Supplement to "Semiparametric estimation of structural functions in nonseparable triangular models"

(*Quantitative Economics*, Vol. 11, No. 2, May 2020, 503–533)

Victor Chernozhukov
Department of Economics, MIT

Iván Fernández-Val
Department of Economics, Boston University

Whitney Newey
Department of Economics, MIT

Sami Stouli
Department of Economics, University of Bristol

Francis Vella
Department of Economics, Georgetown University

Appendix B: Summary

In the Online Supplementary Material, we first give the proofs of asymptotic theory results in the main text. We then report results from a sensitivity analysis we carried out to check the robustness of our empirical results. In Section D.1, we report additional QSF estimates obtained for different regions of interest, with grids of values of X of varying cardinality and length, as well as with additional quantile levels. We also compare ASF estimates obtained by least-squares projection, as described in Remark 4 in the main text, to those obtained by QR. In Section D.2, we report more flexible QSF estimates including additional powers of the control variable $\Phi^{-1}(V)^k$ as well as interaction terms $X \cdot \Phi^{-1}(V)^k$, $Z_1 \cdot \Phi^{-1}(V)^k$, and $X \cdot Z_1 \cdot \Phi^{-1}(V)^k$. The selection of these additional terms is investigated for the ASF by means of a least-squares cross-validation procedure. In Section D.3, we exploit knowledge of the control function distribution and implement a simulation-based integration procedure as an alternative to sample averaging over the estimated control function. Finally, in Section E, we perform several Monte Carlo simulations in order to assess the finite sample performance of our estimators. We compare the DR and QR estimators of structural functions with several designs calibrated to the

Victor Chernozhukov: vchern@mit.edu
Iván Fernández-Val: ivanf@bu.edu
Whitney Newey: wnewey@mit.edu
Sami Stouli: s.stouli@bristol.ac.uk
Francis Vella: Francis.Vella@georgetown.edu

 https://doi.org/10.3982/QE1239

empirical application. We give a detailed description of our calibration procedure for the Monte Carlo simulations. Overall, our robustness checks show that our empirical results are robust to the modeling, estimation, and integration choices, and our additional simulation results confirm the main findings for the ASF discussed in the main text.

Appendix C: Asymptotic theory

C.1 *Notation*

In what follows, ϑ denotes a generic value for the control function. It is convenient also to introduce some additional notation, which will be extensively used in the proofs. Let $V_i(\vartheta) := \vartheta(X_i, Z_i)$, $W_i(\vartheta) := w(X_i, Z_{1i}, V_i(\vartheta))$, and $\dot{W}_i(\vartheta) := \partial_v w(X_i, Z_{1i}, v)|_{v=V_i(\vartheta)}$. When the previous functions are evaluated at the true values, we use $V_i = V_i(\vartheta_0)$, $W_i = W_i(\vartheta_0)$, and $\dot{W}_i = \dot{W}_i(\vartheta_0)$. Also, let $\rho_y(u, v) := -1(u \leq y)\log \Lambda(v) - 1(u > y)\log \Lambda(-v)$. Recall that $A := (Y, X, Z, W, V)$, $T(x) = 1(x \in \overline{\mathcal{X}})$, and $T = T(X)$. For a function $f : \mathcal{A} \mapsto \mathbb{R}$, we use $\|f\|_{T,\infty} = \sup_{a\in\mathcal{A}} |T(x)f(a)|$; for a K-vector of functions $f : \mathcal{A} \mapsto \mathbb{R}^K$, we use $\|f\|_{T,\infty} = \sup_{a\in\mathcal{A}} \|T(x)f(a)\|_2$. We make functions in $\mathcal{Y}$ as well as estimators $\widehat{\vartheta}$ to take values in $[0, 1]$, the support of the control function V. This allows us to simplify notation in what follows.

We adopt the standard notation in the empirical process literature (see, e.g., van der Vaart (2000)),

$$\mathbb{E}_n[f] = \mathbb{E}_n\big[f(A)\big] = n^{-1}\sum_{i=1}^{n} f(A_i),$$

and

$$\mathbb{G}_n[f] = \mathbb{G}_n\big[f(A)\big] = n^{-1/2}\sum_{i=1}^{n}\big(f(A_i) - \mathrm{E}_P\big[f(A)\big]\big).$$

When the function $\widehat{f}$ is estimated, the notation should interpreted as

$$\mathbb{G}_n[\widehat{f}] = \mathbb{G}_n[f]\,|_{f=\widehat{f}} \quad \text{and} \quad \mathrm{E}_P[\widehat{f}] = \mathrm{E}_P[f]\,|_{f=\widehat{f}}.$$

We also use the concepts of covering entropy and bracketing entropy in the proofs. The covering entropy $\log N(\epsilon, \mathcal{F}, \|\cdot\|)$ is the logarithm of the minimal number of $\|\cdot\|$-balls of radius ϵ needed to cover the set of functions $\mathcal{F}$. The bracketing entropy $\log N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|)$ is the logarithm of the minimal number of ϵ-brackets in $\|\cdot\|$ needed to cover the set of functions $\mathcal{F}$. An ϵ-bracket $[\ell, u]$ in $\|\cdot\|$ is the set of functions f with $\ell \leq f \leq u$ and $\|u - \ell\| < \epsilon$.

For a sequence of random functions $y \mapsto f_n(y)$ and a deterministic sequence a_n, we use $f_n(y) = \bar{o}_{\mathbb{P}}(a_n)$ and $f_n(y) = \bar{O}_{\mathbb{P}}(a_n)$ to denote uniform in $y \in \mathcal{Y}$ orders in probability, that is, $\sup_{y\in\mathcal{Y}} f_n(y) = o_{\mathbb{P}}(a_n)$ and $\sup_{y\in\mathcal{Y}} f_n(y) = O_{\mathbb{P}}(a_n)$, respectively. The uniform in $y \in \mathcal{Y}$ deterministic orders $\bar{o}(a_n)$ and $\bar{O}(a_n)$ are defined analogously suppressing the $\mathbb{P}$ subscripts.

We follow the notation and definitions in van der Vaart and Wellner (1996) of bootstrap consistency. Let D_n denote the data vector and E_n be the vector of bootstrap

weights. Consider the random element $Z_n^e = Z_n(D_n, E_n)$ in a normed space $\mathbb{Z}$. We say that the bootstrap law of Z_n^e consistently estimates the law of some tight random element Z and write $Z_n^e \rightsquigarrow_{\mathbb{P}} Z$ in $\mathbb{Z}$ if

$$\sup_{h \in \mathrm{BL}_1(\mathbb{Z})} \left| \mathrm{E}_P^e h\left(Z_n^e\right) - \mathrm{E}_P h(Z) \right| \to_{\mathbb{P}^*} 0, \tag{C.1}$$

where $\mathrm{BL}_1(\mathbb{Z})$ denotes the space of functions with Lipschitz norm at most 1, E_P^e denotes the conditional expectation with respect to E_n given the data D_n, and $\to_{\mathbb{P}^*}$ denotes convergence in (outer) probability.

C.2 *Proof of Lemma 3*

We only consider the case where $\mathcal{Y}$ is a compact interval of $\mathbb{R}$. The case where $\mathcal{Y}$ is finite is simpler and follows similarly.

C.2.1 *Auxiliary lemmas* We start with two results on stochastic equicontinuity and a local expansion for the second stage estimators that will be used in the proof of Lemma 3.

Lemma C.1 (Stochastic equicontinuity). *Let $e \geq 0$ be a positive random variable with $\mathrm{E}_P[e] = 1$, $\mathrm{Var}_P[e] = 1$, and $\mathrm{E}_P|e|^{2+\delta} < \infty$ for some $\delta > 0$, that is independent of (Y, X, Z, W, V), including as a special case $e = 1$, and set, for $A = (e, Y, X, Z, W, V)$,*

$$f_y(A, \vartheta, \beta) := e \cdot \left[\Lambda\left(W(\vartheta)'\beta\right) - 1(Y \leq y)\right] \cdot W(\vartheta) \cdot T.$$

Under Assumptions 3–5, *the following relations are true*:

(a) *Consider the set of functions*

$$\mathcal{F} = \left\{ f_y(A, \vartheta, \beta)'\alpha : (\vartheta, \beta, y) \in \Upsilon_0 \times \mathcal{B} \times \mathcal{Y}, \alpha \in \mathbb{R}^{\dim(W)}, \|\alpha\|_2 \leq 1 \right\},$$

where $\mathcal{Y}$ is a compact subset of $\mathbb{R}$, $\mathcal{B}$ is a compact set under the $\|\cdot\|_2$ metric containing $\beta_0(y)$ for all $y \in \mathcal{Y}$, Υ_0 is the intersection of Υ, defined in Lemma 2, *with a neighborhood of ϑ_0 under the $\|\cdot\|_{T,\infty}$ metric. This class is P-Donsker with a square integrable envelope of the form e times a constant.*

(b) *Moreover, if $(\vartheta, \beta(y)) \to (\vartheta_0, \beta_0(y))$ in the $\|\cdot\|_{T,\infty} \vee \|\cdot\|_2$ metric uniformly in $y \in \mathcal{Y}$, then*

$$\sup_{y \in \mathcal{Y}} \left\| f_y\left(A, \vartheta, \beta(y)\right) - f_y\left(A, \vartheta_0, \beta_0(y)\right) \right\|_{P,2} \to 0.$$

(c) *Hence for any $(\widetilde{\vartheta}, \widetilde{\beta}(y)) \to_{\mathbb{P}} (\vartheta_0, \beta_0(y))$ in the $\|\cdot\|_{T,\infty} \vee \|\cdot\|_2$ metric uniformly in $y \in \mathcal{Y}$ such that $\widetilde{\vartheta} \in \Upsilon_0$,*

$$\sup_{y \in \mathcal{Y}} \left\| \mathbb{G}_n f_y\left(A, \widetilde{\vartheta}, \widetilde{\beta}(y)\right) - \mathbb{G}_n f_y\left(A, \vartheta_0, \beta_0(y)\right) \right\|_2 \to_{\mathbb{P}} 0.$$

(d) *For any* $(\widehat{\vartheta}, \widetilde{\beta}(y)) \to_{\mathbb{P}} (\vartheta_0, \beta_0(y))$ *in the* $\|\cdot\|_{T,\infty} \vee \|\cdot\|_2$ *metric uniformly in* $y \in \mathcal{Y}$, *so that*

$$\|\widehat{\vartheta} - \widetilde{\vartheta}\|_{T,\infty} = o_{\mathbb{P}}(1/\sqrt{n}), \quad \textit{where } \widetilde{\vartheta} \in Y_0,$$

we have that

$$\sup_{y\in\mathcal{Y}} \big\| \mathbb{G}_n f_y\big(A, \widehat{\vartheta}, \widetilde{\beta}(y)\big) - \mathbb{G}_n f_y\big(A, \vartheta_0, \beta_0(y)\big) \big\|_2 \to_{\mathbb{P}} 0.$$

Proof of Lemma C.1. The proof is divided in subproofs of each of the claims.

Proof of Claim (a). The proof proceeds in several steps.

Step 1. Here, we bound the bracketing entropy for

$$\mathcal{I}_1 = \big\{ \big[\Lambda\big(W(\vartheta)'\beta\big) - 1(Y \le y)\big]T : \beta \in \mathcal{B}, \vartheta \in Y_0, y \in \mathcal{Y} \big\}.$$

For this purpose, consider a mesh $\{\vartheta_k\}$ over Y_0 of $\|\cdot\|_{T,\infty}$ width δ, a mesh $\{\beta_l\}$ over $\mathcal{B}$ of $\|\cdot\|_2$ width δ, and a mesh $\{y_j\}$ over $\mathcal{Y}$ of $\|\cdot\|_2$ width δ. A generic bracket over $\mathcal{I}_1$ takes the form

$$\big[i_1^0, i_1^1\big] = \big[\big\{\Lambda\big(W(\vartheta_k)'\beta_l - \kappa\delta\big) - 1(Y \le y_j - \delta)\big\}T, \big\{\Lambda\big(W(\vartheta_k)'\beta_l + \kappa\delta\big) - 1(Y \le y_j + \delta)\big\}T\big],$$

where $\kappa = L_W \max_{\beta\in\mathcal{B}} \|\beta\|_2 + L_W$, and $L_W := \|\partial_v w\|_{T,\infty} \vee \|w\|_{T,\infty}$.

Note that this is a valid bracket for all elements of $\mathcal{I}_1$ because for any ϑ located within δ from ϑ_k and any β located within δ from β_l,

$$\begin{aligned} \big|W(\vartheta)'\beta - W(\vartheta_k)'\beta_l\big|T &\le \big|\big(W(\vartheta) - W(\vartheta_k)\big)'\beta\big|T + \big|W(\vartheta_k)'(\beta - \beta_l)\big|T \\ &\le L_W \delta \max_{\beta\in\mathcal{B}} \|\beta\|_2 + L_W\delta \le \kappa\delta, \end{aligned} \tag{C.2}$$

and the $\|\cdot\|_{P,2}$-size of this bracket is given by

$$\begin{aligned} \big\|i_1^0 - i_1^1\big\|_{P,2} &\le \sqrt{\mathrm{E}_P\big[P\big\{Y \in [y \pm \delta] \mid X, Z\big\}T\big]} \\ &\quad + \sqrt{\mathrm{E}_P\big[\big\{\Lambda\big(W(\vartheta_k)'\beta_l + \kappa\delta\big) - \Lambda\big(W(\vartheta_k)'\beta_l - \kappa\delta\big)\big\}^2 T\big]} \\ &\le \sqrt{\big\|f_Y(\cdot \mid \cdot)\big\|_{T,\infty} 2\delta} + \kappa\delta/2, \end{aligned}$$

because $\|\lambda(\cdot)\|_{T,\infty} \le 1/4$, where $\lambda = \Lambda(1-\Lambda)$ is the derivative of Λ.

