Supplement: A Semiparametric Quantile Single-Index Model for Zero-Inflated Outcomes

Zirui Wang

Department of Statistics and Data Science, Tsinghua University

Tianying Wang

Department of Statistics, Colorado State University

S1 Asymptotic Properties of Estimation

In this section, we present the proofs of the theorems proposed in the main text, as well as the asymptotic property of the Average Quantile Effect (AQE).

S1.1 Proof of Theorems in the main text

In this section, we sketch technical arguments for proving the theorems.

Proof of Theorem 1 First, we consider the asymptotic distribution of

$$\begin{aligned} & \sqrt{n}\left\{\hat{\beta}\circ\Gamma(\tau;\mathbf{x},\hat{\gamma}_n)-\beta\circ\Gamma(\tau;\mathbf{x},\gamma)\right\} \\ = & \sqrt{n}\left\{\hat{\beta}\circ\Gamma(\tau;\mathbf{x},\hat{\gamma}_n)-\beta\circ\Gamma(\tau;\mathbf{x},\hat{\gamma}_n)+\beta\circ\Gamma(\tau;\mathbf{x},\hat{\gamma}_n)-\beta\circ\Gamma(\tau;\mathbf{x},\gamma)\right\}. \end{aligned}$$

By Ma and He (2016a) and Assumption (2.1), we have

$$\sqrt{n}\left(\hat{\beta}\circ\Gamma(\tau;\mathbf{x},\gamma)-\beta\circ\Gamma(\tau;\mathbf{x},\gamma)\right)\xrightarrow{d}\Sigma_1,$$

where Σ_1 is defined in Theorem 1. And by the nature of logistic regression, we have

$$\sqrt{n}\left(\Gamma(\tau;\mathbf{x},\hat{\gamma}_n)-\Gamma(\tau;\mathbf{x},\gamma)\right)\xrightarrow{d}V,$$

where $V=\{1-\Gamma(\tau;\mathbf{x},\gamma)\}^2\{1-\pi(\gamma,\mathbf{x})\}^2\mathbf{x}^\top D_{1,\gamma}^{-1}\mathbf{x}$. Then, we have the asymptotic distribution

$$\sqrt{n}\begin{pmatrix}\hat{\beta}\circ\Gamma(\tau;\mathbf{x},\gamma)-\beta\circ\Gamma(\tau;\mathbf{x},\gamma)\\ \Gamma(\tau;\mathbf{x},\hat{\gamma}_n)-\Gamma(\tau;\mathbf{x},\gamma)\end{pmatrix}\xrightarrow{d}N\begin{pmatrix}\Sigma_1 & 0\\ 0 & V\end{pmatrix}.$$

By delta method, we get the asymptotic distribution of $\hat{\beta}\circ\Gamma(\tau;\mathbf{x},\hat{\gamma}_n)-\beta\circ\Gamma(\tau;\mathbf{x},\gamma)$:

$$\sqrt{n}\left\{\hat{\beta}\circ\Gamma(\tau;\mathbf{x},\hat{\gamma}_n)-\beta\circ\Gamma(\tau;\mathbf{x},\gamma)\right\}\xrightarrow{d}N(0,\Sigma_1+\Sigma_2),$$

where Σ_1 and Σ_2 are defined in Theorem 1.

Proof of Theorem 2 We prove this theorem under three scenarios: $\tau<1-\pi(\gamma,\mathbf{x})$, $\tau>1-\pi(\gamma,\mathbf{x})$, and $\tau=1-\pi(\gamma,\mathbf{x})$.

If $\tau < 1 - \pi(\gamma, \mathbf{x})$, the consistency of the estimator can be shown as in Ling et al. (2022) Theorem 1.

If $\tau > 1 - \pi(\gamma, \mathbf{x})$, $P(R_{3,n}) \to 1$ and $P(R_{1,n} \cup R_{2,n}) \to 0$. We have

$$\begin{aligned} & P\left(\left|\widehat{Q}_Y(\tau \mid \mathbf{x}) - Q_Y(\tau \mid \mathbf{x})\right| > \epsilon\right) \\ & \quad \le P(R_{1,n} \cup R_{2,n}) + P\left(\left|\widehat{Q}_Y(\tau \mid \mathbf{x}) - Q_Y(\tau \mid \mathbf{x})\right| > \epsilon, R_{3,n}\right). \end{aligned}$$

By the definition of $\widehat{Q}_Y(\tau \mid \mathbf{x})$ in eq (2.9), we have

$$\begin{aligned} & P\left(\left|\widehat{Q}_Y(\tau \mid \mathbf{x}) - Q_Y(\tau \mid \mathbf{x})\right| > \epsilon, R_{3,n}\right) \\ & \quad = P\Bigg\{\Bigg| G_{\Gamma(\tau;\mathbf{x},\gamma)}\left(\mathbf{x}^\top \beta \circ \Gamma(\tau; \mathbf{x}, \gamma)\right) \\ & \qquad - B\left(\mathbf{x}^\top \hat{\beta} \circ \Gamma(\tau; \mathbf{x}, \hat{\gamma}_n)\right)^\top \tilde{\theta}_n \left(\hat{\beta} \circ \Gamma(\tau; \mathbf{x}, \hat{\gamma}_n), \Gamma(\tau; \mathbf{x}, \hat{\gamma}_n)\right) \Bigg| > \epsilon \Bigg\}. \end{aligned}$$

For ease of notation, we adopt the definition of $\tau_s = \Gamma(\tau; \mathbf{x}, \gamma)$ and $\hat{\tau}_s = \Gamma(\tau; \mathbf{x}, \hat{\gamma}_n)$, and rewrite the above equation as below:

$$\begin{aligned} & P\left(\left|\widehat{Q}_Y(\tau \mid \mathbf{x}) - Q_Y(\tau \mid \mathbf{x})\right| > \epsilon, R_{3,n}\right) \\ = \; & P\left\{\left| G_{\tau_s}\left(\mathbf{x}^\top \beta_{\tau_s}\right) - B\left(\mathbf{x}^\top \hat{\beta}_{\hat{\tau}_s}\right)^\top \tilde{\theta}_n\left(\hat{\beta}_{\hat{\tau}_s}, \hat{\tau}_s\right)\right| > \epsilon\right\}. \end{aligned} \tag{S1.1}$$

We first define $\tilde{G}_{\tau_s}(u, \beta)$ as the τ_sth quantile function given $\mathbf{x}^\top \beta = u$, where $\tau_s = \Gamma(\tau; \mathbf{x}, \gamma)$. Thus, we can rewrite $G_{\tau_s}\left(\mathbf{x}^\top \beta_{\tau_s}\right) = \tilde{G}_{\tau_s}\left(\mathbf{x}^\top \beta_{\tau_s}, \beta_{\tau_s}\right)$. Also, by Corollary 6.21 of Schumaker (2007), for any $\tau_s \in (0, 1)$ and $\beta \in \Theta$ there exists $\theta(\tau_s, \beta)$, such that $G^0_{\tau_s}(u, \beta) = B(u)^\top \theta(\tau_s, \beta)$, where $G^0_{\tau_s}(u, \beta)$

satisfies

$$\sup_{u\in[a_0,b_0]}\left|\tilde{G}_{\tau_s}(u,\beta)-G^0_{\tau_s}(u,\beta)\right|\leq CJ_n^{-r}, \tag{S1.2}$$

C is a constant, r refers to the same notation in Assumption (3.2), and J_n refers to the dimension of the B-spline function space.

We rewrite $\widehat{Q}_Y(\tau \mid \mathbf{x})-Q_Y(\tau \mid \mathbf{x})$ by several parts

$$\begin{aligned}
& G_{\tau_s}\left(\mathbf{x}^\top\beta_{\tau_s}\right)-B\left(\mathbf{x}^\top\hat{\beta}_{\hat{\tau}_s}\right)^\top\tilde{\theta}_n\left(\hat{\beta}_{\hat{\tau}_s},\hat{\tau}_s\right)\\
=\ & \tilde{G}_{\tau_s}\left(\mathbf{x}^\top\beta_{\tau_s},\beta_{\tau_s}\right)-B\left(\mathbf{x}^\top\hat{\beta}_{\hat{\tau}_s}\right)^\top\tilde{\theta}_n\left(\hat{\beta}_{\hat{\tau}_s},\hat{\tau}_s\right)\\
=\ & \tilde{G}_{\tau_s}\left(\mathbf{x}^\top\beta_{\tau_s},\beta_{\tau_s}\right)-G^0_{\tau_s}\left(\mathbf{x}^\top\beta_{\tau_s},\beta_{\tau_s}\right)\\
& \quad +G^0_{\tau_s}\left(\mathbf{x}^\top\beta_{\tau_s},\beta_{\tau_s}\right)-B\left(\mathbf{x}^\top\hat{\beta}_{\hat{\tau}_s}\right)^\top\tilde{\theta}_n\left(\hat{\beta}_{\hat{\tau}_s},\hat{\tau}_s\right).
\end{aligned}$$

Then by eq (S1.2), we have

$$\tilde{G}_{\tau_s}\left(\mathbf{x}^\top\beta_{\tau_s},\beta_{\tau_s}\right)-G^0_{\tau_s}\left(\mathbf{x}^\top\beta_{\tau_s},\beta_{\tau_s}\right)=o_P(1). \tag{S1.3}$$

By Lemma S.3 in Ma and He (2016b), for any $\beta\in\Theta$, $u\in[a_0,b_0]$, and $\tau\in(0,1)$, we have

$$\left|\widehat{G}_{\tau_s,n}(u,\beta)-G^0_{\tau_s}(u,\beta)\right| = \left|B(u)^\top\left\{\tilde{\theta}_n(\beta,\tau_s)-\theta(\tau_s,\beta)\right\}\right| \tag{S1.4}$$

$$\sim J_n^{3/2}n_0^{-1/2}+J_n^{-r+3/2}=o_P(1), \tag{S1.5}$$

where $\widehat{G}_{\tau_s,n}(u,\beta)$ denotes the spline estimator of $\tilde{G}_{\tau_s}(u,\beta)$ given knowing the real value of the quantile nominal τ_s. Besides, since $\hat{\beta}\circ\Gamma(\tau;\mathbf{x},\hat{\gamma}_n)$ is a consistent estimator of $\beta \circ \Gamma(\tau; \mathbf{x}, \gamma)$(Theorem 1), we have

$$\widehat{G}_{\tau_s,n}\left(\mathbf{x}^\top \beta_{\tau_s}, \beta_{\tau_s}\right) - B\left(\mathbf{x}^\top \hat{\beta}_{\hat{\tau}_s}\right)^\top \tilde{\theta}_n\left(\hat{\beta}_{\hat{\tau}_s}, \hat{\tau}_s\right) = o_P(1). \tag{S1.6}$$

