Supplement to "Decentralization estimators for instrumental variable quantile regression models"

(*Quantitative Economics*, Vol. 12, No. 2, May 2021, 443–475)

Hiroaki Kaido
Department of Economics, Boston University

Kaspar Wüthrich
Department of Economics, University of California San Diego

Contents

Appendix A: Overidentification and efficiency

In the main text, we focus on just-identified moment restrictions with $d_Z = d_D$, for which the construction of an estimator is straightforward. If the model is overidentified (i.e., if $d_Z > d_D$), instead of the original moment conditions,

$$E_P\left[\left(1\left\{Y \le X'\theta_1(\tau) + D_1\theta_2(\tau) + \cdots + D_{d_D}\theta_J(\tau)\right\} - \tau\right)\begin{pmatrix} X \\ Z \end{pmatrix}\right] = 0,$$

we may use a set of just-identified moment conditions

$$E_P\left[\left(1\left\{Y \le X'\theta_1(\tau) + D_1\theta_2(\tau) + \cdots + D_{d_D}\theta_J(\tau)\right\} - \tau\right)\begin{pmatrix} X \\ \tilde{Z} \end{pmatrix}\right] = 0, \tag{A.1}$$

where $\tilde{Z}$ is a $d_D \times 1$ vector of transformations of (X, Z). A practical choice is to construct $\tilde{Z}$ using a least squares projection of D on Z and X (Chernozhukov and Hansen (2006)).

To achieve pointwise (in τ) efficiency, we can employ the following two-step procedure which is based on Remark 5 in Chernozhukov and Hansen (2006):

Hiroaki Kaido: hkaido@bu.edu
Kaspar Wüthrich: kwuthrich@ucsd.edu

Available at http://qeconomics.org. https://doi.org/10.3982/QE1440

Step 1: We first obtain an initial consistent estimate of θ^* using one of our estimators based on a set of just-identified moment conditions such as (A.1). We then use nonparametric estimators to estimate the conditional densities $V(\tau) = f_{\varepsilon(\tau)|X,Z}(0)$ and $v(\tau) = f_{\varepsilon(\tau)|D,X,Z}(0)$, where $\varepsilon(\tau) = Y - X'\theta_1^*(\tau) - D_1\theta_2^*(\tau) - \cdots - D_{d_D}\theta_J^*(\tau)$, and the conditional expectation function $E_P[Dv(\tau) \mid X, Z]$.

Step 2: We apply our procedure to obtain a solution to the following moment conditions:

$$E_P\left[\left(1\{Y \le X'\theta_1(\tau) + D_1\theta_2(\tau) + \cdots + D_{d_D}\theta_J(\tau)\} - \tau\right)\begin{pmatrix} V(\tau)X \\ E_P[Dv(\tau) \mid X, Z]\end{pmatrix}\right] = 0. \quad \text{(A.2)}$$

Consider players $j = 1, \ldots, J$ solving the following optimization problems:

$$\min_{\tilde{\theta}_1 \in \mathbb{R}^{d_X}} Q_{P,1}(\tilde{\theta}_1, \theta_{-1}), \quad \text{(A.3)}$$

$$\min_{\tilde{\theta}_j \in \mathbb{R}} Q_{P,j}(\tilde{\theta}_j, \theta_{-j}), \quad j = 2, \ldots, J, \quad \text{(A.4)}$$

where

$$Q_{P,1}\big(\theta(\tau)\big) := E_P\big[\rho_\tau\big(Y - X'\theta_1(\tau) - D_1\theta_2(\tau) - \cdots - D_{d_D}\theta_J(\tau)\big)V(\tau)\big],$$

$$Q_{P,j}\big(\theta(\tau)\big) := E_P\left[\rho_\tau\big(Y - X'\theta_1(\tau) - D_1\theta_2(\tau) - \cdots - D_{d_D}\theta_J(\tau)\big)\frac{E_P[Dv(\tau) \mid X, Z]_{j-1}}{D_{j-1}}\right],$$
$$j = 2, \ldots, J,$$

and $E_P[Dv(\tau) \mid X, Z]_{j-1}$ is the jth element of $E_P[Dv(\tau) \mid X, Z]$. For each j, the BR function $L_j(\theta_{-j}(\tau))$, defined as a member of the set of minimizers of $Q_{P,j}(\cdot, \theta_{-j})$, solves a suitable subset of the moment conditions in (A.2). The optimization problems in (A.3)–(A.4) are convex population QR problems provided that the model is parametrized such that $E_P[Dv(\tau) \mid X, Z]_{j-1}/D_{j-1}$, $j = 2, \ldots, J$, is positive. Estimation can then proceed by replacing the population QR problems by their sample analogues and applying one of the estimation algorithms discussed in the main text. By Corollary 2, the resulting estimator is asymptotically equivalent to a GMM estimator that uses the optimal instrumental variables and thus achieves pointwise (in τ) efficiency (e.g., Chamberlain (1987)).[26]

Appendix B: Reparametrization

In the main text, we assume that the model is reparametrized such that Z_ℓ/D_ℓ is positive for all $\ell = 1, \ldots, d_D$. This ensures that the weights are well-defined and that the weighted QR problems are convex. However, in empirical applications, the weights may not be well-defined (e.g., if D_ℓ is an indicator variable with $P(D_\ell = 0) > 0$) or negative in

[26] Corollary 2 can be applied to the current setting by replacing the original set of covariates and instruments $\Psi(\tau) = (X', Z')'$ in (6.6) with the optimal instrumental variables $\Psi(\tau) = (V(\tau)', E_P[D_1v(\tau) \mid X, Z]/D_1, \ldots, E_P[D_{d_D}v(\tau) \mid X, Z]/D_{d_D})'$.

some instances. Assuming that Z_ℓ is positive, a simple way to reparametrize the model is to add a large enough constant c to D_ℓ.[27] This transformation is theoretically justified by the compactness of the support of D_ℓ (Assumption 2(2)). To fix ideas, suppose that one is interested in estimating the following linear-in-parameters model with a single endogenous variable:

$$q(D, X, \tau) = \theta_{11} + \tilde{X}'\theta_{12} + D\theta_2,$$

where $\theta_1 = (\theta_{11}, \theta_{12}')'$ and $X = (1, \tilde{X}')'$. Suppose further that the support of D is a compact interval, $[d_{\min}, d_{\max}] \subset \mathbb{R}$, with $d_{\min} < 0$. In this case, we can apply the transformation $D^\star = D + c$, where $c > |d_{\min}|$. The transformed model reads

$$q(D, X, \tau) = \theta_{11}^\star + \tilde{X}'\theta_{12} + D^\star\theta_2,$$

where $\theta_{11}^\star = \theta_{11} - c\theta_2$. Importantly, one can always back out the original parameters, $\theta = (\theta_{11}, \theta_{12}', \theta_2)'$, from the parameters in the reparametrized model, $\theta^\star = (\theta_{11}^\star, \theta_{12}', \theta_2)'$.

Appendix C: Decentralization

C.1 *The domains of M_j-maps*

Recall that, in (3.6), we defined the set

$$\begin{aligned}\tilde{R}_1 := \big\{&\theta_{-1} \in \Theta_{-1} : \Psi_{P,1}(\theta_1, \theta_{-1}) = 0, \\ &\Psi_{P,2}(\theta_1, \theta_2, \pi_{-\{1,2\}}\theta_{-1}) = 0, \exists(\theta_1, \theta_2) \in \Theta_1 \times \Theta_2\big\}.\end{aligned}$$

Similarly, for $k = 2, \dots, d_D - 1$, define

$$\begin{aligned}\tilde{R}_k := \Bigg\{&\theta_{-1} \in \Theta_{-1} : \Psi_{P,1}(\theta_1, \theta_{-1}) = 0, \\ &\Psi_{P,2}(\theta_1, \theta_2, \pi_{-\{1,2\}}\theta_{-1}) = 0, \\ &\vdots \\ &\Psi_{P,k}(\theta_1, \dots, \theta_k, \pi_{-\{1,\dots,k\}}\theta_{-1}) = 0, \exists(\theta_1, \dots, \theta_k) \in \prod_{j=1}^{k}\Theta_j\Bigg\}.\end{aligned}$$

For $k = d_D$, let

$$\begin{aligned}\tilde{R}_{d_D} := \Bigg\{&\theta_{-1} \in \Theta_{-1} : \Psi_{P,1}(\theta_1, \theta_{-1}) = 0, \\ &\Psi_{P,2}(\theta_1, \theta_2, \pi_{-\{1,2\}}\theta_{-1}) = 0,\end{aligned}$$

[27] Since the IVQR model is characterized by conditional moments (as in (2.1)), one may choose transformations of instruments to generate unconditional moment conditions. In case Z_ℓ is not positive *a.s.*, one can use a positive transformation (e.g., a logistic function) of Z_ℓ instead of Z_ℓ itself. The decentralization and identification results then hold with the transformed instruments as long as they satisfy our assumptions.

$$\vdots$$

$$\Psi_{P,J}(\theta_1, \dots, \theta_J) = 0, \exists(\theta_1, \dots, \theta_J) \in \prod_{j=1}^{J} \Theta_j \Bigg\}.$$

For each k, the map M_k is well-defined on $\tilde{R}_k$. Note also that $\tilde{R}_{d_D} \subset \tilde{R}_j$ for all $j \leq d_D$.

C.2 *Local decentralization and local contractions*

We say that an estimation problem admits *local decentralization* if the BR functions L_j, $j = 1, \dots, J$, and the maps K and M are well-defined over a local neighborhood of θ^*. The following weak conditions are sufficient for local decentralization of the IVQR estimation problem.

ASSUMPTION 4. *The following conditions hold*:

(1) *The conditional cdf* $y \mapsto F_{Y|D,X,Z}(y)$ *is continuously differentiable at* $y^* = d'\theta^*_{-1} + x'\theta^*_1$ *for almost all* (d, x, z). *The conditional density* $f_{Y|D,Z,X}$ *is bounded on a neighborhood of* y^* *a.s.*;

(2) *The matrices*

$$E_P\big[f_{Y|D,X,Z}\big(D'\theta^*_{-1} + X'\theta^*_1\big)XX'\big]$$

and

$$E_P\big[f_{Y|D,X,Z}\big(D'\theta^*_{-1} + X'\theta^*_1\big)D_\ell Z_\ell\big], \quad \ell = 1, \dots, d_D,$$

are positive definite.

Assumption 4 is weaker than Assumption 2(3)–2(4) as it only requires the conditions, for example, continuous differentiability of the conditional CDF, at a particular point, for example, y^*. Under this condition, we can study the local properties of our population algorithms. For this, the following lemma ensures that the BR maps are well-defined locally.

LEMMA 3. *Suppose that Assumptions* 1, 2(1)–2(2), *and* 4 *hold. Then there exist open neighborhoods* $\mathcal{N}_{L_{-j}}$, $j = 1, \dots, J$, $\mathcal{N}_K$, $\mathcal{N}_M$ *of* θ^*_{-j}, θ^*, *and* θ^*_{-1} *such that*:

(i) *There exist maps* $L_j : \mathcal{N}_{-j} \to \mathbb{R}^{d_j}$, $j = 1, \dots, J$ *such that, for* $j = 1, \dots, J$,

$$\Psi_{P,j}\big(L_j(\theta_{-j}), \theta_{-j}\big) = 0, \quad \textit{for all } \theta_{-j} \in \mathcal{N}_{-j}.$$

Further, L_j *is continuously differentiable for all* $j = 1, \dots, J$.

(ii) *The maps* $K : \mathcal{N}_K \to \mathbb{R}^{d_X + d_D}$ *and* $M : \mathcal{N}_M \to \mathbb{R}^{d_D}$ *are continuously differentiable.*

Proof. (i) The proof is similar to that of Lemma 1 (see Appendix E). Therefore, we sketch the argument below for $j=1$. By Assumptions 2(2) and 4(1), $\Psi_{P,1}$ is continuously differentiable on a neighborhood V of θ^*. By Assumption 4(2) and the continuity of $\det(\partial \Psi_{P,1}(\theta)/\partial \theta_1')$, one may choose V so that $\det(\partial \Psi_{P,1}(\theta)/\partial \theta_1') \neq 0$ for all $\theta = (\theta_1, \theta_{-1}) \in V$. By the implicit function theorem, there is a continuously differentiable function L_1 and an open set $\mathcal{N}_{-1}$ containing θ_{-1} such that

$$\Psi_{P,1}\big(L_1(\theta_{-1}), \theta_{-1}\big) = 0, \quad \text{for all } \theta_{-1} \in \mathcal{N}_{-1}.$$

The arguments for L_j, $j \neq 1$ are similar.

(ii) Let $\mathcal{N}_K = \{\theta \in \Theta : \pi_{-j}\theta \in \mathcal{N}_{-j}, j = 1, \ldots, J\}$ and let $\mathcal{N}_M$ be defined by mimicking (3.6), while replacing Θ_j with $\mathcal{N}_j$ in the definition of $\tilde{R}_j$ for $j = 1, \ldots, J$. The continuous differentiability of K and M follows from that of L_j, $j = 1, \ldots, J$. □

C.2.1 *Local contractions* Recall that $\rho(A)$ denotes the spectral radius of a square matrix A. The following assumption ensures that K and M are local contractions.

Assumption 5.

(1) $\rho(J_K(\theta^*)) < 1$;

(2) $\rho(J_M(\theta^*_{-1})) < 1$.

Here, we illustrate a primitive condition for Assumption 5. Consider a simple setup without covariates (i.e., $X = 1$), a binary D, and a binary Z. We only analyze Assumption 5(1). A similar result can be derived for Assumption 5(2). In this setting, the Jacobian of K evaluated at θ^* is given by

$$J_K\big(\theta^*\big) = \begin{pmatrix} 0 & -\dfrac{E_P\big[f_{Y|D,Z}\big(D\theta_2^* + \theta_1^*\big)D\big]}{E_P\big[f_{Y|D,Z}\big(D\theta_2^* + \theta_1^*\big)\big]} \\ -\dfrac{E_P\big[f_{Y|D,Z}\big(D\theta_2^* + \theta_1^*\big)Z\big]}{E_P\big[f_{Y|D,Z}\big(D\theta_2^* + \theta_1^*\big)ZD\big]} & 0 \end{pmatrix}.$$

The characteristic polynomial is then given by

$$p_K(\lambda) = \lambda^2 - \frac{E_P\big[f_{Y|D,Z}\big(D\theta_2^* + \theta_1^*\big)D\big]}{E_P\big[f_{Y|D,Z}\big(D\theta_2^* + \theta_1^*\big)\big]} \frac{E_P\big[f_{Y|D,Z}\big(D\theta_2^* + \theta_1^*\big)Z\big]}{E_P\big[f_{Y|D,Z}\big(D\theta_2^* + \theta_1^*\big)ZD\big]}.$$

Hence, Assumption 3(1) holds if all eigenvalues (i.e., the roots λ_K of $p_K(\lambda) = 0$) have modulus less than one, which holds when

$$\left| \frac{E_P\big[f_{Y|D,Z}\big(D\theta_2^* + \theta_1^*\big)D\big]}{E_P\big[f_{Y|D,Z}\big(D\theta_2^* + \theta_1^*\big)\big]} \frac{E_P\big[f_{Y|D,Z}\big(D\theta_2^* + \theta_1^*\big)Z\big]}{E_P\big[f_{Y|D,Z}\big(D\theta_2^* + \theta_1^*\big)ZD\big]} \right| < 1.$$

This condition can be simplified to

$$f_{Y|0,1}\big(\theta_1^*\big) p(0|1) f_{Y|1,0}\big(\theta_2^* + \theta_1^*\big) p(1|0) < f_{Y|1,1}\big(\theta_2^* + \theta_1^*\big) p(1|1) f_{Y|0,0}\big(\theta_1^*\big) p(0|0), \qquad \text{(C.1)}$$

where $f_{Y|d,z}(y) := f_{Y|D=d,Z=z}(y)$ and $p(d|z) := P(D = d \mid Z = z)$. It is instructive to interpret condition (C.1) under the local average treatment effects framework of Imbens and Angrist (1994). Specifically, condition (C.1) holds if (i) their monotonicity assumption is such that there are compliers but no defiers and (ii) the complier potential outcome density functions are strictly positive. Conversely, the condition is violated if there are defiers but no compliers.