Hence, counting the number of brackets induced by the mesh created above, we arrive at the following relationship between the bracketing entropy of $\mathcal{I}_1$ and the covering entropies of Y_0, $\mathcal{B}$, and $\mathcal{Y}$,

$$\begin{aligned} \log N_{[]}\big(\epsilon, \mathcal{I}_1, \|\cdot\|_{P,2}\big) &\lesssim \log N\big(\epsilon^2, Y_0, \|\cdot\|_{T,\infty}\big) + \log N\big(\epsilon^2, \mathcal{B}, \|\cdot\|_2\big) + \log N\big(\epsilon^2, \mathcal{Y}, \|\cdot\|_2\big) \\ &\lesssim 1/\big(\epsilon^2 \log^4 \epsilon\big) + \log(1/\epsilon) + \log(1/\epsilon), \end{aligned}$$

and so $\mathcal{I}_1$ is P-Donsker with a constant envelope.

Step 2. Similar to Step 1, it follows that

$$\mathcal{I}_2 = \left\{W(\vartheta)'\alpha T : \vartheta \in \Upsilon_0, \alpha \in \mathbb{R}^{\dim(W)}, \|\alpha\|_2 \leq 1\right\}$$

also obeys a similar bracketing entropy bound

$$\log N_{[]}\left(\epsilon, \mathcal{I}_2, \|\cdot\|_{P,2}\right) \lesssim 1/\left(\epsilon^2 \log^4 \epsilon\right) + \log(1/\epsilon)$$

with a generic bracket taking the form $[i_2^0, i_2^1] = [\{W(\vartheta_k)'\beta_l - \kappa\delta\}T, \{W(\vartheta_k)'\beta_l + \kappa\delta\}T]$. Hence, this class is also P-Donsker with a constant envelope.

Step 3. In this step, we verify the claim (a). Note that $\mathcal{F} = e \cdot \mathcal{I}_1 \cdot \mathcal{I}_2$. This class has a square-integrable envelope under P. The class $\mathcal{F}$ is P-Donsker by the following argument. Note that the product $\mathcal{I}_1 \cdot \mathcal{I}_2$ of uniformly bounded classes is P-Donsker, for example, by Theorem 2.10.6 of van der Vaart and Wellner (1996). Under the stated assumption, the final product of the random variable e with the P-Donsker class remains to be P-Donsker by the multiplier Donsker theorem, namely Theorem 2.9.2 in van der Vaart and Wellner (1996).

Proof of Claim (b). The claim follows by the dominated convergence theorem, since any $f \in \mathcal{F}$ is dominated by a square-integrable envelope under P, and, uniformly in $y \in \mathcal{Y}$, $\Lambda[W(\vartheta)'\beta(y)]T \to \Lambda[W'\beta_0(y)]T$ and $|W(\vartheta)'\beta(y)T - W'\beta_0(y)T| \to 0$ in view of the relation such as (C.2).

Proof of Claim (c). This claim follows from the asymptotic equicontinuity of the empirical process $(\mathbb{G}_n[f_y], f_y \in \mathcal{F})$ under the $L_2(P)$ metric, and hence also with respect to the $\|\cdot\|_{T,\infty} \vee \|\cdot\|_2$ metric uniformly in $y \in \mathcal{Y}$ in view of Claim (b).

Proof of Claim (d). It is convenient to set $\widehat{f_y} := f_y(A, \widehat{\vartheta}, \widehat{\beta}(y))$ and $\widetilde{f_y} := f_y(A, \widetilde{\vartheta}, \widetilde{\beta}(y))$. Note that

$$\begin{aligned}\max_{1\leq j\leq \dim W} \left|\mathbb{G}_n[\widehat{f_y} - \widetilde{f_y}]\right|_j &\leq \max_{1\leq j\leq \dim W} \left|\sqrt{n}\mathbb{E}_n[\widehat{f_y} - \widetilde{f_y}]\right|_j + \max_{1\leq j\leq \dim W} \left|\sqrt{n}\mathrm{E}_P(\widehat{f_y} - \widetilde{f_y})\right|_j \\ &\lesssim \sqrt{n}\mathbb{E}_n[\widehat{\zeta}] + \sqrt{n}\mathrm{E}_P[\widehat{\zeta}] \lesssim \mathbb{G}_n[\widehat{\zeta}] + 2\sqrt{n}\mathrm{E}_P[\widehat{\zeta}],\end{aligned}$$

where $|f_y|_j$ denotes the jth element of an application of absolute value to each element of the vector f_y, and $\widehat{\zeta}$ is defined by the following relationship, which holds with probability approaching one uniformly in $y \in \mathcal{Y}$,

$$\begin{aligned}\max_{1\leq j\leq \dim W} |\widehat{f_y} - \widetilde{f_y}|_j &\lesssim |e| \cdot \left\{\left\|W(\widehat{\vartheta}) - W(\widetilde{\vartheta})\right\|_2 + \left|\Lambda\left[W(\widehat{\vartheta})'\widetilde{\beta}(y)\right] - \Lambda\left[W(\widetilde{\vartheta})'\widetilde{\beta}(y)\right]\right|\right\} \cdot T \\ &\lesssim \widehat{\zeta} := e \cdot \kappa\Delta_n,\end{aligned}$$

where $\kappa = L_W \max_{\beta\in\mathcal{B}} \|\beta\|_2 + L_W$, $L_W = \|\partial_v w\|_{T,\infty} \vee \|w\|_{T,\infty}$, and $\Delta_n = o(1/\sqrt{n})$ is a deterministic sequence such that

$$\Delta_n \geq \|\widehat{\vartheta} - \widetilde{\vartheta}\|_{T,\infty}.$$

By part (c), the result follows from

$$\mathbb{G}_n[\widehat{\zeta}] = \bar{o}_{\mathbb{P}}(1), \qquad \sqrt{n}\mathrm{E}_P[\widehat{\zeta}] = \bar{o}_{\mathbb{P}}(1).$$

Indeed,

$$\|e \cdot \kappa\Delta_n\|_{P,2} = \bar{o}(1) \quad \Rightarrow \quad \mathbb{G}_n[\widehat{\zeta}] = \bar{o}_{\mathbb{P}}(1),$$

and

$$\|e \cdot \kappa\Delta_n\|_{P,1} \leq \mathrm{E}_P|e| \cdot \kappa\Delta_n = \bar{o}(1/\sqrt{n}) \quad \Rightarrow \quad \mathrm{E}_P|\widehat{\zeta}| = \bar{o}_{\mathbb{P}}(1/\sqrt{n}),$$

since $\Delta_n = o(1/\sqrt{n})$. □

LEMMA C.2 (Local expansion). *Under Assumptions* 3–5, *for*

$$\widehat{\delta}(y) = \sqrt{n}\big(\widetilde{\beta}(y) - \beta_0(y)\big) = \bar{O}_{\mathbb{P}}(1);$$
$$\widehat{\Delta}(x,r) = \sqrt{n}\big(\widehat{\vartheta}(x,r) - \vartheta_0(x,r)\big) = \sqrt{n}\mathbb{E}_n\big[\ell(A,x,r)\big] + o_{\mathbb{P}}(1) \quad \textit{in } \ell^\infty(\overline{\mathcal{X}\mathcal{R}}),$$
$$\big\|\sqrt{n}\mathbb{E}_n\big[\ell(A,\cdot)\big]\big\|_{T,\infty} = O_{\mathbb{P}}(1),$$

we have that

$$\sqrt{n}\mathrm{E}_P\big[\{\Lambda\big[W(\widehat{\vartheta})'\widetilde{\beta}(y)\big] - 1(Y \leq y)\}W(\widehat{\vartheta})T\big] = J(y)\widehat{\delta}(y) + \sqrt{n}\mathbb{E}_n\big[g_y(A)\big] + \bar{o}_{\mathbb{P}}(1),$$

where

$$g_y(a) = \mathrm{E}_P\big\{\big[\Lambda\big(W'\beta_0(y)\big) - 1(Y \leq y)\big]\dot{W} + \lambda\big(W'\beta_0(y)\big)W\dot{W}'\beta_0(y)\big\}T\ell(a,X,R).$$

PROOF OF LEMMA C.2. Uniformly in $\xi := (X,Z) \in \overline{\mathcal{X}\mathcal{Z}}$ and $y \in \mathcal{Y}$,

$$\begin{aligned}
&\sqrt{n}\mathrm{E}_P\big\{\Lambda\big[W(\widehat{\vartheta})'\widetilde{\beta}(y)\big] - 1(Y \leq y) \mid X,Z\big\}T \\
&\quad = \sqrt{n}\mathrm{E}_P\big\{\Lambda\big[W'\beta_0(y)\big] - 1(Y \leq y) \mid X,Z\big\}T \\
&\qquad + \lambda\big[W(\bar{\vartheta}_\xi)'\bar{\beta}_\xi(y)\big]\big\{W(\bar{\vartheta}_\xi)'\widehat{\delta}(y) + \dot{W}(\bar{\vartheta}_\xi)'\bar{\beta}_\xi\widehat{\Delta}(X,R)\big\}T \\
&\quad = \sqrt{n}\mathrm{E}_P\big\{\Lambda\big[W'\beta_0(y)\big] - 1(Y \leq y) \mid X,Z\big\}T \\
&\qquad + \lambda\big[W'\beta_0(y)\big]\big\{W'\widehat{\delta}(y) + \dot{W}'\beta_0(y)\widehat{\Delta}(X,R)\big\}T + R_\xi(y),
\end{aligned}$$

and

$$\bar{R}(y) = \sup_{\{\xi \in \overline{\mathcal{X}\mathcal{Z}}\}} \big|R_\xi(y)\big| = \bar{o}_{\mathbb{P}}(1),$$

where $\bar{\vartheta}_\xi$ is on the line connecting ϑ_0 and $\widehat{\vartheta}$ and $\bar{\beta}_\xi(y)$ is on the line connecting $\beta_0(y)$ and $\widetilde{\beta}(y)$. The first equality follows by the mean value expansion. The second equality follows by uniform continuity of $\lambda(\cdot)$, uniform continuity of $W(\cdot)$ and $\dot{W}(\cdot)$, and by $\|\widehat{\vartheta} - \vartheta_0\|_{T,\infty} \to_{\mathbb{P}} 0$ and $\sup_{y\in\mathcal{Y}}\|\widetilde{\beta}(y) - \beta_0(y)\|_2 \to_{\mathbb{P}} 0$.

Since $\lambda(\cdot)$ and the entries of W and $\dot{W}$ are bounded, $\widehat{\delta}(y) = \bar{O}_{\mathbb{P}}(1)$, and $\|\widehat{\Delta}\|_{T,\infty} = O_{\mathbb{P}}(1)$, with probability approaching one uniformly in $y \in \mathcal{Y}$,

$$\begin{aligned}
&\sqrt{n}\mathrm{E}_P\big\{\Lambda\big[W(\widehat{\vartheta})'\widetilde{\beta}(y)\big] - 1(Y \leq y)\big\}W(\widehat{\vartheta})T \\
&\quad = \mathrm{E}_P\big\{\Lambda\big(W'\beta_0(y)\big) - 1(Y \leq y)\big\}\dot{W}T\widehat{\Delta}(X,R)
\end{aligned}$$

$$
\begin{aligned}
&+ \mathrm{E}_P\{\lambda[W'\beta_0(y)]WW'T\}\widehat{\delta}(y) + \mathrm{E}_P\{\lambda[W'\beta_0(y)]W\dot{W}'\beta_0(y)T\widehat{\Delta}(X,R)\} + O_{\mathbb{P}}(\bar{R}(y)) \\
&= J(y)\widehat{\delta}(y) + \mathrm{E}_P[\{\Lambda(W'\beta_0(y)) - 1(Y \le y)\}\dot{W} + \lambda[W'\beta_0(y)]W\dot{W}'\beta_0(y)]T\widehat{\Delta}(X,R) \\
&\quad + o_{\mathbb{P}}(1).
\end{aligned}
$$

Substituting in $\widehat{\Delta}(x,r) = \sqrt{n}\mathbb{E}_n[\ell(A,x,r)] + o_{\mathbb{P}}(1)$ and interchanging E_P and $\mathbb{E}_n$, we obtain

$$
\begin{aligned}
&\mathrm{E}_P[\{\Lambda(W'\beta_0(y)) - 1(Y \le y)\}\dot{W} + \lambda[W'\beta_0(y)]W\dot{W}'\beta_0(y)]T\widehat{\Delta}(X,R) \\
&\quad = \sqrt{n}\mathbb{E}_n[g_y(A)] + \bar{o}_{\mathbb{P}}(1),
\end{aligned}
$$

since $[\{\Lambda(W'\beta_0(y)) - 1(Y \le y)\}\dot{W} + \lambda[W'\beta_0(y)]W\dot{W}'\beta_0(y)]T$ is bounded uniformly in $y \in \mathcal{Y}$. The claim of the lemma follows. □

C.2.2 *Proof of Lemma 3* The proof is divided in two parts corresponding to the FCLT and bootstrap FCLT.

Part 1: FCLT. In this part, we show $\sqrt{n}(\widehat{\beta}(y) - \beta_0(y)) \rightsquigarrow J(y)^{-1}G(y)$ in $\ell^\infty(\mathcal{Y})^{d_w}$.

Step 1. This step shows that $\sqrt{n}(\widehat{\beta}(y) - \beta_0(y)) = \bar{O}_{\mathbb{P}}(1)$.