From eq (S1.4) and (S1.6), we can deduce that

$$G^0_{\tau_s}\left(\mathbf{x}^\top \beta_{\tau_s}, \beta_{\tau_s}\right) - B\left(\mathbf{x}^\top \hat{\beta}_{\hat{\tau}_s}\right)^\top \tilde{\theta}_n\left(\hat{\beta}_{\hat{\tau}_s}, \hat{\tau}_s\right) = o_P(1). \tag{S1.7}$$

By eq (S1.3) and (S1.7), we have

$$P\left(\left|\widehat{Q}_Y(\tau \mid \mathbf{x}) - Q_Y(\tau \mid \mathbf{x})\right| > \epsilon, R_{3,n}\right) \to 0.$$

Thus, we conclude that when $\tau > 1 - \pi(\gamma, \mathbf{x})$,

$$P\left(\left|\widehat{Q}_Y(\tau \mid \mathbf{x}) - Q_Y(\tau \mid \mathbf{x})\right| > \epsilon\right) \to 0.$$

If $\tau = 1 - \pi(\gamma, \mathbf{x})$, then we have

$$\begin{aligned}
& P\left(\left|\widehat{Q}_Y(\tau \mid \mathbf{x}) - 0\right| > \epsilon\right) \\
= \; & P\left(\left|\widehat{Q}_Y(\tau \mid \mathbf{x}) - 0\right| > \epsilon, R_{1,n}\right) + P\left(\left|\widehat{Q}_Y(\tau \mid \mathbf{x}) - 0\right| > \epsilon, R_{2,n}\right) \\
& \quad + P\left(\widehat{Q}_Y\left(|\tau \mid \mathbf{x}) - 0\right| > \epsilon, R_{3,n}\right) \\
\leq \; & 0 + P\left(\left|\widehat{Q}_Y(\tau \mid \mathbf{x})\right| > \epsilon, R_{2,n}\right) + P(R_{3,n}) \\
= \; & P\left(\left|\widehat{Q}_Y(\tau \mid \mathbf{x})\right| > \epsilon, R_{2,n}\right) + P(R_{3,n}).
\end{aligned} \tag{S1.8}$$

By the definition of $\widehat{Q}_Y(\tau \mid \mathbf{x})$ in eq (2.9), we have

$$
\begin{aligned}
& P\left(\left|\widehat{Q}_Y(\tau \mid \mathbf{x})\right| > \epsilon, R_{2,n}\right) \\
&= P\left\{\left|\widehat{G}_{n^{-\delta}\pi(\hat{\gamma}_n,\mathbf{x})^{-1}}\left(\mathbf{x}^\top \hat{\beta}_{n^{-\delta}\pi(\hat{\gamma}_n,\mathbf{x})^{-1}}\right) \times n^\delta \{\pi(\hat{\gamma}_n, \mathbf{x}) - \pi(\gamma, \mathbf{x})\}\right| > \epsilon\right\} \\
&= P\left\{\left|B\left(\mathbf{x}^\top \hat{\beta}_{n^{-\delta}\pi(\hat{\gamma}_n,\mathbf{x})^{-1}}\right)^\top \tilde{\theta}_n\left(\hat{\beta}_{n^{-\delta}\pi(\hat{\gamma}_n,\mathbf{x})^{-1}}, n^{-\delta}\pi(\hat{\gamma}_n, \mathbf{x})^{-1}\right)\right.\right. \\
&\qquad \left.\left. \times n^\delta \{\pi(\hat{\gamma}_n, \mathbf{x}) - \pi(\gamma, \mathbf{x})\}\right| > \epsilon\right\}.
\end{aligned}
$$

Then, by the proof for the part $\tau > 1 - \pi(\gamma, \mathbf{x})$ in Theorem 2 and the property of logistic regression, we have

$$
B\left(\mathbf{x}^\top \hat{\beta}_{n^{-\delta}\pi(\hat{\gamma}_n,\mathbf{x})^{-1}}\right)^\top \tilde{\theta}_n\left(\hat{\beta}_{n^{-\delta}\pi(\hat{\gamma}_n,\mathbf{x})^{-1}}, n^{-\delta}\pi(\hat{\gamma}_n, \mathbf{x})^{-1}\right) \times n^{-\frac{1}{2}+\delta} = o_P(1),
$$

and

$$
\sqrt{n}\{\pi(\hat{\gamma}_n, \mathbf{x}) - \pi(\gamma, \mathbf{x})\} = O_P(1).
$$

Thus, we have $P\left(\widehat{Q}_Y(\tau \mid \mathbf{x}), R_{2,n}\right) \to 0$ holds. Note that the consistency only depends on the fact that $\delta < 0.5$. Finally, with $P(R_{3,n}) \to 0$, we conclude that

$$
P\left(\left|\widehat{Q}_Y(\tau \mid \mathbf{x}) - 0\right| > \epsilon\right) \to 0.
$$

Proof of Theorem 3 We prove this theorem under three scenarios: $\tau < 1 - \pi(\gamma, \mathbf{x})$, $\tau > 1 - \pi(\gamma, \mathbf{x})$, and $\tau = 1 - \pi(\gamma, \mathbf{x})$.

Proof of the convergence when $\tau < 1-\pi(\mathbf{x};\gamma)$ The proof of $\sqrt{n}\{\widehat{Q}(\tau \mid \mathbf{x}, Y > 0) - 0\} \xrightarrow{p} 0$ is the same as in Ling et al. (2022).

Proof of the asymptotic property when $\tau > 1-\pi(\mathbf{x};\gamma)$ To derive the convergence rate of

$$B\left(\mathbf{x}^\top \hat{\beta}_{\hat{\tau}_s}\right)^\top \tilde{\theta}_n\left(\hat{\beta}_{\hat{\tau}_s}, \hat{\tau}_s\right) - G_{\Gamma(\tau;\mathbf{x},\gamma)}\left(\mathbf{x}^\top \beta \circ \Gamma(\tau;\mathbf{x},\gamma)\right),$$

we rewrite it in several parts:

$$\begin{aligned}
& B\left(\mathbf{x}^\top \hat{\beta}_{\hat{\tau}_s}\right)^\top \tilde{\theta}_n\left(\hat{\beta}_{\hat{\tau}_s}, \hat{\tau}_s\right) - G_{\Gamma(\tau;\mathbf{x},\gamma)}\left(\mathbf{x}^\top \beta \circ \Gamma(\tau;\mathbf{x},\gamma)\right) \\
= \; & B\left(\mathbf{x}^\top \hat{\beta}_{\hat{\tau}_s}\right)^\top \tilde{\theta}_n\left(\hat{\beta}_{\hat{\tau}_s}, \hat{\tau}_s\right) - G_{\hat{\tau}_s}\left(\mathbf{x}^\top \beta_{\hat{\tau}_s}\right) \\
& \quad + G_{\hat{\tau}_s}\left(\mathbf{x}^\top \beta_{\hat{\tau}_s}\right) - G_{\Gamma(\tau;\mathbf{x},\gamma)}\left(\mathbf{x}^\top \beta \circ \Gamma(\tau;\mathbf{x},\gamma)\right).
\end{aligned}$$

We further denote

$$\mathbf{A} = B\left(\mathbf{x}^\top \hat{\beta}_{\hat{\tau}_s}\right)^\top \tilde{\theta}_n\left(\hat{\beta}_n(\hat{\tau}_s), \hat{\tau}_s\right) - G_{\hat{\tau}_s}\left(\mathbf{x}^\top \beta_{\hat{\tau}_s}\right),$$

$$\mathbf{B} = G_{\hat{\tau}_s}\left(\mathbf{x}^\top \beta_{\hat{\tau}_s}\right) - G_{\Gamma(\tau;\mathbf{x},\gamma)}\left(\mathbf{x}^\top \beta \circ \Gamma(\tau;\mathbf{x},\gamma)\right).$$

By Assumption (3.3), we have

$$\begin{aligned}
\mathbf{B} &= G_{\hat{\tau}_s}\left(\mathbf{x}^\top \beta_{\hat{\tau}_s}\right) - G_{\Gamma(\tau;\mathbf{x},\gamma)}\left(\mathbf{x}^\top \beta \circ \Gamma(\tau;\mathbf{x},\gamma)\right) \\
&= \frac{\partial G_\tau\left(\mathbf{x}^\top \beta_{\tau_s}\right)}{\partial \tau}\left(\hat{\tau}_s - \Gamma(\tau;\mathbf{x},\gamma)\right) + o\left(\hat{\tau}_s - \Gamma(\tau;\mathbf{x},\gamma)\right) \\
&= O\left(\hat{\tau}_s - \Gamma(\tau;\mathbf{x},\gamma)\right) \\
&= O(\hat{\tau}_s - \tau_s).
\end{aligned}$$

By the $\sqrt{n}$-convergence of $\hat{\tau}_s \to \tau_s$, as $n \to \infty$, we have

$$\mathbf{B} = O(\hat{\tau}_s - \tau_s) = O_P(n^{-\frac{1}{2}}). \tag{S1.9}$$

Then, we rewrite $\mathbf{A}$ as $\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2$, where

$$\mathbf{A}_1 = B\left(\mathbf{x}^\top \hat{\beta}_{\hat{\tau}_s}\right)^\top \tilde{\theta}_n\left(\hat{\beta}_{\hat{\tau}_s}, \hat{\tau}_s\right) - B\left(\mathbf{x}^\top \beta_{\hat{\tau}_s}\right)^\top \tilde{\theta}_n\left(\beta_{\hat{\tau}_s}, \hat{\tau}_s\right),$$

and

$$\mathbf{A}_2 = B\left(\mathbf{x}^\top \beta_{\hat{\tau}_s}\right)^\top \tilde{\theta}_n\left(\beta_{\hat{\tau}_s}, \hat{\tau}_s\right) - G_{\hat{\tau}_s}\left(\mathbf{x}^\top \beta_{\hat{\tau}_s}\right).$$

The order of $\mathbf{A}_1$ is given by the $\sqrt{n}$-convergence of $\hat{\beta}_{\hat{\tau}_s}$. For the order of the convergence rate of $\mathbf{A}_2$, we further rewrite $\mathbf{A}_2$ by two parts:

$$\mathbf{A}_2 = \mathbf{A}_{21} + \mathbf{A}_{22},$$

where

$$\mathbf{A}_{21} = B\left(\mathbf{x}^\top \beta_{\hat{\tau}_s}\right)^\top \tilde{\theta}_n\left(\beta_{\hat{\tau}_s}, \hat{\tau}_s\right) - G^0_{\hat{\tau}_s}\left(u, \beta_{\hat{\tau}_s}\right)$$

and

$$\mathbf{A}_{22} = G^0_{\hat{\tau}_s}\left(u, \hat{\beta}_{\tau_s}\right) - \tilde{G}_{\hat{\tau}_s}(u, \beta).$$

