BOOTSTRAP CONSISTENCY FOR EMPIRICAL LIKELIHOOD IN DENSITY RATIO MODELS

Weiwei Zhuang[1], Weiqi Yang[1] and Jiahua Chen[2]

[1]*School of Management, University of Science and Technology of China*

[2]*Department of Statistics, The University of British Columbia*

Supplementary Material

This supplementary material contains the detailed proofs of the theoretical results (Theorems 1–5 and auxiliary lemmas) presented in the main text, alongside additional figures and descriptive statistics for the real-data analysis. Throughout this document, equation numbers enclosed in parentheses without a prefix, such as (1.1), refer to equations in the main manuscript. In contrast, equation numbers prefixed with "S", such as (S1.1), refer to equations within this supplementary material.

S1 Proof of Theorem 1

As a technical preparation, we present the following lemma, followed by a straightforward proof. Let $\mathcal{X}_n = \{\mathcal{X}_k,\ k = 0, \ldots, m\}$.

Lemma S.1. *For any function $g(\cdot)$ satisfying $\mathrm{E}_k[g^2(X)] < \infty$ for each $k = 0, 1, \ldots, m$.*

$$\frac{1}{n}\sum_{k,j} g(x^*_{kj}) - \sum_k \rho_k \,\mathrm{E}_k[g(X)] = \mathbf{o}_p(1). \tag{S1.1}$$

Proof. Utilizing the finite second moment condition, it suffices to show

$$\mathrm{E}\left[\left(\frac{1}{n}\sum_{k,j} g(x^*_{kj}) - \sum_k \rho_k \,\mathrm{E}_k\{g(X)\}\right)^2 \,\middle|\, \mathcal{X}_n\right] = o_p(1).$$

It is seen

$$\begin{aligned}
&\mathrm{E}\left[\left(\frac{1}{n}\sum_{k,j} g(x^*_{kj}) - \sum_k \rho_k \,\mathrm{E}_k[g(X)]\right)^2 \,\middle|\, \mathcal{X}_n\right] \\
&\quad = \underbrace{\mathrm{Var}\left[\frac{1}{n}\sum_{k,j} g(x^*_{kj}) \,\middle|\, \mathcal{X}_n\right]}_{\text{(A) Conditional Variance}} + \underbrace{\left(\mathrm{E}\left[\frac{1}{n}\sum_{k,j} g(x^*_{kj}) \,\middle|\, \mathcal{X}_n\right] - \sum_k \rho_k \,\mathrm{E}_k[g(X)]\right)^2}_{\text{(B) Squared Conditional Bias}}.
\end{aligned}$$

Denote the sample mean and sample variance of $\{g(x_{kj}), j = 1, \ldots, n_k\}$ by $\bar{g}_k$ and s_k^2. By the law of large numbers, $\bar{g}_k \to \mathrm{E}_k[g(X)]$ and $s_k^2 \to \mathrm{Var}_k[g(X)]$ almost surely. Because the bootstrap samples x^*_{kj} are conditionally independent given $\mathcal{X}_n$, we have

$$\mathrm{Var}\left[\frac{1}{n}\sum_{k,j} g(x^*_{kj}) \,\middle|\, \mathcal{X}_n\right] = \frac{1}{n^2}\sum_{k,j} \mathrm{Var}\left[g(x^*_{kj}) \,\middle|\, \mathcal{X}_n\right] = \frac{1}{n}\sum_{k=1}^m \rho_k s_k^2 = o_p(1)$$

and

$$\mathrm{E}\left[\frac{1}{n}\sum_{k,j} g(x^*_{kj}) \,\middle|\, \mathcal{X}_n\right] = \sum_{k=1}^m \rho_k \bar{g}_k \to \sum_k \rho_k \,\mathrm{E}_k[g(X)]$$

almost surely. Hence, both Terms (A) and (B) are $o_p(1)$ and the lemma is proven. □

Proof of Theorem 1. For any differentiable $f(\boldsymbol{\theta})$, write the partitioned gradient and Hessian as

$$\nabla f(\boldsymbol{\theta})^{\mathrm{T}} = \left(\frac{\partial f}{\partial \boldsymbol{\theta}_1}^{\mathrm{T}}, \ldots, \frac{\partial f}{\partial \boldsymbol{\theta}_m}^{\mathrm{T}}\right), \qquad \nabla^2 f(\boldsymbol{\theta}) = \left(\frac{\partial^2 f}{\partial \boldsymbol{\theta}_r \partial \boldsymbol{\theta}_s^{\mathrm{T}}}\right)_{1 \le r,s \le m}.$$

For l_n in (2.2), at true value $\boldsymbol{\theta}^\dagger$,

$$\frac{\partial l_n(\boldsymbol{\theta}^\dagger)}{\partial \boldsymbol{\theta}_r} = \sum_{j=1}^{n_r} \mathbf{q}(x_{rj}) - \sum_{k,j} h_r(x_{kj}; \boldsymbol{\theta}^\dagger)\, \mathbf{q}(x_{kj}) = \sum_{k,j} \Big(\delta_{kr} - h_r(x_{kj}; \boldsymbol{\theta}^\dagger)\Big) \mathbf{q}(x_{kj}),$$

$$\frac{\partial^2 l_n(\boldsymbol{\theta}^\dagger)}{\partial \boldsymbol{\theta}_r \partial \boldsymbol{\theta}_s^{\mathrm{T}}} = \sum_{k,j} \mathbf{q}(x_{kj}) \mathbf{q}^{\mathrm{T}}(x_{kj}) \Big(\delta_{rs} h_r(x_{kj}; \boldsymbol{\theta}^\dagger) - h_r(x_{kj}; \boldsymbol{\theta}^\dagger) h_s(x_{kj}; \boldsymbol{\theta}^\dagger)\Big).$$

