}(0).
\end{aligned}$$

We define some notations before presenting the main theorem. Let $P_i = E(D_i|W_i)$ and

$$S(\beta(\tau, \rho)) = E\left[1\{V_i > 1 - P_i\}X_iX_i'f_{Y^*}(X_i'\beta(\tau, \rho), W_i)f_{Y_2^*}(0, W_i)\,\omega(1 - P_i)\right],$$

and

$$\begin{aligned}
\phi_{\beta 0i}(\rho, \tau) = {} & 1\{V_i > 1 - P_i\}\left[1\left\{Y_i - X_i'\beta(\tau, \rho) \le 0\right\} - C(\tau, 1 - P_i, \rho)\right] \\
& \times f_{Y_2^*}(0, W_i)X_i\omega(1 - P_i),
\end{aligned}$$

and

$$\phi_{\beta 1i}(\rho, \tau) = Q_{1i}(\tau, \rho)(D_i - P_i)\tilde{W}_i f_{Y_2^*}(0, W_i)\omega(1 - P_i),$$

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

where $Q_{1i}(\tau,\rho) = -Q_1(\rho,\tau,1-P_i)Q_{1-P_i}^{-1}$ and $f_{Y_2^*}(\cdot,w)$ is the conditional density function of Y_2^* given $W=w$ and

$$Q_1(\rho,\tau,v) = E\left[\frac{\partial}{\partial w_1}G_1(\beta(\tau,\rho),\rho,\tau,v,-\tilde{W}_i'\tilde{\gamma}(v),\tilde{W}_i)\right],$$

with

$$\begin{aligned} &G_1(b,\rho,\tau,v,w)\\ &\quad = E\left[D_i\left(1\left\{Y_i \le X_i'b\right\} - C(\tau,v,\rho)\right)|W=(w_1,\tilde{w})\right]p(w_1,\tilde{w})x(w_1,\tilde{w})\,\tilde{w}', \end{aligned}$$

where $X = x\left(W_1,\tilde{W}\right)$ such that $\left(X',Z'\right)' = \left(W_1,\tilde{W}'\right)'$. We can then define

$$\phi_{\beta i}(\rho,\tau) = S^{-1}(\beta(\tau,\rho))\left(\phi_{\beta 0i}(\tau,\rho)+\phi_{\beta 1i}(\tau,\rho)\right).$$

Furthermore, define

$$\begin{aligned} \phi_{0,\rho i} &:= \sum_{\tau_j\in\mathcal{J}} D_i\left[1\left\{Y_i \le X_i'\beta\left(\tau_j,\rho_0\right)\right\} - C(\tau_j,1-P_i,\rho_0)\right]Z_i\\ &\qquad \times G_{2\rho}(\tau_j,1-P_i)f_{Y_2^*}(0,W_i)\omega(1-P_i), \end{aligned}$$

and

$$\phi_{1,\rho i} := \sum_{\tau_j\in\mathcal{J}}\int_{v_l}^{v_u}\frac{\partial}{\partial\beta'}G_2\left(\beta\left(\tau_j,\rho_0\right),\gamma(v),\rho_0,v,\tau_j\right)G_{2\rho}(\tau_j,v)\phi_{\beta i}\left(\rho_0,\tau_j\right)\omega(v)\,dv,$$

and

$$\begin{aligned} \phi_{2,\rho i} &:= \sum_{\tau_j\in\mathcal{J}}\frac{\partial}{\partial\gamma'}G_2\left(\beta\left(\tau_j,\rho_0\right),\gamma(1-P_i),\rho_0,P_i,\tau_j\right)\\ &\qquad \times G_{2\rho}(\tau_j,1-P_i)Q_{1-P_i}^{-1}(D_i-P_i)\,\tilde{W}_i f_{Y_2^*}(0,W_i)\omega(1-P_i), \end{aligned}$$

where note that $\beta(\tau,\rho_0)=\beta(\tau)$. Finally, we can define

$$\phi_{\rho i} = \left(\sum_{\tau_j\in\mathcal{J}}\int_{v_l}^{v_u}\left[G_{2\rho}^2\left(\tau_j,v\right)\omega(v)\,dv\right]\right)^{-1}\left(\phi_{0,\rho i}+\phi_{1,\rho i}+\phi_{2,\rho i}\right),$$

and

$$\phi_{\beta i}(\tau) = \phi_{\beta i}(\rho_0,\tau)+\beta_\rho(\tau,\rho_0)\,\phi_{\rho i},$$

where

$$\beta_\rho(\tau,\rho) = \left(\frac{\partial\bar{G}_1(\tau,\rho)}{\partial\beta}\right)^{-1}\frac{\partial\bar{G}_1(\tau,\rho)}{\partial\rho},$$

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

and

$$\bar{G}_1(\tau,\rho) = E\left[\left(1\left\{Y_i \le X_i'\beta(\tau,\rho)\right\} - C(\tau,1-P_i,\rho)\right)1\{V_i > 1-P_i\}X_i f_{Y_2^*}(0,W_i)\,\omega(1-P_i)\right].$$

THEOREM 1. *If Assumptions 1–10 hold, then (i) $\hat{\rho}$ is consistent for ρ_0 and for any $\tau \in [\tau_l^*, \tau_u^*]$, and $\hat{\beta}(\tau,\hat{\rho})$ is also consistent for $\beta(\tau)$; (ii) furthermore, $\hat{\rho}$ and $\hat{\beta}(\tau,\hat{\rho})$ have the asymptotic linear representation:*

$$\sqrt{n}(\hat{\rho}-\rho_0) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_{\rho i} + o_p(1),$$

and

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

uniformly in $\tau \in [\tau_l^*, \tau_u^*]$. Consequently, $\hat{\rho}$ and $\hat{\beta}(\tau,\hat{\rho})$ are asymptotically normal with

$$\sqrt{n}(\hat{\rho}-\rho_0) \to^d N\left(0, E\left[\phi_{\rho i}\phi_{\rho i}'\right]\right),$$

and

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

Let $\theta_0 = (\beta(\tau), \rho_0)$, then $\hat{\theta}$ is consistent for θ_0 and jointly asymptotically normal with

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right) \to^d N\left(0, E\left[\phi_{\theta i}(\tau)\phi_{\theta i}'(\tau)\right]\right),$$

where $\phi_{\theta i}(\tau) = \left(\phi_{\beta i}'(\tau), \phi_{\rho i}'\right)'$.

For the purpose of carrying out large-sample statistical inference for $\beta(\tau)$, consistent estimates of the asymptotic variance and covariance matrix need to be constructed. From the proof of the theorem, we can show that all the components in the asymptotic variance can be consistently estimated by the sample analogue replacing the true value of the parameters by their consistent sample analogues. Specifically,

(i) Finite sample estimate of the influence function of $\hat{\gamma}(v)$:

$$\hat{\phi}_{n\gamma i}(v) = -\hat{Q}_{nv}^{-1}(D_i - (1-v))\frac{1}{h}k\left(\frac{W_i'\hat{\gamma}(v)}{h}\right)\tilde{W}_i,$$

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

where

$$\hat{Q}_{nv}^{-1} = \frac{1}{nh^2}\sum_{i=1}^{n}(D_i - (1-v))\,k'\left(\frac{W_i'\hat{\gamma}(v)}{h}\right)\tilde{W}_i\tilde{W}_i'.$$

(ii) Finite sample estimate of the linear representation of $\hat{\beta}(\tau,\rho)$:

$$\hat{\phi}_{n\beta i}\left(\hat{\rho},\tau\right) = \hat{S}_n^{-1}\left(\hat{\beta}\left(\tau,\hat{\rho}\right)\right)\left(\hat{\phi}_{\beta 0i}\left(\hat{\rho},\tau\right)+\hat{\phi}_{\beta 1i}\left(\hat{\rho},\tau\right)\right),$$

where

$$\hat{S}_n\left(\hat{\beta}\left(\tau,\hat{\rho}\right)\right) = \int_{v_l}^{v_u}\hat{S}_n(\hat{\beta}\left(\tau,\hat{\rho}\right),\hat{\gamma}(v))\omega(v)\,dv,$$

with

$$\hat{S}_n(\hat{\beta}\left(\tau,\hat{\rho}\right),\hat{\gamma}(v)) = \frac{1}{nh_1h_2}\sum_{i=1}^{n}D_iX_iX_i'k_2\left(\frac{X_i'\hat{\beta}\left(\tau,\hat{\rho}\right)-Y_i}{h_2}\right)k_1\left(\frac{W_i'\hat{\gamma}(v)}{h_1}\right),$$

$$\hat{\phi}_{\beta 0i}\left(\hat{\rho},\tau\right) = \int_{v_l}^{v_u}D_iX_i\left[K_2\left(\frac{X_i'\hat{\beta}\left(\tau,\hat{\rho}\right)-Y_i}{h_2}\right)-C(\tau,v,\hat{\rho})\right]$$
$$\times\frac{1}{h_1}k_1\left(\frac{W_i'\hat{\gamma}(v)}{h_1}\right)\omega(v)\,dv,$$

and

$$\hat{\phi}_{\beta 1i}\left(\hat{\rho},\tau\right) = \int_{v_l}^{v_u}\hat{Q}_{n1}(\hat{\rho},\tau,v)\phi_{\gamma i}(v)\,\omega(v)\,dv,$$

with

$$\hat{Q}_{n1}(\hat{\rho},\tau,v) = \frac{1}{n}\sum_{i=1}^{n}D_iX_i\tilde{W}_i\left[K_2\left(\frac{X_i'\hat{\beta}\left(\tau,\hat{\rho}\right)-Y_i}{h_2}\right)-C(\tau,v,\hat{\rho})\right]\frac{1}{h_1^2}k_1'\left(\frac{W_i'\hat{\gamma}(v)}{h_1}\right).$$

(iii) Finite sample estimate of the influence function of $\hat{\rho}$,

$$\hat{\phi}_{n\rho i} := \left(\sum_{\tau_j\in\mathcal{J}}\int_{v_l}^{v_u}\left[\hat{G}_{2n,\rho}^2\left(\tau_j,v\right)\omega(v)\,dv\right]\right)^{-1}\left(\hat{\phi}_{0,\rho i}+\hat{\phi}_{1,\rho i}+\hat{\phi}_{2,\rho i}\right),$$

where

$$\hat{G}_{2n,\rho}(\tau,v) = \frac{d}{d\rho}\hat{G}_{2n}\left(\hat{\beta}\left(\tau,\hat{\rho}\right),\gamma(v),\hat{\rho},v,\tau\right),$$

with

$$\hat{G}_{2n}\left(\hat{\beta}(\tau,\rho),\gamma(v),\rho,v,\tau\right)$$
$$= \frac{1}{nh_1}\sum_{i=1}^{n}D_i\left(K_2\left(\frac{X_i'\hat{\beta}(\tau,\rho)-Y_i}{h_2}\right)-C(\tau,v,\rho)\right)k_1\left(\frac{W'\hat{\gamma}(v)}{h_1}\right)Z_i.$$

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

$\hat{G}_{2n,\rho}(\tau,v)$ involves $\hat{\beta}_\rho(\tau,\rho)$ which is given by

$$\hat{\beta}_\rho(\tau,\rho)=\left(\frac{\partial \hat{Q}_{n0}(\rho,\tau)}{\partial\beta}\right)^{-1}\frac{\partial \hat{Q}_{n0}(\rho,\tau)}{\partial\rho},$$

where

$$\hat{Q}_{n0}(\rho,\tau)=\int_{v_l}^{v_u} Q_{n0}(\rho,\tau,v)\omega(v)\,dv,$$

with

$$\hat{Q}_{n0}(\rho,\tau,v)=\frac{1}{n}\sum_{i=1}^{n}D_iX_i\left[K_2\left(\frac{X_i'\hat{\beta}(\tau,\rho)-Y_i}{h_2}\right)-C(\tau,v,\rho)\right]\frac{1}{h_1}k_1\left(\frac{W_i'\hat{\gamma}(v)}{h_1}\right),$$

and furthermore,

$$\hat{\phi}_{0,\rho i}:=\sum_{\tau_j\in\mathcal{J}}\int_{v_l}^{v_u}D_iZ_i\left[K_2\left(\frac{X_i'\hat{\beta}(\tau_j,\hat{\rho})-Y_i}{h_2}\right)-C(\tau_j,v,\hat{\rho})\right]$$
$$\times\frac{1}{h_1}k_1\left(\frac{W_i'\hat{\gamma}(v)}{h_1}\right)\hat{G}_{2n,\rho}\left(\tau_j,v\right)\omega(v)\,dv,$$

$$\hat{\phi}_{1,\rho i}:=\sum_{\tau_j\in\mathcal{J}}\int_{v_l}^{v_u}\frac{\partial}{\partial\beta'}\hat{G}_{2n}\left(\hat{\beta}\left(\tau_j,\rho\right),\gamma(v),\rho,v,\tau_j\right)\hat{\phi}_{n\beta i}\left(\hat{\rho},\tau_j\right)\hat{G}_{2n,\rho}\left(\tau_j,v\right)\omega(v)\,dv,$$

and

$$\hat{\phi}_{2,\rho i}:=\sum_{\tau_j\in\mathcal{J}}\frac{\partial}{\partial\gamma'}\hat{G}_{2n}\left(\hat{\beta}\left(\tau_j,\rho\right),\gamma(v),\rho,v,\tau_j\right)\hat{\phi}_{n\gamma i}(v)\,\hat{G}_{2n,\rho}\left(\tau_j,v\right)\omega(v)\,dv.$$

(iv) Finite sample estimate of the influence function of $\hat{\beta}\left(\tau,\hat{\rho}\right)$:

$$\hat{\phi}_{n\beta i}(\tau):=\hat{\phi}_{n\beta i}\left(\hat{\rho},\tau\right)+\hat{\beta}_\rho\left(\tau,\hat{\rho}\right)\hat{\phi}_{n\rho i}.$$

Therefore, the asymptotic variance and covariance matrix of $\hat{\beta}\left(\tau,\hat{\rho}\right)$ and $\hat{\rho}$ can be consistently estimated by $\frac{1}{n}\sum_{i=1}^{n}\hat{\phi}_{n\beta i}(\tau)\hat{\phi}_{n\beta i}'(\tau)$ and $\frac{1}{n}\sum_{i=1}^{n}\hat{\phi}_{n\rho i}^2$, respectively.

The plug-in method of estimating asymptotic variance involves kernel estimation of the density and derivative functions, among other aspects. Alternatively, given the asymptotic distributional characterization in the main theorem, inference may also be conducted using bootstrap or subsampling methods, as suggested by Politis, Romano, and Wolf (1999). For the unconditional distribution or quantile function of the latent outcome Y^*, we recommend using the systematic resampling-based statistical inference methods developed by Chernozhukov, Fernández-Val, and Melly (2013) and Chernozhukov et al. (2020).

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

4. SIMULATION STUDIES

In this section, we report the results of some Monte Carlo experiments to demonstrate the finite sample performance of our estimator. In the simulations, we choose the Gaussian copula and Frank copula with positive and negative correlations. In the case of the Gaussian copula, we set the correlation coefficient $\rho_0 = 0.7$ and -0.7, respectively. For the Frank copula,

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

where we choose $\eta = -\ln(1-\gamma) = 6$ and -6, which correspond to 0.5141 and -0.5141 for the Kendall's tau, respectively.