Recall that

$$
\widehat{\beta}(y) = \arg\min_{\beta \in \mathbb{R}^{\dim(W)}} \mathbb{E}_n[\rho_y(Y, W(\widehat{\vartheta})'\beta)T].
$$

Due to convexity of the objective function, it suffices to show that for any $\epsilon > 0$ there exists a finite positive constant B_ϵ such that uniformly in $y \in \mathcal{Y}$,

$$
\liminf_{n\to\infty} \mathbb{P}\Big(\inf_{\|\eta\|_2=1} \sqrt{n}\eta'\mathbb{E}_n[\widehat{f}_{\eta,B_\epsilon,y}] > 0\Big) \ge 1 - \epsilon, \tag{C.3}
$$

where

$$
\widehat{f}_{\eta,B_\epsilon,y}(A) := \{\Lambda[W(\widehat{\vartheta})'(\beta_0(y) + B_\epsilon\eta/\sqrt{n})] - 1(Y \le y)\}W(\widehat{\vartheta})T.
$$

Let

$$
f_y(A) := \{\Lambda[W'\beta_0(y)] - 1(Y \le y)\}WT.
$$

Then uniformly in $\|\eta\|_2 = 1$,

$$
\begin{aligned}
\sqrt{n}\eta'\mathbb{E}_n[\widehat{f}_{\eta,B_\epsilon,y}] &= \eta'\mathbb{G}_n[\widehat{f}_{\eta,B_\epsilon,y}] + \sqrt{n}\eta'\mathrm{E}_P[\widehat{f}_{\eta,B_\epsilon,y}] \\
&=_{(1)} \eta'\mathbb{G}_n[f_y] + \bar{o}_{\mathbb{P}}(1) + \eta'\sqrt{n}\mathrm{E}_P[\widehat{f}_{\eta,B_\epsilon,y}] \\
&=_{(2)} \eta'\mathbb{G}_n[f_y] + \bar{o}_{\mathbb{P}}(1) + \eta'J(y)\eta B_\epsilon + \eta'\mathbb{G}_n[g_y] + \bar{o}_{\mathbb{P}}(1) \\
&=_{(3)} \bar{O}_{\mathbb{P}}(1) + \bar{o}_{\mathbb{P}}(1) + \eta'J(y)\eta B_\epsilon + \bar{O}_{\mathbb{P}}(1) + \bar{o}_{\mathbb{P}}(1),
\end{aligned}
$$

where relations (1) and (2) follow by Lemma C.1 and Lemma C.2 with $\widetilde{\beta}(y) = \beta_0(y) + B_\epsilon\eta/\sqrt{n}$, respectively, using that $\|\widehat{\vartheta} - \widetilde{\vartheta}\|_{T,\infty} = o_{\mathbb{P}}(1/\sqrt{n})$, $\widetilde{\vartheta} \in Y$, $\|\widetilde{\vartheta} - \vartheta_0\|_{T,\infty} = O_{\mathbb{P}}(1/\sqrt{n})$ and $\|\beta_0(y) + B_\epsilon\eta/\sqrt{n} - \beta_0(y)\|_2 = \bar{O}(1/\sqrt{n})$; relation (3) holds because f_y and g_y are P-Donsker by step-2 below. Since uniformly in $y \in \mathcal{Y}$, $J(y)$ is positive definite, with

minimal eigenvalue bounded away from zero, the inequality (C.3) follows by choosing B_ϵ as a sufficiently large constant.

Step 2. In this step, we show the main result. Let

$$\widehat{f}_y(A) := \{\Lambda[W(\widehat{\vartheta})'\widehat{\beta}(y)] - 1(Y \le y)\}W(\widehat{\vartheta})T.$$

From the first-order conditions of the distribution regression problem,

$$\begin{aligned}
0 &= \sqrt{n}\mathbb{E}_n[\widehat{f}_y] = \mathbb{G}_n[\widehat{f}_y] + \sqrt{n}\mathrm{E}_P[\widehat{f}_y] \\
&=_{(1)} \mathbb{G}_n[f_y] + \bar{o}_{\mathbb{P}}(1) + \sqrt{n}\mathrm{E}_P[\widehat{f}_y] \\
&=_{(2)} \mathbb{G}_n[f_y] + \bar{o}_{\mathbb{P}}(1) + J(y)\sqrt{n}\big(\widehat{\beta}(y) - \beta_0(y)\big) + \mathbb{G}_n[g_y] + \bar{o}_{\mathbb{P}}(1),
\end{aligned}$$

where relations (1) and (2) follow by Lemma C.1 and Lemma C.2 with $\widetilde{\beta}(y) = \widehat{\beta}(y)$, respectively, using that $\|\widehat{\vartheta} - \widetilde{\vartheta}\|_{T,\infty} = o_{\mathbb{P}}(1/\sqrt{n})$, $\widetilde{\vartheta} \in \Upsilon$, and $\|\widetilde{\vartheta} - \vartheta\|_{T,\infty} = O_{\mathbb{P}}(1/\sqrt{n})$ by Lemma 2, and $\|\widehat{\beta}(y) - \beta_0(y)\|_2 = \bar{O}_{\mathbb{P}}(1/\sqrt{n})$.

Therefore, by uniform invertibility of $J(y)$ in $y \in \mathcal{Y}$,

$$\sqrt{n}\big(\widehat{\beta}(y) - \beta_0(y)\big) = -J(y)^{-1}\mathbb{G}_n(f_y + g_y) + \bar{o}_{\mathbb{P}}(1).$$

The function f_y is P-Donsker by standard argument for distribution regression (e.g., step 3 in the proof of Theorem 5.2 of Chernozhukov, Fernández-Val, and Melly (2013)). Similarly, g_y is P-Donsker by Example 19.7 in van der Vaart (2000) because $g_y \in \{h_y(A) : |h_y(A) - h_v(A)| \le M(A)|y - v|; \mathrm{E}_P M(A)^2 < \infty; y, v \in \mathcal{Y}\}$, since

$$|g_y - g_v| \le L\mathrm{E}_P\big[T\big|\ell(a, X, R)\big]\big|_{a=A}|y - v|,$$

with $L = 2L_W + L_W^2 \max_{\beta\in\mathcal{B}} \|\beta\|_2/4$, $L_W := \|\partial_v w\|_{T,\infty} \vee \|w\|_{T,\infty}$, and $\mathrm{E}_P[T\ell(A, X, R)^2] < \infty$ by Lemma 2. Hence, by the functional central limit theorem

$$\mathbb{G}_n(f_y + g_y) \rightsquigarrow G(y) \quad \text{in } \ell^\infty(\mathcal{Y})^{d_w},$$

where $y \mapsto G(y)$ is a zero mean Gaussian process with uniformly continuous sample paths and the covariance function $C(y, v)$ specified in the lemma. Conclude that

$$\sqrt{n}\big(\widehat{\beta}(y) - \beta_0(y)\big) \rightsquigarrow J(y)^{-1}G(y) \quad \text{in } \ell^\infty(\mathcal{Y})^{d_w}.$$

Part 2: Bootstrap FCLT. In this part, we show $\sqrt{n}(\widehat{\beta}^e(y) - \widehat{\beta}(y)) \rightsquigarrow_{\mathbb{P}} J(y)^{-1}G(y)$ in $\ell^\infty(\mathcal{Y})^{d_w}$.

Step 1. This step shows that $\sqrt{n}(\widehat{\beta}^e(y) - \beta_0(y)) = \bar{O}_{\mathbb{P}}(1)$ under the unconditional probability $\mathbb{P}$.

Recall that

$$\widehat{\beta}^e(y) = \arg\min_{\beta\in\mathbb{R}^{\dim(W)}} \mathbb{E}_n\big[e\rho_y\big(Y, W\big(\widehat{\vartheta}^e\big)'\beta\big)T\big],$$

where e is the random variable used in the weighted bootstrap. Due to convexity of the objective function, it suffices to show that for any $\epsilon > 0$ there exists a finite positive constant B_ϵ such that uniformly in $y \in \mathcal{Y}$,

$$\liminf_{n\to\infty} \mathbb{P}\Big(\inf_{\|\eta\|_2=1} \sqrt{n}\eta'\mathbb{E}_n\big[\widehat{f}^e_{\eta,B_\epsilon,y}\big] > 0\Big) \ge 1 - \epsilon, \tag{C.4}$$

where

$$\widehat{f}^e_{\eta,B_\epsilon,y}(A) := e \cdot \big\{\Lambda\big[W\big(\widehat{\vartheta}^e\big)'\big(\beta_0(y) + B_\epsilon \eta/\sqrt{n}\big)\big] - 1(Y \le y)\big\} W\big(\widehat{\vartheta}^e\big) T.$$

Let

$$f^e_y(A) := e \cdot \big\{\Lambda\big[W'\beta_0(y)\big] - 1(Y \le y)\big\} WT.$$

Then uniformly in $\|\eta\|_2 = 1$,

$$\begin{aligned}
\sqrt{n}\eta'\mathbb{E}_n\big[\widehat{f}^e_{\eta,B_\epsilon,y}\big] &= \eta'\mathbb{G}_n\big[\widehat{f}^e_{\eta,B_\epsilon,y}\big] + \sqrt{n}\eta'\mathrm{E}_P\big[\widehat{f}^e_{\eta,B_\epsilon,y}\big] \\
&=_{(1)} \eta'\mathbb{G}_n\big[f^e_y\big] + \bar{o}_{\mathbb{P}}(1) + \eta'\sqrt{n}\mathrm{E}_P\big[\widehat{f}^e_{\eta,B_\epsilon,y}\big] \\
&=_{(2)} \eta'\mathbb{G}_n\big[f^e_y\big] + \bar{o}_{\mathbb{P}}(1) + \eta' J(y)\eta B_\epsilon + \eta'\mathbb{G}_n\big[g^e_y\big] + \bar{o}_{\mathbb{P}}(1) \\
&=_{(3)} \bar{O}_{\mathbb{P}}(1) + \bar{o}_{\mathbb{P}}(1) + \eta' J(y)\eta B_\epsilon + \bar{O}_{\mathbb{P}}(1) + \bar{o}_{\mathbb{P}}(1),
\end{aligned}$$

where relations (1) and (2) follow by Lemma C.1 and Lemma C.2 with $\widetilde{\beta}(y) = \beta_0(y) + B_\epsilon\eta/\sqrt{n}$, respectively, using that $\|\widehat{\vartheta}^e - \widetilde{\vartheta}^e\|_{T,\infty} = o_{\mathbb{P}}(1/\sqrt{n})$, $\widetilde{\vartheta}^e \in \Upsilon$ and $\|\widetilde{\vartheta}^e - \vartheta_0\|_{T,\infty} = O_{\mathbb{P}}(1/\sqrt{n})$ by Lemma 2, and $\|\beta_0(y) + B_\epsilon\eta/\sqrt{n} - \beta_0(y)\|_2 = \bar{O}(1/\sqrt{n})$; relation (3) holds because $f^e_y = e \cdot f_y$ and $g^e_y = e \cdot g_y$, where f_y and g_y are P-Donsker by step-2 of the proof of Theorem 3 and $\mathrm{E}_P e^2 < \infty$. Since uniformly in $y \in \mathcal{Y}$, $J(y)$ is positive definite, with minimal eigenvalue bounded away from zero, the inequality (C.4) follows by choosing B_ϵ as a sufficiently large constant.

Step 2. In this step, we show that $\sqrt{n}(\widehat{\beta}^e(y) - \beta_0(y)) = -J(y)^{-1}\mathbb{G}_n(f^e_y + g^e_y) + \bar{o}_{\mathbb{P}}(1)$ under the unconditional probability $\mathbb{P}$.

Let

$$\widehat{f}^e_y(A) := e \cdot \big\{\Lambda\big[W\big(\widehat{\vartheta}^e\big)'\widehat{\beta}^e(y)\big] - 1(Y \le y)\big\} W\big(\widehat{\vartheta}^e\big) T.$$

From the first-order conditions of the distribution regression problem in the weighted sample, uniformly in $y \in \mathcal{Y}$,

$$\begin{aligned}
0 = \sqrt{n}\mathbb{E}_n\big[\widehat{f}^e_y\big] &= \mathbb{G}_n\big[\widehat{f}^e_y\big] + \sqrt{n}\mathrm{E}_P\big[\widehat{f}^e_y\big] \\
&=_{(1)} \mathbb{G}_n\big[f^e_y\big] + \bar{o}_{\mathbb{P}}(1) + \sqrt{n}\mathrm{E}_P\big[\widehat{f}^e_y\big] \\
&=_{(2)} \mathbb{G}_n\big[f^e_y\big] + \bar{o}_{\mathbb{P}}(1) + J(y)\sqrt{n}\big(\widehat{\beta}^e(y) - \beta_0(y)\big) + \mathbb{G}_n\big[g^e_y\big] + \bar{o}_{\mathbb{P}}(1),
\end{aligned}$$

where relations (1) and (2) follow by Lemma C.1 and Lemma C.2 with $\widetilde{\beta}(y) = \widehat{\beta}^e(y)$, respectively, using that $\|\widehat{\vartheta}^e - \widetilde{\vartheta}^e\|_{T,\infty} = o_{\mathbb{P}}(1/\sqrt{n})$, $\widetilde{\vartheta}^e \in \Upsilon$ and $\|\widetilde{\vartheta}^e - \vartheta_0\|_{T,\infty} = O_{\mathbb{P}}(1/\sqrt{n})$ by Lemma 2, and $\|\widehat{\beta}^e(y) - \beta_0(y)\|_2 = \bar{O}_{\mathbb{P}}(1/\sqrt{n})$.

