Econometrica Supplementary Material

SUPPLEMENT TO "COUNTERFACTUAL SENSITIVITY AND ROBUSTNESS"

(*Econometrica*, Vol. 91, No. 1, January 2023, 263–298)

TIMOTHY CHRISTENSEN
Department of Economics, New York University and Department of Economics, University College London

BENJAMIN CONNAULT
Latour Trading LLC

THIS SUPPLEMENT PRESENTS EXTENSIONS OF OUR METHODOLOGY in Appendix A, additional results on nonparametric bounds on counterfactuals in Appendix B, connections with local approaches to sensitivity analysis in Appendix C, additional details on the empirical applications in Appendix D, and proofs of results from the main text in Appendix E.

APPENDIX A: EXTENSIONS

This appendix presents three extensions of our methodology. Proofs of all results in this appendix are presented in Appendix G.7 of our working paper version Christensen and Connault (2022).

A.1. *Group Invariance*

In certain settings, it can be attractive to impose shape restrictions on F such as symmetry, exchangeability, or, more generally, invariance to a finite group of transformations. For instance, imposing exchangeability in discrete choice modeling ensures that alternatives' choice probabilities depend on their deterministic components of utility but not their labeling. These shape restrictions can be easily imposed whenever F_* is invariant.

Formally, let J denote the number of elements of U and let Π be a finite commutative group of transformations on $\mathbb{R}^J$—see, for example, Section 1.4 of Lehmann and Casella (1998). We say that a distribution F of U is Π-*invariant* if $\varpi U \sim F$ for all $\varpi \in \Pi$.

EXAMPLE A.1—Symmetry: Central symmetry corresponds to $\Pi = \{I, -I\}$ for I the identity matrix. Sign symmetry corresponds to taking Π to be the collection of all 2^J diagonal matrices with ± 1 in each diagonal entry.

EXAMPLE A.2—Exchangeability: Let Π_J denote the set of all $J!$ permutation matrices of dimension J. Full exchangeability (permutation invariance) corresponds to $\Pi = \Pi_J$. Cyclic exchangeability (rotation invariance) corresponds to $\Pi = \Pi_J^c$, where Π_J^c is the collection of all J cyclic permutation matrices of dimension J ($\Pi_J^c = \Pi_J$ when $J = 2$ and is a strict subset otherwise). When $J \geq 3$, dihedral exchangeability (rotation and reflection invariance) corresponds to taking Π to be the set of all $2J$ permutation matrices representing rotations and reflections of $\{1, \dots, J\}$. These types of exchangeability ensure the elements of U are identically distributed, but they have different implications for the joint

Timothy Christensen: timothy.christensen@nyu.edu
Benjamin Connault: benjamin.connault@gmail.com

 https://doi.org/10.3982/ECTA17232

distribution of the elements of U. For instance, the distribution of (U_i, U_j) for $i \neq j$ depends on $i - j$ and $|i - j|$ under cyclic and dihedral exchangeability, but is independent of (i, j) under full exchangeability.

Let $\mathcal{N}_\delta^\Pi = \{F \in \mathcal{N}_\delta : F \text{ is } \Pi\text{-invariant}\}$. We are interested in

$$\underline{\kappa}_\delta^\Pi := \inf_{\theta \in \Theta, F \in \mathcal{N}_\delta^\Pi} \mathbb{E}^F\big[k(U, \theta, \gamma_0)\big] \quad \text{subject to (1a)–(1d),} \tag{32}$$

and $\overline{\kappa}_\delta^\Pi$ defined as the analogous supremum. One may write $\underline{\kappa}_\delta^\Pi$ and $\overline{\kappa}_\delta^\Pi$ as the value of two optimization problems in which criterion functions $\underline{K}_\delta^\Pi(\theta; \gamma_0, P_0)$ and $\overline{K}_\delta^\Pi(\theta; \gamma_0, P_0)$ are optimized with respect to θ. For a generic (θ, γ, P), define

$$\underline{K}_\delta^\Pi(\theta; \gamma, P) = \inf_{F \in \mathcal{N}_\delta^\Pi} \mathbb{E}^F\big[k(U, \theta, \gamma)\big] \quad \text{subject to (1a)–(1d) holding at } (\theta, \gamma, P), \tag{33}$$

and define $\overline{K}_\delta^\Pi(\theta; \gamma, P)$ as the analogous supremum. These criterions have dual representations as finite-dimensional convex programs when F_* is Π-invariant. Define

$$k^\Pi(U, \theta, \gamma) = \frac{1}{|\Pi|} \sum_{\varpi \in \Pi} k(\varpi U, \theta, \gamma),$$

$$g_j^\Pi(U, \theta, \gamma) = \frac{1}{|\Pi|} \sum_{\varpi \in \Pi} g_j(\varpi U, \theta, \gamma), \quad j = 1, 2, 3, 4,$$

where $|\Pi|$ denotes the cardinality of Π, and let $g^\Pi = (g_1^\Pi, g_2^\Pi, g_3^\Pi, g_4^\Pi)$.