For $\|\boldsymbol{\theta} - \boldsymbol{\theta}^\dagger\| = \mathbf{o}_p(n^{-1/3})$, a second-order Taylor expansion yields

$$l_n(\boldsymbol{\theta}) - l_n(\boldsymbol{\theta}^\dagger) = \nabla l_n(\boldsymbol{\theta}^\dagger)^{\mathrm{T}} (\boldsymbol{\theta} - \boldsymbol{\theta}^\dagger) + \tfrac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}^\dagger)^{\mathrm{T}} \nabla^2 l_n(\boldsymbol{\theta}^\dagger) (\boldsymbol{\theta} - \boldsymbol{\theta}^\dagger) + R_n,$$

with $R_n = \mathbf{o}_p(1)$, since $R_n = \mathbf{O}_p\big(n\|\boldsymbol{\theta} - \boldsymbol{\theta}^\dagger\|^3\big)$. Moreover, by applying the law of large number to the Hessian blocks,

$$\frac{1}{n} \frac{\partial^2 l_n(\boldsymbol{\theta}^\dagger)}{\partial \boldsymbol{\theta}_r \partial \boldsymbol{\theta}_s^{\mathrm{T}}} = \mathbf{W}_{rs} + \mathbf{o}_p(1), \qquad \Rightarrow \qquad \frac{1}{n} \nabla^2 l_n(\boldsymbol{\theta}^\dagger) - \mathbf{W} = \mathbf{o}_p(1).$$

Hence the local quadratic representation

$$l_n(\boldsymbol{\theta}) - l_n(\boldsymbol{\theta}^\dagger) = \nabla l_n(\boldsymbol{\theta}^\dagger)^{\mathrm{T}} (\boldsymbol{\theta} - \boldsymbol{\theta}^\dagger) + \frac{n}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}^\dagger)^{\mathrm{T}} \mathbf{W} (\boldsymbol{\theta} - \boldsymbol{\theta}^\dagger) + \mathbf{o}_p(1). \tag{S1.2}$$

Given that $\nabla l_n(\boldsymbol{\theta}^\dagger)$ is a sum of independent random vectors with finite moments, the central limit theorem ensures that $n^{-1/2} \nabla l_n(\boldsymbol{\theta}^\dagger) \rightsquigarrow N(\mathbf{0}, \mathbf{W} - \mathbf{WSW})$. Hence, maximizing (S1.2) gives

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\dagger) = \mathbf{W}^{-1} n^{-1/2} \nabla l_n(\boldsymbol{\theta}^\dagger) + \mathbf{o}_p(1). \tag{S1.3}$$

For $\boldsymbol{\theta} = \boldsymbol{\theta}^\dagger + \mathbf{o}_p(n^{-1/3})$, the bootstrap log-likelihood (3.3) has Taylor expansion

$$l_n^*(\boldsymbol{\theta}) - l_n^*(\boldsymbol{\theta}^\dagger) = \nabla l_n^*(\boldsymbol{\theta}^\dagger)^{\mathrm{T}} (\boldsymbol{\theta} - \boldsymbol{\theta}^\dagger) + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}^\dagger)^{\mathrm{T}} \nabla^2 l_n^*(\boldsymbol{\theta}^\dagger) (\boldsymbol{\theta} - \boldsymbol{\theta}^\dagger) + R_n^*, \tag{S1.4}$$

where $R_n^* = \mathbf{o}_p(1)$. Within this expansion, the gradient and Hessian are made of

$$\frac{\partial l_n^*(\boldsymbol{\theta}^\dagger)}{\partial \boldsymbol{\theta}_r} = \sum_{j=1}^{n_r} \mathbf{q}(x_{rj}^*) - \sum_{k,j} h_r(x_{kj}^*; \boldsymbol{\theta}^\dagger)\, \mathbf{q}(x_{kj}^*),$$

$$\frac{\partial^2 l_n^*(\boldsymbol{\theta}^\dagger)}{\partial \boldsymbol{\theta}_r \partial \boldsymbol{\theta}_s^{\mathrm{T}}} = \sum_{k,j} \mathbf{q}(x_{kj}^*)\mathbf{q}^{\mathrm{T}}(x_{kj}^*)\Big(\delta_{rs} h_r(x_{kj}^*; \boldsymbol{\theta}^\dagger) - h_r(x_{kj}^*; \boldsymbol{\theta}^\dagger) h_s(x_{kj}^*; \boldsymbol{\theta}^\dagger)\Big).$$

Both of them are sum of conditional i.i.d. terms which validating the conclusion of Lemma S.1. Hence,

$$\frac{1}{n}\nabla^2 l_n^*(\boldsymbol{\theta}^\dagger) - \mathbf{W} = \mathbf{o}_p(1).$$

A Taylor expansion then yields

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^\dagger) = -\mathbf{W}^{-1} n^{-1/2} \nabla l_n^*(\boldsymbol{\theta}^\dagger) + \mathbf{o}_p(1). \tag{S1.5}$$

Combining (S1.3)–(S1.5),

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}}) = \mathbf{W}^{-1} n^{-1/2}\big(\nabla l_n^*(\boldsymbol{\theta}^\dagger) - \nabla l_n(\boldsymbol{\theta}^\dagger)\big) + \mathbf{o}_p(1).$$

Thus (3.4) follows from

$$\sup_{\boldsymbol{\theta}} \big|\mathrm{P}\big\{n^{-1/2}\big(\nabla l_n^*(\boldsymbol{\theta}^\dagger) - \nabla l_n(\boldsymbol{\theta}^\dagger)\big) \le \boldsymbol{\theta} \,\big|\, \mathcal{X}_n\big\} - \mathrm{P}(\boldsymbol{Z}_1 \le \boldsymbol{\theta})\big| = \mathbf{o}_p(1), \tag{S1.6}$$

where $\boldsymbol{Z}_1 \sim N(\mathbf{0}, \mathbf{W} - \mathbf{WSW})$. Since $\nabla l_n(\boldsymbol{\theta}^\dagger)$ is a sum of independent random vectors with finite moments and $\nabla l_n^*(\boldsymbol{\theta}^\dagger)$ is its bootstrap analogue, Theorem 23.4 of Van der Vaart (2000) gives (S1.6), completing the proof.