We consider data generating processes (DGP) for both homoskedastic and heteroskedastic cases, including one homoskedastic design and two heteroskedastic designs. The heteroskedastic models captures the nature of quantile regression where both slope and intercept coefficients vary over quantiles.

DGP1: Homoskedastic model

$$\begin{aligned} Y_1^* &= X_1 + X_2 + 1 + \Phi^{-1}(U), \\ Y_2^* &= Z + 0.5X_1 + X_2 + \Phi^{-1}(V), \\ D &= 1\left\{Y_2^* > 0\right\}, Y_1 = Y_1^* D, \end{aligned}$$

where $X_1 \sim \chi^2_{(1)}, X_2 \sim U(0,1)$ and are mutually independent and the excluded variable $Z \sim N(0,1)$ is independent of (X_1, X_2). In this design, $\beta_1(\tau) = \beta_2(\tau) \equiv 1, \beta_3(\tau) = 1 + \Phi^{-1}(\tau)$. DGP2: Heteroskedastic model

$$\begin{aligned} Y_1^* &= X_1\Phi^{-1}(U) + X_2\left(1+\Phi^{-1}(U)\right) + 1 + \Phi^{-1}(U), \\ Y_2^* &= Z + X_1\left(1 + 0.5F_{t_{(3)}}^{-1}(V)\right) + X_2\left(1 + 0.5F_{t_{(3)}}^{-1}(V)\right) + 0.5F_{t_{(3)}}^{-1}(V) - 1, \\ D &= 1\left\{Y_2^* > 0\right\}, Y_1 = Y_1^* D, \end{aligned}$$

where the distribution of X_1, X_2, and Z is the same as in the homoskedastic case. And $F_{t_{(3)}}$ is the cdf of a t-distributed random variable with three degrees of freedom. In this design, $\beta_1(\tau) = \Phi^{-1}(\tau), \beta_2(\tau) = \beta_3(\tau) = 1 + \Phi^{-1}(\tau)$. In both DGPs, (U,V) are independent of all regressors (X_1, X_2, and Z) and follow either the Gaussian copula $C_G^*(\cdot,\cdot,\pm 0.7)$ or the Frank copula $C_F^*(\cdot,\cdot,\pm 6)$.

Based on the randomly generated samples, we calculate both the AB17 estimators and our estimators to compare their performance under both homoskedastic and heteroskedastic conditions. It is expected that the estimates of the quantile coefficients $\beta(\tau)$ will exhibit a large bias for AB17's rotated quantile regression even for large sample sizes, as the Probit model misspecifies the conditional probability in the heteroskedastic scenario.

Before presenting our simulation results, we first describe the implementation details of the estimation procedures.

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

For AB17's rotated QR (denoted by AB17):

1. For an one-parameter copula function $C^*(u,v,\rho)$, since for all $(u,v) \in (0,1)^2$,

$$u - C^*(u,1-v,\rho) = C^*(u,v,-\rho).$$

Suppose the true conditional probability function follows the Probit model in AB17, $P(W) = \Phi\left(W'\theta_0\right)$, under our selection equation (2), the moment condition for AB17's estimator becomes

$$\begin{aligned}
&\Pr\left(Y^* \le X'\beta(\tau)\,|D=1, W=w\right)\\
&\quad= \Pr\left(Y^* \le X'\beta(\tau)\,|W'\gamma(V) > 0, W=w\right)\\
&\quad= \Pr\left(Y^* \le X'\beta(\tau)\,|V > 1-\Phi\left(W'\theta_0\right), W=w\right)\\
&\quad= \frac{C^*\left(\tau, \Phi\left(w'\theta_0\right), -\rho_0\right)}{\Phi\left(w'\theta_0\right)}.
\end{aligned}$$

Thus, ρ_0 would exhibit opposite signs for our estimators and the AB17 estimators.

For our estimator (denoted by CZ):

1. Kernel functions: in estimating the binary quantile regression, a fourth-order Gaussian kernel is used in SMSC estimator,

$$K(t) = \Phi(t) + \frac{1}{2} t\phi(t),$$

where Φ and ϕ are the standard normal cdf and pdf, respectively. $k_1(\cdot)$ is standard Gaussian density function. For the choice of $k_2(\cdot)$, He et al. (2023) discussed in detail the computation aspect of convolution SQR. In particular, the smoothed objective function (11) can be explicitly expressed for several commonly used kernel functions, for example, Gaussian kernel and Epanechnikov kernel. In this simulation study, we utilize the Gaussian kernel for $k_2(\cdot)$, specifically,

Gaussian kernel: $k_2(t) = \phi(t)$, the resulting smoothed objective function is

$$\begin{aligned}
&M_{n1}(b,\gamma,\rho,\tau,v)\\
&\quad= \frac{1}{nh_1}\sum_{i=1}^{n} D_i \left(\frac{h_2}{2} M^G\left(\frac{Y_i - X_i'b}{h_2}\right) + \left(\tau - \frac{1}{2}\right)\left(Y_i - X_i'b\right)\right) k_1\left(\frac{W_i'\gamma}{h_1}\right),
\end{aligned}$$

where $M^G(u) = (2/\pi)^{1/2}\exp\left(-u^2/2\right) + u(1-2\Phi(-u))$.

2. Bandwidths: We set bandwidths according to a rule of thumb that satisfies Assumption 8: $h = 0.6 \times \mathrm{std}\left(W'\hat{\gamma}(v)\right) \times n^{-2/15}$, where $\hat{\gamma}(v)$ is the initial MSC estimates of $\gamma(v)$; $h_1 = c \times \mathrm{std}\left(W'\gamma(v)\right) \times n^{-11/40}$ (11/40=0.275) and $h_2 = n^{-1/3}$. To save space, we only report the numerical results for the following simulation designs: (i) Gaussian homoskedastic model ($\rho_0 = 0.7, c = 1.2$); (ii) Gaussian heteroskedastic model ($\rho_0 = -0.7, c = 1.2$); and Frank heteroskedastic model ($\rho_0 = -0.5141$(Kendall's tau)$, c = 1.2$). The results for the above

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

TABLE 1. Homoskedastic model: Gaussian copula ($\rho_0 = 0.7$), $n = 1{,}000$, $c = 1.2$

	Bias						SD					
τ	β_1		β_2		β_3		β_1		β_2		β_3	
	CZ	AB17	CZ	AB17	CZ	AB17	CZ	AB17	CZ	AB17	CZ	AB17
0.2	0.008	0.001	−0.008	−0.017	−0.062	0.003	0.106	0.037	0.297	0.205	0.252	0.167
0.3	0.006	−0.001	−0.004	−0.015	−0.049	0.004	0.094	0.034	0.253	0.183	0.208	0.144
0.4	0.006	−0.001	0.001	−0.010	−0.039	0.004	0.087	0.031	0.228	0.169	0.182	0.125
0.5	0.005	−0.001	0.002	−0.015	−0.028	0.008	0.082	0.031	0.223	0.164	0.164	0.117
0.6	0.005	−0.002	0.004	−0.011	−0.022	0.010	0.082	0.030	0.220	0.159	0.155	0.111
0.7	0.001	−0.003	0.008	−0.010	−0.013	0.010	0.084	0.031	0.220	0.163	0.149	0.107
0.8	−0.003	−0.004	0.008	−0.002	−0.006	0.007	0.089	0.033	0.236	0.169	0.151	0.110

	Bias		SD	
	CZ	AB17	CZ	AB17
ρ	0.024	0.001	0.127	0.099

designs using other choices of bandwidth ($c = 0.6, 1.8$) will be presented in the Supplementary Material correspondingly. Additional simulation results will be available upon request from the authors.

3. Optimization algorithm: For the binary quantile regression, we solve the SMSC using the Nelder–Mead algorithm. Since the SMSC is not globally convex, a good initialization is crucial for ensuring global convergence. We initialize the SMSC with Manski's MSC estimator using a simulated annealing algorithm with multiple starting points. Alternatively, the mixed-integer algorithm by Florios and Skouras (2008) offers another suitable option. Second, to compute $\beta(\tau,\rho)$, we basically use the Quasi-Newton algorithm, as the corresponding objective function is differentiable and convex. Finally, similar to Arellano and Bonhomme (2017), $\hat{\rho}$ is obtained through grid search.

In estimating $\beta(\tau,\rho)$, $v_l = \max\left\{0.1, 1-Q_{P(W)}(0.9)\right\}$, and $v_u = \min\left\{0.9, 1-Q_{P(W)}(0.1)\right\}$, thus we are pooling over several quantiles such that $v \in \mathcal{J}_0 = \{\tilde{v}_j\}_{j=1}^{L}$, where $\tilde{v}_1 = v_l$, $\tilde{v}_L = v_u$, and $\tilde{v}_j - \tilde{v}_{j+1} = 0.05$ for $j = 1,\ldots,L-1$ and further let $\mathcal{J} = \mathcal{J}_0$ for the estimate of ρ. We consider two sample sizes of $n = 1{,}000$ and $n = 2{,}000$ and each is replicated 500 times.

We now report the performance of our estimator $\hat{\rho}$ and $\hat{\beta}(\tau,\hat{\rho})$ for $\tau = 0.80, 0.70, 0.60, 0.50, 0.40, 0.30,$ and 0.20 to demonstrate how well our estimation procedure can recover the quantile coefficients at different quantile levels. In particular, we report the bias (Bias) and standard deviation (SD) for both estimators.

Tables 1 and 2 present the simulation results for the Gaussian homoskedastic designs, while Tables 3–6 display the results for the heteroskedastic designs with both Gaussian and Frank copulas. In all scenarios, our estimator performs well. When estimating the copula parameter η, the biases appear relatively small, but the

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

Table 2. Homoskedastic model: Gaussian copula ($\rho_0 = 0.7$), $n = 2{,}000$, $c = 1.2$

τ	Bias						SD					
	β_1		β_2		β_3		β_1		β_2		β_3	
	CZ	AB17	CZ	AB17	CZ	AB17	CZ	AB17	CZ	AB17	CZ	AB17
0.2	0.001	−0.001	−0.002	−0.005	−0.027	0.004	0.073	0.026	0.200	0.143	0.170	0.113
0.3	0.000	−0.001	0.005	−0.004	−0.025	0.003	0.065	0.023	0.177	0.128	0.147	0.097
0.4	−0.000	−0.002	0.009	−0.003	−0.021	0.003	0.061	0.022	0.166	0.120	0.132	0.088
0.5	−0.001	−0.002	0.006	−0.000	−0.015	0.002	0.058	0.021	0.160	0.114	0.118	0.080
0.6	−0.002	−0.001	−0.000	0.000	−0.008	0.002	0.055	0.021	0.157	0.112	0.109	0.079
0.7	−0.003	−0.003	0.003	0.005	−0.006	0.002	0.056	0.022	0.159	0.112	0.104	0.077
0.8	−0.005	−0.003	0.007	0.006	−0.005	0.001	0.059	0.024	0.164	0.113	0.104	0.075

	Bias		SD	
	CZ	AB17	CZ	AB17
ρ	0.011	0.002	0.089	0.070

Table 3. Heteroskedastic model: Gaussian copula ($\rho_0 = -0.7$), $n = 1{,}000$, $c = 1.2$

τ	Bias						SD					
	β_1		β_2		β_3		β_1		β_2		β_3	
	CZ	AB17	CZ	AB17	CZ	AB17	CZ	AB17	CZ	AB17	CZ	AB17
0.2	0.002	0.038	−0.027	−0.066	0.047	−0.004	0.232	0.176	0.433	0.435	0.276	0.276
0.3	−0.014	0.013	−0.041	−0.081	0.073	0.025	0.219	0.165	0.430	0.424	0.279	0.270
0.4	−0.012	−0.011	−0.039	−0.095	0.089	0.050	0.216	0.161	0.444	0.419	0.291	0.282
0.5	−0.015	−0.030	−0.024	−0.118	0.099	0.081	0.219	0.164	0.478	0.424	0.317	0.302
0.6	−0.009	−0.065	−0.030	−0.162	0.109	0.132	0.221	0.174	0.513	0.456	0.339	0.343
0.7	−0.013	−0.102	−0.031	−0.201	0.135	0.175	0.247	0.186	0.575	0.494	0.394	0.384
0.8	−0.007	−0.152	0.014	−0.288	0.129	0.263	0.296	0.212	0.703	0.619	0.468	0.508

	Bias		SD	
	CZ	AB17	CZ	AB17
ρ	−0.022	−0.053	0.113	0.104

standard errors are quite large. This can be misleading due to the extreme nonlinearity inherent in the Frank copula. To provide a clearer understanding, we convert the copula parameter to Kendall's tau, which offers a more intuitive representation of the copula structure, as Kendall's tau is a concrete and intuitive measure of dependence for any copula. The simulation results show that when the model does not exhibit heteroskedasticity, both our estimator and the AB17 estimator have

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

TABLE 4. Heteroskedastic model: Gaussian copula ($\rho_0 = -0.7$), $n = 2{,}000$, $c = 1.2$

τ	Bias						SD					
	β_1		β_2		β_3		β_1		β_2		β_3	
	CZ	AB17	CZ	AB17	CZ	AB17	CZ	AB17	CZ	AB17	CZ	AB17
0.2	−0.013	0.033	−0.006	−0.044	0.028	0.006	0.167	0.123	0.320	0.295	0.200	0.191
0.3	−0.015	0.013	−0.006	−0.078	0.035	0.038	0.156	0.118	0.327	0.279	0.204	0.192
0.4	−0.009	−0.008	−0.008	−0.105	0.046	0.068	0.156	0.116	0.331	0.292	0.216	0.206
0.5	−0.008	−0.029	−0.005	−0.129	0.054	0.098	0.150	0.115	0.337	0.304	0.226	0.223
0.6	−0.011	−0.056	0.005	−0.168	0.059	0.137	0.157	0.120	0.351	0.334	0.239	0.252
0.7	−0.019	−0.088	0.009	−0.228	0.068	0.190	0.171	0.132	0.396	0.369	0.276	0.286
0.8	−0.025	−0.134	−0.017	−0.266	0.097	0.244	0.208	0.148	0.493	0.432	0.358	0.346

	Bias		SD	
	CZ	AB17	CZ	AB17
ρ	−0.012	−0.046	0.078	0.068

TABLE 5. Heteroskedastic model: Frank copula ($\rho_0 = -0.5141$ (Kendall's tau)), $n = 1{,}000$, $c = 1.2$

τ	Bias						SD					
	β_1		β_2		β_3		β_1		β_2		β_3	
	CZ	AB17	CZ	AB17	CZ	AB17	CZ	AB17	CZ	AB17	CZ	AB17
0.2	−0.022	0.039	−0.049	−0.069	0.062	0.010	0.223	0.181	0.437	0.401	0.257	0.263
0.3	−0.026	0.025	−0.049	−0.108	0.074	0.044	0.204	0.172	0.415	0.382	0.263	0.262
0.4	−0.038	0.009	−0.062	−0.161	0.100	0.086	0.202	0.170	0.433	0.390	0.295	0.285
0.5	−0.033	−0.016	−0.068	−0.211	0.117	0.126	0.213	0.176	0.468	0.419	0.335	0.311
0.6	−0.032	−0.048	−0.066	−0.252	0.134	0.181	0.229	0.178	0.549	0.447	0.404	0.346
0.7	−0.036	−0.084	−0.051	−0.329	0.159	0.252	0.258	0.189	0.654	0.541	0.485	0.432
0.8	−0.025	−0.127	−0.008	−0.384	0.165	0.314	0.323	0.231	0.809	0.699	0.569	0.556

	Bias		SD	
	CZ	AB17	CZ	AB17
Kendall's τ	−0.018	−0.046	0.107	0.080
ρ	−0.822	−0.647	2.506	1.399

very small biases and standard deviations, with the AB17 estimator's bias being nearly nonexistent and its standard deviation smaller than ours.