PROPOSITION A.1: *Suppose that Assumption Φ holds and F_* is Π-invariant. Then*

$$\begin{aligned} \underline{K}_\delta^\Pi(\theta; \gamma, P) = \sup_{\eta > 0, \zeta \in \mathbb{R}, \lambda \in \Lambda} & -\eta \mathbb{E}^{F_*}\left[\phi^\star\left(\frac{k^\Pi(U, \theta, \gamma) + \zeta + \lambda' g^\Pi(U, \theta, \gamma)}{-\eta}\right)\right] \\ & - \eta\delta - \zeta - \lambda_{12}' P, \end{aligned} \tag{34}$$

$$\begin{aligned} \overline{K}_\delta^\Pi(\theta; \gamma, P) = \inf_{\eta > 0, \zeta \in \mathbb{R}, \lambda \in \Lambda} & \eta \mathbb{E}^{F_*}\left[\phi^\star\left(\frac{k^\Pi(U, \theta, \gamma) - \zeta - \lambda' g^\Pi(U, \theta, \gamma)}{\eta}\right)\right] \\ & + \eta\delta + \zeta + \lambda_{12}' P. \end{aligned} \tag{35}$$

Moreover, the value of problem (34) *is $+\infty$* (*equivalently, the value of problem* (35) *is $-\infty$*) *if and only if there is no distribution in $\mathcal{N}_\delta^\Pi$ under which* (1a)–(1d) *holds at* (θ, γ, P).

REMARK A.1: If F is Π-invariant and satisfies (1a)–(1d), then it must also satisfy (1a)–(1d) under all $|\Pi|$ transformations of the elements of U. Therefore, in effect there are a total of $|\Pi| \times d$ moment conditions imposed in the inner optimization, namely,

$$\begin{aligned} &\mathbb{E}^F\big[g_1(\varpi U, \theta, \gamma_0)\big] \le P_{10}, \qquad \mathbb{E}^F\big[g_2(\varpi U, \theta, \gamma_0)\big] = P_{20}, \\ &\mathbb{E}^F\big[g_3(\varpi U, \theta, \gamma_0)\big] \le 0, \qquad \mathbb{E}^F\big[g_4(\varpi U, \theta, \gamma_0)\big] = 0, \quad \text{for all } \varpi \in \Pi. \end{aligned} \tag{36}$$

In principle, one could form a criterion by including all $|\Pi| \times d$ moments. By Π-invariance of F_* and convexity of the objective, the multipliers on the moments $g(\varpi U, \theta, \gamma)$ will be identical across all $\varpi \in \Pi$. It therefore suffices to form the criterion using only the d averaged moments g^{Π} rather than the full set of $|\Pi| \times d$ moments, thereby reducing the dimension of the inner optimization by a factor of $|\Pi|$.

REMARK A.2: When Monte Carlo integration is used to compute expectations, taking a sample from F_* and then concatenating the sample across each of its $|\Pi|$ transformations ensures the empirical distribution of the random draws is Π-invariant.

A.2. *Conditional Moment Models*

Consider the conditional moment model

$$\begin{aligned} &\mathbb{E}^F\big[g_1(U, X, \theta, \gamma_0)|X = x\big] \le P_{10,x}, \qquad \mathbb{E}^F\big[g_2(U, X, \theta, \gamma_0)|X = x\big] = P_{20,x}, \\ &\mathbb{E}^F\big[g_3(U, X, \theta, \gamma_0)|X = x\big] \le 0, \qquad \mathbb{E}^F\big[g_4(U, X, \theta, \gamma_0)|X = x\big] = 0, \\ &\quad \text{for all } x \in \mathcal{X}, \end{aligned} \tag{37}$$

where $\mathcal{X}$ is a finite set, and a counterfactual[24]

$$\kappa = \sum_{x \in \mathcal{X}} \mathbb{E}^F\big[k(U, X, \theta, \gamma_0)|X = x\big]. \tag{38}$$

