Supplementary Materials for "Quantile Regression in the Secondary Analysis of Case-Control Data"

1 Additional Simulation Studies

1.1 Simulations results for Models (1) and (3) in Section 3.1

n	Methods	$\tau = 0.5$			$\tau = 0.9$		
		RB (%)	SE	MSE $\times n$	RB (%)	SE	MSE $\times n$
	QR_controls	-13.0	0.044	4.3	-13.2	0.060	8.0
	QR_cases	10.8	0.041	3.8	-0.6	0.055	6.0
2000	QR_case-control	39.4	0.032	6.5	25.2	0.043	6.4
	IPW	-4.8	0.042	3.6	-2.5	0.056	6.3
	KS	-4.7	0.042	3.6	-2.0	0.056	6.3
	SICO (m=1)	-4.6	0.047	4.4	-1.9	0.064	8.1
	SICO (m=10)	-3.5	0.042	3.5	-1.6	0.055	6.0
	SICO (m=100)	-3.9	0.042	3.5	-2.0	0.055	6.0
	QR_controls	-7.3	0.089	4.0	-6.1	0.122	7.4
	QR_cases	12.2	0.087	3.9	4.7	0.116	6.7
500	QR_case-control	41.0	0.064	3.2	31.0	0.086	4.8
	IPW	0.1	0.084	3.5	2.0	0.112	6.3
	KS	0.3	0.086	3.7	-0.3	0.114	6.5
	SICO (m=1)	-0.6	0.088	3.9	-0.6	0.116	6.7
	SICO (m=10)	-1.3	0.082	3.4	-2.9	0.105	5.5
	SICO (m=100)	4.1	0.081	3.3	5.4	0.104	5.4

Table 1: Relative bias (RB), standard error (SE) and mean squared error (MSE) of the estimated quantile coefficients under Model (1) at quantile levels 0.5 and 0.9. In Model (1),$x_i = u_{i,1} + u_{i,2}$ where $u_{i,1}$ and $u_{i,2}$ are iid bernoulli random variables with $p = 0.3$, $z_i \sim N(0, 1)$, and $e_i \sim N(0, 1)$. "QR_controls" stands for unadjusted quantile regression using controls only. "QR_cases" stands for unadjusted quantile regression using cases only. "QR_controls" are unadjusted quantile regression using both case and control samples. IPW is the estimates using inverse probability weighting;"KS" is the KS estimates using kernel smoothing. SICO(m) is the SICO estimates with m replicate.

n	Methods	$\tau = 0.5$			$\tau = 0.9$		
		RB (%)	SE	MSE $\times n$	RB (%)	SE	MSE $\times n$
2000	QR_controls	-7.4	0.030	2.0	-11.3	0.042	4.0
	QR_cases	10.0	0.029	2.0	3.3	0.042	3.6
	QR_case-control	46.0	0.023	7.1	31.8	0.031	6.2
	IPW	1.2	0.029	1.7	0.1	0.039	3.1
	KS	2.8	0.031	1.9	1.2	0.041	3.4
	SICO (m=1)	0.8	0.031	2.0	-1.2	0.044	3.8
	SICO (m=10)	1.5	0.029	1.6	0.0	0.038	2.9
	SICO (m=100)	1.6	0.028	1.6	-0.2	0.038	2.9
500	QR_controls	-9.0	0.059	1.8	-8.9	0.083	3.5
	QR_cases	7.8	0.061	1.9	1.0	0.081	3.3
	QR_case-control	45.0	0.046	2.5	31.6	0.062	2.9
	IPW	-1.2	0.057	1.6	1.1	0.077	2.9
	KS	-0.3	0.062	1.9	-0.3	0.081	3.3
	SICO (m=1)	-0.8	0.061	1.9	-2.1	0.078	3.1
	SICO (m=10)	-2.2	0.054	1.5	-3.6	0.071	2.5
	SICO (m=100)	4.2	0.054	1.5	5.4	0.070	2.5

Table 2: Relative bias (RB), standard error (SE) and mean squared error (MSE) of the estimated quantile coefficients under Model (3) at quantile levels 0.5, and 0.9. In Model (3), $x_i \sim N(0,1)$, $z_i \sim N(0,1)$, and $e_i \sim N(0,1)$. "QR_controls" stands for unadjusted quantile regression using controls only. "QR_cases" stands for unadjusted quantile regression using cases only. "QR_controls" are unadjusted quantile regression using both case and control samples. IPW is the estimates using inverse probability weighting;"KS" is the KS estimates using kernel smoothing. SICO(m) is the SICO estimates with m replicate.

1.2 Bandwidth selection in KS estimates

We use kernel smoothing to approximate the expectation terms in the estimating equations, and proposed a cross-validation (CV) approach to select the *best* bandwidth. To address the question of whether the estimation is sensitive to the choice of bandwidth, and whether the proposed CV-based bandwidth selection provides a reasonable estimation, we simulate 100 Monte-Carlo samples from Models (1) and (2) respectively. For each sample, we repeatedly apply the proposed estimation procedure to a sequence of fixed bandwidths, ranging from 0.02 to 100, and then evaluate the resulting mean absolute bias with each bandwidth. To see whether the estimates from smaller sample sizes are more sensitive to bandwidth selection, we repeat this procedure on a subset of 500 cases and 500 controls. In Figure 1, we plot the mean absolute biases of the estimated quantile coefficients from Model (1) against the logarithm of their corresponding bandwidths. The horizontal line is the mean absolute bias of the estimated coefficients with CV selected bandwidth. Similarly, we plot in Figure 2 the mean absolute biases with fixed and CV selected bandwidth from Model (2). We found that the biases are well controlled within 0.02 regardless of the selection of bandwidth. Hence we conclude that the proposed method is not sensitive to the choice of bandwidth. The estimates are close for a fairly wide range of bandwidth. The estimates using CV selected optimal bandwidth outperform most of those with fixed bandwidths, which suggested that the proposed bandwidth selection works reasonably well. The advantage of CV selected bandwidth is more visible at the 0.5th quantile when the outcome is normally distributed (Model (1)), and at 0.1th quantile when the outcome follows chi-square distribution(Model (2)). In other words, the selection is more helpful for the quantile levels at which the density is higher.