Therefore, by uniform invertibility of $J(y)$ in $y \in \mathcal{Y}$,

$$\sqrt{n}\big(\widehat{\beta}^e(y) - \beta_0(y)\big) = -J(y)^{-1}\mathbb{G}_n\big(f^e_y + g^e_y\big) + \bar{o}_{\mathbb{P}}(1).$$

Step 3. In this final step, we establish the behavior of $\sqrt{n}(\widehat{\beta}^e(y) - \widehat{\beta}(y))$ under $\mathbb{P}^e$. Note that $\mathbb{P}^e$ denotes the conditional probability measure, namely the probability measure

induced by draws of $e_1, \ldots, e_n$ conditional on the data $A_1, \ldots, A_n$. By Step 2 of the proof of Theorem 1 and Step 2 of this proof, we have that under $\mathbb{P}$:

$$\sqrt{n}\big(\widehat{\beta}^e(y) - \beta_0(y)\big) = -J(y)^{-1}\mathbb{G}_n\big(f_y^e + g_y^e\big) + \bar{o}_{\mathbb{P}}(1),$$
$$\sqrt{n}\big(\widehat{\beta}(y) - \beta_0(y)\big) = -J(y)^{-1}\mathbb{G}_n(f_y + g_y) + \bar{o}_{\mathbb{P}}(1).$$

Hence, under $\mathbb{P}$

$$\begin{aligned}\sqrt{n}\big(\widehat{\beta}^e(y) - \widehat{\beta}(y)\big) &= -J(y)^{-1}\mathbb{G}_n\big(f_y^e - f_y + g_y^e - g_y\big) + r_n(y)\\ &= -J(y)^{-1}\mathbb{G}_n\big((e-1)(f_y + g_y)\big) + r_n(y),\end{aligned}$$

where $r_n(y) = \bar{o}_{\mathbb{P}}(1)$. Note that it is also true that

$$r_n(y) = \bar{o}_{\mathbb{P}^e}(1) \quad \text{in } \mathbb{P}\text{-probability},$$

where the latter statement means that for every $\epsilon > 0$, $\mathbb{P}^e(\|r_n(y)\|_2 > \epsilon) = \bar{o}_{\mathbb{P}}(1)$. Indeed, this follows from Markov inequality and by

$$\mathrm{E}_{\mathbb{P}}\big[\mathbb{P}^e\big(\big\|r_n(y)\big\|_2 > \epsilon\big)\big] = \mathbb{P}\big(\big\|r_n(y)\big\|_2 > \epsilon\big) = \bar{o}(1),$$

where the latter holds by the law of iterated expectations and $r_n(y) = \bar{o}_{\mathbb{P}}(1)$.

Note that $f_y^e = e \cdot f_y$ and $g_y^e = e \cdot g_y$, where f_y and g_y are P-Donsker by step-2 of the proof of the first part and $\mathrm{E}_P e^2 < \infty$. Then, by the conditional multiplier functional central limit theorem, for example, Theorem 2.9.6 in van der Vaart and Wellner (1996),

$$G_n^e(y) := \mathbb{G}_n\big((e-1)(f_y + g_y)\big) \rightsquigarrow_{\mathbb{P}} G(y) \quad \text{in } \ell^\infty(\mathcal{Y})^{d_w}.$$

Conclude that

$$\sqrt{n}\big(\widehat{\beta}^e(y) - \widehat{\beta}(y)\big) \rightsquigarrow_{\mathbb{P}} J(y)^{-1}G(y) \quad \text{in } \ell^\infty(\mathcal{Y})^{d_w}.$$

C.3 *Proof of Theorems 2–4*

In this section, we use the notation $W_x(\vartheta) = w(x, Z_1, V(\vartheta))$ such that $W_x = w(x, Z_1, V(\vartheta_0))$. Again we focus on the case where $\mathcal{Y}$ is a compact interval of $\mathbb{R}$.

C.3.1 *Proof of Theorem 2* The result follows by a similar argument to the proof of Lemma 3 using Lemmas C.3 and C.4 in place of Lemmas C.1 and C.2, and the delta method. For the sake of brevity, here we just outline the proof of the FCLT.

Let $\psi_x(A, \vartheta, \beta) := \Lambda(W_x(\vartheta)'\beta)T$ such that $G_T(y, x) = \mathrm{E}_P\psi_x(A, \vartheta_0, \beta_0(y))/\mathrm{E}_P T$ and $\widehat{G}(y, x) = \mathbb{E}_n\psi_x(A, \widehat{\vartheta}, \widehat{\beta}(y))/\mathbb{E}_n T$. Then, for $\widehat{\psi}_{y,x} := \psi_x(A, \widehat{\vartheta}, \widehat{\beta}(y))$ and $\psi_{y,x} := \psi_x(A, \vartheta_0, \beta_0(y))$,

$$\begin{aligned}&\sqrt{n}\big[\mathbb{E}_n\psi_x\big(A, \widehat{\vartheta}, \widehat{\beta}(y)\big) - \mathrm{E}_P\psi_x\big(A, \vartheta_0, \beta_0(y)\big)\big]\\ &\quad= \mathbb{G}_n[\widehat{\psi}_{y,x}] + \sqrt{n}\mathrm{E}_P[\widehat{\psi}_{y,x} - \psi_{y,x}]\\ &\quad=_{(1)} \mathbb{G}_n[\psi_{y,x}] + \bar{o}_{\mathbb{P}}(1) + \sqrt{n}\mathrm{E}_P[\widehat{\psi}_{y,x} - \psi_{y,x}]\\ &\quad=_{(2)} \mathbb{G}_n[\psi_{y,x}] + \bar{o}_{\mathbb{P}}(1) + \mathbb{G}_n[h_{y,x}] + \bar{o}_{\mathbb{P}}(1),\end{aligned}$$

where relations (1) and (2) follow by Lemma C.3 and Lemma C.4 with $\widetilde{\beta}(y) = \widehat{\beta}(y)$, respectively, using that $\|\widehat{\vartheta} - \widetilde{\vartheta}\|_{T,\infty} = o_{\mathbb{P}}(1/\sqrt{n})$, $\widetilde{\vartheta} \in \Upsilon$, and $\|\widetilde{\vartheta} - \vartheta\|_{T,\infty} = O_{\mathbb{P}}(1/\sqrt{n})$ by Lemma 2, and $\sqrt{n}(\widehat{\beta}(y) - \beta_0(y)) = -J(y)^{-1}\mathbb{G}_n(f_y + g_y) + \bar{o}_{\mathbb{P}}(1)$ from step 2 of the proof of Lemma 3.

The functions $(y, x) \mapsto \psi_{y,x}$ and $(y, x) \mapsto h_{y,x}$ are P-Donsker by Example 19.7 in van der Vaart (2000) because they are Lipschitz continuous on $\mathcal{Y}\overline{\mathcal{X}}$. Hence, by the functional central limit theorem

$$\mathbb{G}_n(\psi_{y,x} + h_{y,x}) \rightsquigarrow Z(y, x) \quad \text{in } \ell^\infty(\mathcal{Y}\overline{\mathcal{X}}),$$

where $(y, x) \mapsto Z(y, x)$ is a zero mean Gaussian process with uniformly continuous sample paths and covariance function

$$\mathrm{Cov}_P[\psi_{y,x} + h_{y,x}, \psi_{v,u} + h_{v,u}], \quad (y, x), (v, u) \in \mathcal{Y}\overline{\mathcal{X}}.$$

The result follows by the functional delta method applied to the ratio of $\mathbb{E}_n\psi_x(A, \widehat{\vartheta}, \widehat{\beta}(y))$ and $\mathbb{E}_n T$ using that

$$\begin{pmatrix} \mathbb{G}_n\psi_x\big(A, \widehat{\vartheta}, \widehat{\beta}(y)\big) \\ \mathbb{G}_n T \end{pmatrix} \rightsquigarrow \begin{pmatrix} Z(y, x) \\ Z_T \end{pmatrix},$$

where $Z_T \sim N(0, p_T(1 - p_T))$,

$$\mathrm{Cov}_P\big(Z(y, x), Z_T\big) = G_T(y, x)\, p_T(1 - p_T),$$

and

$$\begin{aligned} &\mathrm{Cov}_P[\psi_{y,x} + h_{y,x}, \psi_{v,u} + h_{v,u} \mid T = 1] \\ &\quad = \frac{\mathrm{Cov}_P[\psi_{y,x} + h_{y,x}, \psi_{v,u} + h_{v,u}] - G_T(y, x)G_T(v, u)\, p_T(1 - p_T)}{p_T}. \end{aligned}$$

Lemma C.3 (Stochastic equicontinuity). *Let $e \geq 0$ be a positive random variable with $\mathrm{E}_P[e] = 1$, $\mathrm{Var}_P[e] = 1$, and $\mathrm{E}_P|e|^{2+\delta} < \infty$ for some $\delta > 0$, that is independent of (Y, X, Z, W, V), including as a special case $e = 1$, and set, for $A = (e, Y, X, Z, W, V)$,*

$$\psi_x(A, \vartheta, \beta) := e \cdot \Lambda\big(W_x(\vartheta)'\beta\big) \cdot T.$$

Under Assumptions 3–5, *the following relations are true*:

(a) *Consider the set of functions*

$$\mathcal{F} := \big\{\psi_x(A, \vartheta, \beta) : (\vartheta, \beta, x) \in \Upsilon_0 \times \mathcal{B} \times \overline{\mathcal{X}}\big\},$$

where $\overline{\mathcal{X}}$ is a compact subset of $\mathbb{R}$, $\mathcal{B}$ is a compact set under the $\|\cdot\|_2$ metric containing $\beta_0(y)$ for all $y \in \mathcal{Y}$, Υ_0 is the intersection of Υ, defined in Lemma 2, *with a neighborhood of ϑ_0 under the $\|\cdot\|_{T,\infty}$ metric. This class is P-Donsker with a square integrable envelope of the form e times a constant.*

(b) *Moreover, if* $(\vartheta, \beta(y)) \to (\vartheta_0, \beta_0(y))$ *in the* $\|\cdot\|_{T,\infty} \vee \|\cdot\|_2$ *metric uniformly in* $y \in \mathcal{Y}$, *then*

$$\sup_{(y,x)\in\mathcal{Y}\overline{\mathcal{X}}} \big\|\psi_x\big(A, \vartheta, \beta(y)\big) - \psi_x\big(A, \vartheta_0, \beta_0(y)\big)\big\|_{P,2} \to 0.$$

(c) *Hence for any* $(\widetilde{\vartheta}, \widetilde{\beta}(y)) \to_{\mathbb{P}} (\vartheta_0, \beta_0(y))$ *in the* $\|\cdot\|_{T,\infty} \vee \|\cdot\|_2$ *metric uniformly in* $y \in \mathcal{Y}$ *such that* $\widetilde{\vartheta} \in \Upsilon_0$,

$$\sup_{(y,x)\in\mathcal{Y}\overline{\mathcal{X}}} \big\|\mathbb{G}_n\psi_x\big(A, \widetilde{\vartheta}, \widetilde{\beta}(y)\big) - \mathbb{G}_n\psi_x\big(A, \vartheta_0, \beta_0(y)\big)\big\|_2 \to_{\mathbb{P}} 0.$$

(d) *For any* $(\widehat{\vartheta}, \widetilde{\beta}(y)) \to_{\mathbb{P}} (\vartheta_0, \beta_0(y))$ *in the* $\|\cdot\|_{T,\infty} \vee \|\cdot\|_2$ *metric uniformly in* $y \in \mathcal{Y}$, *so that*

$$\|\widehat{\vartheta} - \widetilde{\vartheta}\|_{T,\infty} = o_{\mathbb{P}}(1/\sqrt{n}), \quad \text{where } \widetilde{\vartheta} \in \Upsilon_0,$$

we have that

$$\sup_{(y,x)\in\mathcal{Y}\overline{\mathcal{X}}} \big\|\mathbb{G}_n\psi_x\big(A, \widehat{\vartheta}, \widetilde{\beta}(y)\big) - \mathbb{G}_n\psi_x\big(A, \vartheta_0, \beta_0(y)\big)\big\|_2 \to_{\mathbb{P}} 0.$$

Proof of Lemma C.3. The proof is omitted because is similar to the proof of Lemma C.1. □

Lemma C.4 (Local expansion). *Under Assumptions* 3–5, *for*

$$\begin{aligned}
\widehat{\delta}(y) &= \sqrt{n}\big(\widetilde{\beta}(y) - \beta_0(y)\big) = \bar{O}_{\mathbb{P}}(1);\\
\widehat{\Delta}(x,r) &= \sqrt{n}\big(\widehat{\vartheta}(x,r) - \vartheta_0(x,r)\big)\\
&= \sqrt{n}\mathbb{E}_n\big[\ell(A,x,r)\big] + o_{\mathbb{P}}(1) \quad \text{in } \ell^\infty(\overline{\mathcal{X}\mathcal{R}}),\\
\big\|\sqrt{n}\mathbb{E}_n\big[\ell(A,\cdot)\big]\big\|_{T,\infty} &= O_{\mathbb{P}}(1),
\end{aligned}$$

we have that

$$\begin{aligned}
&\sqrt{n}\big\{\mathrm{E}_P\Lambda\big[W_x(\widehat{\vartheta})'\widetilde{\beta}(y)\big]T - \mathrm{E}_P\Lambda\big[W_x'\beta_0(y)\big]T\big\}\\
&\quad = \mathrm{E}_P\big\{\lambda\big[W_x'\beta_0(y)\big]W_xT\big\}'\widehat{\delta}(y) + \mathrm{E}_P\big\{\lambda\big[W_x'\beta_0(y)\big]\dot{W}_x'\beta_0(y)T\ell(a,X,R)\big\}|_{a=A} + \bar{o}_{\mathbb{P}}(1),
\end{aligned}$$

where $\bar{o}_{\mathbb{P}}(1)$ *denotes order in probability uniform in* $(y,x) \in \mathcal{Y}\overline{\mathcal{X}}$.