Similarly, our estimator continues to perform well, for the heteroskedastic designs, with the exception of the intercept coefficients when the sample size n is small. As the sample size increases, the biases are nearly eliminated, accompanied

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

TABLE 6. Heteroskedastic model: Frank copula ($\rho_0 = -0.5141$ (Kendall's tau)), $n = 2{,}000$, $c = 1.2$

τ	Bias						SD					
	β_1		β_2		β_3		β_1		β_2		β_3	
	CZ	AB17	CZ	AB17	CZ	AB17	CZ	AB17	CZ	AB17	CZ	AB17
0.2	−0.008	0.034	0.007	−0.062	0.020	0.016	0.149	0.123	0.290	0.278	0.180	0.183
0.3	−0.012	0.022	−0.001	−0.085	0.033	0.043	0.141	0.117	0.281	0.271	0.178	0.186
0.4	−0.014	0.008	0.001	−0.120	0.041	0.076	0.142	0.115	0.291	0.280	0.192	0.200
0.5	−0.011	−0.017	−0.001	−0.165	0.050	0.120	0.146	0.118	0.326	0.312	0.215	0.227
0.6	−0.006	−0.050	−0.002	−0.212	0.056	0.170	0.158	0.124	0.366	0.338	0.250	0.266
0.7	−0.004	−0.086	−0.018	−0.263	0.079	0.216	0.185	0.132	0.432	0.398	0.312	0.322
0.8	−0.010	−0.127	−0.012	−0.326	0.104	0.269	0.227	0.156	0.590	0.491	0.394	0.391

	Bias		SD	
	CZ	AB17	CZ	AB17
Kendall's τ	−0.011	−0.044	0.067	0.057
ρ	−0.381	−0.707	1.366	0.984

by small standard deviations. However, the biases for AB17's estimator remain unchanged even when the sample size increases to 2,000. The simulation results in the supplement indicate that adjusting the bandwidth selection within a certain range does not affect the estimation results significantly.

5. EMPIRICAL ILLUSTRATION

We next apply our estimation method to study female wages in the US using a subset of the GSS data for women in the US for the years 2006, 2008, 2010, and 2012. Specifically, we use a randomly generated sample of 1,500 observations and estimate the model accordingly.

In this example, all the observations are female. We assume that the hourly wage is a function of education, age, and marital status, whereas the likelihood of working (the selection equation) is a function of the number of children at home (an excluded variable) and the three included variables in the wage equation. Table 7 reports the summary statistics of the sample. We estimate the conditional quantile function of women's latent wages using both AB17's quantile selection regression and our approach.

In this study, the conditional probability of labor participation follows a Probit model for AB17's selection equation. For our approach, we adopt a slightly different scale normalization in the estimation of the binary quantile selection equation. Specifically, the coefficients in the selection equation are standardized to have unit length, ensuring that $\|\gamma(v)\| = 1$ for all v, thereby circumventing the need

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

TABLE 7. Summary statistics

Variable	No. of obs.	Mean	SD	Min	Max
log-wage (observed)	1,500 (1111)	1.9851	0.7447	−2.5771	5.6649
age	1,500	41.5427	12.1925	18	65
edu (edu>0)	1,495	13.7010	3.0842	1	20
Married	1,500	0.4960	0.5002	0	1
No. of kids	1,500	0.8573	1.1755	0	8

to ascertain the sign of $\gamma_1(v)$ in advance. The selection of bandwidths and other computational techniques employed in the estimation procedure are identical to those used in the preceding simulations, namely, $h = 0.6\times\text{std}\big(W'\gamma(v)\big)\times n^{-2/15}$, $h_1 = 1.2\times\text{std}\big(W'\gamma(v)\big)\times n^{-0.275}$ and $h_2 = n^{-1/3}$; $\tau_l = \max\big\{0.1, 1-Q_{P(W)}(0.9)\big\}$, and $\tau_u = \min\big\{0.9, 1-Q_{P(W)}(0.1)\big\}$.

Following AB17, the unconditional cdf of the latent wage may be estimated as a discretized version of

$$\hat{F}_{Y*}(y) = \frac{1}{n}\sum_{i=1}^{n}\int_0^1 1\left\{X_i'\hat{\beta}(\tau)\le y\right\}d\tau.$$

The quantile of the latent wage is then obtained as $\hat{F}_{Y*}^{-}$ after applying a monotone rearrangement. Chernozhukov et al. (2013) and Chernozhukov et al. (2020) proposed a systematic treatment for the statistical inference of the counterfactual distribution or quantile function of the latent outcome, which can be readily applied to our model.[7] Finally, the empirical results for both Gaussian and Frank copula functions are reported below to address robustness concerns.

We first examine the results of the Gaussian copula model. Regarding the copula estimates, we obtain a Spearman's rank correlation (implied by the copula estimate) of 0.756. In contrast, Arellano and Bonhomme (2017) reports a rank correlation of −0.38, which potentially indicates smaller selectivity for women. We then discuss the estimation of the unconditional quantile function, as illustrated in Figure 1. Specifically, Figure 1a presents the result from the Gaussian copula model and Figure 1b shows the result obtained from the Frank copula model. Both models predict a positive selection for females; however, our model indicates a more pronounced selection bias. This is evident as our unconditional quantile curve of the latent log-wage (blue dashed line) is positioned significantly lower than the observed wage quantile curve (yellow solid line), compared to the unconditional quantile curve predicted by AB17's model (red solid line). This pattern remains robust even when we switch to the Frank copula function. The

[7] In this version, we did not incorporate and report the statistical inference for the distribution or quantile function of Y^*.

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

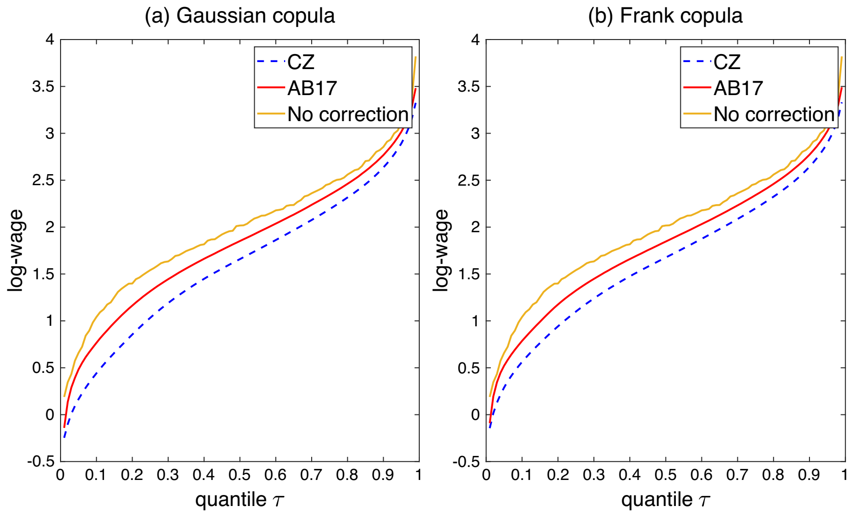

FIGURE 1. Quantile function of log-wage for female after no, AB17 and our (CZ) correction with Gaussian and Frank copula.

correlation parameters, Kendall's τ for the Frank copula are -0.25 for the AB17's model and 0.484 for our model.

6. CONCLUSION

In this article, we have considered semiparametric estimation of the quantile regression model subject to binary quantile sample selection, based on local moment conditions. Compared with Arellano and Bonhomme (2017), we extend the parametric selection equation to semiparametric quantile selection equation and provide a more coherent framework that allows for a general form of heterogeneity for both equations. We derive the large sample properties of our estimator, which performs comparably well in finite samples, especially in heteroskedastic cases where AB17's rotated quantile regression suffers from relatively large biases. We also apply our model and estimator to study the wage distribution among women using a randomly simulated sample from the U.S. GSS. The major finding is that there exists a significant positive selection among women in the US, which is notably more severe than the model prediction by AB17.

One important issue is the presence of endogeneity and censoring in the outcome equation. It is possible to extend the analysis in Arellano and Bonhomme (2017), Chernozhukov and Hansen (2006), Chen (2018), Chen and Wang (2023), and this article. It is an interesting topic for future research.

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

APPENDICES

Notations: Throughout the proofs, we use $d_{\tilde{w}}$ to denote $\dim\left(\tilde{W}\right)$, where $W = \left(W_1, \tilde{W}'\right)'$; γ_j represents the jth component of parameter vector γ. We use $\hat{\beta}_\rho(\tau,\rho)$ and $\hat{\beta}_{\rho\rho}(\tau,\rho)$ to denote $\frac{\partial}{\partial\rho}\hat{\beta}(\tau,\rho)$ and $\frac{\partial^2}{\partial\rho^2}\hat{\beta}(\tau,\rho)$ and similarly $C_\rho(\tau,v,\rho)$ and $C_{\rho\rho}(\tau,v,\rho)$ are used for $\frac{\partial}{\partial\rho}C(\tau,v,\rho)$ and $\frac{\partial^2}{\partial\rho^2}C(\tau,v,\rho)$; and $\delta_{n\gamma} = (nh)^{-1/2}\ln n + h^s$.

Appendix A: Identification

In this appendix, we show the point identification of $\beta(\tau)$ and ρ_0 for all $\tau \in (0,1)$.

Assumption ID.1. $\{X_i, W_i, D_i, Y_i\}_{i=1}^n$ is a random sample generated from (1) and (2).

Assumption ID.2. (U,V) is jointly statistically independent of W and the bivariate distribution of (U,V) is absolutely continuous with respect to the Lebesgue measure with standard uniform marginals and rectangular support. We denote its cdf as $C^*(u,v,\rho_0)$. The conditional distribution $F_{Y^*}(y,x)$ of Y^* given $X = x$ and its inverse $x'\beta(\tau)$ are strictly increasing for any given x. In addition, $C^*(u,v,\rho_0)$ is strictly increasing in u and v.

Assumption ID.3. For $u,\tilde{u} \in (0,1)$ and $\rho \in (-1,1)$, $C^*(u,v,\rho_0) = C^*(\tilde{u},v,\rho)$ for all $v \in (0,1)$, if and only if $u = \tilde{u}$ and $\rho = \rho_0$.

Assumption ID.4. Let $p_{W'\gamma(v)}$ denote the pdf of $W'\gamma(v)$, $E\left(D1\left\{W'\gamma(v) = 0\right\}X'X\right)$ is of full rank for all $v \in (0,1)$ such that $p_{W'\gamma(v)}(0) > 0$.

Assumption ID.3 is the key identification condition that conveys our identification strategy, which, in spirit, follows the common approach used for nonlinear regression models.

PROPOSITION. *Let the Assumptions ID.1–4 hold, and suppose $\gamma(v)$s are identified for all $v \in (0,1)$ such that $p_{W'\gamma(v)}(0) > 0$, then $\beta(\tau)$ and ρ_0 are identified for all $\tau \in (0,1)$.*

Proof. We define $I(\tau,b,\rho)$ and by the Assumptions ID.1–2, we have

$$\begin{aligned} &I(\tau,b,\rho) \\ &:= \int_{v_l}^{v_u} E\left[1\left\{W'\gamma(v) = 0\right\}E\left(D1\left\{Y \le X'b\right\}|X, W'\gamma(v) = 0\right) - (1-v)C(\tau,v,\rho)\right]^2 dv \\ &= \int_{v_l}^{v_u} (1-v)^2 p_{W'\gamma(v)}(0)\left(C\left(F_{Y^*}\left(X'b,X\right),v,\rho_0\right) - C(\tau,v,\rho)\right)^2 dv. \end{aligned}$$

It is straightforward to show that $I(\tau,\beta(\tau),\rho_0) = 0$. Now suppose $I(\tau,b,\rho) = 0$ for some $(b,\rho) \neq (\beta(\tau),\rho_0)$, then for any v such that $p_{W'\gamma(v)}(0) > 0$,

$$C\left(F_{Y^*}\left(X'b,X\right),v,\rho_0\right) = C(\tau,v,\rho)$$

thus

$$X'b = X'\beta(\tau) \text{ and } \rho = \rho_0$$

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

due to Assumption ID.3. Finally, identification follows from the full rank condition in Assumption ID.4. □

Appendix B: Large Sample Properties

For the convenience of readers, we restate the Lemma A1 in Carroll et al. (1997) which is used in the proof of two following lemmas:

LEMMA A0. *Let $\mathcal{C}$ and $\mathcal{D}$ be compact sets in $\mathbb{R}^d$ and $\mathbb{R}^p$, and let $f(x,\theta)$ be a continuous function in $\theta \in \mathcal{C}$ and $x \in \mathcal{D}$. Assume that $\hat{\theta}(x) \in \mathcal{C}$ is continuous in $x \in \mathcal{D}$ and is the unique maximizer of $f(x,\theta)$. Let $\hat{\theta}_n(x) \in \mathcal{C}$ be a maximizer of $f_n(x,\theta)$. If*

$$\sup_{\theta\in\mathcal{C},x\in\mathcal{D}} |f_n(x,\theta) - f(x,\theta)| \to 0,$$

then

$$\sup_{x\in\mathcal{D}} \left\| \hat{\theta}_n(x) - \hat{\theta}(x) \right\| \to 0,$$

as $n \to \infty$.

We now present the following two lemmas, which are useful for proving the main theorem.