By Corollary 6.21 of Schumaker (2007), for any given $\beta \in \Theta$ and $\hat{\tau}_s \in (0,1)$, there exists $\theta^0_{\hat{\tau}_s}(\beta) \in \mathbb{R}^{J_n}$, s.t. $G^0_{\hat{\tau}_s}(u, \beta) = B(u)^\top \theta^0_{\hat{\tau}_s}(\beta) \in \mathcal{H}_r$ and

$$\sup_{u \in [a_0, b_0]} \left|\tilde{G}_{\hat{\tau}_s}(u, \beta) - G^0_{\hat{\tau}_s}(u, \beta)\right| \leq \left|\tilde{C}_m(\beta)\right| J_n^{-r},$$

for some continuous function $\tilde{C}_m(\beta)$ depending on m and C_0, where J_n denotes the dimension of B-spline function $B(u)$ and C_0 is defined in Assumption (3.4). Also, we assume $C_m = \sup_{\beta\in\Theta}\tilde{C}_m(\beta) < \infty$. Thus, we have

$$\mathbf{A}_{22} = G^0_{\hat{\tau}_s}\left(u, \hat{\beta}_{\tau_s}\right) - \tilde{G}_{\hat{\tau}_s}(u,\beta) = O_P(J_n^{-r}). \tag{S1.10}$$

Under Assumption 3, by Ma and He (2016b) eq (S.4), as $n \to \infty$, we have

$$\left\|\tilde{\theta}_{\hat{\tau}_s}(\beta_{\hat{\tau}_s}) - \theta^0_{\hat{\tau}_s}(\beta_{\hat{\tau}_s})\right\|_2 = O_P\left(J_n n_0^{-\frac{1}{2}} + J_n^{-r+\frac{1}{2}}\right).$$

By the condition regarding n and n_0 given in Assumption (2.1), we have

$$\left\|\tilde{\theta}_{\hat{\tau}_s}(\beta_{\hat{\tau}_s}) - \theta^0_{\hat{\tau}_s}(\beta_{\hat{\tau}_s})\right\|_2 = O_P\left(J_n n^{-\frac{1}{2}} + J_n^{-r+\frac{1}{2}}\right). \tag{S1.11}$$

Then, by eq (S1.11) and Cauchy-Schwarz inequality, under Assumption 3, for $\forall u \in [a_0, b_0]$, we have

$$\begin{aligned}\left|B(u)^\top\left\{\tilde{\theta}_{\hat{\tau}_s}(\beta_{\hat{\tau}_s}) - \theta^0_{\hat{\tau}_s}(\beta_{\hat{\tau}_s})\right\}\right| &\le \sqrt{\left(\sum_{i=1}^{J_n} B_i(u)^2\right)} \times \left\|\tilde{\theta}_{\hat{\tau}_s}(\beta_{\hat{\tau}_s}) - \theta^0_{\hat{\tau}_s}(\beta_{\hat{\tau}_s})\right\|_2 \\ &= O_P\left(J_n n^{-\frac{1}{2}} + J_n^{-r+\frac{1}{2}}\right).\end{aligned}$$

Consequently,

$$\sup_{u\in[a_0,b_0]}\left|B(u)^\top\left\{\tilde{\theta}_{\hat{\tau}_s}(\beta(\hat{\tau}_s)) - \theta^0_{\hat{\tau}_s}(\beta(\hat{\tau}_s))\right\}\right| = O_P\left(J_n^{\frac{1}{2}} n^{-\frac{1}{2}} + J_n^{-r}\right). \tag{S1.12}$$

By eq (S1.10) and eq (S1.12), we have

$$\mathbf{A}_2 = O_P\left(J_n^{\frac{1}{2}} n^{-\frac{1}{2}} + J_n^{-r}\right).$$

Since $\mathbf{A}_2$ converges slower than than $\mathbf{A}_1$, we have

$$\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2 = O_P\left(J_n^{\frac{1}{2}} n^{-\frac{1}{2}} + J_n^{-r}\right).$$

Proof of the asymptotic property when $\tau = 1 - \pi(\mathbf{x};\gamma)$ At the change point $\tau = 1 - \pi(\mathbf{x};\gamma)$, we have

$$\begin{aligned}
& \sqrt{n}\left\{\widehat{Q}_Y(\tau \mid Y > 0, \mathbf{x}) - 0\right\} \\
= & \sqrt{n}\left\{\widehat{Q}_Y(\tau \mid Y > 0, \mathbf{x}) I(R_{2,n})\right\} + \sqrt{n}\left\{\widehat{Q}_Y(\tau \mid Y > 0, \mathbf{x}) I(R_{3,n})\right\} \\
= & \sqrt{n}\left\{\widehat{Q}_Y(\tau \mid Y > 0, \mathbf{x}) I(R_{2,n})\right\} + o_P(1) \\
= & n^{\delta} B\left(\mathbf{x}^\top \hat{\beta}_{n^{-\delta}\pi(\hat{\gamma}_n,\mathbf{x})^{-1}}\right)^\top \tilde{\theta}_n\left(\hat{\beta}_{n^{-\delta}\pi(\hat{\gamma}_n,\mathbf{x})^{-1}}, n^{-\delta}\pi(\hat{\gamma}_n,\mathbf{x})^{-1}\right) \\
& \quad \times \sqrt{n}\left\{\pi(\hat{\gamma}_n,\mathbf{x}) - \pi(\gamma,\mathbf{x})\right\} I\left(0 < \pi(\hat{\gamma}_n,\mathbf{x}) - \pi(\gamma,\mathbf{x}) < n^{-\delta}\right) + o_P(1) \\
= & n^{\delta} \widehat{Q}_Y\left(n^{-\delta}\pi(\hat{\gamma}_n,\mathbf{x}) \mid \mathbf{x}, Y > 0\right) \\
& \quad \times \sqrt{n}\left\{\pi(\hat{\gamma}_n,\mathbf{x}) - \pi(\gamma,\mathbf{x})\right\} I(0 < \pi(\hat{\gamma}_n,\mathbf{x}) - \pi(\gamma,\mathbf{x}) < n^{-\delta}) + o_P(1).
\end{aligned}$$

The second equation above comes from the fact that, $\forall \epsilon > 0$,

$$P\left\{\left|\sqrt{n}\left\{\widehat{Q}_Y(\tau \mid Y > 0, \mathbf{x}) I(R_{3,n})\right\}\right| > \epsilon\right\} \leq P(R_{3,n}) \to 0.$$

By Ling et al. (2022), we have

$$\begin{aligned}
& \sqrt{n}\left\{\pi(\hat{\gamma}_n,\mathbf{x}) - \pi(\gamma,\mathbf{x})\right\} I\left(0 < \pi(\hat{\gamma}_n,\mathbf{x}) - \pi(\gamma,\mathbf{x}) < n^{-\delta}\right) \\
& \quad \xrightarrow{d} \pi(\gamma,\mathbf{x})\left\{1 - \pi(\gamma,\mathbf{x})\right\} \sqrt{\mathbf{x}^\top D_{1,\gamma}\mathbf{x}} Z_0 I(Z_0 > 0), \qquad \text{(S1.13)}
\end{aligned}$$

where $Z_0 \sim N(0,1)$. Then, we only need to prove that

$$n^{\delta}\widehat{Q}_Y\left(n^{-\delta}\pi(\hat{\gamma}_n, \mathbf{x}) \mid \mathbf{x}, Y>0\right) \xrightarrow{p} Q_Y^{'}(0 \mid \mathbf{x}, Y>0)\pi(\gamma, \mathbf{x})^{-1}. \quad \text{(S1.14)}$$

To ensure the $\sqrt{n}$-convergence in the proof below, we set $\delta \leq 0.250$ based on the lower and upper bounds for J_n:

$$\max\left\{(n\log(n))^{1/(3r-1/2)}, n^{1/(2r+2)}\right\} \ll J_n \ll n^{1/4}/(\log n)^{5/4}$$

as given in Theorem 2 of Ma and He (2016a). Alternatively, one can also choose $\delta \in (0.25, 0.50)$, though the convergence rate at $\tau = 1-\pi(\mathbf{x};\gamma)$ will adjust accordingly. By the proof of Theorem 3 (iii) and the bounds for J_n above, when $\delta \leq 0.250$, we can deduce that:

$$n^{\delta}\left\{\widehat{Q}_Y\left(n^{-\delta}\pi^{-1}(\hat{\gamma}_n, \mathbf{x}) \mid \mathbf{x}, Y>0\right) - Q_Y\left(n^{-\delta}\pi^{-1}(\gamma, \mathbf{x}) \mid \mathbf{x}, Y>0\right)\right\} \xrightarrow{p} 0 \quad \text{(S1.15)}$$

Then, by Taylor expansion, we have:

$$Q_Y\left(n^{-\delta}\pi^{-1}(\gamma, \mathbf{x}) \mid \mathbf{x}, Y>0\right) = Q_Y(0 \mid \mathbf{x}, Y>0) + Q_Y^{'}(0 \mid \mathbf{x}, Y>0)\pi^{-1}(\gamma, \mathbf{x})n^{-\delta} + o_P(n^{-\delta}).$$

Thus, we have

$$n^{\delta}Q_Y\left(n^{-\delta}\pi^{-1}(\gamma, \mathbf{x}) \mid \mathbf{x}, Y>0\right) \xrightarrow{p} Q_Y^{'}(0 \mid \mathbf{x}, Y>0)\pi^{-1}(\gamma, \mathbf{x}).$$

By the equations above, we can prove eq (S1.14). Finally, by Slutsky's Theorem, we have

$$\sqrt{n}\left\{\widehat{Q}_Y(\tau \mid Y>0, \mathbf{x}) - 0\right\} \xrightarrow{d} \{1-\pi(\gamma, \mathbf{x})\}\sqrt{\mathbf{x}^{\top}D_{1,\gamma}^{-1}\mathbf{x}}Q_Y^{'}(0 \mid \mathbf{x}, Y>0)Z_0 I(Z_0>0).$$

One can also set a larger δ such that $0.250 < \delta < 0.5$. For example, one can select $\delta = 0.499$ and $u > 0$, s.t. $n^u = J_n^{-\frac{1}{2}} n^{\frac{1}{2}} + J_n^r$. Consequently, we have $u \leq \delta < 0.5$. Now we show that

$$n^u \left\{ \widehat{Q}_Y(\tau \mid Y > 0, \mathbf{x}) - 0 \right\} = O_P(1). \tag{S1.16}$$