Under the local contraction conditions in Assumption 5, we have the following results.

PROPOSITION 3. *Suppose that Assumptions* 1, 2(1), 2(2), *and* 4 *hold.*

(i) *Suppose further that Assumption* 5(1) *holds. Then there exists a closed neighborhood* $\bar{\mathcal{N}}_K$ *of* θ^* *such that* $K(\bar{\mathcal{N}}_K) \subset \bar{\mathcal{N}}_K$ *and* K *is a contraction on* $\bar{\mathcal{N}}_K$ *with respect to an adapted norm.*

(ii) *Suppose further that Assumption* 5(2) *holds. Then there exists a closed neighborhood* $\bar{\mathcal{N}}_M$ *of* θ_2^* *such that* $M(\bar{\mathcal{N}}_M) \subset \bar{\mathcal{N}}_M$ *and* M *is a contraction on* $\bar{\mathcal{N}}_M$ *with respect to an adapted norm.*

PROOF. We only prove the result for K, the proof for M is similar. By Lemma 3, L_j is continuously differentiable at θ^*. Note that J_K is given by

$$J_K(\theta) = \begin{bmatrix} 0 & \dfrac{\partial L_1(\theta_{-1})}{\partial \theta_2'} & \cdots & \cdots & \dfrac{\partial L_1(\theta_{-1})}{\partial \theta_J'} \\ \dfrac{\partial L_2(\theta_{-2})}{\partial \theta_1'} & 0 & \dfrac{\partial L_2(\theta_{-2})}{\partial \theta_3'} & \cdots & \dfrac{\partial L_2(\theta_{-2})}{\partial \theta_J'} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \dfrac{\partial L_J(\theta_{-J})}{\partial \theta_1'} & \cdots & \cdots & \dfrac{\partial L_J(\theta_{-J})}{\partial \theta_{J-1}'} & 0 \end{bmatrix},$$

which is continuous at θ^*. The desired result now follows, for instance, from Proposition 2.2.19 in Hasselblatt and Katok (2003). □

C.3 *Nested algorithms: Existence and uniqueness of fixed points in subgames*

Here, we discuss two different sets of conditions which ensure that the nested fixed-point algorithms in Section 4.3 are well-defined. Specifically, we present conditions for the existence and uniqueness of fixed points in the subgames. Section C.3.1 considers contraction-based identification conditions. In Section C.3.2, we briefly discuss global identification conditions. To illustrate, we consider the case with three players ($J = 3$).

C.3.1 *Contraction-based identification* Suppose that Assumptions 1–3 hold. Consider a subgame formed by players 1 and 2 given some $\tilde{\theta}_3$. We assume that $\tilde{\theta}_3$ is chosen so that $\mathsf{D}_{M_{1,2|3}}$ (defined below) is nonempty. Let the moment conditions for the subgame be defined as

$$\Psi_{P,1}(\theta_1, \theta_2) := \Psi_{P,1}(\theta_1, \theta_2, \tilde{\theta}_3) = E_P\big[\big(1\{Y \le X'\theta_1 + D_1\theta_2 + D_2\tilde{\theta}_3\} - \tau\big)X\big],$$
$$\Psi_{P,2}(\theta_1, \theta_2) := \Psi_{P,2}(\theta_1, \theta_2, \tilde{\theta}_3) = E_P\big[\big(1\{Y \le X'\theta_1 + D_1\theta_2 + D_2\tilde{\theta}_3\} - \tau\big)Z_1\big].$$

Note that $\Psi_{P,1}$ and $\Psi_{P,2}$ and other objects below depend on $\tilde{\theta}_3$. We will often suppress this dependence to alleviate the notation. Moreover, define

$$\mathsf{R}_1 := \{\theta_1 \in \Theta_1 : \Psi_{P,2}(\theta_1, \theta_2) = 0, \text{ for some } \theta_2 \in \Theta_2\},$$
$$\mathsf{R}_2 := \{\theta_2 \in \Theta_2 : \Psi_{P,1}(\theta_1, \theta_2) = 0, \text{ for some } \theta_1 \in \Theta_1\}.$$

Assumptions 1–2 and Lemma 1 guarantee that BR functions L_1 and L_2, where

$$\Psi_{P,1}\big(\mathsf{L}_1(\theta_2), \theta_2\big) = 0, \quad \text{for all } \theta_2 \in \mathsf{R}_2,$$
$$\Psi_{P,2}\big(\theta_1, \mathsf{L}_2(\theta_1)\big) = 0, \quad \text{for all } \theta_1 \in \mathsf{R}_1,$$

are well-defined. The M map for the subsystem is

$$M_{1,2|3}(\theta_2 \mid \tilde{\theta}_3) = \mathsf{L}_2\big(\mathsf{L}_1(\theta_2)\big).$$

This map exists and is well-defined on

$$\mathsf{D}_{M_{1,2|3}} := \{\theta_2 \in \Theta_2 : \Psi_{P,1}(\theta_1, \theta_2) = 0, \Psi_{P,2}(\theta_1, \tilde{\theta}_2) = 0, \text{ for some } (\theta_1, \tilde{\theta}_2) \in \Theta_1 \times \Theta_2, (\theta_2, \tilde{\theta}_3) \in \tilde{D}_M\}.$$

Note that, by Assumption 3, for any $(\theta_2, \theta_3) \in \tilde{D}_M$,

$$\big\|J_M(\theta_2, \theta_3)\big\| \le \lambda < 1.$$

Given these definitions, we can now investigate the subsystem. Observe that the derivative $J_{M_{1,2|3}}(\theta_2)$ of $M_{1,2|3}$ with respect to θ_2 is a component of J_M. In particular, for any (θ_2, θ_3), J_M may be written as

$$J_M(\theta_2, \theta_3) = \begin{bmatrix} \frac{\partial M_1(\theta_2, \theta_3)}{\partial \theta_2} & \frac{\partial M_1(\theta_2, \theta_3)}{\partial \theta_3} \\ \frac{\partial M_2(\theta_2, \theta_3)}{\partial \theta_2} & \frac{\partial M_2(\theta_2, \theta_3)}{\partial \theta_3} \end{bmatrix} = \begin{bmatrix} J_{M_{1,2|3}}(\theta_2) & \frac{\partial M_1(\theta_2, \theta_3)}{\partial \theta_3} \\ \frac{\partial M_2(\theta_2, \theta_3)}{\partial \theta_2} & \frac{\partial M_2(\theta_2, \theta_3)}{\partial \theta_3} \end{bmatrix}.$$

Let $V_{12} := \{x \in \mathbb{R}^2 : x = (x_1, 0)', x_1 \in \mathbb{R}\}$. Then, by the definition of the operator norm (see, e.g., Bhatia (1997)),

$$\begin{aligned} \big\|J_M(\theta_2, \theta_3)\big\| &:= \sup_{x,y \in \mathbb{R}^2 : \|x\| = \|y\| = 1} \big|x' J_M(\theta_2, \theta_3) y\big| \\ &\ge \sup_{x,y \in V_{12} : \|x\| = \|y\| = 1} \big|x' J_M(\theta_2, \theta_2) y\big| \\ &= \sup_{\|x_1\| = \|y_1\| = 1} \big|x_1' J_{M_{1,2|3}}(\theta_2) y_1\big| = \big\|J_{M_{1,2|3}}(\theta_2)\big\|. \end{aligned}$$

Hence, it follows that

$$\big\|J_{M_{1,2|3}}(\theta_2)\big\| \le \big\|J_M(\theta_2, \tilde{\theta}_3)\big\| \le \lambda < 1, \quad \text{for all } \theta_2 \in \mathsf{D}_{K_{1,2|3}},$$

which is an analog of Assumption 3 for the subgame. Then Proposition 2 ensures the existence and uniqueness of the fixed point in the subgame. In this section, we focused on identification conditions based on the dynamical system M. Similar arguments can be used to establish identification based on the dynamical system K.

C.3.2 *Global identification conditions* To ensure the existence and uniqueness of the fixed point in the subgame, we can alternatively rely on the existing global identification conditions in Chernozhukov and Hansen (2006) (cf. Section 4.2 and Lemma 2) and Proposition 1. For every value $\tilde{\theta}_3 \in \Theta_3$, this requires analogues of the conditions in Lemma 2 to hold for the subgame between players 1 and 2.

Appendix D: Additional simulation results

D.1 *Bias and RMSE application-based DGP*

Here, we report simulation evidence on the finite sample bias and RMSE of the different IVQR estimators based on the application-based DGPs in Section 8. Tables 5–6 present the results. We find that all the proposed algorithms perform well and exhibit comparable bias and RMSE properties. In particular, the finite sample performances of our preferred estimators are comparable to IQR and the profiling estimator, which shows that their computational advantages do not come at a cost in terms of the finite sample performance.

D.2 *Three endogenous variables*

Here, we present additional simulation evidence with three endogenous variables. We consider the application-based DGP of Section 8, augmented with an additional endogenous variable:

$$Y_i = X_i'\theta_X(U_i) + D_i\theta_D(U_i) + D_{2,i}\theta_{D,2}(U_i) + D_{3,i}\theta_{D,3}(U_i) + G^{-1}(U_i),$$

where $\theta_{D,3}(U_i) = 10{,}000$, $D_{3,i} = 0.8 \cdot Z_{3,i} + 0.2 \cdot \Phi^{-1}(U_i)$, and $Z_{3,i} \sim N(0,1)$. We only report the results based on the contraction algorithm and the nested fixed-point algo-

Table 5. Bias and RMSE, 401(k) DGP with one endogenous regressor.

	Bias/10^2				RMSE/10^3			
τ	Contr	Brent	Profil	InvQR	Contr	Brent	Profil	InvQR
0.15	−4.43	−6.84	−6.71	−7.26	7.58	7.71	7.64	7.87
0.25	−0.28	−1.86	−1.90	−1.67	4.04	4.10	4.11	4.10
0.50	−1.04	−1.46	−1.49	−1.23	2.06	2.07	2.07	2.08
0.75	−1.60	−1.31	−1.48	−1.14	1.85	1.85	1.86	1.85
0.85	0.61	1.17	0.85	1.24	2.06	2.07	2.07	2.08

Note: Monte Carlo simulation with 500 repetitions as described in the main text. Contr: contraction algorithm; Brent: root-finding algorithm based on Brent's method; Profil: profiling estimator based on Brent's method; InvQR: inverse quantile regression. We use 2SLS estimates as starting values.

TABLE 6. Bias and RMSE, 401(k) DGP with two endogenous regressors.

	Bias/10^2					RMSE/10^3				
τ	Contr	NestBr	SimAnn	Profil	InvQR	Contr	NestBr	SimAnn	Profil	InvQR
	Coefficient on binary endogenous variable									
0.15	−4.81	−2.73	3.77	−3.65	−7.77	8.29	7.84	6.74	7.95	8.60
0.25	−3.47	−3.75	−3.17	−3.83	−3.42	4.32	4.31	4.25	4.31	4.32
0.50	0.84	0.56	0.68	0.71	0.74	1.93	1.95	1.95	1.95	1.98
0.75	−0.56	−0.33	−0.37	−0.57	−0.25	1.75	1.74	1.74	1.74	1.78
0.85	−1.03	−0.72	−0.74	−1.28	−0.61	2.18	2.19	2.20	2.19	2.22
	Coefficient on continuous endogenous variable									
0.15	2.00	0.48	4.72	−0.03	0.23	1.07	1.07	2.20	1.07	1.19
0.25	1.72	−0.09	0.25	−0.48	−0.10	1.00	1.02	1.13	1.03	1.13
0.50	0.75	−0.47	−0.45	−0.54	−0.68	0.89	0.97	0.97	0.97	1.11
0.75	−1.43	−0.49	−0.44	−1.57	−0.15	0.99	1.10	1.11	1.12	1.24
0.85	−2.66	−0.89	−0.91	−2.60	−0.40	1.12	1.27	1.27	1.28	1.32

Note: Monte Carlo simulation with 500 repetitions as described in the main text. Contr: contraction algorithm; NestBr: nested algorithm based Brent's method; SimAnn: simulated annealing based optimization algorithm; Profil: nested profiling estimator based on Brent's method; IQR: inverse quantile regression. We use 2SLS estimates as starting values.

rithm. We do not report results for IQR, which we found to be computationally prohibitive with three endogenous regressors. Table 7 shows that both methods exhibit similar performances in terms of bias and RMSE, which are comparable to their respective performances with two endogenous regressors. Table 8 displays average computation times. As expected, the computational advantages of the contraction algorithm relative to the nested fixed-point algorithm are more pronounced than with two endogenous variables.

D.3 *Additional simulations simple location scale DGP*

This section presents some additional simulation evidence based on the following location-scale shift model:

$$Y_i = \gamma_1 + \gamma_2 X_i + \gamma_3 D_{1,i} + \gamma_4 D_{2,i} + (\gamma_5 + \gamma_6 D_{1,i} + \gamma_7 D_{2,i}) U_i.$$

Here, $D_{1,i}$ and $D_{2,i}$ are the endogenous variables of interest and X_i is an exogenous covariate. In addition, we have access to two instruments $Z_{1,i}$ and $Z_{2,i}$. For $\gamma_2 = \gamma_4 = \gamma_7 = 0$, this model reduces to the model considered in Section 6.1 of Andrews and Mikusheva (2016). We set $\gamma_1 = \cdots = \gamma_7 = 1$. To evaluate the performance of our algorithms with one endogenous variable, we set $\gamma_4 = \gamma_7 = 0$ and use Z_{1i} as the instrument. Following Andrews and Mikusheva (2016), we consider a symmetric as well as an asymmetric DGP:

$$\begin{aligned}
&(U_i, D_{1,i}, D_{2,i}, Z_{1,i}, Z_{2,i}, X_i) \\
&\quad = \big(\Phi(\xi_{U,i}), \Phi(\xi_{D_1,i}), \Phi(\xi_{D_2,i}), \Phi(\xi_{Z_1,i}), \Phi(\xi_{Z_2,i}), \Phi(\xi_{X,i})\big) \quad \text{(symmetric)}, \\
&(U_i, D_{1,i}, D_{2,i}, Z_{1,i}, Z_{2,i}, X_i) \\
&\quad = \big(\xi_{U,i}, \exp(2\xi_{D_1,i}), \exp(2\xi_{D_2,i}), \xi_{Z_1,i}, \xi_{Z_2,i}, \xi_{X,i}\big) \quad \text{(asymmetric)},
\end{aligned}$$

TABLE 7. Bias and RMSE, 401(k) DGP with three endogenous regressors.