Proof of Lemma C.4. The proof is omitted because is similar to the proof of Lemma C.2. □

C.3.2 *Proof of Theorem 3* The result follows from Theorem 2 and the functional delta method, because the map $\phi : H \mapsto \int_{\mathcal{Y}^+} 1(H(y,x) \le \tau)\,dy - \int_{\mathcal{Y}^-} 1(H(y,x) \ge \tau)\,dy$ is Hadamard differentiable at $H = G_T$ under the conditions of the theorem by Proposition 2 of Chernozhukov, Fernández-Val, and Galichon (2010) with derivative

$$\phi'_{G_T}(h) = -\frac{h\big(\phi(\cdot,x),x\big)}{g_T\big(\phi(\cdot,x),x\big)}.$$

C.3.3 *Proof of Theorem 4* The result follows from Theorem 2 and the functional delta method, because the map $\varphi : H \mapsto \int_{\mathcal{Y}}[1(y \geq 0) - H(y,x)]\,dy$ is Hadamard differentiable at $H = G_T$ by Lemma C.5 with derivative

$$\varphi'_{G_T}(h) = -\int_{\mathcal{Y}} h(y,x)\nu(dy).$$

LEMMA C.5 (Hadamard differentiability of ASF Map). *Let $\mathcal{Y}$ be bounded. The ASF map $\varphi : \ell^\infty(\mathcal{Y}\overline{\mathcal{X}}) \to \ell^\infty(\overline{\mathcal{X}})$ defined by*

$$H \mapsto \varphi(H) := \int_{\mathcal{Y}} \big[1(y \geq 0) - H(y,x)\big]\nu(dy),$$

is Hadamard-differentiable at $H = G$, tangentially to the set of uniformly continuous functions on $\mathcal{Y}\overline{\mathcal{X}}$, with derivative map $h \mapsto \varphi'_G(h)$ defined by

$$\varphi'_G(h) := -\int_{\mathcal{Y}} h(y,x)\nu(dy),$$

where the derivative is defined and is continuous on $\ell^\infty(\mathcal{Y}\overline{\mathcal{X}})$.

PROOF OF LEMMA C.5. Consider any sequence $H^t \in \ell^\infty(\mathcal{Y}\overline{\mathcal{X}})$ such that for $h^t := (H^t - G)/t$, $h^t \to h$ in $\ell^\infty(\mathcal{Y}\overline{\mathcal{X}})$ as $t \searrow 0$, where h is a uniformly continuous function on $\mathcal{Y}\overline{\mathcal{X}}$. We want to show that as $t \searrow 0$,

$$\frac{\varphi(H^t) - \varphi(G)}{t} - \varphi'_G(h) \to 0 \quad \text{in } \ell^\infty(\mathcal{Y}\overline{\mathcal{X}}).$$

The result follows because by linearity of the map φ

$$\frac{\varphi(H^t) - \varphi(G)}{t} = -\int_{\mathcal{Y}} h^t(y,x)\nu(dy) \to -\int_{\mathcal{Y}} h(y,x)\nu(dy) = \varphi'_G(h).$$

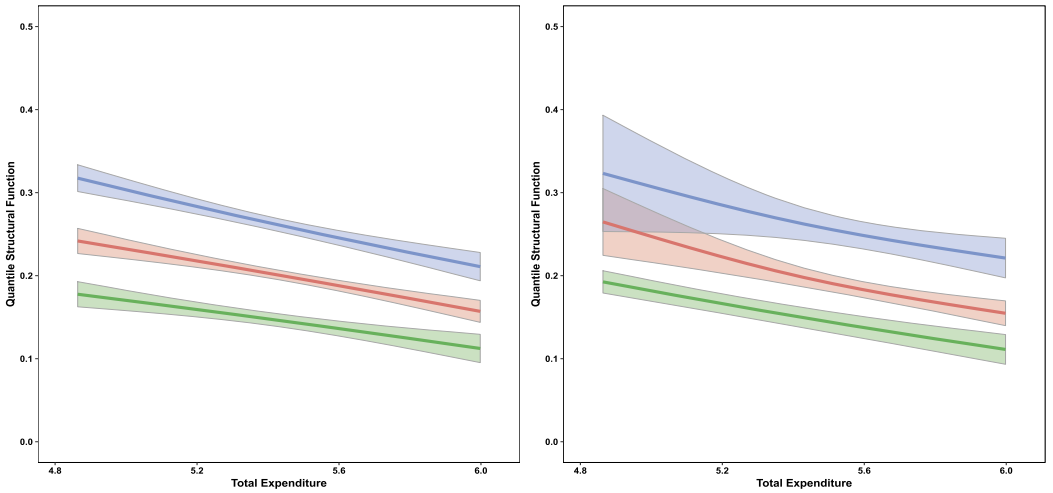
The derivative is well-defined over $\ell^\infty(\mathcal{Y}\overline{\mathcal{X}})$ and continuous with respect to the sup-norm on $\ell^\infty(\mathcal{Y}\overline{\mathcal{X}})$. □

APPENDIX D: ROBUSTNESS OF EMPIRICAL RESULTS

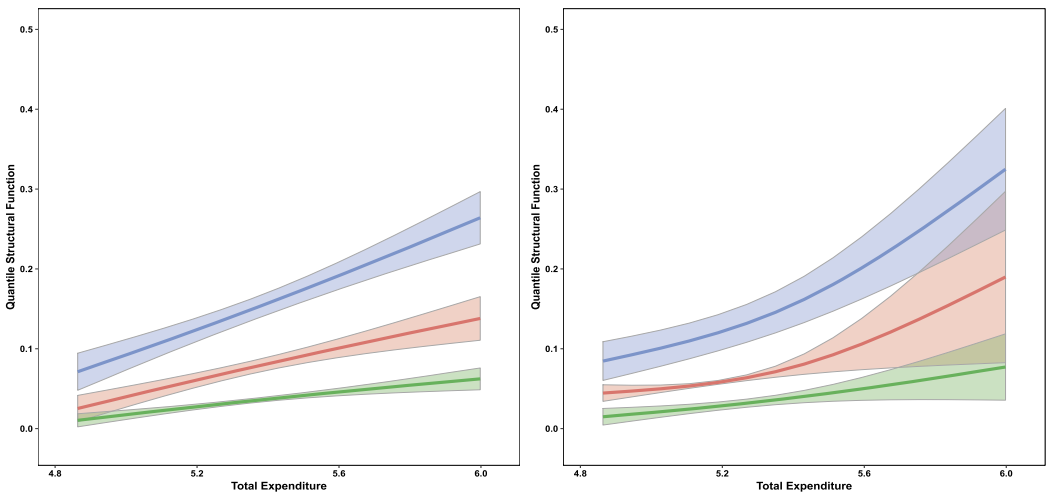
D.1 *Sensitivity analysis*

To further check the robustness of our empirical findings, we implemented a thorough sensitivity analysis and investigated several alternative specifications. We replicated all results of the empirical application in the main text using a probit specification for DR, as well as enforcing a trimming rule, yielding very similar results which we do not report for brevity. For both QR and DR, we report additional QSF estimates with varying grids of values of X, as well as for a different number of quantile levels.

For the equispaced grids $0.1 = t_1 < \cdots < t_K = 0.9$ and $0.15 = t_1^* < \cdots < t_K^* = 0.85$, let $\widetilde{\mathcal{X}}_K = \{\widehat{Q}_X(t_1), \ldots, \widehat{Q}_X(t_K)\}$ and $\widetilde{\mathcal{X}}_K^* = \{\widehat{Q}_X(t_1^*), \ldots, \widehat{Q}_X(t_K^*)\}$. Further let $\widetilde{\mathcal{T}}_3 = \{1/4, 1/2, 3/4\}$ and $\widetilde{\mathcal{T}}_5 = \{1/6, 1/3, 1/2, 2/3, 5/6\}$. Then Figures D.1–D.4 display QSFs and their uniform confidence bands obtained by setting the regions of interest as follows:



(A) Food.

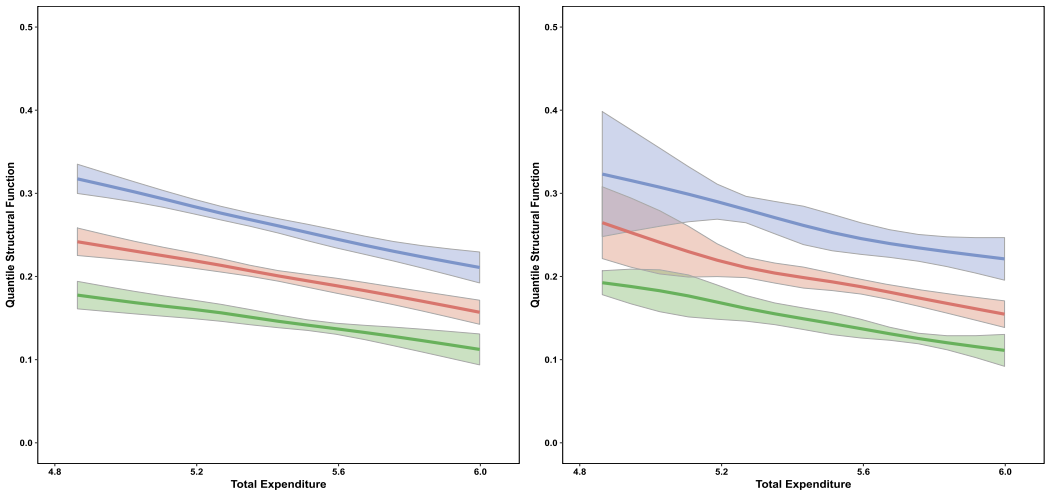


(B) Leisure.

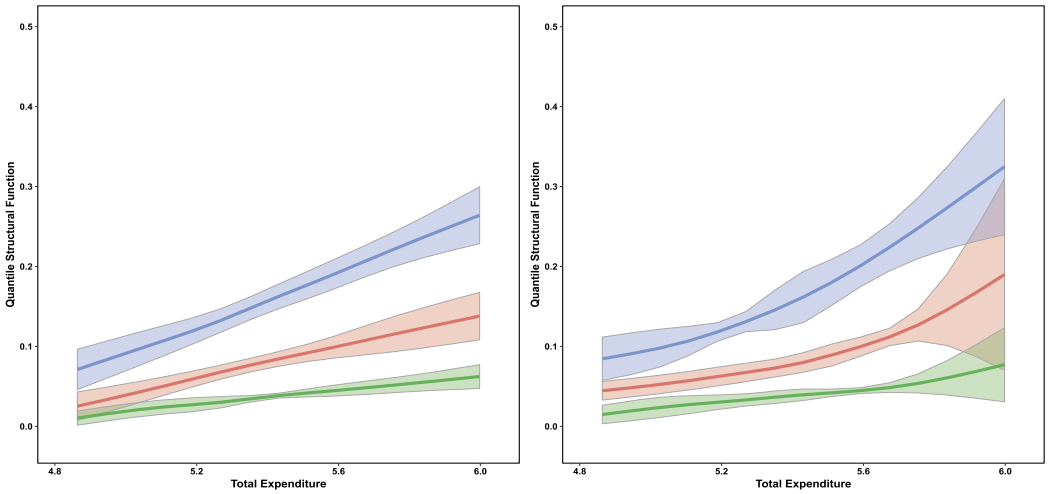
FIGURE D.1. QSF over $\widetilde{\mathcal{T}}_3\widetilde{\mathcal{X}}_3$. Quantile (left) and distribution regression (right).

(1) Figure D.1: we set $\mathcal{I}_Q = \widetilde{\mathcal{T}}_3\widetilde{\mathcal{X}}_3$,
(2) Figure D.2: we set $\mathcal{I}_Q = \widetilde{\mathcal{T}}_3\widetilde{\mathcal{X}}_7$,
(3) Figure D.3: we set $\mathcal{I}_Q = \widetilde{\mathcal{T}}_5\widetilde{\mathcal{X}}_5$,
(4) Figure D.4: we set $\mathcal{I}_Q = \widetilde{\mathcal{T}}_5\widetilde{\mathcal{X}}_5^*$.

QSF estimates across varying regions of interest confirm the results of the empirical application in the main text. For QR, varying the number of grid points has very little effect on the QSF estimates and confidence bands. For DR, QSF estimates are also almost identical across specifications, and only the shape of confidence bands varies according to K. For both goods, and both DR and QR methods, all specifications capture the features



(A) Food.

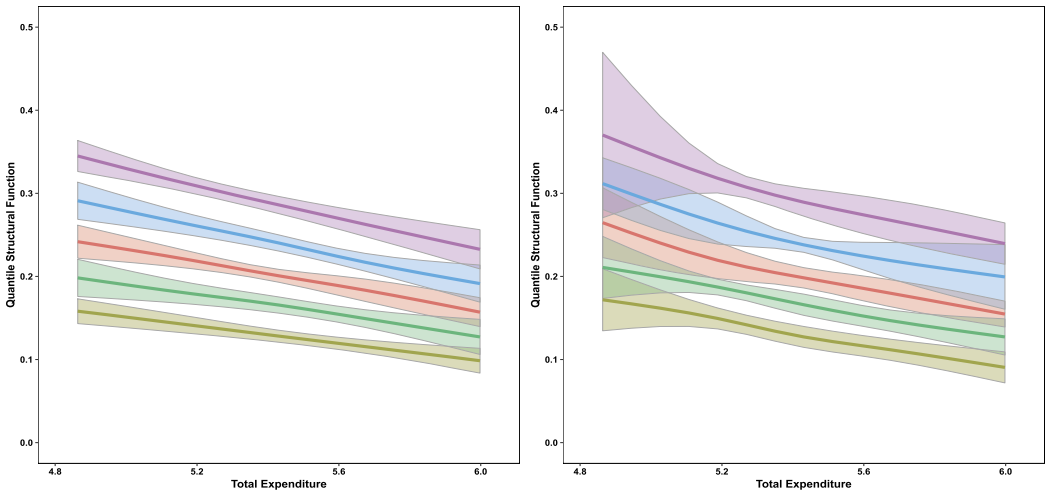


(B) Leisure.

