Supplemental appendix: Nonparametric inference on (conditional) quantile differences and interquantile ranges, using L-statistics

Matt Goldman[†] and David M. Kaplan[‡]

[†] *Microsoft Research, Redmond, WA 98052, USA.*
E-mail: `mattgold@microsoft.com`
[‡] *Department of Economics, University of Missouri, 118 Professional Bldg, 909 University Ave, Columbia, MO 65211, USA.*
E-mail: `kaplandm@missouri.edu`

Received: December 2016

Summary Appendix C contains fully detailed proofs; Appendix D, implementation details; Appendix E, detailed steps for CI construction; Appendix F, further approximation and intuition for the quantile difference CI calibration; Appendix G, simulations; and Appendix H, additional empirical analysis.

C. ADDITIONAL PROOFS

We maintain the same notation and preliminary results from the main appendix. For easier reference, we reproduce the following definitions and lemma (without the accompanying text):

$$\hat{Q}_X^L(k/(n+1)) \equiv \hat{X}_{n:k}^L,\ \tilde{Q}_X^I(k/(n+1)) \equiv \tilde{X}_{n:k}^I,\ \tilde{Q}_U^I(k/(n+1)) \equiv \tilde{U}_{n:k}^I,\ \tilde{Q}_X^I(\cdot) \equiv Q_X(\tilde{Q}_U^I(\cdot)),$$

$$u_j^h(\alpha) \equiv k_j^h(\alpha)/(n+1), \quad u_j^l(\alpha) \equiv k_j^l(\alpha)/(n+1), \tag{C.1}$$

$$\begin{aligned}
k_j &\equiv \lfloor (n+1)u_j \rfloor, & \epsilon_j &\equiv (n+1)u_j - k_j,\\
Y_j^{\mathbf{u}} &\equiv U_{n:k_j} \sim \mathrm{Beta}(k_j, n+1-k_j), & \mathbf{Y}^{\mathbf{u}} &\equiv (Y_1^{\mathbf{u}}, \ldots, Y_J^{\mathbf{u}})',\\
\Delta\mathbf{Y}^{\mathbf{u}} &\equiv (Y_1, Y_2 - Y_1, \ldots, 1 - Y_J)' \sim \mathrm{Dirichlet}(\Delta\mathbf{k}), & &\\
\Lambda_j^{\mathbf{u}} &\equiv U_{n:k_j+1} - U_{n:k_j} \sim \mathrm{Beta}(1, n), & \boldsymbol{\Lambda}^{\mathbf{u}} &\equiv (\Lambda_1^{\mathbf{u}}, \ldots, \Lambda_J^{\mathbf{u}})',\\
\mathbb{W}^{\mathbf{u}} &\equiv \sqrt{n}\left(\sum_{j=1}^J \psi_j Q(Y_j^{\mathbf{u}}) - \sum_{j=1}^J \psi_j Q(u_j)\right), & &
\end{aligned} \tag{C.2}$$

$$\begin{aligned}
\mathbb{W}_{\boldsymbol{\epsilon},\boldsymbol{\Lambda}}^{\mathbf{u}} &\equiv \mathbb{W}^{\mathbf{u}} + n^{1/2}\sum_{j=1}^J \epsilon_j \psi_j \Lambda_j^{\mathbf{u}}(Q'(u_j) + Q''(u_j)(Y_j^{\mathbf{u}} - u_j)),\\
\hat{\mathbf{u}}^H &\equiv \{u_j^H(\tilde{\alpha}(\hat{\gamma}))\}_{j=1}^J, \qquad \mathbf{u}_0^H \equiv \{u_j^H(\tilde{\alpha}(\boldsymbol{\gamma}_0))\}_{j=1}^J,
\end{aligned}$$

$$\max_j\{n\Delta k_j^{-1/2}|\Delta y_j - \Delta k_j/n|\} \le 2\log(n). \qquad \text{Condition } \star$$

Lemma C.1. *Let $z_{1-\alpha}$ denote the $(1-\alpha)$-quantile of a standard normal distribution. From the definitions in* (3.1) *and* (C.1), *the values $u_j^l(\alpha)$ and $u_j^h(\alpha)$ can be approximated as*

$$\begin{aligned}
u_j^l(\alpha) &= \tau_j - n^{-1/2} z_{1-\alpha}\sqrt{\tau_j(1-\tau_j)} - \frac{2\tau_j - 1}{6n}(z_{1-\alpha}^2 + 2) + O(n^{-3/2}),\\
u_j^h(\alpha) &= \tau_j + n^{-1/2} z_{1-\alpha}\sqrt{\tau_j(1-\tau_j)} - \frac{2\tau_j - 1}{6n}(z_{1-\alpha}^2 + 2) + O(n^{-3/2}).
\end{aligned}$$

With $L_0 \equiv \sum_{j=1}^J \psi_j Q_X(u_j)$, uniformly over $\mathbf{u} = \boldsymbol{\tau} + o(1)$,

$$\sup_{K\in\mathbb{R}}\left|\mathrm{P}\left(\sum_{j=1}^J \psi_j \hat{X}_{n:(n+1)u_j}^L < L_0 + n^{-1/2}K\right) - \mathrm{P}\left(\sum_{j=1}^J \psi_j \tilde{X}_{n:(n+1)u_j}^I < L_0 + n^{-1/2}K\right)\right|$$
$$= O(n^{-1}). \tag{C.3}$$

To expand upon the connection between Condition $\star$ and a law of iterated logarithm (LIL), for example, if $\Delta k_j/n = p_j$ is fixed and Δy_j were a sample average of iid random variables with mean p_j and unit variance, then

$$\limsup_{n\to\infty}\frac{n(\Delta y_j - p_j)}{\sqrt{n\log(\log(n))}} = -\liminf_{n\to\infty}\frac{n(\Delta y_j - p_j)}{\sqrt{n\log(\log(n))}} = \sqrt{2}\text{ a.s.},$$

so $|\Delta y_j - \Delta k_j/n| = O(n^{-1/2}\sqrt{\log(\log(n))})$ almost surely, compared to the weaker bound from Condition $\star$, $O(n^{-1/2}\log(n))$ with probability $O(n^{-2})$. Actually, when $\Delta k_j \asymp n$, then Δy_j is equal in distribution to a uniform order statistic, which can be expressed as a sample average by the Bahadur representation (up to a smaller-order error), so the usual LIL could be used. However, we also need Condition $\star$ when $\Delta k_j = o(n)$ and a corresponding LIL is not available, so we use Condition $\star$ in all cases for simplicity since there is no detriment to the final results.

C.1. Proof of Theorem 3.2 and Theorem A.3 (from main appendix)

The proof from the main appendix is complete except for proving the lemmas concerning T_h, T_l, E_h, and E_l. As a reminder,

$$\mathrm{P}(\boldsymbol{\psi}'\hat{Q}^L_X(\hat{\mathbf{u}}^H) > \boldsymbol{\psi}'Q(\boldsymbol{\tau})) = 1-\alpha+T_h+E_h+O(n^{-1}), \tag{C.4}$$

$$T_h = \mathrm{P}\left(\sum_{j=1}^J \psi_j(Q(\tilde{Q}^I_U(u^H_{0,j}))-Q(\tau_j))>0\right) - \mathrm{P}\left(\sum_{j=1}^J \psi_j Q'(\tau_j)(\tilde{Q}^I_U(u^H_{0,j})-\tau_j)>0\right) + O(n^{-3/2}(\log(n))^3), \tag{C.5}$$

$$E_h = E[\mathrm{P}(\mathbb{W}^{\hat{\mathbf{u}}^H}_{\mathbf{C},\boldsymbol{\Lambda}} > \sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau})-Q(\hat{\mathbf{u}}^H)) \mid \hat{\boldsymbol{\gamma}}) - \mathrm{P}(\mathbb{W}^{\mathbf{u}^H_0}_{\mathbf{C},\boldsymbol{\Lambda}} > \sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau})-Q(\mathbf{u}^H_0)) \mid \hat{\boldsymbol{\gamma}})]. \tag{C.6}$$

Proof of Lemma A.4 (main appendix): For this proof, we introduce the scaled and centred

$$\Delta^H_j \equiv \sqrt{n}(\tilde{Q}^I_U(u^H_j(\tilde{\alpha}))-u^H_j(\tilde{\alpha})), \quad \Delta^L_j \equiv \sqrt{n}(\tilde{Q}^I_U(u^L_j(\tilde{\alpha}))-u^L_j(\tilde{\alpha})). \tag{C.7}$$

We split the overall probability into two pieces: one where

$$|\Delta^H_j| \le 2\sqrt{\log(n)} \text{ and } |\Delta^L_j| \le 2\sqrt{\log(n)}, \quad j=1,\ldots,J, \tag{C.8}$$

and one for the rest. Note that (C.8) is just a slight variant of Condition $\star$. The probability of (C.8) being violated is bounded with the help of GK Lemma 7(iv): applied here, for any $\tilde{Q} \sim \mathrm{Beta}(u(n+1),(1-u)(n+1))$, $\mathrm{P}(\sqrt{n}|\tilde{Q}-u|>a_n) = O(a_n^{-1}\exp(-a_n^2/2))$. In terms of $\Delta = \sqrt{n}(\tilde{Q}-u)$, letting $a_n = \sqrt{2\log(n)}$,

$$\mathrm{P}(|\Delta| > 2\sqrt{\log(n)}) = O(\exp(-(1/2)(2\sqrt{\log(n)})^2)) = O(\exp(\log(n)(-2))) = O(n^{-2}). \tag{C.9}$$

So, with $a_n^2 = 2\log(n)$, since J is fixed, the probability that any $\Delta_j^2 > a_n^2$ (i.e., the probability of the union of such events) is of the same order of magnitude by Boole's Inequality. This order of magnitude is much smaller than that of the dominant terms in T_h and T_l stated in the theorem. We can now focus on the case where all the Δ_j are less than $\sqrt{2\log(n)}$ in absolute value.

Let $\phi_{\underline{\mathbf{V}}}(\cdot)$ denote the PDF of a multivariate normal distribution with mean zero and covariance matrix $\underline{\mathbf{V}}$. From equation (A.4) in GK Lemma 7(iii), the PDFs of the corresponding vectors $\boldsymbol{\Delta}^H \equiv (\Delta^H_1,\ldots,\Delta^H_J)'$ and $\boldsymbol{\Delta}^L \equiv (\Delta^L_1,\ldots,\Delta^L_J)'$ are asymptotically normal; specifically, uniformly over values $\mathbf{d}$ satisfying $|d_j| \le \log(n)$ for each $j=1,\ldots,J$,

$$f_{\boldsymbol{\Delta}^H}(\mathbf{d}) = \phi_{\underline{\mathbf{V}}^H}(\mathbf{d})(1+O(n^{-1/2}(\log(n))^3)),$$
$$f_{\boldsymbol{\Delta}^L}(\mathbf{d}) = \phi_{\underline{\mathbf{V}}^L}(\mathbf{d})(1+O(n^{-1/2}(\log(n))^3)),$$

where the row i, column k elements of covariance matrices $\underline{\mathcal{V}}^H$ and $\underline{\mathcal{V}}^L$ are, respectively,

$$\mathcal{V}^H_{i,k} = \min\{u^H_i(\tilde{\alpha}), u^H_k(\tilde{\alpha})\} - u^H_i(\tilde{\alpha})u^H_k(\tilde{\alpha}),$$
$$\mathcal{V}^L_{i,k} = \min\{u^L_i(\tilde{\alpha}), u^L_k(\tilde{\alpha})\} - u^L_i(\tilde{\alpha})u^L_k(\tilde{\alpha}).$$

Further, these PDFs are first-order equivalent. Let $\underline{\mathcal{V}}$ have row i, column k elements $\mathcal{V}_{i,k} = \min\{\tau_i, \tau_k\} - \tau_i\tau_k$. Then, $\mathcal{V}^H_{i,k} = \mathcal{V}_{i,k} + O(n^{-1/2})$ and $\mathcal{V}^L_{i,k} = \mathcal{V}_{i,k} + O(n^{-1/2})$ because both $u^H_j(\tilde{\alpha}) = \tau_j + O(n^{-1/2})$ and $u^L_j(\tilde{\alpha}) = \tau_j - O(n^{-1/2})$ by Lemma C.1. Writing out the formula for a mean-zero multivariate normal PDF, with $|\cdot|$ denoting determinant, when $\mathbf{d} = O(\log(n))$,

$$\begin{aligned}
f_{\mathbf{\Delta}^H}(\mathbf{d}) &= (2\pi)^{-J/2} \overbrace{|\underline{\mathcal{V}}^H|^{-1/2}}^{=|\underline{\mathcal{V}}+O(n^{-1/2})|^{-1/2}} \exp(-(1/2)\mathbf{d}' \overbrace{(\underline{\mathcal{V}}^H)^{-1}}^{=\underline{\mathcal{V}}^{-1}+O(n^{-1/2})} \mathbf{d}) \\
&= (2\pi)^{-J/2}|\underline{\mathcal{V}}(\underline{\mathbf{I}}_J + \underline{\mathcal{V}}^{-1}O(n^{-1/2}))|^{-1/2} \\
&\quad \times \exp(-(1/2)\mathbf{d}'\underline{\mathcal{V}}^{-1}\mathbf{d} + O(\log(n))O(n^{-1/2})O(\log(n))) \\
&= (2\pi)^{-J/2}|\underline{\mathcal{V}}|^{-1/2}((1 + O(n^{-1/2}))^J)^{-1/2} \\
&\quad \times \exp(-(1/2)\mathbf{d}'\underline{\mathcal{V}}^{-1}\mathbf{d} + O(\log(n))O(n^{-1/2})O(\log(n))) \\
&= (2\pi)^{-J/2}|\underline{\mathcal{V}}|^{-1/2}(1 + O(n^{-1/2}))\exp(-(1/2)\mathbf{d}'\underline{\mathcal{V}}^{-1}\mathbf{d})\overbrace{\exp(O(n^{-1/2}(\log(n))^2))}^{=1+O(n^{-1/2}(\log(n))^2)} \\
&= (2\pi)^{-J/2}|\underline{\mathcal{V}}|^{-1/2}\exp(-(1/2)\mathbf{d}'\underline{\mathcal{V}}^{-1}\mathbf{d})(1 + O(n^{-1/2}(\log(n))^2)) \\
&= \phi_{\underline{\mathcal{V}}}(\mathbf{d})(1 + O(n^{-1/2}(\log(n))^2)).
\end{aligned}$$

The same applies to the PDF of $\mathbf{\Delta}^L$, so when each element of the argument is $O(\log(n))$,

$$f_{\mathbf{\Delta}^H}(\mathbf{d}) = \phi_{\underline{\mathcal{V}}}(\mathbf{d})(1 + O(n^{-1/2}(\log(n))^2)), \quad f_{\mathbf{\Delta}^L}(\mathbf{d}) = \phi_{\underline{\mathcal{V}}}(\mathbf{d})(1 + O(n^{-1/2}(\log(n))^2)). \tag{C.10}$$

We also define

$$D^H_j \equiv \sqrt{n}(u^H_j(\tilde{\alpha}) - \tau_j), \quad D^L_j \equiv \sqrt{n}(u^L_j(\tilde{\alpha}) - \tau_j), \tag{C.11}$$

$$D^L_j = \overbrace{-z_{1-\tilde{\alpha}}\sqrt{\tau_j(1-\tau_j)} - n^{-1/2}\frac{2\tau_j - 1}{6}(z^2_{1-\tilde{\alpha}} + 2) + O(n^{-1})}^{\text{Lemma C.1}} = -D^H_j + O(n^{-1/2}), \tag{C.12}$$

$$D^H_0 \equiv \sum_{j=1}^{J}\psi_j\gamma_j D^H_j, \quad D^L_0 \equiv \sum_{j=1}^{J}\psi_j\gamma_j D^L_j = -D^H_0 + O(n^{-1/2}). \tag{C.13}$$

Below, $\tilde{u}_j$ is between τ_j and $u^H_{0,j}$, so $\tilde{u}_j \to \tau_j$. Thus, for large enough n, all $\tilde{u}_j$ are in an arbitrarily small neighbourhood of τ_j, and A2.2 implies $Q'''(\tilde{u}_j)$ is uniformly bounded. We have

$$\sqrt{n}(\tilde{Q}^I_U(u^H_{0,j}) - u^H_{0,j}) = \overbrace{O((\log(n))^{1/2})}^{\text{from (C.8)}},$$

$$\tilde{Q}_U^I(u_{0,j}^H) - \tau_j = (\sqrt{n}(\tilde{Q}_U^I(u_{0,j}^H) - u_{0,j}^H))n^{-1/2} + \overbrace{(u_{0,j}^H - \tau_j)}^{\text{Lemma C.1}}$$
$$= O((\log(n))^{1/2}n^{-1/2}) + O(n^{-1/2}) = O((\log(n))^{1/2}n^{-1/2}). \tag{C.14}$$

For each j, the Taylor expansion is

$$Q(\tilde{Q}_U^I(u_{0,j}^H)) - Q(\tau_j) = Q'(\tau_j)(\tilde{Q}_U^I(u_{0,j}^H) - \tau_j) + \frac{1}{2}Q''(\tau_j)(\tilde{Q}_U^I(u_{0,j}^H) - \tau_j)^2$$
$$+ \frac{1}{6} \overbrace{\overbrace{Q'''(\tilde{u}_j)}^{=O(1) \text{ uniformly by A2.2}} \overbrace{(\tilde{Q}_U^I(u_{0,j}^H) - \tau_j)^3}^{=O(n^{-3/2}(\log(n))^{3/2}) \text{ by (C.14)}}}^{=O(n^{-3/2}(\log(n))^{3/2})} . \tag{C.15}$$

We continue from (C.5), making explicit the dependence on $\tilde{\alpha}$. Recall that for T_h, the true value of $\boldsymbol{\gamma}$ is used, so $\tilde{\alpha}$ is non-random, and $u_{0,j}^H = u_j^H(\tilde{\alpha})$. Plugging in (C.15), the third-order term can be pulled out using the (above) fact that $n^{1/2}\boldsymbol{\psi}'(Q(\tilde{Q}_U^I(\mathbf{u}_0^H)) - Q(\boldsymbol{\tau}))$ has an asymptotically normal PDF:

$$T_h(\tilde{\alpha}) = \mathrm{P}\Biggl(\sum_{j=1}^J \psi_j \overbrace{(Q'(\tau_j)(\tilde{Q}_U^I(u_{0,j}^H) - \tau_j) + \frac{Q''(\tau_j)}{2}(\tilde{Q}_U^I(u_{0,j}^H) - \tau_j)^2 + O(n^{-3/2}(\log(n))^{3/2}))}^{\text{from (C.15)}}$$
$$> 0\Biggr)$$
$$- \mathrm{P}\left(\sum_{j=1}^J \psi_j Q'(\tau_j)(\tilde{Q}_U^I(u_{0,j}^H) - \tau_j) > 0\right) + O(n^{-3/2}(\log(n))^3)$$
$$= \mathrm{P}\Biggl(\sum_{j=1}^J \psi_j Q'(\tau_j) \overbrace{\sqrt{n}(\tilde{Q}_U^I(u_{0,j}^H) - \tau_j)}^{\text{non-degenerate normal PDF}}$$
$$> -\sum_{j=1}^J \psi_j \frac{1}{2} Q''(\tau_j)\sqrt{n}(\tilde{Q}_U^I(u_{0,j}^H) - \tau_j)^2 - \overbrace{O(n^{-1}(\log(n))^{3/2})}^{\text{can pull out}}\Biggr)$$
$$- \mathrm{P}\left(\sum_{j=1}^J \psi_j Q'(\tau_j)(\tilde{Q}_U^I(u_{0,j}^H) - \tau_j) > 0\right)$$
$$+ O(n^{-3/2}(\log(n))^3)$$
$$= \overbrace{\mathrm{P}\left(\sum_{j=1}^J \psi_j Q'(\tau_j)(\tilde{Q}_U^I(u_j^H(\tilde{\alpha})) - \tau_j) > -\sum_{j=1}^J \psi_j \frac{1}{2} Q''(\tau_j)(\tilde{Q}_U^I(u_j^H(\tilde{\alpha})) - \tau_j)^2\right)}^{T_{H,1}}$$
$$- \overbrace{\mathrm{P}\left(\sum_{j=1}^J \psi_j Q'(\tau_j)(\tilde{Q}_U^I(u_j^H(\tilde{\alpha})) - \tau_j) > 0\right)}^{T_{H,2}}$$

$$+ O(n^{-1}(\log(n))^{3/2}).$$
$$= T_{H,1} - T_{H,2} + O(n^{-1}(\log(n))^{3/2}),$$
$$T_{H,1} \equiv \mathrm{P}\left(\sum_{j=1}^{J} \psi_j Q'(\tau_j)(\Delta_j^H + D_j^H) > -\frac{n^{-1/2}}{2}\sum_{j=1}^{J} \psi_j Q''(\tau_j)(\Delta_j^H + D_j^H)^2\right), \tag{C.16}$$
$$T_{H,2} \equiv \mathrm{P}\left(\sum_{j=1}^{J} \psi_j Q'(\tau_j)(\Delta_j^H + D_j^H) > 0\right). \tag{C.17}$$

Following the same steps,

$$T_l(\tilde{\alpha}) = T_{L,1} - T_{L,2} + O(n^{-1}(\log(n))^{3/2}), \tag{C.18}$$
$$T_{L,1} \equiv \mathrm{P}\left(\sum_{j=1}^{J} \psi_j Q'(\tau_j)(\Delta_j^L + D_j^L) < -\frac{n^{-1/2}}{2}\sum_{j=1}^{J} \psi_j Q''(\tau_j)(\Delta_j^L + D_j^L)^2\right), \tag{C.19}$$
$$T_{L,2} \equiv \mathrm{P}\left(\sum_{j=1}^{J} \psi_j Q'(\tau_j)(\Delta_j^L + D_j^L) < 0\right). \tag{C.20}$$

From (C.8), within $T_{H,1}$, $(\Delta_j^H + D_j^H)^2 = O(\log(n))$, and $\psi_j Q''(\tau_j) = O(1)$. Thus, $T_{H,1} - T_{H,2}$ is the probability that the (non-degenerate) Gaussian random variable

$$\sum_{j=1}^{J} \psi_j Q'(\tau_j)(\Delta_j^H + D_j^H)$$

is between zero and

$$-\frac{n^{-1/2}}{2}\sum_{j=1}^{J} \psi_j Q''(\tau_j)(\Delta_j^H + D_j^H)^2 = O(n^{-1/2}(\log(n))).$$

The Gaussian PDF is $O(1)$, and it remains $O(1)$ after conditioning on the event that all $\Delta_j^2 \le 2\log(n)$ (as in (C.8)) since the event's complement has smaller-order probability; heuristically,

$$f_{X|E}(x \mid E) = \mathbb{1}\{E\} f_X(x) / \mathrm{P}(E),$$

where $\mathrm{P}(E) = 1 - O(n^{-2}(\log(n))^{-1/2})$. Consequently, by the MVT, the probability of any interval of length $O(n^{-1/2}\log(n))$ is $O(n^{-1/2}\log(n))$:

$$\mathrm{P}(X \in [a,b] \mid E) = \int_a^b f_{X|E}(x \mid E)\, dx = \overbrace{(b-a)}^{=O(n^{-1/2}\log(n))} \overbrace{f_{X|E}(\tilde{x} \mid E)}^{=O(1)}.$$

Altogether,

$$\begin{aligned} T_h &= T_{H,1} - T_{H,2} + O(n^{-1}(\log(n))^{3/2}) \\ &= \overbrace{(T_{H,1} - T_{H,2} \mid \text{any } (\Delta_j^H)^2 > 2\log(n))}^{\le 1} \overbrace{\mathrm{P}(\text{any } (\Delta_j^H)^2 > 2\log(n))}^{=O(n^{-2}(\log(n))^{-1/2}) \text{ by (C.9)}} \\ &\quad + \overbrace{(T_{H,1} - T_{H,2} \mid \text{all } (\Delta_j^H)^2 \le 2\log(n))}^{=O(n^{-1/2}\log(n))} \overbrace{\mathrm{P}(\text{all } (\Delta_j^H)^2 \le 2\log(n))}^{\le 1} \end{aligned}$$

$$
\begin{aligned}
&+ O(n^{-1}(\log(n))^{3/2}) \\
&= O(n^{-2}(\log(n))^{-1/2} + n^{-1/2}\log(n) + n^{-1}(\log(n))^{3/2}) \\
&= O(n^{-1/2}\log(n)).
\end{aligned}
$$

The orders of the terms for T_l are identical, so the same result follows.

For the two-sided result, we must be more precise. The strategy is to consider the inside of $T_{H,1}$ as a quadratic in Δ_1^H conditional on $\Delta_2^H, \ldots, \Delta_J^H$. The "negative" root is proportional to $n^{1/2}$, so (still conditional on the other Δ_j^H) $T_{H,1}$ can be approximated as the probability that Δ_1^H is above the other root. These conditional probabilities are then integrated over the distribution of the other Δ_j^H, which are asymptotically jointly normal. Again, we restrict attention to when the $|\Delta_j^H| \le \sqrt{2\log(n)}$, which is violated with asymptotically negligible probability. Using a similar argument for $T_{L,1}$ and using the symmetry from, e.g., Lemma C.1 will then complete the proof.

We first rewrite $T_{H,2}$. Without loss of generality, let $\psi_1 = 1$. (If $\psi_1 \neq 1$, then letting $\tilde{\psi}_j = \psi_j/\psi_1$ achieves $\tilde{\psi}_1 = 1$, and the final confidence interval can be multiplied by ψ_1 to reverse the transformation.) Also, let $\gamma_j = Q'(\tau_j)/Q'(\tau_1)$, so $\gamma_1 = 1$. Let

$$
\boldsymbol{\Delta}_{-1}^H \equiv (\Delta_2^H, \ldots, \Delta_J^H)', \quad \boldsymbol{\Delta}_{-1}^L \equiv (\Delta_2^L, \ldots, \Delta_J^L)'. \tag{C.21}
$$

Define the function $\pi^{H,2}(\cdot) : \mathbb{R}^{J-1} \mapsto \mathbb{R}$ as

$$
\pi^{H,2}(\mathbf{v}_{-1}) = -D_0^H - \sum_{j=2}^{J} \psi_j \gamma_j v_j \tag{C.22}
$$

for any argument $\mathbf{v}_{-1} = (v_2, \ldots, v_J)' \in \mathbb{R}^{J-1}$, to help simplify $T_{H,2}$ as follows. Starting from (C.17), now we can rewrite $T_{H,2}$:

$$
\begin{aligned}
T_{H,2} &= \mathrm{P}\left(\sum_{j=1}^{J} \psi_j Q'(\tau_j)(\Delta_j^H + D_j^H) > 0\right) \\
&= \mathrm{P}\left(\sum_{j=1}^{J} \psi_j \gamma_j (\Delta_j^H + D_j^H) > 0\right) \\
&= \mathrm{P}(\Delta_1^H + \sum_{j=2}^{J} \psi_j \gamma_j \Delta_j^H > -D_0^H) \\
&= \mathrm{P}(\Delta_1^H > \pi^{H,2}(\boldsymbol{\Delta}_{-1}^H)).
\end{aligned} \tag{C.23}
$$

Starting from (C.16),

$$
\begin{aligned}
T_{H,1} &= \mathrm{P}\left(\sum_{j=1}^{J} \psi_j Q'(\tau_j)(\Delta_j^H + D_j^H) > -\frac{n^{-1/2}}{2}\sum_{j=1}^{J} \psi_j Q''(\tau_j)(\Delta_j^H + D_j^H)^2\right) \\
&= \mathrm{P}\left(\sum_{j=1}^{J} \psi_j \gamma_j (\Delta_j^H + D_j^H) > -\frac{n^{-1/2}}{2}\sum_{j=1}^{J} \psi_j \frac{Q''(\tau_j)}{Q'(\tau_1)}(\Delta_j^H + D_j^H)^2\right) \\
&= \mathrm{P}(D_0^H + \Delta_1^H + \sum_{j=2}^{J} \psi_j \gamma_j \Delta_j^H > -\frac{n^{-1/2}}{2}\frac{Q''(\tau_1)}{Q'(\tau_1)}((\Delta_1^H)^2 + 2\Delta_1^H D_1^H + (D_1^H)^2)
\end{aligned}
$$

$$
\begin{aligned}
&\quad - \frac{n^{-1/2}}{2}\sum_{j=2}^{J}\psi_j \frac{Q''(\tau_j)}{Q'(\tau_1)}(\Delta_j^H + D_j^H)^2) \\
&= \mathrm{P}\Bigg((\Delta_1^H)^2\left(\frac{n^{-1/2}}{2}\frac{Q''(\tau_1)}{Q'(\tau_1)}\right) + \Delta_1^H\left(1 + n^{-1/2}D_1^H\frac{Q''(\tau_1)}{Q'(\tau_1)}\right) \\
&\quad + (D_0^H + \sum_{j=2}^{J}\psi_j\gamma_j\Delta_j^H + \frac{n^{-1/2}}{2}\frac{Q''(\tau_1)}{Q'(\tau_1)}(D_1^H)^2 + \frac{n^{-1/2}}{2}\sum_{j=2}^{J}\psi_j\frac{Q''(\tau_j)}{Q'(\tau_1)}(\Delta_j^H + D_j^H)^2) \\
&\quad > 0\Bigg) \\
&= \mathrm{P}(a(\Delta_1^H)^2 + b\Delta_1^H + c > 0), \\
&a \equiv n^{-1/2}a_0, \quad a_0 \equiv \frac{Q''(\tau_1)}{2Q'(\tau_1)}, \\
&b \equiv 1 + n^{-1/2}b_0, \quad b_0 \equiv D_1^H\frac{Q''(\tau_1)}{Q'(\tau_1)}, \quad b^{-1} = 1 - n^{-1/2}b_0 + O(n^{-1}), \\
&c \equiv -\pi^{H,2}(\boldsymbol{\Delta}_{-1}^H) + \frac{n^{-1/2}}{2}\frac{Q''(\tau_1)}{Q'(\tau_1)}(D_1^H)^2 + \frac{n^{-1/2}}{2}\sum_{j=2}^{J}\psi_j\frac{Q''(\tau_j)}{Q'(\tau_1)}(\Delta_j^H + D_j^H)^2 \\
&= O((\log(n))^{1/2} + n^{-1/2} + n^{-1/2}\log(n)) = O((\log(n))^{1/2}).
\end{aligned}
$$

It's possible that $a = 0$, if (and only if) $a_0 = 0$, which occurs iff $Q''(\tau_1) = 0$ (e.g., if $\tau_1 = 0.5$ and the population is normal). In that case, the probability simplifies to

$$\mathrm{P}(\Delta_1^H > -c/b). \tag{C.24}$$

As noted below, this results in an equivalent expression (up to smaller-order terms) to the general case below.