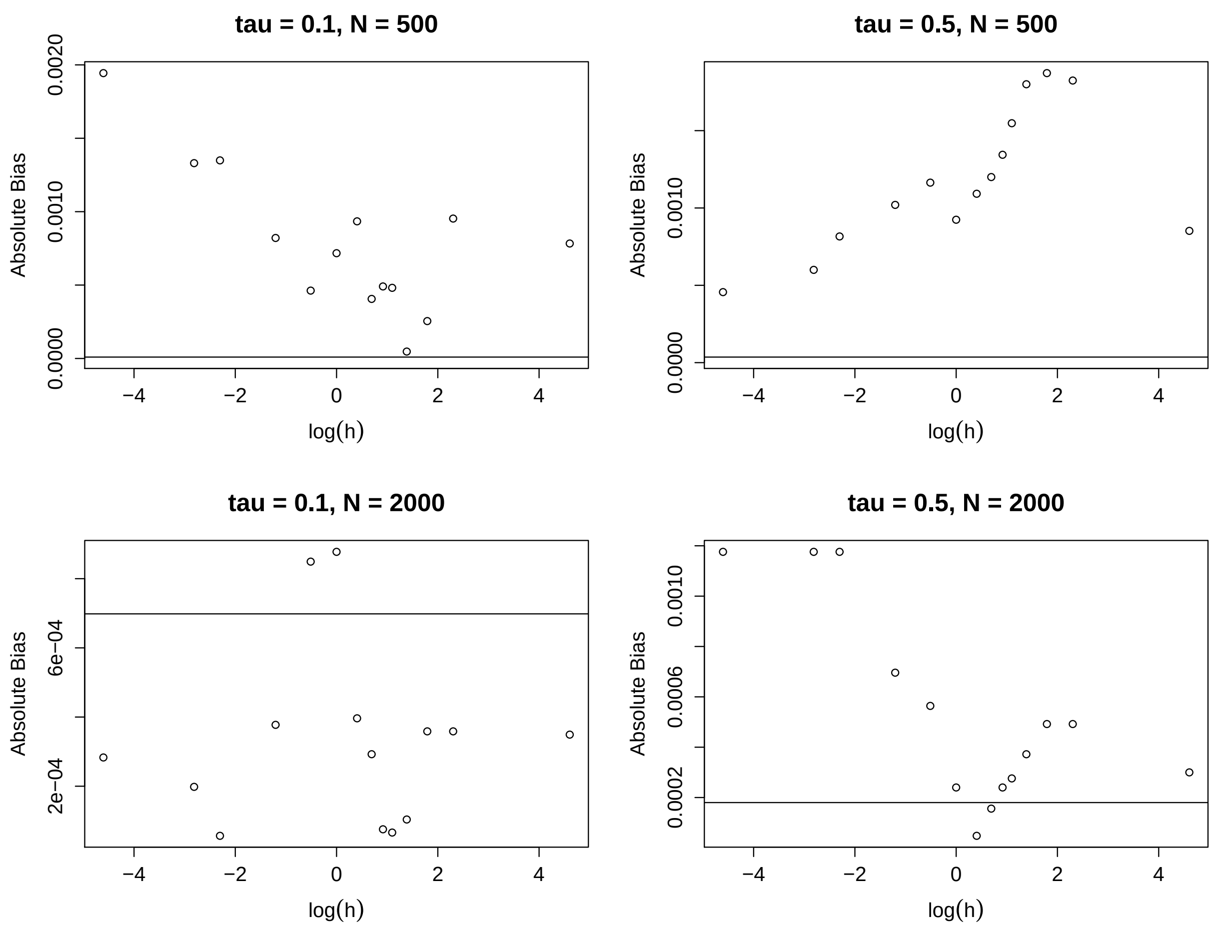

Figure 1: Mean absolute biases of the estimates with different bandwidths from Model (1). The horizontal line is the mean absolute bias of the estimated coefficients with CV selected bandwidth. The dots are the mean absolute biases of the estimated quantile coefficients with fixed bandwidths.