We have

$$\begin{aligned}
& n^u \left\{ \widehat{Q}_Y(\tau \mid Y > 0, \mathbf{x}) - 0 \right\} \\
= \;& n^u \left\{ \widehat{Q}_Y(\tau \mid Y > 0, \mathbf{x}) I(R_{2,n}) \right\} + n^u \{ \widehat{Q}_Y(\tau \mid Y > 0, \mathbf{x}) I(R_{3,n}) \} \\
= \;& n^u \{ \widehat{Q}_Y(\tau \mid Y > 0, \mathbf{x}) I(R_{2,n}) \} + o_P(1) \\
= \;& n^\delta B \left(\mathbf{x}^\top \hat{\beta}_{n^{-\delta}\pi(\hat{\gamma}_n, \mathbf{x})^{-1}} \right)^\top \tilde{\theta}_n \left(\hat{\beta}_{n^{-\delta}\pi(\hat{\gamma}_n, \mathbf{x})^{-1}}, n^{-\delta} \pi(\hat{\gamma}_n, \mathbf{x})^{-1} \right) \\
& \times n^u \{ \pi(\hat{\gamma}_n, \mathbf{x}) - \pi(\gamma, \mathbf{x}) \} I \left(0 < \pi(\hat{\gamma}_n, \mathbf{x}) - \pi(\gamma, \mathbf{x}) < n^{-\delta} \right) + o_P(1) \\
= \;& n^{\delta - 1/2 + u} B \left(\mathbf{x}^\top \hat{\beta}_{n^{-\delta}\pi(\hat{\gamma}_n, \mathbf{x})^{-1}} \right)^\top \tilde{\theta}_n \left(\hat{\beta}_{n^{-\delta}\pi(\hat{\gamma}_n, \mathbf{x})^{-1}}, n^{-\delta} \pi(\hat{\gamma}_n, \mathbf{x})^{-1} \right) \\
& \times \sqrt{n} \{ \pi(\hat{\gamma}_n, \mathbf{x}) - \pi(\gamma, \mathbf{x}) \} I \left(0 < \pi(\hat{\gamma}_n, \mathbf{x}) - \pi(\gamma, \mathbf{x}) < n^{-\delta} \right) + o_P(1) \\
= \;& n^{u - 1/2} \times \left(n^\delta \widehat{Q}_Y \left(n^{-\delta} \pi(\hat{\gamma}_n, \mathbf{x}) \mid \mathbf{x}, Y > 0 \right) \right) \qquad \text{(S1.17)} \\
& \times \sqrt{n} \{ \pi(\hat{\gamma}_n, \mathbf{x}) - \pi(\gamma, \mathbf{x}) \} I \left(0 < \pi(\hat{\gamma}_n, \mathbf{x}) - \pi(\gamma, \mathbf{x}) < n^{-\delta} \right) + o_P(1).
\end{aligned}$$

By the proof of Theorem 3 (iii) and the fact that $\delta < 1/2$, we can prove that

$$n^{\delta - 1/2 + u} \left\{ \widehat{Q}_Y \left(n^{-\delta} \pi^{-1}(\hat{\gamma}_n, \mathbf{x}) \mid \mathbf{x}, Y > 0 \right) - Q_Y(n^{-\delta} \pi^{-1}(\gamma, \mathbf{x}) \mid \mathbf{x}, Y > 0) \right\} = o_P(1).$$

By Taylor expansion, we have

$$n^{\delta-1/2+u}Q_Y\left(n^{-\delta}\pi^{-1}(\gamma,\mathbf{x}) \mid \mathbf{x}, Y>0\right) \\ \xrightarrow{p} n^{u-1/2}Q_Y'(0 \mid \mathbf{x}, Y>0)\pi(\gamma,\mathbf{x})^{-1}. \quad \text{(S1.18)}$$

Due to the fact that $n^u = J_n^{-\frac{1}{2}}n^{\frac{1}{2}} + J_n^r < n^{1/2}$, we know that

$$n^{u-1/2}\widehat{Q}_Y\left(n^{-\delta}\pi^{-1}(\hat{\gamma}_n,\mathbf{x}) \mid \mathbf{x}, Y>0\right) = o_P(1).$$

Similarly, we have

$$\sqrt{n}\left\{\pi(\hat{\gamma}_n,\mathbf{x}) - \pi(\gamma,\mathbf{x})\right\} I\left(0 < \pi(\hat{\gamma}_n,\mathbf{x}) - \pi(\gamma,\mathbf{x}) < n^{-\delta}\right) \quad \text{(S1.19)} \\ \xrightarrow{d} \pi(\gamma,\mathbf{x})\left\{1-\pi(\gamma,\mathbf{x})\right\}\sqrt{\mathbf{x}^\top D_{1,\gamma}\mathbf{x}}Z_0 I(Z_0>0),$$

where $Z_0 \sim N(0,1)$. Using Slutsky's Theorem and the fact $\delta < 0.5$, we have

$$n^u\left\{\widehat{Q}_Y(\tau \mid Y>0,\mathbf{x}) - 0\right\} = O_P(1).$$

Hence, we conclude the proof of Theorem 3.

S1.2 Asymptotic Property of Average Quantile Effect

We provide the asymptotic convergence rate of the estimated AQE constructed in eq (2.12) based on Corollary 1.

Theorem S1. *Given the assumption that the coefficient functions* $\beta \circ \Gamma(\tau; x_j, \boldsymbol{x}^{(-j)}, \gamma)$ *are smooth functions of* $\boldsymbol{x}^{(-j)}$ *with compact supports and* *the conditions in Theorem 1-3, as* $n \to \infty$,

$$\widehat{\Delta}_\tau(x_j; u, v) - \Delta_\tau(x_j; u, v) = O_P\left(J_n^{\frac{1}{2}} n^{-\frac{1}{2}} + J_n^{-r}\right).$$

Of note, hypothesis testing on AQE is possible, but beyond the main scope of this paper. Thus, we do not provide a detailed discussion of it. In Ling et al. (2022), though the asymptotic distribution of estimated AQE is provided for the linear quantile regression model, paired bootstrap is recommended for hypothesis testing of the covariate effect as the asymptotic variance of the estimated AQE is too complicated to calculate. We conjecture that the bootstrap scheme would also work for our ZIQSI model. We provide the proof of Theorem S1 as below.

Proof of Theorem S1 As stated in Ling et al. (2022), we deduce the treatment effect of a binary covariate x_j. By Theorem 3, we have:

$$\begin{aligned}
&\widehat{Q}_Y(\tau \mid x_j = 1, \mathbf{x}^{(-j)}) - \widehat{Q}_Y(\tau \mid x_j = 0, \mathbf{x}^{(-j)}) \\
&\quad - \left[Q_Y(\tau \mid x_j = 1, \mathbf{x}^{(-j)}) - Q_Y(\tau \mid x_j = 0, \mathbf{x}^{(-j)})\right] \\
= &\left\{B^\top\left((1, \mathbf{x}^{(-j)})\hat{\beta} \circ \Gamma(\tau; 1, \mathbf{x}^{(-j)}, \hat{\gamma}_n)\right)\tilde{\theta}_n(\hat{\beta} \circ \Gamma, \Gamma(\tau; 1, \mathbf{x}^{(-j)}, \hat{\gamma}_n))\right. \\
&\quad \left. - Q_Y(\tau \mid x_j = 1, \mathbf{x}^{(-j)})\right\} I\left\{\tau > 1 - \pi(\gamma, 1, \mathbf{x}^{(-j)})\right\} \\
&- \left\{B^\top\left((1, \mathbf{x}^{(-j)})\hat{\beta} \circ \Gamma(\tau; 0, \mathbf{x}^{(-j)}, \hat{\gamma}_n)\right)\tilde{\theta}_n(\hat{\beta} \circ \Gamma, \Gamma(\tau; 0, \mathbf{x}^{(-j)}, \hat{\gamma}_n))\right. \\
&\quad \left. - Q_Y\left(\tau \mid x_j = 0, \mathbf{x}^{(-j)}\right)\right\} I\left\{\tau > 1 - \pi(\gamma, 0, \mathbf{x}^{(-j)})\right\} \\
&\quad + \widehat{Q}_Y\left(n^{-\delta}\pi(\hat{\gamma}_n, 1, \mathbf{x}^{(-j)}) \mid 1, \mathbf{x}^{(-j)}, Y > 0\right)
\end{aligned}$$

$$\times \left[\pi(\hat{\gamma}_n, 1, \mathbf{x}^{(-j)}) - \pi(\gamma, 1, \mathbf{x}^{(-j)})\right]_+ I\left\{\tau = 1 - \pi(\gamma, 1, \mathbf{x}^{(-j)})\right\}$$

$$-\widehat{Q}_Y\left(n^{-\delta}\pi(\hat{\gamma}_n, 0, \mathbf{x}^{(-j)}) \mid 0, \mathbf{x}^{(-j)}, Y > 0\right)$$

$$\times \left[\pi(\hat{\gamma}_n, 0, \mathbf{x}^{(-j)}) - \pi(\gamma, 0, \mathbf{x}^{(-j)})\right]_+ I\left\{\tau = 1 - \pi(\gamma, 0, \mathbf{x}^{(-j)})\right\}$$

$$+O_P\left(J_n^{\frac{1}{2}} n^{-\frac{1}{2}} + J_n^{-r}\right).$$

Denote $\eta = (\beta^\top, \gamma^\top)^\top$ and $\hat{\eta}_n = (\hat{\beta}^\top, \hat{\gamma}_n^\top)^\top$ given τ, we have

$$\begin{aligned} h_n(\eta, \mathbf{x}^{(-j)}) &= B^\top\left((1, \mathbf{x}^{(-j)})\beta \circ \Gamma(\tau; 1, \mathbf{x}^{(-j)}, \gamma)\right)\tilde{\theta}_n\left(\beta, \Gamma(\tau; 1, \mathbf{x}^{(-j)}, \gamma)\right) I\left\{\tau > 1 - \pi(\gamma, 1, \mathbf{x}^{(-j)})\right\} \\ &\quad -B^\top\left((0, \mathbf{x}^{(-j)})\beta \circ \Gamma(\tau; 0, \mathbf{x}^{(-j)}, \gamma)\right)\tilde{\theta}_n(\beta, \Gamma(\tau; 0, \mathbf{x}^{(-j)}, \gamma)) I\{\tau > 1 - \pi(\gamma, 0, \mathbf{x}^{(-j)})\}, \end{aligned}$$

and

$$\begin{aligned} h(\eta, \mathbf{x}^{(-j)}) &= Q_Y(\tau \mid 1, \mathbf{x}^{(-j)}) I\left\{\tau > 1 - \pi(\gamma, 1, \mathbf{x}^{(-j)})\right\} \\ &\quad -Q_Y(\tau \mid 0, \mathbf{x}^{(-j)}) I\left\{\tau > 1 - \pi(\gamma, 0, \mathbf{x}^{(-j)})\right\} \\ &= G_{\Gamma(\tau; 1, \mathbf{x}^{(-j)}, \gamma)}\left\{(1, \mathbf{x}^{(-j)})\beta \circ \Gamma(\tau; 1, \mathbf{x}^{(-j)}, \gamma)\right\} I\left\{\tau > 1 - \pi(\gamma, 1, \mathbf{x}^{(-j)})\right\} \\ &\quad -G_{\Gamma(\tau; 0, \mathbf{x}^{(-j)}, \gamma)}\left\{(0, \mathbf{x}^{(-j)})\beta \circ \Gamma(\tau; 0, \mathbf{x}^{(-j)}, \gamma)\right\} I\left\{\tau > 1 - \pi(\gamma, 0, \mathbf{x}^{(-j)})\right\}. \end{aligned}$$