Suppose the researcher assumes $U|X = x \sim F_*$ for each x. We wish to relax this assumption and allow each conditional distribution of U given $X = x$, say F_x, to vary in a neighborhood $\mathcal{N}_{\delta_x}$ of F_*. In doing so, we are allowing the conditional distributions F_x to vary with x, and therefore relaxing independence of U and X.[25]

We assume each $\mathcal{N}_\delta$ is defined by the same ϕ to simplify the exposition, but we allow the neighborhood size to vary with x. Let $\boldsymbol{\delta} = (\delta_x)_{x \in \mathcal{X}}$. We are interested in

$$\underline{\kappa}_{\boldsymbol{\delta}} := \inf_{\theta \in \Theta, (F_x \in \mathcal{N}_{\delta_x})_{x \in \mathcal{X}}} \sum_{x \in \mathcal{X}} \mathbb{E}^{F_x}\big[k(U, x, \theta, \gamma_0)\big] \quad \text{subject to (37)}, \tag{39}$$

and $\overline{\kappa}_{\boldsymbol{\delta}}$ defined as the analogous supremum. One may write $\underline{\kappa}_{\boldsymbol{\delta}}$ and $\overline{\kappa}_{\boldsymbol{\delta}}$ as the value of two optimization problems where $\underline{K}_{\boldsymbol{\delta}}(\theta; \gamma_0, P_0)$ and $\overline{K}_{\boldsymbol{\delta}}(\theta; \gamma_0, P_0)$ are optimized with respect to θ. Let $P = (P_x)_{x \in \mathcal{X}}$ where $P_x = (P_{1,x}, P_{2,x})$ is partitioned conformably with g_1 and g_2. For a generic (θ, γ, P), define

$$\underline{K}_{\boldsymbol{\delta}}(\theta; \gamma, P) = \inf_{(F_x \in \mathcal{N}_{\delta_x})_{x \in \mathcal{X}}} \sum_{x \in \mathcal{X}} \mathbb{E}^{F_x}\big[k(U, x, \theta, \gamma_0)\big] \quad \text{subject to (37) holding at } (\theta, \gamma, P),$$

and define $\overline{K}_{\boldsymbol{\delta}}(\theta; \gamma, P)$ as the analogous supremum. These criterion functions have dual forms analogous to Proposition 2.1. Let $g(\cdot, x, \theta, \gamma) = (g_1(\cdot, x, \theta, \gamma), \dots, g_4(\cdot, x, \theta, \gamma))$. Recall $d = \sum_{i=1}^4 d_i$ where d_i is the dimension of g_i, and $\Lambda = \mathbb{R}_+^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}_+^{d_3} \times \mathbb{R}^{d_4}$. Let $\lambda_{12,x}$ denote the first $d_1 + d_2$ elements of $\lambda_x \in \Lambda$.

[24]Note κ can be the expected value at a particular x_0 if $k(U, x, \theta, \gamma_0) = 0$ for $x \neq x_0$. More generally, κ can be a weighted average by incorporating the weighting into the definition of $k(u, x, \theta, \gamma_0)$.

[25]The case with U independent of X is subsumed in (1a)–(1d) by stacking the moment functions and reduced-form parameters by values of the conditioning variable, as in Examples 2.1–2.3.

ASSUMPTION Φ-CONDITIONAL: (i) $\phi \in \Phi_0$.
(ii) *$k(\cdot, x, \theta, \gamma)$ and each entry of $g(\cdot, x, \theta, \gamma)$ belong to $\mathcal{E}$ for each $(\theta, \gamma, x) \in \Theta \times \Gamma \times \mathcal{X}$.*