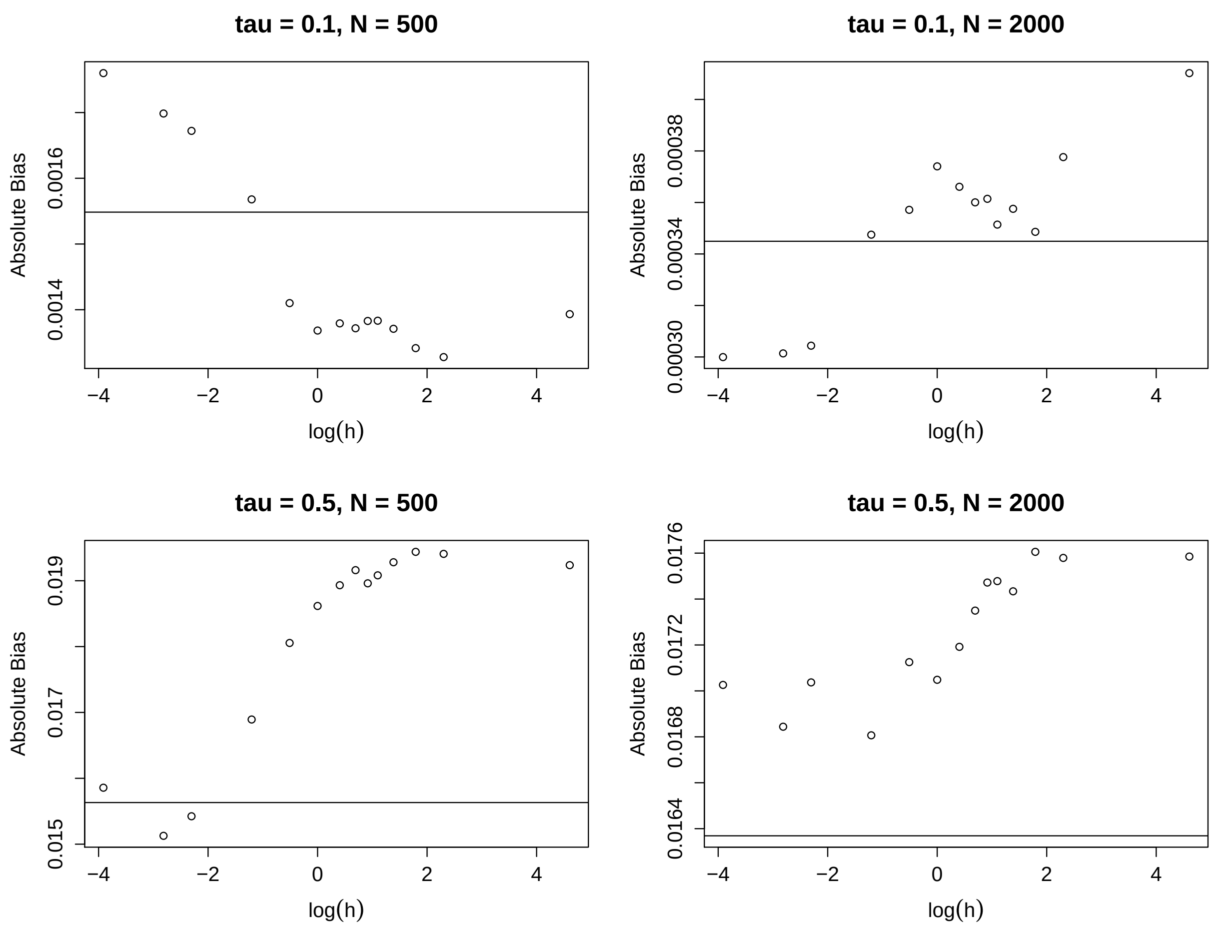

Figure 2: Mean absolute biases of the estimates with different bandwidths from Model (2). The horizontal line is the mean absolute bias of the estimated coefficients with CV selected bandwidth. The dots are the mean absolute biases of the estimated quantile coefficients with fixed bandwidths.

2 Technical Details for Theorems 2.1 and 2.2

In this appendix, we provide the technical proofs of the asymptotic behaviors of SICO estimate $\widehat{\boldsymbol{\beta}}_{n,\tau}$ and KS estimate $\widetilde{\boldsymbol{\beta}}_{n,\tau}$.

2.1 Consistency of $\widehat{\boldsymbol{\beta}}_{n,\tau}^{(\ell)}$

We first establish the following Lemma, which will be used to show the consistency of $\widehat{\boldsymbol{\beta}}^{(\ell)}$.

Lemma 1 *Let* $U_n(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n}\{[\Psi_\tau(\tilde{y}_i, \mathbf{x}_i, \boldsymbol{\beta}) - \Psi_\tau(\widehat{\tilde{y}}_i, \mathbf{x}_i, \boldsymbol{\beta})]p(1-d_i|\mathbf{x}_i)\}$, *where* $\widehat{\tilde{y}}_i$ *is a random draw from the estimated conditional quantile function* $\mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}_n^{(1-d_i)}(\tau)$, *and* $U(\boldsymbol{\beta}) = \lim_{n\to\infty} U_n$, *then under Assumptions 2 and 3, for any* ϵ, *we have*

$$pr(\sup_{\boldsymbol{\beta}\in\Theta} \|U_n(\boldsymbol{\beta}) - U(\boldsymbol{\beta})\| > \epsilon) \to 0 \qquad (1)$$

Proof of Lemma 1: Following the Huber's chaining argument, we partition the parameter space into L_n disjoint small cubes Γ_l with diameters less than C_1/n for some constant C_1. Let ξ_l be the center of the l-th cube, the left side of (1) is bounded by the sum of the following two probabilities, $P_1 + P_2$, where

$$P_1 = \text{Prob}\left(\max_l \sup_{\boldsymbol{\beta}\in\Gamma_l} \|U_n(\boldsymbol{\beta}) - U_n(\xi_l) - U(\boldsymbol{\beta}) + U(\xi_l)\| \geq \epsilon\right) \quad \text{and}$$

$$P_2 = \text{Prob}\left(\max_l \|U_n(\xi_l) - U(\xi_l)\| \geq \epsilon/2\right).$$

We first note that

$$
\begin{aligned}
&\|U_n(\boldsymbol{\beta}) - U_n(\xi_l)\| \\
= \;& n^{-1}\left\|\sum_{i=1}^{n}\left\{\Psi_\tau(\tilde{y}_i,\mathbf{x}_i,\boldsymbol{\beta}) - \Psi_\tau(\tilde{y}_i,\mathbf{x}_i,\xi_l) + \Psi_\tau(\widehat{\tilde{y}}_i,\mathbf{x}_i,\boldsymbol{\beta}) - \Psi_\tau(\widehat{\tilde{y}}_i,\mathbf{x}_i,\xi_l)\right\} p(1-d_i|\mathbf{x}_i)\right\| \\
= \;& n^{-1}\left\|\sum_{i=1}^{n}\left[I\{\tilde{y}_i \le \mathbf{x}_i^\top\boldsymbol{\beta}\} - I\{\tilde{y}_i \le \mathbf{x}_i^\top\xi_l\} + I\{\widehat{\tilde{y}}_i \le \mathbf{x}_i^\top\boldsymbol{\beta}\} - I\{\widehat{\tilde{y}}_i \le \mathbf{x}_i^\top\xi_l\}\right] p(1-d_i|\mathbf{x}_i)\mathbf{x}_i\right\| \\
\le \;& n^{-1}\left\|\sum_{i=1}^{n}\left[I\{|\tilde{y}_i - \mathbf{x}_i^\top\xi_l| \le |\mathbf{x}_i^\top(\boldsymbol{\beta}-\xi_l)|\} + I\{|\widehat{\tilde{y}}_i - \mathbf{x}_i^\top\xi_l| \le |\mathbf{x}_i^\top(\boldsymbol{\beta}-\xi_l)|\}\right] p(1-d_i|\mathbf{x}_i)\mathbf{x}_i\right\| \\
\le \;& n^{-1}\left\|\sum_{i=1}^{n}\left[I\{|\tilde{y}_i - \mathbf{x}_i^\top\xi_l| \le \|\mathbf{x}_i\|q_n\} + I\{|\widehat{\tilde{y}}_i - \mathbf{x}_i^\top\xi_l| \le \|\mathbf{x}_i\|q_n|\}\right] p(1-d_i|\mathbf{x}_i)\mathbf{x}_i\right\|
\end{aligned}
$$

Let $g_i()$ be the density of $\widetilde{y}_i$, then g_i is a continuous function and bounded away from zero and infinity according to Assumption 3. Following the mean value theorem, there exists an z_i^* such that

$$
\text{prob}(|\widetilde{y}_i - \mathbf{x}_i^\top\xi_l| \le \|\mathbf{x}_i\|q_n) = 2\|x_i\|q_n g_i(z_i^* - \mathbf{x}_i^\top\xi_l). \tag{2}
$$