LEMMA A1. *If Assumptions 1–9 are satisfied,*

(a). *$\beta(\tau,\rho)$ is uniquely identified for all $(\tau,\rho) \in [\tau_l^*,\tau_u^*] \times \varrho$, and uniformly over $(\tau,\rho) \in [\tau_l^*,\tau_u^*] \times \varrho$,*

$$\sup_{\tau\in[\tau_l^*,\tau_u^*]} \sup_{\rho\in\varrho} \left\| \hat{\beta}(\tau,\rho) - \beta(\tau,\rho) \right\| = o_p(1). \tag{a.1}$$

(b).

$$\sqrt{n}\left(\hat{\beta}(\tau,\rho) - \beta(\tau,\rho)\right) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \phi_{\beta i}(\rho,\tau) + o_p(1), \tag{a.2}$$

where

$$\phi_{\beta i}(\rho,\tau) = S^{-1}(\beta(\tau,\rho))\left[\phi_{\beta 0i}(\rho,\tau) + \phi_{\beta 1i}(\rho,\tau)\right],$$

with

$$\phi_{\beta 0i}(\rho,\tau) = D_i\left[1\left\{Y_i - X_i'\beta(\tau,\rho) \le 0\right\} - C(\tau, 1-P_i,\rho)\right] X_i f_{Y_2^*}(0,W_i)\omega(1-P_i),$$
$$\phi_{\beta 1i}(\rho,\tau) = Q_{1i}(\tau,\rho)(D_i - P_i) f_{Y_2^*}(0,W_i)\tilde{W}_i\omega(1-P_i),$$

and

$$S(b) = \int_{v_l}^{v_u} E\left[1\{V_i > v\} X_i X_i' f_{Y^*}(X_i'b, W_i) | W_i'\gamma(v) = 0\right] p_{W'\gamma(v)}(0)\omega(v)\,dv$$
$$= E\left[1\{V_i > 1-P_i\} X_i X_i' f_{Y^*}(X_i'b, W_i) f_{Y_2^*}(0,W_i)\,\omega(1-P_i)\right],$$

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

where $Q_{1i}(\tau,\rho) = -Q_1(\rho,\tau,1-P_i)Q^{-1}_{1-P_i}$ *and* $P_i = E(D_i|W_i)$ *and* $f_{Y^*}(\cdot,w)$ *denotes the density of* Y^* *given* $W = w$*;* $Y_2^* = W'\gamma(V)$ *and* $f_{Y_2^*}(\cdot,w)$ *denotes the conditional density function of* Y_2^* *given* $W = w$*, and*

$$Q_1(\rho,\tau,v) = E\left[\frac{\partial}{\partial w_1}G_1(\beta(\tau,\rho),\rho,\tau,v,-\tilde{W}_i'\tilde{\gamma}(v),\tilde{W}_i)\right],$$

and let $x = x(w_1,\tilde{w})$

$$G_1(b,\rho,\tau,v,w_1,\tilde{w}) = E\left[D_i\left(1\left\{Y_i \le X_i'b\right\} - C(\tau,v,\rho)\right)|W = (w_1,\tilde{w})\right]p(w_1,\tilde{w})x(w_1,\tilde{w})\,w'.$$

For the convenience of presenting Lemma 2, we also define

$$\begin{aligned}&\bar{G}_1(\tau,\rho)\\ &= E\left[1\{V_i > 1-P_i\}\left(1\left\{Y_i \le X_i'\beta(\tau,\rho)\right\} - C(\tau,1-P_i,\rho)\right)X_i f_{Y_2^*}(0,W_i)\,\omega(1-P_i)\right],\end{aligned}$$

and

$$\hat{Q}_{n0}(\rho,\tau) = \int_{v_l}^{v_u} Q_{n0}(\rho,\tau,v)\omega(v)\,dv,$$

with

$$\hat{Q}_{n0}(\rho,\tau,v) = \frac{1}{n}\sum_{i=1}^{n} D_i\left[K_2\left(\frac{X_i'\hat{\beta}(\tau,\rho)-Y_i}{h_2}\right) - C(\tau,v,\rho)\right]X_i\frac{1}{h_1}k_1\left(\frac{W_i'\hat{\gamma}(v)}{h_1}\right).$$

We further define $\hat{\beta}_\rho(\tau,\rho)$ and $\hat{\beta}_{\rho\rho}(\tau,\rho)$ by

$$\hat{\beta}_\rho(\tau,\rho) = \left(\frac{\partial\hat{Q}_{n0}(\rho,\tau)}{\partial\beta}\right)^{-1}\frac{\partial\hat{Q}_{n0}(\rho,\tau)}{\partial\rho},$$

$$\hat{\beta}_{\rho\rho}(\tau,\rho) = \left(\frac{\partial\hat{Q}_{n0}(\rho,\tau)}{\partial\beta}\right)^{-1}\left(\frac{\partial^2\hat{Q}_{n0}(\rho,\tau)}{\partial\rho} + \hat{\beta}_\rho'(\tau,\rho)\frac{\partial^2\hat{Q}_{n0}(\rho,\tau)}{\partial\beta\partial\beta'}\hat{\beta}_\rho(\tau,\rho)\right).$$

LEMMA A2. *If Assumptions 1–9 are satisfied, then uniformly over* $(\tau,\rho)\in\left[\tau_l^*,\tau_u^*\right]\times\varrho$

$$\hat{\beta}_\rho(\tau,\rho) - \beta_\rho(\tau,\rho) = O_p\left(\delta_{n\gamma} + (nh_1h_2)^{-1/2}\ln n + h_1^{s_1} + h_2^{s_2}\right), \quad \textbf{(a.3)}$$

and

$$\hat{\beta}_{\rho\rho}(\tau,\rho) - \beta_{\rho\rho}(\tau,\rho) = O_p\left(\delta_{n\gamma} + \left(nh_1h_2^3\right)^{-1/2}\ln n + h_1^{s_1} + h_2^{s_2}\right), \quad \textbf{(a.4)}$$

where by differentiating $\bar{G}_1(\tau,\rho)$ *with respect to* ρ *twice,*

$$\beta_\rho(\tau,\rho) = \left(\frac{\partial\bar{G}_1(\tau,\rho)}{\partial\beta}\right)^{-1}\frac{\partial\bar{G}_1(\tau,\rho)}{\partial\rho},$$

$$\beta_{\rho\rho}(\tau,\rho) = \left(\frac{\partial\bar{G}_1(\tau,\rho)}{\partial\beta}\right)^{-1}\left(\frac{\partial^2\bar{G}_1(\tau,\rho)}{\partial\rho} + \beta_\rho'(\tau,\rho)\frac{\partial^2\bar{G}_1(\tau,\rho)}{\partial\beta\partial\beta'}\beta_\rho(\tau,\rho)\right).$$

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

Proof of Lemma A1. To prove (a). Following the standard arguments for the quantile regression, the full rank Assumption 6 ensures that for any given $\tau \in \left[\tau_l^*, \tau_u^*\right]$, and $\rho \in \varrho$, $\int M_1\left(b, \gamma\left(v\right), \rho, \tau, v\right)\omega\left(v\right)dv$ is continuous in b with an the unique minimizer $\beta\left(\tau, \rho\right)$.

Define the class of functions

$$\mathcal{F}_1 = \Big\{m(D, Y, W; h_1, h_2, b, \gamma, \rho, \tau, v) \colon h_1, h_2 \in \mathbb{R}^+, \\ b \in B, \gamma \in G, \rho \in \varrho, \tau, v \in \left[\tau_l^*, \tau_u^*\right] \times [\tau_l, \tau_u]\Big\},$$

where

$$\begin{aligned} &m(D, Y, W; h_1, h_2, b, \gamma, \rho, \tau, v) \\ &= \frac{D}{h_2} k_1\left(\frac{W'\gamma}{h_1}\right)\int_{-\infty}^{\infty}\left[\rho_{C(\tau,v,\rho)}\left(u\right) - \rho_{C(\tau,v,\rho)}\left(u + X'b\right)\right] k_2\left(\frac{u - \left(Y - X'b\right)}{h_2}\right)du \\ &= Dk_1\left(\frac{W'\gamma}{h_1}\right)\int_{-\infty}^{\infty}\left[\rho_{C(\tau,v,\rho)}\left(Y - X'b + sh_2\right) - \rho_{C(\tau,v,\rho)}\left(Y + sh_2\right)\right] k_2\left(s\right)ds. \end{aligned}$$

First, note that

$$\mathcal{F}_K = \left\{k_1\left(\frac{W'\gamma}{h_1}\right) : \gamma \in G, h_1 \in \mathbb{R}^+\right\}$$

is Euclidean for the constant envelope (Example 2.10, Pakes and Pollard, 1989). Next, following Lemma 6 in Chetverikov, Larsen, and Palmer (2016), shows that

$$\mathcal{F}_\rho = \left\{\rho_\upsilon\left(Y - X'b - s\right) - \rho_\upsilon\left(Y - s\right) : s \in \mathbb{R}, b \in B, \upsilon \in (0, 1)\right\},$$

is VC subgragh with a bounded envelope function $\left(\|X\|^2 \sup_{b\in B}\|b\|^2\right)^{1/2}$ as X has a bounded support and thus $\mathcal{F}_\rho$ is an Euclidean class with a bounded envelope (Lemma 2.12, Pakes and Pollard, 1989), and by Corollary 21 in Nolan and Pollard (1987) and Lemma 2.14 in Pakes and Pollard (1989), we can deduce that $\mathcal{F}_1$ is Euclidean with a bounded envelope. Consequently, use the uniform law of large numbers in Pollard (1995) (Example 2 on page 273), we can deduce that

$$M_{n1}\left(b, \gamma, \rho, \tau, v\right) - M_1\left(b, \gamma, \rho, \tau, v\right) = O_p\left(\left(nh_1\right)^{-1/2}\ln n\right), \quad \textbf{(a.5)}$$

uniformly over $(b, \gamma, \rho, \tau, v)$, where[8]

$$\begin{aligned} &M_{n1}\left(b, \gamma, \rho, \tau, v\right) \\ &= \frac{1}{nh_1h_2}\sum_{i=1}^{n}\int_{-\infty}^{\infty} D\left[\rho_{C(\tau,v,\rho)}\left(u\right) - \rho_{C(\tau,v,\rho)}\left(u + X'b\right)\right] \\ &\quad \times k_2\left(\frac{u - \left(Y - X'b\right)}{h_2}\right)k_1\left(\frac{W'\gamma}{h_1}\right)du \\ &= \frac{1}{nh_1}\sum_{i=1}^{n} Dk_1\left(\frac{W'\gamma}{h_1}\right)\int_{-\infty}^{\infty}\left[\rho_{C(\tau,v,\rho)}\left(Y - X'b + sh_2\right) - \rho_{C(\tau,v,\rho)}\left(Y + sh_2\right)\right]k_2\left(s\right)ds, \end{aligned}$$

[8]The definition of M_{n1} is slightly different from the original definition in the main body of the article. However, they are equivalent in the sense that $\rho_{C(\tau,\tau_0,\rho)}\left(Y + sh_2\right)$ does not involve the parameter b, and thus can be suppressed.

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

and

$$M_1(b,\gamma,\rho,\tau,v) = E[M_{n1}(b,\gamma,\rho,\tau,v)].$$

In addition, under Assumptions 3 and 7, we have

$$\begin{aligned}
&E[M_{n1}(b,\gamma,\rho,\tau,v)] \\
&= \int_{-\infty}^{\infty} E\left[D\rho_{C(\tau,v,\rho)}\left(Y - X'b + sh_2\right)\frac{1}{h_1}k_1\left(\frac{W'\gamma}{h_1}\right)\right]k_2(s)\,ds \\
&- \int_{-\infty}^{\infty} E\left[D\rho_{C(\tau,v,\rho)}\left(Y + sh_2\right)\frac{1}{h_1}k_1\left(\frac{W'\gamma}{h_1}\right)\right]k_2(s)\,ds \\
&= \int_{-\infty}^{\infty} E\left[D\rho_{C(\rho,\tau,v)}\left(Y - X'b + sh_2\right)|W'\gamma = 0\right]p_{W'\gamma}(0)k_2(s)\,ds \\
&- \int_{-\infty}^{\infty} E\left[D\rho_{C(\rho,\tau,v)}\left(Y + sh_2\right)|W'\gamma = 0\right]p_{W'\gamma}(0)k_2(s)\,ds + O(h_1^{s_1}) \\
&= \int_{-\infty}^{\infty} E\left[D\rho_{C(\rho,\tau,v)}\left(Y - X'b\right)|W'\gamma = 0\right]p_{W'\gamma}(0)k_2(s)\,ds \\
&- \int_{-\infty}^{\infty} E\left[D\rho_{C(\rho,\tau,v)}\left(Y\right)|W'\gamma = 0\right]p_{W'\gamma}(0)k_2(s)\,ds + O(h_1^{s_1} + h_2^{s_2}),
\end{aligned} \tag{a.6}$$

where $p_{W'\gamma}(0) = p\left(-\tilde{w}'\gamma,\tilde{w}\right)$. Let

$$\begin{aligned}
M_1^0(b,\gamma,\rho,\tau,v) = {}& E\left[D\rho_{C(\tau,v,\rho)}\left(Y - X'b\right)|W'\gamma = 0\right]p_{W'\gamma}(0) \\
& - E\left[D\rho_{C(\tau,v,\rho)}\left(Y\right)|W'\gamma = 0\right]p_{W'\gamma}(0),
\end{aligned}$$

with

$$\begin{aligned}
&M_1^0(b,\gamma(v),\rho,\tau,v) \\
&= E\left[1\{V > v\}\rho_{C(\tau,v,\rho)}\left(Y - X'b\right)|W'\gamma(v) = 0\right]p_{W'\gamma(v)}(0) \\
&\quad - E\left[1\{V > v\}\rho_{C(\tau,v,\rho)}\left(Y\right)|W'\gamma(v) = 0\right]p_{W'\gamma(v)}(0).
\end{aligned}$$

Then, (a.5) and (a.6), together with Assumption 5, imply that

$$\begin{aligned}
&\int_{v_l}^{v_u} M_{n1}\left(b,\hat{\gamma}(v),\rho,\tau,v\right)\omega(v)\,dv \\
&= \int_{v_l}^{v_u} M_1^0(b,\gamma(v),\rho,\tau,v)\,\omega(v)\,dv + O_p\left(\delta_{n\gamma} + (nh_1)^{-1/2}\ln n + h_1^{s_1} + h_2^{s_2}\right),
\end{aligned}$$

uniformly in (b,ρ,τ). And $\int_{v_l}^{v_u} M_1^0(b,\gamma(v),\rho,\tau,v)\,\omega(v)\,dv$ is continuous in b with a unique minimizer $\beta(\tau,\rho)$ as all terms involving with $\rho_{C(\tau,v,\rho)}(Y)$ are independent of b. Thus, an application of the standard argument for consistency (e.g., Newey and McFadden, 1994) implies

$$\hat{\beta}(\tau,\rho) - \beta(\tau,\rho) = o_p(1).$$

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

Furthermore, by Lemma A0 (Lemma A1 in Carroll et al., 1997), we obtain the uniform consistency

$$\sup_{\tau\in\left[\tau_l^*,\tau_u^*\right]}\sup_{\rho\in\varrho}\left\|\hat{\beta}(\tau,\rho)-\beta(\tau,\rho)\right\|=o_p(1).$$

Next, we proceed to prove (b), the asymptotic linearity (a.2). The first-order condition of the objective function (10) with respect to $\hat{\beta}(\tau,\rho)$, obtains the estimating equation,

$$\frac{1}{n}\sum_{i=1}^{n}\int_{v_l}^{v_u}D_iX_i\left(K_2\left(\frac{X_i'\hat{\beta}(\tau,\rho)-Y_i}{h_2}\right)-C(\tau,v,\rho)\right)\frac{1}{h_1}k_1\left(\frac{W_i'\hat{\gamma}(v)}{h_1}\right)\omega(v)\,dv=0. \tag{a.7}$$