PROPOSITION A.2: *Suppose that Assumption Φ-conditional holds. Then*

$$\begin{aligned}&\underline{K}_{\boldsymbol{\delta}}(\theta; \gamma, P)\\ &\quad= \sup_{(\eta_x>0,\zeta_x\in\mathbb{R},\lambda_x\in\Lambda)_{x\in\mathcal{X}}} \sum_{x\in\mathcal{X}}\left(-\eta_x \mathbb{E}^{F_*}\left[\phi^\star\left(\frac{k(U,x,\theta,\gamma)+\zeta_x+\lambda_x' g(U,x,\theta,\gamma)}{-\eta_x}\right)\right]\right.\\ &\qquad\left. -\eta_x\delta_x-\zeta_x-\lambda_{12,x}' P_x\right), \end{aligned} \tag{40}$$

$$\begin{aligned}&\overline{K}_{\boldsymbol{\delta}}(\theta; \gamma, P)\\ &\quad= \inf_{(\eta_x>0,\zeta_x\in\mathbb{R},\lambda_x\in\Lambda)_{x\in\mathcal{X}}} \sum_{x\in\mathcal{X}}\left(\eta_x \mathbb{E}^{F_*}\left[\phi^\star\left(\frac{k(U,x,\theta,\gamma)-\zeta_x-\lambda_x' g(U,x,\theta,\gamma)}{\eta_x}\right)\right]\right.\\ &\qquad\left. +\eta_x\delta_x+\zeta_x+\lambda_{12,x}' P_x\right). \end{aligned} \tag{41}$$

Moreover, the value of (40) is $+\infty$ (equivalently, the value of (41) is $-\infty$) if and only if, for some $x \in \mathcal{X}$, there is no distribution in $\mathcal{N}_{\delta_x}$ under which (37) holds at (θ, γ, P).

As before, estimators $\hat{\underline{\kappa}}_{\boldsymbol{\delta}}$ and $\hat{\overline{\kappa}}_{\boldsymbol{\delta}}$ of $\underline{\kappa}_{\boldsymbol{\delta}}$ and $\overline{\kappa}_{\boldsymbol{\delta}}$ are computed by optimizing sample criterion functions with respect to θ. Let $\hat{P} = (\hat{P}_x)_{x\in\mathcal{X}}$. The sample criterion functions are

$$\hat{\underline{K}}_{\boldsymbol{\delta}}(\theta) = \begin{cases}\underline{K}_{\boldsymbol{\delta}}(\theta; \hat\gamma, \hat P), \\ +\infty,\end{cases} \qquad \hat{\overline{K}}_{\boldsymbol{\delta}}(\theta) = \begin{cases}\overline{K}_{\boldsymbol{\delta}}(\theta; \hat\gamma, \hat P) & \text{if } \Delta_x(\theta;\hat\gamma,\hat P_x) < \delta_x \text{ for each } x \in \mathcal{X}, \\ -\infty & \text{if } \Delta_x(\theta;\hat\gamma,\hat P_x) \ge \delta_x \text{ for some } x \in \mathcal{X},\end{cases}$$

where $\overline{K}_{\boldsymbol{\delta}}(\theta; \hat\gamma, \hat P)$ and $\overline{K}_{\boldsymbol{\delta}}(\theta; \hat\gamma, \hat P)$ denote the programs in Proposition A.2 evaluated at $(\hat\gamma, \hat P)$, and

$$\Delta_x(\theta;\hat\gamma,\hat P_x) = \sup_{\zeta_x\in\mathbb{R},\lambda_x\in\Lambda} -\mathbb{E}^{F_*}\left[\phi^\star\left(-\zeta_x-\lambda_x' g(U,x,\theta,\hat\gamma)\right)\right]-\zeta_x-\lambda_{12,x}'\hat P_x.$$

A.3. *Non-Separable Models*

Consider the model

$$\begin{aligned}&\mathbb{E}^H\left[\tilde g_1(U,X,\theta,\tilde\gamma_0)\right] \le P_{10}, \qquad \mathbb{E}^H\left[\tilde g_2(U,X,\theta,\tilde\gamma_0)\right] = P_{20},\\ &\mathbb{E}^H\left[\tilde g_3(U,X,\theta,\tilde\gamma_0)\right] \le 0, \qquad \mathbb{E}^H\left[\tilde g_4(U,X,\theta,\tilde\gamma_0)\right] = 0,\end{aligned} \tag{42}$$

and counterfactual

$$\kappa = \mathbb{E}^H\left[\tilde k(U,X,\theta,\tilde\gamma_0)\right], \tag{43}$$

where the expectation is with respect to the distribution H of (U, X) and X takes values in a finite set $\mathcal{X}$. Suppose the researcher assumes $U|X = x \sim F_*$ for each x. We wish to relax this assumption and allow the conditional distribution of U given $X = x$, say F_x, to vary in a neighborhood $\mathcal{N}_{\delta_x}$ of F_*.