Denote $\mathbf{x}^{0,(-j)}$ as the covariates from the current data excluding x_j. $\mathbf{x}^{0,(-j)}$ has the same distribution $P_{\mathbf{x}^{(-j)}}$ as the new covariates $\mathbf{x}^{(-j)}$ defined in Section 2.3. We can write:

$$
\begin{aligned}
&\widehat{\Delta}_\tau(x_j;1,0) - \Delta_\tau(x_j;1,0)\\
= &\int h_n(\hat{\eta}_n, \mathbf{x}^{0,(-j)}) d\mathbb{P}_{n\mathbf{x}^{0,(-j)}} - \int h(\eta, \mathbf{x}^{(-j)}) dP_{\mathbf{x}^{(-j)}}\\
= &\int \left\{ h_n(\hat{\eta}_n, \mathbf{x}^{0,(-j)}) - h(\eta, \mathbf{x}^{0,(-j)}) d\mathbb{P}_{n\mathbf{x}^{0,(-j)}} \right\}\\
&\quad + \int h(\eta, \mathbf{x}^{0,(-j)}) d(\mathbb{P}_{n\mathbf{x}^{0,(-j)}} - P_{\mathbf{x}^{(-j)}}). \qquad \text{(S1.20)}
\end{aligned}
$$

For the first part of eq (S1.20), we write:

$$
\begin{aligned}
&\int h_n(\hat{\eta}_n, \mathbf{x}^{0,(-j)}) - h(\eta, \mathbf{x}^{0,(-j)}) d\mathbb{P}_{n\mathbf{x}^{0,(-j)}}\\
= &\frac{1}{n}\sum_{i=1}^n \Bigg\{ B^\top \left((1, \mathbf{x}_i^{(-j)})\hat{\beta} \circ \Gamma(\tau; 1, \mathbf{x}_i^{(-j)}, \hat{\gamma}_n)\right) \tilde{\theta}_n \left(\hat{\beta} \circ \Gamma, \Gamma(\tau; 1, \mathbf{x}_i^{(-j)}, \hat{\gamma}_n)\right) I\left\{\tau > 1 - \pi(\hat{\gamma}_n, 1, \mathbf{x}_i^{(-j)})\right\}\\
&\quad - B^\top \left((0, \mathbf{x}^{(-j)})\hat{\beta} \circ \Gamma(\tau; 0, \mathbf{x}^{(-j)}, \hat{\gamma}_n)\right) \tilde{\theta}_n \left(\hat{\beta} \circ \Gamma, \Gamma(\tau; 0, \mathbf{x}^{(-j)}, \hat{\gamma}_n)\right) I\left\{\tau > 1 - \pi(\hat{\gamma}_n, 0, \mathbf{x}^{(-j)})\right\} \Bigg\}\\
&\quad - \frac{1}{n}\sum_{i=1}^n \Bigg\{ G_{\Gamma(\tau;1,\mathbf{x}_i^{(-j)},\gamma)} \left\{(1, \mathbf{x}_i^{(-j)})\beta \circ \Gamma(\tau; 1, \mathbf{x}_i^{(-j)}, \gamma)\right\} I\left\{\tau > 1 - \pi(\gamma, 1, \mathbf{x}_i^{(-j)})\right\}\\
&\quad - G_{\Gamma(\tau;0,\mathbf{x}_i^{(-j)},\gamma)} \left\{(0, \mathbf{x}^{(-j)})\beta \circ \Gamma(\tau; 0, \mathbf{x}_i^{(-j)}, \gamma)\right\} I\left\{\tau > 1 - \pi(\gamma, 0, \mathbf{x}_i^{(-j)})\right\} \Bigg\}\\
= &\frac{1}{n}\sum_{i=1}^n \Bigg\{ B^\top((1, \mathbf{x}_i^{(-j)})\hat{\beta} \circ \Gamma(\tau; 1, \mathbf{x}_i^{(-j)}, \hat{\gamma}_n))\tilde{\theta}_n(\hat{\beta} \circ \Gamma, \Gamma(\tau; 1, \mathbf{x}_i^{(-j)}, \hat{\gamma}_n))\\
&\quad - G_{\Gamma(\tau;1,\mathbf{x}_i^{(-j)},\gamma)} \left\{(1, \mathbf{x}_i^{(-j)})\beta \circ \Gamma(\tau; 1, \mathbf{x}_i^{(-j)}, \gamma)\right\} \Bigg\} I\left\{\tau > 1 - \pi(\gamma, 1, \mathbf{x}_i^{(-j)})\right\}\\
&\quad - \frac{1}{n}\sum_{i=1}^n \Bigg\{ B^\top \left((0, \mathbf{x}^{(-j)})\hat{\beta} \circ \Gamma(\tau; 0, \mathbf{x}^{(-j)}, \hat{\gamma}_n)\right) \tilde{\theta}_n \left(\hat{\beta} \circ \Gamma, \Gamma(\tau; 0, \mathbf{x}^{(-j)}, \hat{\gamma}_n)\right)\\
&\quad - G_{\Gamma(\tau;0,\mathbf{x}_i^{(-j)},\gamma)} \left\{(0, \mathbf{x}_i^{(-j)})\beta \circ \Gamma(\tau; 0, \mathbf{x}_i^{(-j)}, \gamma)\right\} \Bigg\} I\left\{\tau > 1 - \pi(\gamma, 0, \mathbf{x}_i^{(-j)})\right\} + o_P(n^{-1/2})\\
= &O_P\left(J_n^{\frac{1}{2}} n^{-\frac{1}{2}} + J_n^{-r}\right).
\end{aligned}
$$

The last equation above is deduced from Corollary 1. For the second part of eq (S1.20), we can write

$$\begin{aligned} & \int h(\eta, \mathbf{x}^{0,(-j)}) d(\mathbb{P}_{n\mathbf{x}^{0,(-j)}} - P_{\mathbf{x}_i^{(-j)}}) \\ = \ & \frac{1}{n}\Bigg[\sum_{i=1}^{n} G_{\Gamma(\tau;1,\mathbf{x}_i^{(-j)},\gamma)} \left\{ (1, \mathbf{x}_i^{(-j)})\beta \circ \Gamma(\tau; 1, \mathbf{x}_i^{(-j)}, \gamma) \right\} I \left\{ \tau > 1 - \pi(\gamma, 1, \mathbf{x}_i^{(-j)}) \right\} \\ & \quad - G_{\Gamma(\tau;0,\mathbf{x}_i^{(-j)},\gamma)} \left\{ (0, \mathbf{x}_i^{(-j)})\beta \circ \Gamma(\tau; 0, \mathbf{x}_i^{(-j)}, \gamma) \right\} I \left\{ \tau > 1 - \pi(\gamma, 0, \mathbf{x}_i^{(-j)}) \right\} \Bigg] \\ & \quad - \Delta(x_j; 1, 0). \end{aligned} \tag{S1.21}$$

By the same strategy in Proof of Theorem 2 in Ling et al. (2022), we can derive that the convergence rate of eq (S1.21) is $\sqrt{n}$. Thus, we conclude the proof of Theorem S1.

S2 Additional Simulation Results

S2.1 Simulation Settings

Here, we represent the health-related covariates of 12 individuals, whose quantile functions are estimated in Section 3.2 in our main text.

In Figure S2.1, we represent the taxon we simulated in Section 3. We observe that the simulated data mimics the real data well.

Table S2.1: The summary of covariates information for 12 individuals.

ID	medicament	BMI	waist circumstance	diastolic blood pressure	systolic blood pressure
1	0	26.65	83.73	71.91	111.52
2	1	26.65	83.73	71.91	111.52
3	0	28	92.5	80	124
4	1	28	92.5	80	124
5	0	29.68	103.44	90.10	139.57
6	1	29.68	103.44	90.10	139.57
7	0	26.32	81.56	90.10	139.57
8	1	26.32	81.56	90.10	139.57
9	0	28	92.5	69.90	108.43
10	1	28	92.5	69.90	108.43
11	0	28.51	95.79	73.71	114.30
12	1	28.51	95.79	73.71	114.30

S2.2 Average proportion of negative predicted counts over the quantile process

We further report the average proportion of negative predicted counts over the quantile process $\tau \in (0, 1)$ for three methods. As none of them enforce the constraint of positive prediction, it is possible to have negative predicted counts. However, negative predictions are unreasonable in micro-

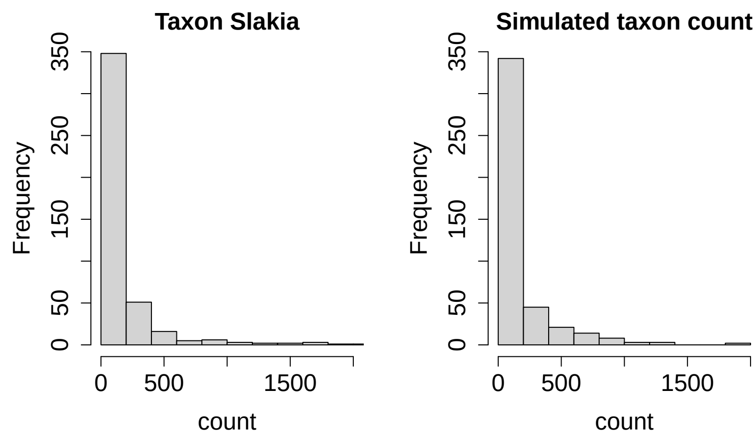


Figure S2.1: The histogram of real taxon count and the taxon count of our simulation setting.

biome studies. As shown in Table S2.2, Quantile Single-index has nearly 20% negative predicted counts among all τ's for every subject, whereas the two-part modeling approaches (i.e., ZIQSI and ZIQ-linear) have a much lower percentage of negative predictions.

S2.3 Additional simulation results for RIBIAS, RIVAR, and RIMSE truncated at zero

As we did not impose a non-negativity constraint when optimizing the quantile function (2.5), our method may produce negative values in the estimated quantile curve. Table S2.2 shows that this issue also arises with other methods. In practice, one can impose truncation at zero if strictly non-negativity is desired. Here, we present the RIBIAS, RIVAR, and RIMSE

Table S2.2: Summary of the average proportion of negative predicted counts over the quantile process $\tau \in (0, 1)$.

ID	ZIQSI	ZIQ-linear	Quantile Single-index
1	0.04	0.05	0.23
2	0.04	0.04	0.23
3	0.03	0.00	0.17
4	0.06	0.00	0.17
5	0.07	0.00	0.21
6	0.04	0.00	0.21
7	0.03	0.05	0.19
8	0.04	0.05	0.19
9	0.04	0.04	0.22
10	0.04	0.03	0.22
11	0.04	0.00	0.19
12	0.03	0.00	0.18

for each method after truncating at zero in Table S2.3, demonstrating that our method retains its advantage even after truncation.