	Bias/10^2		RMSE/10^3	
τ	Contr	Nested	Contr	Nested
		Coefficient on D		
0.15	−3.40	−5.02	7.52	7.62
0.25	−0.98	−1.52	4.11	4.12
0.50	−1.04	−1.42	2.03	2.06
0.75	−1.67	−1.50	1.86	1.86
0.85	0.86	1.03	2.05	2.05
		Coefficient on D_2		
0.15	1.12	−0.13	1.01	1.03
0.25	2.11	0.74	1.01	1.00
0.50	0.80	−0.28	0.94	0.99
0.75	−0.39	0.48	1.00	1.09
0.85	−2.64	−0.83	1.13	1.25
		Coefficient on D_3		
0.15	1.70	−0.25	1.08	1.13
0.25	1.57	−0.21	0.99	1.01
0.50	0.95	−0.06	0.92	0.96
0.75	−1.15	−0.33	1.02	1.11
0.85	−1.01	0.23	1.22	1.37

Note: Monte Carlo simulation with 500 repetitions as described in the main text. Contr: contraction algorithm; Nested: nested algorithm based on Brent's method. We use 2SLS estimates as starting values.

where $(\xi_{U,i}, \xi_{D_1,i}, \xi_{D_2,i}, \xi_{Z_1,i}, \xi_{Z_2,i}, \xi_{X,i})$ is a Gaussian vector with mean zero, all variances are set equal to one, $\mathrm{Cov}(\xi_U, \xi_{D_1}) = \mathrm{Cov}(\xi_U, \xi_{D_2}) = 0.5$, $\mathrm{Cov}(\xi_{D_1}, \xi_{Z_1}) = 0.8$, $\mathrm{Cov}(\xi_{D_2}, \xi_{Z_2}) = 0.4$, which allows us to investigate the impact of instrument strength,

TABLE 8. Computation time, 401(k) DGP with three endogenous regressors.

N	Contr	Nested
1000	0.36	6.29
5000	4.47	42.93
10,000	10.11	145.58

Note: The table reports average computation time in seconds at $\tau = 0.5$ over 20 simulation repetitions based on the DGP described in the main text. Contr: contraction algorithm; Nested: nested algorithm based on Brent's method. We use 2SLS estimates as starting values.

TABLE 9. Bias and RMSE, symmetric design with one endogenous regressor.

	$N = 500$							
	Bias				RMSE			
τ	Contr	Brent	Profil	InvQR	Contr	Brent	Profil	InvQR
0.15	0.03	−0.00	−0.02	−0.00	0.10	0.10	0.11	0.10
0.25	0.03	0.00	−0.01	0.00	0.12	0.12	0.12	0.12
0.50	−0.00	−0.00	−0.02	−0.00	0.12	0.14	0.14	0.14
0.75	−0.04	−0.01	−0.03	−0.01	0.13	0.12	0.12	0.12
0.85	−0.04	−0.00	−0.02	−0.00	0.11	0.11	0.11	0.11

	$N = 1000$							
	Bias				RMSE			
τ	Contr	Brent	Profil	InvQR	Contr	Brent	Profil	InvQR
0.15	0.02	0.00	−0.00	0.00	0.07	0.07	0.07	0.07
0.25	0.01	−0.00	−0.01	−0.00	0.08	0.08	0.08	0.08
0.50	−0.01	−0.01	−0.01	−0.01	0.09	0.10	0.10	0.10
0.75	−0.02	−0.00	−0.01	−0.00	0.09	0.08	0.08	0.08
0.85	−0.02	−0.00	−0.01	−0.00	0.08	0.08	0.08	0.08

Note: Monte Carlo simulation with 500 repetitions as described in the main text. Contr: contraction algorithm; Brent: root-finding algorithm based on Brent's method; Profil: profiling estimator based on Brent's method; InvQR: inverse quantile regression. We use 2SLS estimates as starting values.

all other covariances are equal to zero, and Φ is the cumulative distribution function of the standard normal distribution.[28]

We first investigate the bias and RMSE of the different methods. Tables 9–12 present the results. With one endogenous variable, the performances of the root-finding algorithm using Brent's method, the profiling estimator, and IQR are similar both in terms of bias and RMSE. The contraction algorithm performs well, but exhibits some bias at the tail quantiles. Turning to the results with two endogenous variables, we can see that the nested algorithm exhibits the best overall performance, both in terms of bias and RMSE. The performances of the SA-based optimization algorithm, IQR, and the profiling estimator are similar and only slightly worse than that of the nested algorithm. The contraction algorithm tends to exhibit some bias at the tail quantiles. However, this bias decreases substantially as the sample size gets larger. Finally, comparing the results for the coefficients on D_1 and D_2, we can see that the instrument strength matters for the performance of all estimators (including IQR), suggesting that weak identification can have implications for the estimation of IVQR models.

Table 13 displays the empirical coverage probabilities of the bootstrap confidence intervals. The results show that the our bootstrap procedure exhibits excellent size properties.

[28] To ensure that the weights are positive, we transform $Z_{1,i}$ and $Z_{2,i}$ by subtracting the minimum over each sample and adding 0.1 under the asymmetric DGP.

TABLE 10. Bias and RMSE, asymmetric design with one endogenous regressor.

	$N = 500$							
	Bias				RMSE			
τ	Contr	Brent	Profil	InvQR	Contr	Brent	Profil	InvQR
0.15	0.11	0.01	−0.04	−0.00	0.22	0.20	0.21	0.20
0.25	0.07	0.00	−0.02	−0.00	0.17	0.16	0.17	0.16
0.50	0.04	−0.00	−0.02	−0.00	0.13	0.12	0.12	0.12
0.75	0.03	0.00	−0.01	0.00	0.11	0.11	0.11	0.11
0.85	−0.03	−0.01	−0.03	−0.00	0.12	0.11	0.12	0.11

	$N = 1000$							
	Bias				RMSE			
τ	Contr	Brent	Profil	InvQR	Contr	Brent	Profil	InvQR
0.15	0.05	−0.01	−0.03	−0.01	0.16	0.15	0.15	0.15
0.25	0.04	0.00	−0.01	0.00	0.11	0.11	0.11	0.11
0.50	0.03	0.00	−0.01	0.00	0.08	0.08	0.08	0.08
0.75	0.01	−0.01	−0.02	−0.01	0.08	0.08	0.08	0.08
0.85	−0.03	−0.01	−0.02	−0.01	0.09	0.09	0.09	0.09

Note: Monte Carlo simulation with 500 repetitions as described in the main text. Contr: contraction algorithm; Brent: root-finding algorithm based on Brent's method; Profil: profiling estimator based on Brent's method; InvQR: inverse quantile regression. We use 2SLS estimates as starting values.

APPENDIX E: PROOFS OF THEORETICAL RESULTS IN SECTION 3

PROOF OF LEMMA 1. (i) We first show that L_1 is well-defined. For a given $\theta_{-1} \in \mathbb{R}^{d_D}$, let $\theta_1^* \in \arg\min_{\tilde{\theta}_1 \in \mathbb{R}^{d_X}} Q_{P,1}(\tilde{\theta}_1, \theta_{-1})$. Under Assumption 2, the objective function $\tilde{\theta}_1 \mapsto Q_{P,1}(\tilde{\theta}_1, \theta_{-1})$ is convex and differentiable with respect to $\tilde{\theta}_1$. Therefore, by the necessary and sufficient condition of minimization, θ_1^* solves

$$E_P\big[\big(1\{Y \le D'\theta_{-1} + X'\theta_1^*\}\big)X\big] = 0.$$

In what follows, we show that the map $L_1 : \theta_{-1} \mapsto \theta_1^*$ is well-defined on R_{-1} using a global inverse function theorem. Recall that

$$\Psi_{P,1}(\theta) = E_P\big[\big(1\{Y \le D'\theta_{-1} + X'\theta_1\}\big)X\big].$$

This function is continuously differentiable with respect to θ. The Jacobian is given by

$$J_{\Psi_{P,1}}(\theta) = \frac{\partial}{\partial\theta'} E_P\big[F_{Y|D,X,Z}\big(D'\theta_{-1} + X'\theta_1\big)X\big] = E_P\big[f_{Y|D,X,Z}\big(D'\theta_{-1} + X'\theta_1\big)X\big(X', D'\big)\big],$$

where the second equality follows from Assumption 2 and the dominated convergence theorem. Define a transform $\Xi : \Theta \to \mathbb{R}^{d_X + d_D}$ by

$$\Xi(\theta) := \big(\Psi_{P,1}(\theta)', \theta_{-1}{}'\big)'. \tag{E.1}$$

We follow Krantz and Parks (2003, Section 3.3) to obtain an implicit function L_1 on a suitable domain such that $\theta_1 = L_1(\theta_2)$ if and only if $\Psi_{P,1}(\theta) = 0$. The key is to apply a

TABLE 11. Bias and RMSE, symmetric design with two endogenous regressors.

	$N = 500$									
	Bias					RMSE				
τ	Contr	NestBr	SimAnn	Profil	InvQR	Contr	NestBr	SimAnn	Profil	InvQR
	Coefficient on D_1									
0.15	0.00	−0.00	0.00	−0.02	−0.01	0.11	0.12	0.14	0.13	0.13
0.25	0.01	−0.00	−0.01	−0.02	−0.01	0.15	0.16	0.17	0.16	0.16
0.50	−0.02	−0.02	−0.02	−0.04	−0.02	0.17	0.19	0.19	0.19	0.20
0.75	−0.04	−0.03	−0.03	−0.05	−0.03	0.21	0.20	0.21	0.20	0.20
0.85	−0.05	−0.03	−0.03	−0.05	−0.03	0.18	0.17	0.18	0.18	0.17
	Coefficient on D_2									
0.15	0.10	−0.01	−0.01	−0.05	−0.02	0.27	0.27	0.29	0.28	0.31
0.25	0.10	−0.00	−0.02	−0.04	−0.02	0.29	0.29	0.30	0.30	0.30
0.50	−0.01	−0.02	−0.02	−0.06	−0.02	0.33	0.38	0.39	0.40	0.39
0.75	−0.15	−0.04	−0.06	−0.08	−0.05	0.40	0.40	0.41	0.41	0.41
0.85	−0.19	−0.05	−0.06	−0.10	−0.07	0.39	0.36	0.40	0.38	0.43

	$N = 1000$									
	Bias					RMSE				
τ	Contr	NestBr	SimAnn	Profil	InvQR	Contr	NestBr	SimAnn	Profil	InvQR
	Coefficient on D_1									
0.15	−0.00	−0.00	−0.01	−0.01	−0.00	0.08	0.09	0.10	0.09	0.10
0.25	−0.00	−0.00	−0.01	−0.01	−0.01	0.10	0.11	0.12	0.11	0.13
0.50	−0.01	−0.01	−0.01	−0.02	−0.01	0.12	0.13	0.13	0.13	0.16
0.75	−0.01	−0.01	−0.01	−0.01	−0.00	0.13	0.13	0.14	0.13	0.14
0.85	−0.02	−0.01	−0.02	−0.02	−0.02	0.12	0.12	0.13	0.12	0.13
	Coefficient on D_2									
0.15	0.05	−0.01	−0.01	−0.02	−0.02	0.19	0.19	0.21	0.19	0.20
0.25	0.05	−0.00	−0.01	−0.02	−0.01	0.22	0.21	0.23	0.22	0.23
0.50	−0.02	−0.02	−0.02	−0.04	−0.03	0.25	0.27	0.27	0.28	0.29
0.75	−0.09	−0.02	−0.02	−0.04	−0.03	0.27	0.25	0.28	0.25	0.26
0.85	−0.09	−0.01	−0.03	−0.04	−0.03	0.26	0.23	0.25	0.24	0.24

Note: Monte Carlo simulation with 500 repetitions as described in the main text. Contr: contraction algorithm; NestBr: nested algorithm based on Brent's method; SimAnn: simulated annealing based optimization algorithm; Profil: nested profiling estimator based on Brent's method; InvQR: inverse quantile regression. We use 2SLS estimates as starting values.

global inverse function theorem to Ξ. Toward this end, we analyze the Jacobian of Ξ, which is given as

$$\begin{aligned} J_{\Xi}(\theta) &= \begin{bmatrix} \partial \Psi_{P,1}(\theta_1, \theta_{-1})/\partial \theta_1' & \partial \Psi_{P,1}(\theta_1, \theta_{-1})/\partial \theta_{-1}' \\ 0_{d_{-1}\times d_1} & I_{d_{-1}} \end{bmatrix} \\ &= \begin{bmatrix} E_P\big[f_{Y|D,X,Z}\big(D'\theta_{-1} + X'\theta_1\big)XX'\big] & E_P\big[f_{Y|D,X,Z}\big(D'\theta_{-1} + X'\theta_1\big)XD'\big] \\ 0_{d_{-1}\times d_1} & I_{d_{-1}} \end{bmatrix}, \end{aligned} \tag{E.2}$$

where, for any $d \in \mathbb{N}$, I_d denotes the $d \times d$ identity matrix.

TABLE 12. Bias and RMSE, asymmetric design with two endogenous regressors.