Then, since the above uniform consistency result of $\hat{\beta}(\tau,\rho)$, a mean-value expansion of (a.7) with respect to $\hat{\beta}(\tau,\rho)$ at $\beta(\tau,\rho)$, yields

$$\hat{\beta}(\tau,\rho)-\beta(\tau,\rho)=-S_n^{-1}(\hat{b}^*(\tau,\rho))Q_n(\beta(\tau,\rho),\rho,\tau),$$

where $\hat{b}^*(\tau,\rho)$ lies on the line segment between $\hat{\beta}(\tau,\rho)$ and $\beta(\tau,\rho)$ and

$$S_n(b)=\int_{v_l}^{v_u}S_n(b,\hat{\gamma}(v))\omega(v)\,dv,$$

and

$$Q_n(b,\rho,\tau)=\int_{v_l}^{v_u}Q_n(b,\hat{\gamma}(v),\rho,\tau,v)\omega(v)\,dv,$$

where

$$S_n(b,\gamma)=\frac{1}{n}\sum_{i=1}^{n}D_iX_iX_i'\frac{1}{h_1h_2}k_2\left(\frac{X_i'b-Y_i}{h_2}\right)k_1\left(\frac{W_i'\gamma}{h_1}\right),$$

and

$$Q_n(b,\gamma,\rho,\tau,v)=\frac{1}{n}\sum_{i=1}^{n}D_iX_i\left(K_2\left(\frac{X_i'b-Y_i}{h_2}\right)-C(\tau,v,\rho)\right)\frac{1}{h_1}k_1\left(\frac{W_i'\gamma}{h_1}\right).$$

Then, similar to (a.5) and (a.6), the following class of functions:

$$\mathcal{F}_2=\left\{s(D,Y,W;h_1,h_2,b,\gamma):h_1,h_2\in\mathbb{R}^+,b\in B,\gamma\in G\right\},$$

with

$$s(D,Y,W;h_1,h_2,b,\gamma)=DXX'k_2\left(\frac{X'b-Y}{h_2}\right)k_1\left(\frac{W'\gamma}{h_1}\right).$$

Thus similar to $\mathcal{F}_1$, $\mathcal{F}_2$ is Euclidean with a bounded envelope function. Hence, using the uniform law of large numbers in Pollard (1995), under Assumptions 3 and 7, we can show that uniformly in (b,γ)

$$S_n(b,\gamma)=S(b,\gamma)+O_p\left((nh_1h_2)^{-1/2}\ln n+h_1^{s_1}+h_2^{s_2}\right), \tag{a.8}$$

where

$$S(b,\gamma)=E\left[D_iX_iX_i'f_{Y^*}(X_i'b,W_i)|W_i'\gamma=0\right]p_{W'\gamma}(0), \tag{a.9}$$

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

where $p_{W'\gamma}(0)$ denotes the density of $W'\gamma$ at 0. Using (a.8), (a.9), (a.1) of Lemma A1, and Assumption 5,

$$\begin{aligned} S_n(\hat{b}^*(\tau,\rho),\hat{\gamma}(v)) - S(\beta(\tau,\rho),\gamma(v)) &= S_n(\hat{b}^*(\tau,\rho),\hat{\gamma}(v)) - S(\hat{b}^*(\tau,\rho),\hat{\gamma}(v)) \\ &\quad + S(\hat{b}^*(\tau,\rho),\hat{\gamma}(v)) - S(\beta(\tau,\rho),\gamma(v)) \\ &= o_p(1), \end{aligned}$$

uniformly over (ρ,τ,v). Thus

$$S_n\left(\hat{b}^*(\tau,\rho)\right) = S(\beta(\tau,\rho)) + o_p(1),$$

uniformly over (ρ,τ).

Next, we consider $Q_n(\beta(\tau,\rho),\rho,\tau)$, where

$$Q_n(b,\rho,\tau) = \int_{v_l}^{v_u} Q_n(\rho,b,\hat{\gamma}(v),\tau,v)\omega(v)\,dv,$$

with

$$Q_n(\rho,b,\gamma,\tau,v) = \frac{1}{n}\sum_{i=1}^{n} D_iX_i\left[K_2\left(\frac{X_i'b-Y_i}{h_2}\right) - C(\tau,v,\rho)\right]\frac{1}{h_1}k_1\left(\frac{W_i'\gamma}{h_1}\right).$$

Note that $Q_n(\beta(\tau,\rho),\rho,\tau)$ can be further decomposed into four components after a Taylor expansion with respect to γ at $\gamma(v)$,

$$Q_n(\beta(\tau,\rho),\rho,\tau) = Q_{n0}(\rho,\tau) + Q_{n1}(\rho,\tau) + Q_{n2}(\rho,\tau) + Q_{n3}(\rho,\tau),$$

where

$$Q_{n0}(\rho,\tau) = \int_{v_l}^{v_u} Q_{n0}(\rho,\tau,v)\omega(v)\,dv,$$

and

$$Q_{n1}(\rho,\tau) = \int_{v_l}^{v_u} Q_{n1}(\rho,\tau,v)\left(\hat{\gamma}(v)-\gamma(v)\right)\omega(v)\,dv,$$

and

$$Q_{n2}(\rho,\tau) = \int_{v_l}^{v_u} Q_{n2}(\rho,\tau,v)\omega(v)\,dv,$$

and

$$Q_{n3}(\rho,\tau) = \int_{v_l}^{v_u} Q_{n3}(\rho,\tau,v)\omega(v)\,dv,$$

with

$$Q_{n0}(\rho,\tau,v) = \frac{1}{n}\sum_{i=1}^{n} D_iX_i\left[K_2\left(\frac{X_i'\beta(\tau,\rho)-Y_i}{h_2}\right) - C(\tau,v,\rho)\right]\frac{1}{h_1}k_1\left(\frac{W_i'\gamma(v)}{h_1}\right),$$

$$Q_{n1}(\rho,\tau,v) = \frac{1}{n}\sum_{i=1}^{n} D_iX_i\tilde{W}_i'\left[K_2\left(\frac{X_i'\beta(\tau,\rho)-Y_i}{h_2}\right) - C(\tau,v,\rho)\right]\frac{1}{h_1^2}k_1'\left(\frac{W_i'\gamma(v)}{h_1}\right),$$

$$Q_{n2}(\rho,\tau,v) = \frac{1}{n}\sum_{i=1}^{n} D_i X_i \left[K_2\left(\frac{X_i'\beta(\tau,\rho) - Y_i}{h_2}\right) - C(\tau,v,\rho)\right] \times \frac{1}{h_1^3} k_1''\left(\frac{W_i'\gamma(v)}{h_1}\right)(W_i'\hat{\gamma}(v) - W_i\gamma(v))^2,$$

and

$$Q_{n3}(\rho,\tau,v) = \frac{1}{n}\sum_{i=1}^{n} D_i X_i \left[K_2\left(\frac{X_i'\beta(\tau,\rho) - Y_i}{h_2}\right) - C(\tau,v,\rho)\right] \times \frac{1}{h_1^4} k_1'''\left(\frac{W_i'\bar{\gamma}(v)}{h_1}\right)(W_i'\hat{\gamma}(v) - W_i'\gamma(v))^3,$$

where $\bar{\gamma}(v)$ is on the line segment between $\gamma(v)$ and $\hat{\gamma}(v)$.

i. First, we consider $Q_{n0}(\rho,\tau)$: Note that $Y_2^* = W'\gamma(V)$, and let $F_{Y_2^*}(y_2,w)$ denote the cdf of Y_2^* evaluated at y_2, given $W = w$, then we have $F_{Y_2^*}(w'\gamma(v), w) = v$ or $w'\gamma(v) = F_{Y_2^*}^{-1}(v,w)$, with $P(w) := E[D|W=w] = 1 - F_{Y_2^*}(0,w)$. With a change of variable $\frac{W_i'\gamma(v)}{h_1} = u_1$, or $v = F_{Y_2^*}(h_1u_1, W_i)$, we have

$$Q_{n0}(\rho,\tau) = \frac{1}{n}\sum_{i=1}^{n}\int D_i\left[K_2\left(\frac{X_i'\beta(\tau,\rho) - Y_i}{h_2}\right) - C(\tau, F_{Y_2^*}(h_1u_1,W_i),\rho)\right] X_i \times k_1(u_1)\,\omega\left(F_{Y_2^*}(h_1u_1,W_i)\right) f_{Y_2^*}(h_1u_1,W_i))\,du_1.$$

Let

$$q_0(D,Y,W;h_1,h_2,b,\rho,\tau) = \int \begin{pmatrix} D_i\left[K_2\left(\frac{X_i'b - Y_i}{h_2}\right) - C(\tau, F_{Y_2^*}(h_1u_1,W_i),\rho)\right] \\ \times X_i k_1(u_1)\,\omega\left(F_{Y_2^*}(h_1u_1,W_i)\right) f_{Y_2^*}(h_1u_1,W_i)) \end{pmatrix} du_1,$$

and also note that

$$K_2\left(\frac{X_i'\beta(\tau,\rho) - Y_i}{h_2}\right) = \int_{-\infty}^{\frac{X_i'\beta(\tau,\rho)-Y_i}{h_2}} k_2(u_2)du_2 = \int 1\left\{Y_i \le X_i'\beta(\tau,\rho) - h_2u_2\right\} k_2(u_2)du_2.$$

Define

$$\mathcal{F}_3 = \left\{q_0(D,Y,W;h_1,h_2,b,\rho,\tau) \colon h_1,h_2 \in \mathbb{R}^+, b \in B, \rho \in \varrho, \tau \in \left[\tau_l^*, \tau_u^*\right]\right\},$$

and

$$\mathcal{F}_3^* = \left\{q_0(D,Y,W;\rho,\tau) \colon \rho \in \varrho, \tau \in \left[\tau_l^*, \tau_u^*\right]\right\},$$

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

and

$$\bar{Q}_{n0}(\rho,\tau) = \frac{1}{n}\sum_{i=1}^{n} q_0(D_i, Y_i, W_i; \rho, \tau),$$

where

$$q_0(D,Y,W;\rho,\tau) = D\left[1\left\{Y - X'\beta(\tau,\rho) \leq 0\right\} - C(\tau, 1-P, \rho)\right] f_{Y_2^*}(0,W) X\omega(1-P).$$

Similar to $\mathcal{F}_1$ and $\mathcal{F}_2$, we can show $\mathcal{F}_3$ and $\mathcal{F}_3^*$ are Euclidean with a bounded envelope. Then, by Theorem 4.4 in Pollard (1989), we obtain uniformly over (ρ,τ)

$$(Q_{n0}(\rho,\tau) - Q_0(\rho,\tau)) - \left(\bar{Q}_{n0}(\rho,\tau) - \bar{Q}_0(\rho,\tau)\right) = o_p\left(n^{-1/2}\right), \quad \textbf{(a.10)}$$

where $Q_0(\rho,\tau) = EQ_{n0}(\rho,\tau)$ and $\bar{Q}_0(\rho,\tau) = E\bar{Q}_{n0}(\rho,\tau) = 0$. Furthermore, using Assumption 8, we have $Q_0(\rho,\tau) - \bar{Q}_0(\rho,\tau) = O\left(h_1^{s_1} + h_2^{s_2}\right) = o\left(n^{-1/2}\right)$ uniformly in (ρ,τ). Hence $Q_{n0}(\rho,\tau) = \bar{Q}_{n0}(\rho,\tau) + o_p\left(n^{-1/2}\right)$ uniformly over (ρ,τ).

ii. Now we consider $Q_{n1}(\rho,\tau)$. Note that

$$Q_{n1}(\rho,\tau) = \frac{1}{n}\sum_{i=1}^{n}\int_{v_l}^{v_u} Q_{n1}(\rho,\tau,v)\left(\hat{\gamma}(v) - \gamma(v)\right)\omega(v)\,dv,$$

where

$$Q_{n1}(\rho,\tau,v) = \frac{1}{n}\sum_{i=1}^{n} D_i X_i \tilde{W}_i\left[K_2\left(\frac{X_i'\beta(\tau,\rho) - Y_i}{h_2}\right) - C(\tau,v,\rho)\right]\frac{1}{h_1^2}k_1'\left(\frac{W_i'\gamma(v)}{h_1}\right).$$

Note that

$$\begin{aligned} &G_1(b,\rho,\tau,v,w_1,\tilde{w}) \\ &\quad = E\left[D_i\left(1\left\{Y_i < X_i'b\right\} - C(\tau,v,\rho)\right)|W = (w_1,\tilde{w})\right]p(w_1,\tilde{w})x(w_1,\tilde{w})\tilde{w}'. \end{aligned}$$

Similar to (a.5), we can show that

$$Q_{n1}(\rho,\tau,v) = Q_1(\rho,\tau,v) + O_p\left(\left(nh_1^3\right)^{-1/2}\ln n + h_1^{s_1} + h_2^{s_2}\right),$$

where

$$Q_1(\rho,\tau,v) = E\left[\frac{\partial}{\partial w_1}G_1(\beta(\tau,\rho),\rho,\tau,v,-\tilde{W}_i'\tilde{\gamma}(v),\tilde{W}_i)\right],$$

which, together with Assumptions 5 and 8, implies

$$\begin{aligned} &Q_{n1}(\rho,\tau,v)\left(\hat{\gamma}(v) - \gamma(v)\right) \\ &= Q_1(\rho,\tau,v)\left(\hat{\gamma}(v) - \gamma(v)\right) + o_p\left(n^{-1/2}\right) \\ &= -\frac{1}{n}\sum_{i=1}^{n} Q_1(\rho,\tau,v)Q_v^{-1}\left(D_i - (1-v)\right)\frac{1}{h}k\left(\frac{W_i'\gamma(v)}{h}\right)\tilde{W}_i + o_p\left(n^{-1/2}\right). \end{aligned}$$

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

Furthermore, similar to (a.10) in (i), using Assumption 8, we can show that uniformly over (τ,ρ),

$$
\begin{aligned}
&-\frac{1}{n}\sum_{i=1}^{n}\int_{v_l}^{v_u} Q_1(\rho,\tau,v)Q_v^{-1}\left(D_i-(1-v)\right)\frac{1}{h}k\left(\frac{W_i'\gamma(v)}{h}\right)\tilde{W}_i\omega(v)\,dv\\
&=\frac{1}{n}\sum_{i=1}^{n}Q_{1i}(\tau,\rho)\left(D_i-P_i\right)\tilde{W}_i f_{Y_2^*}(0,W_i)\omega(1-P_i)+O\left(h^s\right)+o_p\left(n^{-1/2}\right)\\
&:=\frac{1}{n}\sum_{i=1}^{n}q_1\left(D_i,W_i;\rho,\tau\right)+o_p\left(n^{-1/2}\right),
\end{aligned}
$$

where $Q_{1i}(\tau,\rho)=-Q_1(\rho,\tau,1-P_i)Q_{1-P_i}^{-1}$.