Write $H(u,x) = q_{0,x} \cdot F_x(u)$ where $q_{0,x} = \Pr(X = x)$. The vector $q_0 = (q_{0,x})_{x\in\mathcal{X}}$ can be consistently estimated from data on X. Let $\gamma_0 = (\tilde{\gamma}_0, q_0)$. Define $g_1(U,x,\theta,\gamma_0) = q_{0,x} \cdot \tilde{g}_1(U,x,\theta,\tilde{\gamma}_0)$ and similarly for g_2, g_3, g_4, and k. The model (42) and counterfactual (43) can then be written as

$$\begin{aligned} &\sum_{x\in\mathcal{X}} \mathbb{E}^{F_x}\big[g_1(U,x,\theta,\gamma_0)\big] \le P_{10}, \qquad \sum_{x\in\mathcal{X}} \mathbb{E}^{F_x}\big[g_2(U,x,\theta,\gamma_0)\big] = P_{20}, \\ &\sum_{x\in\mathcal{X}} \mathbb{E}^{F_x}\big[g_3(U,x,\theta,\gamma_0)\big] \le 0, \qquad \sum_{x\in\mathcal{X}} \mathbb{E}^{F_x}\big[g_4(U,x,\theta,\gamma_0)\big] = 0, \end{aligned} \tag{44}$$

and $\kappa = \sum_{x\in\mathcal{X}} \mathbb{E}^{F_x}[k(U,x,\theta,\gamma_0)]$. We again assume each $\mathcal{N}_\delta$ is defined by the same ϕ function, but allow the neighborhood size to vary with x. Let $\boldsymbol{\delta} = (\delta_x)_{x\in\mathcal{X}}$. We are interested in

$$\underline{\kappa}_{\boldsymbol{\delta}} := \inf_{\theta\in\Theta,(F_x\in\mathcal{N}_{\delta_x})_{x\in\mathcal{X}}} \sum_{x\in\mathcal{X}} \mathbb{E}^{F_x}\big[k(U,x,\theta,\gamma_0)\big] \quad \text{subject to (44)},$$

and $\overline{\kappa}_{\boldsymbol{\delta}}$ defined as the analogous supremum. One may write $\underline{\kappa}_{\boldsymbol{\delta}}$ and $\overline{\kappa}_{\boldsymbol{\delta}}$ as the value of two optimization problems where criterion functions $\underline{K}_{\boldsymbol{\delta}}(\theta;\gamma_0,P_0)$ and $\overline{K}_{\boldsymbol{\delta}}(\theta;\gamma_0,P_0)$ are optimized with respect to θ. For a generic (θ,γ,P), define

$$\underline{K}_{\boldsymbol{\delta}}(\theta;\gamma,P) = \inf_{(F_x\in\mathcal{N}_{\delta_x})_{x\in\mathcal{X}}} \sum_{x\in\mathcal{X}} \mathbb{E}^{F_x}\big[k(U,x,\theta,\gamma_0)\big] \quad \text{s.t. (44) holding at } (\theta,\gamma,P),$$

and define $\overline{K}_{\boldsymbol{\delta}}(\theta;\gamma,P)$ as the analogous supremum. These criterion functions have dual forms analogous to Proposition 2.1. Let $g(\cdot,x,\theta,\gamma) = (g_1(\cdot,x,\theta,\gamma),\dots,g_4(\cdot,x,\theta,\gamma))$. The remaining notation is the same as Proposition 2.1.

PROPOSITION A.3: *Suppose that Assumption Φ-conditional holds. Then*

$$\begin{aligned} &\underline{K}_{\boldsymbol{\delta}}(\theta;\gamma,P) \\ &\quad = \sup_{(\eta_x>0,\zeta_x\in\mathbb{R})_{x\in\mathcal{X}},\lambda\in\Lambda} \sum_{x\in\mathcal{X}} \Bigg(-\eta_x \mathbb{E}^{F_*}\bigg[\phi^\star\bigg(\frac{k(U,x,\theta,\gamma)+\zeta_x+\lambda' g(U,x,\theta,\gamma)}{-\eta_x}\bigg)\bigg] \\ &\qquad - \eta_x\delta_x - \zeta_x - \lambda_{12}' P\Bigg), \end{aligned} \tag{45}$$

$$\begin{aligned} &\overline{K}_{\boldsymbol{\delta}}(\theta;\gamma,P) \\ &\quad = \inf_{(\eta_x>0,\zeta_x\in\mathbb{R})_{x\in\mathcal{X}},\lambda\in\Lambda} \sum_{x\in\mathcal{X}} \Bigg(\eta_x \mathbb{E}^{F_*}\bigg[\phi^\star\bigg(\frac{k(U,x,\theta,\gamma)-\zeta_x-\lambda' g(U,x,\theta,\gamma)}{\eta_x}\bigg)\bigg] \\ &\qquad + \eta_x\delta_x + \zeta_x + \lambda_{12}' P\Bigg). \end{aligned} \tag{46}$$