	$N = 500$									
	Bias					RMSE				
τ	Contr	NestBr	SimAnn	Profil	InvQR	Contr	NestBr	SimAnn	Profil	InvQR
	Coefficient on D_1									
0.15	−0.02	0.02	0.00	−0.03	0.01	0.25	0.26	0.28	0.27	0.26
0.25	−0.05	0.01	0.00	−0.01	−0.00	0.20	0.20	0.21	0.20	0.21
0.50	−0.04	−0.00	0.00	−0.01	0.00	0.16	0.17	0.21	0.17	0.19
0.75	−0.02	−0.02	−0.01	−0.02	−0.02	0.17	0.17	0.18	0.18	0.19
0.85	−0.01	−0.01	−0.02	−0.02	−0.02	0.20	0.19	0.19	0.20	0.19
	Coefficient on D_2									
0.15	0.26	−0.06	−0.11	−0.16	−0.13	0.57	0.52	0.58	0.53	0.59
0.25	0.23	−0.01	−0.02	−0.06	−0.01	0.45	0.41	0.43	0.40	0.44
0.50	0.12	−0.03	−0.04	−0.07	−0.07	0.34	0.32	0.48	0.33	0.73
0.75	0.04	−0.05	−0.06	−0.11	−0.05	0.32	0.31	0.34	0.33	0.34
0.85	−0.13	−0.03	−0.01	−0.08	0.01	0.40	0.34	0.38	0.34	0.36

	$N = 1000$									
	Bias					RMSE				
τ	Contr	NestBr	SimAnn	Profil	InvQR	Contr	NestBr	SimAnn	Profil	InvQR
	Coefficient on D_1									
0.15	−0.03	0.01	−0.00	−0.02	−0.01	0.18	0.19	0.19	0.18	0.19
0.25	−0.04	−0.00	−0.01	−0.02	−0.01	0.15	0.15	0.16	0.15	0.16
0.50	−0.03	−0.01	−0.01	−0.01	−0.01	0.13	0.13	0.14	0.13	0.14
0.75	−0.03	−0.01	−0.01	−0.02	−0.01	0.12	0.12	0.13	0.13	0.14
0.85	0.01	0.00	0.00	−0.01	−0.00	0.14	0.13	0.15	0.14	0.15
	Coefficient on D_2									
0.15	0.15	−0.03	−0.03	−0.07	−0.04	0.37	0.37	0.38	0.37	0.39
0.25	0.10	−0.01	−0.01	−0.05	−0.02	0.28	0.28	0.30	0.28	0.28
0.50	0.05	−0.02	−0.03	−0.04	−0.03	0.22	0.22	0.23	0.22	0.24
0.75	0.06	−0.01	−0.02	−0.03	−0.01	0.24	0.22	0.24	0.23	0.24
0.85	−0.08	−0.03	−0.04	−0.07	−0.03	0.27	0.24	0.26	0.25	0.24

Note: Monte Carlo simulation with 500 repetitions as described in the main text. Contr: contraction algorithm; NestBr: nested algorithm based on Brent's method; SimAnn: simulated annealing based optimization algorithm; Profil: nested profiling estimator based on Brent's method; InvQR: inverse quantile regression. We use 2SLS estimates as starting values.

Let $I \subset \{1, \dots, d_X + d_D\}$. For any matrix A, let $[A]_{I,I}$ denote a principal minor of A, which collects the rows and columns of A whose indices belong to the index set I. By (E.2), if $I \subset \{1, \dots, d_1\}$,

$$[J_{\Xi}(\theta)]_{I,I} = E_P\big[f_{Y|D,X,Z}\big(D'\theta_{-1} + X'\theta_1\big)\tilde{X}\tilde{X}'\big]$$

for a subvector $\tilde{X}$ of X, which is positive definite by Assumption 2 and Lemma 4. If $I \subset \{d_1 + 1, \dots, d_X + d_D\}$, $[J_{\Xi}(\theta)]_{I,I} = I_\ell$ for some $1 \le \ell \le d_D$ and is hence positive definite.

TABLE 13. Coverage, location-scale DGP with one endogenous regressor.

	$N=500$							
	Symmetric Design				Asymmetric Design			
	$1-\alpha=0.95$		$1-\alpha=0.9$		$1-\alpha=0.95$		$1-\alpha=0.9$	
τ	Contr	Brent	Contr	Brent	Contr	Brent	Contr	Brent
0.15	0.93	0.98	0.88	0.94	0.89	0.97	0.85	0.95
0.25	0.94	0.96	0.89	0.93	0.93	0.97	0.88	0.95
0.50	0.96	0.96	0.91	0.91	0.95	0.97	0.91	0.93
0.75	0.94	0.97	0.90	0.94	0.95	0.97	0.91	0.93
0.85	0.95	0.98	0.90	0.96	0.96	0.98	0.92	0.96

	$N=1000$							
	Symmetric Design				Asymmetric Design			
	$1-\alpha=0.95$		$1-\alpha=0.9$		$1-\alpha=0.95$		$1-\alpha=0.9$	
τ	Contr	Brent	Contr	Brent	Contr	Brent	Contr	Brent
0.15	0.95	0.96	0.90	0.92	0.92	0.97	0.85	0.93
0.25	0.95	0.96	0.91	0.91	0.92	0.95	0.87	0.91
0.50	0.96	0.96	0.90	0.90	0.94	0.96	0.90	0.93
0.75	0.95	0.96	0.90	0.91	0.95	0.95	0.91	0.91
0.85	0.96	0.97	0.93	0.94	0.96	0.95	0.92	0.91

Note: Monte Carlo simulation with 1000 repetitions as described in the main text. Contr: contraction algorithm; Brent: root-finding algorithm based on Brent's method. We use 2SLS estimates as starting values.

Otherwise, any principal minor is of the following form:

$$[J_{\Xi}(\theta)]_{I,I} = \begin{bmatrix} E_P[f_{Y|D,X,Z}(D'\theta_{-1} + X'\theta_1)\tilde{X}\tilde{X}'] & B \\ 0_{\ell\times m} & I_\ell \end{bmatrix}$$

for some subvector $\tilde{X}$ of X and a $m\times\ell$ matrix B. Note that

$$\begin{aligned} \det([J_{\Xi}(\theta)]_{I,I}) &= \det(E_P[f_{Y|D,X,Z}(D'\theta_{-1} + X'\theta_1)\tilde{X}\tilde{X}'] - BI_\ell^{-1}\times 0_{\ell\times m})\det(I_\ell) \\ &= \det(E_P[f_{Y|D,X,Z}(D'\theta_{-1} + X'\theta_1)\tilde{X}\tilde{X}']) > 0, \end{aligned}$$

where the last inequality follows again from Assumption 2 and Lemma 4. Hence, $J_{\Xi}(\theta)$ is a P-matrix. Note that Θ is a closed rectangle. By Theorem 4 in Gale and Nikaido (1965), Ξ is univalent, and hence the inverse map Ξ^{-1} is well-defined.

Let

$$\begin{aligned} R_{-1} &= \{\theta_{-1}\in\mathbb{R}^{d_{-1}} : (0, \theta_{-1}) \in \Xi(\Theta)\} \\ &= \{\theta_{-1}\in\mathbb{R}^{d_{-1}} : \Psi_{P,1}(\theta_1, \theta_{-1}) = 0, \text{for some } (\theta_1, \theta_{-1}) \in \Theta\}, \end{aligned}$$

which coincides with the definition in (3.5) with $j=1$. Let $F_1 = [I_{d_1}, 0_{d_1\times d_{-1}}]$. For each $\theta_{-1}\in R_{-1}$, define

$$L_1(\theta_{-1}) := F_1\Xi^{-1}(0, \theta_{-1}).$$

Then, for any $\theta \in \Theta$, $\Psi_{P,1}(\theta) = 0$ if and only if $\theta_{-1} \in R_{-1}$ and $\Xi(\theta) = (0, \theta_{-1})$. By the univalence of Ξ, this is true if and only if $\theta = \Xi^{-1}(0, \theta_{-1})$, and the first d_1 components extracted by applying F_1 is θ_1. This ensures L_1 is well-defined on R_{-1}.

Below, for any set A, let A^o denote the interior of A. Let $R^o_{-1} = \{\theta_{-1} \in \mathbb{R}^{d_{-1}} : (0, \theta_{-1}) \in \Xi(\Theta^o)\}$. Note that $\Psi_{P,1}$ is $\mathcal{C}^1$ on Θ^o and, for each $\theta = (\theta_1, \theta_{-1}) \in \Theta$ with $\theta_{-1} \in R^o_{d_{-1}}$, $\det(\partial \Psi_{P,1}(\theta)/\partial \theta_1') \neq 0$. Therefore, by the implicit function theorem, there is a $\mathcal{C}^1$-function $\tilde{L}_1$ and an open set V containing θ_{-1} such that

$$\Psi_{P,1}\big(\tilde{L}_1(\theta_{-1}), \theta_{-1}\big) = 0, \quad \text{for all } \theta_{-1} \in V.$$

However, such a local implicit function must coincide with the unique global map L_1 on V. Hence, $L_1|_V = \tilde{L}_1$ and, therefore L_1 is continuously differentiable at θ_{-1}. Since the choice of θ_{-1} is arbitrary, L_1 is continuously differentiable for all $\theta_{-1} \in R^o_2$.

Showing that the conclusion holds for any other L_j for $j = 2, \ldots, J$ is similar, and hence we omit the proof. □

Lemma 4. *Suppose $E_P[f_{Y|D,X,Z}(D'\theta_{-1} + X'\theta_1)XX']$ is positive definite. Then, for any subvector $\tilde{X}$ of X with dimension $\tilde{d}_X \leq d_X$, $E_P[f_{Y|D,X,Z}(D'\theta_{-1} + X'\theta_1)\tilde{X}\tilde{X}']$ is positive definite.*

Proof. In what follows, let $W = f_{Y|D,X,Z}(D'\theta_{-1} + X'\theta_1)$ and let

$$A := E_P\big[f_{Y|D,X,Z}\big(D'\theta_{-1} + X'\theta_1\big)XX'\big] = E\big[WXX'\big].$$

Let $\tilde{X}$ be a subvector of X with $\tilde{d}_X$ components. Then there exists a $d_X \times d_X$ permutation matrix P_π such that the first $\tilde{d}_X$ components of $P_\pi X$ is $\tilde{X}$.

Let $B := E[WP_\pi XX'P_\pi']$ and note that

$$B = P_\pi E\big[WXX'\big]P_\pi' = P_\pi A P_\pi', \tag{E.3}$$

by the linearity of the expectation operator and W being a scalar. Let λ be an eigenvalue of B such that

$$Bz = \lambda z, \tag{E.4}$$

for the corresponding eigenvector $z \in \mathbb{R}^{d_X}$. By (E.3)–(E.4),

$$P_\pi A P_\pi' z = \lambda z \quad \Leftrightarrow \quad A P_\pi' z = \lambda P_\pi^{-1} z.$$

Note that $P_\pi^{-1} = P_\pi'$ due to P_π being a permutation matrix. Letting $y := P_\pi' z$ then yields

$$Ay = \lambda y,$$

which in turn shows that λ is an eigenvalue of A. For any eigenvalue of A, the argument above can be reversed to show that it is also an eigenvalue of B. Since the choice of the eigenvalue is arbitrary, A and B share the same eigenvalues.

Now let $C := E[W\tilde{X}\tilde{X}']$ and note that it is a leading principal submatrix of B. Then, by the eigenvalue inclusion principle (Horn and Johnson (2012, Theorem 4.3.28)),

$$\lambda_{\min}(C) \geq \lambda_{\min}(B) = \lambda_{\min}(A) > 0,$$

where the last inequality follows from the positive definiteness of A. This completes the claim of the lemma. □

Proof of Corollary 1. The existence of K and its continuous differentiability follows immediately from Lemma 1. For M, by the definition of $\tilde{R}_1$, for any $\theta_{-1} \in \tilde{R}_j$, there exists $(\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$ such that

$$\Psi_{P,1}(\theta_1, \theta_{-1}) = 0,$$
$$\Psi_{P,2}(\theta_1, \theta_2, \pi_{-\{1,2\}}\theta_{-1}) = 0.$$

By (i), one may then write $\theta_1 = L_1(\theta_{-1})$ and $\theta_2 = L_2(L_1(\theta_{-1}), \pi_{-\{1,2\}}\theta_{-1})$. Hence, the map $M_1 : \tilde{R}_1 \to \Theta_2$ below is well-defined:

$$M_1(\theta_{-1}) = L_2\big(L_1(\theta_{-1}), \pi_{-\{1,2\}}\theta_{-1}\big).$$

Recursively, arguing in the same way, the maps

$$\begin{aligned} M_2(\theta_{-1}) &= L_3\big(L_1(\theta_{-1}), M_1(\theta_{-1}), \pi_{-\{1,2,3\}}\theta_{-1}\big) \\ &\vdots \\ M_j(\theta_{-1}) &= L_{j+1}\big(L_1(\theta_{-1}), M_1(\theta_{-1}), \dots, M_{j-1}(\theta_{-1}), \pi_{-\{1,\dots,j+1\}}\theta_{-1}\big) \\ &\vdots \\ M_{d_D}(\theta_{-1}) &= L_J\big(L_1(\theta_{-1}), M_1(\theta_{-1}), \dots, M_{d_D-1}(\theta_{-1})\big) \end{aligned}$$

are well-defined on $\tilde{R}_2, \dots, \tilde{R}_{d_D}$, respectively. The continuous differentiability of M follows from that of L_js and the chain rule. □

Proof of Proposition 1. $\Longrightarrow$: For every solution, $\Psi_P(\theta^*) = 0$, $\theta_j^* = L_j(\theta_{-j}^*)$ by construction under Assumptions 1 and 2. It follows that $K(\theta^*) = \theta^*$ and $M(\theta_{-1}^*) = \theta_{-1}^*$.

$\Longleftarrow$: For the simultaneous response, note that $K(\bar{\theta}) = \bar{\theta}$ implies that $\bar{\theta}_j = L_j(\bar{\theta}_{-j})$ for all $j \in \{1, \dots, J\}$. Thus, $\bar{\theta}$ solves $\Psi_P(\bar{\theta}) = 0$ by Lemma 1. Consider next the sequential response. Let $\tilde{\theta}, \bar{\theta} \in \Theta$ be such that $\tilde{\theta}_j = L_j(\bar{\theta}_{-j})$ for $j = 1, \dots, J$. By Lemma 1, they satisfy

$$\begin{aligned} \Psi_{P,1}(\tilde{\theta}_1, \bar{\theta}_2, \bar{\theta}_3, \dots, \bar{\theta}_J) &= 0 \\ \Psi_{P,2}(\tilde{\theta}_1, \tilde{\theta}_2, \bar{\theta}_3, \dots, \bar{\theta}_J) &= 0 \\ &\vdots \\ \Psi_{P,J}(\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\theta}_3, \dots, \tilde{\theta}_J) &= 0. \end{aligned}$$

Thus, a fixed point $\tilde{\theta} = \bar{\theta}$ satisfies $\Psi_P(\bar{\theta}) = 0$. □

Appendix F: Proofs of theoretical results in Section 4

Proof of Proposition 2. We prove the result for K. By Assumption 3, there exists a strictly convex set $\tilde{D}_K$ on which the spectral norm of the Jacobian of K is uniformly bounded by 1. This ensures that K is a contraction map on $cl(\tilde{D}_K)$, and the claim of the proposition now follows from Theorem 2.2.16 in Hasselblatt and Katok (2003). □

Appendix G: Proofs of theoretical results in Section 6

To state and prove results in a concise manner, we use the population and sample simultaneous response maps K and $\hat{K}$ below to define our estimand θ^* and estimator $\hat{\theta}_N$. Namely, θ^* is the fixed point of K, and $\hat{\theta}_N$ solves

$$\|\hat{\theta}_N - \hat{K}(\hat{\theta}_N)\| \le \inf_{\theta' \in \Theta} \|\theta' - \hat{K}(\theta')\| + o_p(N^{-1/2}).$$

Note that the fixed-point estimator defined in (6.1)–(6.2) is asymptotically equivalent to the estimator above due to Lemma 5.