On the other hand, $\text{prob}(|\widehat{\widetilde{y}}_i - \mathbf{x}_i^\top\xi_l| \le \|\mathbf{x}_i\|q_n) = \int_{\mathbf{x}_i^\top\xi_l - \|\mathbf{x}_i\|q_n}^{\mathbf{x}_i^\top\xi_l + \|\mathbf{x}_i\|q_n} \widehat{f}_n^{1-d_i}(\widetilde{y}|\mathbf{x}_i)d\widetilde{y}$ where $\widehat{f}_n(\widetilde{y}|\mathbf{x}_i)$ is the density function of $\widetilde{y}_i$ estimated from alternative population using quantile regressions. Following (A.2) in Wei, Ma and Carroll (2011), $\sup_{(y,\mathbf{x})}|\widehat{f}_n(\widetilde{y}|\mathbf{x}_i) - f(\widetilde{y}|\mathbf{x}_i)| = o_p(1)$ under Assumptions 2 and 3. Therefore, the

$$
\text{prob}(|\widehat{\widetilde{y}}_i - \mathbf{x}_i^\top\xi_l| \le \|\mathbf{x}_i\|q_n) \le 2\|x_i\|q_n g_i(z_i^* - \mathbf{x}_i^\top\xi_l) + o_p(1) \tag{3}
$$

Combing (2), (3), and the fact that $\max_i \|\mathbf{x}_i\| = O_p(1)$, we have

$$
\text{prob}(\max_l \sup_{\boldsymbol{\beta}\in\Gamma_l} \|U_n(\boldsymbol{\beta}) - U_n(\xi_l)\| \ge \epsilon/4) \to 0
$$

Following similar arguments, we can also show that $\text{prob}(\max_l \sup_{\boldsymbol{\beta}\in\Gamma_l} \|U(\boldsymbol{\beta}) - U(\xi_l)\| \ge \epsilon/4) \to 0$, which in turn implies $P_1 = o_p(1)$. On the other hand, let $u_i(l,m) = [I\{\tilde{y}_i \le$

$\mathbf{x}_i^\top \boldsymbol{\beta}\} - I\{\widehat{y}_i \le \mathbf{x}_i^\top \boldsymbol{\beta}\}]x_{m,i}p(1-d_i|\mathbf{x}_i)$, where $x_{m,i}$ is the m-th component of $\mathbf{x}_i$ with $m = 1, ..., p$. A sufficient condition for $P_2 = o_p(1)$ is that, for any x_m, we have

$$\text{Prob}\left(\max_{1\le l\le L_n; 1\le m\le p}\left|\sum_{i=1}^n u_i(l,m) - Eu_i(l,m)\right| \ge \epsilon/2\right) = o_p(1).$$

Under Assumption 2, $u_i(l,m)$ is bounded for all i's. Applying Bernstein's inequality to the probability term above, we have

$$\begin{aligned}
&\text{Prob}\left(\max_{l,m} n^{-1}\left|\sum_{i=1}^n [u_i(l,m) - Eu_i(l,m)]\right| > \epsilon/2\right) \\
\le\ & \sum_{l=1}^{L_n}\sum_{m=1}^{p} \text{Prob}\left(n^{-1}\left|\sum_{i=1}^n [u_i(l,m) - Eu_i(l,m)]\right| > \epsilon/2\right) \\
\le\ & L_n p \exp\left\{-\frac{n^2\epsilon^2}{2n\max_i x_{m,i}^2 + 2/3\max_i x_{m,i} n\epsilon}\right\} \to 0
\end{aligned}$$

when $\max_i \|\mathbf{x}_i\|^2/n \to 0$. We now have shown that both P_1 and $P_2 = o(1)$, which in turn implies the uniform convergence (1). Lemma 1 is hence proved.

Proof of the consistency of $\widehat{\boldsymbol{\beta}}_n^{(\ell)}$

Recall that $\widehat{\boldsymbol{\beta}}_{n,\tau}^{(\ell)}$ is the solution to

$$\widehat{\mathscr{S}_{n,\tau}}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ \Psi_\tau(y_i, \mathbf{x}_i, \boldsymbol{\beta}) \widehat{p}(d_i \mid \mathbf{x}_i) + \Psi_\tau(\widehat{\widetilde{y}}_i^{(\ell)}, \mathbf{x}_i, \boldsymbol{\beta}) \widehat{p}(1 - d_i \mid \mathbf{x}_i) \right\} = 0.$$

We can equivalently define

$$\widehat{\boldsymbol{\beta}}_{n,\tau}^{(\ell)} = \arg\min_{\boldsymbol{\beta}} \left\| \sum_{i=1}^{n} \left\{ \Psi_\tau(y_i, \mathbf{x}_i, \boldsymbol{\beta}) \widehat{p}(d_i \mid \mathbf{x}_i) + \Psi_\tau(\widehat{\widetilde{y}}_i^{(\ell)}, \mathbf{x}_i, \boldsymbol{\beta}) \widehat{p}(1 - d_i \mid \mathbf{x}_i) \right\} \right\|$$

We also define

$$\mathscr{S}_{n,\tau}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \{ \Psi_\tau(y_i, \mathbf{x}_i, \boldsymbol{\beta}) p(d_i \mid \mathbf{x}_i) + \Psi_\tau(\widetilde{y}_i, \mathbf{x}_i, \boldsymbol{\beta}) p(1 - d_i \mid \mathbf{x}_i) \},$$

where $\widetilde{y}_i$ is actual unobserved counter-factual outcome, and p is true conditional disease probability, and its liming function $\mathscr{S}_\tau(\boldsymbol{\beta}) = \lim_{n\to\infty} \mathscr{S}_{n,\tau}(\boldsymbol{\beta})$. Following Assumption 1, $\boldsymbol{\beta}_{0,\tau}$ uniquely minimizes $\|\mathscr{S}(\boldsymbol{\beta})\|$. Since $\mathscr{S}_\tau(\boldsymbol{\beta})$ is continuous function in $\boldsymbol{\beta}$, a sufficient condition for the consistency of $\widehat{\boldsymbol{\beta}}_{n,\tau}^{(\ell)}$ is

$$\sup_{\boldsymbol{\beta}\in\Theta} n^{-1} \|\widehat{\mathscr{S}}_{n,\tau}^{(\ell)}(\boldsymbol{\beta}) - \mathscr{S}_\tau(\boldsymbol{\beta})\| = o_p(1) \tag{4}$$