The roots of $ax^2 + bx + c = 0$ are

$$r_- \equiv \frac{-b - \sqrt{b^2 - 4ac}}{2a}, \quad r_+ \equiv \frac{-b + \sqrt{b^2 - 4ac}}{2a}. \tag{C.25}$$

The subscript refers only to the $\pm$ in the quadratic formula; we may have either $r_- < r_+$ or $r_- > r_+$ depending on the sign of a. Both roots are real (for large enough n) since $b^2 \approx 1$ while $4ac = O(n^{-1/2}(\log(n))^{1/2}) \to 0$. Since $a = O(n^{-1/2})$, we approximate $g(a) = \sqrt{b^2 - 4ac}$ around $a = 0$:

$$
\begin{aligned}
g'(a) &= (1/2)(b^2 - 4ac)^{-1/2}(-4c), \quad g''(a) = (-1/4)(b^2 - 4ac)^{-3/2}(-4c)^2, \\
g'''(a) &= (3/8)(b^2 - 4ac)^{-5/2}(-4c)^3, \\
g(a) &= g(0) + ag'(0) + (1/2)a^2 g''(0) + (1/6)a^3 g'''(\tilde{a}), \\
\sqrt{b^2 - 4ac} &= b + a(1/2)b^{-1}(-4c) + (1/2)a^2(-1/4)b^{-3}(-4c)^2 + O(n^{-3/2}(\log(n))^{3/2}), \\
r_+ &= \frac{-b + b - 2ac/b - 2a^2c^2/b^3 + O(n^{-3/2}(\log(n))^{3/2})}{2a} = -\frac{c}{b} - \frac{ac^2}{b^3} + O(n^{-1}(\log(n))^{3/2}) \\
&= -c(1 - n^{-1/2}b_0) - n^{-1/2}a_0(\pi^{H,2}(\boldsymbol{\Delta}_{-1}^H))^2 + O(n^{-1}(\log(n))^{3/2}).
\end{aligned}
$$

Similar to (C.22), define the function $\pi^{H,1}(\cdot) : \mathbb{R}^{J-1} \mapsto \mathbb{R}$ so that

$$r_+ = \pi^{H,1}(\mathbf{\Delta}_{-1}^H) = -c(1-n^{-1/2}b_0)-n^{-1/2}a_0(\pi^{H,2}(\mathbf{\Delta}_{-1}^H))^2+O(n^{-1}(\log(n))^{3/2}), \quad \text{(C.26)}$$

where c also depends implicitly on the argument.

For r_-,

$$\begin{aligned} r_- &= \frac{-b-b+2ac/b+2a^2c^2/b^3+O(n^{-3/2}(\log(n))^{3/2})}{2a} \\ &= -b/a+c/b+ac^2/b^3+O(n^{-1}(\log(n))^{3/2}) \\ &= -n^{1/2}(1+n^{-1/2}b_0)/a_0+O((\log(n))^{1/2})+O(n^{-1/2}\log(n))+O(n^{-3/2}(\log(n))^{3/2}) \\ &= -n^{1/2}/a_0+O((\log(n))^{1/2}). \end{aligned}$$

If $a_0 = 0$, then there is essentially only r_+, as seen in (C.24). If $a_0 \neq 0$, then $r_- \asymp n^{1/2}$ since a_0 is a finite, fixed constant by Assumption A2.2.

If $a > 0$, then $r_- < r_+$ and the function $ax^2 + bx + c$ is positive when either $x < r_-$ or $x > r_+$:

$$\begin{aligned} T_{H,1} &= \mathrm{P}(a(\Delta_1^H)^2 + b\Delta_1^H + c > 0) \\ &= \mathrm{P}(\Delta_1^H > r_+) + \mathrm{P}(\Delta_1^H < r_-). \end{aligned}$$

Conditional on the other Δ_j^H, even if they are all as big as allowed by (C.8), they are only of order $\sqrt{\log(n)}$ while r_- is of polynomial $n^{1/2}$ order (and the variance of Δ_1^H is still $O(1)$), so we always have an exponentially small $\mathrm{P}(\Delta_1^H < r_-) = O(e^{-0.99n})$, to give a loose but sufficiently small bound.

If $a < 0$, then $r_+ < r_-$ and the function $ax^2 + bx + c$ is positive iff $r_+ < x < r_-$, so

$$\begin{aligned} T_{H,1} &= \mathrm{P}(a(\Delta_1^H)^2 + b\Delta_1^H + c > 0) \\ &= \mathrm{P}(\Delta_1^H > r_+) - \mathrm{P}(\Delta_1^H > r_-). \end{aligned}$$

Again, given any $\mathbf{\Delta}_{-1}^H$ satisfying the bound in (C.8), the $\mathrm{P}(\Delta_1^H > r_-)$ term is exponentially small. Thus, regardless of $a > 0$ or $a < 0$,

$$T_{H,1} = \mathrm{P}(\Delta_1^H > r_+) + O(e^{-0.99n}). \quad \text{(C.27)}$$

When $a = 0$, (C.24) had $\mathrm{P}(\Delta_1^H > -c/b)$. This is the same as above but without the exponentially small error and without the $-ac^2/b^3$ or remainder term inside r_+.

To fully treat $T_{H,1}$, we must account for the fact that r_+ contains random variables, by integrating a conditional probability over the distribution of the conditioning variable, $\mathbf{\Delta}_{-1}^H$:

$$\begin{aligned} T_{H,1} &= E_{\mathbf{\Delta}_{-1}^H}[\mathrm{P}(\Delta_1^H > r_+ \mid \mathbf{\Delta}_{-1}^H)] \\ &= \underset{v_j^2 \leq 2\log(n),\, j\geq 2}{\int \cdots \int} \int_{r_+}^{\infty} f_{\Delta_1^H \mid \mathbf{\Delta}_{-1}^H}(v_1 \mid \mathbf{v}_{-1}) f_{\mathbf{\Delta}_{-1}^H}(\mathbf{v}_{-1})\, dv_1\, dv_2 \ldots dv_J. \end{aligned} \quad \text{(C.28)}$$

Overall,

$$
\begin{aligned}
T_h &= T_{H,1} - T_{H,2} + O(n^{-1}(\log(n))^{3/2}) \\
&= \overbrace{\mathrm{P}(\Delta_1^H > \pi^{H,1}(\boldsymbol{\Delta}_{-1}^H)) + O(e^{-0.99n})}^{\text{from (C.26) and (C.27)}} - \overbrace{\mathrm{P}(\Delta_1^H > \pi^{H,2}(\boldsymbol{\Delta}_{-1}^H))}^{\text{from (C.23)}} + O(n^{-1}(\log(n))^{3/2}),
\end{aligned} \tag{C.29}
$$

so we need the probability that Δ_1^H is between $\pi^{H,1}(\boldsymbol{\Delta}_{-1}^H)$ and $\pi^{H,2}(\boldsymbol{\Delta}_{-1}^H)$. Towards that end,

$$
\begin{aligned}
&\overbrace{\pi^{H,1}(\boldsymbol{\Delta}_{-1}^H)}^{\text{from (C.26)}} - \pi^{H,2}(\boldsymbol{\Delta}_{-1}^H) \\
&= -c(1 - n^{-1/2} b_0) - n^{-1/2} a_0 (\pi^{H,2}(\boldsymbol{\Delta}_{-1}^H))^2 + O(n^{-1}(\log(n))^{3/2}) - \pi^{H,2}(\boldsymbol{\Delta}_{-1}^H) \\
&= \overbrace{(\pi^{H,2}(\boldsymbol{\Delta}_{-1}^H) - \frac{n^{-1/2} Q''(\tau_1)}{2Q'(\tau_1)} (D_1^H)^2 - \frac{n^{-1/2}}{2} \sum_{j=2}^J \psi_j \frac{Q''(\tau_j)}{Q'(\tau_1)} (\Delta_j^H + D_j^H)^2)}^{-c} \\
&\quad + n^{-1/2} \overbrace{(-\pi^{H,2}(\boldsymbol{\Delta}_{-1}^H) + O(n^{-1/2}\log(n)))}^{c} b_0 - n^{-1/2} a_0 (\pi^{H,2}(\boldsymbol{\Delta}_{-1}^H))^2 \\
&\quad - \pi^{H,2}(\boldsymbol{\Delta}_{-1}^H) + O(n^{-1}(\log(n))^{3/2}) \\
&= -\frac{n^{-1/2} Q''(\tau_1)}{2Q'(\tau_1)} (D_1^H)^2 - \frac{n^{-1/2}}{2} \sum_{j=2}^J \psi_j \frac{Q''(\tau_j)}{Q'(\tau_1)} (\Delta_j^H + D_j^H)^2 \\
&\quad - n^{-1/2} \pi^{H,2}(\boldsymbol{\Delta}_{-1}^H) b_0 - n^{-1/2} a_0 (\pi^{H,2}(\boldsymbol{\Delta}_{-1}^H))^2 + O(n^{-1}(\log(n))^{3/2}).
\end{aligned}
$$

Plugging in for a_0 and b_0, and emphasising that $\pi^{H,1}(\cdot)$ and $\pi^{H,2}(\cdot)$ are functions taking any argument $\mathbf{v}_{-1}$,

$$
\begin{aligned}
&\pi^{H,2}(\mathbf{v}_{-1}) - \pi^{H,1}(\mathbf{v}_{-1}) \\
&= -(\pi^{H,1}(\mathbf{v}_{-1}) - \pi^{H,2}(\mathbf{v}_{-1})) \\
&= \frac{n^{-1/2} Q''(\tau_1)}{2Q'(\tau_1)} (D_1^H)^2 + \frac{n^{-1/2}}{2} \sum_{j=2}^J \psi_j \frac{Q''(\tau_j)}{Q'(\tau_1)} (v_j + D_j^H)^2 \\
&\quad + n^{-1/2} \pi^{H,2}(\mathbf{v}_{-1}) D_1^H \frac{Q''(\tau_1)}{Q'(\tau_1)} + n^{-1/2} \frac{Q''(\tau_1)}{Q'(\tau_1)} (\pi^{H,2}(\mathbf{v}_{-1}))^2 + O(n^{-1}(\log(n))^{3/2}) \\
&= O(n^{-1/2}\log(n)).
\end{aligned}
$$

Then, using (C.28) and (C.29), letting $\underline{\mathcal{V}}_{-1}$ be the $(J-1)\times(J-1)$ lower-right submatrix of $\underline{\mathcal{V}}$, and letting $\phi_{V_1|\mathbf{V}_{-1}}(v_1 \mid \mathbf{v}_{-1})$ denote the conditional PDF of V_1 given $\mathbf{V}_{-1}$ for random vector $\mathbf{V}$ with PDF $\phi_{\underline{\mathcal{V}}}(\cdot)$,

$$
T_h = \int \cdots \int_{v_j^2 \le 2\log(n),\, j\ge 2} \int_{\pi^{H,1}(\mathbf{v}_{-1})}^{\pi^{H,2}(\mathbf{v}_{-1})} f_{\Delta_1^H|\boldsymbol{\Delta}_{-1}^H}(v_1 \mid \mathbf{v}_{-1}) f_{\boldsymbol{\Delta}_{-1}^H}(\mathbf{v}_{-1})\, dv_1\, dv_2 \dots dv_J + O(n^{-1}(\log(n))^{3/2})
$$

$$
\begin{aligned}
&= \int\limits_{v_j^2 \le 2\log(n),\, j\ge 2} \cdots \int \int_{\pi^{H,1}(\mathbf{v}_{-1})}^{\pi^{H,2}(\mathbf{v}_{-1})} f_{\boldsymbol{\Delta}^H}(\mathbf{v})\, dv_1\, dv_2 \ldots dv_J + O(n^{-1}(\log(n))^{3/2}) \\
&= \int\limits_{v_j^2 \le 2\log(n),\, j\ge 2} \cdots \int \int_{\pi^{H,1}(\mathbf{v}_{-1})}^{\pi^{H,2}(\mathbf{v}_{-1})} \overbrace{\phi_{\underline{\mathcal{V}}}(\mathbf{v})(1 + O(n^{-1/2}\log(n)))}^{\text{by (C.10) but with } x = O((\log(n))^{1/2})}\, dv_1\, dv_2 \ldots dv_J + O(n^{-1}(\log(n))^{3/2}) \\
&= \int\limits_{v_j^2 \le 2\log(n),\, j\ge 2} \cdots \int \int_{\pi^{H,1}(\mathbf{v}_{-1})}^{\pi^{H,2}(\mathbf{v}_{-1})} \phi_{V_1 \mid \mathbf{V}_{-1}}(v_1 \mid \mathbf{v}_{-1}) \phi_{\underline{\mathcal{V}}_{-1}}(\mathbf{v}_{-1})\, dv_1\, dv_2 \ldots dv_J (1 + O(n^{-1/2}\log(n))) \\
&\quad + O(n^{-1}(\log(n))^{3/2}) \\
&= \int\limits_{v_j^2 \le 2\log(n),\, j\ge 2} \cdots \int \overbrace{(\pi^{H,2}(\mathbf{v}_{-1}) - \pi^{H,1}(\mathbf{v}_{-1}))\phi_{V_1 \mid \mathbf{V}_{-1}}(\tilde{v}_1 \mid \mathbf{v}_{-1})}^{\text{mean value theorem}} \\
&\qquad \times \phi_{\underline{\mathcal{V}}_{-1}}(\mathbf{v}_{-1})\, dv_2 \ldots dv_J (1 + O(n^{-1/2}\log(n))) \\
&\quad + O(n^{-1}(\log(n))^{3/2}) \\
&= \int\limits_{v_j^2 \le 2\log(n),\, j\ge 2} \cdots \int \overbrace{(\pi^{H,2}(\mathbf{v}_{-1}) - \pi^{H,1}(\mathbf{v}_{-1}))}^{= O(n^{-1/2}\log(n))} (\phi_{V_1 \mid \mathbf{V}_{-1}}(\pi^{H,2}(\mathbf{v}_{-1}) \mid \mathbf{v}_{-1}) + O(n^{-1/2}\log(n))) \\
&\qquad \times \phi_{\underline{\mathcal{V}}_{-1}}(\mathbf{v}_{-1})\, dv_2 \ldots dv_J \\
&\quad + O(n^{-1}(\log(n))^2) \\
&= \int\limits_{v_j^2 \le 2\log(n),\, j\ge 2} \cdots \int (\pi^{H,2}(\mathbf{v}_{-1}) - \pi^{H,1}(\mathbf{v}_{-1}))\phi_{V_1 \mid \mathbf{V}_{-1}}(\pi^{H,2}(\mathbf{v}_{-1}) \mid \mathbf{v}_{-1}) \phi_{\underline{\mathcal{V}}}(\mathbf{v}_{-1})\, dv_2 \ldots dv_J \\
&\quad + O(n^{-1}(\log(n))^2). \qquad\qquad \text{(C.30)}
\end{aligned}
$$

For T_l, the results are parallel, eventually leading to $T_l = -T_h + O(n^{-1}(\log(n))^2)$ in (C.34). Returning to (C.18),

$$
\begin{aligned}
T_l &= T_{L,1} - T_{L,2} + O(n^{-1}(\log(n))^{3/2}), \\
T_{L,2} &= \mathrm{P}(\Delta_1^L + \sum_{j=2}^{J} \psi_j \gamma_j \Delta_j^L < -D_0^L) \\
&= \mathrm{P}(\Delta_1^L < \pi^{L,2}(\boldsymbol{\Delta}_{-1}^L)), \\
\pi^{L,2}(\mathbf{v}_{-1}) &\equiv -\sum_{j=2}^{J} \psi_j \gamma_j v_j - D_0^L, \\
T_{L,1} &= \mathrm{P}\left(\sum_{j=1}^{J} \psi_j Q'(\tau_j)(\Delta_j^L + D_j^L) < -\frac{n^{-1/2}}{2} \sum_{j=1}^{J} \psi_j Q''(\tau_j)(\Delta_j^L + D_j^L)^2 \right) \\
&= \mathrm{P}\left(\sum_{j=1}^{J} \psi_j \gamma_j (\Delta_j^L + D_j^L) < -\frac{n^{-1/2}}{2} \sum_{j=1}^{J} \psi_j \frac{Q''(\tau_j)}{Q'(\tau_1)} (\Delta_j^L + D_j^L)^2 \right)
\end{aligned}
$$

$$
\begin{aligned}
&= \mathrm{P}(D_0^L + \Delta_1^L + \sum_{j=2}^{J} \psi_j \gamma_j \Delta_j^L < -\frac{n^{-1/2}}{2} \frac{Q''(\tau_1)}{Q'(\tau_1)} ((\Delta_1^L)^2 + 2\Delta_1^L D_1^L + (D_1^L)^2) \\
&\qquad - \frac{n^{-1/2}}{2} \sum_{j=2}^{J} \psi_j \frac{Q''(\tau_j)}{Q'(\tau_1)} (\Delta_j^L + D_j^L)^2) \\
&= \mathrm{P}\Bigg((\Delta_1^L)^2 \left(\frac{n^{-1/2}}{2} \frac{Q''(\tau_1)}{Q'(\tau_1)} \right) + \Delta_1^L \left(1 + n^{-1/2} D_1^L \frac{Q''(\tau_1)}{Q'(\tau_1)} \right) \\
&\qquad + (D_0^L + \sum_{j=2}^{J} \psi_j \gamma_j \Delta_j^L + \frac{n^{-1/2}}{2} \frac{Q''(\tau_1)}{Q'(\tau_1)} (D_1^L)^2 \\
&\qquad + \frac{n^{-1/2}}{2} \sum_{j=2}^{J} \psi_j \frac{Q''(\tau_j)}{Q'(\tau_1)} (\Delta_j^L + D_j^L)^2) \\
&\qquad < 0 \Bigg) \\
&= \mathrm{P}(a(\Delta_1^L)^2 + b\Delta_1^L + c < 0), \\
& a \equiv n^{-1/2} a_0, \quad a_0 \equiv \frac{Q''(\tau_1)}{2Q'(\tau_1)}, \\
& b \equiv 1 + n^{-1/2} b_0, \quad b_0 \equiv D_1^L \frac{Q''(\tau_1)}{Q'(\tau_1)}, \\
& c \equiv -\pi^{L,2}(\boldsymbol{\Delta}_{-1}^L) + \frac{n^{-1/2}}{2} \frac{Q''(\tau_1)}{Q'(\tau_1)} (D_1^L)^2 + \frac{n^{-1/2}}{2} \sum_{j=2}^{J} \psi_j \frac{Q''(\tau_j)}{Q'(\tau_1)} (\Delta_j^L + D_j^L)^2,
\end{aligned}
$$

with rates identical to those for $T_{H,1}$. Thus, the arguments concerning the roots of the quadratic are identical, only changing $>$ to $<$:

$$
\begin{aligned}
T_{L,1} &= \mathrm{P}(\Delta_1^L < \pi^{L,1}(\boldsymbol{\Delta}_{-1}^L)) + O(e^{-0.99n}), \\
\pi^{L,1}(\boldsymbol{\Delta}_{-1}^L) &\equiv -c(1 - n^{-1/2} b_0) - n^{-1/2} a_0 (\pi^{L,2}(\boldsymbol{\Delta}_{-1}^L))^2 + O(n^{-1}(\log(n))^{3/2}).
\end{aligned}
$$

Since the inequality is the opposite direction from T_h, instead of $\pi^{H,2}(\mathbf{v}_{-1}) - \pi^{H,1}(\mathbf{v}_{-1})$ we compute

$$
\begin{aligned}
&\pi^{L,1}(\mathbf{v}_{-1}) - \pi^{L,2}(\mathbf{v}_{-1}) \\
&= -c(1 - n^{-1/2} b_0) - n^{-1/2} a_0 (\pi^{L,2}(\mathbf{v}_{-1}))^2 + O(n^{-1}(\log(n))^{3/2}) - \pi^{L,2}(\mathbf{v}_{-1}) \\
&= -\frac{n^{-1/2} Q''(\tau_1)}{2Q'(\tau_1)} (D_1^L)^2 - \frac{n^{-1/2}}{2} \sum_{j=2}^{J} \psi_j \frac{Q''(\tau_j)}{Q'(\tau_1)} (v_j + D_j^L)^2 \\
&\quad - n^{-1/2} \pi^{L,2}(\mathbf{v}_{-1}) D_1^L \frac{Q''(\tau_1)}{Q'(\tau_1)} - n^{-1/2} (\pi^{L,2}(\mathbf{v}_{-1}))^2 \frac{Q''(\tau_1)}{2Q'(\tau_1)} + O(n^{-1}(\log(n))^{3/2}).
\end{aligned}
$$

Repeating the steps in the derivation of (C.30) yields

$$
T_l = \underset{v_j^2 \le 2\log(n),\, j \ge 2}{\int \cdots \int} (\pi^{L,1}(\mathbf{v}_{-1}) - \pi^{L,2}(\mathbf{v}_{-1})) \phi_{V_1 | \mathbf{V}_{-1}} (\pi^{L,2}(\mathbf{v}_{-1}) \mid \mathbf{v}_{-1}) \phi_{\underline{\mathcal{V}}_{-1}}(\mathbf{v}_{-1})\, dv_2 \ldots dv_J
$$

$$+ O(n^{-1}(\log(n))^2). \tag{C.31}$$

Since $D_0^L = -D_0^H + O(n^{-1/2})$,

$$\pi^{L,2}(-\mathbf{v}_{-1}) = -\pi^{H,2}(\mathbf{v}_{-1}) + O(n^{-1/2}). \tag{C.32}$$

Also,

$$\begin{aligned}
&\pi^{L,1}(-\mathbf{v}_{-1}) - \pi^{L,2}(-\mathbf{v}_{-1}) \\
&= -\frac{n^{-1/2}Q''(\tau_1)}{2Q'(\tau_1)}(D_1^L)^2 - \frac{n^{-1/2}}{2}\sum_{j=2}^{J}\psi_j\frac{Q''(\tau_j)}{Q'(\tau_1)}(v_j^2 - 2v_jD_j^L + (D_j^L)^2) \\
&\quad - n^{-1/2}\pi^{L,2}(-\mathbf{v}_{-1})D_1^L\frac{Q''(\tau_1)}{Q'(\tau_1)} - n^{-1/2}(\pi^{L,2}(-\mathbf{v}_{-1}))^2\frac{Q''(\tau_1)}{2Q'(\tau_1)} + O(n^{-1}(\log(n))^{3/2}) \\
&= -\frac{n^{-1/2}Q''(\tau_1)}{2Q'(\tau_1)}(-D_1^H + O(n^{-1/2}))^2 \\
&\quad - \frac{n^{-1/2}}{2}\sum_{j=2}^{J}\psi_j\frac{Q''(\tau_j)}{Q'(\tau_1)}(v_j^2 - 2v_j(-D_j^H + O(n^{-1/2})) + (-D_j^H + O(n^{-1/2}))^2) \\
&\quad - n^{-1/2}(-\pi^{H,2}(\mathbf{v}_{-1}) + O(n^{-1/2}))(-D_1^H + O(n^{-1/2}))\frac{Q''(\tau_1)}{Q'(\tau_1)} \\
&\quad - n^{-1/2}(-\pi^{H,2}(\mathbf{v}_{-1}) + O(n^{-1/2}))^2\frac{Q''(\tau_1)}{2Q'(\tau_1)} + O(n^{-1}(\log(n))^{3/2}) \\
&= -\frac{n^{-1/2}Q''(\tau_1)}{2Q'(\tau_1)}(D_1^H)^2 - \frac{n^{-1/2}}{2}\sum_{j=2}^{J}\psi_j\frac{Q''(\tau_j)}{Q'(\tau_1)}(v_j^2 + 2v_jD_j^H + (D_j^H)^2) \\
&\quad - n^{-1/2}\pi^{H,2}(\mathbf{v}_{-1})D_1^H\frac{Q''(\tau_1)}{Q'(\tau_1)} - n^{-1/2}(\pi^{H,2}(\mathbf{v}_{-1}))^2\frac{Q''(\tau_1)}{2Q'(\tau_1)} + O(n^{-1}(\log(n))^{3/2}) \\
&= \pi^{H,1}(\mathbf{v}_{-1}) - \pi^{H,2}(\mathbf{v}_{-1}) + O(n^{-1}(\log(n))^{3/2}).
\end{aligned} \tag{C.33}$$

Because a mean-zero normal distribution is symmetric about zero, and the area of integration is also symmetric about zero, we can replace $\mathbf{v}_{-1}$ with $-\mathbf{v}_{-1}$. Heuristically, in one dimension, with the change of variables $y = -x$,

$$\int_a^b g(x)\phi(x)\,dx = \int_{-a}^{-b} g(-y)\phi(-y)\,(-dy) = \int_{-b}^{-a} g(-y)\phi(y)\,dy,$$

and if $b = -a$ then $[-b, -a] = [a, b]$. Thus, continuing,

$$\begin{aligned}
T_l &= \int\!\!\cdots\!\!\int_{v_j^2 \le 2\log(n),\, j\ge 2} \overbrace{(\pi^{L,1}(-\mathbf{v}_{-1}) - \pi^{L,2}(-\mathbf{v}_{-1}))}^{\text{apply (C.33)}}\overbrace{\phi_{V_1|\mathbf{V}_{-1}}(\pi^{L,2}(-\mathbf{v}_{-1}) \mid -\mathbf{v}_{-1})}^{\text{apply (C.32)}}\phi_{\underline{\mathcal{V}}_{-1}}(\mathbf{v}_{-1})\,dv_2\ldots dv_J \\
&\quad + O(n^{-1}(\log(n))^2) \\
&= \int\!\!\cdots\!\!\int_{v_j^2 \le 2\log(n),\, j\ge 2} (\pi^{H,1}(\mathbf{v}_{-1}) - \pi^{H,2}(\mathbf{v}_{-1}) + O(n^{-1}(\log(n))^{3/2})) \\
&\qquad\qquad \times \phi_{V_1|\mathbf{V}_{-1}}(-\pi^{H,2}(\mathbf{v}_{-1}) + O(n^{-1/2}) \mid -\mathbf{v}_{-1})\phi_{\underline{\mathcal{V}}_{-1}}(\mathbf{v}_{-1})\,dv_2\ldots dv_J
\end{aligned}$$

$$
\begin{aligned}
&+ O(n^{-1}(\log(n))^2) \\
&= - \int\cdots\int_{v_j^2 \le 2\log(n),\, j\ge 2} (\pi^{H,2}(\mathbf{v}_{-1}) - \pi^{H,1}(\mathbf{v}_{-1}))\phi_{V_1|\mathbf{V}_{-1}}(\pi^{H,2}(\mathbf{v}_{-1}) \mid \mathbf{v}_{-1})\phi_{\underline{\mathcal{V}}_{-1}}(\mathbf{v}_{-1})\, dv_2 \dots dv_J \\
&+ O(n^{-1}(\log(n))^2) \\
&= -T_h + O(n^{-1}(\log(n))^2) \qquad \text{(C.34)}
\end{aligned}
$$

since $\phi_{V_1|\mathbf{V}_{-1}}(v_1 \mid \mathbf{v}_{-1}) = \phi_{V_1|\mathbf{V}_{-1}}(-v_1 \mid -\mathbf{v}_{-1})$ by symmetry, using the formula for the conditional distribution from a mean-zero multivariate normal distribution. Thus,

$$
T_h(\tilde{\alpha}) + T_l(\tilde{\alpha}) = T_h(\tilde{\alpha}) + (-T_h(\tilde{\alpha}) + O(n^{-1}(\log(n))^2)) = O(n^{-1}(\log(n))^2)
$$

for any $\tilde{\alpha}$ in the range of possible values (which is bounded away from zero and fixed as $n \to \infty$). □

Proof of Lemma A.5 (main appendix): We continue to use notation from (C.2). Additionally, for the sparsity (nuisance parameter) estimator, used only with $\mathbf{u} = \boldsymbol{\tau}$, with smoothing parameters m_j,

$$
\begin{aligned}
\Omega_j &\equiv U_{n:k_j+m_j} - U_{n:k_j-m_j} = \Omega_j^+ + \Omega_j^-, \\
\Omega_j^- &\equiv U_{n:k_j} - U_{n:k_j-m_j}, \quad \Omega_j^+ \equiv U_{n:k_j+m_j} - U_{n:k_j}.
\end{aligned} \qquad \text{(C.35)}
$$

As in (C.2), let $\mathbf{Y}^{\boldsymbol{\tau}} \equiv (U_{n:\lfloor(n+1)\tau_j\rfloor})_{j=1}^J$, with Ω_j^+ and Ω_j^- defined with respect to this $\mathbf{Y}^{\boldsymbol{\tau}}$ as in (C.35). For simplicity, we write m instead of m_j since all m_j have the same rate. We return to using $\boldsymbol{\gamma} \equiv Q'(\boldsymbol{\tau})$ instead of the normalised version.

We continue to assume Condition $\star$ holds for realisations of all random variables, so

$$
\mathbf{Y}^{\boldsymbol{\tau}} - \boldsymbol{\tau} = O(n^{-1/2}\log(n)), \quad \boldsymbol{\Omega} - 2m/(n+1) = O(n^{-1}m^{1/2}\log(n)) \implies \boldsymbol{\Omega} = O(m/n)
$$

since $\log(n) \lesssim m^{1/2}$ as assumed in Theorem A.3.

We start from the sparsity estimator in (3.5), using a mean value expansion where $\tilde{Y}_j$ and $\tilde{\tilde{Y}}_j$ are between $Y_j - \Omega_j^-$ and $Y_j + \Omega_j^+$. Under Condition $\star$, $\tilde{Y}_j \to \tau_j$, so Assumption A2.2 implies $Q'(\tilde{Y}_j)$ and $Q''(\tilde{Y}_j)$ exist and are uniformly bounded for large enough n, and similarly for $\tilde{\tilde{Y}}_j$. So,[22]

$$
\begin{aligned}
\widehat{Q'}(\tau_j) &= \frac{n}{2m}(X_{n:\lfloor(n+1)\tau_j\rfloor+m_j} - X_{n:\lfloor(n+1)\tau_j\rfloor-m_j}) \\
&= \frac{n}{2m}(Q(Y_j + \Omega_j^+) - Q(Y_j - \Omega_j^-)) \\
&= \frac{n}{2m}(Q'(Y_j)\Omega_j + (1/2)Q''(\tilde{Y}_j)\Omega_j^2) \\
&= \frac{n}{2m}\Omega_j(Q'(\tau_j) + Q''(\tilde{\tilde{Y}}_j)(Y_j - \tau_j)) + O((n/m)\Omega_j^2) \\
&= \frac{n}{2m}(2m/(n+1) + O(n^{-1}m^{1/2}\log(n)))(Q'(\tau_j) + Q''(\tilde{\tilde{Y}}_j)(Y_j - \tau_j)) \\
&\quad + O((n/m)(m/n)^2)
\end{aligned}
$$

[22] For comparison, from equations (2.5) and (2.6) in Bloch and Gastwirth (1968), the bias and standard deviation of $\widehat{Q'}(\tau_j)$ are respectively $O(m^2/n^2)$ and $O(m^{-1/2})$, so $\widehat{Q'}(\tau_j) = O_p(m^{-1/2} + (m/n)^2)$. The m^2/n^2 comes from the next term in the Taylor approximation because the second-order term above zeroes out when taking an expectation.

$$
\begin{aligned}
&= (1 + O(m^{-1/2}\log(n)))(Q'(\tau_j) + O(1)O(n^{-1/2}\log(n))) + O(m/n) \\
&= Q'(\tau_j) + O(m^{-1/2}\log(n)) + O(n^{-1/2}\log(n)) + O(m/n) \\
&= Q'(\tau_j) + O(m^{-1/2}\log(n) + m/n). \qquad \text{(C.36)}
\end{aligned}
$$

From Lemma C.1, where the $O(n^{-1})$ remainder is uniform since $\tilde{\alpha}$ is bounded away from zero and thus $z_{1-\tilde{\alpha}}$ is bounded,

$$
\begin{aligned}
\frac{d}{d\tilde{\alpha}} u_j^h(\tilde{\alpha}) &= \frac{d}{d\tilde{\alpha}} (\tau_j + n^{-1/2}\Phi^{-1}(1-\tilde{\alpha})\sqrt{\tau_j(1-\tau_j)} - \frac{2\tau_j - 1}{6n}(z_{1-\tilde{\alpha}}^2 + 2) + O(n^{-3/2})) \\
&= 0 - n^{-1/2}\frac{\sqrt{\tau_j(1-\tau_j)}}{\phi(\Phi^{-1}(1-\tilde{\alpha}))} + O(n^{-1}) = O(n^{-1/2}),
\end{aligned}
\qquad \text{(C.37)}
$$

and the order is the same for the derivative of $u_j^l(\tilde{\alpha})$.