Proof of Theorem 1. Let $H := I_{d_X+d_D} - K$. A fixed point θ^* of K then satisfies

$$H(\theta^*) = 0.$$

Similarly, let $\hat{H} := I_{d_X+d_D} - \hat{K}$. The estimator $\hat{\theta}_N$ satisfies

$$\|\hat{H}(\hat{\theta}_N)\|^2 \le \inf_{\theta' \in \Theta} \|\hat{H}(\theta')\|^2 + r_N^2,$$

where $r_N = o_p(N^{-1/2})$. Let $\varphi : \ell^\infty(\Theta)^{d_X+d_D} \times \mathbb{R} \to \mathbb{R}^{d_X+d_D}$ be a map such that, for each $(H, r) \in \ell^\infty(\Theta)^{d_X+d_D} \times \mathbb{R}$, $\tilde{\theta} = \varphi(H, r)$ is an r-approximate solution, which satisfies

$$\|H(\tilde{\theta})\|^2 \le \inf_{\theta' \in \Theta} \|H(\theta')\|^2 + r^2.$$

One may then write

$$\sqrt{N}(\hat{\theta}_N - \theta^*) = \sqrt{N}(\varphi(\hat{H}, \hat{r}) - \varphi(H, 0)).$$

By Lemma 12, $\sqrt{N}(\hat{K} - K) \rightsquigarrow \mathbb{W}$ in $\ell^\infty(\Theta)^{d_X+d_D}$, where $\mathbb{W}$ is a Gaussian process defined in Lemma 12. Assumption 2(4) and $J_{\Psi_P}(\theta^*)$ being full rank imply $\det(I_{d_X+d_D} - J_K(\theta^*)) \neq 0$ (see (G.4)), which ensures the condition of Lemma 7. By Lemmas 6–7, Condition Z in Chernozhukov, Fernandez-Val, and Melly (2013) (CFM henceforth) holds, which in turn ensures that one may apply Lemmas E.2 and E.3 in CFM. This ensures

$$\sqrt{N}(\varphi(\hat{H}, \hat{r}) - \varphi(H, 0)) \rightsquigarrow \varphi'_{H,0}(\mathbb{W}, 0) = -\dot{H}_{\theta^*}^{-1}\mathbb{W}(\theta^*).$$

Hence, we obtain (6.3) with

$$V = \dot{H}_{\theta^*}^{-1} E[\mathbb{W}(\theta^*)\mathbb{W}(\theta^*)'](\dot{H}_{\theta^*}^{-1})'.$$

Finally, note that $\dot{H}_{\theta^*} = I_{d_X+d_D} - J_K(\theta^*)$ by Lemma 7. This establishes the theorem. □

PROOF OF THEOREM 2. Recall that $\hat{H} = I_{d_X+d_D} - \hat{K}$. The estimator $\hat{\theta}_N$ satisfies

$$\|\hat{H}(\hat{\theta}_N)\|^2 \le \inf_{\theta'\in\Theta} \|\hat{H}(\theta')\|^2 + r_N^2,$$

where $r_N = o_p(N^{-1/2})$. Similarly, let $\hat{H}^* = I_{d_X+d_D} - \hat{K}^*$. Let P^* denote the law of $\hat{H}^*$ conditional on $\{W_i\}_{i=1}^\infty$. The bootstrap estimator $\hat{\theta}_N^*$ satisfies

$$\|\hat{H}^*(\hat{\theta}_N^*)\|^2 \le \inf_{\theta'\in\Theta} \|\hat{H}^*(\theta')\|^2 + (r_N^*)^2,$$

where $r_N^* = o_{P^*}(N^{-1/2})$ conditional on $\{W_i\}_{i=1}^\infty$.

Using the r-approximation, one may therefore write

$$\sqrt{N}(\hat{\theta}_N^* - \hat{\theta}_N) = \sqrt{N}(\varphi(\hat{H}^*, r_N^*) - \varphi(\hat{H}, r_N)).$$

Let E_{P^*} denote the conditional expectation with respect to P^*. Let BL_1 denote the space of bounded Lipschitz functions on $\mathbb{R}^{d_X+d_D}$ with Lipschitz constant 1. Then, for any $\epsilon > 0$,

$$\begin{aligned}
&\sup_{h\in BL_1} \left|E_{P^*} h\big(\sqrt{N}[\varphi(\hat{H}^*, r_N^*) - \varphi(\hat{H}, r_N)]\big) - E_{P^*} h\big(\varphi'_{H,0}(\sqrt{N}[(\hat{H}^*, r_N^*)' - (\hat{H}, r_N)'])\big)\right| \\
&\quad \le \epsilon + 2P^*\big(\|\sqrt{N}[\varphi(\hat{H}^*, r_N^*) - \varphi(\hat{H}, r_N)] - \varphi'_{H,0}(\sqrt{N}[(\hat{H}^*, r_N^*) - (\hat{H}, r_N)])\| > \epsilon\big). \quad \text{(G.1)}
\end{aligned}$$

By Lemma 12, $\sqrt{N}(\hat{H}^* - \hat{H}) = -\sqrt{N}(\hat{K}^* - \hat{K}) \overset{L^*}{\rightsquigarrow} -\mathbb{W} \overset{d}{=} \mathbb{W}$. Noting that $h \circ \varphi'_{H,0} \in BL_1(\ell^\infty(\Theta) \times \mathbb{R})$ and $r_N = o_p(N^{-1/2})$, it follows that

$$\sup_{h\in BL_1} \left|E_{P^*} h \circ \varphi'_{H,0}\big(\sqrt{N}[(\hat{H}^*, r_N^*) - (\hat{H}, r_N)]\big) - E_{P^*} h \circ \varphi'_{H,0}(\mathbb{W}, 0)\right| \to 0,$$

with probability approaching 1 due to $r_N = o_P(N^{-1/2})$. Hence, for the conclusion of the theorem, it suffices to show that the second term on the right-hand side of (G.1) tends to 0 in probability.

For this, as shown in the proof of Theorem 1, φ is Hadamard differentiable at $(H, 0)$. Hence, by Theorem 3.9.4 in Van der Vaart and Wellner (1996),

$$\begin{aligned}
\sqrt{N}[\varphi(\hat{H}^*, r_N^*) - \varphi(H, 0)] &= \varphi'_{H,0}\big(\sqrt{N}[(\hat{H}^*, r_N^*) - (H, 0)]\big) + o_{P^*}(1), \\
\sqrt{N}[\varphi(\hat{H}, r_N) - \varphi(H, 0)] &= \varphi'_{H,0}\big(\sqrt{N}[(\hat{H}, r_N) - (H, 0)]\big) + o_P(1).
\end{aligned}$$

Take the difference of the left- and right-hand sides of the equations above respectively and note that $\varphi'_{H,0}$ is linear. This implies the second term on the right-hand side of (G.1) tends to 0 in probability. This ensures

$$\sqrt{N}(\varphi(\hat{H}, r_N^*) - \varphi(\hat{H}, r_N)) \overset{L^*}{\rightsquigarrow} \varphi'_{H,0}(\mathbb{W}, 0) = -\dot{H}_{\theta^*}^{-1}\mathbb{W}(\theta^*). \qquad \square$$

PROOF OF COROLLARY 2. Note that V and g may be written as

$$V = \big(I_{d_X+d_D} - J_K(\theta^*)\big)^{-1} E[g(W;\theta^*)g(W;\theta^*)'][\big(I_{d_X+d_D} - J_K(\theta^*)\big)^{-1}]', \quad \text{(G.2)}$$

$$g(w;\theta^*) = R^{-1}(\theta^*) f(w;\theta^*), \quad \text{(G.3)}$$

where $R(\theta^*)$ is a $d_X + d_D$-by-$d_X + d_D$ matrix given by

$$
\begin{aligned}
R(\theta^*) &= \begin{pmatrix}
\frac{\partial^2}{\partial\theta_1\partial\theta_1'}Q_{P,1}(\theta^*) & 0 & \cdots & \cdots & 0 \\
0 & \frac{\partial^2}{\partial\theta_2\partial\theta_2'}Q_{P,2}(\theta^*) & 0 & \cdots & 0 \\
\vdots & 0 & \ddots & & \vdots \\
0 & \cdots & \cdots & & \frac{\partial^2}{\partial\theta_J\partial\theta_J'}Q_{P,J}(\theta^*)
\end{pmatrix} \\
&= \begin{pmatrix}
\frac{\partial}{\partial\theta_1'}\Psi_{P,1}(\theta^*) & 0 & \cdots & \cdots & 0 \\
0 & \frac{\partial}{\partial\theta_2'}\Psi_{P,2}(\theta^*) & 0 & \cdots & 0 \\
\vdots & 0 & \ddots & & \vdots \\
0 & \cdots & \cdots & & \frac{\partial}{\partial\theta_J'}\Psi_{P,J}(\theta^*)
\end{pmatrix}.
\end{aligned}
$$

Further, by Lemma 1 and the form of $J_{L_{-j}}(\theta_j^*)$ given in (4.1),

$$
\begin{aligned}
&J_K(\theta^*) \\
&= \begin{pmatrix}
0_{d_X\times d_X} & \frac{\partial L_1(\theta^*)}{\partial\theta_2'} & \cdots & \frac{\partial L_1(\theta^*)}{\partial\theta_J'} \\
\frac{\partial L_2(\theta^*)}{\partial\theta_1'} & 0 & \cdots & \frac{\partial L_2(\theta^*)}{\partial\theta_J'} \\
\vdots & & \ddots & \vdots \\
\frac{\partial L_J(\theta^*)}{\partial\theta_1'} & \frac{\partial L_J(\theta^*)}{\partial\theta_2'} & \cdots & 0
\end{pmatrix} \\
&= \begin{pmatrix}
0_{d_X\times d_X} & -\left(\frac{\partial\Psi_{P,1}(\theta^*)}{\partial\theta_1'}\right)^{-1}\frac{\partial\Psi_{P,1}(\theta*)}{\partial\theta_2'} & \cdots & -\left(\frac{\partial\Psi_{P,1}(\theta^*)}{\partial\theta_1'}\right)^{-1}\frac{\partial\Psi_{P,1}(\theta^*)}{\partial\theta_J'} \\
-\left(\frac{\partial\Psi_{P,2}(\theta^*)}{\partial\theta_2'}\right)^{-1}\frac{\partial\Psi_{P,2}(\theta^*)}{\partial\theta_1'} & 0 & \cdots & -\left(\frac{\partial\Psi_{P,2}(\theta^*)}{\partial\theta_2'}\right)^{-1}\frac{\partial\Psi_{P,2}(\theta^*)}{\partial\theta_J'} \\
\vdots & & \ddots & \vdots \\
-\left(\frac{\partial\Psi_{P,J}(\theta^*)}{\partial\theta_J'}\right)^{-1}\frac{\partial\Psi_{P,J}(\theta^*)}{\partial\theta_1'} & -\left(\frac{\partial\Psi_{P,J}(\theta^*)}{\partial\theta_J'}\right)^{-1}\frac{\partial\Psi_{P,2}(\theta^*)}{\partial\theta_2'} & \cdots & 0
\end{pmatrix}.
\end{aligned}
$$

The form of $R(\theta^*)$ and $J_K(\theta^*)$ imply

$$
R(\theta^*)\big(I_{d_X+d_D} - J_K(\theta^*)\big) = J_{\Psi_P}(\theta^*), \tag{G.4}
$$

where J_{Ψ_P} is the Jacobian of the estimating equations. Equations (G.2)–(G.3) and (G.4) ensure that one may also write

$$
V = J_{\Psi_P}(\theta^*)^{-1}E\big[f(W;\theta^*)f(W;\theta^*)'\big]\big[J_{\Psi_P}(\theta^*)^{-1}\big]'. \tag{G.5}
$$

As shown in Chernozhukov and Hansen (2006), the Jacobian of Ψ_P is given by

$$J_{\Psi_P}(\theta^*) = E\big[f_{\varepsilon(\tau)|X,D,Z}(0)\Psi(\tau)\big[X', D'\big]\big], \tag{G.6}$$

where $\Psi(\tau) = (X', Z')'$. Furthermore,

$$E\big[f\big(W; \theta^*\big)f\big(W; \theta^*\big)'\big] = \tau(1-\tau)E\big[\Psi(\tau)\Psi(\tau)'\big]. \tag{G.7}$$

Hence, (G.5)–(G.7) show that V coincides with the asymptotic variance of the estimator that solves the estimating equations in (6.5). □

LEMMA 5. *Suppose Assumptions* 1–2 *hold.* (i) *Let* $\hat{\theta}_N$ *be an estimator of* θ^* *that satisfies*

$$\|\hat{\theta}_N - \hat{K}(\hat{\theta}_N)\| \le \inf_{\theta'\in\Theta}\|\theta' - \hat{K}(\theta')\| + o_p(N^{-1/2}). \tag{G.8}$$

Then, it also satisfies (6.1)–(6.2); (ii) *Let* $\hat{\theta}_N$ *be an estimator of* θ^* *that satisfies* (6.1)–(6.2). *Then it also satisfies* (G.8).

PROOF. (i) Consider the case $j = 2$. Note that, by (G.8),

$$\begin{aligned}
&\hat{\theta}_{N,2} - \hat{L}_2\big(\hat{L}_1(\hat{\theta}_{N,-1}), \hat{\theta}_{N,3}, \dots, \hat{\theta}_{N,J}\big) \\
&\quad = \hat{\theta}_{N,2} - \hat{L}_2(\hat{\theta}_{N,1} + r_{N,1}, \hat{\theta}_{N,3}, \dots, \hat{\theta}_{N,J}) &\text{(G.9)}\\
&\quad = \hat{L}_2(\hat{\theta}_{N,1}, \hat{\theta}_{N,3}, \dots, \hat{\theta}_{N,J}) - \hat{L}_2(\hat{\theta}_{N,1} + r_{N,1}, \hat{\theta}_{N,3}, \dots, \hat{\theta}_{N,J}), &\text{(G.10)}
\end{aligned}$$

where $r_{N,1} = o_p(N^{-1/2})$, and the second equality follows from the definition of $\hat{\theta}_{N,2}$. The right-hand side of (G.10) can be written as

$$\begin{aligned}
&\hat{L}_2(\hat{\theta}_{N,1}, \hat{\theta}_{N,3}, \dots, \hat{\theta}_{N,J}) - \hat{L}_2(\hat{\theta}_{N,1} + r_{N,1}, \hat{\theta}_{N,3}, \dots, \hat{\theta}_{N,J}) \\
&\quad = \big(\big[\hat{L}_2(\hat{\theta}_{N,1}, \hat{\theta}_{N,3}, \dots, \hat{\theta}_{N,J}) - L_2(\hat{\theta}_{N,1}, \hat{\theta}_{N,3}, \dots, \hat{\theta}_{N,J})\big] \\
&\qquad - \big[\hat{L}_2(\hat{\theta}_{N,1} + r_{N,1}, \hat{\theta}_{N,3}, \dots, \hat{\theta}_{N,J}) - L_2(\hat{\theta}_{N,1} + r_{N,1}, \hat{\theta}_{N,3}, \dots, \hat{\theta}_{N,J})\big]\big) \\
&\qquad + \big[L_2(\hat{\theta}_{N,1} + r_{N,1}, \hat{\theta}_{N,3}, \dots, \hat{\theta}_{N,J}) - L_2(\hat{\theta}_{N,1}, \hat{\theta}_{N,3}, \dots, \hat{\theta}_{N,J})\big] \\
&\quad = o_p\big(N^{-1/2}\big) + O_P(r_{N,1}), &\text{(G.11)}
\end{aligned}$$

where the last equality follows from the stochastic equicontinuity of $\mathcal{L}_N$ shown in the proof of Lemma 11 and L_2 being Lipschitz since L_2 is continuously differentiable with a derivative that is uniformly bounded on the compact set Θ. By (G.9)–(G.11), it holds that $\hat{\theta}_{N,j} = M_j(\hat{\theta}_{N,-1}) + o_p(N^{-1/2})$ for $j = 2$. Repeat the same argument sequentially for $j = 3, \dots, J$. The first conclusion of the lemma then follows.