as $n \to \infty$, $k_n \to \infty$ and $k_n/n \to 0$. We first note that

$$\begin{aligned}
&\widehat{\mathscr{S}}_{n,\tau}^{(\ell)}(\boldsymbol{\beta}) - \mathscr{S}_{n,\tau}(\boldsymbol{\beta}) \\
&= \sum_{i=1}^{n} \Psi_\tau(y_i, \mathbf{x}_i, \boldsymbol{\beta}) \{\widehat{p}(d_i \mid \mathbf{x}_i) - p(d_i \mid \mathbf{x}_i)\} + \sum_{i=1}^{n} \Psi_\tau(\widehat{\widetilde{y}}_i^{(\ell)}, \mathbf{x}_i, \boldsymbol{\beta}) \{\widehat{p}(1 - d_i \mid \mathbf{x}_i) - p(1 - d_i \mid \mathbf{x}_i)\} \\
&\quad + \sum_{i=1}^{n} \{\Psi_\tau(\widehat{\widetilde{y}}_i^{(\ell)}, \mathbf{x}_i, \boldsymbol{\beta}) - \Psi_\tau(\widetilde{y}_i, \mathbf{x}_i, \boldsymbol{\beta})\} p(1 - d_i \mid \mathbf{x}_i).
\end{aligned}$$

The left side of (4) is then bounded by

$$\begin{aligned}
&\sup_{\boldsymbol{\beta}\in\Theta} n^{-1} \| \sum_{i=1}^{n} \Psi_\tau(y_i, \mathbf{x}_i, \boldsymbol{\beta}) \{\widehat{p}(d_i \mid \mathbf{x}_i) - p(d_i \mid \mathbf{x}_i)\} \| \\
&\quad + \sup_{\boldsymbol{\beta}\in\Theta} n^{-1} \| \Psi_\tau(\widehat{\widetilde{y}}_i^{(\ell)}, \mathbf{x}_i, \boldsymbol{\beta}) \{\widehat{p}(1 - d_i \mid \mathbf{x}_i) - p(1 - d_i \mid \mathbf{x}_i)\} \| + \sup_{\boldsymbol{\beta}\in\Theta} n^{-1} \|U_n\|. \\
&= I_1 + I_2 + \sup_{\boldsymbol{\beta}\in\Theta} n^{-1} \|U_n\|
\end{aligned}$$

The uniform convergency of I_1 and I_2 follows readily from Assumption 2. So we only need to show $\sup_{\boldsymbol\beta\in\Theta} n^{-1}\|U_n\| = o_p(1)$.

In what follows, we show that

$$\sup_{\boldsymbol\beta\in\Theta}\|U(\boldsymbol\beta)\| = o(1) \tag{5}$$

Following the law of large number,

$$\begin{aligned} U(\boldsymbol\beta) &= \lim_{n\to\infty} n^{-1}\left\{\sum_{i=1}^n [I\{\tilde y_i \le \mathbf{x}_i^\top\boldsymbol\beta\} - I\{\widehat{\tilde y}_i - \mathbf{x}_i^\top\boldsymbol\beta\}]\mathbf{x}_i p(1-d_i|\mathbf{x}_i)\right\} \\ &= \lim_{n\to\infty} n^{-1}\left\{\sum_{i=1}^n [F_{\tilde y_i}(\mathbf{x}_i^\top\boldsymbol\beta) - \widehat F_n(\mathbf{x}_i^\top\boldsymbol\beta)]\mathbf{x}_i p(1-d_i|\mathbf{x}_i)\right\} \end{aligned}$$

Due to the uniform convergence of $\widehat{\boldsymbol\beta}_n(\tau)$, $\sup_{\boldsymbol\beta\in\Theta}|\widehat F_n(\mathbf{x}_i^\top\boldsymbol\beta) - F_{\tilde y_i}(\mathbf{x}_i^\top\boldsymbol\beta)| = o_p(k_n^{1/2}n^{-1/2})$ as shown in Wei and Carroll (2009). Combing Assumption 2, we have $\sup_{\boldsymbol\beta\in\Theta}\|U(\boldsymbol\beta)\| = o_p(k_n^{1/2}n^{-1/2})$, which together with Lemma 1 imply (5).

2.2 Asymptotic normality of $\widehat{\boldsymbol\beta}_{n,\tau}^{(\ell)}$

Similar arguments as those used in proving Lemma 4.6 of He and Shao(1996) yield the following uniform convergence results. For any descending sequence $\delta_n \to 0$,

$$\sup_{\|\boldsymbol\beta-\boldsymbol\beta_{0,\tau}\|<\delta_n} n^{-1/2}\left\|\widehat{\mathscr{S}}_{n,\tau}^{(\ell)}(\boldsymbol\beta) - \widehat{\mathscr{S}}_{n,\tau}^{(\ell)}(\boldsymbol\beta_{0,\tau}) - E\widehat{\mathscr{S}}_{n,\tau}^{(\ell)}(\boldsymbol\beta) + E\widehat{\mathscr{S}}_{n,\tau}^{(\ell)}(\boldsymbol\beta_{0,\tau})\right\| = o_p(1).$$

Due to the consistency of $\widehat{\boldsymbol\beta}_{n,\tau}^{(\ell)}$,

$$n^{-1/2}\left\|\widehat{\mathscr{S}}_{n,\tau}^{(\ell)}(\widehat{\boldsymbol\beta}_{n,\tau}^{(\ell)}) - \widehat{\mathscr{S}}_{n,\tau}^{(\ell)}(\boldsymbol\beta_{0,\tau}) - E\widehat{\mathscr{S}}_{n,\tau}^{(\ell)}(\widehat{\boldsymbol\beta}_{n,\tau}^{(\ell)}) + E\widehat{\mathscr{S}}_{n,\tau}^{(\ell)}(\boldsymbol\beta_{0,\tau})\right\| = o_p(1).$$