S2 Proof of Theorem 2

The proof proceeds in three steps: establishing asymptotic joint normality, working out the covariance function $\omega_r(x, y)$, and verifying tightness. Together, these results establish the claimed weak convergence.

Step 1 (finite-dimensional convergence). From (S1.3),

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\dagger = n^{-1}\mathbf{W}^{-1}\nabla l_n(\boldsymbol{\theta}^\dagger) + \mathbf{o}_p(n^{-1/2}). \tag{S2.1}$$

Recall $\widehat{F}_r(x) = n_r^{-1}\sum_{k,j} h_r(x_{kj}; \widehat{\boldsymbol{\theta}})\,\mathbf{I}(x_{kj} \le x)$. A Taylor expansion of $h_r(\cdot; \widehat{\boldsymbol{\theta}})$ at $\boldsymbol{\theta}^\dagger$ gives

$$h_r(x_{kj}; \widehat{\boldsymbol{\theta}}) = h_r(x_{kj}; \boldsymbol{\theta}^\dagger) + \dot{h}_r(x_{kj}; \boldsymbol{\theta}^\dagger)^{\mathrm{T}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\dagger) + \mathbf{o}_p(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\dagger\|),$$

with $\dot{h}_r = \partial h_r/\partial\boldsymbol{\theta}$ at $\boldsymbol{\theta}^\dagger$. Since $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\dagger = \mathbf{O}_p(n^{-1/2})$, the remainder is $\mathbf{o}_p(n^{-1/2})$. Hence

$$\widehat{F}_r(x) = n_r^{-1}\sum_{k,j} h_r(x_{kj}; \boldsymbol{\theta}^\dagger)\mathbf{I}(x_{kj} \le x) + n_r^{-1}\Big\{n^{-1}\sum_{k,j}\dot{h}_r(x_{kj}; \boldsymbol{\theta}^\dagger)\mathbf{I}(x_{kj} \le x)\Big\}^{\mathrm{T}}\mathbf{W}^{-1}\nabla l_n(\boldsymbol{\theta}^\dagger) + \mathbf{o}_p(n^{-1/2}).$$

By the law of large numbers,

$$n^{-1}\sum_{k,j}\dot{h}_r(x_{kj}; \boldsymbol{\theta}^\dagger)\mathbf{I}(x_{kj} \le x) = \mathbf{B}_r(x) + \mathbf{o}_p(1).$$

Also

$$F_r(x) = n_r^{-1}\ \mathrm{E}\Big[\sum_{k,j} h_r(x_{kj}; \boldsymbol{\theta}^\dagger)\mathbf{I}(x_{kj} \le x)\Big].$$

Define

$$\Delta_{r1}(x) = \sum_{k,j}\Big\{h_r(x_{kj}; \boldsymbol{\theta}^\dagger)\mathbf{I}(x_{kj} \le x) - \mathrm{E}[h_r(x_{kj}; \boldsymbol{\theta}^\dagger)\mathbf{I}(x_{kj} \le x)]\Big\}, \quad \Delta_{r2}(x) = \mathbf{B}_r^{\mathrm{T}}(x)\,\mathbf{W}^{-1}\nabla l_n(\boldsymbol{\theta}^\dagger).$$

Then

$$\sqrt{n}\big(\widehat{F}_r(x) - F_r(x)\big) = \frac{\sqrt{n}}{n_r}\,\Delta_{r1}(x) + \frac{\sqrt{n}}{n_r}\,\Delta_{r2}(x) + \mathbf{o}_p(1).$$

For any fixed $x_1, \ldots, x_t$, the vector formed by $\{\Delta_{r1}(x_i)\}$ and $\{\Delta_{r2}(x_i)\}$ is a sum of independent mean-zero terms; by the multivariate central limit theorem, the finite-dimensional distributions are normal with mean zero. We compute the limiting covariance next.

Step 2 (covariance function). Introduce

$$c_{rk}(x) = \int_{-\infty}^{x} h_r(t;\boldsymbol{\theta}^\dagger)\, h_k(t;\boldsymbol{\theta}^\dagger)\, d\bar{F}(t), \qquad k = 0, \ldots, m.$$

(i) For $\mathrm{Cov}\,\big(\Delta_{r1}(x), \Delta_{r1}(y)\big)$, direct calculation gives

$$\mathrm{Cov}\,\big(\Delta_{r1}(x), \Delta_{r1}(y)\big) = n\Big\{c_{rr}(x \wedge y) - \sum_{k=0}^{m} \rho_k^{-1} c_{rk}(x) c_{rk}(y)\Big\}.$$

(ii) For $\mathrm{Cov}\,\big(\Delta_{r1}(x), \Delta_{r2}(y)\big)$, note that

$$\mathrm{Cov}\,\big(\Delta_{r1}(x), \nabla l_n(\boldsymbol{\theta}^\dagger)\big) = -\, n \sum_{k=0}^{m} \rho_k^{-1} c_{rk}(x)\, \mathbf{B}_k^{\mathrm{T}}.$$

This follows from writing $\nabla l_n(\boldsymbol{\theta}^\dagger)$ as a sum of centered scores and using independence across samples. Multiplying by $\mathbf{W}^{-1}\mathbf{B}_r(y)$ yields

$$\mathrm{Cov}\,\big(\Delta_{r1}(x), \Delta_{r2}(y)\big) = -\, n \sum_{k=0}^{m} \rho_k^{-1} c_{rk}(x)\, \mathbf{B}_k^{\mathrm{T}} \mathbf{W}^{-1} \mathbf{B}_r(y).$$

(iii) For $\mathrm{Cov}(\Delta_{r2}(x), \Delta_{r2}(y))$, using $\mathrm{Var}\{\nabla l_n(\boldsymbol{\theta}^\dagger)\} = n(\mathbf{W} - \mathbf{WSW})$,

$$\mathrm{Cov}\,\big(\Delta_{r2}(x), \Delta_{r2}(y)\big) = n\, \mathbf{B}_r^{\mathrm{T}}(x)\, (\mathbf{W}^{-1} - \mathbf{S})\, \mathbf{B}_r(y).$$

Combining (i)–(iii), dividing by $n_r^2 = (\rho_r n)^2$, and using the identity $c_{rr}(x) = \rho_r F_r(x) - a_r(x)$, one obtains (3.5).