(ii) Suppose now that $r_{N,1} := \hat{\theta}_{N,1} - \hat{L}_1(\hat{\theta}_{N,-1}) \neq o_P(N^{-1/2})$. Then, there is a subsequence k_N along which, for any $\eta > 0$, $\sqrt{k_N} r_{k_N,1} > \eta$ for all k_N with positive probability. Then the $O_P(r_{k_N,1})$-term in (G.11) is not $o_p(k_N^{-1/2})$, which therefore implies $\hat{\theta}_{N,j} \neq M_j(\hat{\theta}_{N,-1}) + o_p(N^{-1/2})$ for $j = 2$. The second conclusion of the lemma then follows. □

LEMMA 6. *Let $\Lambda \subset \mathbb{R}^p$ be a compact set, and let $K : \Lambda \to \mathbb{R}^p$ be a map that has a unique fixed point $\lambda_0 \in \Lambda$. let $H : \Lambda \to \mathbb{R}^p$ be defined by $H(\lambda) := \lambda - K(\lambda)$. Then $H^{-1}(x) = \{\lambda \in \Lambda : H(\lambda) = x\}$ is continuous at $x = 0$ in Hausdorff distance.*

PROOF. For any x, write

$$H^{-1}(x) = \{\lambda : \lambda - K(\lambda) = x\}.$$

Let $x_n \to 0$. Since λ_0 is the unique fixed point of K, $H^{-1}(0) = \{\lambda_0\}$. Therefore,

$$\begin{aligned} d_H\big(H^{-1}(0), H^{-1}(x_n)\big) &= \max\Big\{\inf_{\lambda \in H^{-1}(x_n)} \|\lambda - \lambda_0\|, \sup_{\lambda \in H^{-1}(x_n)} \|\lambda - \lambda_0\|\Big\} \\ &= \sup_{\lambda \in H^{-1}(x_n)} \|\lambda - \lambda_0\|. \end{aligned}$$

Hence, it suffices to show that $\sup_{\lambda \in H^{-1}(x_n)} \|\lambda - \lambda_0\| = o(1)$. We show this by contradiction. Suppose that there is a sequence $\{\lambda_n\} \subset \Lambda$ and $\delta > 0$ such that $\lambda_n \in H^{-1}(x_n)$ for all n and $\{\lambda_n\}$ has a subsequence $\{\lambda_{k_n}\}$ such that $\|\lambda_{k_n} - \lambda_0\| > \delta$ for all n. $\lambda_{k_n} \in \Lambda$ is a sequence in a compact space, and hence there is a further subsequence λ_{h_n} such that $\lambda_{h_n} \to \lambda^*$ for some $\lambda^* \in \Lambda$ with $\lambda^* \neq \lambda_0$. By the continuity of K, one then has

$$\lambda_{h_n} - K(\lambda_{h_n}) \to \lambda^* - K\big(\lambda^*\big).$$

By $\lambda_{h_n} - K(\lambda_{h_n}) = x_n$ and $x_n \to 0$, it must hold that

$$\lambda^* - K\big(\lambda^*\big) = 0.$$

However this contradicts the fact that λ_0 is the unique fixed point, and hence the conclusion follows. □

LEMMA 7. *Suppose $H = I - K$ and $K : \mathbb{R}^p \to \mathbb{R}^p$ is continuously differentiable at λ_0. Suppose further that* $\det(I - J_K(\lambda_0)) \neq 0$. *Let $\dot{H}_{\lambda_0} := I - J_K(\lambda_0)$. Then*

$$\lim_{t \downarrow 0} \sup_{h : \|h\| = 1} \big\| t^{-1}\big[H(\lambda_0 + th) - H(\lambda_0)\big] - \dot{H}_{\lambda_0} h\big\| = 0,$$

and

$$\inf_{h : \|h\| = 1} \|\dot{H}_{\lambda_0} h\| > 0.$$

PROOF. Let $\{h_n\} \subset \mathbb{S}^p$ be a sequence in the unit sphere $\mathbb{S}^p = \{x \in \mathbb{R}^p : \|x\| = 1\}$. Then

$$\begin{aligned} &t^{-1}\big[H(\lambda_0 + th_n) - H(\lambda_0)\big] - \dot{H}_{\lambda_0} h_n \\ &\quad = t^{-1}\big[\lambda_0 + th_n + K(\lambda_0 + th_n) - \lambda_0 - K(\lambda_0)\big] - h_n - J_K(\lambda_0) h_n \\ &\quad = t^{-1}\big[K(\lambda_0 + th_n) - K(\lambda_0)\big] - J_K(\lambda_0) h_n \\ &\quad = \big(J_K(\bar{\lambda}_n) - J_K(\lambda_0)\big) h_n, \end{aligned}$$

where $\bar{\lambda}_n$ is a mean value between $\lambda_0 + th_n$ and λ_0. Therefore, by the Cauchy–Schwarz inequality,

$$\left\| \left(J_K(\bar{\lambda}_n) - J_K(\lambda_0)\right)h_n \right\| \leq \left\| J_K(\bar{\lambda}_n) - J_K(\lambda_0) \right\| \|h_n\| \to 0,$$

where we used $\|h_n\| = 1$, $\bar{\lambda}_n \to \lambda_0$, and the continuity of the Jacobian.

For the second claim, note that

$$\|\dot{H}_{\lambda_0} h\| = \left\| \left(I - J_K(\lambda_0)\right)h \right\|,$$

and $h \mapsto \|(I - J_K(\lambda_0))h\|$ is continuous. Since the domain of h is compact, there is $h^* \in \mathbb{S}^p$ such that $\inf_{\|h\|=1} \|\dot{H}_{\lambda_0} h\| = \|(I - J_K(\lambda_0))h^*\|$. Let $q = (I - J_K(\lambda_0))h^*$ and note that $I - J_K(\lambda_0)$ is linearly independent (due to $\det(I - J_K(\lambda_0)) \neq 0$), and hence $q \neq 0$. Hence $\inf_{\|h\|=1} \|\dot{H}_{\lambda_0} h\| = \|q\| > 0$. Hence, the second conclusion follows. $\square$

The following result is a slight extension of Lemma E.1 in CFM.

LEMMA 8. *Suppose that $\Lambda \subset \mathbb{R}^p$ and $\mathcal{U}$ is a compact and convex set in $\mathbb{R}^q$. Let $\mathcal{I}$ be an open set containing $\mathcal{U}$. Suppose that* (a) *$\Psi : \Lambda \times \mathcal{I} \to \mathbb{R}^p$ is continuous and $\lambda \mapsto \Psi(\lambda, u)$ is the gradient of a convex function in λ for each $u \in \mathcal{U}$*; (b) *for each $u \in \mathcal{U}$, $\Psi(\lambda_0(u), u) = 0$*; (c) *$\frac{\partial}{\partial(\lambda', u')}\Psi(\lambda, u)$ exists at $(\lambda_0(u), u)$ and is continuous at $(\lambda_0(u), u)$ for each $u \in \mathcal{U}$ and $\dot{\Psi}_{\lambda_0(u),u} := \frac{\partial}{\partial \lambda'}\Psi(\lambda, u)|_{\lambda_0(u)}$ obeys $\inf_{u\in\mathcal{U}} \inf_{\|h\|=1} \|\dot{\Psi}_{\lambda_0(u),u} h\| > c_0 > 0$. Then Condition Z in CFM holds and $u \mapsto \lambda_0(u)$ is continuously differentiable with derivative $J_{\lambda_0}(u) = -\dot{\Psi}^{-1}_{\lambda_0(u)u} \frac{\partial}{\partial u'}\Psi(\lambda_0(u), u)$.*

PROOF. The proof is the same as that of Lemma E.1 in CFM, in which $\mathcal{U}$ is a compact interval in $\mathbb{R}$. A slight modification is needed when one computes the derivative of $\lambda_0(u)$ with respect to u. Since u is allowed to be multidimensional, the implicit function theorem gives

$$J_{\lambda_0}(u) = -\dot{\Psi}^{-1}_{\lambda_0(u)u} \frac{\partial}{\partial u'} \Psi\big(\lambda_0(u), u\big),$$

which is uniformly bounded and continuous in u by condition (c), which ensures continuous differentiability of $u \mapsto \lambda_0(u)$. Note that for any $\delta > 0$ and $\lambda \in B_\delta(\lambda_0(u))$, there is $\eta > 0$ and u' such that $\|u' - u\| \leq \eta$ so that

$$\left\| \lambda - \lambda_0\big(u'\big) \right\| \leq \left\| \lambda - \lambda_0(u) \right\| + \left\| \lambda_0(u) - \lambda_0\big(u'\big) \right\| \leq 2\delta.$$

Since $\mathcal{U}$ is compact (and hence totally bounded), there is a finite set $\{u_j\}_{j=1}^J \subset \mathcal{U}$ such that $\mathcal{U} \subset \bigcup_j B_\eta(u_j)$. The argument above then shows that $\mathcal{N} = \bigcup_{u\in\mathcal{U}} B_\delta(\lambda_0(u)) \subset \bigcup_j B_{2\delta}(\lambda_0(u_j))$, which ensures that $\mathcal{N}$ is totally bounded. Since $\mathcal{N}$ is a subset of a Euclidean space (equipped with a complete metric), it follows that $\mathcal{N}$ is compact. This ensures condition Z (i) in CFM. The rest of the proof is essentially the same as the case in which $\mathcal{U}$ being a compact interval. $\square$

Lemma 9. *Suppose Assumption* 2 *holds. Let* $w=(y,d',x',z')$ *and let* $\tau\in(0,1)$. *Define*

$$\mathcal{M} := \big\{f : f(w;\theta) = \big(\big(1\big\{y \le d'\theta_{-1} + x'\theta_1\big\} - \tau\big)x,\\ \big(1\big\{y \le d'\theta_{-1} + x'\theta_1\big\} - \tau\big)z_1, \dots, \big(1\big\{u \le d'\theta_{-1} + x'\theta_1\big\} - \tau\big)z_{d_D}\big), \theta \in \Theta\big\}.$$

Then $\mathcal{M}$ *is a Donsker-class.*

Proof. The proof is standard, and hence we give a brief sketch for the first component of f, $f_1(w;\theta) = (1\{y \le d'\theta_{-1} + x'\theta_1\} - \tau)x$. Note that $w \mapsto 1\{y \le d'\theta_{-1} + x'\theta_1\} - \tau$ belongs to Type I-class in Andrews (1994), and the map $w \mapsto x$ does not depend on the parameter. By Theorems 2 and 3 in Andrews (1994), this function then satisfies the uniform entropy condition with the envelope function $\bar{M}(w) = x$, which is square integrable by assumption. Similar arguments apply to the other components of f. By Theorem 1 in Andrews (1994), the empirical process: $\mathbb{G}_n f$ is stochastically equicontinuous, and $\mathbb{G}_n f(\cdot,\theta)$ obeys the classical central limit theorem for each $\theta\in\Theta$. Hence, we conclude that $\mathcal{M}$ is Donsker. □

Below, let $g(w;\theta) = (g_1(w;\theta)', \dots, g_J(w;\theta))'$ be a vector such that

$$g_j(w;\theta) = \left(\frac{\partial^2}{\partial\theta_j\partial\theta_j'} Q_{P,j}\big(L_j(\theta_{-j}),\theta_{-j}\big)\right)^{-1} f_j\big(w; L_j(\theta_{-j}),\theta_{-j}\big), \quad j=1,\dots,J.$$

Let $\rho(\theta,\tilde{\theta}) := \|\operatorname{diag}(E_P[(g(W;\theta) - E_P[g(W;\theta)])(g(W;\tilde{\theta}) - E_P[g(W;\tilde{\theta})])'])\|$ be the variance semimetric. Let $W_i = (Y_i, D_i', X_i', Z_i')$, $i=1,\dots,N$ be an i.i.d. sample generated from the IVQR model. Define

$$\mathcal{L}_{N,j}(\theta_{-j}) := \sqrt{N}\big(\hat{L}_j(\theta_{-j}) - L_j(\theta_{-j})\big), \quad j=1,\dots,J. \tag{G.12}$$

Similarly, let $W_i^* = (Y_i^*, D_i^{*\prime}, X_i^{*\prime}, Z_i^{*\prime})'$, $i=1,\dots,N$ be a bootstrap sample from the empirical distribution P_N of $\{W_i\}$. Define

$$\mathcal{L}^*_{N,j}(\theta_{-j}) := \sqrt{N}\big(\hat{L}^*_j(\theta_{-j}) - \hat{L}_j(\theta_{-j})\big), \quad j=1,\dots,J,$$

where $\hat{L}_j^*$ is the sample best response map of player j, which is defined as in (5.3)–(5.4) while replacing W_i with the bootstrap sample W_i^* in (5.1)–(5.2).

Lemma 10 below shows that the sample BR functions approximately solve sample estimating equations and Lemma 11 characterizes the limiting distributions of $\mathcal{L}_N$ and $\mathcal{L}_N^*$.