Moreover, the value of (45) is $+\infty$ (equivalently, the value of (46) is $-\infty$) if and only if there is no $H(u,x) = q_{0,x}\cdot F_x(u)$ with $F_x \in \mathcal{N}_{\delta_x}$ under which (42) holds at (θ,γ,P).

As before, estimators $\hat{\underline{\kappa}}_\delta$ and $\hat{\overline{\kappa}}_\delta$ of $\underline{\kappa}_\delta$ and $\overline{\kappa}_\delta$ are computed by optimizing sample criterion functions with respect to θ. The sample criterion functions are

$$\hat{\underline{K}}_\delta(\theta) = \begin{cases} \underline{K}_\delta(\theta; \hat{\gamma}, \hat{P}), \\ +\infty, \end{cases} \qquad \hat{\overline{K}}_\delta(\theta) = \begin{cases} \overline{K}_\delta(\theta; \hat{\gamma}, \hat{P}) & \text{if } \Delta_{\text{nonsep}}(\theta; \hat{\gamma}, \hat{P}) < 0, \\ -\infty & \text{if } \Delta_{\text{nonsep}}(\theta; \hat{\gamma}, \hat{P}) \geq 0, \end{cases}$$

where $\overline{K}_\delta(\theta; \hat{\gamma}, \hat{P})$ and $\overline{K}_\delta(\theta; \hat{\gamma}, \hat{P})$ denote the programs in Proposition A.3 evaluated at $(\hat{\gamma}, \hat{P})$ with $\hat{\gamma} = (\hat{\tilde{\gamma}}, \hat{q})$ for estimators $\hat{\tilde{\gamma}}$ of $\tilde{\gamma}_0$ and $\hat{q}$ of q_0, and

$$\begin{aligned} &\Delta_{\text{nonsep}}(\theta; \gamma, P) \\ &= \sup_{\substack{(\eta_x \geq 0, \zeta_x \in \mathbb{R})_{x \in \mathcal{X}}, \lambda \in \Lambda \\ \sum_{x \in \mathcal{X}} \eta_x \leq 1}} \left(-\sum_{x \in \mathcal{X}} \mathbb{E}^{F_*}\left[(\eta_x \phi)^\star\left(-\zeta_x - \lambda_x' g(U, x, \theta, \gamma)\right)\right] - \eta_x \delta_x - \zeta_x \right) - \lambda_{12}' P. \end{aligned}$$

By similar arguments to Appendix G.3 of Christensen and Connault (2022), $\Delta_{\text{nonsep}}(\theta; \gamma, P)$ may be shown to be the dual of

$$\inf_{t \in \mathbb{R}, (F_x)_{x \in \mathcal{X}}} t \quad \text{s.t.} \quad D_\phi(F_x \| F_*) \leq \delta_x + t \text{ for each } x \in \mathcal{X} \text{ and (44) holding at } (\theta, \gamma, P).$$

Therefore, if there exists F_x with $D_\phi(F_x \| F_*) < \delta_x$ for each x such that (44) holds at (θ, γ, P), then $\Delta_{\text{nonsep}}(\theta; \gamma, P) < 0$.

APPENDIX B: ADDITIONAL RESULTS ON NONPARAMETRIC BOUNDS

This appendix presents further details to supplement Section 2.5. Proofs of all results in this appendix are presented in Appendix G.8 of Christensen and Connault (2022). Our first result concerns the behavior of $\underline{\kappa}_\delta$ and $\overline{\kappa}_\delta$ as the neighborhood size δ becomes large. Recall $\mathcal{N}_\infty = \{F : D_\phi(F \| F_*) < \infty\}$. Let

$$\mathcal{K}_\infty = \left\{\mathbb{E}^F\left[k(U, \theta, \gamma_0)\right] : \text{(1a)–(1d) holds at } (\theta, \gamma_0, P_0) \text{ for some } \theta \in \Theta, F \in \mathcal{N}_\infty\right\}.$$