Step 3 (tightness). Write

$$\tilde{F}_r(x) = n_r^{-1} \sum_{k,j} h_r(x_{kj}; \boldsymbol{\theta}^\dagger)\, \mathbf{I}(x_{kj} \le x).$$

For $x \ge y$,

$$\sqrt{n}\Big[\big(\widehat{F}_r(x) - F_r(x)\big) - \big(\widehat{F}_r(y) - F_r(y)\big)\Big] = A_n(x,y) + B_n(x,y),$$

where

$$A_n(x,y) = \sqrt{n}\Big[(\widehat{F}_r(x) - \tilde{F}_r(x)) - (\widehat{F}_r(y) - \tilde{F}_r(y))\Big],\ B_n(x,y) = \sqrt{n}\Big[(\tilde{F}_r(x) - F_r(x)) - (\tilde{F}_r(y) - F_r(y))\Big].$$

For A_n, by the mean value theorem in $\boldsymbol{\theta}$,

$$\sup_{|x-y|\le\delta} |A_n(x,y)| \ \le\ \big(\sqrt{n}\,\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\dagger\|\big) \sup_{|x-y|\le\delta} n_r^{-1} \sum_{k,j} C\, \|\mathbf{q}(x_{kj})\|\, \mathbf{I}(y < x_{kj} \le x).$$

Here C is a generic constant. Since $\sqrt{n}\,\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\dagger\| = \mathbf{O}_p(1)$, it suffices to bound the empirical increment. By the Cauchy–Schwarz inequality,

$$n_r^{-1} \sum_{k,j} \|\mathbf{q}(x_{kj})\| \mathbf{I}(y < x_{kj} \le x) \ \le\ \Big(n_r^{-1} \sum_{k,j} \|\mathbf{q}(x_{kj})\|^2\Big)^{1/2} \Big(n_r^{-1} \sum_{k,j} \mathbf{I}(y < x_{kj} \le x)\Big)^{1/2}.$$

The first factor is $\mathbf{O}_p(1)$ by the law of large numbers, and the second factor is $\mathbf{O}_p(\delta^{1/2})$.

Thus

$$\sup_{|x-y|\le\delta} |A_n(x,y)| = \mathbf{O}_p(\delta^{1/2}).$$

For B_n, it is an empirical-type process with bounded increments. Using the moment criterion (e.g., Billingsley, 2013, Thm. 13.5),

$$\mathrm{E}\left[|B_n(x,y)|^2\right] = \mathrm{Var}\Big[\sqrt{n}(\tilde{F}_r(x) - \tilde{F}_r(y))\Big] \;\le\; C\,|F_r(x) - F_r(y)| \;\le\; C'\,|x-y|,$$

for constants C, C'. Hence B_n is tight.

Conclusion. We have shown that A_n is tight and B_n is tight, and that the finite-dimensional convergence has been established in Step 1. Together these results prove that the process

$$\sqrt{n}\left(\widehat{F}_r(\cdot) - F_r(\cdot)\right)$$

converges weakly in $l^\infty(\mathbb{R})$ to the Gaussian process $\mathcal{G}_r$ with covariance function (3.5).

S3 Proof of Theorem 3

The argument parallels that of Theorem 2, and we highlight only the main steps.

Step 1 (finite-dimensional convergence). From (S1.5),

$$\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^\dagger = n^{-1}\mathbf{W}^{-1}\nabla l_n^*(\boldsymbol{\theta}^\dagger) + \mathbf{o}_p(n^{-1/2}).$$

Expanding $h_r(x_{kj}^*; \widehat{\boldsymbol{\theta}}^*)$ at $\boldsymbol{\theta}^\dagger$ gives

$$h_r(x_{kj}^*; \widehat{\boldsymbol{\theta}}^*) = h_r(x_{kj}^*; \boldsymbol{\theta}^\dagger) + \dot{h}_r(x_{kj}^*; \boldsymbol{\theta}^\dagger)^{\mathrm{T}}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^\dagger) + \mathbf{o}_p(n^{-1/2}),$$

so that

$$\widehat{F}_r^*(x) = n_r^{-1}\sum_{k,j} h_r(x_{kj}^*; \boldsymbol{\theta}^\dagger)\mathbf{I}(x_{kj}^* \le x) + n_r^{-1}\Big\{n^{-1}\sum_{k,j}\dot{h}_r(x_{kj}^*; \boldsymbol{\theta}^\dagger)\mathbf{I}(x_{kj}^* \le x)\Big\}^{\mathrm{T}}\mathbf{W}^{-1}\nabla l_n^*(\boldsymbol{\theta}^\dagger) + \mathbf{o}_p(n^{-1/2}).$$

By (S1.1),

$$n^{-1}\sum_{k,j}\dot{h}_r(x^*_{kj};\boldsymbol{\theta}^\dagger)\mathbf{I}(x^*_{kj}\le x)=\mathbf{B}_r(x)+\mathbf{o}_p(1).$$