Lemma 10. *Let the sample BR functions be* $\hat{L}_j(\theta_{-j}) \in \operatorname{argmin}_{\tilde{\theta}_j} Q_{N,j}(\tilde{\theta}_j,\theta_{-j})$, $j=1,\dots,J$. *Let* $\hat{L}_j^*(\theta_{-j})$ *be an analog of* $\hat{L}_j(\theta_{-j})$ *for the bootstrap sample. Then* (i) *the sample BR* *functions satisfy*

$$\begin{aligned}&\left|\frac{1}{N}\sum_{i=1}^{N}\bigl(1\{Y_i \le D_i'\theta_{-1} + X_i'\hat{L}_1(\theta_{-1})\} - \tau\bigr)X_i\right|^2 \\ &\quad \le \inf_{\theta_1\in\Theta_1}\left|\frac{1}{N}\sum_{i=1}^{N}\bigl(1\{Y_i \le D_i'\theta_{-1} + X_i'\theta_1\} - \tau\bigr)X_i\right|^2 + r_{N,1}^2(\theta_{-1}), \qquad \text{(G.13)}\end{aligned}$$

and

$$\begin{aligned}&\left|\frac{1}{N}\sum_{i=1}^{N}\bigl(1\{Y_i \le (X_i', D_{i,-(j-1)}')'\theta_{-j} + D_{i,j}'\hat{L}_j(\theta_{-j})\} - \tau\bigr)Z_{i,j}\right|^2 \\ &\quad \le \inf_{\theta_j\in\Theta_j}\left|\frac{1}{N}\sum_{i=1}^{N}\bigl(1\{Y_i \le (X_i', D_{i,-(j-1)}')'\theta_{-j} + D_{i,j}'\hat{L}_j(\theta_{-j})\} - \tau\bigr)Z_{i,j}\right|^2 \\ &\qquad + r_{N,j}^2(\theta_{-j}), \quad j = 2, \ldots, J, \qquad \text{(G.14)}\end{aligned}$$

where $\sup_{\theta_{-j}\in\Theta_{-j}} |r_{N,j}(\theta_{-j})| = o_P(N^{-1/2})$ *for all* j; (ii) *the sample BR functions* $\hat{L}_j^*(\theta_{-j})$, $j = 1, \ldots, J$ *in the bootstrap sample satisfy* (G.13)–(G.14) *while replacing* (Y_i, D_i, X_i, Z_i) *with a bootstrap sample* $(Y_i^*, D_i^*, X_i^*, Z_i^*)$, *each* $\hat{L}_j$ *with* $\hat{L}_j^*$, *and each* $r_{N,j}$ *with* $r_{N,j}^*$ *such that* $\sup_{\theta_{-j}\in\Theta_{-j}} |r_{N,j}^*(\theta_{-j})| = o_{P^*}(N^{-1/2})$.

Proof. (i) For $j \ge 2$, the subgradient of $Q_{N,j}$ is

$$\xi_j = \frac{1}{N}\sum_{i=1}^{N}\bigl(1\{Y_i \le (X_i', D_{i,-(j-1)}')'\theta_{-j} + D_{i,j-1}'\hat{L}_j(\theta_{-j})\} - \tau\bigr)Z_{i,j-1},$$

and hence by the property of the subgradient, for any $v \in \mathbb{R}$, one has

$$\xi_j v \le \nabla_{\theta_j} Q_{N,j}\bigl(\hat{L}_j(\theta_{-j}), \theta_{-j}, v\bigr),$$

where $\nabla_{\theta_j} Q_{N,j}(\hat{L}_j(\theta_{-j}), \theta_{-j}, v)$ is the directional derivative of $Q_{N,j}(\theta_j, \theta_{-j})$ with respect to θ_j toward direction $v \in \mathbb{R}$ evaluated at $(\hat{L}_j(\theta_{-j}), \theta_{-j})$. Note that the directional derivative is given by

$$\begin{aligned}&\nabla_{\theta_j} Q_{N,j}\bigl(\hat{L}_j(\theta_{-j}), \theta_{-j}, v\bigr) \\ &\quad = -\frac{1}{N}\sum_{i=1}^{N}\psi_\tau^*\bigl(Y_i - (X_i', D_{i,-(j-1)}')'\theta_{-j} - D_{i,j-1}'\hat{L}_j(\theta_{-j}), -Z_{i,j-1}v\bigr)Z_{i,j-1}v,\end{aligned}$$

where

$$\psi_\tau^*(u, w) = \begin{cases}\tau - 1\{u < 0\}, & u \ne 0, \\ \tau - 1\{w < 0\}, & u = 0.\end{cases}$$

Observe that $-\nabla_{\theta_j} Q_{N,j}(\hat{L}_j(\theta_{-j}), \theta_{-j}, -v) \le \xi v \le \nabla_{\theta_j} Q_{N,j}(\hat{L}_j(\theta_{-j}), \theta_{-j}, v)$. This implies

$$
\begin{aligned}
|\xi_j v| &\le \nabla_{\theta_j} Q_{N,j}\big(\hat{L}_j(\theta_{-j}), \theta_{-j}, v\big) - \big(-\nabla_{\theta_j} Q_{N,j}\big(\hat{L}_j(\theta_{-j}), \theta_{-j}, -v\big)\big) \\
&= \frac{1}{N}\sum_{i=1}^{N}\big(-\psi_\tau^*\big(Y_i - \big(X_i', D_{i,-(j-1)}'\big)'\theta_{-j} - D_{i,j}'\hat{L}_j(\theta_{-j}), -Z_{i,j-1}v\big) \\
&\quad + \psi_\tau^*\big(Y_i - \big(X_i', D_{i,-(j-1)}'\big)'\theta_{-j} - D_{i,j}'\hat{L}_j(\theta_{-j}), Z_{i,j-1}v\big)\big)Z_{i,j-1}v \\
&= \frac{1}{N}\sum_{i=1}^{N} 1\big\{Y_i = \big(X_i', D_{i,-(j-1)}'\big)'\theta_{-j} + D_{i,j}'\hat{L}_j(\theta_{-j})\big\}\mathrm{sgn}(Z_{i,j-1}v)Z_{i,j-1}v \\
&= \frac{1}{N}\sum_{i=1}^{N} 1\big\{Y_i = \big(X_i', D_{i,-(j-1)}'\big)'\theta_{-j} + D_{i,j}'\hat{L}_j(\theta_{-j})\big\}|Z_{i,j-1}v| \\
&\le \left(\sum_{i=1}^{N} 1\big\{Y_i = \big(X_i', D_{i,-(j-1)}'\big)'\theta_{-j} + D_{i,j}'\hat{L}_j(\theta_{-j})\big\}\right)\max_{i=1,\dots,N}\frac{|Z_{i,j-1}v|}{N}.
\end{aligned}
$$

Noting that $\sum_{i=1}^{N} 1\{Y_i = (X_i', D_{i,-(j-1)}')'\theta_{-j} + D_{i,j}'\hat{L}_j(\theta_{-j})\} = \dim(\theta_j) = 1$ and taking $v = 1$, we obtain

$$
\left|\frac{1}{N}\sum_{i=1}^{N}\big(1\big\{Y_i \le \big(X_i', D_{i,-(j-1)}'\big)'\theta_{-j} + D_{i,j}'\hat{L}_j(\theta_{-j})\big\} - \tau\big)Z_{i,j}\right| \le \max_{i=1,\dots,N}\frac{|Z_{i,j-1}|}{N} = o_P\big(N^{-1/2}\big),
$$

uniformly in θ_{-j}, where the last equality is due to $E[|Z_{i,j-1}|^2] < \infty$ by Assumption 2(2). Therefore, for some $r_{N,j}$ satisfying the assumption of the lemma, we may write

$$
\begin{aligned}
&\left|\frac{1}{N}\sum_{i=1}^{N}\big(1\big\{Y_i \le \big(X_i', D_{i,-(j-1)}'\big)'\theta_{-j} + D_{i,j}'\hat{L}_j(\theta_{-j})\big\} - \tau\big)Z_{i,j}\right|^2 \\
&\quad \le r_{N,j}^2(\theta_{-j}) \le \inf_{\theta_j \in \Theta_j}\left|\frac{1}{N}\sum_{i=1}^{N}\big(1\big\{Y_i \le \big(X_i', D_{i,-(j-1)}'\big)'\theta_{-j} + D_{i,j}'\theta_j\big\} - \tau\big)Z_{i,j}\right|^2 + r_{N,j}^2(\theta_{-j}).
\end{aligned}
$$

The proof for $j = 1$ is similar. Also, (ii) can be shown by mimicking the argument above. □

LEMMA 11. *Suppose that Assumptions* 1 *and* 2 *hold. Then* (i) $\mathcal{L}_N := (\mathcal{L}_{N,1}, \dots, \mathcal{L}_{N,J})$ *defined in* (G.12) *satisfies*

$$
\mathcal{L}_N(\cdot) \rightsquigarrow \mathbb{W},
$$

where $\mathbb{W}$ *is a tight Gaussian process in* $\ell^\infty(\Theta)^{d_X + d_D}$ *with the covariance kernel*

$$
\mathrm{Cov}\big(\mathbb{W}(\theta), \mathbb{W}(\tilde{\theta})\big) = E_P\big[\big(g(W;\theta) - E_P\big[g(W;\theta)\big]\big)\big(g(W;\tilde{\theta}) - E_P\big[g(W;\tilde{\theta})\big]\big)'\big]; \tag{G.15}
$$

$\mathcal{L}_N$ is stochastically equicontinuous with respect to the variance semimetric ρ; (ii) $\mathcal{L}_N^* := (\mathcal{L}_{N,1}^*, \dots, \mathcal{L}_{N,J}^*)$ *satisfies*

$$\mathcal{L}_N^*(\cdot) \overset{L^*}{\rightsquigarrow} \mathbb{W};$$

(iii) ρ *satisfies* $\lim_{\delta\downarrow 0}\sup_{\|\theta-\tilde\theta\|<\delta}\rho(\theta,\tilde\theta)\to 0$.

Proof. (i) We first work with $\mathcal{L}_{N,1}$. For this, we establish that L_1 is Hadamard differentiable. Note that $\theta_1 = L_1(\theta_{-1})$ solves

$$E_P\big[\big(1\big\{Y \le D'\theta_{-1} + X'\theta_1\big\} - \tau\big)X\big] = 0.$$

Take $\mathcal{U} = \Theta_{-1}$, $\Xi = \Theta_1$, $\psi(\lambda, u) = E_P[(1\{Y \le D'u + X'\lambda\} - \tau)X]$. Define $\phi : \ell^\infty(\Xi\times\mathcal{U})^{k_b} \times \ell^\infty(\mathcal{U}) \to \ell^\infty(\mathcal{U})$, which maps (ψ, r) to a solution $\phi(\psi, r) = \lambda(\cdot)$ such that

$$\big\|\psi\big(\lambda(u), u\big)\big\|^2 \le \inf_{\lambda'\in\Theta}\big\|\psi\big(\lambda', u\big)\big\|^2 + r(u)^2. \tag{G.16}$$

Then one may write $L_1(\cdot) = \phi(\psi, 0)$. We then show that ψ satisfies the conditions of Lemma 8. Note first that $\mathcal{U}$ and Ξ are compact. ψ is continuous and $\lambda \mapsto \psi(\lambda, u)$ is the gradient of the convex function $\lambda \mapsto E_P[\rho_\tau(Y - D'u - X'\lambda)]$. The function $L_1(u) = \lambda_0(u)$ is defined as the exact solution of $\psi(\lambda, u) = 0$. Note also that, by Assumption 2,

$$\begin{aligned}
\frac{\partial^2}{\partial\theta_1\partial\theta_1'}Q_{P,1}(\theta_1,\theta_{-1}) &= \frac{\partial}{\partial\theta_1'}E_P\big[\big(1\big\{Y\le D'\theta_{-1}+X'\theta_1\big\}-\tau\big)X\big] \\
&= E_P\left[\frac{\partial}{\partial\theta_1'}\big(F_{Y|D,X,Z}\big(D'\theta_{-1}+X'\theta_1\big)-\tau\big)X\right] \\
&= E_P\big[f_{Y|D,X,Z}\big(D'\theta_{-1}+X'\theta_1\big)XX'\big],
\end{aligned}$$

where the second equality follows from the dominated convergence theorem, and the last display is well-defined by the square integrability of X. Similarly,

$$\frac{\partial^2}{\partial\theta_1\partial\theta_{-1}'}Q_{P,1}(\theta_1,\theta_{-1}) = E_P\big[f_{Y|D,X,Z}\big(D'\theta_{-1}+X'\theta_1\big)XD'\big].$$

Hence, the derivative

$$\frac{\partial}{\partial\big(\lambda',u'\big)}\Psi(\lambda,u) = \left(\frac{\partial^2}{\partial\theta_1\partial\theta_1'}Q_{P,1}(\theta_1,\theta_{-1}), \frac{\partial^2}{\partial\theta_1\partial\theta_{-1}'}Q_{P,1}(\theta_1,\theta_{-1})\right)$$

exists and is continuous by Assumption 2. By Assumption 2(4), $\dot\Psi_{\lambda_0(u),u} = \frac{\partial^2}{\partial\theta_1\partial\theta_1'} \times Q_{P,1}(L_1(\theta_{-1}),\theta_{-1})$ obeys

$$\inf_{u\in\mathcal{U}}\inf_{\|h\|=1}\|\dot\Psi_{\lambda_0(u),u}h\| = \inf_{\theta_{-1}\in\Theta_{-1}}\inf_{\|h\|=1}\big\|E_P\big[f_{Y|D,X,Z}\big(D'\theta_{-1}+X'\theta_1\big)XX'\big]h\big\| > 0.$$

Then, by Lemma 8 and Lemma E.2 in CFM, ϕ is Hadamard differentiable tangentially to $\mathcal{C}(\mathcal{N}\times\mathcal{U})^K\times\{0\}$ with the Hadamard derivative (of L_1)

$$\phi'_{\Psi,0}(z,0)=-\left(\frac{\partial^2}{\partial\theta_1\partial\theta_1'}Q_{P,1}\big(L_1(\cdot),\cdot\big)\right)^{-1}z\big(L_1(\cdot),\cdot\big),$$

where $(z,0)\mapsto\phi'_{\Psi,0}(z,0)$ is continuous over $z\in\ell^\infty(\Theta)^K$.

For $j\geq 2$, the argument is similar. For example, for $j=2$, one may take $\mathcal{U}=\Theta_{-2}$, $\Xi=\Theta_2$ and $\psi(\lambda,u)=E_P[(1\{Y\leq D_1\theta_2+(D_{-1},X)'u\}-\tau)Z_1]$ and write $L_2(\cdot)=\phi(\psi,0)$. The rest of the argument is the same.

Continuing with $j=1$, by Lemma 10, one may write $\hat{L}_j(\cdot)=\phi(\psi_N,r_{N,1})$ with $\psi_N(\lambda,u)=\frac{1}{N}\sum_{i=1}^N 1\{Y_i\leq D_i'u+X_i'\lambda\}X_i$ and $\sup_{\theta_{-1}\in\Theta_{-1}}|r_{N,1}(\theta_{-1})|=o_p(N^{-1/2})$. By Lemma 9 and applying the δ-method (as in Lemma E.3 in CFM), we obtain

$$\mathcal{L}_N(\cdot)\rightsquigarrow\mathbb{W},$$

where $\mathbb{W}=(\mathbb{W}_1',\dots,\mathbb{W}_J')'$ is a tight Gaussian process in $\ell^\infty(\Theta)^{d_X+d_D}$, where for each j, $\mathbb{W}_j\in\ell^\infty(\Theta_{-j})^{d_j}$ is given pointwise by

$$\mathbb{W}_j(\theta_{-j})=-\left(\frac{\partial^2}{\partial\theta_j\partial\theta_j'}Q_{P,j}\big(L_j(\theta_{-j}),\theta_{-j}\big)\right)^{-1}\mathbb{G}f_j\big(w;L_j(\theta_{-j}),\theta_{-j}\big),\quad j=1,\dots,J;$$

Hence, its covariance kernel is as given in (G.15). By Lemma 1.3.8. in Van der Vaart and Wellner (1996), $\{\mathcal{L}_N\}$ is asymptotically tight, which in turn means that $\{\mathcal{L}_N\}$ is stochastically equicontinuous with respect to ρ by Theorem 1.5.7 in Van der Vaart and Wellner (1996).