We also will need the derivative of the function $\tilde{\alpha}(\cdot)$ to be $O(1)$, with generic argument $\mathbf{g}$. In the subsequent application, $\tilde{g}_j$ is a value $\tilde{Q}'(\tau_j)$ between the true $Q'(\tau_j)$ and estimated $\widehat{Q'}(\tau_j)$ that arises from the MVT. Consequently, restrictions on the range of possible values $\widehat{Q'}(\tau_j)$ under Condition $\star$ apply to $\tilde{g}_j$, too. Now, the result that $\frac{\partial\tilde{\alpha}}{\partial g_j}$ evaluated at $\tilde{g}_j$ is uniformly bounded for all j can be shown with the implicit function theorem. Let P denote (temporarily) the RHS of equation (A.7) or (A.8) from the main appendix, so

$$
\frac{\partial\tilde{\alpha}}{\partial g_j} = -\frac{\partial P/\partial g_j}{\partial P/\partial\tilde{\alpha}}.
$$

Thus, it suffices to show that $\frac{\partial P}{\partial g_j} = O(1)$ uniformly over possible $\tilde{g}_j$ values and that $\frac{\partial P}{\partial\tilde{\alpha}}$ is (uniformly) bounded away from zero. Heuristically, note that $g_j \in (0,\infty)$ while $\tilde{\alpha} \in [\alpha, 1]$, and asymptotically, (near) singularity points occur with smaller-order probability, so it is intuitive that $\frac{\partial\tilde{\alpha}}{\partial g_j}$ is bounded.

To the first order, $P = \mathrm{P}(n^{-1/2}Z < 0) = \mathrm{P}(Z < 0)$ (or > 0) where $Z \overset{a}{\sim} \mathrm{N}(\mu(\mathbf{g},\tilde{\alpha}), \sigma^2(\mathbf{g},\tilde{\alpha}))$, since Z is a linear combination of (asymptotically) jointly normal random variables with means and variances depending on $\mathbf{g}$ and $\tilde{\alpha}$. More explicitly, (A.8) from the main appendix is

$$
\begin{aligned}
0 &= \mathrm{P}\left(\sum_{j=1}^{J} \psi_j g_j \overbrace{\sqrt{n}(\tilde{Q}_U^I(u_j^L(\tilde{\alpha})) - \tau_j)}^{\equiv Z_j \overset{a}{\sim}\text{Normal}} < 0\right) - (1-\alpha) \\
&= \mathrm{P}(\overbrace{\mathbf{c}'\mathbf{Z}}^{=Z \overset{a}{\sim}\mathrm{N}(\mu(\mathbf{g},\tilde{\alpha}),\sigma^2(\mathbf{g},\tilde{\alpha}))} < 0) - (1-\alpha) \\
&\doteq \Phi\left(-\frac{\mu(\mathbf{g},\tilde{\alpha})}{\sigma(\mathbf{g},\tilde{\alpha})}\right) - (1-\alpha),
\end{aligned}
$$

where the linear combination weights are $c_j = \psi_j g_j$. Using (C.7) and (C.11), the normality is from

$$
\begin{aligned}
&\sqrt{n}(\tilde{Q}_U^I(\mathbf{u}^L(\tilde{\alpha})) - \boldsymbol{\tau}) = \boldsymbol{\Delta}^L + \mathbf{D}^L, \\
&D_j^L = \overbrace{-\Phi^{-1}(1-\tilde{\alpha})\sqrt{\tau_j(1-\tau_j)} + O(n^{-1/2})}^{\text{from (C.12)}}. \qquad \text{(C.38)}
\end{aligned}
$$

From (C.10), $\mathbf{\Delta}^L$ is asymptotically multivariate normal with covariance matrix $\underline{\mathcal{V}}$ having elements $\mathcal{V}_{i,k} = \min\{\tau_i, \tau_k\} - \tau_i\tau_k$.

To show $\frac{\partial P}{\partial g_j} = O(1)$,

$$\begin{aligned}\frac{\partial P}{\partial g_j} &= \frac{\partial}{\partial g_j}\Phi\left(-\frac{\mu(\mathbf{g}, \tilde{\alpha})}{\sigma(\mathbf{g}, \tilde{\alpha})}\right) \\ &= \phi(-\mu/\sigma)\left(-\frac{\partial \mu}{\partial g_j}\sigma^{-1} + (-\mu)(-\sigma^{-2})\frac{\partial \sigma}{\partial g_j}\right) \\ &= \overbrace{\phi(-\mu/\sigma)}^{=O(1)}\left(\frac{\mu}{\sigma^2}\frac{\partial \sigma}{\partial g_j} - \frac{1}{\sigma}\frac{\partial \mu}{\partial g_j}\right).\end{aligned}$$

Now we must show that σ is uniformly bounded away from zero and that the other terms are uniformly bounded (for large enough n, under Condition $\star$).

For σ being uniformly bounded away from zero,

$$\sigma^2 = \sum_{i=1}^{J}\sum_{j=1}^{J} \psi_i\psi_j g_i g_j(\min\{\tau_i, \tau_j\} - \tau_i\tau_j), \tag{C.39}$$

which is bounded away from zero as long as the g_j are, which they are under Condition $\star$ for large enough n.

For $\mu = O(1)$ uniformly,

$$\begin{aligned}\mu &= \sum_{j=1}^{J} \psi_j g_j \overbrace{\left((\mathbb{1}\{\psi_j < 0\} - \mathbb{1}\{\psi_j > 0\})\Phi^{-1}(1-\tilde{\alpha})\sqrt{\tau_j(1-\tau_j)} + O(n^{-1/2})\right)}^{\text{by (C.38)}} \\ &= -\Phi^{-1}(1-\tilde{\alpha})\sum_{j=1}^{J}|\psi_j| g_j\sqrt{\tau_j(1-\tau_j)} + O(n^{-1/2}),\end{aligned} \tag{C.40}$$

which is bounded as long as $\tilde{\alpha}$ is bounded away from zero and the g_j are bounded. By A2.2, the true $\gamma_j = Q'(\tau_j)$ are indeed bounded away from zero, and under Condition $\star$, the estimated $\widehat{Q'}(\tau_j) \to Q'(\tau_j)$ and thus are also bounded away from zero for large enough n (and since there are a finite number J of these). So, under Condition $\star$, μ is uniformly bounded for large enough n.

For $\frac{\partial \sigma}{\partial g_j} = O(1)$ uniformly,

$$\begin{aligned}\frac{\partial\sqrt{\sigma^2}}{\partial g_j} &= \frac{1}{2\sigma}\frac{\partial \sigma^2}{\partial g_j} \\ &= \frac{1}{2\sigma}\left(\sum_{i\neq j}\psi_i\psi_j g_i(\min\{\tau_i, \tau_j\} - \tau_i\tau_j) + 2\psi_j^2 g_j\tau_j(1-\tau_j)\right) \\ &= O(1)\end{aligned}$$

uniformly since σ is uniformly bounded away from zero (per the above argument) and the g_j are uniformly bounded under Condition $\star$ for large enough n.

For $\frac{\partial\mu}{\partial g_j} = O(1)$ uniformly,

$$\frac{\partial\mu}{\partial g_j} = |\psi_j|\Phi^{-1}(1-\tilde{\alpha})\sqrt{\tau_j(1-\tau_j)} = O(1)$$

uniformly since $\tilde{\alpha} \geq \alpha$, so $\Phi^{-1}(1-\tilde{\alpha}) \leq \Phi^{-1}(1-\alpha) < \infty$.

To show that $\frac{\partial P}{\partial\tilde{\alpha}}$ is (uniformly) bounded away from zero,

$$\begin{aligned}\frac{\partial P}{\partial\tilde{\alpha}} &= \frac{\partial}{\partial\tilde{\alpha}}\Phi\left(-\frac{\mu(\mathbf{g},\tilde{\alpha})}{\sigma(\mathbf{g},\tilde{\alpha})}\right) \\ &= \phi(-\mu/\sigma)\left(-\frac{\partial\mu}{\partial\tilde{\alpha}}\sigma^{-1} + (-\mu)(-\sigma^{-2})\overbrace{\frac{\partial\sigma}{\partial\tilde{\alpha}}}^{=0}\right).\end{aligned}$$

From arguments above, μ is uniformly bounded, and σ is uniformly bounded away from zero, so $-\mu/\sigma$ is uniformly bounded; hence, $\phi(-\mu/\sigma)$ is uniformly bounded away from zero. Also, from (C.39), σ is uniformly bounded as long as the g_j are, which they are for large enough n under Condition $\star$. Finally, using (C.40),

$$\frac{\partial\mu}{\partial\tilde{\alpha}} = \frac{1}{\phi(\Phi^{-1}(1-\tilde{\alpha}))}\sum_{j=1}^{J}|\psi_j|g_j\sqrt{\tau_j(1-\tau_j)} + O(n^{-1/2}).$$

The sum is uniformly bounded away from zero for large enough n under Condition $\star$ since the g_j are. The leading coefficient is, too, since $\phi(\cdot)$ is uniformly (over its argument in $\mathbb{R}$) bounded.

Using mean value expansions where $\tilde{a}$ is between $\tilde{\alpha}(\widehat{Q'}(\boldsymbol{\tau}))$ and $\tilde{\alpha}(Q'(\boldsymbol{\tau}))$, and each element of $\tilde{Q}'(\boldsymbol{\tau})$ is between the corresponding elements of the true and estimated vectors, using (C.37),

$$\begin{aligned}\hat{u}_j^h - u_{0,j}^h &\equiv u_j^h(\tilde{\alpha}(\widehat{Q'}(\boldsymbol{\tau}))) - u_j^h(\tilde{\alpha}(Q'(\boldsymbol{\tau}))) \\ &= (\tilde{\alpha}(\widehat{Q'}(\boldsymbol{\tau})) - \tilde{\alpha}(Q'(\boldsymbol{\tau})))u_j^{h\prime}(\tilde{a}) \\ &= \overbrace{(\widehat{Q'}(\boldsymbol{\tau}) - Q'(\boldsymbol{\tau}))'}^{=O(m^{-1/2}\log(n)+m/n)\text{ by (C.36)}}\ \overbrace{\tilde{\alpha}'(\tilde{Q}'(\boldsymbol{\tau}))}^{=O(1)\text{ from above}}\ \overbrace{u_j^{h\prime}(\tilde{a})}^{=O(n^{-1/2})\text{ by (C.37)}} \\ &= O(m^{-1/2}n^{-1/2}\log(n) + mn^{-3/2}),\end{aligned} \tag{C.41}$$

$$Q(\hat{u}_j^h) - Q(u_{0,j}^h) = (\hat{u}_j^h - u_{0,j}^h)Q'(\tilde{u}) = O(m^{-1/2}n^{-1/2}\log(n) + mn^{-3/2}), \tag{C.42}$$

where $\tilde{u}_j$ is between $\hat{u}_j^h$ and $u_{0,j}^h$ and thus $\tilde{u}_j \to \tau_j$, so for large enough n, $Q'(\tilde{u})$ is uniformly bounded by A2.2; and similarly with $\hat{u}_j^l$ and $u_{0,j}^l$.

Returning to (C.6), the term of ultimate interest can be decomposed into

$$\begin{aligned}E_h &= E_{\hat{\boldsymbol{\gamma}}}[\mathrm{P}(\mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\hat{\mathbf{u}}^H} > \sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau}) - Q(\hat{\mathbf{u}}^H)) \mid \hat{\boldsymbol{\gamma}}) - \mathrm{P}(\mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\mathbf{u}_0^H} > \sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau}) - Q(\mathbf{u}_0^H)) \mid \hat{\boldsymbol{\gamma}})] \\ &= E_{\hat{\boldsymbol{\gamma}}}[\mathrm{P}(\mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\mathbf{u}_0^H} < \sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau}) - Q(\mathbf{u}_0^H)) \mid \hat{\boldsymbol{\gamma}}) - \mathrm{P}(\mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\hat{\mathbf{u}}^H} < \sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau}) - Q(\hat{\mathbf{u}}^H)) \mid \hat{\boldsymbol{\gamma}})] \\ &= \overbrace{E_{\hat{\boldsymbol{\gamma}}}[\mathrm{P}(\mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\hat{\mathbf{u}}^H} < \sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau}) - Q(\mathbf{u}_0^H)) \mid \hat{\boldsymbol{\gamma}}) - \mathrm{P}(\mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\hat{\mathbf{u}}^H} < \sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau}) - Q(\hat{\mathbf{u}}^H)) \mid \hat{\boldsymbol{\gamma}})]}^{E_h^1}\end{aligned}$$

$$+\overbrace{E_{\hat{\gamma}}[\mathrm{P}(\mathbb{W}^{\mathbf{u}_0^H}_{\mathbf{C},\mathbf{\Lambda}} < \sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau}) - Q(\mathbf{u}_0^H)) \mid \hat{\gamma}) - \mathrm{P}(\mathbb{W}^{\hat{\mathbf{u}}^H}_{\mathbf{C},\mathbf{\Lambda}} < \sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau}) - Q(\mathbf{u}_0^H)) \mid \hat{\gamma})]}^{E_h^2}.$$

For E_h^1, GK Lemma 8(ii) is helpful: uniformly over any $\mathbf{u} = \boldsymbol{\tau} + o(1)$, which includes all possible $\hat{\mathbf{u}}^H = \boldsymbol{\tau} + O(n^{-1/2})$, the PDF of $\mathbb{W}^{\hat{\mathbf{u}}^H}_{\mathbf{C},\mathbf{\Lambda}}$ is approximately that of a mean-zero normal distribution with variance $\mathcal{V}^{\hat{\mathbf{u}}^H}_{\boldsymbol{\psi}}$:

$$\begin{aligned}
f_{\mathbb{W}^{\hat{\mathbf{u}}^H}_{\mathbf{C},\mathbf{\Lambda}}}(w) &= \phi_{\mathcal{V}^{\hat{\mathbf{u}}^H}_{\boldsymbol{\psi}}}(w)(1 + O(n^{-1/2}(\log(n))^3)),\\
\mathcal{V}^{\hat{\mathbf{u}}^H}_{\boldsymbol{\psi}} &= \sum_{i=1}^{J}\sum_{j=1}^{J}\psi_i\psi_j\frac{\min\{\hat{u}_i^H, \hat{u}_j^H\} - \hat{u}_i^H\hat{u}_j^H}{f(Q(\hat{u}_i^H))f(Q(\hat{u}_j^H))}\\
&= \sum_{i=1}^{J}\sum_{j=1}^{J}\psi_i\psi_j\Bigg(\frac{\min\{u_{0,i}^H, u_{0,j}^H\} - u_{0,i}^H u_{0,j}^H + \overbrace{O(m^{-1/2}n^{-1/2}\log(n) + mn^{-3/2})}^{\text{(C.41)}}}{f(Q(u_{0,i}^H))f(Q(u_{0,j}^H))}\\
&\qquad + \overbrace{O(m^{-1/2}n^{-1/2}\log(n) + mn^{-3/2})}^{\text{(C.41), A2.2}}\Bigg)\\
&= \overbrace{\sum_{i=1}^{J}\sum_{j=1}^{J}\psi_i\psi_j\frac{\min\{u_{0,i}^H, u_{0,j}^H\} - u_{0,i}^H u_{0,j}^H}{f(Q(u_{0,i}^H))f(Q(u_{0,j}^H))}}^{\equiv \mathcal{V}^{0,H}_{\boldsymbol{\psi}}} + O(m^{-1/2}n^{-1/2}\log(n) + mn^{-3/2})\\
&= \mathcal{V}^{0,H}_{\boldsymbol{\psi}} + O(m^{-1/2}n^{-1/2}\log(n) + mn^{-3/2}),
\end{aligned}$$

using (for the denominator) the mean value expansion $Q'(\hat{u}_j^H) = Q'(u_{0,j}^H) + Q''(\tilde{u}_j)(\hat{u}_j^H - u_{0,j}^H)$, where $\tilde{u}_j$ is between $\hat{u}_j^H$ and $u_{0,j}^H$ and thus converges to τ_j, so for large enough n, A2.2 ensures $Q''(\tilde{u}_j)$ is uniformly $O(1)$. In terms of the standard normal PDF $\phi(\cdot)$, a mean-zero normal PDF with variance σ^2 can be written $\sigma^{-1}\phi(\cdot/\sigma)$. Thus, in terms of the reference PDF with $\mathbf{u}_0^H$,

$$\begin{aligned}
\phi_{\mathcal{V}^{0,H}_{\boldsymbol{\psi}}}(w) &= \sigma_0^{-1}\phi(w/\sigma_0),\\
\phi_{\mathcal{V}^{\hat{\mathbf{u}}^H}_{\boldsymbol{\psi}}}(w) &= \hat{\sigma}^{-1}\phi(w/\hat{\sigma}) = (\sigma_0/\hat{\sigma})\phi_{\mathcal{V}^{0,H}_{\boldsymbol{\psi}}}(w\sigma_0/\hat{\sigma}),\\
(\sigma_0/\hat{\sigma}) &\equiv \sqrt{\frac{\mathcal{V}^{0,H}_{\boldsymbol{\psi}}}{\mathcal{V}^{\hat{\mathbf{u}}^H}_{\boldsymbol{\psi}}}} = \sqrt{\frac{\mathcal{V}^{0,H}_{\boldsymbol{\psi}}}{\mathcal{V}^{0,H}_{\boldsymbol{\psi}} + O(m^{-1/2}n^{-1/2}\log(n) + mn^{-3/2})}}\\
&= \sqrt{1 + O(m^{-1/2}n^{-1/2}\log(n) + mn^{-3/2})}\\
&= 1 + O(m^{-1/2}n^{-1/2}\log(n) + mn^{-3/2}).
\end{aligned}$$

Since the points of evaluation of interest w are uniformly $O(1)$,

$$w(1 + O(m^{-1/2}n^{-1/2}\log(n) + mn^{-3/2})) = w + O(m^{-1/2}n^{-1/2}\log(n) + mn^{-3/2}),$$

so altogether,

$$f_{\mathbb{W}^{\hat{\mathbf{u}}^H}_{\mathbf{C},\mathbf{\Lambda}}}(w) = \phi_{\mathcal{V}^{\hat{\mathbf{u}}^H}_{\boldsymbol{\psi}}}(w)(1 + O(n^{-1/2}(\log(n))^3))$$

$$
\begin{aligned}
&= (\sigma_0/\hat{\sigma})\phi_{\mathcal{V}_{\boldsymbol{\psi}}^{0,H}}(w\sigma_0/\hat{\sigma})(1+O(n^{-1/2}(\log(n))^3)) \\
&= (1+O(m^{-1/2}n^{-1/2}\log(n)+mn^{-3/2})) \\
&\quad \times \phi_{\mathcal{V}_{\boldsymbol{\psi}}^{0,H}}(w+O(m^{-1/2}n^{-1/2}\log(n)+mn^{-3/2})) \\
&\quad \times (1+O(n^{-1/2}(\log(n))^3)) \\
&= \phi_{\mathcal{V}_{\boldsymbol{\psi}}^{0,H}}(w)(1+O(n^{-1/2}(\log(n))^3))
\end{aligned} \tag{C.43}
$$

since $m\to\infty$ implies $m^{-1/2}n^{-1/2}=o(n^{-1/2})$ and $m=o(n)$ implies $mn^{-3/2}=o(n^{-1/2})$. The error inside the argument to the PDF can be pulled out multiplicatively as in (C.43) because w is $O(1)$ and the variance $\mathcal{V}_{\boldsymbol{\psi}}^{0,H}$ is fixed, so

$$
\begin{aligned}
&\exp(-(w+O(m^{-1/2}n^{-1/2}\log(n)+mn^{-3/2}))^2/(2\mathcal{V}_{\boldsymbol{\psi}}^{0,H})) \\
&= \exp(-w^2/(2\mathcal{V}_{\boldsymbol{\psi}}^{0,H})+O(m^{-1/2}n^{-1/2}\log(n)+mn^{-3/2})) \\
&= \exp(-w^2/(2\mathcal{V}_{\boldsymbol{\psi}}^{0,H}))\overbrace{\exp(O(m^{-1/2}n^{-1/2}\log(n)+mn^{-3/2}))}^{=1+O(m^{-1/2}n^{-1/2}\log(n)+mn^{-3/2})}.
\end{aligned} \tag{C.44}
$$

(Alternatively, one can use the MVT and the fact that the normal PDF derivative is proportional to the PDF.) So, the PDF evaluated over a set of uniformly bounded values is uniformly approximated over all possible $\hat{\mathbf{u}}^H$ by that of a mean-zero normal with variance $\mathcal{V}_{\boldsymbol{\psi}}^{0,H}$, up to a multiplicative error.

Given a value of $\hat{\mathbf{u}}^H$, the mean value theorem gives

$$
\begin{aligned}
E_h^1(\hat{\mathbf{u}}^H) &= \int_{\sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau})-Q(\hat{\mathbf{u}}^H))}^{\sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau})-Q(\mathbf{u}_0^H))} f_{\mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\hat{\mathbf{u}}^H}}(w)\,dw \\
&= (\sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau})-Q(\mathbf{u}_0^H))-\sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau})-Q(\hat{\mathbf{u}}^H)))f_{\mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\hat{\mathbf{u}}^H}}(\tilde{w}) \\
&= \sqrt{n}\boldsymbol{\psi}'(Q(\hat{\mathbf{u}}^H)-Q(\mathbf{u}_0^H))\overbrace{\phi_{\mathcal{V}_{\boldsymbol{\psi}}^{0,H}}(\tilde{w})(1+O(n^{-1/2}(\log(n))^3))}^{\text{by (C.43)}} \\
&= \sqrt{n}\boldsymbol{\psi}'\overbrace{(\widehat{Q'}(\boldsymbol{\tau})-Q'(\boldsymbol{\tau}))O(1)O(n^{-1/2})O(1)}^{\text{by (C.41) and (C.42)}}\phi_{\mathcal{V}_{\boldsymbol{\psi}}^{0,H}}(\tilde{w})(1+O(n^{-1/2}(\log(n))^3)) \\
&= \boldsymbol{\psi}'(\widehat{Q'}(\boldsymbol{\tau})-Q'(\boldsymbol{\tau}))\phi_{\mathcal{V}_{\boldsymbol{\psi}}^{0,H}}(\tilde{w})(1+O(n^{-1/2}(\log(n))^3))
\end{aligned} \tag{C.45}
$$

where $\tilde{w}$ is between the limits of integration and thus $O(1)$. Actually, $\tilde{w}$ is pinned down much more precisely: using (C.42),

$$
\begin{aligned}
\tilde{w} &= \sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau})-Q(\mathbf{u}_0^H))+O(\sqrt{n}\boldsymbol{\psi}'\overbrace{(Q(\hat{\mathbf{u}}^H)-Q(\mathbf{u}_0^H))}^{=O(m^{-1/2}n^{-1/2}\log(n)+mn^{-3/2})\text{ by (C.42)}}) \\
&= \overbrace{\sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau})-Q(\mathbf{u}_0^H))}^{=O(1)}+O(m^{-1/2}\log(n)+mn^{-1}).
\end{aligned}
$$

By the same arguments leading to (C.44), this error becomes multiplicative when pulling it out of the normal PDF:

$$
\phi_{\mathcal{V}_{\boldsymbol{\psi}}^{0,H}}(\tilde{w}) = \phi_{\mathcal{V}_{\boldsymbol{\psi}}^{0,H}}(\sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau})-Q(\mathbf{u}_0^H))+O(m^{-1/2}\log(n)+mn^{-1}))
$$

$$= \phi_{\mathcal{V}_{\boldsymbol{\psi}}^{0,H}}(\sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau}) - Q(\mathbf{u}_0^H)))(1 + O(m^{-1/2}\log(n) + mn^{-1})).$$

So, up to a uniform multiplicative error, the only random variables left in E_h^1 are the sparsity estimators, $\widehat{Q'}(\tau_j)$. Altogether,

$$\begin{aligned}
E_h^1 &= E[\boldsymbol{\psi}'(\widehat{Q'}(\boldsymbol{\tau}) - Q'(\boldsymbol{\tau}))\phi_{\mathcal{V}_{\boldsymbol{\psi}}^{0,H}}(\sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau}) - Q(\mathbf{u}_0^H))) \\
&\qquad \times \overbrace{(1 + O(m^{-1/2}\log(n) + mn^{-1}))(1 + O(n^{-1/2}(\log(n))^3))}^{\text{uniform over } \hat{\mathbf{u}}^H}] \\
&= \overbrace{\boldsymbol{\psi}'}^{=O(1)} E[(1 + O(m^{-1/2}\log(n) + mn^{-1}))(\widehat{Q'}(\boldsymbol{\tau}) - Q'(\boldsymbol{\tau}))]\overbrace{\phi_{\mathcal{V}_{\boldsymbol{\psi}}^{0,H}}(\sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau}) - Q(\mathbf{u}_0^H)))}^{=O(1)} \\
&= O(E[AB]), \qquad\qquad \text{(C.46)} \\
A &= 1 + O(m^{-1/2}\log(n) + mn^{-1}), \\
B &= \max_j\{\widehat{Q'}(\tau_j) - Q'(\tau_j)\}.
\end{aligned}$$

Now,

$$\begin{aligned}
E[AB] &= \mathrm{Cov}(A, B) + E[A]E[B], \\
|\mathrm{Cov}(A, B)| &= |\mathrm{Corr}(A, B)\sqrt{\mathrm{Var}(A)\,\mathrm{Var}(B)}| \le \sqrt{\mathrm{Var}(A)\,\mathrm{Var}(B)}.
\end{aligned}$$

From equations (2.5) and (2.6) in Bloch and Gastwirth (1968, p. 1084),

$$E[B] = O(m^2/n^2), \quad \mathrm{Var}(B) = O(m^{-1}).$$

Also,

$$\begin{aligned}
E[A] &= E[1 + O(m^{-1/2}\log(n) + mn^{-1})] = O(1), \\
\mathrm{Var}(A) &\le (1 + O(m^{-1/2}\log(n) + mn^{-1}) - (1 + O(m^{-1/2}\log(n) + mn^{-1})))^2 \\
&= O(m^{-1}(\log(n))^2 + m^2n^{-2}), \\
|E[AB]| &\le \sqrt{O(m^{-1}(\log(n))^2 + m^2n^{-2})O(m^{-1})} + O(1)O(m^2/n^2) \\
&= O(m^{-1}\log(n) + m^2/n^2).
\end{aligned}$$

Using $m \asymp n^{2/3}$ attains the (nearly) minimum rate of $O(n^{-2/3}\log(n))$. This can be slightly improved to $O(n^{-2/3}(\log(n))^{2/3})$ with $m \asymp n^{2/3}(\log(n))^{1/3}$, but the practical difference is negligible, so we prefer $m \asymp n^{2/3}$ for simplicity. With any $n^{1/2} \lesssim m \lesssim n^{3/4}$, the rate is no greater than $T_h = O(n^{-1/2}\log(n))$.

Finally, for E_h^2, recall that for any $\mathbf{t} = \boldsymbol{\tau} + o(1)$, $\mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\mathbf{t}}$ has (asymptotically) a mean-zero normal distribution with variance $\mathcal{V}_{\boldsymbol{\psi}}^{\mathbf{t}}$, so its CDF can be written in terms of the standard normal CDF $\Phi(\cdot)$, as $\Phi(\cdot/\sqrt{\mathcal{V}_{\boldsymbol{\psi}}^{\mathbf{t}}})$. Denote the fixed (wrt $\hat{\mathbf{u}}^H$), $O(1)$ point of evaluation $w = \sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau}) - Q(\mathbf{u}_0^H))$. Then,

$$\frac{\partial\Phi(w/\sqrt{\mathcal{V}_{\boldsymbol{\psi}}^{\mathbf{t}}})}{\partial\mathbf{t}} = \phi(w/\sqrt{\mathcal{V}_{\boldsymbol{\psi}}^{\mathbf{t}}})\frac{\partial w(\mathcal{V}_{\boldsymbol{\psi}}^{\mathbf{t}})^{-1/2}}{\partial\mathbf{t}} = \phi(w/\sqrt{\mathcal{V}_{\boldsymbol{\psi}}^{\mathbf{t}}})\frac{-w}{2(\mathcal{V}_{\boldsymbol{\psi}}^{\mathbf{t}})^{3/2}}\frac{\partial\mathcal{V}_{\boldsymbol{\psi}}^{\mathbf{t}}}{\partial\mathbf{t}} = O(1) \qquad \text{(C.47)}$$

since all three terms in the product are $O(1)$.