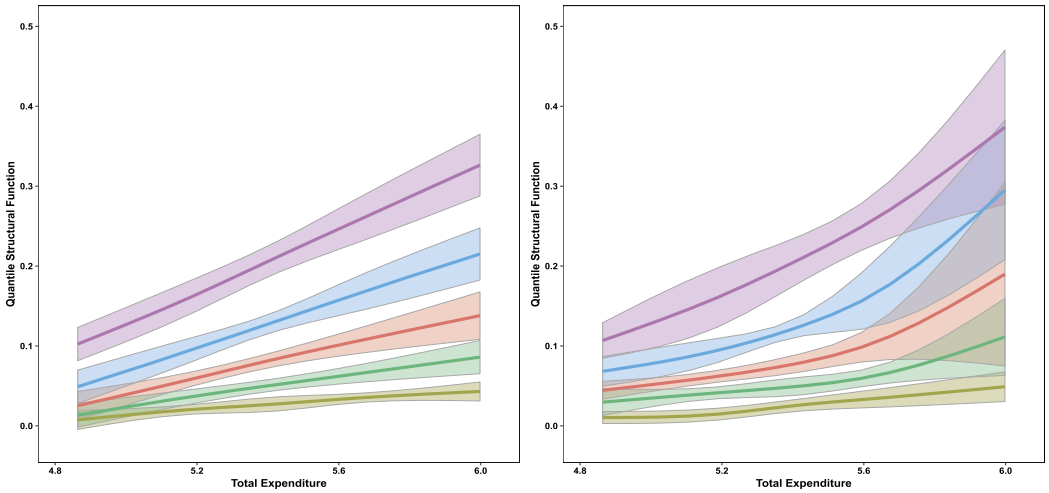
FIGURE D.2. QSF over $\widetilde{\mathcal{T}}_3\widetilde{\mathcal{X}}_7$. Quantile (left) and distribution regression (right).

emphasized in the main text: for both goods QSF estimates display heteroskedasticity, and estimates for leisure display asymmetry. These features are especially apparent in Figure D.3 which shows the QSF at 5 different quantile levels. Finally, comparing Figures D.1–D.4 shows that the length of confidence bands over $\widetilde{\mathcal{X}}$ is affected by the choice of end-points for $\widetilde{\mathcal{X}}_K$, especially so for DR estimates, but is robust to the choice of K.

For QR, we also check the robustness of our ASF estimates by comparing them to those obtained based on the least-squares projection characterization of the ASF for the QR baseline given in Remark 4 in the main text. Figure D.5 shows that the two estimates are very similar for both food and leisure share expenditure.



(A) Food.



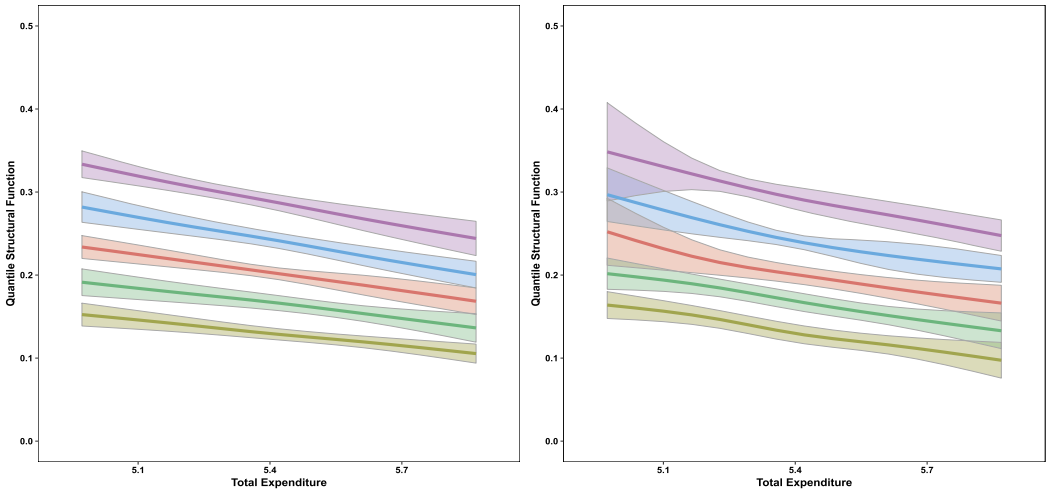
(B) Leisure.

FIGURE D.3. QSF over $\widetilde{\mathcal{T}}_5\widetilde{\mathcal{X}}_5$. Quantile (left) and distribution regression (right).

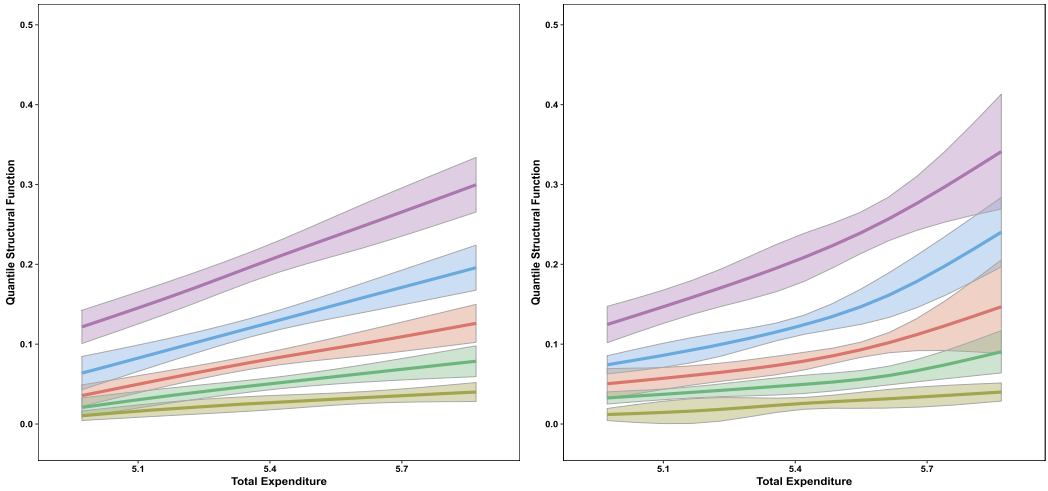
D.2 *Flexible specifications*

Our estimators can also easily accommodate additional powers of the control function $\Phi^{-1}(V)^k$ as well as interaction terms $X \cdot \Phi^{-1}(V)^k$, $Z_1 \cdot \Phi^{-1}(V)^k$, and $X \cdot Z_1 \cdot \Phi^{-1}(V)^k$. We consider augmenting our baseline specifications by adding quadratic and both quadratic and cubic transformations of the control function and associated interactions terms. The control function and its powers are interacted with total expenditure (X), the children variable (Z_1), and their interaction ($X \cdot Z_1$).

In Figure D.6, we display the corresponding QSF for $k = 2, 3$, for both methods. The main difference with the baseline specifications is the increased curvature in the DR-based 0.75-QSFs in the right panel of Figures D.6(A)–(B). Augmenting the model pro-



(A) Food.



(B) Leisure.

FIGURE D.4. QSF over $\widetilde{\mathcal{T}}_5\widetilde{\mathcal{X}}_5^*$. Quantile (left) and distribution regression (right).

vides further evidence that our QSF estimates are robust are robust to the inclusion of higher-order terms in the control function.

To investigate further the selection of higher-order control function terms, we implement a leave-one-out cross-validation (CV) procedure for the ASF. For conditional mean specification of the QR baseline model, powers of the control function $\Phi^{-1}(V)^k$ as well as interaction terms $X \cdot \Phi^{-1}(V)^k$, $Z_1 \cdot \Phi^{-1}(V)^k$, and $X \cdot Z_1 \cdot \Phi^{-1}(V)^k$, $k = 2, \ldots, 5$, are added in increasing order to the specification of $E[Y \mid X, Z_1, V]$. Figure D.7 displays the CV criterion values for each specification. The results confirm that adding additional powers of the control function does not improve the model fit markedly for the quantile regression specification of the ASF, for both food and leisure.

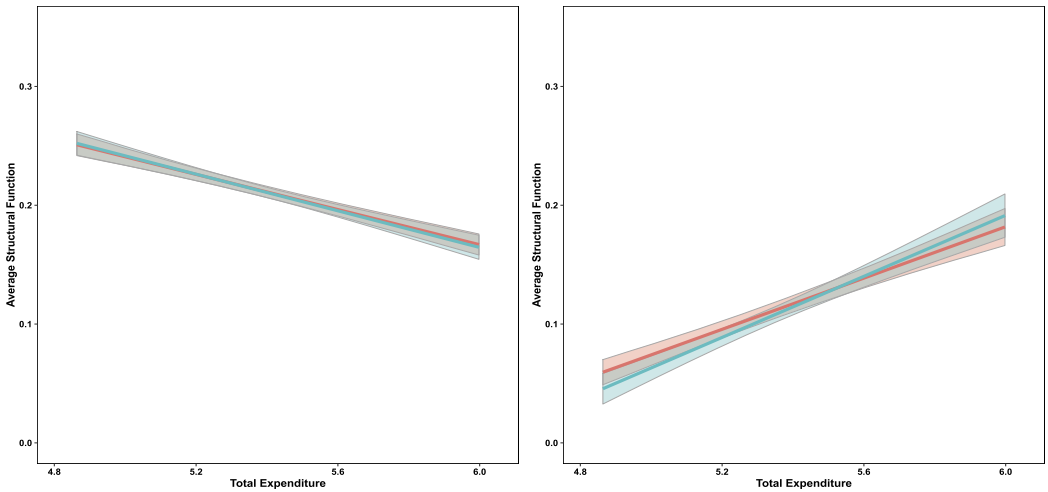


FIGURE D.5. Comparison of ASF estimates. Food (left) and leisure (right); QR (red) and OLS (blue).

The extensions considered in this section further illustrate the complementarity of our estimation methods and their relevance for empirical work.

D.3 *Robustness to integration method*

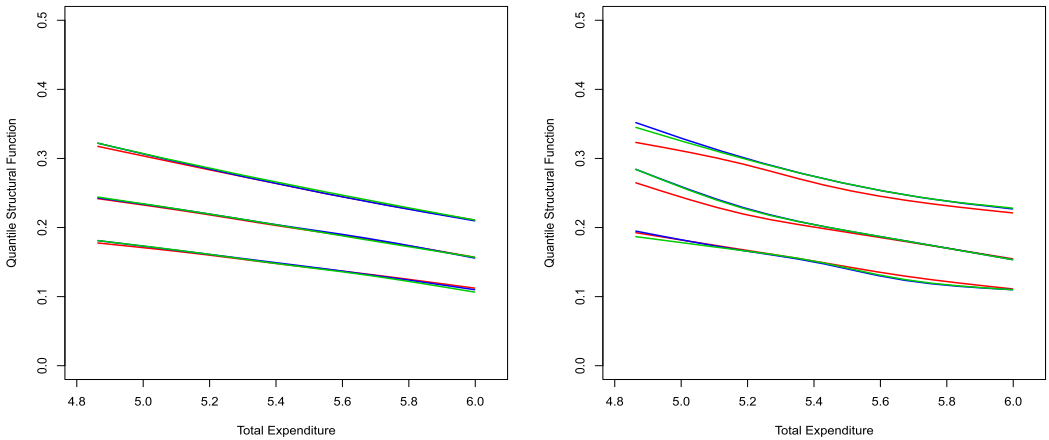
In this section, we compare the QSF obtained by replacing integration over the control function by sample averaging in the construction of the DSF by integration based on simulation of the control function. This can be done since the control function has a known distribution. Let $\widetilde{\mathcal{Y}}_S$, $\widetilde{\mathcal{X}}_K$ and $\widetilde{\mathcal{T}}_N$ be defined as in Algorithm 1 in the main text. Algorithm D.1 describes estimation of structural functions with control function integration by simulation.

Figure D.8 compares QSF estimates for QR and DR methods, obtained by implementing sample averaging (red) and integration by simulation (blue), for both food and leisure. The obtained ASFs are very similar, with slight differences for the DR based 0.75-QSF for food and the QR based 0.5, 0.75-QSF for leisure which are slightly steeper. Overall, the shape of the estimated QSF is essentially the same for both integration methods.

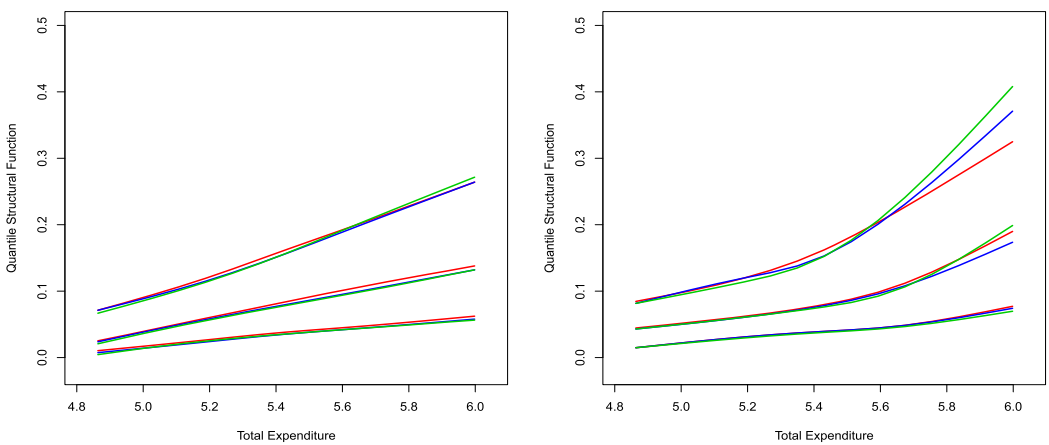
APPENDIX E: NUMERICAL SIMULATIONS

E.1 *Simulation results*

To assess the performance of our estimators, we implement Monte Carlo experiments based on three different designs, calibrated to the leisure empirical application. The first two experiments are based on Gaussian location-scale and DR triangular models, designed to reflect the respective strengths of the QR and DR estimators. The third experiment is a location triangular model, for which both estimators are consistent for the corresponding structural functions.



(A) Food.



(B) Leisure.