S2.4 Estimated quantile curves for other individuals

Here we present the average estimated quantile curves and their 95% confidence intervals for other individuals in Figure S2.2 − S2.5, except subject

Table S2.3: Summary of RIMSE(%), RIBIAS(%), RIVAR(%) of the estimated conditional quantile functions truncated at zero by ZIQSI, ZIQ-linear(ZIQ), and Quantile Single-index(QSI).

	RIBIAS			RIVAR			RIMSE		
ID	ZIQSI	ZIQ	QSI	ZIQSI	ZIQ	QSI	ZIQSI	ZIQ	QSI
1	0.19	21.18	1.14	3.20	5.25	2.81	3.39	26.43	3.95
2	0.07	21.29	0.44	3.96	6.19	3.94	4.03	27.48	4.38
3	0.24	4.07	1.10	1.54	1.63	1.49	1.78	5.70	2.59
4	0.04	4.17	0.13	1.68	1.66	1.82	1.72	5.83	1.95
5	0.10	2.53	0.76	3.31	1.02	3.62	3.41	3.55	4.38
6	0.04	2.34	0.13	3.80	1.20	3.80	3.84	3.54	3.93
7	0.34	1.23	0.84	3.09	2.00	2.95	3.43	3.23	3.79
8	0.12	1.27	0.65	3.54	2.27	3.55	3.66	3.54	4.20
9	0.13	19.12	1.01	1.57	4.30	2.36	2.70	23.42	3.37
10	0.02	18.88	0.15	3.01	4.83	3.14	3.03	23.71	3.29
11	0.02	9.04	0.99	1.98	2.22	1.53	2.00	11.26	2.96
12	$6.26e^{-5}$	9.93	0.12	2.25	2.55	2.42	2.25	12.48	2.54

11, which has already been reported in the main text. For each individual, we based the results on the 500 estimations of its quantile curve. The confidence interval is constructed based on the percentile of the empirical distribution of $\widehat{Q}_Y(\tau \mid \mathbf{x})$ at a given τ. We observe that both ZIQ-linear and Quantile Single-index have more prominent bias compared to ZIQSI.

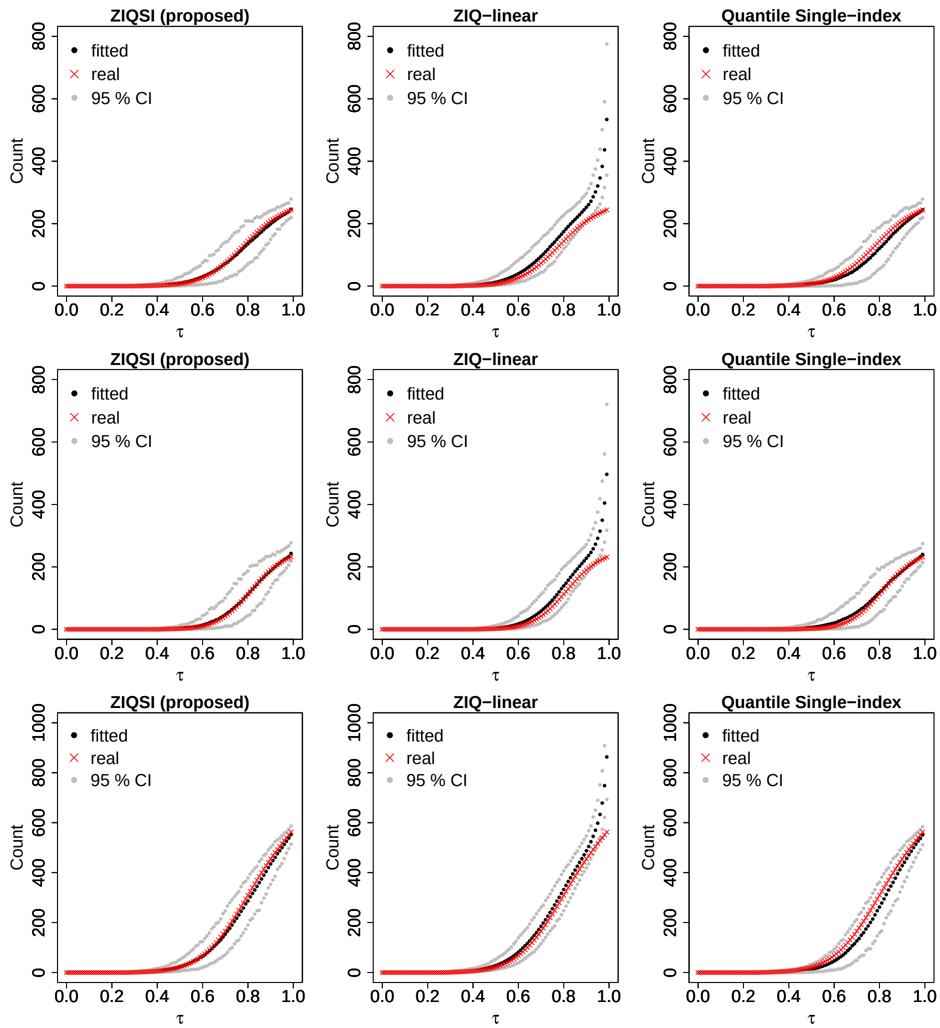


Figure S2.2: Estimated quantile curves based on 500 predictions for subjects $1-3$ (from top to bottom).

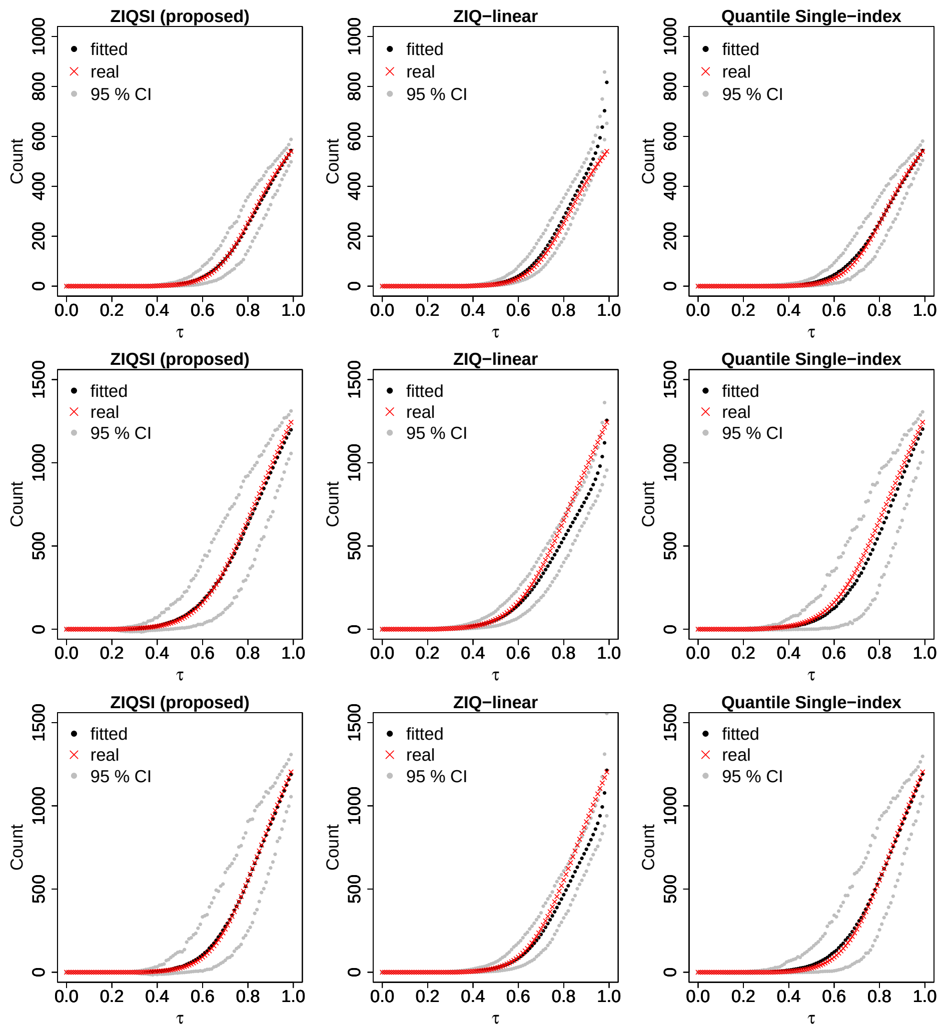


Figure S2.3: Estimated quantile curves based on 500 predictions for subjects $4-6$ (from top to bottom).

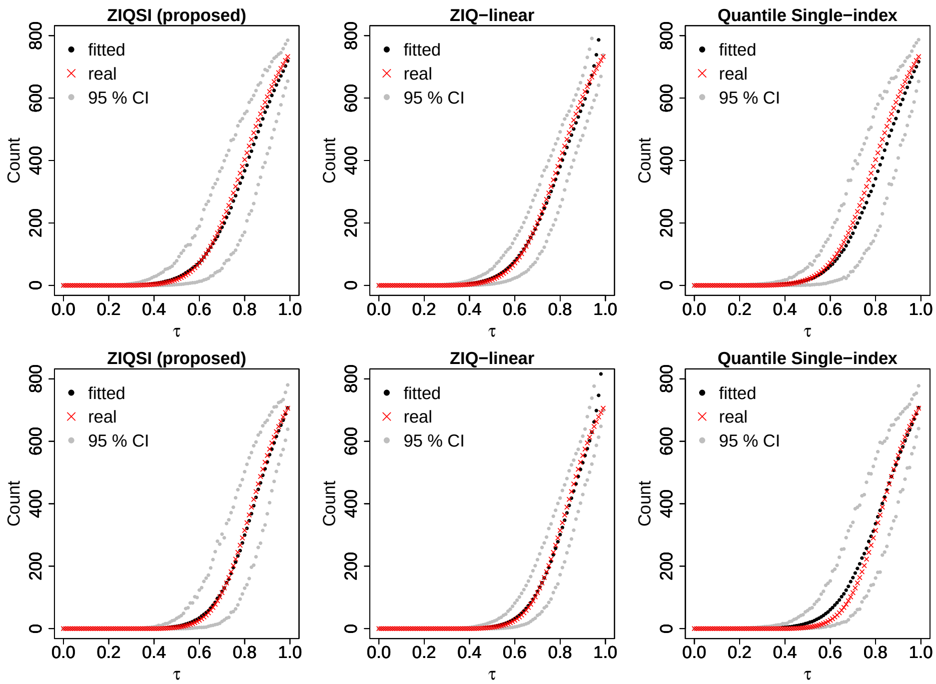


Figure S2.4: Estimated quantile curves based on 500 predictions for subjects $7-8$ (from top to bottom).