Also $\widehat{F}_r(x)=n_r^{-1}\sum_{k,j}h_r(x_{kj};\widehat{\boldsymbol{\theta}})\mathbf{I}(x_{kj}\le x)$. Hence

$$\widehat{F}^*_r(x)-\widehat{F}_r(x)=n_r^{-1}\Delta^*_{r1}(x)+n_r^{-1}\Delta^*_{r2}(x)+\mathbf{o}_p(n^{-1/2}),$$

where

$$\begin{aligned}\Delta^*_{r1}(x) &= \sum_{k,j}\Big\{h_r(x^*_{kj};\boldsymbol{\theta}^\dagger)\mathbf{I}(x^*_{kj}\le x)-h_r(x_{kj};\boldsymbol{\theta}^\dagger)\mathbf{I}(x_{kj}\le x)\Big\},\\ \Delta^*_{r2}(x) &= \mathbf{B}_r^{\mathrm{T}}(x)\,\mathbf{W}^{-1}\big\{\nabla l^*_n(\boldsymbol{\theta}^\dagger)-\nabla l_n(\boldsymbol{\theta}^\dagger)\big\}.\end{aligned}$$

By Theorem 23.4 of Van der Vaart (2000), conditionally on $\mathcal{X}_n$ the finite-dimensional distributions of $n^{-1/2}\Delta^*_{ri}(\cdot)$ $(i=1,2)$ converge to those of their population counterparts, yielding the same limiting covariance $\omega_r(x,y)$ as in Theorem 2.

Step 2 (tightness). Let $\tilde{F}^*_r(x)=n_r^{-1}\sum_{k,j}h_r(x^*_{kj};\boldsymbol{\theta}^\dagger)\mathbf{I}(x^*_{kj}\le x)$. For $x\ge y$,

$$\sqrt{n}\Big[\big(\widehat{F}^*_r(x)-\widehat{F}_r(x)\big)-\big(\widehat{F}^*_r(y)-\widehat{F}_r(y)\big)\Big]=A^*_n(x,y)+B^*_n(x,y),$$

with

$$A^*_n(x,y)=\sqrt{n}\big[(\widehat{F}^*_r-\tilde{F}^*_r)(x)-(\widehat{F}^*_r-\tilde{F}^*_r)(y)\big],\quad B^*_n(x,y)=\sqrt{n}\big[(\tilde{F}^*_r-F_r)(x)-(\tilde{F}^*_r-F_r)(y)\big].$$

For A^*_n, conditioning on $\mathcal{X}_n$ and using the mean value theorem,

$$\sup_{|x-y|\le\delta}|A^*_n(x,y)|\le\big(\sqrt{n}\,\|\widehat{\boldsymbol{\theta}}^*-\boldsymbol{\theta}^\dagger\|\big)\sup_{|x-y|\le\delta}n_r^{-1}\sum_{k,j}C\,\|\mathbf{q}(x^*_{kj})\|\mathbf{I}(y<x^*_{kj}\le x).$$

By (S1.5), $\sqrt{n}\,\|\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^\dagger\| = \mathbf{O}_p(1)$. For the empirical increment, apply Cauchy–Schwarz:

$$n_r^{-1}\sum_{k,j}\|\mathbf{q}(x_{kj}^*)\|\mathbf{I}(y < x_{kj}^* \le x) \;\le\; \Big(n_r^{-1}\sum_{k,j}\|\mathbf{q}(x_{kj}^*)\|^2\Big)^{1/2}\Big(n_r^{-1}\sum_{k,j}\mathbf{I}(y < x_{kj}^* \le x)\Big)^{1/2}.$$

The first factor is $O_p(1)$ conditionally, and the second is $O_p(\delta^{1/2})$. Thus

$$\sup_{|x-y|\le\delta}|A_n^*(x,y)| = O_p(\delta^{1/2}).$$

For B_n^*, a conditional variance bound analogous to Step 3 of Theorem 2 gives

$$\mathrm{E}^*\left[|B_n^*(x,y)|^2\right] \;\le\; C\,|\widehat{F}_r(x) - \widehat{F}_r(y)| \;\le\; C'\,|x-y|,$$

in probability, where E^* denotes conditional expectation given $\mathcal{X}_n$. Thus B_n^* is conditionally tight.

Conclusion. Both A_n^* and B_n^* are tight, and Step 1 established conditional finite-dimensional convergence with the same covariance as in Theorem 2. Therefore,

$$\sqrt{n}\left(\widehat{F}_r^*(\cdot) - \widehat{F}_r(\cdot)\right)$$

converges weakly in $l^\infty(\mathbb{R})$ to $\mathcal{G}_r$, validating the bootstrap procedure.

S4 Proof of Theorem 4

(1) Convergence in $\mathbb{L}$. By Theorem 2, $\sqrt{n}(\widehat{F}_r - F_r) \rightsquigarrow \mathcal{G}_r$ in $l^\infty(\mathbb{R})$. To strengthen this to $\mathbb{L}$, we verify the equal-integrability condition of Kaji (2018), Theorem 1.1: for every $\eta, \epsilon > 0$ there exists $\delta > 0$ such that

$$\limsup_{n\to\infty}\mathrm{P}\left(\sup_x \int_{x-\delta}^{x+\delta}\Big|\sqrt{n}\big(\widehat{F}_r(y) - F_r(y)\big)\Big|\,dy > \eta\right) \le \epsilon. \tag{S4.1}$$

We illustrate for $r = 0$; other groups are analogous. Write

$$\widehat{F}_0 - F_0 = (\widehat{F}_0 - \tilde{F}_0) + (\tilde{F}_0 - F_0), \qquad \tilde{F}_0(x) = n_0^{-1} \sum_{k,j} h_0(x_{kj}; \boldsymbol{\theta}^\dagger)\, \mathbf{I}(x_{kj} \le x).$$

A mean-value expansion of $h_0(\cdot; \widehat{\boldsymbol{\theta}})$ about $\boldsymbol{\theta}^\dagger$ shows

$$\sup_y |\widehat{F}_0(y) - \tilde{F}_0(y)| = \mathbf{O}_p(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\dagger\|) = \mathbf{O}_p(n^{-1/2}),$$

under $\mathrm{E}\,\|\mathbf{q}(X)\| < \infty$. Standard empirical-process bounds yield $\sup_y |\tilde{F}_0(y) - F_0(y)| = O_p(n^{-1/2})$. Consequently,

$$\sup_x \int_{x-\delta}^{x+\delta} \Big| \sqrt{n}\big(\widehat{F}_0(y) - F_0(y)\big) \Big|\, dy \ \le \ 2\sqrt{n}\delta \, \sup_y |\widehat{F}_0(y) - F_0(y)| = \mathbf{O}_p(\delta).$$

Choosing δ sufficiently small ensures (S4.1). Thus $\sqrt{n}(\widehat{F}_r - F_r) \rightsquigarrow \mathcal{G}_r$ in $\mathbb{L}$.