Since $\widehat{\mathscr{S}}_{n,\tau}^{(\ell)}(\widehat{\boldsymbol\beta}_{n,\tau}^{(\ell)}) \approx 0$, we Taylor expand $E\widehat{\mathscr{S}}_{n,\tau}^{(\ell)}(\widehat{\boldsymbol\beta}_{n,\tau}^{(\ell)})$ around $\boldsymbol\beta_{0,\tau}$, so that

$$
\begin{aligned}
0 &\approx n^{-1/2}\widehat{\mathscr{S}}^{(\ell)}_{n,\tau}(\boldsymbol{\beta}_{0,\tau}) + n^{-1}\frac{\partial E\widehat{\mathscr{S}}^{(\ell)}_{n,\tau}(\boldsymbol{\beta}_{0,\tau})}{\partial \boldsymbol{\beta}_{0,\tau}} n^{1/2}(\widehat{\boldsymbol{\beta}}^{(\ell)}_{n,\tau} - \boldsymbol{\beta}_{0,\tau}) + o_p(1) \\
&= n^{-1/2}\sum_{i=1}^{n}\left\{\Psi_\tau(y_i,\mathbf{x}_i,\boldsymbol{\beta}_{0,\tau})p(d_i \mid \mathbf{x}_i) + \Psi_\tau(\widehat{\widetilde{y}}^{(\ell)}_i,\mathbf{x}_i,\boldsymbol{\beta}_{0,\tau})p(1-d_i \mid \mathbf{x}_i)\right\} \\
&\quad + n^{-1}\left\{\sum_{i=1}^{n}\left.\frac{E_{y_i}[\Psi_\tau(y_i,\mathbf{x}_i,\boldsymbol{\beta})|\mathbf{x}_i]}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{0,\tau}} p(d_i|\mathbf{x}_i) + \left.\frac{E_{y_i}[\Psi_\tau(\widehat{\widetilde{y}}^{(\ell)}_i,\mathbf{x}_i,\boldsymbol{\beta})|\mathbf{x}_i]}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{0,\tau}} p(1-d_i|\mathbf{x}_i)\right\} \\
&\qquad \times n^{1/2}(\widehat{\boldsymbol{\beta}}^{(\ell)}_{n,\tau} - \boldsymbol{\beta}_{0,\tau}) + o_p(1)
\end{aligned}
$$

Note that

$$
\begin{aligned}
& n^{-1}\frac{\partial E\widehat{\mathscr{S}}^{(\ell)}_{n,\tau}(\boldsymbol{\beta}_{0,\tau})}{\partial \boldsymbol{\beta}_{0,\tau}} \\
&= n^{-1}\left\{\sum_{i=1}^{n}\left.\frac{E_{y_i}[\Psi_\tau(y_i,\mathbf{x}_i,\boldsymbol{\beta})|\mathbf{x}_i]}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{0,\tau}} p(d_i|\mathbf{x}_i) + \left.\frac{E_{y_i}[\Psi_\tau(\widehat{\widetilde{y}}^{(\ell)}_i,\mathbf{x}_i,\boldsymbol{\beta})|\mathbf{x}_i]}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{0,\tau}} p(1-d_i|\mathbf{x}_i)\right\} \\
&= n^{-1}\sum_{i=1}^{n}\left\{f_{y_i}(\mathbf{x}_i^\top\boldsymbol{\beta}_{0,\tau})p(d_i|\mathbf{x}_i) + f_{\widehat{\widetilde{y}}_i}(\mathbf{x}_i^\top\boldsymbol{\beta}_{0,\tau})p(1-d_i|\mathbf{x}_i)\right\}\mathbf{x}_i \\
&= n^{-1}\sum_{i=1}^{n}\left[f_{y_i}(\mathbf{x}_i^\top\boldsymbol{\beta}_{0,\tau})p(d_i|\mathbf{x}_i) + \{f_{\widetilde{y}_i}(\mathbf{x}_i^\top\boldsymbol{\beta}_{0,\tau}) + o_p(1)\}(\mathbf{x}_i^\top\boldsymbol{\beta}_{0,\tau})p(1-d_i|\mathbf{x}_i)\right]\mathbf{x}_i \\
&= n^{-1}\sum_{i=1}^{n}\left[f_{y_i}(\mathbf{x}_i^\top\boldsymbol{\beta}_{0,\tau})p(d_i|\mathbf{x}_i) + f_{\widetilde{y}_i}(\mathbf{x}_i^\top\boldsymbol{\beta}_{0,\tau})p(1-d_i|\mathbf{x}_i)\right]\mathbf{x}_i + o_p(1) \\
&\widehat{=}\; G_n + o_p(1)
\end{aligned}
$$

On the other hand,

$$
\mathrm{var}\left[n^{-1/2}\sum_{i=1}^{n}\left\{\Psi_\tau(y_i,\mathbf{x}_i,\boldsymbol{\beta}_{0,\tau})p(d_i \mid \mathbf{x}_i) + \Psi_\tau(\widehat{\widetilde{y}}^{(\ell)}_i,\mathbf{x}_i,\boldsymbol{\beta}_{0,\tau})p(1-d_i \mid \mathbf{x}_i)\right\}\right] = V_{n,1} + V_{n,2} + 2U_{n,1},
$$

where $V_{n,1}$, $V_{n,2}$ and $U_{n,1}$ are defined in Section 2.6. Consequently,

$$
\begin{aligned}
n^{1/2}(\widehat{\boldsymbol{\beta}}_{n,\tau}^{(\ell)} - \boldsymbol{\beta}_{0,\tau}) &= -G_n^{-1} n^{-1/2} \sum_{i=1}^{n} \Big\{ \Psi_\tau(y_i, \mathbf{x}_i, \boldsymbol{\beta}_{0,\tau}) p(d_i \mid \mathbf{x}_i) + \Psi_\tau(\widehat{\widetilde{y}}_i^{(\ell)}, \mathbf{x}_i, \boldsymbol{\beta}_{0,\tau}) p(1 - d_i \mid \mathbf{x}_i) \Big\} + o_p(1) \\
&= AN(0, G_n^{-1}(V_{n,1} + V_{n,2} + 2U_{n,1}) G_n^{-1}) \qquad (6)
\end{aligned}
$$

$$
\begin{aligned}
& n^{1/2}(\widehat{\boldsymbol{\beta}}_{n,\tau} - \boldsymbol{\beta}_{0,\tau}) \\
&= m^{-1} \sum_{\ell=1}^{m} n^{1/2}(\widehat{\boldsymbol{\beta}}^{(\ell)} - \boldsymbol{\beta}_{0,\tau}) \\
&= -D_n^{-1} \frac{1}{m} \sum_{\ell=1}^{m} \left[n^{-1/2} \sum_{i=1}^{n} \Big\{ \Psi_\tau(y_i, \mathbf{x}_i, \boldsymbol{\beta}_{0,\tau}) p(d_i \mid \mathbf{x}_i) + \Psi_\tau(\widehat{\widetilde{y}}_i^{(\ell)}, \mathbf{x}_i, \boldsymbol{\beta}_{0,\tau}) p(1 - d_i \mid \mathbf{x}_i) \Big\} + o_p(1) \right] \\
&= -D_n^{-1} n^{-1/2} \sum_{i=1}^{n} \left\{ \Psi_\tau(y_i, \mathbf{x}_i, \boldsymbol{\beta}_{0,\tau}) p(d_i \mid \mathbf{x}_i) + \frac{1}{m} \sum_{\ell=1}^{m} \Psi_\tau(\widehat{\widetilde{y}}_i^{(\ell)}, \mathbf{x}_i, \boldsymbol{\beta}_{0,\tau}) p(1 - d_i \mid \mathbf{x}_i) \right\} + o_p(1) \\
&=
\end{aligned}
$$