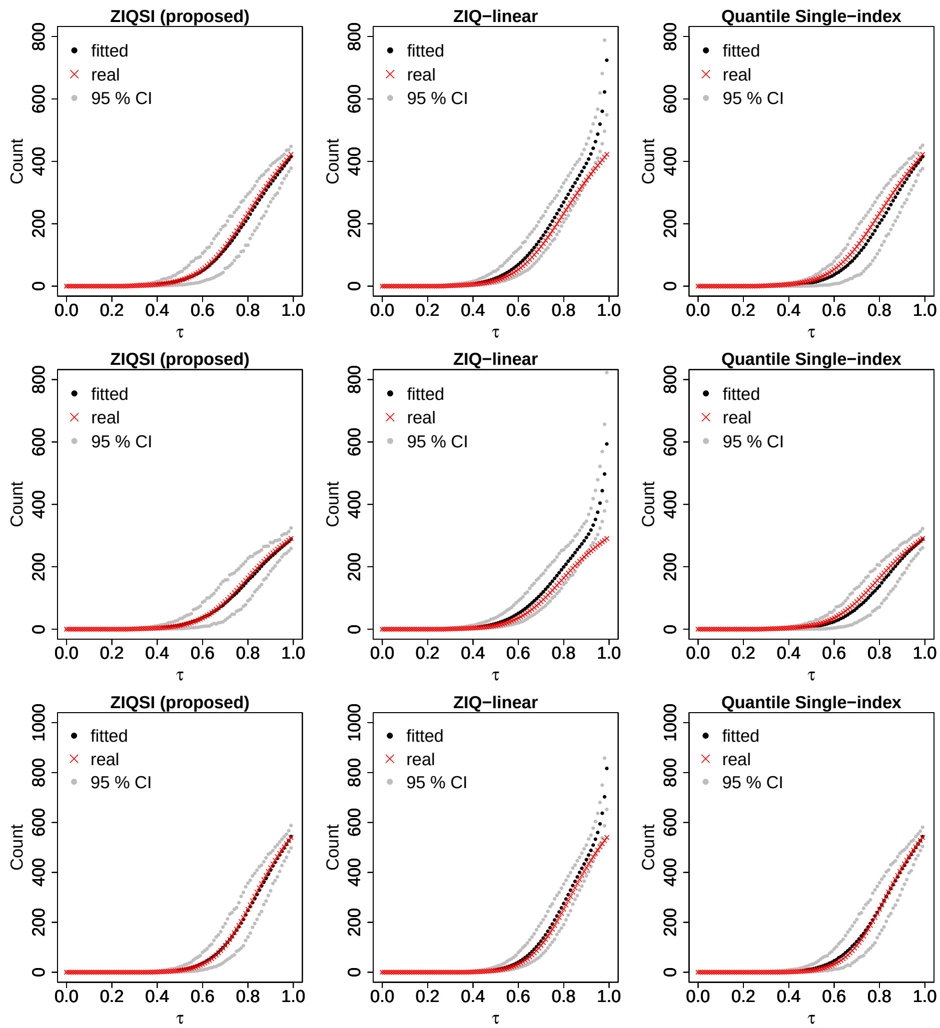


Figure S2.5: Estimated quantile curves based on 500 predictions for subjects 9, 10, and 12 (from top to bottom).

S2.5 Model fitting results for our proposed ZIQSI method with $\delta = 0.250$

Here we present additional simulation results for $\delta = 0.25$. The results suggest that using $\delta = 0.25$ will lead to a minor increase of bias and decrease of variance (Table S2.4). However, compared to Table 1, the change of results due to δ is negligible.

S2.6 Simulation results for the average quantile effect (AQE)

In this section, we present simulation results comparing the point estimates of the average quantile effect (AQE) of BMI across various methods. The AQE is computed using equation (2.10), with x_j as BMI and $u = 23\text{kg/m}^2$, $v = 28\text{kg/m}^2$. We generate samples $(\mathbf{x}_i, y_i)$ for $i = 1, \cdots, 500$ as described in Section 3.1, and measure AQE using the ZIQSI method, ZIQ-linear (ZIQ) method, and Quantile Single-index (QSI) method at representative quantile levels: $\tau = 0.25, 0.50, 0.75$. This procedure is repeated 500 times for each quantile level. For each method, we calculate the trimmed relative bias and trimmed relative standard deviation (sd) based on the 500 estimations. To mitigate the impact of extreme estimations caused by large variance at higher quantile levels (Figure S2.2 to Figure S2.5), particularly at $\tau = 0.50$ and $\tau = 0.75$, we remove the largest and smallest 10% of estimated values

Table S2.4: Summary of RIMSE(%), RIBIAS(%), RIVAR(%) of the estimated conditional quantile function $Q_Y(\tau \mid \mathbf{x})$ using ZIQSI with $\delta = 0.250$.

	ZIQSI (proposed)		
ID	RIMSE	RIBIAS	RIVAR
1	3.22	0.29	2.93
2	3.76	0.15	3.61
3	2.04	0.26	1.78
4	2.03	0.06	1.93
5	3.40	0.12	3.28
6	3.92	0.07	3.85
7	3.31	0.29	3.02
8	3.67	0.17	3.50
9	2.73	0.18	2.55
10	3.19	0.05	3.14
11	2.07	0.12	1.95
12	2.27	0.03	2.24

before analyzing relative bias and sd. Table S2.5 shows that our method yields the lowest bias at each quantile level compared to others. While the Quantile Single-index method demonstrates a relatively smaller standard deviation, it introduces significant bias, especially at $\tau = 0.25$ and $\tau = 0.50$, due to its failure to adjust the quantile level τ, which can distort the estimated impact of BMI on the response Y, as shown in Figure 6.

Table S2.5: The trimmed Relative Bias and trimmed Relative Standard deviation for estimated AQE of BMI. "ZIQ" represents ZIQ-linear, and "QSI" represents Quantile Single-index method.

τ	true AQE	measure	ZIQSI	ZIQ	QSI
		Estimate	2.59	2.64	-0.46
0.25	2.39	Relative Bias	0.08	0.10	-1.19
		Relative Sd	2.06	1.97	1.03
		Estimate	15.30	16.24	-5.95
0.5	15.56	Relative Bias	0.02	0.06	-1.38
		Relative Sd	2.25	2.85	1.55
		Estimate	-27.83	-42.24	-38.27
0.75	-24.56	Relative Bias	0.13	0.72	0.56
		Relative Sd	2.65	4.48	2.62

S2.7 Additional Simulation results with linear G_τ

In this section, we assess the performance of all methods using a linear function $G_\tau(x) = \tau x$, a simpler case compared to the non-linear functions considered in Section 3. We continue to use the data generation scheme for covariates from Section 3.1, and we conduct prediction regarding individuals presented in Table S2.1. From Table S2.6, we observe that our method achieves a RIBIAS comparable to the ZIQ-linear method but shows a slightly higher RIVAR. This is expected, as the ZIQ-linear method takes advantage of the model's linearity, which matches the true model, leading to more efficient estimation.

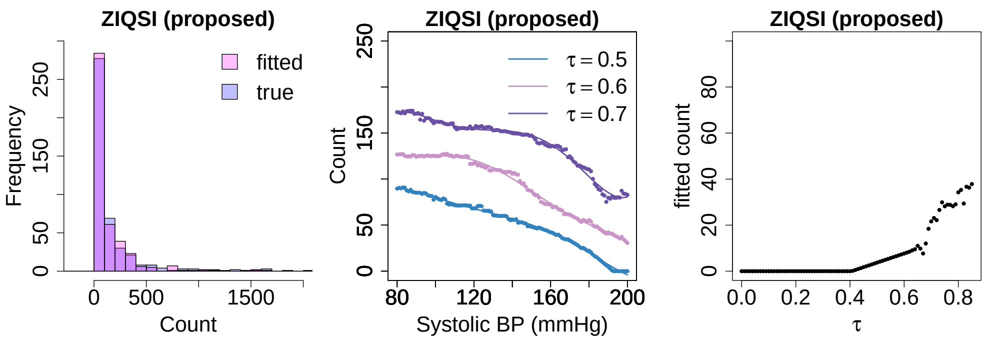
Table S2.6: Summary of RIMSE(%), RIBIAS(%), RIVAR(%) of the estimated conditional quantile functions under the linear setting by ZIQSI, ZIQ-linear(ZIQ), and Quantile Single-index(QSI).

	RIBIAS			RIVAR			RIMSE		
ID	ZIQSI	ZIQ	QSI	ZIQSI	ZIQ	QSI	ZIQSI	ZIQ	QSI
1	0.11	0.04	1.30	2.24	1.41	2.39	2.35	1.45	3.69
2	0.02	0.06	0.24	2.51	1.74	3.06	2.53	1.80	3.30
3	0.07	0.02	0.73	1.35	0.65	1.34	1.42	0.67	2.07
4	0.01	0.02	0.23	1.50	0.68	1.65	1.51	0.70	1.88
5	0.05	0.01	0.61	2.35	0.86	2.93	2.40	0.87	3.54
6	0.05	0.02	0.19	2.54	1.08	3.28	2.59	1.10	3.47
7	0.11	0.02	0.50	2.15	1.21	1.96	2.26	1.23	2.46
8	0.05	0.02	0.73	2.55	1.56	2.36	2.60	1.58	3.09
9	0.11	0.02	0.91	1.88	1.07	1.84	1.99	1.09	2.75
10	0.01	0.03	0.19	2.12	1.35	2.33	2.13	1.38	2.52
11	0.06	0.02	1.78	1.29	0.78	4.84	1.35	0.80	5.62
12	0.01	0.03	0.20	1.51	1.01	4.78	1.52	1.04	5.02

S3 Additional Results for Applications

S3.1 Results for ZIQSI estimation with $\delta = 0.25$

In this section, we present the results of taxon *Slackia* with $\delta = 0.25$. A smaller δ represents a larger region for interpolation. Compared to Figure 4, Figure 5 and Figure 7, results suggest that the value of δ do not affect the model fitting with noticeable change (Figure S3.6).



(a) Histogram of the observed and fitted taxon counts.

(b) Change of predicted counts regarding systolic bp(mmHg).

(c) Predicted quantile curve of subject X11993.MI385H.

Figure S3.6: Fitted results for the taxon *Slackia* using ZIQSI with $\delta = 0.250$.

S3.2 Model fitting results for other taxa

In Figure S3.7 – S3.8 we represent the histogram of 5 taxa whose probability of observing zero is around respectively 0.3, 0.4, 0.5, 0.6, and 0.7, fitted by three methods. We observe similar patterns as shown in the main text.

Our proposed ZIQSI fits the data much better than the other two methods. ZIQ-linear commonly has a small proportion of negative predicted counts, but it could be as small as -500 (**Top** figure of Figure S3.8). Quantile Single-index often provides a large proportion of negative predicted counts, consistent with the simulation results.