FIGURE D.6. QSF including additional powers of the control function. Baseline (red): $\{p(X) \otimes r_1(Z_1)\} \cdot \Phi^{-1}(V)$; Quadratic specification (blue): baseline spec. $+ \{p(X) \otimes r_1(Z_1)\} \cdot \Phi^{-1}(V)^2$; cubic specification (green): quadratic spec. $+ \{p(X) \otimes r_1(Z_1)\} \cdot \Phi^{-1}(V)^3$. Quantile (left) and distribution (right) regression.

DESIGN QR. Our first design is the linear location-scale shift system of equations

$$X = \pi_{11} + \pi_{21} Z + (\pi_{12} + \pi_{22} Z)\eta,$$
$$Y = \theta_{11} + \theta_{21} X + (\theta_{12} + \theta_{22} X)\varepsilon.$$

The ASF and QSF of this model are linear,

$$\mu(x) = \theta_{11} + \theta_{21} x, \qquad Q(\tau, x) = \theta_{11} + \theta_{21} x + (\theta_{12} + \theta_{22} x)\Phi^{-1}(\tau).$$

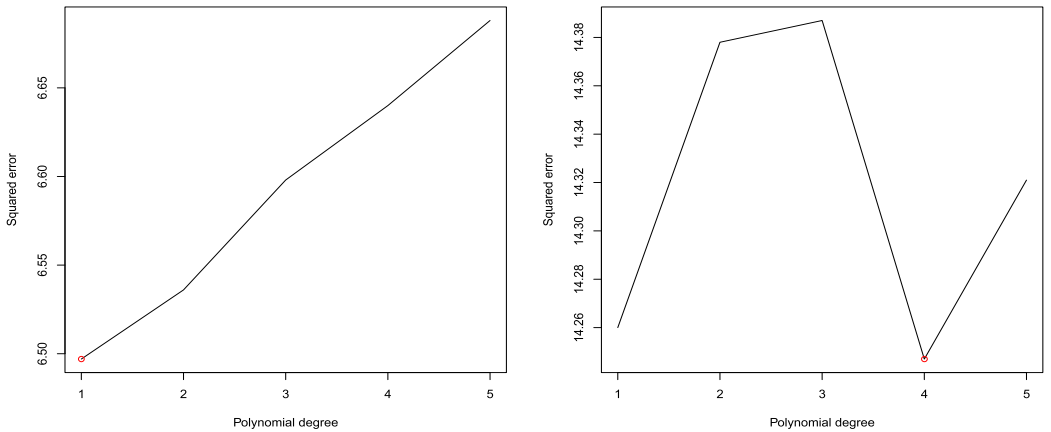


FIGURE D.7. Least-squares CV criterion $\times 1000$. Food (left) and leisure (right).

Algorithm D.1 Estimation of structural functions—Integration by simulation.

First and second stages. Repeat the first and second stages in Algorithm 1.

Third stage. [Structural functions via integration by simulation]

(1) Draw n realizations $\{\check{V}_i\}_{i=1}^n$ from a Uniform $[0,1]$.

(2) For the DSF, set, for $(y,x) \in \widetilde{\mathcal{Y}}_S \widetilde{\mathcal{X}}_K$, $\widehat{G}(y,x) = \sum_{i=1}^n \widehat{F}_Y(y \mid x, \check{V}_i) T_i/n$. For the ASF and QSF, set, for $(\tau,x) \in \widetilde{\mathcal{T}}_N \widetilde{\mathcal{X}}_K$,

$$\widehat{Q}_S(\tau,x) = \delta \sum_{s=1}^{S} \big[1(y_s \geq 0) - 1\big\{\widehat{G}(y_s,x) \geq \tau\big\}\big], \qquad \widehat{\mu}_S(x) = \delta \sum_{s=1}^{S} \big[1(y_s \geq 0) - \widehat{G}(y_s,x)\big].$$

DESIGN DR. Our second design is the nonlinear location-scale shift system of equations

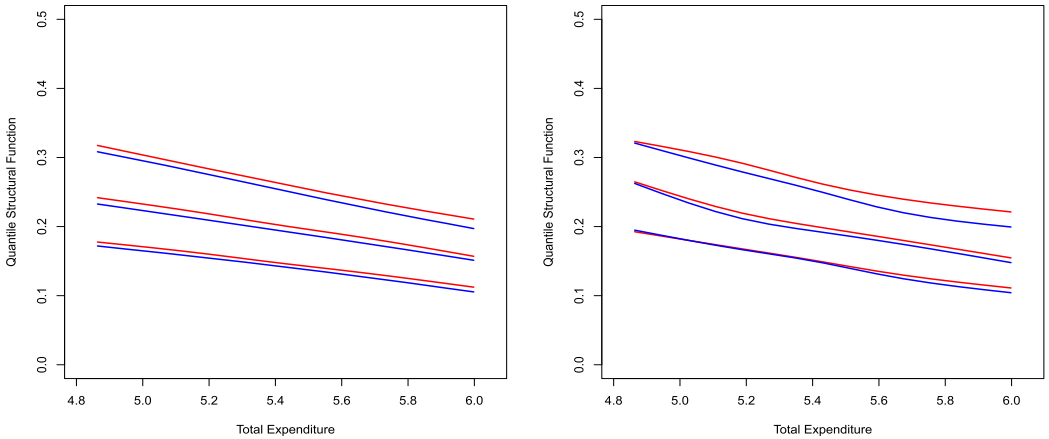
$$X = -\left(\frac{\pi_{11} + \pi_{12} Z}{\pi_{21} + \pi_{22} Z}\right) + \left(\frac{1}{\pi_{21} + \pi_{22} Z}\right)\eta,$$
$$Y = -\left(\frac{\theta_{11} + \theta_{12} x}{\theta_{21} + \theta_{22} x}\right) + \left(\frac{1}{\theta_{21} + \theta_{22} x}\right)\varepsilon.$$

The ASF and QSF of this model are nonlinear,

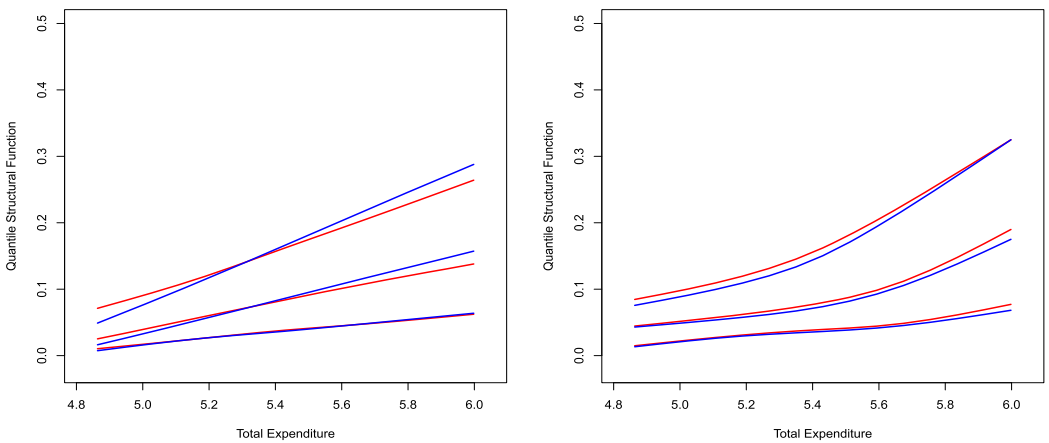
$$\mu(x) = -\left(\frac{\theta_{11} + \theta_{12} x}{\theta_{21} + \theta_{22} x}\right), \qquad Q(\tau,x) = -\left(\frac{\theta_{11} + \theta_{12} x}{\theta_{21} + \theta_{22} x}\right) + \left(\frac{1}{\theta_{21} + \theta_{22} x}\right)\Phi^{-1}(\tau).$$

DESIGN LOC. Our third design is the linear location shift system of equations

$$X = \pi_{11} + \pi_{21} Z + \sigma_\eta \eta,$$
$$Y = \theta_{11} + \theta_{21} X + \sigma_\varepsilon \varepsilon,$$



(A) Food.



(B) Leisure.

FIGURE D.8. QSF by sample averaging (red) and integration by simulation (blue). Quantile (left) and distribution regression (right).

for which the QR and DR models are correctly specified. The ASF and QSF of this model are

$$\mu(x) = \theta_{11} + \theta_{21}x, \qquad Q(\tau, x) = \theta_{11} + \theta_{21}x + \sigma_\varepsilon \Phi^{-1}(\tau).$$

For all three experiments, the sample size is set to $n = 1655$, the number of observations in the empirical application, and 500 simulations are performed. For the regions of interest, we use the same $\mathcal{T}_3$ and $\overline{\mathcal{X}}_5$ as in the empirical application. We let (η, ε) be jointly normal scalar random variables with zero means, unit variances and correlation ρ, and assess the performance of our estimators under two different levels of endogeneity by setting $\rho = -0.2$, for low endogeneity, and $\rho = -0.9$, for extreme endogeneity. Ac-

TABLE 1. Average L^p estimation errors of ASF $\times 1000$ for the DR and QR estimators and their ratio $\times 100$, for $p = 1, 2$ and ∞. Average L^p estimation errors of ASF $\times 1000$ for OLS are included as a benchmark.

	QR			LOC			DR		
Design	L_1	L_2	L_∞	L_1	L_2	L_∞	L_1	L_2	L_∞
					$\rho = -0.2$				
DR	6.9	8.1	15.2	5.2	6.7	8.3	5.9	7.4	10.2
QR	2.7	3.4	3.5	4.7	6.0	6.9	8.2	9.7	15.4
Ratio $\times 100$	251.1	237.2	426.8	110.3	111.9	121.2	72.4	76.6	66.2
OLS	10.6	10.9	22.4	14.9	15.4	26.0	15.4	16.0	33.0
					$\rho = -0.9$				
DR	4.7	6.0	7.7	6.4	7.9	10.5	7.8	9.5	13.6
QR	3.8	4.5	9.4	4.9	6.0	7.3	9.2	10.5	24.4
Ratio $\times 100$	123.6	132.7	82.0	131.1	131.9	144.7	84.4	90.4	56.0
OLS	47.2	47.3	100.2	66.2	66.3	117.9	73.2	73.3	152.8

cordingly, the DR estimator is implemented with the probit link function. We report simulation results for the ASF and the QSF.

Table 1 reports a first set of results regarding the accuracy of ASF estimates by DR and QR. For comparison purposes, Table 1 also includes ASF estimates by ordinary least-squares (OLS), providing a benchmark with no correction for endogeneity. We report average estimation errors across simulations of QR and DR estimators, and their ratio in percentage terms. Estimation errors are measured in L^p norms $\|\cdot\|_p$, $p = 1, 2$, and ∞, where for a function $f : \mathcal{X} \mapsto \mathbb{R}$, $\|f\|_p = \{\int_{\mathbb{R}} |f(s)|^p \, ds\}^{1/p}$, and are then averaged over the 500 simulations.

For this design, DR and QR-based estimators both perform very well and significantly improve over the OLS benchmark, including for $\rho = -0.2$. As expected, the accuracy of the estimates obtained by each method dominates for the corresponding design. For the QR design, the ratio of average estimation errors ranges from 82 to 426.8. Interestingly, the relative accuracy of DR-based estimates for $\rho = -0.9$ is close to the accuracy of QR estimates, with the ratio of average estimation errors ranging from 82 to 132.7, across norms; this feature is specific to the ASF and does not apply to the QSF. For the DR design, the ratio of average estimation errors ranges from 56 to 90.4. The larger reduction in average errors in L^∞ norm reflects the higher accuracy in estimation of extreme parts of the support where the ASF displays some curvature. Finally, for the LOC design, the performance of both methods is very similar for $\rho = -0.2$, and the QR-based estimator dominates more markedly for $\rho = -0.9$.

Table 2 reports the results regarding the accuracy of DR and QR estimates of the QSF, for quantile levels 0.25, 0.5, and 0.75, respectively. Compared to the results for the ASF, the main feature of the results is the stronger relative performance of the QR-based estimator for all three designs, although the DR-based estimator still dominates for the DR design.

Overall, the simulations show that both DR- and QR-based estimation methods perform well for their respective designs, and yield substantial correction for endogeneity.

TABLE 2. Average L^p estimation errors of $\{0.25, 0.5, 0.75\}$-QSF $\times 1000$ for the DR and QR estimators and their ratio $\times 100$, for $p = 1, 2$, and ∞.

	Design	QR			LOC			DR		
		L_1	L_2	L_∞	L_1	L_2	L_∞	L_1	L_2	L_∞
					$\tau = 0.25$					
$\rho = -0.2$	DR	10.4	11.5	28.2	7.0	8.9	12.6	7.9	9.9	12.5
	QR	3.4	4.3	5.6	6.0	7.5	8.6	9.1	11.2	15.9
	Ratio $\times 100$	302.0	266.1	506.6	117.4	118.3	145.6	86.3	88.7	78.6
$\rho = -0.9$	DR	5.7	6.8	11.5	7.4	9.4	15.3	7.7	9.6	14.0
	QR	2.6	3.3	3.2	5.3	6.7	9.0	8.8	10.6	18.6
	Ratio $\times 100$	218.4	207.5	355.9	139.7	140.1	170.1	87.2	91.0	75.2
					$\tau = 0.50$					
$\rho = -0.2$	DR	7.9	9.3	20.0	6.5	8.3	9.9	7.2	9.2	13.2
	QR	2.7	3.5	4.0	5.6	7.1	7.8	9.4	11.1	16.5
	Ratio $\times 100$	289.2	267.8	497.5	116.4	117.1	127.6	77.0	83.0	80.1
$\rho = -0.9$	DR	6.1	7.3	17.0	5.4	6.8	7.9	6.2	7.8	10.8
	QR	2.4	3.0	3.5	4.6	5.7	6.5	8.3	9.6	18.5
	Ratio $\times 100$	256.4	246.1	491.6	118.5	118.9	120.9	74.9	81.1	58.7
					$\tau = 0.75$					
$\rho = -0.2$	DR	10.0	11.6	16.5	7.3	9.3	14.3	7.6	9.5	15.7
	QR	3.4	4.3	5.0	6.1	7.6	8.6	10.3	11.9	21.1
	Ratio $\times 100$	297.0	271.9	330.7	119.6	122.1	165.8	73.6	79.5	74.6
$\rho = -0.9$	DR	8.1	10.0	20.9	7.0	8.9	15.4	9.2	11.5	24.8
	QR	2.9	3.6	4.6	5.3	6.6	9.1	10.3	11.7	29.3
	Ratio $\times 100$	279.3	276.7	457.0	131.6	135.4	168.7	89.5	98.7	84.6

QR-based estimation dominates for both the QR and LOC designs, but the DR estimator is able to correct for endogeneity in data generating processes displaying nonlinearities in the structural functions. These simulation results illustrate further the complementarity of the two estimation methods introduced in this paper.