Technical proofs

Proof of consistency Recall that

$$S_{n,\tau}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\Psi_\tau(y_i, \mathbf{x}_i, \boldsymbol{\beta})p(d_i \mid \mathbf{x}_i) + \widetilde{\Psi}_\tau(\mathbf{x}_i, \boldsymbol{\beta})p(1-d_i \mid \mathbf{x}_i),$$

where $\widetilde{\Psi}_\tau(\mathbf{x}_i, \boldsymbol{\beta})$ is the expected quantile regression estimating function given the opposite disease status of the ith subject and his covariates $\mathbf{x}_i$, i.e.,

$$\widetilde{\Psi}_\tau(\mathbf{x}_i, \boldsymbol{\beta}) = E_{\tilde{y}_i}\left\{\Psi_\tau(\tilde{y}_i, \mathbf{x}_i, \boldsymbol{\beta}) \mid \mathbf{x}_i, D_i = 1-d_i\right\},$$

and $p(d_i \mid \mathbf{x}_i)$ is the true probability of being disease given $\mathbf{x}_i$. Then the true parameter $\boldsymbol{\beta}_{0,\tau}$ is the unique solution to $S(\boldsymbol{\beta}) = 0$, where $S_\tau(\boldsymbol{\beta}) = \lim_{n\to} S_{n,\tau}(\boldsymbol{\beta})$. Further recall that $\widehat{p}(d_i|\mathbf{x}_i)$ is the estimated probability and $\widehat{\widetilde{\Psi}}_\tau(\mathbf{x}_i, \boldsymbol{\beta})$ is kernel estimated $\widetilde{\Psi}_\tau(\mathbf{x}_i, \boldsymbol{\beta})$ as in (**??**). The working estimation equations are

$$\widehat{S}_{n,\tau} = \frac{1}{n}\sum_{i=1}^{n}\Psi_\tau(y_i, \mathbf{x}_i, \boldsymbol{\beta})\widehat{p}(d_i \mid \mathbf{x}_i) + \widehat{\widetilde{\Psi}}_\tau(\mathbf{x}_i, \boldsymbol{\beta})\widehat{p}(1-d_i \mid \mathbf{x}_i),$$

and the estimator $\widehat{\boldsymbol{\beta}}_{n,\tau}$ is the solution of $\widehat{S}_{n,\tau}(\boldsymbol{\beta}) = 0$.

Following the arguments in Amemiya (1985, pp.106-108), the following conditions suffices the consistency of $\widehat{\boldsymbol{\beta}}_{n,\tau}$.

(1) $\boldsymbol{\beta}_{0,\tau} \in \Theta$ is the unique solution to $S_\tau(\boldsymbol{\beta}) = 0$;

(2) $S_\tau(\boldsymbol{\beta})$ is a continuous function for $\boldsymbol{\beta} \in \Theta$;

(3) $\sup_{\boldsymbol{\beta}\in\Theta} \|\widehat{S}_{n,\tau}(\boldsymbol{\beta}) - S_\tau(\boldsymbol{\beta})\| = o_p(1)$.

The uniqueness of $\boldsymbol{\beta}_{0,\tau}$ is assumed in Assumption 1, and the continuity of $S_\tau(\boldsymbol{\beta})$ is implied readily from the continuity of the conditional density $f(Y|X, D)$, which is assumed in Assumption 3. Hence, we only need to show the condition (3) to obtain the consistency. To do so, we bound

$$\sup_{\boldsymbol{\beta}\in\Theta}\|\widehat{S}_{n,\tau}(\boldsymbol{\beta})-S_{\tau}(\boldsymbol{\beta})\| \leq \sup_{\boldsymbol{\beta}\in\Theta}\|\widehat{S}_{n,\tau}(\boldsymbol{\beta})-S_{n,\tau}(\boldsymbol{\beta})\| + \sup_{\boldsymbol{\beta}\in\Theta}\|S_{n,\tau}(\boldsymbol{\beta})-S_{\tau}(\boldsymbol{\beta})\|. \tag{7}$$

Note that

$$\begin{aligned}\|\widehat{S}_{n,\tau}(\boldsymbol{\beta})-S_{n,\tau}(\boldsymbol{\beta})\| &= \frac{1}{n}\sum_{i=1}^{n}\left\|\Psi_{\tau}(y_i,\mathbf{x}_i,\boldsymbol{\beta})-\widetilde{\Psi}_{\tau}(\mathbf{x}_i,\boldsymbol{\beta})\right\|\{\widehat{p}(d_i|\mathbf{x}_i)-p(d_i|\mathbf{x}_i)\} \\ &\quad +\frac{1}{n}\sum_{i=1}^{n}\left\|\widehat{\widetilde{\Psi}}_{\tau}(\mathbf{x}_i,\boldsymbol{\beta})-\widetilde{\Psi}_{\tau}(\mathbf{x}_i,\boldsymbol{\beta})\right\|\widehat{p}(1-d_i|\mathbf{x}_i) \end{aligned} \tag{8}$$

Following Fan and Gilbels (1996), the bias of the kernel estimated $\widehat{\widetilde{\Psi}}_{\tau}(\mathbf{x}_i,\boldsymbol{\beta})$ converges to zero at the rate of $O(h^2)$. Since $\widetilde{\Psi}_{\tau}(\mathbf{x}_i,\boldsymbol{\beta})$ is a continuous function, the bias convergences to zero uniformly as $h \to 0$. Moreover, when $h = o(n^{-1/5})$, and $hn \to \infty$, we consequently have

$$\max_{\mathbf{x}_i}\sup_{\boldsymbol{\beta}\in\Theta}\sqrt{nh}\left\|\widehat{\widetilde{\Psi}}_{\tau}(\mathbf{x}_i,\boldsymbol{\beta})-\widetilde{\Psi}_{\tau}(\mathbf{x}_i,\boldsymbol{\beta})\right\| = o_p(1),$$

which further leads to,

$$\sup_{\boldsymbol{\beta}\in\Theta}\frac{1}{n}\sum_{i=1}^{n}\left\|\widehat{\widetilde{\Psi}}_{\tau}(\mathbf{x}_i,\boldsymbol{\beta})-\widetilde{\Psi}_{\tau}(\mathbf{x}_i,\boldsymbol{\beta})\right\|\widehat{p}(1-d_i|\mathbf{x}_i) = o(n^{-1/2}h^{-1/2}) = o_p(1). \tag{9}$$