(2) Functional delta method. By Kaji (2018), Theorem 1.3, the map $\phi : \mathbb{L} \to L_1[0,1]$, $F \mapsto Q$, is Hadamard differentiable at F_r with derivative

$$\phi'_{F_r}(h)(p) = -\frac{h(Q_r(p))}{F'_r(Q_r(p))} = -\, Q'_r(p)\, h(Q_r(p)).$$

Applying the functional delta method (Kosorok, 2008, Theorem 2.8) to $\sqrt{n}(\widehat{F}_r - F_r)$ yields (4.7).

(3) Bootstrap. A conditional analog of Step (1) (using bootstrap maximal inequalities; e.g., Ahmed et al. 2001) shows $\sqrt{n}(\widehat{F}^*_r - \widehat{F}_r) \rightsquigarrow \mathcal{G}_r$ in $\mathbb{L}$, conditionally on $\mathcal{X}_n$. The bootstrap functional delta method (Kosorok, 2008, Theorem 2.9) then yields the stated $L_1[0,1]$ limit for $\sqrt{n}(\widehat{Q}^*_r - \widehat{Q}_r)$.

S5 Proof of Theorem 5

As a technical preparation, we first present two lemmas, followed by straightforward proofs.

Lemma S.2 (Localization of bootstrap crossings)**.** *Assume the conditions of Lemma 3. Work under the probability integral transform so that $F_0(u) = u$ on $[0,1]$, and (4.8) becomes $f_1(0) \neq 1$ and $f_1(1) \neq 1$. Let $b_n := n^{-1/2}\log n \to 0$. If $\tilde{x}$ is the first crossing point of F_0 and F_1 in $(0,1)$, then, with probability tending to one (conditionally for the bootstrap),*

$$\widehat{F}_0^* \text{ and } \widehat{F}_1^* \text{ do not cross on } (b_n,\ \tilde{x} - \epsilon)$$

for any fixed small $\epsilon > 0$. Symmetrically, if $\tilde{y}$ is the last crossing in $(0,1)$, then there is no bootstrap crossing on $(\tilde{y} + \epsilon,\ 1 - b_n)$ with probability tending to one.

Proof. We prove the claim to the left of the first crossing; the right-side statement is analogous. Since $F_0(u) = u$ and $f_1(0) \neq 1$, continuity implies the existence of $\delta > 0$ and $\eta \in (0, \epsilon)$ such that

$$F_1(u) - F_0(u) = F_1(u) - u \ \geq\ \delta\, u \quad \text{for all } u \in [0, \eta].$$

From Theorem 2 and its bootstrap version (Theorem 3), we have the uniform conditional rate

$$\sup_{u\in[0,1]} \left|\widehat{F}_r^*(u) - F_r(u)\right| = \mathbf{O}_p(n^{-1/2}), \qquad r = 0, 1,$$

hence, uniformly in $u \in [0,1]$,

$$\widehat{F}_1^*(u) - \widehat{F}_0^*(u) = \{F_1(u) - F_0(u)\} + R_n^*(u), \qquad \sup_u |R_n^*(u)| = \mathbf{O}_p(n^{-1/2}).$$

Fix n large so that $b_n < \eta$ and $|R_n^*(u)| \le \frac{1}{2}\delta b_n$ with conditional probability at least $1 - \varepsilon$ (for any fixed $\varepsilon > 0$). Then, for all $u \in [b_n, \eta]$,

$$\widehat{F}_1^*(u) - \widehat{F}_0^*(u) \;\ge\; \delta u - \tfrac{1}{2}\delta b_n \;\ge\; \tfrac{1}{2}\delta b_n \;>\; 0.$$

On $[\eta, \tilde{x} - \epsilon]$, $F_1 - F_0$ is bounded away from 0 (by definition of the first crossing), so the same uniform $O_p(n^{-1/2})$ control implies $\widehat{F}_1^* - \widehat{F}_0^* > 0$ there for n large. Hence no bootstrap crossing occurs on $(b_n, \tilde{x} - \epsilon)$ with probability tending to one. □

Lemma S.3 (Uniqueness of bootstrap crossing near each true crossing)**.** *Assume Lemma S.2. Let $\tilde{x} \in (0,1)$ be a crossing of (F_0, F_1) such that $f_1(\tilde{x}) \neq f_0(\tilde{x})$. Then there exists $\eta > 0$ such that, with probability tending to one (conditionally for the bootstrap), the function $u \mapsto \widehat{F}_1^*(u) - \widehat{F}_0^*(u)$ is strictly monotone on $(\tilde{x} - \eta, \tilde{x} + \eta)$ and therefore has a unique zero in this interval. The same conclusion holds for $u \mapsto \widehat{F}_1(u) - \widehat{F}_0(u)$.*

Proof. Let $G(u) := F_1(u) - F_0(u)$. By assumption, $G(\tilde{x}) = 0$ and $G'(\tilde{x}) = f_1(\tilde{x}) - f_0(\tilde{x}) \neq 0$. By continuity of G', there exist $\eta > 0$ and $c > 0$ such that either $G'(u) \ge c > 0$ for all $u \in (\tilde{x} - \eta, \tilde{x} + \eta)$ or $G'(u) \le -c < 0$ there. Hence G is strictly monotone on $(\tilde{x} - \eta, \tilde{x} + \eta)$ and has a unique zero at $u = \tilde{x}$.