iii. We now consider $Q_{n2}(\rho,\tau)$ and $Q_{n3}(\rho,\tau)$ and show they are negligible. Note that

$$
\begin{aligned}
Q_{n2}(\rho,\tau,v)=&\sum_{j,k}^{d_{\tilde{w}}}\left(\hat{\gamma}(v)-\gamma(v)\right)_j\left(\hat{\gamma}(v)-\gamma(v)\right)_k\\
&\times\frac{1}{n}\sum_{i=1}^{n}q_{2ijk}(\beta(\tau,\rho),\gamma(v),\rho,\tau,v),
\end{aligned}
$$

and

$$
\begin{aligned}
Q_{n3}(\rho,\tau,v)=&\sum_{j,k,l}^{d_{\tilde{w}}}\left(\hat{\gamma}(v)-\gamma(v)\right)_j\left(\hat{\gamma}(v)-\gamma(v)\right)_k\left(\hat{\gamma}(v)-\gamma(v)\right)_l\\
&\times\frac{1}{n}\sum_{i=1}^{n}q_{3ijkl}(\beta(\tau,\rho),\bar{\gamma}(v),\rho,\tau,v),
\end{aligned}
$$

where $\bar{\gamma}(v)$ lies on the line segment between $\hat{\gamma}(v)$ and $\gamma(v)$ and

$$
q_{2ijk}(b,\gamma,\rho,\tau,v)=D_i\left[K_2\left(\frac{X_i'b-Y_i}{h_2}\right)-C(\tau,v,\rho)\right]X_i\frac{1}{h_1^3}k_1''\left(\frac{W_i'\gamma}{h_1}\right)\tilde{W}_{ij}\tilde{W}_{ik},
$$

and

$$
q_{3ijkl}(b,\gamma,\rho,\tau,v)=D_i\left[K_2\left(\frac{X_i'b-Y_i}{h_2}\right)-C(\tau,v,\rho)\right]X_i\frac{1}{h_1^4}k_1'''\left(\frac{W_i'\gamma}{h_1}\right)\tilde{W}_{ij}\tilde{W}_{ik}\tilde{W}_{il}.
$$

Again similar to the arguments in (i), we can show that uniformly over (ρ,τ,v),

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}q_{2ijk}^*(\beta(\tau,\rho),\gamma(v),\rho,\tau,v)\\
&=Eq_{2ijk}^*(\beta(\tau,\rho),\gamma(v),\rho,\tau,v)+O_p\left((h_1/n)^{1/2}\ln n\right),
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}q_{3ijkl}^*(\beta(\tau,\rho),\bar{\gamma}(v),\rho,\tau,v)\\
&=Eq_{3ijkl}^*(\beta(\tau,\rho),\gamma,\rho,\tau,v)|_{\gamma=\bar{\gamma}(v)}+O_p\left((h_1/n)^{1/2}\ln n\right),
\end{aligned}
$$

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

where

$$q_{2ijk}(b,\gamma,\rho,\tau,v) = h_1^{-3} q_{2ijk}^{*}(b,\gamma,\rho,\tau,v),$$

and

$$q_{3ijkl}(b,\gamma,\rho,\tau,v) = h_1^{-4} q_{3ijkl}^{*}(b,\gamma,\rho,\tau,v).$$

On the other hand, since uniformly over (ρ,τ,v),

$$Eq_{2ijk}(\beta(\tau,\rho),\gamma(v),\rho,\tau,v) = E\left[\frac{\partial^2}{\partial^2 w_1} G_1\left(\beta(\tau,\rho),\rho,\tau,v,-\tilde{W}_i'\tilde{\gamma}(v),\tilde{W}_i\right)\right] + O\left(h_1^{s_1}+h_2^{s_2}\right),$$

and

$$Eq_{3ijkl}(\beta(\tau,\rho),\gamma,\rho,\tau,v)|_{\gamma=\bar{\gamma}(v)} = E\left[\frac{\partial^3}{\partial^3 w_1} G_1\left(\beta(\tau,\rho),\rho,\tau,v,-\tilde{W}_i'\bar{\gamma}(v),\tilde{W}_i\right)\right] + O\left(h_1^{s_1}+h_2^{s_2}\right).$$

Therefore,

$$\frac{1}{n}\sum_{i=1}^{n} q_{2ijk}(\beta(\tau,\rho),\gamma(v),\rho,\tau,v) = E\left[\frac{\partial^2}{\partial^2 w_1} G_1\left(\beta(\tau,\rho),\rho,\tau,v,-\tilde{W}_i'\tilde{\gamma}(v),\tilde{W}_i\right)\right] + O_p\left(h_1^{s_1}+h_2^{s_2}+\left(nh_1^5\right)^{-1/2}\ln n\right),$$

and

$$\frac{1}{n}\sum_{i=1}^{n} q_{3ijkl}(\beta(\tau,\rho),\gamma(v),\rho,\tau,v) = E\left[\frac{\partial^3}{\partial^3 w_1} G_1\left(\beta(\tau,\rho),\rho,\tau,v,-\tilde{W}_i'\tilde{\gamma}(v),\tilde{W}_i\right)\right] + O_p\left(\delta_{n\gamma}+h_1^{s_1}+h_2^{s_2}+\left(nh_1^7\right)^{-1/2}\ln n\right).$$

Consequently, by Assumption 8, we have uniformly over (τ,ρ),

$$Q_{n2}(\rho,\tau) = O_p\left(\delta_{n\gamma}^2+\delta_{n\gamma}^2\left(nh_1^5\right)^{-1/2}\ln n\right) = o_p\left(n^{-1/2}\right),$$

and

$$Q_{n3}(\rho,\tau) = O_p\left(\delta_{n\gamma}^3+\delta_{n\gamma}^3\left(nh_1^7\right)^{-1/2}\ln n\right) = o_p\left(n^{-1/2}\right).$$

Finally, combining the above analysis (i)–(iii), we obtain uniformly in $(\tau,\rho) \in \left[\tau_l^*,\tau_u^*\right]\times\varrho$,

$$\hat{\beta}(\tau,\rho)-\beta(\tau,\rho) = -S^{-1}(\beta(\rho,\tau))\frac{1}{n}\sum_{i=1}^{n}(q_0(D_i,Y_i,W_i;\rho,\tau)+q_1(D_i,W_i;\rho,\tau)) + o_p\left(n^{-1/2}\right).$$

(b) is concluded, if denote $\phi_{\beta 0i}(\rho,\tau) = q_0(D_i,Y_i,W_i;\rho,\tau)$ and $\phi_{\beta 1i}(\rho,\tau) = q_1(D_i,W_i;\rho,\tau)$. □

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

Proof of Lemma A2. Given the uniform consistency of $\hat{\beta}(\tau,\rho)$ and $\hat{\gamma}(v)$, by differentiating the estimating equation for $\hat{\beta}(\tau,\rho)$ with respect to ρ, we obtain

$$\left[\frac{1}{n}\sum_{i=1}^{n}\int_{v_l}^{v_u} D_i X_i X_i' \frac{1}{h_2} k_2\left(\frac{X_i'\hat{\beta}(\tau,\rho)-Y_i}{h_2}\right)\frac{1}{h_1}k_1\left(\frac{W_i'\hat{\gamma}(v)}{h_1}\right)\omega(v)\,dv\right]\hat{\beta}_\rho(\tau,\rho) \tag{a.11}$$

$$=\frac{1}{n}\sum_{i=1}^{n}\int_{v_l}^{v_u} D_i C_\rho(\tau,v,\rho) X_i \frac{1}{h_1}k_1\left(\frac{W_i'\hat{\gamma}(v)}{h_1}\right)\omega(v)\,dv,$$

with probability approaching one when sample size increases to infinity. Therefore, we have

$$\hat{\beta}_\rho(\tau,\rho) = S_{nd\rho}^{-1}\left(\hat{\beta}(\tau,\rho)\right)Q_{nd\rho}(\rho),$$

where

$$S_{nd\rho}\left(\hat{\beta}(\tau,\rho)\right)$$
$$=\int_{v_l}^{v_u} S_n\left(\hat{\beta}(\tau,\rho),\hat{\gamma}(v)\right)\omega(v)\,dv$$
$$=\frac{1}{n}\sum_{i=1}^{n}\int_{v_l}^{v_u} D_i X_i X_i' \frac{1}{h_1 h_2} k_2\left(\frac{X_i'\hat{\beta}(\tau,\rho)-Y_i}{h_2}\right)k_1\left(\frac{W_i'\hat{\gamma}(v)}{h_1}\right)\omega(v)\,dv,$$

and

$$Q_{nd\rho}(\rho) = \frac{1}{n}\sum_{i=1}^{n}\int_{v_l}^{v_u} D_i X_i C_\rho(\tau,v,\rho)\frac{1}{h_1}k_1\left(\frac{W_i'\hat{\gamma}(v)}{h_1}\right)\omega(v)\,dv.$$

In the proof of the Lemma A1, we've shown that uniformly over (τ,v,ρ),

$$S_n\left(\hat{\beta}(\tau,\rho),\hat{\gamma}(v)\right) - S\left(\hat{\beta}(\tau,\rho),\hat{\gamma}(v)\right) = O_p\left((nh_1h_2)^{-1/2}\ln n + h_1^{s_1} + h_2^{s_2}\right).$$

Thus, together with Assumption 5 and the (a.1) of Lemma A1, we have

$$S_{nd\rho}\left(\left(\hat{\beta}(\tau,\rho)\right)\right) = S_{d\rho}(\beta(\tau,\rho)) + O_p\left(\delta_{n\gamma} + (nh_1h_2)^{-1/2}\ln n + h_1^{s_1} + h_2^{s_2}\right),$$

and similarly

$$Q_{nd\rho}(\rho) = Q_{d\rho}(\rho) + O_p\left(\delta_{n\gamma} + (nh_1)^{-1/2}\ln n + h_1^{s_1}\right),$$

where

$$S_{d\rho}(\beta(\tau,\rho)) = E\left(D_i f_{Y^*}\left(X_i'\beta(\tau,\rho),W_i\right)\omega(1-P_i)f_{Y_2^*}(0,W_i)X_iX_i'\right),$$

and

$$Q_{d\rho}(\rho) = E\left(D_i C_\rho(\tau,1-P_i,\rho)\,\omega(1-P_i)f_{Y_2^*}(0,W_i)X_i\right),$$

and

$$S_{d\rho}^{-1}(\beta(\tau,\rho))Q_{d\rho}(\rho) = \left(\frac{\partial\bar{G}_1(\tau,\rho)}{\partial\beta}\right)^{-1}\frac{\partial\bar{G}_1(\tau,\rho)}{\partial\rho}.$$

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

Thus

$$\hat{\beta}_\rho(\tau,\rho)-\beta_\rho(\tau,\rho)=O_p\left(\delta_{n\gamma}+(nh_1h_2)^{-1/2}\ln n+h_1^{s_1}+h_2^{s_2}\right).$$

Next, differentiating equation (a.11) with respect to ρ again, we obtain

$$\left[\frac{1}{n}\sum_{i=1}^{n}\int_{v_l}^{v_u}D_i\frac{1}{h_2}k_2\left(\frac{X_i'\hat{\beta}(\tau,\rho)-Y_i}{h_2}\right)X_iX_i'\frac{1}{h_1}k_1\left(\frac{W_i'\hat{\gamma}(v)}{h_1}\right)\omega(v)\,dv\right]\hat{\beta}_{\rho\rho}(\tau,\rho)$$
$$+\frac{1}{n}\sum_{i=1}^{n}\int_{v_l}^{v_u}D_i\frac{1}{h_2^2}k_2'\left(\frac{Y_i-X_i'\hat{\beta}(\tau,\rho)}{h_2}\right)\left(X_i'\hat{\beta}_\rho(\tau,\rho)\right)^2X_i\frac{1}{h_1}k_1\left(\frac{W_i'\hat{\gamma}(v)}{h_1}\right)\omega(v)\,dv$$
$$=\frac{1}{n}\sum_{i=1}^{n}\int_{v_l}^{v_u}D_iC_{\rho\rho}(\tau,v,\rho)X_i\frac{1}{h_1}k_1\left(\frac{W_i'\hat{\gamma}(v)}{h_1}\right)\omega(v)\,dv. \quad \textbf{(a.12)}$$

Thus,

$$\hat{\beta}_{\rho\rho}(\tau,\rho)=S_{nd\rho}^{-1}\left(\hat{\beta}(\tau,\rho)\right)Q_{ndd\rho}\left(\left(\hat{\beta}(\tau,\rho)\right),\rho\right),$$

where

$$Q_{ndd\rho}\left(\left(\hat{\beta}(\tau,\rho)\right),\rho\right)=Q_{n2dd\rho}\left(\hat{\beta}(\tau,\rho)\right)-Q_{n1dd\rho}(\rho),$$

with

$$Q_{n1dd\rho}\left(\hat{\beta}(\tau,\rho)\right)$$
$$=\frac{1}{n}\sum_{i=1}^{n}\int_{v_l}^{v_u}D_i\frac{1}{h_2^2}k_2'\left(\frac{Y_i-X_i'\hat{\beta}(\tau,\rho)}{h_2}\right)\left(X_i'\hat{\beta}_\rho(\tau,\rho)\right)^2X_i\frac{1}{h_1}k_1\left(\frac{W_i'\hat{\gamma}(v)}{h_1}\right)\omega(v)\,dv,$$

and

$$Q_{n2dd\rho}(\rho)=\frac{1}{n}\sum_{i=1}^{n}\int_{v_l}^{v_u}D_iC_{\rho\rho}(\tau,v,\rho)\frac{1}{h_1}k_1\left(\frac{W_i'\hat{\gamma}(v)}{h_1}\right)X_i\omega(v)\,dv.$$

Then similar to the above proof of $\hat{\beta}_\rho(\tau,\rho)-\beta_\rho(\tau,\rho)$, it can be shown uniformly over (τ,ρ),

$$Q_{n1dd\rho}\left(\hat{\beta}(\tau,\rho)\right)=Q_{1dd\rho}(\beta(\tau,\rho))+O_p\left(\delta_{n\gamma}+(nh_1h_2^3)^{-1/2}\ln n+h_1^{s_1}+h_2^{s_2}\right),$$

and

$$Q_{n2dd\rho}(\rho)=Q_{2dd\rho}(\rho)+O_p\left(\delta_{n\gamma}+(nh_1)^{-1/2}\ln n+h_1^{s_1}\right),$$

where

$$Q_{1dd\rho}(\beta(\tau,\rho))=E\left(D_if_{Y^*}'\left(X_i'\beta(\tau,\rho),W_i\right)\left(X_i'\beta_\rho(\tau,\rho)\right)^2\omega(1-P_i)f_{Y_2^*}(0,W_i)X_i\right),$$

and

$$Q_{2dd\rho}(\rho)=E\left(D_iC_{\rho\rho}(\tau,1-P_i,\rho)\,\omega(1-P_i)f_{Y_2^*}(0,W_i)X_i\right),$$

where $f_{Y^*}'(t,w)=\partial f_{Y^*}(t,w)/\partial t$.