E.2 *Calibration*

In this section, we give a detailed description of how the three data generating processes used for the Monte Carlo were calibrated to our empirical application.

E.2.1 *Linear Gaussian location-scale model (QR specification)* Consider the heteroscedastic normal system of equations

$$X = \pi_{11} + \pi_{21} Z + (\pi_{12} + \pi_{22} Z)\eta,$$
$$Y = \theta_{11} + \theta_{21} X + (\theta_{12} + \theta_{22} X)\varepsilon,$$

where (ξ, ε) are jointly normal with zero means, unit variances and correlation ρ. The reduced form of this system is

$$Q_X(v \mid z) = \pi_{11} + \pi_{21} z + (\pi_{12} + \pi_{22} z)\Phi^{-1}(v),$$

$$Q_Y(u \mid x, v) = \theta_{11} + \theta_{21}x + (\theta_{12} + \theta_{22}x)\big(\rho\Phi^{-1}(v) + \big(1-\rho^2\big)^{1/2}\Phi^{-1}(u)\big).$$

This system thus admits the QR representation

$$\begin{aligned} Q_X(v \mid z) &= \pi_1(v) + \pi_2(v)z, \\ Q_Y(u \mid x, v) &= \theta_1(u) + \gamma_1(u)\Phi^{-1}(v) + \theta_2(u)x + \gamma_2(u)\Phi^{-1}(v)x, \end{aligned}$$

with

$$\begin{aligned} \pi_1(v) &= \pi_{11} + \pi_{12}\Phi^{-1}(v), \\ \pi_2(v) &= \pi_{21} + \pi_{22}\Phi^{-1}(v), \\ \theta_1(u) &= \theta_{11} + \theta_{12}\big(1-\rho^2\big)^{1/2}\Phi^{-1}(u), \\ \theta_2(u) &= \theta_{21} + \theta_{22}\big(1-\rho^2\big)^{1/2}\Phi^{-1}(u), \\ \gamma_1(u) &= \theta_{12}\rho, \\ \gamma_2(u) &= \theta_{22}\rho. \end{aligned}$$

Define the fine meshes of M values $0.01 = v_1 < \cdots < v_M = 0.99$ and $0.01 = u_1 < \cdots < u_M = 0.99$, with $M = 599$, as in the empirical application. The vectors of parameter values are calibrated following the method suggested in Koenker and Xiao (2002).

For the first stage parameters, we estimate the QR coefficients $(\widehat{\pi}_1(v), \widehat{\pi}_2(v))$ for quantile indexes $v \in \{v_1, \ldots, v_M\}$. The value of the coefficients $\pi_1 = (\pi_{11}, \pi_{12})'$ and $\pi_2 = (\pi_{21}, \pi_{22})'$ are then set to the estimates obtained from linear regression of $(\widehat{\pi}_1(v_m), \widehat{\pi}_2(v_m))$ on $(1, \Phi^{-1}(v_m))$.

For the second stage, we estimate the QR coefficients $(\widehat{\theta}_1(u), \widehat{\theta}_2(u))$ for quantile indexes $u \in \{u_1, \ldots, u_M\}$. The value of the coefficients $\theta_1 = (\theta_{11}, \theta_{12})'$ and $\theta_2 = (\theta_{21}, \theta_{22})'$ are then set to the estimates obtained from linear regression of $(\widehat{\theta}_1(u_m), \widehat{\theta}_2(u_m))$ on $(1, (1-\widehat{\rho}^2)^{1/2}\Phi^{-1}(u_m))$. The correlation coefficient $\widehat{\rho}$ is calibrated to the correlation between $\widehat{\varepsilon}_i = \Phi^{-1}(\widehat{G}(Y_i, X_i))$ and $\widehat{\eta} = \Phi^{-1}(\widehat{F}_X(X_i \mid Z_i))$, where $\widehat{G}(Y_i, X_i)$ and $\widehat{F}_X(X_i \mid Z_i)$ are the DSF and F_X QR-based estimates evaluated at the n sample points. We find $\hat{\rho} \approx -0.1$.

E.2.2 *Nonlinear Gaussian location-scale model (DR specification)* Consider the linear DR system of equations:

$$\begin{aligned} \eta &= (\pi_{11} + \pi_{21}X) + (\pi_{12} + \pi_{22}X)Z, \\ \varepsilon &= (\theta_{11} + \theta_{21}Y) + (\theta_{12} + \theta_{22}Y)X, \end{aligned}$$

where (η, ε) are jointly normal with zero means, unit variances and correlation ρ. The reduced form of this system is

$$F_X(x \mid z) = \Phi\big((\pi_{11} + \pi_{21}x) + (\pi_{12} + \pi_{22}x)z\big),$$

$$F_Y(y \mid x, v) = \Phi\Bigg(\frac{1}{\left(1-\rho^2\right)^{1/2}}(\theta_{11} + \theta_{21} y) + \frac{1}{\left(1-\rho^2\right)^{1/2}}(\theta_{12} + \theta_{22} y)x - \frac{\rho}{\left(1-\rho^2\right)^{1/2}}\Phi^{-1}(v) \Bigg).$$

This system thus admits the Gaussian DR representation

$$F_X(x \mid z) = \Phi\big(\pi_1(x) + \pi_2(x) z\big),$$
$$F_Y(y \mid x, v) = \Phi\big(\theta_1(y) + \gamma_1(y)\Phi^{-1}(v) + \theta_2(y)x\big)$$

with

$$\pi_1(x) = \pi_{11} + \pi_{12} x,$$
$$\pi_2(x) = \pi_{21} + \pi_{22} x,$$
$$\theta_1(y) = \frac{1}{\left(1-\rho^2\right)^{1/2}}[\theta_{11} + \theta_{12} y],$$
$$\theta_2(y) = \frac{1}{\left(1-\rho^2\right)^{1/2}}[\theta_{21} + \theta_{22} y],$$
$$\gamma_1(y) = -\frac{\rho}{\left(1-\rho^2\right)^{1/2}}.$$

Define the fine mesh of M values $0.01 = t_1 < \cdots < t_M = 0.99$, with $M = 599$, as in the empirical application. The vectors of parameter values are calibrated by implementing the DR analog to the method suggested in Koenker and Xiao (2002).

For the first stage parameters, we estimate the DR coefficients $(\widehat{\pi}_1(x), \widehat{\pi}_2(x))$ for a fine mesh of X values $x \in \{\widehat{Q}_X(t_1), \ldots, \widehat{Q}_X(t_M)\}$, setting the link function to the gaussian CDF. The value of the coefficients $\pi_1 = (\pi_{11}, \pi_{12})'$ and $\pi_2 = (\pi_{21}, \pi_{22})'$ are then set to the estimates obtained from linear regression of $(\widehat{\pi}_1(\widehat{Q}_X(t_m)), \widehat{\pi}_2(\widehat{Q}_X(t_m)))$ on $(1, Q_X(t_m))$.

For the second stage, we estimate the DR coefficients $(\widehat{\theta}_1(y), \widehat{\theta}_2(y))$ for the fine mesh of Y values $y \in \{\widehat{Q}_Y(t_1), \ldots, \widehat{Q}_Y(t_M)\}$, setting the link function to the Gaussian CDF. The value of the coefficients $\theta_1 = (\theta_{11}, \theta_{12})'$ and $\theta_2 = (\theta_{21}, \theta_{22})'$ are then set to the estimates obtained from linear regression of $(\widehat{\theta}_1(\widehat{Q}_Y(t_m)), \widehat{\theta}_2(\widehat{Q}_Y(t_m)))$ on $((1-\widehat{\rho}^2)^{-1/2}, (1-\widehat{\rho}^2)^{-1/2}\widehat{Q}_Y(t_m))$. The correlation coefficient $\widehat{\rho}$ is calibrated to the correlation between $\widehat{\varepsilon}_i = \Phi^{-1}(\widehat{G}(Y_i, X_i))$ and $\widehat{\eta} = \Phi^{-1}(\widehat{F}_X(X_i \mid Z_i))$, where $\widehat{G}(Y_i, X_i)$ and $\widehat{F}_X(X_i \mid Z_i)$ are the DSF and F_X DR-based estimates evaluated at the n sample points. We find $\hat{\rho} \approx -0.1$.

In practice, the corresponding ASF and QSFs calibrated to our data are very close to being linear over the range of values of X considered, as can be seen from Figure E.1 which displays the true $\{0.25, 0.5, 0.75\}$-QSFs generated by our calibration (black lines). This yields simulation results favorable to our QR-based estimators. In order to assess our methods for a data generating process which is less favorable to the QR specification, the parameter values for θ_{12} and θ_{22} used in the simulations are set to the parameter values from our initial calibration multiplied by 1.5 and 2.25, respectively. This modi-

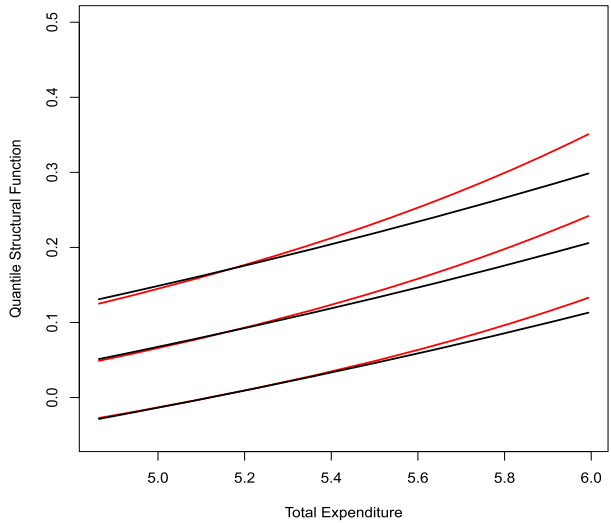


FIGURE E.1. True QSFs for DR Calibration. Initial calibration (black) and modified calibration used in simulations (red).

fication generates some curvature in the QSF as shown in Figure E.1 which also displays the true $\{0.25, 0.5, 0.75\}$-QSFs generated by our adjusted calibration (red lines).

E.2.3 *Linear Gaussian location model (LOC specification)* The QR and DR specifications coincide for the location shift model

$$X = \pi_{11} + \pi_{21} Z + \sigma_\eta \eta,$$
$$Y = \theta_{11} + \theta_{21} X + \sigma_\varepsilon \varepsilon.$$

For this model, both the conditional quantile and distribution functions are linear in the covariate, and so are the structural functions. After substitution, the reduced-form equation for the second stage is

$$\begin{aligned} Y &= \theta_{11} + \theta_{21} X + \left[\rho\eta + \left(1-\rho^2\right)^{1/2}\xi\right] \\ &= \theta_{11} + \theta_{21} X + \sigma_\varepsilon \rho \eta + \sigma_\varepsilon \left(1-\rho^2\right)^{1/2}\xi. \end{aligned}$$

The value of the coefficients $(\pi_{11}, \pi_{21})'$ are then set to the estimates obtained from linear regression of X_i on $(1, Z_i)$, and the scale parameter σ_η is set to the corresponding estimate $\widehat{\sigma}_\eta = [(n-1)^{-1}\sum_{i=1}^n (X_i - \widehat{\pi}_{11} - \widehat{\pi}_{21} Z_i)^2]^{1/2}$. Letting $\widehat{\eta}_i = (X_i - \widehat{\pi}_{11} - \widehat{\pi}_{21} Z_i)/\widehat{\sigma}_\eta$, the value of the coefficients $(\theta_{11}, \theta_{21})'$ are then set to the estimates of the first two coefficients in the linear regression of Y_i on $(1, X_i, \widehat{\eta}_i)$. We then set the scale parameter σ_ε to the corresponding estimate $\widehat{\sigma}_\varepsilon = [(n-1)^{-1}\sum_{i=1}^n (Y_i - \widehat{\theta}_{11} - \widehat{\theta}_{21} X_i)^2]^{1/2}$. The correlation coefficient $\widehat{\rho}$ is calibrated to the correlation between $\widehat{\varepsilon}_i = (Y_i - \widehat{\theta}_{11} - \widehat{\theta}_{21} X_i)/\widehat{\sigma}_\varepsilon$ and $\widehat{\eta}$. We find $\hat{\rho} \approx -0.2$.

References

Chernozhukov, V., I. Fernandez-Val, and A. Galichon (2010), "Quantile and probability curves without crossing." *Econometrica*, 78 (3), 1093–1125. [12]

Chernozhukov, V., I. Fernandez-Val, and B. Melly (2013), "Inference on counterfactual distributions." *Econometrica*, 81 (6), 2205–2268. [8]

Koenker, R. and Z. Xiao (2002), "Inference on the quantile regression process." *Econometrica*, 70 (4), 1583–1612. [24, 25]

van der Vaart, A. W. (2000), *Asymptotic Statistics*. Cambridge University Press, Cambridge. [2, 8, 11]

van der Vaart, A. W. and J. A. Wellner (1996), *Weak Convergence and Empirical Processes*. Springer. [2, 5, 10]

Co-editor Christopher Taber handled this manuscript.

Manuscript received 5 November, 2018; final version accepted 24 April, 2019; available online 1 July, 2019.