Write the bootstrap perturbation as

$$\widehat{G}^*(u) := \widehat{F}_1^*(u) - \widehat{F}_0^*(u) = G(u) + R_n^*(u), \qquad \sup_{u\in[0,1]} |R_n^*(u)| = \mathbf{O}_p(n^{-1/2}).$$

Shrink η if necessary so that $|G(u)| \geq c_1|u-\tilde{x}|$ on $(\tilde{x}-\eta, \tilde{x}+\eta)$ for some $c_1 > 0$ (mean value theorem with G' bounded away from 0). For n large, with probability tending to one we have $\sup_{|u-\tilde{x}|\leq\eta} |R_n^*(u)| \leq \frac{1}{2}c_1\eta$. Then, for $u_1 < u_2$ in $(\tilde{x}-\eta, \tilde{x}+\eta)$,

$$\widehat{G}^*(u_2) - \widehat{G}^*(u_1) = \{G(u_2) - G(u_1)\} + \{R_n^*(u_2) - R_n^*(u_1)\}.$$

Since G is strictly monotone, $G(u_2) - G(u_1)$ has fixed sign and magnitude at least $c_1(u_2 - u_1)$, which dominates the $\mathbf{o}_p(1)$ perturbation $R_n^*(u_2) - R_n^*(u_1)$. Thus $\widehat{G}^*$ is strictly monotone on $(\tilde{x}-\eta, \tilde{x}+\eta)$ for n large, and it has exactly one zero there. The same argument (without conditional probability) applies to $\widehat{G}(u) := \widehat{F}_1(u) - \widehat{F}_0(u)$. □

Proof of Theorem 5. *Roadmap.* We (i) reduce by the PIT normalization used above, (ii) localize bootstrap crossings to the boundary strips or small neighborhoods of the true crossings (Lemmas S.2–S.3), (iii) linearize crossing locations via an argmax expansion driven by the quantile processes (Theorem 4), and (iv) conclude by conditional weak convergence of the bootstrap process.

Reduction via PIT. As above, work on $[0,1]$ with $F_0(u) = u$ and $f_1(0) \neq 1 \neq f_1(1)$.

Localization of bootstrap crossings. Let $b_n := n^{-1/2}\log n$ as in Lemma S.2. The lemmas confine any bootstrap crossings of $\widehat{F}_0^*$ and $\widehat{F}_1^*$ to $(0, b_n) \cup (1-b_n, 1)$ or to small neighborhoods of the true crossings, and ensure uniqueness there.

Case 1: no crossings. If F_0 and F_1 never cross, then $\gamma^\dagger \in \{0,1\}$. By Lemma S.2, any bootstrap crossing can only occur in boundary strips of length $O(b_n)$, so $\widehat{\gamma}^* = \mathbf{o}_p(n^{-1/2})$ or $1 - \mathbf{o}_p(n^{-1/2})$. The same holds for $\widehat{\gamma}$ (cf. the supplement of Zhuang et al., 2019). Thus

$$n^{1/2}(\widehat{\gamma}^* - \widehat{\gamma}) = \mathbf{o}_p(1),$$

and (4.9) is immediate.

Case 2: one crossing. Suppose F_0 and F_1 cross once at $\tilde{x}_1 \in (0,1)$, so $Q_0(t) > Q_1(t)$ iff $t < \tilde{x}_1$. By Lemma S.3 (and the analogous result for $\widehat{F}_r$), there exist unique $\widehat{x}_1, \widehat{x}_1^* \in (b_n, 1-b_n)$ with

$$\widehat{\gamma} = \widehat{x}_1 + \mathbf{o}_p(n^{-1/2}), \qquad \widehat{\gamma}^* = \widehat{x}_1^* + \mathbf{o}_p(n^{-1/2}).$$

For small $\epsilon > 0$, define the integrated contrasts

$$\Phi(u) = \int_{\tilde{x}_1-\epsilon}^{u} \{Q_0(t) - Q_1(t)\}\, dt, \quad \widehat{\Phi}_n(u) = \int_{\tilde{x}_1-\epsilon}^{u} \{\widehat{Q}_0(t) - \widehat{Q}_1(t)\}\, dt,$$

$$\widehat{\Phi}_n^*(u) = \int_{\tilde{x}_1-\epsilon}^{u} \{\widehat{Q}_0^*(t) - \widehat{Q}_1^*(t)\}\, dt.$$

By Theorem 4, the centered processes

$$W_n(t) = n^{1/2}\Big[(\widehat{Q}_0 - \widehat{Q}_1)(t) - (Q_0 - Q_1)(t)\Big], \quad W_n^*(t) = n^{1/2}\Big[(\widehat{Q}_0^* - \widehat{Q}_1^*)(t) - (\widehat{Q}_0 - \widehat{Q}_1)(t)\Big]$$

converge weakly (conditionally for W_n^*) in $L_1[0,1]$ to the same Gaussian limit. Hence

$$\Delta_n(u) := \widehat{\Phi}_n(u) - \Phi(u) = n^{-1/2}\int_{\tilde{x}_1-\epsilon}^{u} W_n(t)\, dt, \quad \Delta_n^*(u) := \widehat{\Phi}_n^*(u) - \widehat{\Phi}_n(u) = n^{-1/2}\int_{\tilde{x}_1-\epsilon}^{u} W_n^*(t)\, dt.$$