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

Consequently, we obtain

$$\hat{\beta}_{\rho\rho}(\tau,\rho)-\beta_{\rho\rho}(\tau,\rho)=O_p\left(\delta_{n\gamma}+(nh_1h_2^3)^{-1/2}\ln n+h_1^{s_1}+h_2^{s_2}\right).$$

□

Proof of Theorem 1. Recall that

$$\hat{\rho}=\arg\min_{\rho\in\varrho}\bar{G}_{2n}(\rho),$$

where

$$\bar{G}_{2n}(\rho)=\sum_{\tau\in\mathcal{J}}\int_{v_l}^{v_u}\left\|G_{2n}\left(\hat{\beta}(\tau,\rho),\hat{\gamma}(v),\rho,v,\tau\right)\right\|^2\omega(v)\,dv,$$

with

$$G_{2n}(b,\gamma,\rho,v,\tau)=\frac{1}{nh_1}\sum_{i=1}^{n}D_i\left(K_2\left(\frac{X_i'b-Y_i}{h_2}\right)-C(\tau,v,\rho)\right)k_1\left(\frac{W_i'\gamma}{h_1}\right)Z_i.$$

Define the class of functions

$$\mathcal{G}_2=\Big\{m(D,Y,W;h_1,h_2,b,\gamma,v,\tau):$$
$$(h_1,h_2,b,\gamma,v,\tau)\in\mathbb{R}^2\times B\times G\times\varrho\times[\tau_l,\tau_u]\times\left[\tau_l^*,\tau_u^*\right]\Big\},$$

where

$$m(D,Y,W;h_1,h_2,b,\gamma,v,\tau)=D\left(K_2\left(\frac{X'b-Y}{h_2}\right)-C(\tau,v,\rho)\right)k_1\left(\frac{W'\gamma}{h_1}\right)Z.$$

(i). Consistency: Similar $\mathcal{F}_1,\mathcal{F}_2$ and $\mathcal{F}_3,\mathcal{G}_2$ is the Euclidean class with a uniformly bounded envelope, thus we have

$$G_{2n}(b,\gamma,\rho,v,\tau)=G_2(b,\gamma,\rho,v,\tau)+o_p(1),$$

uniformly in $(b,\gamma,\rho,v,\tau)\in B\times G\times\varrho\times[\tau_l,\tau_u]\times\left[\tau_l^*,\tau_u^*\right]$, where

$$G_2(b,\gamma,\rho,v,\tau)=E\left[D_i\left[1\left\{Y_i\le X_i'b\right\}-C(\tau,v,\rho)\right]Z_i|W_i'\gamma=0\right]p_{W'\gamma}(0).$$

Then using Lemma A1 and Assumption 5, we have

$$\sup_{\rho\in\varrho}\sup_{v,\tau\in[v_l,v_u]\times\left[\tau_l^*,\tau_u^*\right]}\left\|G_{2n}\left(\hat{\beta}(\tau,\rho),\hat{\gamma}(v),\rho,v,\tau\right)-G_2\left(\hat{\beta}(\tau,\rho),\hat{\gamma}(v),\rho,v,\tau\right)\right\|=o_p(1),$$

and

$$\sup_{\rho\in\varrho}\sup_{v,\tau\in[v_l,v_u]\times\left[\tau_l^*,\tau_u^*\right]}\left\|G_2\left(\hat{\beta}(\tau,\rho),\hat{\gamma}(v),\rho,v,\tau\right)-G_2(\beta(\tau,\rho),\gamma(v),\rho,v,\tau)\right\|=o_p(1).$$

Consequently, we have

$$\bar{G}_{2n}(\rho)=\bar{G}_2(\rho)+o_p(1),$$

uniformly over $\rho\in\varrho$. In addition, $\bar{G}_2(\rho)$ is continuous and achieves the unique minimum at ρ_0 (Assumption 9). Consequently, consistency follows from some standard arguments (e.g., Newey and McFadden, 1994).

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

(ii). Asymptotic normality. Given the consistency result, with probability approaching one as the sample size increases, we obtain the following first-order condition:

$$\frac{d}{d\rho}\bar{G}_{2n}\left(\hat{\rho}\right)=0, \tag{a.13}$$

where

$$\frac{d}{d\rho}\bar{G}_{2n}\left(\rho\right)$$
$$=\sum_{\tau_j\in\mathcal{J}}\int_{v_l}^{v_u}\left[G_{2n}\left(\hat{\beta}\left(\tau_j,\rho\right),\hat{\gamma}\left(v\right),\rho,v,\tau_j\right)\right]\frac{dG_{2n}\left(\hat{\beta}\left(\tau_j,\rho\right),\hat{\gamma}\left(v\right),\rho,v,\tau_j\right)}{d\rho}\omega\left(v\right)dv,$$

and

$$\frac{dG_{2n}\left(\hat{\beta}\left(\tau,\rho\right),\hat{\gamma}\left(v\right),\rho,v,\tau\right)}{d\rho}=\frac{\partial}{\partial\beta'}G_{2n}\left(\hat{\beta}\left(\tau,\rho\right),\hat{\gamma}\left(v\right),\rho,v,\tau\right)\hat{\beta}_{\rho}\left(\tau,\rho\right)$$
$$+\frac{\partial}{\partial\rho}G_{2n}\left(\hat{\beta}\left(\tau,\rho\right),\hat{\gamma}\left(v\right),\rho,v,\tau\right).$$

A mean-value expansion of (a.13) with respect to $\hat{\rho}$ at ρ_0 yields

$$\sqrt{n}(\hat{\rho}-\rho_0)=-\left[\frac{d^2}{d\rho^2}\bar{G}_{2n}\left(\bar{\rho}\right)\right]^{-1}\sqrt{n}\frac{d}{d\rho}\bar{G}_{2n}\left(\rho_0\right),$$

where $\bar{\rho}$ lies on the line segment between $\hat{\rho}$ and ρ_0.

We first consider $\sqrt{n}\frac{d}{d\rho}\bar{G}_{2n}\left(\rho_0\right)$.

Similar to (a.8) and from Assumptions 8 and 8, using Pollard (1995)

$$\frac{\partial}{\partial\beta}G_{2n}\left(\hat{\beta}\left(\tau,\rho_0\right),\hat{\gamma}\left(v\right),\rho_0,v,\tau\right)$$
$$=\frac{\partial}{\partial\beta}G_{2}\left(\hat{\beta}\left(\tau,\rho_0\right),\hat{\gamma}\left(v\right),\rho_0,v,\tau\right)+O_p\left((nh_1h_2)^{-1/2}\ln n+h_1^{s_1}+h_2^{s_2}\right),$$

and

$$\frac{\partial}{\partial\rho}G_{2n}\left(\hat{\beta}\left(\tau,\rho_0\right),\hat{\gamma}\left(v\right),\rho_0,v,\tau\right)$$
$$=\frac{\partial}{\partial\rho}G_{2}\left(\hat{\beta}\left(\tau,\rho_0\right),\hat{\gamma}\left(v\right),\rho_0,v,\tau\right)+O_p\left((nh_1)^{-1/2}\ln n+h_1^{s_1}+h_2^{s_2}\right),$$

where because of Lemma A1 and Assumption 5

$$\frac{\partial}{\partial\beta}G_{2}\left(\hat{\beta}\left(\tau,\rho_0\right),\hat{\gamma}\left(v\right),\rho_0,v,\tau\right)$$
$$=\frac{\partial}{\partial\beta}G_{2}\left(\beta\left(\tau,\rho_0\right),\gamma\left(v\right),\rho_0,v,\tau\right)+O_p\left(\delta_{n\gamma}\right),$$

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

and similarly,

$$\frac{\partial}{\partial\rho}G_2\left(\hat{\beta}(\tau,\rho_0),\hat{\gamma}(v),\rho_0,v,\tau\right)$$
$$=\frac{\partial}{\partial\rho}G_2(\beta(\tau,\rho_0),\gamma(v),\rho_0,v,\tau)+O_p(\delta_{n\gamma}).$$

Hence, using (a.3) of Lemma A2, uniformly over (τ,v),

$$\frac{dG_{2n}\left(\hat{\beta}(\tau,\rho_0),\hat{\gamma}(v),\rho_0,v,\tau\right)}{d\rho}=G_{2\rho}(\tau,v)+O_p\left(\delta_{n\gamma}+(nh_1h_2)^{-1/2}\ln n+h_1^{s_1}+h_2^{s_2}\right), \tag{a.14}$$

where

$$\begin{aligned}G_{2\rho}(\tau,v)&=\frac{d}{d\rho}G_2(\beta(\tau,\rho_0),\gamma(v),\rho_0,v,\tau)\\&=\frac{\partial}{\partial\beta'}G_2(\beta(\tau,\rho_0),\gamma(v),\rho_0,v,\tau)\beta_\rho(\tau,\rho)\\&\quad+\frac{\partial}{\partial\rho}G_2(\beta(\tau,\rho_0),\gamma(v),\rho_0,v,\tau).\end{aligned}$$

Therefore,

$$\begin{aligned}\frac{d}{d\rho}\bar{G}_{2n}(\rho_0)=&\sum_{\tau_j\in\mathcal{J}}\int_{v_l}^{v_u}G_{2n}\left(\hat{\beta}(\tau_j,\rho_0),\hat{\gamma}(v),\rho_0,v,\tau_j\right)G_{2\rho}(\tau_j,v)\omega(v)\,dv\\&+O_p\left(\delta_{n\gamma}+(nh_1h_2)^{-1/2}\ln n+h_1^{s_1}+h_2^{s_2}\right)\\&\times\sum_{\tau_j\in\mathcal{J}}\int_{v_l}^{v_u}G_{2n}\left(\hat{\beta}(\tau_j,\rho_0),\hat{\gamma}(v),\rho_0,v,\tau_j\right)\omega(v)\,dv.\end{aligned} \tag{a.15}$$

Using (a.1) and (a.2) of Lemma A1 and Assumption 5, the first term on the r.h.s of (a.15) becomes

$$\begin{aligned}&\sum_{\tau_j\in\mathcal{J}}\int_{v_l}^{v_u}G_{2n}\left(\hat{\beta}(\tau_j,\rho_0),\hat{\gamma}(v),\rho_0,v,\tau_j\right)G_{2\rho}(\tau_j,v)\omega(v)\,dv\\&=\sum_{\tau_j\in\mathcal{J}}\int_{v_l}^{v_u}G_{2n}(\beta(\tau_j,\rho_0),\gamma(v),\rho_0,v,\tau_j)G_{2\rho}(\tau_j,v)\omega(v)\,dv\\&+\sum_{\tau_j\in\mathcal{J}}\left[\hat{\beta}(\tau_j,\rho_0)-\beta(\tau_j,\rho_0)\right]'\int_{v_l}^{v_u}\frac{\partial}{\partial\beta}G_{2n}(\beta(\tau_j,\rho_0),\gamma(v),\rho_0,v,\tau_j)G_{2\rho}(\tau_j,v)\omega(v)\,dv\\&+\sum_{\tau_j\in\mathcal{J}}\int_{v_l}^{v_u}\frac{\partial}{\partial\gamma'}G_{2n}(\beta(\tau_j,\rho_0),\gamma(v),\rho_0,v,\tau_j)G_{2\rho}(\tau_j,v)(\hat{\gamma}(v)-\gamma(v))\omega(v)\,dv\\&+o_p\left(n^{-1/2}\right).\end{aligned}$$

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

Then, similar to (a.10) of (i) and (ii) in proof of Lemma A1, under Assumption 8, we obtain

$$\begin{aligned}
&\sum_{\tau_j \in \mathcal{J}} \int_{v_l}^{v_u} G_{2n}\left(\beta\left(\tau_j, \rho_0\right), \gamma\left(v\right), \rho_0, v, \tau_j\right) G_{2\rho}(\tau_j, v)\omega\left(v\right) dv \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{\tau_j \in \mathcal{J}} \int_{v_l}^{v_u} D_i Z_i \left[K_2\left(\frac{X_i'\beta(\tau_j,\rho_0) - Y_i}{h_2}\right) - C(\tau_j, v, \rho_0)\right] \\
&\quad \times \frac{1}{h_1} k_1\left(\frac{W_i'\gamma\left(v\right)}{h_1}\right) G_{2\rho}(\tau_j, v)\omega\left(v\right) dv \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{\tau_j \in \mathcal{J}} D_i Z_i \left[1\left\{Y_i \le X_i'\beta\left(\tau_j,\rho_0\right)\right\} - C(\tau_j, 1-P_i, \rho_0)\right] \\
&\quad \times G_{2\rho}(\tau_j, 1-P_i) f_{Y_2^*}(0, W_i)\omega\left(1-P_i\right) + o_p\left(n^{-1/2}\right) \\
&:= \frac{1}{n}\sum_{i=1}^{n}\phi_{0,\rho i} + o_p\left(n^{-1/2}\right)
\end{aligned} \tag{a.16}$$

and

$$\begin{aligned}
&\sum_{\tau_j \in \mathcal{J}} \left[\hat{\beta}\left(\tau_j,\rho_0\right) - \beta\left(\tau_j,\rho_0\right)\right]' \int_{v_l}^{v_u} \frac{\partial}{\partial \beta} G_{2n}\left(\beta\left(\tau_j, \rho_0\right), \gamma\left(v\right), \rho_0, v, \tau_j\right) G_{2\rho}(\tau_j, v)\omega\left(v\right) dv \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{\tau_j \in \mathcal{J}} \int_{v_l}^{v_u} \frac{\partial}{\partial \beta'} G_2\left(\beta\left(\tau_j,\rho_0\right), \gamma\left(v\right), \rho_0, v, \tau_j\right) \\
&\quad \times G_{2\rho}(\tau_j, v)\phi_{\beta i}\left(\rho_0, \tau_j\right)\omega\left(v\right) dv + o_p(n^{-1/2}) \\
&:= \frac{1}{n}\sum_{i=1}^{n}\phi_{1,\rho i} + o_p(n^{-1/2}),
\end{aligned} \tag{a.17}$$

and

$$\begin{aligned}
&\sum_{\tau_j \in \mathcal{J}} \int_{v_l}^{v_u} \frac{\partial}{\partial \gamma'} G_{2n}\left(\beta\left(\tau_j, \rho_0\right), \gamma\left(v\right), \rho_0, v, \tau_j\right) G_{2\rho}(\tau_j, v)\left(\hat{\gamma}\left(v\right) - \gamma\left(v\right)\right)\omega\left(v\right) dv \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{\tau_j \in \mathcal{J}} \frac{\partial}{\partial \gamma'} G_2\left(\beta\left(\tau_j,\rho_0\right), \gamma\left(1-P_i\right), \rho_0, 1-P_i, \tau_j\right) \\
&\times G_{2\rho}(\tau_j, 1-P_i) Q_{1-P_i}^{-1}\left(D_i - P_i\right)\tilde{W}_i f_{Y_2^*}(0, W_i)\omega\left(1-P_i\right) + o_p\left(n^{-1/2}\right) \\
&:= \frac{1}{n}\sum_{i=1}^{n}\phi_{2,\rho i} + o_p\left(n^{-1/2}\right).
\end{aligned} \tag{a.18}$$

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

By similar arguments of showing (a.16)–(a.18), $\sum_{\tau_j \in \mathcal{J}} \int_{v_l}^{v_u} G_{2n}(\hat{\beta}(\tau_j, \rho_0), \hat{\gamma}(v), \rho_0, v, \tau_j)\omega(v)dv$ can be treated similarly. Thus the second term in the r.h.s of (a.15) is of the order $o_p\left(n^{-1/2}\right)$ since $(nh_1h_2)^{-1/2}\ln n + h_1^{s_1} + h_2^{s_2} + \delta_{n\gamma} = o_p(1)$ by Assumption 8.