Altogether,

$$
\begin{aligned}
|E_h^2| &= \left|E_{\hat{\boldsymbol{\gamma}}}[\mathrm{P}(\mathbb{W}^{\mathbf{u}_0^H}_{\mathbf{C},\boldsymbol{\Lambda}} < \sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau}) - Q(\mathbf{u}_0^H)) \mid \hat{\boldsymbol{\gamma}}) - \mathrm{P}(\mathbb{W}^{\hat{\mathbf{u}}^H}_{\mathbf{C},\boldsymbol{\Lambda}} < \sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau}) - Q(\mathbf{u}_0^H)) \mid \hat{\boldsymbol{\gamma}})]\right| \\
&= \left|E_{\hat{\boldsymbol{\gamma}}}[F_{\mathbb{W}^{\mathbf{u}_0^H}_{\mathbf{C},\boldsymbol{\Lambda}}|\hat{\boldsymbol{\gamma}}}(\sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau}) - Q(\mathbf{u}_0^H)) \mid \hat{\boldsymbol{\gamma}}) - F_{\mathbb{W}^{\hat{\mathbf{u}}^H}_{\mathbf{C},\boldsymbol{\Lambda}}|\hat{\boldsymbol{\gamma}}}(\sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau}) - Q(\mathbf{u}_0^H)) \mid \hat{\boldsymbol{\gamma}})]\right| \\
&= \left|E_{\hat{\boldsymbol{\gamma}}}\left[\overbrace{(\mathbf{u}_0^H - \hat{\mathbf{u}}^H)'}^{=O(m^{-1/2}n^{-1/2}\log(n)+mn^{-3/2}) \text{ by (C.41)}} \overbrace{\left.\frac{dF_{\mathbb{W}^{\mathbf{t}}_{\mathbf{C},\boldsymbol{\Lambda}}|\hat{\boldsymbol{\gamma}}}(\sqrt{n}\boldsymbol{\psi}'(Q(\boldsymbol{\tau}) - Q(\mathbf{u}_0^H)) \mid \hat{\boldsymbol{\gamma}})}{d\mathbf{t}}\right|_{\mathbf{t}=\tilde{\mathbf{u}}}}^{=O(1) \text{ by (C.47)}}\right]\right| \\
&= O(m^{-1/2}n^{-1/2}\log(n) + mn^{-3/2}), \qquad \text{(C.48)}
\end{aligned}
$$

where $\tilde{\mathbf{u}}$ is between $\mathbf{u}_0^H$ and $\hat{\mathbf{u}}^H$. The rate in (C.48) is even smaller than necessary. For the two-sided case, we only need $E_h^2 = O(n^{-2/3})$, but with $m \asymp n^{2/3}$ we get $O(n^{-5/6}\log(n))$. For the one-sided case, we need $E_h^2 = O(n^{-1/2}\log(n))$, but over $m \in [n^{1/2}, n^{3/4}]$ the maximum of (C.48) is $O(n^{-3/4}\log(n))$.

The foregoing arguments and rates are all the same for E_l. □

C.2. Theorem for CI for difference of linear combination of quantiles (and QD)

We restate the more general theorem as a reminder of what we will prove. The object of interest is

$$D = \sum_{j=1}^{J} \psi_j(Q_Y(\tau_j) - Q_X(\tau_j)).$$

The lower one-sided CI is

$$\left(-\infty, \sum_{j=1}^{J} \psi_j(\hat{Q}_Y^L(u_{y,j}^H(\tilde{\alpha})) - \hat{Q}_X^L(u_{x,j}^L(\tilde{\alpha})))\right), \qquad \text{(C.49)}$$

where $\tilde{\alpha}$ satisfies

$$1-\alpha = \mathrm{P}\left(\sum_{j=1}^{J} \psi_j(\widehat{Q_Y'}(\tau_j)(\tilde{Q}_{U_y}^I(u_{y,j}^H(\tilde{\alpha}))-\tau_j)-\widehat{Q_X'}(\tau_j)(\tilde{Q}_{U_x}^I(u_{x,j}^L(\tilde{\alpha}))-\tau_j)) > 0\right). \quad \text{(C.50)}$$

For an upper one-sided CI, the analogues of (C.49) and (C.50) are

$$1-\alpha = \mathrm{P}\left(\sum_{j=1}^{J} \psi_j(\widehat{Q_Y'}(\tau_j)(\tilde{Q}_{U_y}^I(u_{y,j}^L(\tilde{\alpha})) - \tau_j) - \widehat{Q_X'}(\tau_j)(\tilde{Q}_{U_x}^I(u_{x,j}^H(\tilde{\alpha})) - \tau_j)) < 0\right), \qquad \text{(C.51)}$$

$$\left(\sum_{j=1}^{J} \psi_j(\hat{Q}_Y^L(u_{y,j}^L(\tilde{\alpha})) - \hat{Q}_X^L(u_{x,j}^H(\tilde{\alpha}))), \infty\right). \qquad \text{(C.52)}$$

Theorem C.2. *Let Assumptions A2.1 and A2.2 hold.*

(a) The one-sided CIs in (C.49) *and* (C.52) *both have CPE of order* $O(n^{-1/2}\log(n))$ *if all* $\widehat{Q_X'}(\tau_j)$ *and* $\widehat{Q_Y'}(\tau_j)$ *are estimated by* (3.5) *with smoothing parameters* $m_{x,j}$ *and* $m_{y,j}$ *having rates larger than* $n^{1/2}$ *and smaller than* $n^{3/4}$.

(b) Two-sided CIs, formed by the intersection of upper and lower one-sided $1-\alpha/2$ CIs, have CPE of order $O(n^{-2/3}\log(n))$ if all $\widehat{Q'_X}(\tau_j)$ and $\widehat{Q'_Y}(\tau_j)$ are estimated by (3.5) with $m_{x,j} \asymp n^{2/3}$ and $m_{y,j} \asymp n^{2/3}$.

(c) The asymptotic probabilities of excluding $D_n = \boldsymbol{\psi}'(Q_Y(\boldsymbol{\tau}) - Q_X(\boldsymbol{\tau}) + \boldsymbol{\kappa} n_y^{-1/2})$ from lower one-sided (l), upper one-sided (u), and equal-tailed two-sided (t) CIs (i.e., asymptotic power of the corresponding hypothesis tests) are

$$\mathcal{P}_n^l(D_n) \to \Phi(z_\alpha + S),\ \ \mathcal{P}_n^u(D_n) \to \Phi(z_\alpha - S),\ \ \mathcal{P}_n^t(D_n) \to \Phi(z_{\alpha/2} + S) + \Phi(z_{\alpha/2} - S),$$

where $S \equiv \boldsymbol{\psi}'\boldsymbol{\kappa}/\sqrt{\mathcal{V}_{\boldsymbol{\psi},x} + \delta^2 \mathcal{V}_{\boldsymbol{\psi},y}}$, and $\mathcal{V}_{\boldsymbol{\psi},x}$ and $\mathcal{V}_{\boldsymbol{\psi},y}$ are as defined in Theorem A.3 for the X and Y population distributions, respectively.

Proof: Let

$$\begin{aligned} \hat{\mathbf{u}}_y^H &= \{u_{y,j}^H(\tilde{\alpha}(\hat{\boldsymbol{\gamma}}_x, \hat{\boldsymbol{\gamma}}_y))\}_{j=1}^J, \quad \mathbf{u}_{0,y}^H = \{u_{y,j}^H(\tilde{\alpha}(\boldsymbol{\gamma}_x, \boldsymbol{\gamma}_y))\}_{j=1}^J, \\ \hat{\mathbf{u}}_x^H &= \{u_{x,j}^L(\tilde{\alpha}(\hat{\boldsymbol{\gamma}}_x, \hat{\boldsymbol{\gamma}}_y))\}_{j=1}^J, \quad \mathbf{u}_{0,x}^H = \{u_{x,j}^L(\tilde{\alpha}(\boldsymbol{\gamma}_x, \boldsymbol{\gamma}_y))\}_{j=1}^J. \end{aligned}$$

From A2.1, $\sqrt{n_x/n_y} = \delta + O(n^{-1})$ with $0 < \delta < \infty$. Referring back to (C.49), and with steps similar to the derivation of (C.4), true CP is

$$\begin{aligned}
&\mathrm{P}(\boldsymbol{\psi}'(\hat{Q}_Y^L(\hat{\mathbf{u}}_y^H) - \hat{Q}_X^L(\hat{\mathbf{u}}_x^H)) > \boldsymbol{\psi}'(Q_Y(\boldsymbol{\tau}) - Q_X(\boldsymbol{\tau})))\\
&= \mathrm{P}(\boldsymbol{\psi}'(\tilde{Q}_Y^I(\hat{\mathbf{u}}_y^H) - \tilde{Q}_X^I(\hat{\mathbf{u}}_x^H)) > \boldsymbol{\psi}'(Q_Y(\boldsymbol{\tau}) - Q_X(\boldsymbol{\tau}))) + \overbrace{O(n^{-1})}^{\text{by Theorem 2.1}}\\
&= \mathrm{P}\Bigg(\overbrace{\sqrt{\frac{n_x}{n_y}}}^{=\delta+O(n^{-1})}\sqrt{n_y}\boldsymbol{\psi}'(\tilde{Q}_Y^I(\hat{\mathbf{u}}_y^H) - Q_Y(\hat{\mathbf{u}}_y^H)) - \sqrt{n_x}\boldsymbol{\psi}'(\tilde{Q}_X^I(\hat{\mathbf{u}}_x^H) - Q_X(\hat{\mathbf{u}}_x^H))\\
&\qquad > \overbrace{\sqrt{\frac{n_x}{n_y}}}^{=\delta+O(n^{-1})}\sqrt{n_y}\boldsymbol{\psi}'(Q_Y(\boldsymbol{\tau}) - Q_Y(\hat{\mathbf{u}}_y^H)) - \sqrt{n_x}\boldsymbol{\psi}'(Q_X(\boldsymbol{\tau}) - Q_X(\hat{\mathbf{u}}_x^H))\Bigg) + O(n^{-1})\\
&= \mathrm{P}(\delta\mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\hat{\mathbf{u}}_y^H} - \mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\hat{\mathbf{u}}_x^H}\\
&\qquad > \delta\sqrt{n_y}\boldsymbol{\psi}'(Q_Y(\boldsymbol{\tau}) - Q_Y(\hat{\mathbf{u}}_y^H)) - \sqrt{n_x}\boldsymbol{\psi}'(Q_X(\boldsymbol{\tau}) - Q_X(\hat{\mathbf{u}}_x^H)))\\
&\quad + (\mathrm{P}(\delta\sqrt{n_y}\boldsymbol{\psi}'(\tilde{Q}_Y^I(\hat{\mathbf{u}}_y^H) - Q_Y(\hat{\mathbf{u}}_y^H)) - \sqrt{n_x}\boldsymbol{\psi}'(\tilde{Q}_X^I(\hat{\mathbf{u}}_x^H) - Q_X(\hat{\mathbf{u}}_x^H))\\
&\qquad > \delta\sqrt{n_y}\boldsymbol{\psi}'(Q_Y(\boldsymbol{\tau}) - Q_Y(\hat{\mathbf{u}}_y^H)) - \sqrt{n_x}\boldsymbol{\psi}'(Q_X(\boldsymbol{\tau}) - Q_X(\hat{\mathbf{u}}_x^H)))\\
&\quad - \mathrm{P}(\delta\mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\hat{\mathbf{u}}_y^H} - \mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\hat{\mathbf{u}}_x^H}\\
&\qquad > \delta\sqrt{n_y}\boldsymbol{\psi}'(Q_Y(\boldsymbol{\tau}) - Q_Y(\hat{\mathbf{u}}_y^H)) - \sqrt{n_x}\boldsymbol{\psi}'(Q_X(\boldsymbol{\tau}) - Q_X(\hat{\mathbf{u}}_x^H))))\\
&\quad + O(n^{-1})\\
&= \mathrm{P}(\delta\mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\hat{\mathbf{u}}_y^H} - \mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\hat{\mathbf{u}}_x^H}\\
&\qquad > \delta\sqrt{n_y}\boldsymbol{\psi}'(Q_Y(\boldsymbol{\tau}) - Q_Y(\hat{\mathbf{u}}_y^H)) - \sqrt{n_x}\boldsymbol{\psi}'(Q_X(\boldsymbol{\tau}) - Q_X(\hat{\mathbf{u}}_x^H)))\\
&\quad + \overbrace{O(n^{-3/2}(\log(n))^3)}^{\text{by (A.13) and independence (A2.1)}} + O(n^{-1})\\
&= \mathrm{P}(\delta\mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\hat{\mathbf{u}}_y^H} - \mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\hat{\mathbf{u}}_x^H}
\end{aligned}$$

$$
\begin{aligned}
&\quad > \delta\sqrt{n_y}\boldsymbol{\psi}'(Q_Y(\boldsymbol{\tau}) - Q_Y(\hat{\mathbf{u}}_y^H)) - \sqrt{n_x}\boldsymbol{\psi}'(Q_X(\boldsymbol{\tau}) - Q_X(\hat{\mathbf{u}}_x^H)))\\
&+ O(n^{-1})\\
&= \mathrm{P}(\delta\mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\mathbf{u}_{0,y}^H} - \mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\mathbf{u}_{0,x}^H}\\
&\qquad > \delta\sqrt{n_y}\boldsymbol{\psi}'(Q_Y(\boldsymbol{\tau}) - Q_Y(\mathbf{u}_{0,y}^H)) - \sqrt{n_x}\boldsymbol{\psi}'(Q_X(\boldsymbol{\tau}) - Q_X(\mathbf{u}_{0,x}^H)))\\
&+ E_{h,2} + O(n^{-1})\\
&= \overbrace{\mathrm{P}\left(\sum_{j=1}^{J}\psi_j(Q_Y'(\tau_j)(\tilde{Q}_{U_y}^I(u_{0,y,j}^H) - \tau_j) - Q_X'(\tau_j)(\tilde{Q}_{U_x}^I(u_{0,x,j}^H) - \tau_j)) > 0\right)}^{=1-\alpha \text{ by (C.50)}}\\
&+ T_{h,2} + E_{h,2} + O(n^{-1}), \qquad\qquad (C.53)\\
E_{h,2} &\equiv \mathrm{P}(\delta\mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\hat{\mathbf{u}}_y^H} - \mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\hat{\mathbf{u}}_x^H}\\
&\quad > \delta\sqrt{n_y}\boldsymbol{\psi}'(Q_Y(\boldsymbol{\tau}) - Q_Y(\hat{\mathbf{u}}_y^H)) - \sqrt{n_x}\boldsymbol{\psi}'(Q_X(\boldsymbol{\tau}) - Q_X(\hat{\mathbf{u}}_x^H)))\\
&- \mathrm{P}(\delta\mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\mathbf{u}_{0,y}^H} - \mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\mathbf{u}_{0,x}^H}\\
&\qquad > \delta\sqrt{n_y}\boldsymbol{\psi}'(Q_Y(\boldsymbol{\tau}) - Q_Y(\mathbf{u}_{0,y}^H)) - \sqrt{n_x}\boldsymbol{\psi}'(Q_X(\boldsymbol{\tau}) - Q_X(\mathbf{u}_{0,x}^H))),\\
T_{h,2} &\equiv \mathrm{P}(\delta\mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\mathbf{u}_{0,y}^H} - \mathbb{W}_{\mathbf{C},\boldsymbol{\Lambda}}^{\mathbf{u}_{0,x}^H}\\
&\qquad > \delta\sqrt{n_y}\boldsymbol{\psi}'(Q_Y(\boldsymbol{\tau}) - Q_Y(\mathbf{u}_{0,y}^H)) - \sqrt{n_x}\boldsymbol{\psi}'(Q_X(\boldsymbol{\tau}) - Q_X(\mathbf{u}_{0,x}^H)))\\
&- \mathrm{P}\left(\sum_{j=1}^{J}\psi_j(Q_Y'(\tau_j)(\tilde{Q}_{U_y}^I(u_{0,y,j}^H) - \tau_j) - Q_X'(\tau_j)(\tilde{Q}_{U_x}^I(u_{0,x,j}^H) - \tau_j)) > 0\right)\\
&= \mathrm{P}(\delta\sqrt{n_y}\boldsymbol{\psi}'(Q_Y(\tilde{Q}_{U_y}^I(\mathbf{u}_{0,y}^H)) - Q_Y(\mathbf{u}_{0,y}^H)) - \sqrt{n_x}\boldsymbol{\psi}'(Q_X(\tilde{Q}_{U_x}^I(\mathbf{u}_{0,x}^H)) - Q_X(\mathbf{u}_{0,x}^H))\\
&\quad > \delta\sqrt{n_y}\boldsymbol{\psi}'(Q_Y(\boldsymbol{\tau}) - Q_Y(\mathbf{u}_{0,y}^H)) - \sqrt{n_x}\boldsymbol{\psi}'(Q_X(\boldsymbol{\tau}) - Q_X(\mathbf{u}_{0,x}^H)))\\
&+ \qquad \overbrace{O(n^{-3/2}(\log(n))^3)}^{\text{by (A.13) and independence (A2.1)}}\\
&- \mathrm{P}\left(\sum_{j=1}^{J}\psi_j(Q_Y'(\tau_j)(\tilde{Q}_{U_y}^I(u_{0,y,j}^H) - \tau_j) - Q_X'(\tau_j)(\tilde{Q}_{U_x}^I(u_{0,x,j}^H) - \tau_j)) > 0\right)\\
&= \mathrm{P}(\delta\sqrt{n_y}\boldsymbol{\psi}'(Q_Y(\tilde{Q}_{U_y}^I(\mathbf{u}_{0,y}^H)) - Q_Y(\boldsymbol{\tau})) - \sqrt{n_x}\boldsymbol{\psi}'(Q_X(\tilde{Q}_{U_x}^I(\mathbf{u}_{0,x}^H)) - Q_X(\boldsymbol{\tau}))\\
&\quad > 0)\\
&- \mathrm{P}\left(\sum_{j=1}^{J}\psi_j(Q_Y'(\tau_j)(\tilde{Q}_{U_y}^I(u_{0,y,j}^H) - \tau_j) - Q_X'(\tau_j)(\tilde{Q}_{U_x}^I(u_{0,x,j}^H) - \tau_j)) > 0\right)\\
&+ O(n^{-3/2}(\log(n))^3)\\
&= \mathrm{P}\left(\sum_{j=1}^{J}\psi_j((Q_Y(\tilde{Q}_{U_y}^I(u_{0,y,j}^H)) - Q_Y(\tau_j)) - (Q_X(\tilde{Q}_{U_x}^I(u_{0,x,j}^H)) - Q_X(\tau_j))) > 0\right)\\
&- \mathrm{P}\left(\sum_{j=1}^{J}\psi_j(Q_Y'(\tau_j)(\tilde{Q}_{U_y}^I(u_{0,y,j}^H) - \tau_j) - Q_X'(\tau_j)(\tilde{Q}_{U_x}^I(u_{0,x,j}^H) - \tau_j)) > 0\right)
\end{aligned}
$$

$$+ O(n^{-3/2}(\log(n))^3 + n^{-1}),$$

where the final n^{-1} in the remainder could be avoided by rewriting everything with $\sqrt{n_x/n_y}$ instead of introducing δ, but it is smaller-order anyway. Similar to before, the term $T_{h,2}$ is the CPE induced by the linearisations of the quantile functions, and $E_{h,2}$ is CPE induced by estimation error of the nuisance parameters. The upper one-sided derivation yields similar terms, denoted $T_{l,2}$ and $E_{l,2}$. The derivations of the orders of magnitude of $T_{h,2}$ and $E_{h,2}$ parallel those of T_h and E_h in the proof of Theorem A.3.

The proof of part (a) follows by applying Lemmas C.3 and C.4, which respectively have $T_{h,2} = O(n^{-1/2}\log(n))$ and $E_{h,2} = O(m^{-1}\log(n) + (m/n)^2)$ for common smoothing parameter rate m and common sample size rate n (as in A2.1), and similarly for $T_{l,2}$ and $E_{l,2}$. As long as $n^{1/2} \lesssim m \lesssim n^{3/4}$, the dominant CPE term is order $O(n^{-1/2}\log(n))$.

The proof of part (b) also follows by applying Lemmas A.4 and A.5, which additionally give $T_{h,2} + T_{l,2} = O(n^{-1}\log(n))$. Thus, CPE is $O(n^{-1}\log(n)) + O(m^{-1}\log(n) + (m/n)^2)$. Now, the second term dominates, and it is minimised by $m \asymp n^{2/3}$, leaving CPE of order $O(n^{-2/3}\log(n))$.

The proof of part (c) remains. It parallels the proof of Theorem A.3(c). The addition of Y variables provides little complication since the samples are assumed independent, so the asymptotic normal distributions from GK Lemma 8 are independent, which implies their sum is normal with variance equal to the sum of variances. Also, the sample size ratio δ^2 must be incorporated. Otherwise, the steps are identical.

One-sided power against

$$H_0 : D_n = \boldsymbol{\psi}'(Q_Y(\boldsymbol{\tau}) - Q_X(\boldsymbol{\tau}) + \boldsymbol{\kappa} n^{-1/2})$$

with $\boldsymbol{\psi}'\boldsymbol{\kappa} > 0$ is the probability that D_n is not contained in the lower one-sided CI. Below, $\tilde{u}_{y,j}$ comes from the mean value theorem and lies between τ_j and $u^H_{y,j}$. Since $u^H_{y,j} \to \tau_j$ by Lemma C.1, $\tilde{u}_{y,j} \to \tau_j$, so for large enough n, all $\tilde{u}_{y,j}$ lie within an arbitrarily small neighbourhood of τ_j and thus A2.2 uniformly bounds $Q''_Y(\tilde{u}_{y,j}) = O(1)$. The same argument applies to $Q''_X(\tilde{u}_{x,j}) = O(1)$. The CI exclusion probability is

$$\begin{aligned}
&\mathcal{P}^l_n(D_n) \\
&= \mathrm{P}\left(\sum_{j=1}^J \psi_j(\hat{Q}^L_Y(u^H_{y,j}(\tilde{\alpha}_j)) - Q_Y(\tau_j) - (\hat{Q}^L_X(u^L_{x,j}(\tilde{\alpha}_j)) - Q_X(\tau_j))) < n^{-1/2}\boldsymbol{\psi}'\boldsymbol{\kappa}\right) \\
&= \mathrm{P}(\boldsymbol{\psi}'(\hat{Q}^L_Y(\mathbf{u}^H_y(\tilde{\boldsymbol{\alpha}})) - Q_Y(\mathbf{u}^H_y(\tilde{\boldsymbol{\alpha}}))) - \boldsymbol{\psi}'(\hat{Q}^L_X(\mathbf{u}^L_x(\tilde{\boldsymbol{\alpha}})) - Q_X(\mathbf{u}^L_x(\tilde{\boldsymbol{\alpha}}))) \\
&\qquad < n^{-1/2}\boldsymbol{\psi}'\boldsymbol{\kappa} - \boldsymbol{\psi}'(Q_Y(\mathbf{u}^H_y(\tilde{\boldsymbol{\alpha}})) - Q_Y(\boldsymbol{\tau})) - \boldsymbol{\psi}'(Q_X(\mathbf{u}^L_x(\tilde{\boldsymbol{\alpha}})) - Q_X(\boldsymbol{\tau}))) \\
&= \mathrm{P}(\delta\sqrt{n_y}\boldsymbol{\psi}'(\hat{Q}^L_Y(\mathbf{u}^H_y(\tilde{\boldsymbol{\alpha}})) - Q_Y(\mathbf{u}^H_y(\tilde{\boldsymbol{\alpha}}))) - \sqrt{n_x}\boldsymbol{\psi}'(\hat{Q}^L_X(\mathbf{u}^L_x(\tilde{\boldsymbol{\alpha}})) - Q_X(\mathbf{u}^L_x(\tilde{\boldsymbol{\alpha}}))) \\
&\qquad < \boldsymbol{\psi}'\boldsymbol{\kappa} - \delta\sqrt{n_y}\sum_{j=1}^J \psi_j(Q'_Y(\tau_j)(u^H_{y,j} - \tau_j) + (1/2)\overbrace{Q''_Y(\tilde{u}_j)}^{=O(1)}\overbrace{(u^H_{y,j} - \tau_j)^2}^{=O(n^{-1})\text{ by Lemma C.1}}) \\
&\qquad - \sqrt{n_x}\sum_{j=1}^J \psi_j(Q'_X(\tau_j)(u^L_{x,j} - \tau_j) + (1/2)\overbrace{Q''_X(\tilde{u}_{x,j})}^{=O(1)}\overbrace{(u^L_{x,j} - \tau_j)^2}^{=O(n^{-1})\text{ by Lemma C.1}}))
\end{aligned}$$

$$
= \Phi\Biggl(\frac{ \overbrace{ \sum_{j=1}^{J} \psi_j \kappa_j - \overbrace{\psi_j Q'_Y(\tau_j)\delta}^{=O(1)} \overbrace{\sqrt{n_y}(u^H_{y,j}(\tilde{\alpha}_j) - \tau_j)}^{\text{apply Lemma C.1}} - \overbrace{\psi_j Q'_X(\tau_j)}^{=O(1)} \overbrace{\sqrt{n_x}(u^L_{x,j}(\tilde{\alpha}_j) - \tau_j)}^{\text{apply Lemma C.1}} }^{\text{by GK Lemma 8 and independence (A2.1)}} + O(n^{-1/2}) }{ \sqrt{\hat{\mathcal{V}}_{\boldsymbol{\psi},x} + \delta^2 \hat{\mathcal{V}}_{\boldsymbol{\psi},y}} } \Biggr)
$$

$$
+ O(n^{-1/2}(\log(n))^3)
$$

$$
= \Phi\Biggl(\frac{\boldsymbol{\psi}'\boldsymbol{\kappa}}{\sqrt{\mathcal{V}_{\boldsymbol{\psi},\mathbf{x}} + \delta^2 \mathcal{V}_{\boldsymbol{\psi},y}}}
$$

$$
- \overbrace{ \frac{1}{\sqrt{\mathcal{V}_{\boldsymbol{\psi},\mathbf{x}} + \delta^2 \mathcal{V}_{\boldsymbol{\psi},y}}} \sum_{j=1}^{J} \psi_j z_{1-\tilde{\alpha}_j} \sqrt{\tau_j(1-\tau_j)}(\delta Q'_Y(\tau_j) + Q'_X(\tau_j)) }^{\to z_{1-\alpha} \text{ to control size when } \boldsymbol{\kappa}=\mathbf{0}} + O(n^{-1/2}) \Biggr)
$$

$$
+ O(n^{-1/2}(\log(n))^3)
$$

$$
\to \Phi\Biggl(\frac{\boldsymbol{\psi}'\boldsymbol{\kappa}}{\sqrt{\mathcal{V}_{\boldsymbol{\psi},\mathbf{x}} + \delta^2 \mathcal{V}_{\boldsymbol{\psi},y}}} - z_{1-\alpha} \Biggr) = \Phi\Biggl(\frac{\boldsymbol{\psi}'\boldsymbol{\kappa}}{\sqrt{\mathcal{V}_{\boldsymbol{\psi},\mathbf{x}} + \delta^2 \mathcal{V}_{\boldsymbol{\psi},y}}} + z_{\alpha} \Biggr),
$$

where

$$
\hat{\mathcal{V}}_{\boldsymbol{\psi},y} \equiv \sum_{i=1}^{J} \sum_{j=1}^{J} \psi_i \psi_j \frac{\min\{u^H_{y,i}(\tilde{\alpha}_i), u^H_{y,j}(\tilde{\alpha}_j)\} - u^H_{y,i}(\tilde{\alpha}_i) u^H_{y,j}(\tilde{\alpha}_j)}{f_Y(Q_Y(u^H_{y,i}(\tilde{\alpha}_i))) f_Y(Q_Y(u^H_{y,j}(\tilde{\alpha}_j)))} \to \mathcal{V}_{\boldsymbol{\psi},y},
$$

$$
\hat{\mathcal{V}}_{\boldsymbol{\psi},x} \equiv \sum_{i=1}^{J} \sum_{j=1}^{J} \psi_i \psi_j \frac{\min\{u^L_{x,i}(\tilde{\alpha}_i), u^L_{x,j}(\tilde{\alpha}_j)\} - u^L_{x,i}(\tilde{\alpha}_i) u^L_{x,j}(\tilde{\alpha}_j)}{f_X(Q_X(u^L_{x,i}(\tilde{\alpha}_i))) f_X(Q_X(u^L_{x,j}(\tilde{\alpha}_j)))} \to \mathcal{V}_{\boldsymbol{\psi},x},
$$

$$
\sqrt{\hat{\mathcal{V}}_{\boldsymbol{\psi},x} + \delta^2 \hat{\mathcal{V}}_{\boldsymbol{\psi},y}} \to \sqrt{\mathcal{V}_{\boldsymbol{\psi},\mathbf{x}} + \delta^2 \mathcal{V}_{\boldsymbol{\psi},y}},
$$

applying the continuous mapping theorem in the final line.

The upper one-sided case follows similarly.

For the two-sided case, since the two-sided CI is the intersection of the upper and lower one-sided $1-\alpha/2$ CIs, the exclusion probability is

$$
\begin{aligned}
\mathcal{P}^t_n(D_n, \alpha) &= \mathcal{P}^u_n(D_n, \alpha/2) + \mathcal{P}^l_n(D_n, \alpha/2) \\
&\to \Phi\Biggl(z_{\alpha/2} + \frac{\boldsymbol{\psi}'\boldsymbol{\kappa}}{\sqrt{\mathcal{V}_{\boldsymbol{\psi},\mathbf{x}} + \delta^2 \mathcal{V}_{\boldsymbol{\psi},y}}} \Biggr) + \Phi\Biggl(z_{\alpha/2} - \frac{\boldsymbol{\psi}'\boldsymbol{\kappa}}{\sqrt{\mathcal{V}_{\boldsymbol{\psi},\mathbf{x}} + \delta^2 \mathcal{V}_{\boldsymbol{\psi},y}}} \Biggr). \qquad \square
\end{aligned}
$$

C.2.1. CPE from two-sample Taylor approximations: $T_{h,2}$*,* $T_{l,2}$

Lemma C.3. *Under the assumptions of Theorem C.2, the term* $T_{h,2}$ *from* (C.53) *is of order* $T_{h,2} = O(n^{-1/2}\log(n))$*, and similarly* $T_{l,2} = O(n^{-1/2}\log(n))$ *for the corresponding upper one-sided term. Additionally,* $T_{h,2} + T_{l,2} = O(n^{-1}(\log(n))^2)$*.*

Proof: The proof largely parallels that of Lemma A.4; here, we point out non-trivial differences. Because the samples are assumed independent in A2.1, joint PDFs of objects involving both samples are simply the product of the marginal PDFs. For example, in the proof of Lemma A.4, the PDFs of $\boldsymbol{\Delta}^H$ and $\boldsymbol{\Delta}^L$ were given. Continuing to use subscripts to denote the sample (x or y), since $\boldsymbol{\Delta}^H_y \perp\!\!\!\perp \boldsymbol{\Delta}^L_x$ by A2.1, their joint PDF is the product of their PDFs. Also, the conditions in A2.2 for both population distributions are the same as for the population distribution in the proof of Lemma A.4, so remainder terms have the same bounds as before.