Because Φ has a unique maximizer at $\tilde{x}_1$ with negative second derivative there ($f_0(Q_0(\tilde{x}_1)) \neq f_1(Q_1(\tilde{x}_1))$), the argmax map is Hadamard differentiable. By Theorem A.2 of Álvarez-Esteban et al. (2017),

$$n^{1/2}(\widehat{x}_1 - \tilde{x}_1) = -C\, W_n(\tilde{x}_1) + \mathbf{o}_p(1), \qquad n^{1/2}(\widehat{x}_1^* - \widehat{x}_1) = -C\, \{W_n^*(\tilde{x}_1) - W_n(\tilde{x}_1)\} + \mathbf{o}_p(1),$$

with

$$C = \frac{f_0(Q_0(\tilde{x}_1))\, f_1(Q_1(\tilde{x}_1))}{f_0(Q_0(\tilde{x}_1)) - f_1(Q_1(\tilde{x}_1))}.$$

Therefore,

$$n^{1/2}(\widehat{\gamma}^* - \widehat{\gamma}) = -C\big\{W_n^*(\tilde{x}_1) - W_n(\tilde{x}_1)\big\} + \mathbf{o}_p(1).$$

By Theorem 4 (bootstrap convergence of the quantile process) together with Theorem 1 (bootstrap consistency for the parameter estimation embedded in $\widehat{Q}_r^*$), the conditional limit of $W_n^*(\tilde{x}_1) - W_n(\tilde{x}_1)$ equals the (unconditional) limit of $W_n(\tilde{x}_1)$. Hence $n^{1/2}(\widehat{\gamma}^* - \widehat{\gamma}) \rightsquigarrow N(0, \sigma^2)$ conditionally on $\mathcal{X}_n$, where $\sigma^2 = C^2\, \mathrm{Var}\{W(\tilde{x}_1)\}$ matches Lemma 3.

Case 3: finitely many crossings. Let the crossings be $\tilde{x}_1, \ldots, \tilde{x}_K$; then

$$\gamma^\dagger = \sum_{j=1}^{K} s_j \tilde{x}_j, \qquad s_j \in \{+1, -1\},$$

with s_j determined by the local sign of $Q_0 - Q_1$. By Lemmas S.2–S.3, there are unique $\widehat{x}_j, \widehat{x}_j^*$ near each $\tilde{x}_j$, and

$$\widehat{\gamma} = \sum_{j=1}^{K} s_j \widehat{x}_j + \mathbf{o}_p(n^{-1/2}), \qquad \widehat{\gamma}^* = \sum_{j=1}^{K} s_j \widehat{x}_j^* + \mathbf{o}_p(n^{-1/2}).$$

Applying the same argmax linearization uniformly over j yields

$$n^{1/2}(\widehat{x}_j^* - \widehat{x}_j) = -C_j\{W_n^*(\tilde{x}_j) - W_n(\tilde{x}_j)\} + \mathbf{o}_p(1),$$

with C_j defined as above at $\tilde{x}_j$. Summing over j,

$$n^{1/2}(\widehat{\gamma}^* - \widehat{\gamma}) = -\sum_{j=1}^{K} s_j C_j\{W_n^*(\tilde{x}_j) - W_n(\tilde{x}_j)\} + \mathbf{o}_p(1).$$

By joint (bootstrap) weak convergence of the quantile process at finitely many points (Theorem 4), the conditional limit coincides with the Gaussian limit of $\sum_{j=1}^{K} s_j C_j W(\tilde{x}_j)$, giving (4.9). $\square$

S6 Real-data analysis

Figure 1 and Table 1 present the empirical distribution functions and descriptive statistics for the six provinces, respectively.

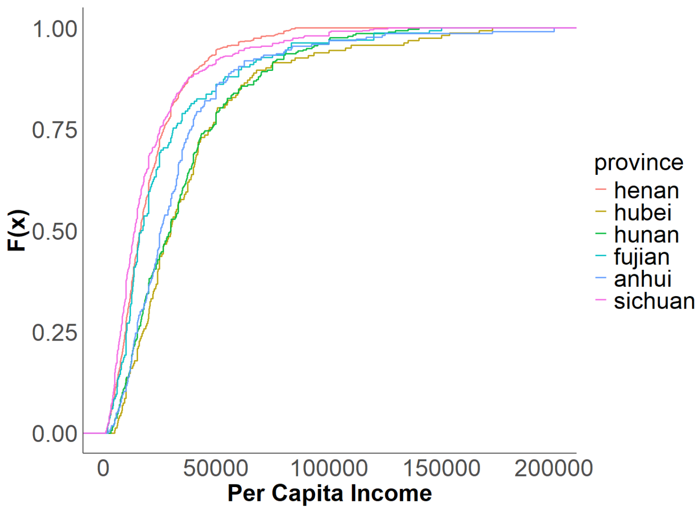

Figure 1: Empirical distribution functions for the six provinces

Table 1: Descriptive statistics for the six provinces.

Province	**N**	**Mean**	**Median**	**StdDev**	**IQR**	**Min**	**Max**
Henan	1160	20737	16362	16050	16794	1600	85000
Hubei	163	38334	30000	32028	29150	5150	172500
Hunan	299	35275	30000	26478	32550	3333	140000
Fujian	166	26722	17083	27135	20575	2000	150000
Anhui	223	32745	25483	28908	23767	2050	200000
Sichuan	510	20558	14000	20780	17419	1089	126300

Bibliography

Ahmed, S., D. Li, A. Rosalsky, and A. I. Volodin (2001). Almost sure lim sup behavior of bootstrapped means with applications to pairwise iid sequences and stationary ergodic sequences. *Journal of statistical planning and inference 98*(1-2), 1–14.

Álvarez-Esteban, P., E. Barrio, J. Cuesta-Albertos, and C. Matrán (2017). Models for the assessment of treatment improvement: The ideal and the feasible. *Statistical Science 32*(3), 469–485.

Billingsley, P. (2013). *Convergence of probability measures.* John Wiley & Sons.

Kaji, T. (2018). *Essays on Asymptotic Methods in Econometrics.* Ph. D. thesis, Massachusetts Institute of Technology.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics. New York: Springer.

Van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge University Press.

Zhuang, W., B. Hu, and R. Chen (2019). Semiparametric inference for the dominance index under the density ratio model. *Biometrika 106*(1), 229–241.