On the other hand, since $\left\|\Psi_{\tau}(y_i,\mathbf{x}_i,\boldsymbol{\beta})-\widetilde{\Psi}_{\tau}(\mathbf{x}_i,\boldsymbol{\beta})\right\|$ is bounded away from infinity, Assumption 3 also implies

$$\sup_{\boldsymbol{\beta}\in\Theta}\frac{1}{n}\sum_{i=1}^{n}\left\|\Psi_{\tau}(y_i,\mathbf{x}_i,\boldsymbol{\beta})-\widetilde{\Psi}_{\tau}(\mathbf{x}_i,\boldsymbol{\beta})\right\|\{\widehat{p}(d_i|\mathbf{x}_i)-p(d_i|\mathbf{x}_i)\} = o_p(1). \tag{10}$$

Combining (8) -(10), we have

$$\sup_{\boldsymbol{\beta}\in\Theta}\|\widehat{S}_n(\boldsymbol{\beta})-S_n(\boldsymbol{\beta})\| = o_p(1). \tag{11}$$

On the other hand, following the similar chaining argument in Lemma 2 of Wei and Carroll (2009), we have

$$\sup_{\boldsymbol{\beta}\in\Theta}\|S_n(\boldsymbol{\beta})-S(\boldsymbol{\beta})\| = o_p(1). \tag{12}$$

The uniform convergence (11) and (12) together implies the condition (3) holds. The convergence of $\widehat{\boldsymbol{\beta}}_{n,\tau}$ is approved.

Proof of normality Following the similar argument as in Lemma 2 of Wei and Carroll(2009), we can show that for any positive descending sequence $d_n \to 0$, we have

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_{0,\tau}\|>d_n} n^{1/2}\|S_n(\boldsymbol{\beta})-S_n(\boldsymbol{\beta}_{0,\tau})-S(\boldsymbol{\beta})+S(\boldsymbol{\beta}_{0,\tau})\| = o_p(1) \tag{13}$$

Due to the consistency of $\widehat{\boldsymbol{\beta}}_{n,\tau}$, the uniform convergence (13) implies that

$$n^{1/2}\|S_n(\widehat{\boldsymbol{\beta}}_{n,\tau})-S_n(\boldsymbol{\beta}_{0,\tau})-S(\widehat{\boldsymbol{\beta}}_{n,\tau})+S(\boldsymbol{\beta}_{0,\tau})\| = o_p(1) \tag{14}$$

Note that $\widehat{\boldsymbol{\beta}}_{n,\tau}$ is the solution to $\widehat{S}_n(\boldsymbol{\beta})=0$, we have

$$\begin{aligned} o_p(1)=n^{1/2}\widehat{S}_n(\widehat{\boldsymbol{\beta}}_{n,\tau}) &= n^{1/2}S_n(\widehat{\boldsymbol{\beta}}_{n,\tau})+n^{1/2}\{\widehat{S}_n(\widehat{\boldsymbol{\beta}}_{n,\tau})-S_n(\widehat{\boldsymbol{\beta}}_{n,\tau})\} \\ &= n^{1/2}S_n(\widehat{\boldsymbol{\beta}}_{n,\tau})+n^{-1/2}\sum_{i=1}^{n}[\widehat{\widetilde{\Psi}}_\tau(\mathbf{x}_i,\boldsymbol{\beta})-\widetilde{\Psi}_\tau(\mathbf{x}_i,\boldsymbol{\beta})]p(1-d_i|\mathbf{x}_i) \\ &= n^{1/2}S_n(\widehat{\boldsymbol{\beta}}_{n,\tau})+o_p(h^{-1/2}). \end{aligned}$$

The last equality follows from (9). Consequently, $n^{1/2}S_n(\widehat{\boldsymbol{\beta}}_{n,\tau})=o_p(1)$, which in turn implies the following convergence.

$$n^{1/2}\|S_n(\boldsymbol{\beta}_{0,\tau})-S(\widehat{\boldsymbol{\beta}}_{n,\tau})\| = o_p(1) \tag{15}$$

Taylor expending $S(\widehat{\boldsymbol{\beta}}_{n,\tau})$ around $S(\boldsymbol{\beta}_{0,\tau})$, we have

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_{n,\tau}-\boldsymbol{\beta}_{0,\tau}) = -G_n^{-1}S_n(\boldsymbol{\beta}_{0,\tau})+o_p(1),$$

where

$$G_n = n^{-1}\sum_{i=1}^{n}\left\{ \frac{\partial E_{y_i}[\Psi_\tau(y_i,\mathbf{x}_i,\boldsymbol{\beta})|\mathbf{x}_i,d_i]}{\partial\boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{0,\tau}} p(d_i|\mathbf{x}_i)+\frac{\partial\widetilde{\Psi}_\tau(\mathbf{x}_i,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{0,\tau}} p(1-d_i|\mathbf{x}_i)\right\}.$$

Under the Assumptions 4 and 5, $\lim_{n\to\infty} G_n \to G_0$ in probability, and $\lim_{n\to\infty} V_n = V$. It follows immediately from the Central Limit Theorem that

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_{n,\tau} - \boldsymbol{\beta}_{n,\tau}) = AN(0, G_0 V_0^{-1} G_0).$$

References

[1] Amemiya, T. (1985), *Advanced Econometrics*. Harvard University Press.

[2] Fan, J. and Gilbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London.

[3] He, X. and Shao, Q. (1996). A general Bahadur representation of M-estimators and its appli- cation to linear regression with nonstochastic designs. *Annals of Statistics*, Vol. 24, 2608-2630.

[4] Wei, Y. and Carroll, R. J. (2009), Quantile Regression With Measurement Error. *Journal of the American Statistical Association*. Vol. 104(487), 1129-1143.

[5] Wei, Y., Ma, Y, and Carroll, R. J. (2012), Multiple Imputation in quantile regression. *Biometrka*, 1-16.