Applying the same arguments as before, $T_{h,2}(\tilde\alpha)$ can be decomposed into

$$
\begin{aligned}
T_{h,2}(\tilde\alpha) &= T_{H,2,1} - T_{H,2,2} + O(n^{-1}(\log(n))^{3/2}),\\
T_{H,2,2} &\equiv \mathrm{P}\left(\sum_{j=1}^{J}\psi_j(Q'_Y(\tau_j)(\tilde Q^I_{U_y}(u^H_{y,j}(\tilde\alpha))-\tau_j) - Q'_X(\tau_j)(\tilde Q^I_{U_x}(u^L_{x,j}(\tilde\alpha))-\tau_j)) > 0\right)\\
&= \mathrm{P}\left(\sum_{j=1}^{J}\psi_j(\delta Q'_Y(\tau_j)(\Delta^H_{y,j}+D^H_{y,j}) - Q'_X(\tau_j)(\Delta^L_{x,j}+D^L_{x,j})) > 0\right),\\
T_{H,2,1} &\equiv \mathrm{P}\left(\sum_{j=1}^{J}\psi_j(\delta Q'_Y(\tau_j)(\Delta^H_{y,j}+D^H_{y,j}) - Q'_X(\tau_j)(\Delta^L_{x,j}+D^L_{x,j}))\right.\\
&\qquad\left. > -\frac{n_x^{-1/2}}{2}\sum_{j=1}^{J}\psi_j(\delta^2 Q''_Y(\tau_j)(\Delta^H_{y,j}+D^H_{y,j})^2 - Q''_X(\tau_j)(\Delta^L_{x,j}+D^L_{x,j})^2)\right),\\
\Delta^H_{y,j} &\equiv \sqrt{n_y}(\tilde Q^I_{U_y}(u^H_{y,j}(\tilde\alpha)) - u^H_{y,j}(\tilde\alpha)),\quad \Delta^L_{x,j} \equiv \sqrt{n_x}(\tilde Q^I_{U_x}(u^L_{x,j}(\tilde\alpha)) - u^L_{x,j}(\tilde\alpha)),\\
D^H_{y,j} &\equiv \sqrt{n_y}(u^H_{y,j}(\tilde\alpha)-\tau_j),\quad D^L_{x,j} \equiv \sqrt{n_x}(u^L_{x,j}(\tilde\alpha)-\tau_j).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
T_{l,2}(\tilde\alpha) &= T_{L,2,1} - T_{L,2,2} + O(n^{-1}(\log(n))^{3/2}),\\
T_{L,2,2} &\equiv \mathrm{P}\left(\sum_{j=1}^{J}\psi_j(Q'_Y(\tau_j)(\tilde Q^I_{U_y}(u^L_{y,j}(\tilde\alpha))-\tau_j) - Q'_X(\tau_j)(\tilde Q^I_{U_x}(u^H_{x,j}(\tilde\alpha))-\tau_j)) < 0\right)\\
&= \mathrm{P}\left(\sum_{j=1}^{J}\psi_j(\delta Q'_Y(\tau_j)(\Delta^L_{y,j}+D^L_{y,j}) - Q'_X(\tau_j)(\Delta^H_{x,j}+D^H_{x,j})) < 0\right),\\
T_{L,2,1} &\equiv \mathrm{P}\left(\sum_{j=1}^{J}\psi_j(\delta Q'_Y(\tau_j)(\Delta^L_{y,j}+D^L_{y,j}) - Q'_X(\tau_j)(\Delta^H_{x,j}+D^H_{x,j}))\right.\\
&\qquad\left. < -\frac{n_x^{-1/2}}{2}\sum_{j=1}^{J}\psi_j(\delta^2 Q''_Y(\tau_j)(\Delta^L_{y,j}+D^L_{y,j})^2 - Q''_X(\tau_j)(\Delta^H_{x,j}+D^H_{x,j})^2)\right),\\
\Delta^L_{y,j} &\equiv \sqrt{n_y}(\tilde Q^I_{U_y}(u^L_{y,j}(\tilde\alpha)) - u^L_{y,j}(\tilde\alpha)),\quad \Delta^H_{x,j} \equiv \sqrt{n_x}(\tilde Q^I_{U_x}(u^H_{x,j}(\tilde\alpha)) - u^H_{x,j}(\tilde\alpha)),\\
D^L_{y,j} &\equiv \sqrt{n_y}(u^L_{y,j}(\tilde\alpha)-\tau_j),\quad D^H_{x,j} \equiv \sqrt{n_x}(u^H_{x,j}(\tilde\alpha)-\tau_j).
\end{aligned}
$$

Also, adding x and y subscripts to (C.13), let

$$
\begin{aligned}
D^H_{0,y} &\equiv \sum_{j=1}^{J}\psi_j\gamma_j D^H_{y,j},\quad D^L_{0,y} \equiv \sum_{j=1}^{J}\psi_j\gamma_j D^L_{y,j} = -D^H_{0,y} + O(n^{-1/2}),\\
D^L_{0,x} &\equiv \sum_{j=1}^{J}\psi_j\gamma_j D^L_{x,j},\quad D^H_{0,x} \equiv \sum_{j=1}^{J}\psi_j\gamma_j D^H_{x,j} = -D^L_{0,x} + O(n^{-1/2}).
\end{aligned}
\tag{C.54}
$$

Using (C.8) and (C.9), the probability that any $\Delta_j^2 > 2\log(n)$ ("Δ_j" including $\Delta^H_{y,j}$,

$\Delta^H_{x,j}$, $\Delta^L_{y,j}$, and $\Delta^L_{x,j}$) is again $O(n^{-2}(\log(n))^{1/2})$, much smaller than the order of magnitude in the statement of this lemma, so we can again focus on the case where all $|\Delta_j| < \sqrt{2\log(n)}$.

For the one-sided result, the remaining bounds and arguments are identical to the proof of Lemma A.4.

For the two-sided result, the strategy is the same, but there are a few differences in the details. Consider $T_{H,2,2}$ first. Before, the strategy was to write the probability in terms of a quadratic function of Δ^H_1, conditional on the other Δ^H_j values. Let $\psi_1 = 1$ again, and let $\gamma_{y,j} = Q'_Y(\tau_j)/Q'_Y(\tau_1)$ and $\gamma_{x,j} = Q'_X(\tau_j)/Q'_Y(\tau_1)$, so $\gamma_{y,1} = 1$. Adding a y subscript to (C.21), let

$$\boldsymbol{\Delta}^H_{y,-1} \equiv (\Delta^H_{y,2}, \ldots, \Delta^H_{y,J})', \quad \boldsymbol{\Delta}^L_{y,-1} \equiv (\Delta^L_{y,2}, \ldots, \Delta^L_{y,J})'. \tag{C.55}$$

Analogous to (C.22), using (C.54), define the function $\pi^{H,2}_y(\cdot,\cdot) : \mathbb{R}^{J-1} \times \mathbb{R}^J \mapsto \mathbb{R}$ as

$$\pi^{H,2}_y(\mathbf{v}_{-1}, \mathbf{w}) = D^L_{0,x} - D^H_{0,y} - \sum_{j=2}^{J} \psi_j \gamma_{y,j} v_j + \sum_{j=1}^{J} \psi_j \gamma_{x,j} w_j \tag{C.56}$$

for any arguments $\mathbf{v}_{-1} = (v_2, \ldots, v_J)' \in \mathbb{R}^{J-1}$ and $\mathbf{w} = (w_1, \ldots, w_J)' \in \mathbb{R}^J$. Thus, analogous to (C.23),

$$T_{H,2,2} = \mathrm{P}(\Delta^H_{y,1} > \delta^{-1} \pi^{H,2}_y(\boldsymbol{\Delta}^H_{y,-1}, \boldsymbol{\Delta}^L_x)). \tag{C.57}$$

Also,

$$\begin{aligned}
T_{H,2,1} &= \mathrm{P}(a(\Delta^H_{y,1})^2 + b\Delta^H_{y,1} + c > 0), \\
&a \equiv n_x^{-1/2} a_0, \quad a_0 \equiv \frac{\delta^2 Q''_Y(\tau_1)}{2Q'_Y(\tau_1)}, \\
&b \equiv \delta + n_x^{-1/2} b_0, \quad b_0 \equiv D^H_{y,1} \frac{\delta^2 Q''_Y(\tau_1)}{Q'_Y(\tau_1)}, \quad b^{-1} = \delta^{-1} - n_x^{-1/2} b_0 \delta^{-2} + O(n_x^{-1}), \\
&c \equiv -\pi^{H,2}_y(\boldsymbol{\Delta}^H_{y,-1}, \boldsymbol{\Delta}^L_x) + \frac{n_x^{-1/2}\delta^2 Q''_Y(\tau_1)}{2Q'_Y(\tau_1)} (D^H_{y,1})^2 + \frac{n_x^{-1/2}}{2} \sum_{j=2}^{J} \psi_j \delta^2 \frac{Q''_Y(\tau_j)}{Q'_Y(\tau_1)} (\Delta^H_{y,j} + D^H_{y,j})^2 \\
&\quad - \frac{n_x^{-1/2}}{2} \sum_{j=1}^{J} \psi_j \frac{Q''_X(\tau_j)}{Q'_Y(\tau_1)} (\Delta^L_{x,j} + D^L_{x,j})^2 \\
&\quad = O((\log(n))^{1/2}).
\end{aligned}$$

From the same arguments as before, the important root of the quadratic is

$$\begin{aligned}
r_+ &= \frac{-b + b - 2ac/b - 2a^2c^2/b^3 + O(n^{-3/2}(\log(n))^{3/2})}{2a} = -\frac{c}{b} - \frac{ac^2}{b^3} + O(n^{-1}(\log(n))^{3/2}) \\
&= -c(\delta^{-1} - n_x^{-1/2} b_0 \delta^{-2}) - n_x^{-1/2} \frac{\delta^2 Q''_Y(\tau_1)}{2Q'_Y(\tau_1)} (\pi^{H,2}_y(\boldsymbol{\Delta}^H_{y,-1}, \boldsymbol{\Delta}^L_x))^2 \delta^{-3} + O(n^{-1}(\log(n))^{3/2})
\end{aligned}$$

Similar to (C.56), define the function $\pi^{H,1}_y(\cdot,\cdot) : \mathbb{R}^{J-1} \times \mathbb{R}^J \mapsto \mathbb{R}$ so that

$$\begin{aligned}
r_+ &= \pi^{H,1}_y(\boldsymbol{\Delta}^H_{y,-1}, \boldsymbol{\Delta}^L_x) \\
&= -c(\delta^{-1} - n_x^{-1/2} b_0 \delta^{-2}) - n_x^{-1/2} \frac{\delta^2 Q''_Y(\tau_1)}{2Q'_Y(\tau_1)} (\pi^{H,2}_y(\boldsymbol{\Delta}^H_{y,-1}, \boldsymbol{\Delta}^L_x))^2 \delta^{-3}
\end{aligned}$$

$$+ O(n^{-1}(\log(n))^{3/2}),$$

where c implicitly depends on the arguments, too. Similar to before, the other root's impact is smaller-order: $a_0 = O(1)$ by A2.2, so r_- is of order at least $n^{1/2}$, and the corresponding tail probability is exponentially small, a loose bound for which is $O(e^{-0.99n})$. Altogether,

$$T_{H,2,1} = \mathrm{P}(\Delta_{y,1}^H > \pi_y^{H,1}(\boldsymbol{\Delta}_{y,-1}^H, \boldsymbol{\Delta}_x^L)) + O(e^{-0.99n}),$$
$$T_{h,2} = \mathrm{P}(\Delta_{y,1}^H > \pi_y^{H,1}(\boldsymbol{\Delta}_{y,-1}^H, \boldsymbol{\Delta}_x^L)) - \mathrm{P}(\Delta_{y,1}^H > \pi_y^{H,2}(\boldsymbol{\Delta}_{y,-1}^H, \boldsymbol{\Delta}_x^L)) + O(n^{-1}(\log(n))^{3/2}),$$

and

$$\begin{aligned}
&\delta^{-1}\pi_y^{H,2}(\mathbf{v}_{-1}, \mathbf{w}) - \pi_y^{H,1}(\mathbf{v}_{-1}, \mathbf{w}) \\
&= \delta^{-1}\pi_y^{H,2}(\mathbf{v}_{-1}, \mathbf{w}) + \delta^{-1}c - n_x^{-1/2}\delta^{-2}cb_0 + n_x^{-1/2}\delta^{-1}\frac{Q_Y''(\tau_1)}{2Q_Y'(\tau_1)}(\pi_y^{H,2}(\mathbf{v}_{-1}, \mathbf{w}))^2 + O(n^{-1}(\log(n))^{3/2}) \\
&= \delta^{-1}\frac{n_x^{-1/2}\delta^2 Q_Y''(\tau_1)}{2Q_Y'(\tau_1)}(D_{y,1}^H)^2 + \delta^{-1}\frac{n_x^{-1/2}}{2}\sum_{j=2}^J \psi_j \delta^2 \frac{Q_Y''(\tau_j)}{Q_Y'(\tau_1)}(v_j + D_{y,j}^H)^2 \\
&\quad - \delta^{-1}\frac{n_x^{-1/2}}{2}\sum_{j=1}^J \psi_j \frac{Q_X''(\tau_j)}{Q_Y'(\tau_1)}(w_j + D_{x,j}^L)^2 \\
&\quad - n_x^{-1/2}\delta^{-2}\overbrace{D_{y,1}^H \frac{\delta^2 Q_Y''(\tau_1)}{Q_Y'(\tau_1)}}^{b_0}\overbrace{(-\pi_y^{H,2}(\mathbf{v}_{-1}, \mathbf{w}) + O(n_x^{-1/2}(\log(n))))}^{c} \\
&\quad + n_x^{-1/2}\delta^{-1}\frac{Q_Y''(\tau_1)}{2Q_Y'(\tau_1)}(\pi_y^{H,2}(\mathbf{v}_{-1}, \mathbf{w}))^2 \\
&\quad + O(n^{-1}(\log(n))^{3/2})
\end{aligned}$$

when all $|v_j| < (\log(n))^{1/2}$ and $|w_j| < (\log(n))^{1/2}$.

The above arguments can be repeated for $T_{l,2}$. First, the analogue of (C.56) is

$$\pi_y^{L,2}(\mathbf{v}_{-1}, \mathbf{w}) = D_{0,x}^H - D_{0,y}^L - \sum_{j=2}^J \psi_j \gamma_{y,j} v_j + \sum_{j=1}^J \psi_j \gamma_{x,j} w_j. \tag{C.58}$$

Second,

$$\begin{aligned}
T_{L,2,1} &= \mathrm{P}(a(\Delta_{y,1}^L)^2 + b\Delta_{y,1}^L + c < 0), \\
&\quad a \equiv n_x^{-1/2}a_0, \quad a_0 \equiv \frac{\delta^2 Q_Y''(\tau_1)}{2Q_Y'(\tau_1)}, \\
&\quad b \equiv \delta + n_x^{-1/2}b_0, \quad b_0 \equiv D_{y,1}^L \frac{\delta^2 Q_Y''(\tau_1)}{Q_Y'(\tau_1)}, \quad b^{-1} = \delta^{-1} - n_x^{-1/2}b_0\delta^{-2} + O(n_x^{-1}), \\
&\quad c \equiv -\pi_y^{L,2}(\boldsymbol{\Delta}_{y,-1}^L, \boldsymbol{\Delta}_x^H) + \frac{n_x^{-1/2}\delta^2 Q_Y''(\tau_1)}{2Q_Y'(\tau_1)}(D_{y,1}^L)^2 + \frac{n_x^{-1/2}}{2}\sum_{j=2}^J \psi_j \delta^2 \frac{Q_Y''(\tau_j)}{Q_Y'(\tau_1)}(\Delta_{y,j}^L + D_{y,j}^L)^2 \\
&\qquad - \frac{n_x^{-1/2}}{2}\sum_{j=1}^J \psi_j \frac{Q_X''(\tau_j)}{Q_Y'(\tau_1)}(\Delta_{x,j}^H + D_{x,j}^H)^2.
\end{aligned}$$

Consequently,

$$\begin{aligned}
T_{L,2,1} &= \mathrm{P}(\Delta_{y,1}^{L} < \pi_y^{L,1}(\boldsymbol{\Delta}_{y,-1}^{L}, \boldsymbol{\Delta}_x^{H})) + O(e^{-0.99n}),\\
\pi_y^{L,1}(\boldsymbol{\Delta}_{y,-1}^{L}, \boldsymbol{\Delta}_x^{H}) &= -c(\delta^{-1} - n_x^{-1/2} b_0 \delta^{-2}) - n_x^{-1/2} \frac{\delta^2 Q_Y''(\tau_1)}{2Q_Y'(\tau_1)} (\pi_y^{L,2}(\boldsymbol{\Delta}_{y,-1}^{L}, \boldsymbol{\Delta}_x^{H}))^2 \delta^{-3}\\
&\quad + O(n^{-1}(\log(n))^{3/2}),
\end{aligned}$$

where c also (implicitly) depends on the argument. Third, using (C.54), (C.56), and (C.58), analogous to (C.33) and (C.32),

$$\begin{aligned}
&\pi_y^{L,2}(-\mathbf{v}_{-1}, -\mathbf{w}) = -\pi_y^{H,2}(\mathbf{v}_{-1}, \mathbf{w}) + O(n^{-1/2}),\\
&\pi_y^{L,1}(-\mathbf{v}_{-1}, -\mathbf{w}) - \delta^{-1}\pi_y^{L,2}(-\mathbf{v}_{-1}, -\mathbf{w})\\
&= -\delta^{-1} \frac{n_x^{-1/2} \delta^2 Q_Y''(\tau_1)}{2Q_Y'(\tau_1)} (D_{y,1}^{L})^2 - \delta^{-1} \frac{n_x^{-1/2}}{2} \sum_{j=2}^{J} \psi_j \delta^2 \frac{Q_Y''(\tau_j)}{Q_Y'(\tau_1)} (-v_j + D_{y,j}^{L})^2\\
&\quad + \delta^{-1} \frac{n_x^{-1/2}}{2} \sum_{j=1}^{J} \psi_j \frac{Q_X''(\tau_j)}{Q_Y'(\tau_1)} (-w_j + D_{x,j}^{H})^2\\
&\quad + n_x^{-1/2} \delta^{-2} D_{y,1}^{L} \frac{\delta^2 Q_Y''(\tau_1)}{Q_Y'(\tau_1)} (-\pi_y^{L,2}(-\mathbf{v}_{-1}, -\mathbf{w}) + O(n^{-1/2}))\\
&\quad - n_x^{-1/2} \frac{\delta^2 Q_Y''(\tau_1)}{2Q_Y'(\tau_1)} (\pi_y^{L,2}(-\mathbf{v}_{-1}, -\mathbf{w}))^2 \delta^{-3} + O(n^{-1}(\log(n))^{3/2})\\
&= -\delta^{-1} \frac{n_x^{-1/2} \delta^2 Q_Y''(\tau_1)}{2Q_Y'(\tau_1)} (-D_{y,1}^{H})^2 - \delta^{-1} \frac{n_x^{-1/2}}{2} \sum_{j=2}^{J} \psi_j \delta^2 \frac{Q_Y''(\tau_j)}{Q_Y'(\tau_1)} (-v_j - D_{y,j}^{H})^2\\
&\quad + \delta^{-1} \frac{n_x^{-1/2}}{2} \sum_{j=1}^{J} \psi_j \frac{Q_X''(\tau_j)}{Q_Y'(\tau_1)} (-w_j - D_{x,j}^{L})^2\\
&\quad + n_x^{-1/2} \delta^{-2} (-D_{y,1}^{H}) \frac{\delta^2 Q_Y''(\tau_1)}{Q_Y'(\tau_1)} \pi_y^{H,2}(\mathbf{v}_{-1}, \mathbf{w})\\
&\quad - n_x^{-1/2} \frac{\delta^2 Q_Y''(\tau_1)}{2Q_Y'(\tau_1)} (-\pi_y^{H,2}(\mathbf{v}_{-1}, \mathbf{w}))^2 \delta^{-3} + O(n^{-1}(\log(n))^{3/2})\\
&= -\delta^{-1} \frac{n_x^{-1/2} \delta^2 Q_Y''(\tau_1)}{2Q_Y'(\tau_1)} (D_{y,1}^{H})^2 - \delta^{-1} \frac{n_x^{-1/2}}{2} \sum_{j=2}^{J} \psi_j \delta^2 \frac{Q_Y''(\tau_j)}{Q_Y'(\tau_1)} (v_j + D_{y,j}^{H})^2\\
&\quad + \delta^{-1} \frac{n_x^{-1/2}}{2} \sum_{j=1}^{J} \psi_j \frac{Q_X''(\tau_j)}{Q_Y'(\tau_1)} (w_j + D_{x,j}^{L})^2\\
&\quad - n_x^{-1/2} \delta^{-2} D_{y,1}^{H} \frac{\delta^2 Q_Y''(\tau_1)}{Q_Y'(\tau_1)} \pi_y^{H,2}(\mathbf{v}_{-1}, \mathbf{w})\\
&\quad - n_x^{-1/2} \frac{\delta^2 Q_Y''(\tau_1)}{2Q_Y'(\tau_1)} (\pi_y^{H,2}(\mathbf{v}_{-1}, \mathbf{w}))^2 \delta^{-3} + O(n^{-1}(\log(n))^{3/2})\\
&= \pi_y^{H,1}(\mathbf{v}_{-1}, \mathbf{w}) - \delta^{-1}\pi_y^{H,2}(\mathbf{v}_{-1}, \mathbf{w}) + O(n^{-1}(\log(n))^{3/2}).
\end{aligned}$$

The final obstacle is the approximation of the joint PDF of $(\boldsymbol{\Delta}_y^{H}, \boldsymbol{\Delta}_x^{L})$ and of $(\boldsymbol{\Delta}_y^{L}, \boldsymbol{\Delta}_x^{H})$.

Since $\mathbf{\Delta}_y^H \perp\!\!\!\perp \mathbf{\Delta}_x^L$ and $\mathbf{\Delta}_y^L \perp\!\!\!\perp \mathbf{\Delta}_x^H$ by A2.1, using the PDF approximations in (C.10),

$$
\begin{aligned}
f_{(\mathbf{\Delta}_y^H,\mathbf{\Delta}_x^L)}(\mathbf{v},\mathbf{w}) &= \overbrace{\phi_{\underline{\mathcal{V}}}(\mathbf{v})(1+O(n^{-1/2}(\log(n))^2))}^{\text{from (C.10)}}\overbrace{\phi_{\underline{\mathcal{V}}}(\mathbf{w})(1+O(n^{-1/2}(\log(n))^2))}^{\text{from (C.10)}} \\
&= \phi_{\underline{\mathcal{V}},\underline{\mathcal{V}}}(\mathbf{v},\mathbf{w})(1+O(n^{-1/2}(\log(n))^2)), \\
f_{(\mathbf{\Delta}_y^L,\mathbf{\Delta}_x^H)}(\mathbf{v},\mathbf{w}) &= \phi_{\underline{\mathcal{V}},\underline{\mathcal{V}}}(\mathbf{v},\mathbf{w})(1+O(n^{-1/2}(\log(n))^2)),
\end{aligned}
$$

with $\phi_{\underline{\mathcal{V}},\underline{\mathcal{V}}}$ indicating the PDF of a mean-zero normal distribution with block diagonal covariance matrix with blocks $\underline{\mathcal{V}}$ and (again) $\underline{\mathcal{V}}$.

Now, the same arguments as before apply. To be explicit, following the derivation of (C.30) with analogous notation yields the parallel result

$$
\begin{aligned}
T_{h,2} = &\int\!\!\cdots\!\!\int_{\substack{v_j^2\le 2\log(n),j\ge 2\\ w_j^2\le 2\log(n)}} \int_{\pi_y^{H,1}(\mathbf{v}_{-1},\mathbf{w})}^{\delta^{-1}\pi_y^{H,2}(\mathbf{v}_{-1},\mathbf{w})} f_{\Delta_{y,1}^H|\mathbf{\Delta}_{y,-1}^H,\mathbf{\Delta}_x^L}(v_1\mid \mathbf{v}_{-1},\mathbf{w}) f_{\mathbf{\Delta}_{y,-1}^H,\mathbf{\Delta}_x^L}(\mathbf{v}_{-1},\mathbf{w}) \\
&\qquad\times dv_1\,dv_2\cdots dv_J\,dw_1\cdots dw_J \\
&+O(n^{-1}(\log(n))^{3/2}) \\
= &\int\!\!\cdots\!\!\int_{\substack{v_j^2\le 2\log(n),j\ge 2\\ w_j^2\le 2\log(n)}} (\delta^{-1}\pi_y^{H,2}(\mathbf{v}_{-1},\mathbf{w})-\pi_y^{H,1}(\mathbf{v}_{-1},\mathbf{w}))\phi_{V_1|\mathbf{V}_{-1},\mathbf{W}}(\delta^{-1}\pi_y^{H,2}(\mathbf{v}_{-1},\mathbf{w})\mid \mathbf{v}_{-1},\mathbf{w}) \\
&\qquad\times \phi_{\underline{\mathcal{V}}_{-1},\underline{\mathcal{V}}}(\mathbf{v}_{-1},\mathbf{w})\,dv_2\cdots dv_J\,dw_1\cdots dw_J \\
&+O(n^{-1}(\log(n))^2).
\end{aligned}
$$

Then, following the same steps as in the derivation of (C.34),

$$
\begin{aligned}
T_{l,2} = &\int\!\!\cdots\!\!\int_{\substack{v_j^2\le 2\log(n),j\ge 2\\ w_j^2\le 2\log(n)}} (\pi_y^{L,1}(-\mathbf{v}_{-1},-\mathbf{w})-\delta^{-1}\pi_y^{L,2}(-\mathbf{v}_{-1},-\mathbf{w})) \\
&\qquad\times \phi_{V_1|\mathbf{V}_{-1},\mathbf{W}}(\pi_y^{L,2}(-\mathbf{v}_{-1},-\mathbf{w})\mid -\mathbf{v}_{-1},-\mathbf{w})\phi_{\underline{\mathcal{V}}_{-1},\underline{\mathcal{V}}}(\mathbf{v}_{-1},\mathbf{w}) \\
&\qquad\times dv_2\cdots dv_J\,dw_1\cdots dw_J \\
&+O(n^{-1}(\log(n))^2) \\
= &\int\!\!\cdots\!\!\int_{\substack{v_j^2\le 2\log(n),j\ge 2\\ w_j^2\le 2\log(n)}} (\pi_y^{H,1}(\mathbf{v}_{-1},\mathbf{w})-\delta^{-1}\pi_y^{H,2}(\mathbf{v}_{-1},\mathbf{w})+O(n^{-1}(\log(n))^{3/2})) \\
&\qquad\times \phi_{V_1|\mathbf{V}_{-1},\mathbf{W}}(-\pi_y^{H,2}(\mathbf{v}_{-1},\mathbf{w})\mid -\mathbf{v}_{-1},-\mathbf{w})\phi_{\underline{\mathcal{V}}_{-1},\underline{\mathcal{V}}}(\mathbf{v}_{-1},\mathbf{w}) \\
&\qquad\times dv_2\cdots dv_J\,dw_1\cdots dw_J \\
&+O(n^{-1}(\log(n))^2) \\
= -&\int\!\!\cdots\!\!\int_{\substack{v_j^2\le 2\log(n),j\ge 2\\ w_j^2\le 2\log(n)}} (\pi_y^{H,2}(\mathbf{v}_{-1},\mathbf{w})-\delta^{-1}\pi_y^{H,1}(\mathbf{v}_{-1},\mathbf{w}))\phi_{V_1|\mathbf{V}_{-1},\mathbf{W}}(\pi_y^{H,2}(\mathbf{v}_{-1},\mathbf{w})\mid \mathbf{v}_{-1},\mathbf{w})
\end{aligned}
$$

$$
\begin{aligned}
&\quad \times \phi_{\underline{\mathcal{V}}_{-1},\underline{\mathcal{V}}}(\mathbf{v}_{-1},\mathbf{w})dv_2\cdots dv_J\,dw_1\cdots dw_J \\
&\quad + O(n^{-1}(\log(n))^2) \\
&= -T_{h,2} + O(n^{-1}(\log(n))^2),
\end{aligned}
$$

so $T_{l,2}+T_{h,2}=O(n^{-1}(\log(n))^2)$. □

C.2.2. CPE from two-sample nuisance parameter estimation error: $E_{h,2}$, $E_{l,2}$

Lemma C.4. *Under the assumptions of Theorem C.2, the term $E_{h,2}$ from* (C.53) *is of order $E_{h,2}=O(m^{-1}\log(n)+(m/n)^2)$, and similarly $E_{l,2}=O(m^{-1}\log(n)+(m/n)^2)$ for the corresponding upper one-sided term, where m is the common rate of smoothing parameters, $m_j \asymp m$ for all j.*

Proof: The proof largely parallels that of Lemma A.5; here, we walk through the differences. Notation remains the same, but with subscripts x and y referring to the two samples and populations.