(ii) For each j, let $\mathcal{L}^*_{N,j}\in\ell^\infty(\Theta_{-j})^{d_j}$ be defined pointwise by

$$\mathcal{L}^*_{N,j}(\theta_{-j})=\sqrt{N}\big(\hat{L}^*_j(\theta_{-j})-\hat{L}_j(\theta_{-j})\big).$$

Below, again we work with the case $j=1$. Using ϕ (the solution to (G.16)) and applying Lemma 10, we may write

$$\mathcal{L}^*_{N,1}(\theta_{-1})=\sqrt{N}\big(\phi\big(\hat{\psi}^*_N,r^*_N\big)-\phi(\hat{\psi}_N,r_N)\big),$$

where $\hat{\psi}_N(\lambda,u)=N^{-1}\sum_{i=1}^N(1\{Y_i\leq D_iu+X_i'\lambda\}-\tau)X_i$, and $\hat{\psi}^*_N$ is defined similarly for the bootstrap sample. Let E_{P^*} denote the conditional expectation with respect to P^*, the law of $\{W_i^*\}_{i=1}^N$ conditional on the sample path. Let BL_1 denote the space of bounded Lipschitz functions on $\mathbb{R}^{d_X}$ with Lipschitz constant 1. Then, for any $\epsilon>0$,

$$\begin{aligned}\sup_{h\in BL_1}&\big|E_{P^*}h\big(\sqrt{N}\big[\phi\big(\hat{\psi}^*_N,r^*_N\big)-\phi(\hat{\psi}_N,r_N)\big]\big)-E_{P^*}h\big(\phi'_{\Psi,0}\big(\sqrt{N}\big[\big(\hat{\psi}^*_N,r^*_N\big)-(\hat{\psi}_N,r_N)\big]\big)\big)\big|\\&\leq\epsilon+2P^*\big(\big\|\sqrt{N}\big[\phi\big(\hat{\psi}^*_N,r^*_N\big)-\phi(\hat{\psi}_N,r_N)\big]\\&\qquad-\phi'_{\Psi,0}\big(\sqrt{N}\big[\big(\hat{\psi}^*_N,r^*_N\big)-(\hat{\psi}_N,r_N)\big]\big)\big\|>\epsilon\big).\end{aligned}\tag{G.17}$$

By Lemma 9 and Theorem 3.6.2 in Van der Vaart and Wellner (1996), $\sqrt{N}(\hat{\psi}_N^* - \hat{\psi}_N) \overset{L^*}{\rightsquigarrow} \mathbb{G}f_1$. Noting that $h \circ \phi'_{\Psi,0} \in BL_1(\ell^\infty(\Theta_{-1})^{d_X} \times \mathbb{R})$ and $r_N = o_p(N^{-1/2})$, it follows that

$$\sup_{h \in BL_1} \left| E_{P^*} h\left(\phi'_{\Psi,0}\left(\sqrt{N}\left[\left(\hat{\psi}_N^*, r_N^*\right) - (\hat{\psi}_N, r_N)\right]\right)\right) - E_{P^*} h \circ \phi'_{\Psi,0}(\mathbb{G}f_1, 0)\right| \to 0, \tag{G.18}$$

with probability approaching 1 due to $r_N = o_P(N^{-1/2})$. Hence, for the conclusion of the theorem, it suffices to show that the second term on the right-hand side of (G.17) tends to 0.

As shown in the proof of (i), ϕ is Hadamard differentiable at $(\psi, 0)$. Hence, by Theorem 3.9.4 in Van der Vaart and Wellner (1996),

$$\begin{aligned}
\sqrt{N}\left[\phi\left(\hat{\psi}_N^*, r_N^*\right) - \phi(\psi, 0)\right] &= \phi'_{\Psi,0}\left(\sqrt{N}\left[\left(\hat{\psi}_N^*, r_N^*\right) - (\psi, 0)\right]\right) + o_{P^*}(1), \\
\sqrt{N}\left[\phi(\hat{\psi}_N, r_N) - \phi(\psi, 0)\right] &= \phi'_{\Psi,0}\left(\sqrt{N}\left[(\hat{\psi}_N, r_N) - (\psi, 0)\right]\right) + o_P(1).
\end{aligned}$$

Take the difference of the left- and right-hand sides, respectively, and note that $\phi'_{\Psi,0}$ is linear. This implies the right-hand side of (G.17) tends to 0 in probability. This, together with (G.17)–(G.18), ensures

$$\mathcal{L}_{N,1}^* \overset{L^*}{\rightsquigarrow} \mathbb{W}_1,$$

where $\mathbb{W}_1(\theta_{-1}) = -\frac{\partial^2}{\partial\theta_1\partial\theta_1'} Q_{P,1}(L_1(\theta_{-1}), \theta_{-1})^{-1} \mathbb{G}f_j(\cdot; L_1(\theta_{-1}), \theta_{-1})$. The analysis for any $j \neq 1$ is similar, and one may apply the arguments above jointly across $j = 1, \ldots, J$, which yields the second claim of the lemma.

(iii) Consider the first submatrix of $E_P[(g(W;\theta) - E_P[g(W;\theta)])(g(W;\tilde{\theta}) - E_P[g(W;\tilde{\theta})])']$. It is given by

$$\begin{aligned}
&\mathrm{Var}\left(-\left(\frac{\partial^2}{\partial\theta_1\partial\theta_1'} Q_{P,1}\left(L_1(\theta_{-1}), \theta_{-1}\right)\right)^{-1} f_1\left(w; L_1(\theta_{-1}), \theta_{-1}\right)\right) \\
&\quad - \mathrm{Var}\left(-\left(\frac{\partial^2}{\partial\theta_1\partial\theta_1'} Q_{P,1}\left(L_1(\tilde{\theta}_{-1}), \tilde{\theta}_{-1}\right)\right)^{-1} f_1\left(w; L_1(\tilde{\theta}_{-1}), \tilde{\theta}_{-1}\right)\right) \\
&= \left(\frac{\partial^2}{\partial\theta_1\partial\theta_1'} Q_{P,1}\left(L_1(\theta_{-1}), \theta_{-1}\right)\right)^{-1} \mathrm{Var}\left(f_1\left(w; L_1(\theta_{-1}), \theta_{-1}\right)\right) \\
&\quad \times \left(\frac{\partial^2}{\partial\theta_1\partial\theta_1'} Q_{P,1}\left(L_1(\theta_{-1}), \theta_{-1}\right)\right)^{-1} \\
&\quad - \left(\frac{\partial^2}{\partial\theta_1\partial\theta_1'} Q_{P,1}\left(L_1(\tilde{\theta}_{-1}), \tilde{\theta}_{-1}\right)\right)^{-1} \mathrm{Var}\left(f_1\left(w; L_1(\tilde{\theta}_{-1}), \tilde{\theta}_{-1}\right)\right) \\
&\quad \times \left(\frac{\partial^2}{\partial\theta_1\partial\theta_1'} Q_{P,1}\left(L_1(\tilde{\theta}_{-1}), \tilde{\theta}_{-1}\right)\right)^{-1}.
\end{aligned}$$

Note that Θ is compact and $\theta_{-1} \mapsto (\frac{\partial^2}{\partial\theta_1\partial\theta_1'} Q_{P,1}(L_1(\theta_{-1}), \theta_{-1}))^{-1}$ is continuous by Lemma 1, which implies that this map is uniformly continuous. Therefore, it remains to show the uniform continuity of $\theta \mapsto \mathrm{Var}(f_1(w;\theta))$. Note that

$$\begin{aligned}
&\mathrm{Var}\big(f_1\big(w; L_1(\theta_{-1}), \theta_{-1}\big)\big) \\
&\quad = E_P\big[\big(1\big\{Y \le D'\theta_{-1} + X'L_1(\theta_{-1})\big\} - \tau\big)XX'\big] \\
&\qquad - E_P\big[\big(1\big\{Y \le D'\theta_{-1} + X'L_1(\theta_{-1})\big\} - \tau\big)X\big]E_P\big[\big(1\big\{Y \le D'\theta_{-1} + X'L_1(\theta_{-1})\big\} - \tau\big)X\big]'.
\end{aligned}$$

The right-hand side of the display above is continuous on the compact domain Θ, and hence it is uniformly continuous. One can argue the same way for the other subcomponents of $\mathrm{diag}(E_P[(g(W;\theta) - E_P[g(W;\theta)])(g(W;\tilde{\theta}) - E_P[g(W;\tilde{\theta})])'])$. This completes the proof. □

Lemma 12. *Suppose that Assumptions* 1 *and* 2 *hold.* (i) *Let* $W_i = (Y_i, D_i', X_i', Z_i')'$, $i = 1, \ldots, N$ *be an i.i.d. sample generated from the IVQR model. Then*

$$\sqrt{N}(\hat{K} - K) \rightsquigarrow \mathbb{W}.$$

(ii) *Let* $W_i^* = (Y_i^*, D_i^{*\prime}, X_i^{*\prime}, Z_i^{*\prime})'$, $i = 1, \ldots, N$ *be an bootstrap sample from the empirical distribution* P_N *of* $\{W_i\}_{i=1}^N$. *Then*

$$\sqrt{N}\big(\hat{K}^* - \hat{K}\big) \overset{L^*}{\rightsquigarrow} \mathbb{W}.$$

Proof. (i) By Lemma 11, it follows that

$$\sqrt{N}\big(\hat{L}_1(\cdot) - L_1(\cdot), \ldots, \hat{L}_J(\cdot) - L_J(\cdot)\big)' \rightsquigarrow \mathbb{W}.$$

Note that, by the definition of $\hat{L}$ and L, one has

$$\sqrt{N}\big(\hat{K}_j(\theta) - K_j(\theta)\big) = \sqrt{N}\big(\hat{L}_j(\theta_{-j}) - L_j(\theta_{-j})\big), \quad j = 1, \ldots, J.$$

The conclusion of the lemma then follows. The proof of (ii) is similar, and is therefore omitted. □

Appendix H: Consistency of the contraction estimator

Below, we adopt the framework of Dominitz and Sherman (2005). Let $(\mathcal{X}, d)$ be a metric space. For a contraction map $F : \mathcal{X} \to \mathcal{X}$, let c_F be the modulus of contraction such that

$$d\big(F(x), F(x')\big) \le c_F d\big(x, x'\big),$$

for any $x, x' \in \mathcal{X}$. As discussed in Section 5 the fixed-point estimator $\hat{\theta}_N$ can be computed using the sample sequential dynamical system (in (5.5)) or the following sample simultaneous dynamical system:

$$\theta^{(s+1)} = \hat{K}\big(\theta^{(s)}\big), \quad s = 0, 1, 2, \ldots, \theta^{(0)} \text{ given.} \tag{H.1}$$

Lemma 13. *Suppose Assumptions* 1, 2, *and* 3 *hold. Let* $\hat{\theta}_N$ *be an estimator constructed by iterating the dynamical system in* (H.1) *or in* (5.5) s_N *times, where* $s_N \geq -\frac{1}{2}\ln N/\ln c_K$. *Then*

$$\hat{\theta}_N - \theta^* = O_p\big(N^{-1/2}\big).$$

Proof. We show the result by applying Theorem 1 in Dominitz and Sherman (2005) to the estimator obtained from the simultaneous dynamical system. The argument for the sequential system is similar.

By Assumption 3, K is a contraction map on D_K. Let $\theta^{(s)}$ be obtained from iterating s-times the population dynamical system in (3.7). The iteration on the dynamical system is covergent at least linearly (Bertsekas and Tsitsiklis (1989, Proposition 1.1)). Under the condition on s_N, arguing as in Dominitz and Sherman (2005, p. 842), it follows that $N^{1/2}\|\theta^{(s_N)} - \theta^*\| \leq \|\theta^{(0)} - \theta^*\|$. Finally, by Lemma 12 and tightness of $\mathbb{W}$, $N^{1/2}\sup_{\theta\in D_K}\|\hat{K}(\theta) - K(\theta)\| = O_p(1)$. These imply the conditions of Theorem 1 in Dominitz and Sherman (2005) with $\delta = 1/2$. The claim of the lemma then follows. □

References

Andrews, D. W. (1994), "Chapter 37 Empirical process methods in econometrics." In *Handbook of Econometrics*, Vol. 4 (R. F. Engle and D. L. McFadden, eds.), 2247–2294, Elsevier, North Holland, Amsterdam. [24]

Andrews, I. and A. Mikusheva (2016), "Conditional inference with a functional nuisance parameter." *Econometrica*, 84 (4), 1571–1612. [9]

Bertsekas, D. P. and J. N. Tsitsiklis (1989), *Parallel and Distributed Computation: Numerical Methods*, Vol. 23. Prentice hall, Englewood Cliffs, NJ. [31]

Bhatia, R. (1997), *Matrix Analysis*, Vol. 169. Springer Science & Business Media, New York, NY, Berlin, Heidelberg. [7]

Chamberlain, G. (1987), "Asymptotic efficiency in estimation with conditional moment restrictions." *Journal of Econometrics*, 34 (3), 305–334. [2]

Chernozhukov, V., I. Fernandez-Val, and B. Melly (2013), "Inference on counterfactual distributions." *Econometrica*, 81 (6), 2205–2268. [18]

Chernozhukov, V. and C. Hansen (2006), "Instrumental quantile regression inference for structural and treatment effects models." *Journal of Econometrics*, 132, 491–525. [1, 8, 21]

Dominitz, J. and R. P. Sherman (2005), "Some convergence theory for iterative estimation procedures with an application to semiparametric estimation." *Econometric Theory*, 21 (04), 838–863. [30, 31]

Gale, D. and H. Nikaido (1965), "The Jacobian matrix and global univalence of mappings." *Mathematische Annalen*, 159 (2), 81–93. [15]

Hasselblatt, B. and A. Katok (2003), *A First Course in Dynamics With a Panorama of Recent Developments*. Cambridge University Press, Cambridge. [6, 18]

Horn, R. A. and C. R. Johnson (2012), *Matrix Analysis*, second edition. Cambridge University Press, New York, NY. [17]

Imbens, G. W. and J. D. Angrist (1994), "Identification and estimation of local average treatment effects." *Econometrica*, 62 (2), 467–475. [6]

Krantz, S. G. and H. R. Parks (2003), *The Implicit Function Theorem: History, Theory, and Applications*. Springer Science & Business Media, New York, NY. [12]

Van der Vaart, A. and J. Wellner (1996), *Weak Convergence and Empirical Processes: With Application to Statistics*. Springer Science+Business Media, New York, NY. [19, 28, 29]

Co-editor Christopher Taber handled this manuscript.

Manuscript received 11 September, 2019; final version accepted 13 November, 2020; available online 21 December, 2020.