We then analyze $\frac{d^2}{d\rho^2}\bar{G}_{2n}(\bar{\rho})$. Note that

$$\frac{d^2}{d\rho^2}\bar{G}_{2n}(\bar{\rho}) = \sum_{\tau_j \in \mathcal{J}} \int_{v_l}^{v_u} \left[\left(\frac{dG_{2n}\left(\hat{\beta}\left(\tau_j, \bar{\rho}\right), \hat{\gamma}(v), \bar{\rho}, v, \tau_j\right)}{d\rho} \right)^2 \omega(v)\,dv \right] \tag{a.19}$$
$$+ \sum_{\tau_j \in \mathcal{J}} \int_{v_l}^{v_u} \left[G_{2n}\left(\hat{\beta}\left(\tau_j, \bar{\rho}\right), \hat{\gamma}(v), \bar{\rho}, v, \tau_j\right) \right] \frac{d^2 G_{2n}\left(\hat{\beta}\left(\tau_j, \bar{\rho}\right), \hat{\gamma}(v), \bar{\rho}, v, \tau_j\right)}{d\rho^2} \omega(v)\,dv,$$

where

$$\begin{aligned}
&\frac{d^2 G_{2n}\left(\hat{\beta}(\tau, \bar{\rho}), \hat{\gamma}(v), \bar{\rho}, v, \tau\right)}{d\rho^2} \\
&= \left(\hat{\beta}_\rho(\tau, \bar{\rho})\right)' \frac{\partial^2}{\partial\beta\partial\beta'} G_{2n}\left(\hat{\beta}(\tau, \bar{\rho}), \hat{\gamma}(v), \bar{\rho}, v, \tau\right) \hat{\beta}_\rho(\tau, \bar{\rho}) \\
&\quad + \frac{\partial}{\partial\beta'} G_{2n}\left(\hat{\beta}(\tau, \bar{\rho}), \hat{\gamma}(v), \bar{\rho}, v, \tau\right) \hat{\beta}_{\rho\rho}(\tau, \bar{\rho}) \\
&\quad + \frac{\partial^2}{\partial\beta'\partial\rho} G_{2n}\left(\hat{\beta}(\tau, \bar{\rho}), \hat{\gamma}(v), \bar{\rho}, v, \tau\right) \hat{\beta}_\rho(\tau, \bar{\rho}) \\
&\quad + \frac{\partial^2 G_{2n}\left(\hat{\beta}(\tau, \bar{\rho}), \hat{\gamma}(v), \bar{\rho}, v, \tau\right)}{\partial\rho^2}.
\end{aligned}$$

Similar to the proof of (a.3) of Lemma A2, under Assumptions 8 and 9, we can show that uniformly over τ, v and ρ in the $o_p(1)$ neighborhood of ρ_0,

$$\begin{aligned}
&G_{2n}\left(\hat{\beta}(\tau, \rho), \hat{\gamma}(v), \rho, v, \tau\right) \\
&= G_2\left(\hat{\beta}(\tau, \rho), \hat{\gamma}(v), \rho, v, \tau\right) + O_p\left((nh_1)^{-1/2}\ln n + h_1^{s_1} + h_2^{s_2}\right) \\
&= G_2(\beta(\tau, \rho), \gamma(v), \rho, v, \tau) + O_p\left(\delta_{n\gamma} + (nh_1)^{-1/2}\ln n + h_1^{s_1} + h_2^{s_2}\right),
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\partial}{\partial\beta} G_{2n}\left(\hat{\beta}(\tau, \rho), \hat{\gamma}(v), \rho, v, \tau\right) \\
&= \frac{\partial}{\partial\beta} G_2\left(\hat{\beta}(\tau, \rho), \hat{\gamma}(v), \rho, v, \tau\right) + O_p\left((nh_1h_2)^{-1/2}\ln n + h_1^{s_1} + h_2^{s_2}\right) \\
&= \frac{\partial}{\partial\beta} G_2(\beta(\tau, \rho), \gamma(v), \rho, v, \tau) + O_p\left(\delta_{n\gamma} + (nh_1h_2)^{-1/2}\ln n + h_1^{s_1} + h_2^{s_2}\right),
\end{aligned}$$

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

and

$$\begin{aligned}
&\frac{\partial^2}{\partial\beta'\partial\rho}G_{2n}\left(\hat{\beta}(\tau,\rho),\hat{\gamma}(v),\rho,v,\tau\right)\\
&=\frac{\partial^2}{\partial\beta'\partial\rho}G_2\left(\hat{\beta}(\tau,\rho),\hat{\gamma}(v),\rho,v,\tau\right)+O_p\left((nh_1h_2)^{-1/2}\ln n+h_1^{s_1}+h_2^{s_2}\right)\\
&=\frac{\partial^2}{\partial\beta'\partial\rho}G_2\left(\beta(\tau,\rho),\gamma(v),\rho,v,\tau\right)+O_p\left(\delta_{n\gamma}+(nh_1h_2)^{-1/2}\ln n+h_1^{s_1}+h_2^{s_2}\right),
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\partial^2}{\partial\beta\partial\beta'}G_{2n}\left(\hat{\beta}(\tau,\rho),\hat{\gamma}(v),\rho,v,\tau\right)\\
&=\frac{\partial^2}{\partial\beta\partial\beta'}G_2\left(\hat{\beta}(\tau,\rho),\hat{\gamma}(v),\rho,v,\tau\right)+O_p\left(\left(nh_1h_2^3\right)^{-1/2}\ln n+h_1^{s_1}+h_2^{s_2}\right)\\
&=\frac{\partial^2}{\partial\beta\partial\beta'}G_2\left(\beta(\tau,\rho),\gamma(v),\rho,v,\tau\right)+O_p\left(\delta_{n\gamma}+\left(nh_1h_2^3\right)^{-1/2}\ln n+h_1^{s_1}+h_2^{s_2}\right),
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\partial^2 G_{2n}\left(\hat{\beta}(\tau,\rho),\hat{\gamma}(v),\rho,v,\tau\right)}{\partial\rho^2}\\
&=\frac{\partial^2 G_2\left(\hat{\beta}(\tau,\rho),\hat{\gamma}(v),\rho,v,\tau\right)}{\partial\rho^2}+O_p\left((nh_1)^{-1/2}\ln n+h_1^{s_1}+h_2^{s_2}\right)\\
&=\frac{\partial^2 G_2\left(\beta(\tau,\rho),\gamma(v),\rho,v,\tau\right)}{\partial\rho^2}+O_p\left(\delta_{n\gamma}+(nh_1)^{-1/2}\ln n+h_1^{s_1}+h_2^{s_2}\right).
\end{aligned}$$

Since $G_2(\beta(\tau,\rho_0),\gamma(v),\rho_0,v,\tau)=0$, then, together with (a.2) of Lemma A1, (a.4) of Lemma 2, and Assumptions 5, 8,

$$\sum_{\tau_j\in\mathcal{J}}\int_{v_l}^{v_u}\left[G_{2n}\left(\hat{\beta}\left(\tau_j,\bar{\rho}\right),\hat{\gamma}(v),\bar{\rho},v,\tau_j\right)\right]\frac{d^2G_{2n}\left(\hat{\beta}\left(\tau_j,\bar{\rho}\right),\hat{\gamma}(v),\bar{\rho},v,\tau_j\right)}{d\rho^2}\omega(v)\,dv=o_p(1). \tag{a.20}$$

Consequently, using Assumption 8, (a.14), (a.19), and (a.20) imply

$$\frac{d^2}{d\rho^2}\bar{G}_{2n}(\bar{\rho})=\frac{d^2}{d\rho^2}\bar{G}_2(\rho_0)+o_p(1),$$

where

$$\frac{d^2}{d\rho^2}\bar{G}_2(\rho_0)=\sum_{\tau_j\in\mathcal{J}}\int_{v_l}^{v_u}\left[G_{2\rho}^2\left(\tau_j,v\right)\omega(v)\,dv\right], \tag{a.21}$$

and where

$$G_{2\rho}(\tau,v)=\frac{dG_2\left(\beta(\tau,\rho_0),\gamma(v),\rho_0,v,\tau\right)}{d\rho}.$$

https://doi.org/10.1017/S0266466625100121 Published online by Cambridge University Press

Combining (a.16)–(a.21), we obtain

$$\sqrt{n}\left(\hat{\rho}-\rho\right)=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{d^2}{d\rho^2}\bar{G}_2\left(\rho_0\right)\right)^{-1}\left(\phi_{0,\rho i}+\phi_{1,\rho i}+\phi_{2,\rho i}\right)+o_p\left(1\right).$$

In the following, we denote

$$\phi_{\rho i}=\left(\frac{d^2}{d\rho^2}\bar{G}_2\left(\rho_0\right)\right)^{-1}\left(\phi_{0,\rho i}+\phi_{1,\rho i}+\phi_{2,\rho i}\right).$$

Finally, given the asymptotic linear representation of $\sqrt{n}\left(\hat{\rho}-\rho\right)$, for any $\tau\in\left[\tau_l^*,\tau_u^*\right]$, using (a.2) of Lemma A1, we obtain

$$\begin{aligned}\hat{\beta}\left(\tau,\hat{\rho}\right)-\beta\left(\tau\right)&=\hat{\beta}\left(\tau,\rho_0\right)-\beta\left(\tau\right)+\hat{\beta}_\rho\left(\tau,\rho_0\right)\left(\hat{\rho}-\rho_0\right)+o\left(\left\|\hat{\rho}-\rho_0\right\|\right)\\&=\hat{\beta}\left(\tau,\rho_0\right)-\beta\left(\tau,\rho_0\right)+\beta_\rho\left(\tau,\rho_0\right)\left(\hat{\rho}-\rho_0\right)+o_p\left(n^{-1/2}\right).\end{aligned}$$

Then

$$\begin{aligned}&\sqrt{n}\left(\hat{\beta}\left(\tau,\hat{\rho}\right)-\beta\left(\tau\right)\right)\\&=\sqrt{n}\left(\hat{\beta}\left(\tau,\hat{\rho}\right)-\beta\left(\tau,\hat{\rho}\right)+\beta\left(\tau,\hat{\rho}\right)-\beta\left(\tau,\rho_0\right)\right)\\&=\sqrt{n}\left(\hat{\beta}\left(\tau,\rho_0\right)-\beta\left(\tau,\rho_0\right)\right)+\beta_\rho\left(\tau,\rho_0\right)\sqrt{n}\left(\hat{\rho}-\rho_0\right)+o_p\left(1\right)\\&=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\phi_{\beta i}\left(\rho_0,\tau\right)+\beta_\rho\left(\tau,\rho_0\right)\phi_{\rho i}\right]+o_p\left(1\right)\\&:=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_{\beta i}\left(\tau\right)+o_p\left(1\right).\end{aligned}$$

Therefore, we have

$$\sqrt{n}\begin{bmatrix}\hat{\beta}\left(\tau,\hat{\rho}\right)-\beta\left(\tau\right)\\\hat{\rho}-\rho_0\end{bmatrix}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_{\theta i}\left(\tau\right)+o_p\left(1\right),$$

where $\phi_{\theta i}\left(\tau\right)=\left(\phi_{\beta i}'\left(\tau\right),\phi_{\rho i}'\right)'$. Consequently, Theorem 1 follows by applying the Central Limit Theorem to the asymptotic linear representations above. □

Supplementary Material

Songnian Chen and Hanghui Zhang (June 2025): Supplement to “Semiparametric Estimation of Quantile Regression with Binary Quantile Selection”, Econometric Theory Supplementary Material. To view, please visit: https://doi.org/10.1017/S0266466625100121.

REFERENCES

Angrist, J., Chernozhukov, V., & Fernandez-Val, I. (2006). Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica*, 74, 539–563.

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

Carroll, R. J., Fan, J., Gijbels, I., & Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, 92(438), 477–489.

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

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

Chen, S., & Zhang, H. (2015). Binary quantile regression with local polynomial smoothing. *Journal of Econometrics*, 189(1), 24–40.

Chernozhukov, V., Fernández-Val, I., & Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, 81(6), 2205–2268.

Chernozhukov, V., Fernández-Val, I., Melly, B., & Wüthrich, K. (2020). Generic inference on quantile and quantile effect functions for discrete outcomes. *Journal of the American Statistical Association*, 115(529), 123–137.

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

Chetverikov, D., Larsen, B., & Palmer, C. (2016). IV quantile regression for group-level treatments, with an application to the distributional effects of trade. *Econometrica*, 84(2), 809–833.

Fernandes, M., Guerre, E., & Horta, E. (2019). Smoothing quantile regression. *Journal of Business and Economic Statistics*, 39(2), 338–357.

Florios, K., & Skouras, S. (2008). Exact computation of max weighted score estimators. *Journal of Econometrics*, 146(1), 86–91.

Gao, Y., Xu, S., & Xu, K. (2022). Two-stage maximum score estimator, arxiv working paper.

Galvao, A. F., & Kato, K. (2016). Smoothed quantile regression for panel data. *Journal of Econometrics*, 193(1), 92–112.

He, X., Pan, X., Tian, M., & Zhou, W. (2023). Smoothed quantile regression for large scale inference. *Journal of Econometrics*, 232(2), 367–388.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153–161.

Horowitz, J. L. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica*, 63(3), 505–531.

Horowitz, J. L. (1993). Semiparametric and nonparametric estimation of quantal response models. In G.S. Maddala, C.R. Rao and H.D. Vinod, Eds. *Handbook of Statistics*, Vol. 11, 45–72. North Holland, Amsterdam.

Horowitz, J. L. (1998). Bootstrap methods for median regression models. *Econometrica*, 66, 1327–1351.

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1), 33–50.

Kaplan, D. M., & Sun, Y. (2017). Smoothed estimating equations for instrumental variables quantile regression. *Econometric Theory*, 33(1), 105–157.

Kordas, G. (2006). Smoothed binary regression quantiles. *Journal of Applied Econometrics*, 21(3), 387–407.

Manski, C. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics*, 3(3), 205–228.

Manski, C. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of Econometrics*, 27(3), 313–333.

Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. In R.F Engle and D L. McFadden, Eds. *Handbook of Ecomometrics*, Vol. 4, 2111–2245. North Holland, Amsterdam.

Nolan, D., & Pollard, D. (1987). U-processes: Rates of convergence. *Annals of Statistics*, 15(2), 780–799.

Phillips, P. C. B. (2015). Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics*, 13(3), 521–555.

Pakes, A., & Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, 57(5), 1027–1057.

Politis, D. N., Romano, J. P., & Wolf, M. (1999). *Subsampling*. Springer Series in Statistics. Springer New York.

Pollard, D. (1989). Asymptotics via empirical processes. *Statistical Science*, 4(4), 341–354.

Pollard, D. (1995). Uniform ratio limit theorems for empirical processes. *Scandinavian Journal of Statistics*, 22(3), 271–278.

Volgushev, S. (2020). Smoothed quantile regression processes for binary response models. *Econometric Theory*, 36(2), 292–330.