LEMMA B.1: *Suppose that Assumption* Φ *holds. Then*

$$\lim_{\delta \to \infty} \underline{\kappa}_\delta = \inf \mathcal{K}_\infty, \qquad \lim_{\delta \to \infty} \overline{\kappa}_\delta = \sup \mathcal{K}_\infty.$$

Next, we characterize bounds on $\mathcal{K}_\infty$ using profiled optimization problems and derive their dual forms. Define

$$\underline{K}_\infty(\theta; \gamma_0, P_0) = \inf_{F \in \mathcal{N}_\infty} \mathbb{E}^F\left[k(U, \theta, \gamma_0)\right] \quad \text{subject to (1a)–(1d) holding at } (\theta, F), \tag{47}$$

and let $\overline{K}_\infty(\theta; \gamma_0, P_0)$ denote the analogous supremum. By definition, we have

$$\inf \mathcal{K}_\infty = \inf_{\theta \in \Theta} \underline{K}_\infty(\theta; \gamma_0, P_0), \qquad \sup \mathcal{K}_\infty = \sup_{\theta \in \Theta} \overline{K}_\infty(\theta; \gamma_0, P_0).$$

Let F_*-ess inf and F_*-ess sup denote the F_*-essential infimum and supremum, respectively.

LEMMA B.2: *Suppose that Assumption* Φ *holds and Condition* S *holds at* (θ, γ, P). *Then*

$$\underline{K}_\infty(\theta; \gamma, P) = \sup_{\lambda \in \Lambda : F_*\text{-ess}\inf(k(\cdot,\theta,\gamma)+\lambda' g(\cdot,\theta,\gamma)) > -\infty} \big(F_*\text{-ess}\inf\big(k(\cdot, \theta, \gamma) + \lambda' g(\cdot, \theta, \gamma)\big) - \lambda_{12}' P\big),$$

$$\overline{K}_\infty(\theta; \gamma, P) = \inf_{\lambda \in \Lambda : F_*\text{-ess}\sup(k(\cdot,\theta,\gamma)-\lambda' g(\cdot,\theta,\gamma)) < +\infty} \big(F_*\text{-ess}\sup\big(k(\cdot, \theta, \gamma) - \lambda' g(\cdot, \theta, \gamma)\big) + \lambda_{12}' P\big).$$

We now derive analogous dual representations for the criterion functions $\underline{K}_{\text{np}}$ and $\overline{K}_{\text{np}}$ from Section 2.5 (see display (18)). We require a slightly different constraint qualification:

DEFINITION B.1: *Condition* S_{np} holds at (θ, γ, P) if $\vec{P} \in \text{ri}(\{\mathbb{E}^F[g(U, \theta, \gamma)] : F \in \mathcal{F}_\theta\} + \mathcal{C})$.

If F_* and μ are mutually absolutely continuous, then Condition S_{np} is equivalent to Condition S from Section 2.5 (see Lemma E.1).

LEMMA B.3: *Suppose that Condition* S_{np} *holds at* (θ, γ, P) *and* k *is* μ*-essentially bounded. Then*

$$\underline{K}_{\text{np}}(\theta; \gamma, P) = \sup_{\lambda \in \Lambda : \mu\text{-ess}\inf(k(\cdot,\theta,\gamma)+\lambda' g(\cdot,\theta,\gamma)) > -\infty} \big(\mu\text{-ess}\inf\big(k(\cdot, \theta, \gamma) + \lambda' g(\cdot, \theta, \gamma)\big) - \lambda_{12}' P\big),$$

$$\overline{K}_{\text{np}}(\theta; \gamma, P) = \inf_{\lambda \in \Lambda : \mu\text{-ess}\sup(k(\cdot,\theta,\gamma)-\lambda' g(\cdot,\theta,\gamma)) < +\infty} \big(\mu\text{-ess}\sup\big(k(\cdot, \theta, \gamma) - \lambda' g(\cdot, \theta, \gamma)\big) + \lambda_{12}' P\big).$$

APPENDIX C: LOCAL SENSITIVITY

In this appendix, we first introduce a measure of local sensitivity of the counterfactual with respect to F. We then contrast our approach with recent work on local sensitivity.