Notational modifications aside, the first difference is in the decomposition of $E_{h,2}$:

$$
\begin{aligned}
E_{h,2} &= \mathrm{P}(\delta\mathbb{W}^{\hat{\mathbf{u}}_y^H}_{\mathbf{C},\mathbf{\Lambda}} - \mathbb{W}^{\hat{\mathbf{u}}_x^H}_{\mathbf{C},\mathbf{\Lambda}} > \delta\sqrt{n_y}\boldsymbol{\psi}'(Q_Y(\boldsymbol{\tau})-Q_Y(\hat{\mathbf{u}}_y^H)) - \sqrt{n_x}\boldsymbol{\psi}'(Q_X(\boldsymbol{\tau})-Q_X(\hat{\mathbf{u}}_x^H))) \\
&\quad - \mathrm{P}(\delta\mathbb{W}^{\mathbf{u}_{0,y}^H}_{\mathbf{C},\mathbf{\Lambda}} - \mathbb{W}^{\mathbf{u}_{0,x}^H}_{\mathbf{C},\mathbf{\Lambda}} > \delta\sqrt{n_y}\boldsymbol{\psi}'(Q_Y(\boldsymbol{\tau})-Q_Y(\mathbf{u}_{0,y}^H)) - \sqrt{n_x}\boldsymbol{\psi}'(Q_X(\boldsymbol{\tau})-Q_X(\mathbf{u}_{0,x}^H))) \\
&= E_{h,2}^1 + E_{h,2}^2, \\
E_{h,2}^1 &\equiv E[\mathrm{P}(\delta\mathbb{W}^{\hat{\mathbf{u}}_y^H}_{\mathbf{C},\mathbf{\Lambda}} - \mathbb{W}^{\hat{\mathbf{u}}_x^H}_{\mathbf{C},\mathbf{\Lambda}} \\
&\qquad < \delta\sqrt{n_y}\boldsymbol{\psi}'(Q_Y(\boldsymbol{\tau})-Q_Y(\mathbf{u}_{0,y}^H)) - \sqrt{n_x}\boldsymbol{\psi}'(Q_X(\boldsymbol{\tau})-Q_X(\mathbf{u}_{0,x}^H)) \mid \hat{\boldsymbol{\gamma}}) \\
&\quad - \mathrm{P}(\delta\mathbb{W}^{\hat{\mathbf{u}}_y^H}_{\mathbf{C},\mathbf{\Lambda}} - \mathbb{W}^{\hat{\mathbf{u}}_x^H}_{\mathbf{C},\mathbf{\Lambda}} \\
&\qquad < \delta\sqrt{n_y}\boldsymbol{\psi}'(Q_Y(\boldsymbol{\tau})-Q_Y(\hat{\mathbf{u}}_y^H)) - \sqrt{n_x}\boldsymbol{\psi}'(Q_X(\boldsymbol{\tau})-Q_X(\hat{\mathbf{u}}_x^H)) \mid \hat{\boldsymbol{\gamma}})], \\
E_{h,2}^2 &\equiv E[\mathrm{P}(\delta\mathbb{W}^{\mathbf{u}_{0,y}^H}_{\mathbf{C},\mathbf{\Lambda}} - \mathbb{W}^{\mathbf{u}_{0,x}^H}_{\mathbf{C},\mathbf{\Lambda}} \\
&\qquad < \delta\sqrt{n_y}\boldsymbol{\psi}'(Q_Y(\boldsymbol{\tau})-Q_Y(\mathbf{u}_{0,y}^H)) - \sqrt{n_x}\boldsymbol{\psi}'(Q_X(\boldsymbol{\tau})-Q_X(\mathbf{u}_{0,x}^H)) \mid \hat{\boldsymbol{\gamma}}) \\
&\quad - \mathrm{P}(\delta\mathbb{W}^{\hat{\mathbf{u}}_y^H}_{\mathbf{C},\mathbf{\Lambda}} - \mathbb{W}^{\hat{\mathbf{u}}_x^H}_{\mathbf{C},\mathbf{\Lambda}} \\
&\qquad < \delta\sqrt{n_y}\boldsymbol{\psi}'(Q_Y(\boldsymbol{\tau})-Q_Y(\mathbf{u}_{0,y}^H)) - \sqrt{n_x}\boldsymbol{\psi}'(Q_X(\boldsymbol{\tau})-Q_X(\mathbf{u}_{0,x}^H)) \mid \hat{\boldsymbol{\gamma}})].
\end{aligned}
$$

In the one-sample proof, there was a single $\mathbb{W}^{\hat{\mathbf{u}}^H}_{\mathbf{C},\mathbf{\Lambda}}$, whose PDF is normal up to a multiplicative approximation error. Here, we have such a random variable for each sample, and we need the PDF of the difference $\delta\mathbb{W}^{\hat{\mathbf{u}}_y^H}_{\mathbf{C},\mathbf{\Lambda}} - \mathbb{W}^{\hat{\mathbf{u}}_x^H}_{\mathbf{C},\mathbf{\Lambda}}$. By A2.1, the samples are drawn independently, so the random variables are independent, and the PDF of the difference can be derived via convolution. Heuristically (to simplify notation), if random variable W has PDF $\phi_{\sigma_W^2}(w)(1+O(r_n))$ and Z has PDF $\phi_{\sigma_Z^2}(z)(1+O(r_n))$, with $W \perp\!\!\!\perp Z$, then $\delta W - Z$ has PDF

$$
\begin{aligned}
f_{\delta W - Z}(t) &= \int_{\mathbb{R}} f_{W,Z}(w,\delta w - t)\,dw \\
&= \int_{\mathbb{R}} \overbrace{f_W(w) f_Z(\delta w - t)}^{\text{since } W \perp\!\!\!\perp Z}\,dw
\end{aligned}
$$

$$
\begin{aligned}
&= \int_{\mathbb{R}} \phi_{\sigma_W^2}(w)(1+O(r_n))\phi_{\sigma_Z^2}(\delta w - t)(1+O(r_n))\,dw \\
&= \overbrace{\int_{\mathbb{R}} \phi_{\sigma_W^2}(w)\phi_{\sigma_Z^2}(\delta w - t)\,dw}^{\text{difference of normals has normal PDF}} \quad (1+O(r_n)) \\
&= \phi_{\delta^2\sigma_W^2+\sigma_Z^2}(t)(1+O(r_n)).
\end{aligned}
$$

So again we have a normal PDF, up to a multiplicative error of the same order as before. Adding x and y subscripts to the notation from the proof of Lemma A.5, the final variance is

$$
\mathcal{V}^{0,H} \equiv \delta^2 \mathcal{V}^{0,H}_{\boldsymbol{\psi},y} + \mathcal{V}^{0,H}_{\boldsymbol{\psi},x} + O(m^{-1/2}n^{-1/2}\log(n) + mn^{-3/2}). \tag{C.59}
$$

The same arguments as in the proof of Lemma A.5 now apply. Following the derivation of (C.45), given any values of $\hat{\mathbf{u}}_x^H$ and $\hat{\mathbf{u}}_y^H$,

$$
\begin{aligned}
E^1_{h,2}(\hat{\mathbf{u}}_x^H, \hat{\mathbf{u}}_y^H) &= \sqrt{n_x}\boldsymbol{\psi}'((Q_Y(\hat{\mathbf{u}}_y^H) - Q_Y(\mathbf{u}_{0,y}^H)) - (Q_X(\hat{\mathbf{u}}_x^H) - Q_X(\mathbf{u}_{0,x}^H)))\phi_{\mathcal{V}^{0,H}}(\tilde{w}) \\
&\quad \times (1+O(n^{-1/2}(\log(n))^3)) \\
&= \boldsymbol{\psi}'(\delta(\widehat{Q'_Y}(\boldsymbol{\tau}) - Q'_Y(\boldsymbol{\tau})) - (\widehat{Q'_X}(\boldsymbol{\tau}) - Q'_X(\boldsymbol{\tau})))\phi_{\mathcal{V}^{0,H}}(\tilde{w}) \\
&\quad \times (1+O(n^{-1/2}(\log(n))^3)),
\end{aligned}
$$

where $\tilde{w}$ (from the mean value theorem) is again $O(1)$. Then, mirroring the derivation of (C.46),

$$
\begin{aligned}
E^1_{h,2} &= O(E[AB]), & A &= 1 + O(m^{-1/2}\log(n) + mn^{-1}), \\
B &= \max_j B_j, & B_j &= \delta\,\overbrace{(\widehat{Q'_Y}(\tau_j) - Q'_Y(\tau_j))}^{B_y} - \overbrace{(\widehat{Q'_X}(\tau_j) - Q'_X(\tau_j))}^{B_x}.
\end{aligned}
$$

By linearity of the expectation operator,

$$
E[\delta B_y - B_x] = \delta E[B_y] - E[B_x] = O(m^2/n^2)
$$

like before. Since $B_x \perp\!\!\!\perp B_y$, $\mathrm{Var}(\delta B_y - B_x) = \delta^2 \mathrm{Var}(B_y) + \mathrm{Var}(B_x) = O(m^{-1})$ like before, too. The remainder of the arguments about $E^1_{h,2}$ are identical to those for E^1_h in the proof of Lemma A.5.

For $E^2_{h,2}$, other than notational changes, the arguments are identical to those for E^2_h in the proof of Lemma A.5, with the covariance matrix changing from $\mathcal{V}^{\mathbf{t}}_{\boldsymbol{\psi}}$ to $\delta^2\mathcal{V}^{\mathbf{t}_y}_{\boldsymbol{\psi},y} + \mathcal{V}^{\mathbf{t}_x}_{\boldsymbol{\psi},x}$. □

D. NUISANCE PARAMETER ESTIMATION: ADDITIONAL CALCULATIONS AND COMPARISONS

The derivatives of a quantile function $Q(\cdot) = F^{-1}(\cdot)$ are (mostly just applying the chain rule for derivatives)

$$
Q'(\tau) = \frac{1}{f(Q(\tau))} = (f(Q(\tau)))^{-1}, \tag{D.1}
$$

$$
Q''(\tau) = -(f(Q(\tau)))^{-2} f'(Q(\tau))Q'(\tau) = -\frac{1}{(f(Q(\tau)))^2}\frac{f'(Q(\tau))}{f(Q(\tau))} = -\frac{f'(Q(\tau))}{(f(Q(\tau)))^3}, \tag{D.2}
$$

$$
\begin{aligned}
Q'''(\tau) &= -\frac{f''(Q(\tau))}{f(Q(\tau))}\frac{1}{(f(Q(\tau)))^3} + (-f'(Q(\tau)))(-3)\frac{1}{(f(Q(\tau)))^4}\frac{f'(Q(\tau))}{f(Q(\tau))} \\
&= \frac{-f''(Q(\tau))f(Q(\tau)) + 3(f'(Q(\tau)))^2}{(f(Q(\tau)))^5}.
\end{aligned} \tag{D.3}
$$

For a normal distribution with mean μ and variance σ^2, the PDF can be written in terms of the standard normal PDF $\phi(\cdot)$, as $\sigma^{-1}\phi((x-\mu)/\sigma)$, and the quantile function as

$$
Q(\tau) = \mu + \sigma\Phi^{-1}(\tau), \tag{D.4}
$$

so

$$
f(Q(\tau)) = \sigma^{-1}\phi\left(\frac{\mu + \sigma\Phi^{-1}(\tau) - \mu}{\sigma}\right) = \sigma^{-1}\phi(\Phi^{-1}(\tau)). \tag{D.5}
$$

Taking a derivative,

$$
f'(x) = \sigma^{-1}\phi'((x-\mu)/\sigma)\sigma^{-1} = \sigma^{-2}\frac{-(x-\mu)}{\sigma}\phi((x-\mu)/\sigma) = -\frac{x-\mu}{\sigma^2}f(x), \tag{D.6}
$$

which can also be seen from taking a derivative of $\exp(-(x-\mu)^2/(2\sigma^2))$ in x. For the second derivative, using the product rule,

$$
f''(x) = -\sigma^{-2}f(x) - \frac{x-\mu}{\sigma^2}f'(x) = f(x)\left(-\sigma^{-2} + \frac{(x-\mu)^2}{\sigma^4}\right) = \sigma^{-2}\left(\left(\frac{x-\mu}{\sigma}\right)^2 - 1\right)f(x). \tag{D.7}
$$

Substituting (D.4)–(D.7) into (D.3), for a $\mathrm{N}(\mu,\sigma^2)$ distribution,

$$
\begin{aligned}
Q'''(\tau) &= \frac{-\sigma^{-2}((\frac{Q(\tau)-\mu}{\sigma})^2 - 1)f(Q(\tau))f(Q(\tau)) + 3(-\frac{Q(\tau)-\mu}{\sigma^2}f(Q(\tau)))^2}{(f(Q(\tau)))^5} \\
&= \frac{-\sigma^{-2}((\frac{Q(\tau)-\mu}{\sigma})^2 - 1) + 3\sigma^{-2}(\frac{Q(\tau)-\mu}{\sigma})^2}{(f(Q(\tau)))^3} \\
&= \frac{\sigma^{-2} - \sigma^{-2}(\frac{Q(\tau)-\mu}{\sigma})^2 + 3\sigma^{-2}(\frac{Q(\tau)-\mu}{\sigma})^2}{(\sigma^{-1}\phi(\Phi^{-1}(\tau)))^3} \\
&= \frac{\sigma^{-2} + 2\sigma^{-2}(\Phi^{-1}(\tau))^2}{\sigma^{-3}(\phi(\Phi^{-1}(\tau)))^3} \\
&= \frac{1 + 2(\Phi^{-1}(\tau))^2}{\sigma^{-1}(\phi(\Phi^{-1}(\tau)))^3}
\end{aligned}
$$

One option for m that achieves the CPE-optimal rate is to minimise the sum of the bias and variance of $\widehat{Q'}(\tau)$. The FOC is

$$
\begin{aligned}
0 &= \frac{\partial}{\partial m}(m/n)^2\frac{Q'''(\tau)}{6} + m^{-1}\frac{(Q'(\tau))^2}{2} = 2m/n^2\frac{Q'''(\tau)}{6} - m^{-2}\frac{(Q'(\tau))^2}{2}, \\
&\implies 2(m/n)^2\frac{Q'''(\tau)}{6} = m^{-1}\frac{(Q'(\tau))^2}{2}.
\end{aligned} \tag{D.8}
$$

Using a "Gaussian plug-in" approach like in the famous bandwidth of Silverman (1986), i.e., computing the quantile derivatives for a $\mathrm{N}(\mu,\sigma^2)$ distribution,

$$
m^3 = n^2\frac{6(Q'(\tau))^2}{4Q'''(\tau)} = n^2\frac{3(Q'(\tau))^2}{2Q'''(\tau)} = n^2\frac{3\sigma^{-1}(\phi(\Phi^{-1}(\tau)))^3}{2(1+2(\Phi^{-1}(\tau))^2)(\sigma^{-1}\phi(\Phi^{-1}(\tau)))^2}
$$

$$
\begin{aligned}
&= n^2 \frac{3\sigma\phi(\Phi^{-1}(\tau))}{2+4(\Phi^{-1}(\tau))^2}, \\
m &= n^{2/3}\left(1.5\frac{\sigma\phi(\Phi^{-1}(\tau))}{1+2(\Phi^{-1}(\tau))^2}\right)^{1/3}.
\end{aligned}
\tag{D.9}
$$

Similar to the suggestion by Silverman (1986), for robustness to distributions lacking a second moment, one can take $\hat{\sigma}$ to be the interquartile range divided by 1.349 (which equals σ for a normal distribution). Note that the same $\hat{\sigma}$ is used for all m_j, while different $\tau = \tau_j$ are used in (D.9).

Alternatively, to remove the dependence on scale through σ, we may normalise by the true $Q'(\tau_j)$ and consider the bias and variance of $\widehat{Q'}(\tau_j)/Q'(\tau_j)$, as done in the main paper's appendix.

Interestingly, the expression for m in the main paper's appendix,

$$m = n^{2/3}\left(1.5\frac{(\phi(\Phi^{-1}(\tau)))^2}{1+2(\Phi^{-1}(\tau))^2}\right)^{1/3}, \tag{D.10}$$

is very similar to the Gaussian plug-in bandwidth based on (3.1) in Hall and Sheather (1988), as in (16) of Kaplan (2015):

$$m_{HS} = n^{2/3} z_{1-\alpha/2}^{2/3}\left(1.5\frac{(\phi(\Phi^{-1}(\tau)))^2}{1+2(\Phi^{-1}(\tau))^2}\right)^{1/3}.$$

This is also the same as the bandwidth m_K proposed by Kaplan (2015) up to a constant multiple. Their suggestions are for a different setting (i.e., inference with a single, Studentised quantile), but they provide the CPE-optimal rate from our Theorems A.3 and C.2 and are scale-invariant (i.e., no σ). Consequently, we suggest using

$$m_j = n^{2/3}\left(1.5\frac{(\phi(\Phi^{-1}(\tau_j)))^2}{1+2(\Phi^{-1}(\tau_j))^2}\right)^{1/3}. \tag{D.11}$$

In simulations, our overall quantile inference method was not sensitive to either the choice of bandwidth or even the choice of sparsity estimation method (spacing estimator versus inverse of kernel estimator). The results were quite good with our quantile spacing estimators and plug-in smoothing parameters, and they were extremely similar with kernel density estimators using Silverman's (1986) rule of thumb bandwidth, which minimises the kernel density estimator's (asymptotic) mean squared error under normality.

E. IMPLEMENTATION OF METHODS

In Supplemental Appendices E.1–E.4, we provide detailed steps for implementing our methods. For practical use, computer code implementing all methods is available in the replication material from the journal's (or latter author's) website. The remaining subsections in this section discuss nuisance parameter estimation and our plug-in bandwidth; these are also implemented in the available code.

E.1. Steps for unconditional joint inference

To construct a $100(1-\alpha)\%$ confidence set for $Q(\boldsymbol{\tau}) = (Q(\tau_1), \ldots, Q(\tau_J))$:

Step 1. Parameters: determine the sample size n, the J quantile indices of interest $\tau_j \in (0,1)$, and the desired coverage level $1-\alpha$.

Step 2. Calibration of $\tilde{\alpha}$: using a numerical solver, plus simulated random variables from a beta distribution or numeric integration, solve for $\tilde{\alpha}$ in (3.3). Let $B_j^h \equiv \tilde{Q}_U^I(u_j^h(\tilde{\alpha}))$ and $B_j^l \equiv \tilde{Q}_U^I(u_j^l(\tilde{\alpha}))$, where $u_j^h(\tilde{\alpha})$ and $u_j^l(\tilde{\alpha})$ are as defined in (C.1). To explain the simulation step, define $\tilde{\mathbf{u}}$ as a $2J \times 1$ vector containing all the elements $u_j^h(\tilde{\alpha})$, $u_j^l(\tilde{\alpha})$ for $j \in \{1,\ldots,J\}$ sorted in ascending order, and let $\tilde{\mathbf{B}} = \tilde{Q}_U^I(\tilde{\mathbf{u}})$ be the corresponding vector containing elements of type B_j^l and B_j^h. We simulate draws of $\tilde{\mathbf{B}}$ according to

$$\tilde{B}_j = \tilde{B}_{j-1} + (1-\tilde{B}_{j-1})\Delta_j, \quad \Delta_j \sim \mathrm{Beta}((n+1)(\tilde{u}_j - \tilde{u}_{j-1}), (n+1)(1-\tilde{u}_j)), \quad j = 1,\ldots,2J,$$

where $\tilde{B}_0 = 0$, $\tilde{u}_0 = 0$, and the Δ_j are all independent; any alternative method of generating random Dirichlet draws (e.g., with Gamma random variables) is also fine.

For any given $\tilde{\alpha}$, many (e.g., 10^5 or 10^6) random samples can be drawn,[23] and the probability on the RHS of (3.3) is the proportion of samples in which $\{\cap_{j=1}^J \{B_j^h > \tau_j\}\} \cap \{\cap_{j=1}^J \{B_j^l < \tau_j\}\}$. The calibrated $\tilde{\alpha}$ is the value that solves (3.3), which can be found by numerical search.

Step 3. CI construction: individual $1-\tilde{\alpha}$ CIs are constructed for each $Q(\tau_j)$ as in Section 3 of Goldman and Kaplan (2017), i.e., by solving (3.1) numerically for the k_j^h and k_j^l (given $\alpha = \tilde{\alpha}$) and then plugging the solutions into (2.5) to compute the CI endpoints. The Cartesian product of the individual CIs is the overall $1-\alpha$ confidence set for the vector $Q(\boldsymbol{\tau})$.

E.2. Steps for unconditional linear combination inference

To construct one-sample $100(1-\alpha)\%$ CIs for linear combinations of quantiles:

Step 1. Parameters: determine the sample size n, the J quantile indices of interest $\tau_j \in (0,1)$, and the desired coverage level $1-\alpha$.

Step 2. Nuisance parameter estimation: using the method in Supplemental Appendix D, estimate $Q'(\tau_j)$ for all $j = 1,\ldots,J$; i.e., using the estimator in (3.5) with the formula for m_j in (D.11). Note: in simulations, using a standard kernel density estimator (evaluated at estimated quantiles $\hat{Q}_X^L(\tau_j)$) with MSE-optimal bandwidth has also worked well.

Step 3. Calibration of $\tilde{\alpha}$: using a numerical solver, plus simulated random variables from a beta distribution or numeric integration, solve for $\tilde{\alpha}$ in equation (A.7) in the main appendix for a lower one-sided CI or (A.8) for an upper one-sided

[23] Let $1-\hat{\alpha}(\tilde{\alpha})$ denote the simulated overall CP given $\tilde{\alpha}$, T the search tolerance such that the search for $\tilde{\alpha}$ stops when $|\hat{\alpha}(\tilde{\alpha}) - \alpha| < T$. With M random draws, for the $\tilde{\alpha}$ such that the true $\alpha(\tilde{\alpha}) = \alpha + K$, and assuming M is large enough to approximate the binomial distribution with the following normal,

$$\hat{\alpha}(\tilde{\alpha}) \sim \mathrm{N}(\alpha + K, (\alpha + K)(1 - \alpha - K)/M).$$

The probability of the numerical search not accepting this $\tilde{\alpha}$ is at least

$$\mathrm{P}(\hat{\alpha} > \alpha + T) = \mathrm{P}(\hat{\alpha} - \alpha - K > T - K) = 1 - \Phi\left(\frac{\sqrt{M}(T-K)}{\sqrt{(\alpha+K)(1-\alpha-K)}}\right).$$

For example, one could let $M = 10^5$ and solve for T such that this probability is 95% for $K = 0.001$.

CI. A two-sided CI is the intersection of upper and lower one-sided $1-\alpha/2$ CIs. To simulate equation (A.7) in the main appendix, for example, ψ_j, τ_j, and $\widehat{Q'}(\tau_j)$ are all known values, and let $B_j \equiv \tilde{Q}_U^I(u_j^H(\tilde{\alpha}))$, where the $u_j^H(\tilde{\alpha})$ are in ascending order (by j) and defined as in equation (A.5) of the main appendix (adjusting for negative ψ_j), which by reference to (C.1) gives u_j^H as an implicit function of $\tilde{\alpha}$. As in Appendix E.1, let $\tilde{\mathbf{u}}$ contain all the $u_j^H(\tilde{\alpha})$ values sorted into ascending order, and let $\tilde{\mathbf{B}}$ be the corresponding vector of $\tilde{Q}_U^I(u_j^H(\tilde{\alpha}))$ random variables. Then for $j=1,\ldots,J$, draws of $\tilde{B}_j$ may be simulated by

$$\tilde{B}_j = \tilde{B}_{j-1}+(1-\tilde{B}_{j-1})\Delta_j, \quad \Delta_j \sim \mathrm{Beta}((n+1)(\tilde{u}_j^H(\tilde{\alpha})-\tilde{u}_{j-1}^H(\tilde{\alpha})), (n+1)(1-\tilde{u}_j^H(\tilde{\alpha}))),$$

where again $\tilde{B}_0=0$, $\tilde{u}_0=0$, and the Δ_j are all independent, and again any alternative method of generating Dirichlet draws is fine. For any $\tilde{\alpha}$ considered, many (e.g., 10^5 or 10^6) random samples can be drawn.[23] The probability on the RHS of equation (A.7) in the main appendix is estimated by the proportion of samples in which $\sum_{j=1}^J \psi_j \widehat{Q'}(\tau_j)(B_j-\tau_j)>0$, and then $\tilde{\alpha}$ may be found by numerical search.

Step 4. CI construction: individual $1-\tilde{\alpha}$ CIs are constructed for each $Q(\tau_j)$ as in Section 3 of Goldman and Kaplan (2017) or Step 3 in Appendix E.1. The overall $1-\alpha$ CI for the linear combination contains all values of the form $\boldsymbol{\psi}'\mathbf{q}$ with $\mathbf{q}$ in the Cartesian product of the individual CIs. For a lower one-sided CI, the jth individual CI is lower one-sided if $\psi_j>0$ and upper one-sided otherwise; for an upper one-sided CI, the opposite is true. The overall two-sided CI is the intersection of the overall lower and upper one-sided CIs.

E.3. Steps for unconditional QD inference

To construct two-sample $100(1-\alpha)\%$ CIs for differences of linear combinations of quantiles:

Step 1. Parameters: determine the sample sizes n_x and n_y, the J quantile indices of interest $\tau_j \in (0,1)$, and the desired coverage level $1-\alpha$.

Step 2. Nuisance parameter estimation: using the method in Supplemental Appendix D, estimate $Q'_X(\tau_j)$ and $Q'_Y(\tau_j)$ for all $j=1,\ldots,J$; i.e., using the estimator in (3.5) with the formula for m_j in (D.11). Note: in simulations, using a standard kernel density estimator (evaluated at estimated quantiles $\hat{Q}_X^L(\tau_j)$) with MSE-optimal bandwidth has also worked well.

Step 3. Calibration of $\tilde{\alpha}$: using a numerical solver, plus simulated random variables from a beta distribution or numeric integration, solve for $\tilde{\alpha}$ in (C.50) for a lower one-sided CI or (C.51) for an upper one-sided CI. Simulation can proceed as in Appendix E.2, with the independence of the two samples allowing us to separately draw realisations from the two Dirichlet distributions. Similar to Step 2 of Appendix E.1 or Step 3 of Appendix E.2, the $\tilde{\mathbf{u}}_x^H$, $\tilde{\mathbf{u}}_y^L$, $\tilde{\mathbf{B}}_x$, and $\tilde{\mathbf{B}}_y$ contain the values of, respectively, $\mathbf{u}_x^H$, $\mathbf{u}_y^L$, $\mathbf{B}_x$, and $\mathbf{B}_y$ sorted in ascending order. In the case of lower one-sided CI we draw

$$\begin{aligned}
&\tilde{B}_{x,j} = \tilde{B}_{x,j-1} + (1 - \tilde{B}_{x,j-1})\Delta_{x,j},\\
&\Delta_{x,j} \sim \mathrm{Beta}((n_x+1)(\tilde{u}^H_{x,j}(\tilde{\alpha}) - \tilde{u}^H_{x,j-1}(\tilde{\alpha})), (n_x+1)(1-\tilde{u}^H_{x,j}(\tilde{\alpha}))),\\
&\tilde{B}_{y,j} = \tilde{B}_{y,j-1} + (1 - \tilde{B}_{y,j-1})\Delta_{y,j},\\
&\Delta_{y,j} \sim \mathrm{Beta}((n_y+1)(\tilde{u}^L_{y,j}(\tilde{\alpha}) - \tilde{u}^L_{y,j-1}(\tilde{\alpha})), (n_y+1)(1-\tilde{u}^L_{y,j}(\tilde{\alpha})))
\end{aligned}$$

where $\tilde{B}_{x,0} = \tilde{B}_{y,0} = 0$, $\tilde{u}^H_{x,0} = \tilde{u}^L_{y,0} = 0$, and the Δ are all independent; alternative methods of generating Dirichlet draws are also fine. Drawing many samples allows us to calculate the RHS of (C.50) as the proportion of samples in which

$$\sum_{j=1}^{J} \psi_j(\widehat{Q'_Y}(\tau_j)(B_{y,j} - \tau_j) - \widehat{Q'_X}(\tau_j)(B_{x,j} - \tau_j)) > 0,$$

which is implicitly a function of $\tilde{\alpha}$. Then, the calibrated value of $\tilde{\alpha}$ that solves (C.50) may be found by numerical search. For $J = 1$, the normal approximation discussed in Appendix F also yields an approximate solution to $\tilde{\alpha}$. Setting $\hat{\theta} \equiv (1 + \hat{\gamma}/\mu)/\sqrt{1 + (\hat{\gamma}/\mu)^2}$, $\tilde{\alpha} = \Phi(z_\alpha/\hat{\theta})$ for one-sided CIs or $\tilde{\alpha}/2 = \Phi(z_{\alpha/2}/\hat{\theta})$ for two-sided; $\mu \equiv \sqrt{n_y/n_x}$, and $\hat{\gamma}$ is the estimator of $f_X(Q_X(\tau))/f_Y(Q_Y(\tau))$.

STEP 4. CI construction: individual $1 - \tilde{\alpha}$ CIs are constructed for each $Q_X(\tau_j)$ and $Q_Y(\tau_j)$ as in Section 3 of Goldman and Kaplan (2017) or Step 3 in Appendix E.1. The overall $1 - \alpha$ CI for the linear combination is given by (C.49) or (C.52) or the intersection of the two. For a lower one-sided CI, the individual CI for $Q_Y(\tau_j)$ is lower one-sided if $\psi_j > 0$ and upper one-sided otherwise, while the individual CI for $Q_X(\tau_j)$ is upper one-sided if $\psi_j > 0$ and lower one-sided otherwise; for an upper one-sided CI, the opposite is true. The overall two-sided CI is the intersection of the overall lower and upper one-sided $1 - \alpha/2$ CIs.

E.4. Steps for conditional inference

To construct $100(1 - \alpha)\%$ CIs for conditional versions of the objects of interest in Supplemental Appendices E.1–E.3:

STEP 1. Discrete covariates (if applicable): for the discrete components of $\mathbf{W}$, restrict the sample to observations where the discrete components of $\mathbf{W}_i$ equal those of $\mathbf{w}_0$, the point of interest. In the following, treat this subsample as the full sample, and treat $\mathbf{W}$ as only having the remaining continuous components.

STEP 2. Bandwidth: let $p = \min_j \tau_j$, and compute the plug-in bandwidth b_{GK} from Goldman and Kaplan (2017, §4.3, p. 336). For the one-sided plug-in bandwidth, a value of α is required; for the confidence set, this can simply be the overall α, but otherwise we suggest using the $\tilde{\alpha}$ that would be computed if $N_n = n$ (using the unconditional steps above). To compute the final b, multiply b_{GK} by the following adjustment, where d is the number of (continuous) components in $\mathbf{W}$: $n^{-2/((2+d)(4+3d))}$ for two-sided joint inference; $n^{8/((12+7d)(4+3d))}$ for a one-sided linear combination CI; $n^{-2/((12+7d)(2+d))}$ for a two-sided linear combination CI; and none for one-sided joint inference. For a conditional quantile difference, do this for the $T_i = 0$ subsample to get b_0 and for the $T_i = 1$ subsample to get b_1, using the same adjustments as for linear combination CIs. With $d > 1$, there is no plug-in b_{GK}, but one could a) normalise all components of $\mathbf{W}$ to have the same variance, b) compute b_{GK} separately for each component and pick the median value, c) multiply the value from (b) by the following adjustment: $n^{3/(2+d)}$ for a two-sided CI, or $n^{7/(4+3d)}$ for a one-sided CI.

Step 3. Local sample: collect the values $\{Y_i : \|\mathbf{W}_i - \mathbf{w}_0\|_\infty \leq b\}$. For a conditional quantile difference, do this separately for the subsamples with $T_i = 0$ and $T_i = 1$, using respective bandwidths b_0 and b_1.

Step 4. CI construction: using the local sample(s), follow the steps for the corresponding unconditional CI from Supplemental Appendices E.1–E.3.

F. FURTHER APPROXIMATION AND INTUITION: TWO-SAMPLE QD

We now provide a more detailed version of the discussion at the end of Section 3.3. In the two-sample QD case with $J = 1$, which is similar to the Behrens–Fisher problem, we explore further approximations that have computational benefits and theoretical insights. The upper one-sided example is used for clarity.

Consider calibrating $\tilde{\alpha}$ by approximating the two independent beta random variables with normals. Let $B_x \equiv \mathrm{Beta}(u_x^h(n_x+1), (1-u_x^h)(n_x+1))$ and $B_y \equiv \mathrm{Beta}(u_y^l(n_y+1), (1-u_y^l)(n_y+1))$. By Theorem 2.1, convolution, $B_x \perp\!\!\!\perp B_y$, Lemma C.1, and Assumption A2.1, and omitting smaller-order remainder terms from interpolation, estimation of γ, and local linearisation of the distribution, the CP in (C.51) is

$$\begin{aligned}
\mathrm{P}(B_x - \tau < \gamma(B_y - \tau)) &= \mathrm{P}((B_x - u_x^h) - \gamma(B_y - u_y^l) < (\tau - u_x^h) + \gamma(u_y^l - \tau)) \\
&= \Phi\left(\frac{(\tau - u_x^h) + \gamma(u_y^l - \tau)}{\sqrt{\gamma^2 u_y^l(1-u_y^l)/n_y + u_x^h(1-u_x^h)/n_x}}\right) + O(n^{-1/2}\log(n)) \\
&= \Phi\left(z_{1-\tilde{\alpha}}\frac{1 + (\gamma/\delta)}{\sqrt{\frac{u_x^h(1-u_x^h)}{\tau(1-\tau)} + (\gamma/\delta)^2\frac{u_y^l(1-u_y^l)}{\tau(1-\tau)}}}\right) + O(n^{-1/2}\log(n)) \\
&= \Phi\left(z_{1-\tilde{\alpha}}\frac{1 + (\gamma/\delta)}{\sqrt{1 + (\gamma/\delta)^2}}\right) + O(n^{-1/2}\log(n)).
\end{aligned}$$

Under exchangeability (which implies $\gamma = 1$) and equal sample sizes ($\delta = 1$), $\tilde{\alpha} = \Phi(z_\alpha/\sqrt{2})$, the biggest possible value.