S3.3 Average Quantile effects

In this section, we present the average quantile effects of the taxon *Slackia*. Though the estimated quantile curve is individual-specific, one can assess the quantile effect of a covariate through the AQE by integrating all possible values of the covariate of interest (Section 2.3). We again use the taxon *Slackia* as an example. As reported in De la Cuesta-Zuluaga et al. (2018), "Adiponectin" has a statistically significant negative effect on the abundance of *Prevotella*, while "Insulin" has not been detected as significantly associated with *Prevotella*. As *Prevotella* is the co-abundance group that the taxon *Slackia* belongs to, we expect to observe consistent AQE as reported in De la Cuesta-Zuluaga et al. (2018). For "Adiponectin", we compare the two levels, namely $15\mu g/ml$ and $2\mu g/ml$, as the normal and the low adiponectin levels (Cleveland Clinic, 2023). For "Insulin", we compare the two levels, namely $10\mu g/ml$ and $25\mu g/ml$, as the normal and the high

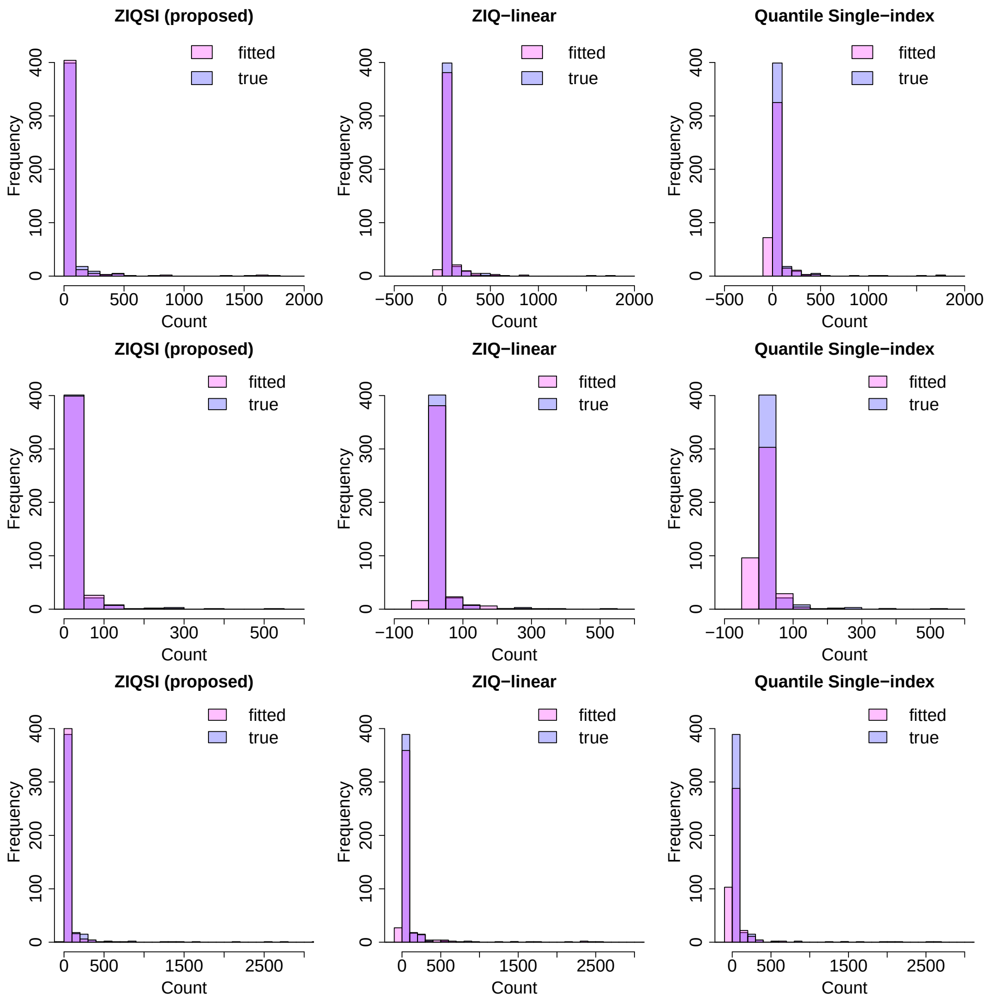


Figure S3.7: **Top**: *Coprococcus* (observed zero proportion: 30%); **Middle**: *Prevotella* (observed zero proportion: 40%); **Bottom**: *Clostridiales* (observed zero proportion: 50%)

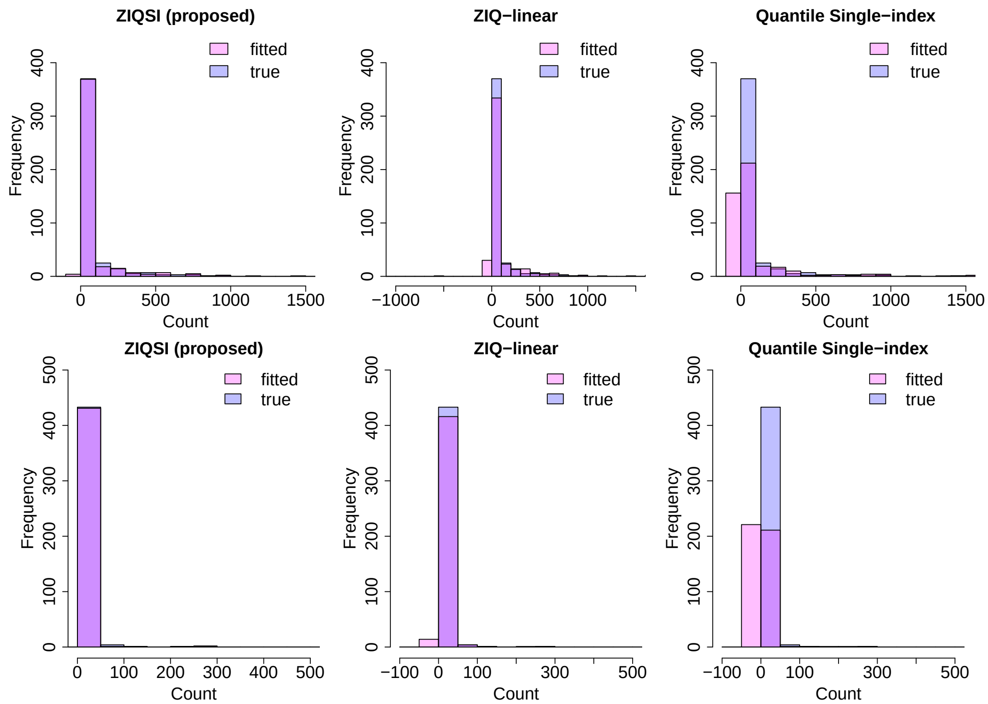


Figure S3.8: **Top**: *RF*39 (observed zero proportion: 60%); **Low**: *Ruminococcaceae* (observed zero proportion: 70%)

insulin levels (Uttekar, 2021). These two variables are closely associated with the host's health and represent how the estimated AQE given three methods perform under two different circumstances where the variable has or does not have a significant effect on the abundance of the taxon.

The results of ZIQSI and ZIQ-linear show that the average quantile effect of adiponectin on the count of *Slackia* is much more prominent than that of insulin, whereas Quantile Single-index does not show any evidence of the quantile effect of adiponectin (Figure S3.9). Further, ZIQSI and ZIQ-linear suggest that adiponectin has a negative effect on the taxa count, which is consistent with the results in De la Cuesta-Zuluaga et al. (2018).

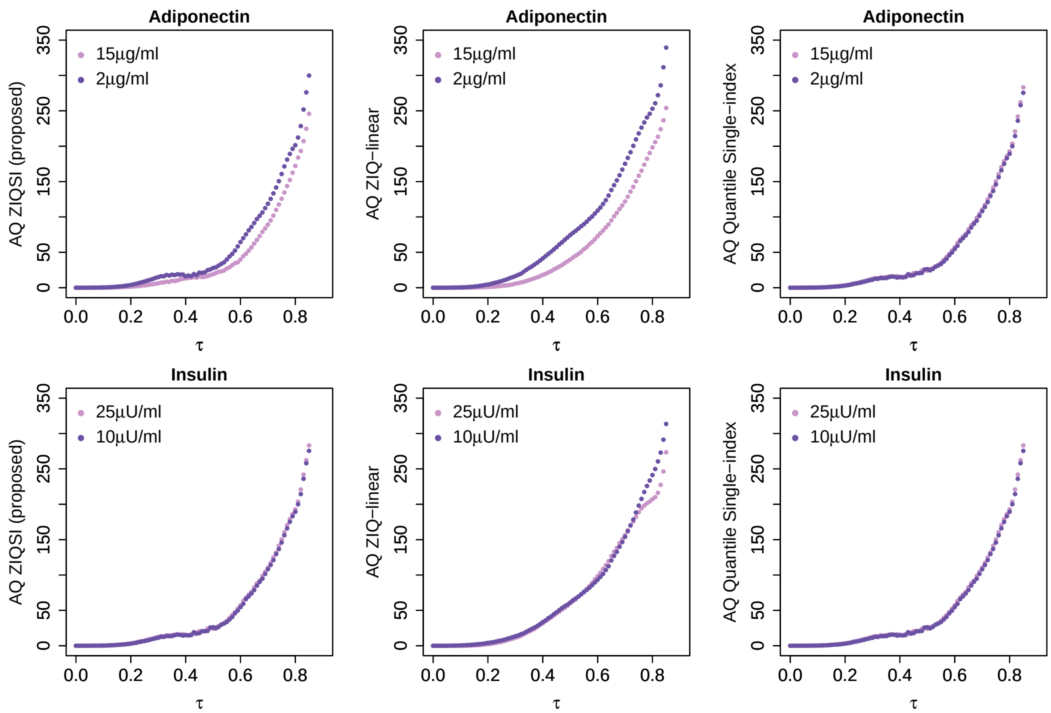


Figure S3.9: Estimated AQEs of selected covariates by three methods.

References

Cleveland Clinic (2023). Adiponectin. `https://my.clevelandclinic.org/health/articles/22439-adiponectin`. [Online; accessed 17-January-2023].

De la Cuesta-Zuluaga, J., V. Corrales-Agudelo, E. P. Velásquez-Mejía, J. A. Carmona, J. M. Abad, and J. S. Escobar (2018). Gut microbiota is associated with obesity and cardiometabolic disease in a population in the midst of westernization. *Scientific reports 8*(1), 1–14.

Ling, W., B. Cheng, Y. Wei, J. Z. Willey, and Y. K. Cheung (2022). Statistical inference in quantile regression for zero-inflated outcomes. *Statistica Sinica 32*(3), 1411.

Ma, S. and X. He (2016a). Inference for single-index quantile regression models with profile optimization. *The Annals of Statistics 44*(3).

Ma, S. and X. He (2016b). Supplement to "inference for single-index quantile regression models with profile optimization". *the Annals of Statistic 44*.

Schumaker, L. (2007). *Spline functions: basic theory.* Cambridge university press.

Uttekar, P. S. (2021). What Is a High Insulin Level? `https://www.medicinenet.com/what_is_a_high_insulin_level/article.htm`. [Online; accessed 25-January-2023].