The calibration equation turns out to be identical for the lower one-sided case. Consequently, using $\alpha/2$ for the two one-sided cases yields the two-sided

$$\tilde{\alpha}/2 = \Phi(z_{\alpha/2}/\theta^*), \quad \theta^* \equiv \frac{1 + \gamma/\delta}{\sqrt{1 + (\gamma/\delta)^2}}.$$

G. SIMULATIONS

Code for all methods is available in the replication materials on the journal's (or latter author's) website, as is code for replicating simulation results.

G.1. Unconditional simulations

We implement our L-statistic method (abbreviated "L-stat") by using numerical integration to solve (3.12) and (3.14), using the smoothing parameter in (D.10) to estimate the nuisance parameter as in (3.5). For comparison, we consider the asymptotic normal approach with kernel-estimated standard errors ("normal"), the Edgeworth expansion-based method of Kaplan (2015), a symmetric percentile-t (Studentised) bootstrap (99 replications) using bootstrapped variance (100 replications), and the permutation test of Chung and Romano (2013) Studentised with a kernel density estimator and using 999 randomly drawn permutations. As detailed in Chung and Romano (2013), it is generally inappropriate to interpret a permutation test as a test of a specific distributional feature (such as quantile equality) due to its reliance on exchangeability. This can cause asymptotic size distortion. In contrast, their proposed test achieves asymptotically exact size without exchangeability, while retaining exact finite-sample size under exchangeability. Due to the computational burden of inverting the test into a CI, we only implement the permutation test of equality, so reported CP is one minus type I error rate and CI length is not reported for their method.

Unless otherwise noted, we consider two-sided CIs and use 10 000 simulation replications.

Table 1. CP and mean length of median difference CIs.

Method	N(0, 1)	Logistic(0,1)	Unif(0,1)	Exp(1)	LogN(0, 1)
		Coverage Probability			
L-stat	0.959	0.959	0.960	0.962	0.964
Normal	0.958	0.962	0.927	0.947	0.964
Kaplan (2015)	0.969	0.970	0.960	0.970	0.980
Bootstrap	0.953	0.951	0.950	0.955	0.959
Permutation	0.953	0.952	0.953	0.949	0.953
		Mean Interval Length			
L-stat	1.43	2.30	0.54	1.16	1.53
Normal	1.47	2.43	0.52	1.11	1.47
Kaplan (2015)	1.59	2.60	0.59	1.31	1.76
Bootstrap	1.60	2.55	0.62	1.30	1.69

Note: $1-\alpha=0.95$, $F_X=F_Y$ shown in column headers, $n_x=n_y=25$.

Table 1 shows CP of nominal 95% CIs for median differences when the two population distributions are identical and sample sizes are both 25, for a variety of distributions. Since exchangeability holds (i.e., $F_X = F_Y$), the permutation test has exact size; the (very slight) deviations from 0.95 could be decreased by using more permutations (and simulation replications). All methods meet or exceed 0.95 CP for all distributions, with some small exceptions for the normal.

In Table 1, L-stat has the shortest mean CI length in two DGPs and second-shortest in the other three. In two of the latter three, the normal CI is shortest but has CP slightly below 0.95. Furthermore, as seen in Table 6, the normal CI is the longest of all methods with $p = 0.1$ for the uniform, exponential, and log-normal distributions. Overall, the L-stat CI most consistently delivers coverage accuracy and short length.

Table 2. CP and mean length of median difference CIs.

Method	$F_X = \mathrm{N}(0,1)$ $F_Y = \mathrm{N}(0,5)$	$F_X = \mathrm{N}(0,1)$ $F_Y = t_5$	$F_X = \mathrm{Logistic}(0,1)$ $F_Y = \mathrm{Unif}(-10,10)$
		Coverage Probability	
L-stat	0.901	0.914	0.895
Normal	0.894	0.909	0.856
Kaplan (2015)	0.926	0.942	0.914
Bootstrap	0.884	0.906	0.885
Permutation (rep)[a]	0.776	0.897	0.752
Permutation (CR13)[b]	0.708	0.689	0.750
		Mean Interval Length	
L-stat	4.72	1.54	7.08
Normal	4.84	1.59	6.85
Kaplan (2015)	5.49	1.76	8.01
Bootstrap	5.32	1.68	8.31

Note: $1-\alpha = 0.90$, F_X and F_Y shown in column headers, $n_x = 13$ and $n_y = 21$.
[a] Our replication, with kernel-based Studentisation.
[b] Calculated as $1-2R$ from the one-sided, level 5% test rejection rates, R, published in Chung and Romano (2013, Table 1), noting that all distributions are symmetric.

Table 2 shows 90% CIs for DGPs with $F_X \neq F_Y$, using a subset of the DGPs in Table 1 of Chung and Romano (2013). With exchangeability violated, although asymptotic size is exact, the permutation test can have significant finite-sample size distortion. L-stat CP is very close to nominal, 89.5–91.4%. In contrast, most other methods have under-coverage with the logistic F_X and uniform F_Y: normal, 85.6%; bootstrap, 88.5%; and permutation, 75.2%. Only Kaplan (2015) always achieves nominal coverage, but with longer CIs than L-stat.

In each DGP in Table 2, the L-stat CI is shortest among methods achieving 90% CP. This is in part due to the L-stat CI not being proportional to an inverse PDF-based estimated standard error (that can be arbitrarily large), as discussed in the introduction. L-stat CIs are even shorter than those with under-coverage in some cases, such as the bootstrap in the first and third DGPs. Results are qualitatively similar (with attenuated differences) with larger sample sizes, as seen in Table 7.

G.2. Conditional simulations

Our CQD simulations use the conditional quantile function from the `rqss` vignette in Koenker (2016), where different points on the conditional quantile function represent the variety of function curvatures one may encounter; see Figure 4. All simulations have $n_1 = n_0 = 200$ where $n_t \equiv \sum_{i=1}^{n} \mathbb{1}\{T_i = t\}$, $\alpha = 0.05$, and 1000 replications each.[24]

[24] If the local sample size is too small for L-stat or even our backup approach of Kaplan (2015) to be computed at a certain w_0, that replication is discarded for that w_0 for L-stat. To be fair, we discard such replications for all methods; whether such are discarded or not makes little difference except where shown.

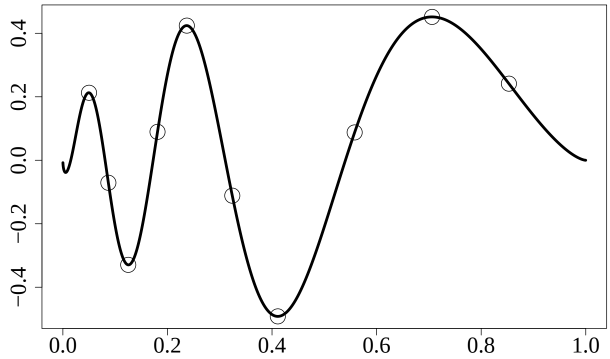

Figure 4. Conditional quantile function for simulations, with circles at w_0 values.

Let $W_i \overset{iid}{\sim} \mathrm{Unif}(0,1)$ and

$$Y_i = g(T_i)h(W_i) + \sigma(W_i)U_i, \quad h(w) = \sqrt{w(1-w)}\sin\left(\frac{2\pi(1+2^{-7/5})}{w+2^{-7/5}}\right), \tag{G.1}$$

where the U_i are sampled iid from one of several distributions satisfying $F_U(0) = \tau$, including re-centreed normal, t_3, Cauchy, χ^2_3, and $\mathrm{Unif}(-\sqrt{3}, \sqrt{3})$ distributions. There is either homoskedasticity with $\sigma(w) = 0.2$ or heteroskedasticity with $\sigma(w) = 0.2(1+w)$. Either $g(T_i) = 1$ so that the CQD is zero everywhere, or $g(T_i) = 1-2T_i$ so that the CQD is $-2h(w)$. To show performance at points where the conditional quantile function has a variety of curvatures, we alternate w_0 between local extrema of $h(w)$ and midpoints between the extrema, starting with the local maximum at $w_0 = 0.050$, and ending with $w_0 = 0.853$, as plotted in Figure 4.

We compare L-stat to local polynomial methods. The first is a local linear method with MSE-optimal bandwidth and bias correction. The second is a local cubic method with CPE-optimal bandwidth (based on our bandwidth) and no bias correction.

First, we implement a local linear method from Qu and Yoon (2015), with either a Gaussian ("QYg") or uniform ("QYu") kernel. QYg usually outperforms QYu, so we relegate the latter's results to the Supplement. We derive their CQD estimator's analytic limit distribution based on their Theorem 3 (or equation (25)), applying the modified bias correction suggested in their Remark 7 (adapted to the CQD case), using the bandwidth suggested by their Corollary 1 and equation (17). Since the simulations consider multiple w_0 values, we use a cubic spline quantile regression to estimate the second derivative at all w_0 concurrently; details of this and other plug-in estimation can be seen in the provided code.

Second, we implement a local cubic method based on Chaudhuri (1991). We use the Gaussian limit in his Proposition 4.2, estimating the standard error with a basic paired bootstrap (labeled "CBS" in our tables). Directly estimating the standard error with a kernel method ("C91" in tables) does not work as well; such results are relegated to the Supplement. Since the MSE-optimal bandwidth rate for the local cubic estimator leads to $O(1)$ CPE, we use our L-stat bandwidth multiplied by a (positive) power of n to achieve the CPE-optimal rate for the local cubic. This assumes three Lipschitz-continuous derivatives of the conditional quantile function, which is stronger than the L-stat assumption but satisfied by our DGP.

If any bandwidth (for any method) extends past the minimum or maximum observed W_i, it is truncated so that the window is not asymmetric (since an asymmetric window could increase the order of magnitude of the bias), although this rarely occurs for the DGPs we show. Again, further details may be seen in the available code.

Table 3. CP and median length for conditional median difference CIs.

		w_0 value									
Method	U_i	0.050	0.087	0.125	0.181	0.237	0.324	0.411	0.558	0.706	0.853
		Coverage Probability									
L-stat	$N(0,1)$	0.922	0.953	0.957	0.958	0.954	0.947	0.972	0.955	0.953	0.941
QYg	$N(0,1)$	0.955	0.986	0.997	0.972	1.000	0.981	1.000	0.999	0.952	0.987
CBS	$N(0,1)$	0.978	0.968	0.980	0.978	0.980	0.967	0.985	0.963	0.976	0.952
L-stat	Cauchy	0.942	0.964	0.941	0.958	0.962	0.962	0.962	0.960	0.957	0.965
QYg	Cauchy	0.989	0.990	0.971	0.990	0.985	0.993	0.975	0.998	0.974	0.982
CBS	Cauchy	0.999	0.995	0.981	0.991	0.996	0.986	0.997	0.988	0.992	0.993
L-stat	χ^2_3	0.951	0.958	0.935	0.956	0.958	0.948	0.952	0.958	0.951	0.965
QYg	χ^2_3	0.956	0.961	0.948	0.962	0.992	0.971	0.964	0.976	0.948	0.967
CBS	χ^2_3	0.986	0.977	0.954	0.973	0.973	0.974	0.970	0.972	0.970	0.975
		Median Interval Length									
L-stat	$N(0,1)$	0.545	0.532	0.515	0.594	0.514	0.576	0.465	0.450	0.366	0.348
QYg	$N(0,1)$	0.472	0.507	0.665	0.434	0.795	0.470	0.683	0.495	0.400	0.343
CBS	$N(0,1)$	0.729	0.449	0.610	0.504	0.621	0.619	0.525	0.466	0.393	0.365
L-stat	Cauchy	0.783	0.759	0.752	0.803	0.813	0.752	0.687	0.626	0.523	0.515
QYg	Cauchy	0.945	0.804	0.927	0.774	0.957	0.818	0.895	0.837	0.696	0.660
CBS	Cauchy	1.661	0.874	1.203	0.760	1.350	0.754	0.990	0.695	0.724	0.649
L-stat	χ^2_3	0.870	0.776	0.959	0.877	0.912	0.955	0.869	0.778	0.715	0.702
QYg	χ^2_3	0.677	0.684	0.760	0.606	0.857	0.630	0.794	0.631	0.622	0.556
CBS	χ^2_3	1.222	0.894	1.145	0.890	1.009	1.051	0.987	0.823	0.768	0.769

Note: $g(T_i) = 1$, $1 - \alpha = 0.95$, $n_1 = n_0 = 200$.

Table 3 shows that all methods avoid significant under-coverage for conditional median differences when $g(T_i) = 1$, but with varied lengths. Compared to QYg, L-stat is longer for χ^2_3, shorter for Cauchy, and shorter or longer for normal U_i. Compared to CBS, L-stat is almost always shorter. CBS is often longer by 10–40%, and even more in four cases.

Table 4 uses normal U_i with $\tau \in \{0.25, 0.50, 0.75\}$ and $g(T_i) = 1 - 2T_i$, so the CQD is no longer zero everywhere and there is CQD bias. At the first and third w_0, QYg (or QYg(f)) has significant under-coverage, with CP below 80% or even 40%. CP for QYu is worse (Supplemental Appendix Table 11). Recall from Figure 4 that both those points are local extrema where the second derivative of the conditional quantile function is large. For this particular DGP, it seems that the use of a CPE-optimal bandwidth (which CBS also uses) works better than the modified bias correction and MSE-optimal bandwidth (as in QYg). L-stat is far less affected than QYg, although at the smallest w_0 its coverage falls below 90%. CBS comes closest to correct coverage.

Table 4 shows CI lengths similar to those in Table 3, but with an extra advantage for symmetric (instead of equal-tailed) CIs when $\tau \neq 0.5$. Despite QYg having some under-

Table 4. CP and median length for CQD CIs.

		w_0 value									
Method	τ	0.050	0.087	0.125	0.181	0.237	0.324	0.411	0.558	0.706	0.853
		Coverage Probability									
L-stat	0.25	0.851	0.959	0.915	0.948	0.954	0.955	0.955	0.952	0.942	0.959
QYg	0.25	0.273	0.991	0.763	0.901	0.981	0.897	0.999	0.987	0.942	0.982
QYg(f)	0.25	0.342	0.991	0.779	0.923	0.983	0.926	0.999	0.988	0.942	0.982
CBS	0.25	0.936	0.968	0.929	0.980	0.967	0.966	0.962	0.976	0.968	0.965
L-stat	0.50	0.895	0.965	0.926	0.953	0.951	0.952	0.954	0.957	0.940	0.946
QYg	0.50	0.321	0.990	0.885	0.929	0.998	0.952	0.999	0.999	0.928	0.987
QYg(f)	0.50	0.343	0.990	0.887	0.932	0.998	0.963	0.999	0.999	0.928	0.987
CBS	0.50	0.976	0.970	0.977	0.966	0.980	0.962	0.977	0.969	0.962	0.957
L-stat	0.75	0.869	0.956	0.928	0.962	0.952	0.969	0.955	0.957	0.931	0.948
QYg	0.75	0.212	0.985	0.756	0.881	0.970	0.868	0.999	0.976	0.953	0.989
QYg(f)	0.75	0.293	0.987	0.781	0.916	0.976	0.915	0.999	0.979	0.953	0.989
CBS	0.75	0.935	0.962	0.938	0.974	0.973	0.977	0.969	0.962	0.956	0.955
		Median Interval Length									
L-stat	0.25	0.638	0.638	0.607	0.823	0.660	0.760	0.526	0.576	0.407	0.405
QYg	0.25	0.493	0.573	0.727	0.436	0.871	0.462	0.767	0.517	0.458	0.354
QYg(f)	0.25	0.499	0.576	0.730	0.438	0.873	0.459	0.770	0.519	0.458	0.354
CBS	0.25	0.693	0.558	0.597	0.696	0.611	0.687	0.526	0.550	0.419	0.413
L-stat	0.50	0.545	0.540	0.511	0.583	0.526	0.591	0.459	0.461	0.366	0.355
QYg	0.50	0.514	0.587	0.780	0.436	0.916	0.472	0.789	0.543	0.454	0.336
QYg(f)	0.50	0.514	0.587	0.780	0.437	0.917	0.477	0.791	0.543	0.454	0.336
CBS	0.50	0.750	0.452	0.609	0.498	0.630	0.610	0.519	0.475	0.396	0.371
L-stat	0.75	0.627	0.632	0.672	0.918	0.721	0.813	0.534	0.572	0.409	0.398
QYg	0.75	0.495	0.570	0.736	0.443	0.857	0.460	0.770	0.512	0.470	0.355
QYg(f)	0.75	0.497	0.578	0.736	0.444	0.857	0.466	0.775	0.517	0.470	0.355
CBS	0.75	0.672	0.549	0.584	0.678	0.605	0.654	0.521	0.521	0.419	0.414

Note: Gaussian U_i, $g(T_i) = 1 - 2T_i$ so that CQD$= -2h(w_0)$, $1 - \alpha = 0.95$, $n_1 = n_0 = 200$. "QYg(f)" is QYg on all replications instead of just those where L-stat is computable.

coverage, L-stat's CIs are still shorter at most of the local extrema (and some midpoints when $\tau = 0.5$). -stat CIs are generally shorter than CBS CIs for $\tau = 0.5$, and the reverse for other τ, although there are exceptions in both directions.

Table 5 shows L-stat and QYg results for conditional IQR inference, where $Y_i = h(W_i) + 0.2U_i$ and $F_U(0) = 0.5$. (CBS cannot be used since it relies on results in Chaudhuri (1991), which only allows a single quantile.) QYg can have severe under-coverage at the fourth and sixth w_0, whereas the very worst L-stat coverage is 89.3%, and next-worst is 93.0%, among the 40 values shown. Even with superior coverage, L-stat has shorter CIs at three w_0 with normal U_i, although QYg intervals are almost always shorter for other U_i distributions.

Among all of Tables 3–5, L-stat CP is never below 85% and is above 92% in 95/100 cases (10 DGPs, 10 w_0 each). In contrast, QYg CP can go below 50% (for both CQD and

Table 5. CP and median length for conditional IQR CIs.

Method	U_i	w_0 value 0.050	0.087	0.125	0.181	0.237	0.324	0.411	0.558	0.706	0.853
		Coverage Probability									
L-stat	N(0, 1)	0.980	0.964	0.979	0.974	0.981	0.966	0.977	0.953	0.968	0.966
QYg	N(0, 1)	0.906	0.919	0.989	0.352	1.000	0.656	1.000	0.952	0.978	0.963
QYg(f)	N(0, 1)	0.921	0.925	0.990	0.428	1.000	0.710	1.000	0.954	0.978	0.963
L-stat	t_3	0.986	0.935	0.981	0.966	0.975	0.964	0.968	0.952	0.979	0.970
QYg	t_3	0.892	0.917	0.967	0.441	0.998	0.709	1.000	0.950	0.972	0.952
QYg(f)	t_3	0.901	0.921	0.970	0.479	0.998	0.737	1.000	0.951	0.972	0.952
L-stat	Cauchy	0.978	0.893	0.969	0.930	0.974	0.966	0.977	0.934	0.965	0.973
QYg	Cauchy	0.940	0.887	0.921	0.804	0.951	0.892	0.934	0.954	0.920	0.935
QYg(f)	Cauchy	0.942	0.889	0.928	0.824	0.952	0.901	0.936	0.955	0.920	0.936
L-stat	χ^2_3	0.948	0.945	0.968	0.963	0.963	0.959	0.952	0.951	0.967	0.953
QYg	χ^2_3	0.905	0.952	0.972	0.912	0.979	0.879	0.968	0.933	0.977	0.939
QYg(f)	χ^2_3	0.909	0.952	0.973	0.915	0.979	0.882	0.969	0.933	0.977	0.939
		Median Interval Length									
L-stat	N(0, 1)	0.379	0.422	0.408	0.451	0.362	0.405	0.350	0.324	0.251	0.278
QYg	N(0, 1)	0.282	0.292	0.424	0.244	0.491	0.270	0.402	0.274	0.242	0.198
QYg(f)	N(0, 1)	0.284	0.293	0.422	0.245	0.497	0.271	0.402	0.275	0.242	0.198
L-stat	t_3	0.482	0.487	0.539	0.560	0.451	0.532	0.463	0.409	0.308	0.357
QYg	t_3	0.326	0.333	0.438	0.281	0.510	0.303	0.432	0.306	0.273	0.239
QYg(f)	t_3	0.327	0.334	0.439	0.281	0.511	0.303	0.434	0.306	0.273	0.239
L-stat	Cauchy	0.902	0.856	1.245	1.158	0.786	1.088	0.968	0.789	0.599	0.647
QYg	Cauchy	0.626	0.550	0.630	0.519	0.627	0.507	0.559	0.503	0.467	0.440
QYg(f)	Cauchy	0.628	0.549	0.632	0.518	0.627	0.508	0.559	0.504	0.467	0.439
L-stat	χ^2_3	0.766	0.640	0.796	0.769	0.692	0.742	0.709	0.643	0.593	0.614
QYg	χ^2_3	0.467	0.457	0.543	0.412	0.610	0.401	0.502	0.406	0.445	0.390
QYg(f)	χ^2_3	0.466	0.457	0.543	0.413	0.610	0.401	0.502	0.406	0.445	0.390

Note: $1 - \alpha = 0.95$, $n = 400$.

conditional IQR) and is above 92% in only 81/100 cases. Even with superior CP, L-stat CIs are often shorter than QYg CIs, as in 14/30 cases in Table 3.

For CQD inference (Tables 3 and 4), CBS CP is always above 92%, but L-stat generally has shorter CIs, as in 27/30 cases in Table 3. Away from the median, L-stat CIs are more often longer than CBS CIs since the former are equal-tailed while the latter are symmetric (by construction), but L-stat is still shorter (or tied) in 7/20 cases with $\tau \neq 0.5$ in Table 4. Although L-stat does not strictly dominate CBS (which also uses a variant of our newly proposed bandwidth), it provides much shorter CIs near $\tau = 0.5$ with only slightly worse CP in a few cases, and L-stat's equal-tailed property may also be preferred.[25]

A CQD simulation based on the empirical application in Section 5 is described in Appendix G.4. In Table 14, L-stat has CP between 93.8% and 96.4% among 18 combinations

[25] Viewing the CI as an interval estimator, equal-tailedness is equivalent to median-unbiasedness; see for example Footnote 11 in Andrews and Guggenberger (2014).

of w_0 and τ. Compared to the worst-case L-stat CP of 93.8%, CBS had CP below 93.8% in 13/18 cases, and QYg in 12/18 cases. The L-stat CIs are naturally longer since their CP is higher (more correct), but recall that in Table 3, when L-stat and CBS both have correct CP, L-stat usually has shorter length. The primary result in Table 14 not found in Tables 3 and 4 is that CBS may under-cover in cases where L-stat does not.

G.3. Additional unconditional simulations

Tables 6 and 7 show additional unconditional results.

Table 6. Mean length of 0.1-quantile difference CIs.

	$\mathrm{N}(0,1)$	Logistic(0,1)	Unif(0,1)	Exp(1)	$\mathrm{LogN}(0,1)$
L-stat	2.58	5.71	0.38	0.45	0.64
Normal	2.13	4.68[a]	0.45	0.77	1.02
Kaplan (2015)	2.96	6.44	0.43	0.51	0.73
Bootstrap	2.54	5.53	0.38	0.44	0.63

Note: $1-\alpha=0.95$, $F_X=F_Y$ shown in column headers, $n_x=n_y=25$.
[a] With 2000 replications. With 10 000 replications, one CI has "infinite" (to the computer's precision) length, so the mean length is infinity.

Table 7. CP and mean length of median difference CIs.

	$F_X=\mathrm{N}(0,1)$ $F_Y=\mathrm{N}(0,5)$	$F_X=\mathrm{N}(0,1)$ $F_Y=t_5$	$F_X=\mathrm{Logistic}(0,1)$ $F_Y=\mathrm{Unif}(-10,10)$
		Coverage Probability	
L-stat	0.899	0.902	0.896
Normal	0.910	0.912	0.887
Kaplan (2015)	0.911	0.919	0.902
Bootstrap	0.893	0.895	0.890
Permutation (rep)[a]	0.809	0.893	0.778
Permutation (CR13)[b]	0.810	0.819	0.815
		Mean Interval Length	
L-stat	2.14	0.73	3.36
Normal	2.24	0.77	3.35
Kaplan (2015)	2.27	0.78	3.53
Bootstrap	2.34	0.78	3.73

Note: $1-\alpha=0.90$, F_X and F_Y shown in column headers, $n_x=51$ and $n_y=101$.
[a] Our replication, with kernel-based Studentisation.
[b] Calculated as $1-2R$ from the one-sided, level 5% test rejection rates, R, published in Chung and Romano (2013, Table 1), noting that all distributions are symmetric.

G.4. Additional conditional simulations

Tables 8 and 9 are (together) an expanded version of Table 3 from the main text, Tables 11 and 12 an expansion of Table 4, and Table 13 an expansion of Table 5. Table 10 is new.

Table 8. CP for conditional median difference CIs.

		w_0 value									
	U_i	0.050	0.087	0.125	0.181	0.237	0.324	0.411	0.558	0.706	0.853
		Coverage Probability									
L-stat	N(0, 1)	0.922	0.953	0.957	0.958	0.954	0.947	0.972	0.955	0.953	0.941
QYg	N(0, 1)	0.955	0.986	0.997	0.972	1.000	0.981	1.000	0.999	0.952	0.987
CBS	N(0, 1)	0.978	0.968	0.980	0.978	0.980	0.967	0.985	0.963	0.976	0.952
QYg(f)	N(0, 1)	0.959	0.986	0.997	0.974	1.000	0.985	1.000	0.999	0.952	0.987
C91	N(0, 1)	0.989	0.991	0.992	0.999	0.986	0.997	0.988	0.994	0.978	0.980
QYu	N(0, 1)	0.947	0.982	0.994	0.962	1.000	0.984	1.000	1.000	0.929	0.992
L-stat	t_3	0.940	0.950	0.951	0.953	0.965	0.966	0.957	0.961	0.955	0.964
QYg	t_3	0.971	0.985	0.989	0.974	0.999	0.991	1.000	0.999	0.951	0.993
CBS	t_3	0.989	0.979	0.977	0.973	0.986	0.979	0.983	0.965	0.971	0.970
QYg(f)	t_3	0.972	0.985	0.989	0.974	0.999	0.991	1.000	0.999	0.951	0.993
C91	t_3	0.993	0.997	0.992	0.999	0.995	0.999	0.991	0.998	0.992	0.992
QYu	t_3	0.970	0.984	0.987	0.961	1.000	0.985	1.000	0.996	0.923	0.991
L-stat	Cauchy	0.942	0.964	0.941	0.958	0.962	0.962	0.962	0.960	0.957	0.965
QYg	Cauchy	0.989	0.990	0.971	0.990	0.985	0.993	0.975	0.998	0.974	0.982
CBS	Cauchy	0.999	0.995	0.981	0.991	0.996	0.986	0.997	0.988	0.992	0.993
QYg(f)	Cauchy	0.989	0.990	0.971	0.990	0.985	0.993	0.975	0.998	0.974	0.982
C91	Cauchy	0.998	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000
QYu	Cauchy	0.985	0.986	0.963	0.987	0.977	0.992	0.955	0.994	0.954	0.985
L-stat	χ^2_3	0.951	0.958	0.935	0.956	0.958	0.948	0.952	0.958	0.951	0.965
QYg	χ^2_3	0.956	0.961	0.948	0.962	0.992	0.971	0.964	0.976	0.948	0.967
CBS	χ^2_3	0.986	0.977	0.954	0.973	0.973	0.974	0.970	0.972	0.970	0.975
QYg(f)	χ^2_3	0.956	0.961	0.949	0.962	0.992	0.972	0.964	0.976	0.948	0.967
C91	χ^2_3	0.996	0.987	0.950	0.994	0.994	0.993	0.977	0.989	0.983	0.977
QYu	χ^2_3	0.955	0.961	0.946	0.975	0.990	0.971	0.955	0.980	0.927	0.963
L-stat	Uniform	0.944	0.967	0.956	0.959	0.955	0.954	0.965	0.946	0.943	0.952
QYg	Uniform	0.957	0.988	0.995	0.975	1.000	0.987	0.999	0.992	0.901	0.966
CBS	Uniform	0.963	0.970	0.960	0.967	0.971	0.968	0.968	0.962	0.932	0.950
QYg(f)	Uniform	0.961	0.988	0.995	0.976	1.000	0.990	0.999	0.992	0.901	0.966
C91	Uniform	0.966	0.991	0.974	0.997	0.974	0.996	0.959	0.983	0.943	0.957
QYu	Uniform	0.949	0.986	0.993	0.966	1.000	0.977	0.999	0.993	0.893	0.976

Note: $g(T_i) = 1$, $1 - \alpha = 0.95$, $n_1 = n_0 = 200$. "QYg(f)" is QYg on all replications instead of just those where L-stat may be used. The number of replications where L-stat was used for normal U_i, by w_0 in ascending order, was 902, 996, 980, 945, 966, 794, 984, 966, 999, 1000; for t_3, 949, 999, 991, 988, 989, 958, 994, 995, 998, 998; for Cauchy, 986, 999, 996, 995, 993, 995, 1000, 995, 998, 1000; for χ^2_3, 990, 997, 987, 996, 999, 980, 993, 997, 1000, 997; for uniform, 893, 994, 982, 946, 973, 777, 993, 963, 1000, 1000.

Table 14 shows results for a new simulation DGP based on the empirical application.

Table 9. Median CI lengths, corresponding to Table 8.

		w_0 value									
	U_i	0.050	0.087	0.125	0.181	0.237	0.324	0.411	0.558	0.706	0.853
		Median Interval Length									
L-stat	$\mathrm{N}(0,1)$	0.545	0.532	0.515	0.594	0.514	0.576	0.465	0.450	0.366	0.348
QYg	$\mathrm{N}(0,1)$	0.472	0.507	0.665	0.434	0.795	0.470	0.683	0.495	0.400	0.343
CBS	$\mathrm{N}(0,1)$	0.729	0.449	0.610	0.504	0.621	0.619	0.525	0.466	0.393	0.365
QYg(f)	$\mathrm{N}(0,1)$	0.472	0.508	0.665	0.435	0.795	0.473	0.684	0.497	0.400	0.343
C91	$\mathrm{N}(0,1)$	0.719	0.592	0.723	0.812	0.621	0.857	0.523	0.651	0.399	0.422
QYu	$\mathrm{N}(0,1)$	0.476	0.512	0.671	0.438	0.802	0.475	0.690	0.500	0.404	0.346
L-stat	t_3	0.588	0.592	0.568	0.656	0.599	0.625	0.526	0.501	0.405	0.397
QYg	t_3	0.536	0.561	0.698	0.489	0.811	0.512	0.727	0.544	0.445	0.402
CBS	t_3	0.825	0.531	0.690	0.521	0.732	0.596	0.612	0.498	0.449	0.429
QYg(f)	t_3	0.536	0.561	0.698	0.490	0.811	0.513	0.727	0.544	0.445	0.401
C91	t_3	0.885	0.682	0.861	0.840	0.781	0.921	0.659	0.735	0.503	0.523
QYu	t_3	0.541	0.567	0.705	0.494	0.818	0.517	0.733	0.549	0.449	0.405
L-stat	Cauchy	0.783	0.759	0.752	0.803	0.813	0.752	0.687	0.626	0.523	0.515
QYg	Cauchy	0.945	0.804	0.927	0.774	0.957	0.818	0.895	0.837	0.696	0.660
CBS	Cauchy	1.661	0.874	1.203	0.760	1.350	0.754	0.990	0.695	0.724	0.649
QYg(f)	Cauchy	0.945	0.804	0.927	0.774	0.957	0.818	0.895	0.837	0.696	0.660
C91	Cauchy	2.586	2.127	2.302	2.137	2.396	2.052	2.234	2.110	2.021	2.014
QYu	Cauchy	0.954	0.812	0.935	0.781	0.965	0.826	0.903	0.845	0.703	0.666
L-stat	χ^2_3	0.870	0.776	0.959	0.877	0.912	0.955	0.869	0.778	0.715	0.702
QYg	χ^2_3	0.677	0.684	0.760	0.606	0.857	0.630	0.794	0.631	0.622	0.556
CBS	χ^2_3	1.222	0.894	1.145	0.890	1.009	1.051	0.987	0.823	0.768	0.769
QYg(f)	χ^2_3	0.677	0.685	0.759	0.606	0.857	0.630	0.795	0.631	0.622	0.556
C91	χ^2_3	1.233	0.898	1.085	0.996	1.089	1.181	0.889	0.926	0.783	0.758
QYu	χ^2_3	0.683	0.691	0.767	0.612	0.865	0.636	0.801	0.637	0.628	0.561
L-stat	Uniform	0.602	0.522	0.602	0.610	0.604	0.637	0.551	0.537	0.443	0.430
QYg	Uniform	0.463	0.513	0.648	0.441	0.777	0.463	0.686	0.496	0.416	0.360
CBS	Uniform	0.771	0.520	0.679	0.588	0.694	0.733	0.589	0.588	0.460	0.453
QYg(f)	Uniform	0.465	0.513	0.649	0.442	0.779	0.465	0.686	0.496	0.416	0.360
C91	Uniform	0.738	0.595	0.738	0.805	0.672	0.866	0.558	0.681	0.441	0.456
QYu	Uniform	0.467	0.518	0.654	0.445	0.784	0.467	0.692	0.501	0.420	0.363

We focused on food budget share for 2-adult, 2-child households and 1-adult, 1-child households. Separately for the two subsamples, we fit a Gaussian mixture model to the data (using R package `mclust`), using a mixture of 10 Gaussian distributions for the larger subsample and seven for the smaller subsample (with no restrictions placed on the component distributions' means or covariances). These Gaussian mixture distributions were taken to be the "true" distributions/DGP for the simulation. As in the application, we used the empirical deciles (for each simulated dataset) for the nine w_0 values. For any w_0, integrating the conditional PDF from the mixture Gaussian distribution determines the "true" (for the simulation DGP) conditional quantile. We ran 1000 replications, sampling a new dataset from the Gaussian mixtures each time (with `mclust::sim`), which

Table 10. CP and median length for CQD CIs.

	τ	w_0 value: 0.050	0.087	0.125	0.181	0.237	0.324	0.411	0.558	0.706	0.853
						Coverage Probability					
L-stat	0.25	0.938	0.956	0.958	0.969	0.966	0.948	0.950	0.942	0.948	0.939
QYg	0.25	0.948	0.996	0.998	0.920	0.994	0.951	1.000	0.997	0.938	0.975
CBS	0.25	0.968	0.964	0.957	0.976	0.981	0.970	0.961	0.973	0.966	0.956
QYg(f)	0.25	0.953	0.996	0.998	0.936	0.994	0.962	1.000	0.997	0.938	0.975
C91	0.25	0.921	0.987	1.000	0.994	0.965	0.990	0.995	0.986	0.941	0.971
QYu	0.25	0.905	0.991	0.996	0.908	0.985	0.939	1.000	1.000	0.917	0.976
L-stat	0.75	0.934	0.948	0.966	0.969	0.964	0.951	0.971	0.956	0.955	0.960
QYg	0.75	0.951	0.963	0.962	0.965	1.000	0.906	0.998	0.977	0.992	0.979
CBS	0.75	0.959	0.972	0.964	0.977	0.955	0.957	0.965	0.970	0.963	0.970
QYg(f)	0.75	0.972	0.965	0.966	0.977	1.000	0.946	0.998	0.979	0.992	0.979
C91	0.75	0.996	0.989	0.947	0.992	0.991	0.986	0.946	0.988	0.984	0.979
QYu	0.75	0.938	0.963	0.942	0.958	1.000	0.923	0.995	0.971	0.989	0.986
						Median Interval Length					
L-stat	0.25	0.651	0.689	0.579	0.880	0.722	0.756	0.516	0.597	0.414	0.400
QYg	0.25	0.468	0.552	0.748	0.423	0.575	0.438	0.811	0.513	0.337	0.368
CBS	0.25	0.698	0.594	0.565	0.708	0.642	0.669	0.524	0.567	0.427	0.407
QYg(f)	0.25	0.468	0.554	0.753	0.424	0.577	0.439	0.815	0.515	0.337	0.368
C91	0.25	0.570	0.688	0.904	0.816	0.516	0.765	0.634	0.612	0.368	0.421
QYu	0.25	0.472	0.557	0.755	0.427	0.580	0.442	0.818	0.518	0.340	0.371
L-stat	0.75	0.641	0.642	0.741	0.896	0.628	0.828	0.589	0.539	0.388	0.426
QYg	0.75	0.477	0.460	0.527	0.443	0.904	0.478	0.509	0.440	0.480	0.349
CBS	0.75	0.654	0.521	0.624	0.677	0.587	0.670	0.547	0.497	0.407	0.426
QYg(f)	0.75	0.482	0.460	0.527	0.449	0.919	0.484	0.510	0.441	0.480	0.350
C91	0.75	0.872	0.602	0.553	0.816	0.791	0.754	0.456	0.593	0.475	0.432
QYu	0.75	0.481	0.464	0.532	0.447	0.912	0.482	0.514	0.444	0.484	0.352

Note: normal U_i, $g(T_i) = 1$, $1 - \alpha = 0.95$, $n_1 = n_0 = 200$. "QYg(f)" is QYg on all replications instead of just those where L-stat may be used. The number of replications where L-stat was used for $\tau = 0.25$, by w_0 in ascending order, was 902, 911, 837, 786, 939, 775, 926, 920, 997, 994; for $\tau = 0.75$, 558, 940, 893, 654, 778, 547, 950, 904, 987, 983.

also determines slightly different w_0 each time. Our method (L-stat) has CP between 93.8% and 96.4% across all w_0 and τ. The other methods have more under-coverage: across all w_0 and τ, the range of CP for CBS is $[0.916, 0.948]$, and the CP range for QYg is $[0.887, 0.974]$. Compared to the worst L-stat CP of 93.8%, CBS had CP below 93.8% in 13/18 cases, and QYg had CP below 93.8% in 12/18 cases. The trade-off is that the L-stat CIs are always longest in this case. Of course, if one tried to make the CBS and QYg intervals long enough to attain correct coverage (like L-stat), then they may get long enough that L-stat is the shortest.

Table 11. CP and median length for CQD CIs.

	τ	0.050	0.087	0.125	0.181	0.237	0.324	0.411	0.558	0.706	0.853
						w_0 value					
						Coverage Probability					
L-stat	0.25	0.851	0.959	0.915	0.948	0.954	0.955	0.955	0.952	0.942	0.959
QYg	0.25	0.273	0.991	0.763	0.901	0.981	0.897	0.999	0.987	0.942	0.982
CBS	0.25	0.936	0.968	0.929	0.980	0.967	0.966	0.962	0.976	0.968	0.965
QYg(f)	0.25	0.342	0.991	0.779	0.923	0.983	0.926	0.999	0.988	0.942	0.982
C91	0.25	0.936	0.991	0.958	0.988	0.976	0.981	0.984	0.991	0.970	0.975
QYu	0.25	0.202	0.990	0.623	0.910	0.970	0.902	1.000	0.982	0.929	0.973
L-stat	0.50	0.895	0.965	0.926	0.953	0.951	0.952	0.954	0.957	0.940	0.946
QYg	0.50	0.321	0.990	0.885	0.929	0.998	0.952	0.999	0.999	0.928	0.987
CBS	0.50	0.976	0.970	0.977	0.966	0.980	0.962	0.977	0.969	0.962	0.957
QYg(f)	0.50	0.343	0.990	0.887	0.932	0.998	0.963	0.999	0.999	0.928	0.987
C91	0.50	0.967	0.997	0.980	0.997	0.983	0.997	0.981	0.993	0.974	0.985
QYu	0.50	0.248	0.990	0.769	0.913	0.994	0.939	0.999	0.998	0.916	0.990
L-stat	0.75	0.869	0.956	0.928	0.962	0.952	0.969	0.955	0.957	0.931	0.948
QYg	0.75	0.212	0.985	0.756	0.881	0.970	0.868	0.999	0.976	0.953	0.989
CBS	0.75	0.935	0.962	0.938	0.974	0.973	0.977	0.969	0.962	0.956	0.955
QYg(f)	0.75	0.293	0.987	0.781	0.916	0.976	0.915	0.999	0.979	0.953	0.989
C91	0.75	0.929	0.980	0.951	0.980	0.989	0.990	0.958	0.985	0.975	0.970
QYu	0.75	0.161	0.987	0.633	0.891	0.967	0.888	0.999	0.978	0.951	0.987

Note: normal U_i, $g(T_i) = 1 - 2T_i$; $1 - \alpha = 0.95$, $n_1 = n_0 = 200$. "QYg(f)" is QYg on all replications instead of just those where L-stat may be used. The number of replications where L-stat was used for $\tau = 0.25$, by w_0 in ascending order, was 760, 962, 909, 757, 887, 694, 970, 923, 992, 995; for $\tau = 0.50$, 901, 994, 972, 948, 973, 764, 988, 961, 1000, 1000; for $p = 0.75$, 672, 855, 856, 678, 806, 614, 936, 868, 992, 990.

H. ADDITIONAL EMPIRICAL RESULTS

H.1. Quantile Engel curve differences

Figure 5 contains additional analysis.

H.2. Gift exchange

To demonstrate our two-sample QTE inference, we use the experimental data from Gneezy and List (2006, Tables I and V). In short, individuals in the control group work for a certain advertised hourly wage, while individuals in the treatment group are surprised with a larger hourly wage upon arrival. The goal is to investigate "gift exchange," specifically whether the higher wages induce higher effort (as measured by productivity). The experiment is run separately for two different tasks: data entry for a library (typing in a book's author, title, etc.), and door-to-door fundraising for a non-profit. Productivity is measured for each participant, as the number of books entered in each of four 90-minute segments for the library task, and as dollars raised before/after lunch for fundraising. The sample sizes are small: 10 and 9 for control and treatment, respectively, for the library task, and 10 and 13 for control and treatment for fundraising.

The main result of Gneezy and List (2006) is that the "gift wage" treatment raises

Table 12. Median CI lengths corresponding to Table 11.

		w_0 value									
	τ	0.050	0.087	0.125	0.181	0.237	0.324	0.411	0.558	0.706	0.853
		Median Interval Length									
L-stat	0.25	0.638	0.638	0.607	0.823	0.660	0.760	0.526	0.576	0.407	0.405
QYg	0.25	0.493	0.573	0.727	0.436	0.871	0.462	0.767	0.517	0.458	0.354
CBS	0.25	0.693	0.558	0.597	0.696	0.611	0.687	0.526	0.550	0.419	0.413
QYg(f)	0.25	0.499	0.576	0.730	0.438	0.873	0.459	0.770	0.519	0.458	0.354
C91	0.25	0.716	0.649	0.724	0.821	0.645	0.762	0.559	0.619	0.408	0.425
QYu	0.25	0.498	0.578	0.734	0.440	0.879	0.467	0.775	0.522	0.462	0.357
L-stat	0.50	0.545	0.540	0.511	0.583	0.526	0.591	0.459	0.461	0.366	0.355
QYg	0.50	0.514	0.587	0.780	0.436	0.916	0.472	0.789	0.543	0.454	0.336
CBS	0.50	0.750	0.452	0.609	0.498	0.630	0.610	0.519	0.475	0.396	0.371
QYg(f)	0.50	0.514	0.587	0.780	0.437	0.917	0.477	0.791	0.543	0.454	0.336
C91	0.50	0.723	0.603	0.719	0.808	0.628	0.856	0.519	0.662	0.401	0.423
QYu	0.50	0.519	0.592	0.787	0.440	0.925	0.476	0.797	0.548	0.458	0.339
L-stat	0.75	0.627	0.632	0.672	0.918	0.721	0.813	0.534	0.572	0.409	0.398
QYg	0.75	0.495	0.570	0.736	0.443	0.857	0.460	0.770	0.512	0.470	0.355
CBS	0.75	0.672	0.549	0.584	0.678	0.605	0.654	0.521	0.521	0.419	0.414
QYg(f)	0.75	0.497	0.578	0.736	0.444	0.857	0.466	0.775	0.517	0.470	0.355
C91	0.75	0.711	0.622	0.723	0.798	0.664	0.771	0.514	0.600	0.435	0.421
QYu	0.75	0.500	0.576	0.743	0.447	0.864	0.465	0.777	0.517	0.474	0.358

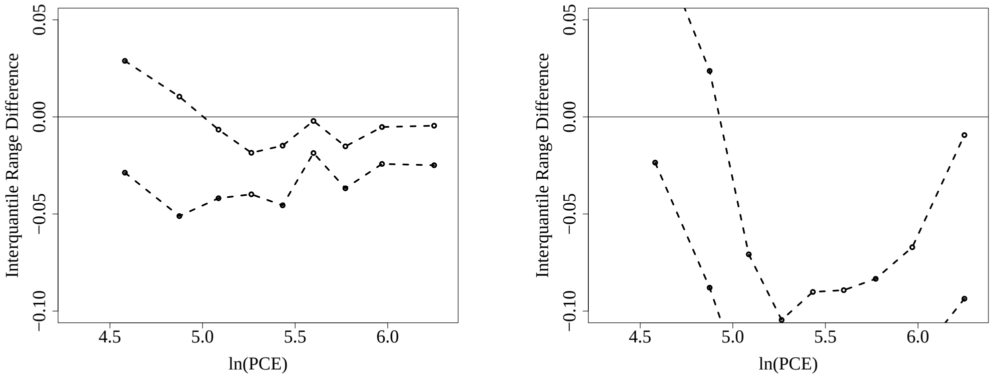

Figure 5. Pointwise 90% confidence intervals (connected for visual ease) for conditional interquantile range (0.9-quantile minus median) differences in alcohol (left) or housing and utilities (right) budget share between two-adult and one-adult childless households, conditional on real log total per capita expenditure.

Table 13. CP and median length for conditional IQR CIs.

		w_0 value									
	U_i	0.050	0.087	0.125	0.181	0.237	0.324	0.411	0.558	0.706	0.853
		Coverage Probability									
L-stat	N(0, 1)	0.980	0.964	0.979	0.974	0.981	0.966	0.977	0.953	0.968	0.966
QYg	N(0, 1)	0.906	0.919	0.989	0.352	1.000	0.656	1.000	0.952	0.978	0.963
QYg(f)	N(0, 1)	0.921	0.925	0.990	0.428	1.000	0.710	1.000	0.954	0.978	0.963
QYu	N(0, 1)	0.857	0.951	0.985	0.462	1.000	0.750	1.000	0.965	0.979	0.968
L-stat	t_3	0.986	0.935	0.981	0.966	0.975	0.964	0.968	0.952	0.979	0.970
QYg	t_3	0.892	0.917	0.967	0.441	0.998	0.709	1.000	0.950	0.972	0.952
QYg(f)	t_3	0.901	0.921	0.970	0.479	0.998	0.737	1.000	0.951	0.972	0.952
QYu	t_3	0.871	0.937	0.986	0.551	1.000	0.791	1.000	0.961	0.975	0.952
L-stat	Cauchy	0.978	0.893	0.969	0.930	0.974	0.966	0.977	0.934	0.965	0.973
QYg	Cauchy	0.940	0.887	0.921	0.804	0.951	0.892	0.934	0.954	0.920	0.935
QYg(f)	Cauchy	0.942	0.889	0.928	0.824	0.952	0.901	0.936	0.955	0.920	0.936
QYu	Cauchy	0.932	0.890	0.897	0.846	0.929	0.901	0.944	0.938	0.909	0.914
L-stat	χ^2_3	0.948	0.945	0.968	0.963	0.963	0.959	0.952	0.951	0.967	0.953
QYg	χ^2_3	0.905	0.952	0.972	0.912	0.979	0.879	0.968	0.933	0.977	0.939
QYg(f)	χ^2_3	0.909	0.952	0.973	0.915	0.979	0.882	0.969	0.933	0.977	0.939
QYu	χ^2_3	0.904	0.953	0.962	0.914	0.982	0.917	0.966	0.954	0.977	0.951
		Median Interval Length									
L-stat	N(0, 1)	0.379	0.422	0.408	0.451	0.362	0.405	0.350	0.324	0.251	0.278
QYg	N(0, 1)	0.282	0.292	0.424	0.244	0.491	0.270	0.402	0.274	0.242	0.198
QYg(f)	N(0, 1)	0.284	0.293	0.422	0.245	0.497	0.271	0.402	0.275	0.242	0.198
QYu	N(0, 1)	0.285	0.295	0.428	0.246	0.496	0.272	0.406	0.277	0.244	0.200
L-stat	t_3	0.482	0.487	0.539	0.560	0.451	0.532	0.463	0.409	0.308	0.357
QYg	t_3	0.326	0.333	0.438	0.281	0.510	0.303	0.432	0.306	0.273	0.239
QYg(f)	t_3	0.327	0.334	0.439	0.281	0.511	0.303	0.434	0.306	0.273	0.239
QYu	t_3	0.329	0.336	0.442	0.283	0.515	0.306	0.436	0.308	0.275	0.241
L-stat	Cauchy	0.902	0.856	1.245	1.158	0.786	1.088	0.968	0.789	0.599	0.647
QYg	Cauchy	0.626	0.550	0.630	0.519	0.627	0.507	0.559	0.503	0.467	0.440
QYg(f)	Cauchy	0.628	0.549	0.632	0.518	0.627	0.508	0.559	0.504	0.467	0.439
QYu	Cauchy	0.632	0.555	0.636	0.524	0.633	0.512	0.564	0.508	0.472	0.444
L-stat	χ^2_3	0.766	0.640	0.796	0.769	0.692	0.742	0.709	0.643	0.593	0.614
QYg	χ^2_3	0.467	0.457	0.543	0.412	0.610	0.401	0.502	0.406	0.445	0.390
QYg(f)	χ^2_3	0.466	0.457	0.543	0.413	0.610	0.401	0.502	0.406	0.445	0.390
QYu	χ^2_3	0.471	0.461	0.548	0.416	0.616	0.405	0.507	0.410	0.449	0.393

Note: $1-\alpha = 0.95$, $n_1 = n_0 = 200$. "QYg(f)" is QYg on all replications instead of just those where L-stat may be used. The number of replications where L-stat was used for normal U_i, by w_0 in ascending order, was 813, 922, 872, 691, 918, 773, 962, 963, 999, 994; for t_3, 907, 953, 898, 815, 964, 868, 967, 970, 997, 996; for Cauchy, 953, 976, 910, 872, 976, 903, 964, 977, 996, 988; for χ^2_3, 959, 993, 970, 963, 988, 970, 984, 993, 999, 997.

Table 14. CP and median length for conditional median difference CIs, empirical DGP.

		w_0 empirical quantile								
Method	τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
		Coverage Probability								
L-stat	0.50	0.944	0.958	0.956	0.952	0.948	0.950	0.952	0.940	0.950
QYg	0.50	0.907	0.900	0.938	0.930	0.924	0.974	0.939	0.967	0.925
CBS	0.50	0.933	0.925	0.929	0.930	0.916	0.942	0.925	0.919	0.936
L-stat	0.90	0.964	0.944	0.948	0.945	0.946	0.953	0.953	0.938	0.956
QYg	0.90	0.945	0.952	0.922	0.926	0.921	0.928	0.926	0.887	0.905
CBS	0.90	0.947	0.944	0.932	0.935	0.932	0.941	0.935	0.927	0.948
		Median Interval Length								
L-stat	0.50	0.070	0.049	0.040	0.035	0.029	0.026	0.027	0.027	0.036
QYg	0.50	0.046	0.037	0.031	0.027	0.024	0.023	0.023	0.024	0.028
CBS	0.50	0.056	0.040	0.032	0.028	0.024	0.021	0.022	0.022	0.030
L-stat	0.90	0.099	0.081	0.064	0.054	0.051	0.047	0.045	0.053	0.048
QYg	0.90	0.062	0.055	0.046	0.041	0.037	0.033	0.032	0.030	0.029
CBS	0.90	0.077	0.063	0.052	0.044	0.041	0.037	0.037	0.042	0.039

Note: w_0 values are empirical (within each simulated dataset) deciles. DGP based on Gaussian mixture model fitted to food expenditure data for 2-adult/2-child ($T = 1$) and 1-adult/1-child ($T = 0$) households; $1 - \alpha = 0.95$, $n_0 = 2490$, $n_1 = 7095$, 1000 simulation replications.

productivity significantly in the first period, but not significantly thereafter. We do not simply re-test this main result (though we indeed support it) but rather offer complementary analysis on quantile treatment effects.

For the library task, Gneezy and List (2006) performed two types of one-sided 5% tests: a Wilcoxon rank-sum (a.k.a. Mann–Whitney–Wilcoxon or Mann–Whitney U) test, and an unequal variances t-test. For the first 90-minute period, the null was (barely) rejected by each test in favour of the treatment productivity being higher. For the remaining 90-minute periods, the null was not rejected by either test. The goal of the rank-sum test is to reject if $\mathrm{P}(T > C) > 0.5$, where T is a random variable corresponding to treatment group productivity and similarly C is control group productivity. The t-test instead tests for equality of means, though the assumption of normality of the sample average productivities is questionable with such a small sample size (too small to rely on the CLT).

Complementing these original tests for the library task, our method tests for equality at a chosen quantile of the productivity distribution. Also using a one-sided 5% test, we do not reject the null of equality at the lower quartile or the median, but we do reject at the upper quartile. Our two-sided, equal-tailed, 90% CIs are given in Table 15. The results are consistent with the rank-sum result that the first period treatment productivity is higher overall in some sense, and consistent with the t-test result that the mean is higher. Our test further suggests that the shift comes primarily (though not exclusively) from the upper part of the distribution: for the library task in the first period, the gift wage seems to induce the most productive workers to become extremely productive, while the effect is much less (if any) on less productive workers. For periods 2–4 (period 2 shown in table), our test fails to reject the null at any of these quartiles, consistent with the original results.

Table 15. Two-sided 90% confidence intervals for quartile treatment effects in Gneezy and List (2006).

Period (method)[a]	Lower quartile	Median	Upper quartile
		Library task[b]	
1 (kern/MSE)	(-10.25,21.42)	(-2.45,26.27)	(2.31,30.01)
1 (spac/bias)	(-10.45,21.67)	(-2.40,26.19)	(2.46,29.76)
2 (kern/MSE)	(-15.20,8.48)	(-7.23,25.53)	(-15.76,28.08)
2 (spac/bias)	(-16.56,8.96)	(-7.66,28.31)	(-16.05,28.25)
		Fundraising task[c]	
1 (kern/MSE)	(7.45,26.50)	(-1.45,27.11)	(-14.11,22.96)
1 (spac/bias)	(7.54,26.32)	(-0.98,27.00)	(-14.06,22.76)
2 (kern/MSE)	(-3.96,13.70)	(-5.58,11.47)	(-14.93,10.96)
2 (spac/bias)	(-3.86,13.67)	(-5.68,11.53)	(-20.13,12.04)

[a] Methods for $\hat{\gamma}$ estimation are abbreviated "kern/MSE" for a kernel density estimator with Silverman's (1986) rule of thumb bandwidth (MSE-optimal under normality), and "spac/bias" for a quantile spacing estimator with our zero-bias smoothing parameter.
[b] Units: books logged per period (90 minutes).
[c] Units: dollars raised per period (three hours).

For fundraising, Gneezy and List (2006) report one-sided 1% significance for the rank-sum test in period one.[26] We again use a 5% one-sided QTE test (10% two-sided), this time finding significance at the *lower* quartile (and almost at the median), but not at the upper quartile. The two-sided CIs are in Table 15. This is consistent with the original results, as is our failure to reject the null at any quartile in period two. Most interestingly, our results suggest that the *less* productive workers are most affected by the gift wage for the fundraising task, whereas the more productive workers were most affected in the library task.

In cases like Gneezy and List (2006), it may be even more appropriate to test for treatment effects across the entire distribution, which can be interpreted as multiple QTE testing that controls the familywise error rate (FWER). Such a method is developed in Goldman and Kaplan (2016), who find that the fundraising effect at lower quantiles is still significant with FWER controlled at 10% (two-sided), although the library effect is (barely) not.

REFERENCES

Andrews, D. W. K. and P. Guggenberger (2014). A conditional-heteroskedasticity-robust confidence interval for the autoregressive parameter. *Review of Economics and Statistics 96*(2), 376–381.

Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006). Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica 74*(2), 539–563.

Björkman, M. and J. Svensson (2009). Power to the people: Evidence from a randomized field experiment on community-based monitoring in Uganda. *Quarterly Journal of Economics 124*(2), 735–769.

[26] We compute a slightly higher p-value of 0.03. They do not report a Wilcoxon result for the second period, but we compute a p-value of 0.19. They also do not report a t-test this time.

Bloch, D. A. and J. L. Gastwirth (1968). On a simple estimate of the reciprocal of the density function. *Annals of Mathematical Statistics 39*(3), 1083–1085.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2017). On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association XX*(XX), XXX–XXX. Forthcoming.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295–2326.

Charness, G. and U. Gneezy (2009). Incentives to exercise. *Econometrica 77*(3), 909–931.

Chaudhuri, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Annals of Statistics 19*(2), 760–777.

Chu, J. T. (1957). Some uses of quasi-ranges. *Annals of Mathematical Statistics 28*(1), 173–180.

Chung, E. and J. P. Romano (2013). Exact and asymptotically robust permutation tests. *Annals of Statistics 41*(2), 484–507.

DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability*. New York: Springer.

Deaton, A. and C. Paxson (1998). Economies of scale, household size, and the demand for food. *Journal of Political Economy 106*(5), 897–930.

Doksum, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Annals of Statistics 2*(2), 267–277.

Duong, T. (2017). *ks: Kernel smoothing*. R package version 1.10.6.

Fan, Y. and R. Liu (2016). A direct approach to inference in nonparametric and semiparametric quantile models. *Journal of Econometrics 191*(1), 196–216.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics 1*(2), 209–230.

Fisher, R. A. (1932). *Statistical Methods for Research Workers* (4th ed.). Edinburgh: Oliver and Boyd.

Furrer, R., D. Nychka, and S. Sain (2012). *fields: Tools for spatial data*. R package version 6.6.3.

Gneezy, U. and J. A. List (2006). Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica 74*(5), 1365–1384.

Goldman, M. and D. M. Kaplan (2016). Comparing distributions by multiple testing across quantiles. Working Paper WP 16-19, Department of Economics, University of Missouri, available at `http://faculty.missouri.edu/~kaplandm`.

Goldman, M. and D. M. Kaplan (2017). Fractional order statistic approximation for nonparametric conditional quantile inference. *Journal of Econometrics 196*(2), 331–346.

Hall, P. and S. J. Sheather (1988). On the distribution of a Studentized quantile. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 50*(3), 381–391.

Hayfield, T. and J. S. Racine (2008). Nonparametric econometrics: The np package. *Journal of Statistical Software 27*(5), 1–32.

Hutson, A. D. (1999). Calculating nonparametric confidence intervals for quantiles using fractional order statistics. *Journal of Applied Statistics 26*(3), 343–353.

Jones, M. C. (2002). On fractional uniform order statistics. *Statistics and Probability Letters 58*(1), 93–96.

Kaplan, D. M. (2014). Nonparametric inference on quantile marginal effects. Working Paper WP 14-13, Department of Economics, University of Missouri, available at `http://faculty.missouri.edu/~kaplandm`.

Kaplan, D. M. (2015). Improved quantile inference via fixed-smoothing asymptotics and Edgeworth expansion. *Journal of Econometrics 185*(1), 20–32.

Koenker, R. (2016). *quantreg: Quantile Regression*. R package version 5.29.

Kopczuk, W., E. Saez, and J. Song (2010). Earnings inequality and mobility in the United States: Evidence from Social Security data since 1937. *Quarterly Journal of Economics 125*(1), 91–128.

Krewski, D. (1976). Distribution-free confidence intervals for quantile intervals. *Journal of the American Statistical Association 71*(354), 420–422.

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco: Holden-Day.

Lehmann, E. L. and J. P. Romano (2005). *Testing Statistical Hypotheses* (3rd ed.). Springer Texts in Statistics. Springer.

MaCurdy, T., X. Chen, and H. Hong (2011). Flexible estimation of treatment effect parameters. *American Economic Review (Papers and Proceedings) 101*(3), 544–551.

Neyman, J. (1937). »Smooth test» for goodness of fit. *Skandinavisk Aktuarietidskrift 20*(3–4), 149–199.

Office for National Statistics and Department for Environment, Food and Rural Affairs (2012). Living Costs and Food Survey. 2nd Edition. Colchester, Essex: UK Data Archive. `https://doi.org/10.5255/UKDA-SN-7472-2`.

Pearson, K. (1933). On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. *Biometrika 25*, 379–410.

Qu, Z. and J. Yoon (2015). Nonparametric estimation and inference on conditional quantile processes. *Journal of Econometrics 185*(1), 1–19.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Sathe, Y. S. and S. R. Lingras (1981). Bounds for the confidence coefficients of outer and inner confidence intervals for quantile intervals. *Journal of the American Statistical Association 76*(374), 473–475.

Siddiqui, M. M. (1960). Distribution of quantiles in samples from a bivariate population. *Journal of Research of the National Bureau of Standards—B. Mathematics and Mathematical Physics 64B*(3), 145–150.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Volume 26 of *Monographs on Statistics and Applied Probability*. London: Chapman & Hall/CRC.

Stigler, S. M. (1977). Fractional order statistics, with applications. *Journal of the American Statistical Association 72*(359), 544–550.

Wilks, S. S. (1962). *Mathematical Statistics*. New York: Wiley.