C.1. *Measure of Local Sensitivity*

Our *measure of local sensitivity* of the counterfactual κ with respect to F at F_* is

$$s = \lim_{\delta \downarrow 0} \frac{(\overline{\kappa}_\delta - \underline{\kappa}_\delta)^2}{4\delta}.$$

If s is finite, then under the regularity conditions below,

$$\underline{\kappa}_\delta = \kappa_* - \sqrt{\delta s} + o(\sqrt{\delta}), \qquad \overline{\kappa}_\delta = \kappa_* + \sqrt{\delta s} + o(\sqrt{\delta}), \quad \text{as } \delta \downarrow 0,$$

where $\kappa_* = \mathbb{E}^{F_*}[k(U, \theta_*, \gamma_0)]$ and θ_* solves (1a)–(1d) under F_*.

To draw connections with the local sensitivity literature, we restrict attention to moment equality models and impose (standard) regularity conditions. These conditions allow us to characterize s very tractably via an influence function representation, which leads to a simple estimator $\hat{s}$ of s. Assume that under F_*, the moment conditions (1b) and (1d) point-identify a structural parameter $\theta_* \in \text{int}(\Theta)$, where we again assume Θ is compact. With some abuse of notation, let

$$g(u, \theta, \gamma, P_2) = \begin{bmatrix} g_2(u, \theta, \gamma) - P_2 \\ g_4(u, \theta, \gamma) \end{bmatrix},$$

$g_*(u) = g(u, \theta_*, \gamma_0, P_{20})$, and $k_*(u) = k(u, \theta_*, \gamma_0)$. Let $\mathbb{E}^{F_*}[g(U, \theta, \gamma_0, P_{20})]$ and $\mathbb{E}^{F_*}[k(U, \theta, \gamma_0)]$ be continuously differentiable with respect to θ at θ_*, $G = \frac{\partial}{\partial \theta'}\mathbb{E}^{F_*}[g(U, \theta, \gamma_0, P_{20})]|_{\theta=\theta_*}$ have full rank, $V = \mathbb{E}^{F_*}[g_*(U)g_*(U)']$ be finite and positive definite, $\mathbb{E}^{F_*}[k(U, \theta_*, \gamma_0)^2]$ be finite, and $k(\cdot, \theta, \gamma_0)$ and $g(\cdot, \theta, \gamma_0, P_{20})$ be $L^2(F_*)$-continuous in θ at θ_*.

Define the *influence function* of the counterfactual κ with respect to F at F_* as

$$\iota(u) = \mathbb{M}k_*(u) - J'(G'V^{-1}G)^{-1}G'V^{-1}g_*(u),$$

where $\mathbb{M}k_*(u) = k_*(u) - \kappa_* - \mathbb{E}^{F_*}[k_*(U)g_*(U)'](V^{-1} - V^{-1}G(G'V^{-1}G)^{-1}G'V^{-1})g_*(u)$ and $J = \frac{\partial}{\partial \theta}\mathbb{E}^{F_*}[k(U, \theta, \gamma_0)]|_{\theta=\theta_*}$. The following theorem relates s and ι. We restrict attention to neighborhoods characterized by χ^2 divergence. Other ϕ-divergences are locally equivalent to χ^2 divergence, so this restriction entails no great loss of generality.[26]

THEOREM C.1: *Suppose that the above GMM-type regularity conditions hold and neighborhoods are defined using χ^2 divergence. Then $s = 2\mathbb{E}^{F_*}[\iota(U)^2]$.*

The proof of Theorem C.1 is presented in Appendix G.9 of our working paper version Christensen and Connault (2022). In addition to reporting an estimated counterfactual $\hat{\kappa} = \mathbb{E}^{F_*}[k(U, \hat{\theta}, \hat{\gamma})]$, researchers could also report an estimate of its local sensitivity to F:

$$\hat{s} = 2\mathbb{E}^{F_*}\big[\big(\hat{k}(U) - \hat{\kappa}\big)^2\big] + 2\hat{Q}'\hat{V}\hat{Q} - 4\mathbb{E}^{F_*}\big[\hat{g}(U)\big(\hat{k}(U) - \hat{\kappa}\big)\big]'\hat{Q},$$

where $\hat{k}(u) = k(u, \hat{\theta}, \hat{\gamma})$, $\hat{g}(u) = g(u, \hat{\theta}, \hat{\gamma}, \hat{P}_2)$, $\hat{V} = \mathbb{E}^{F_*}[\hat{g}(U)\hat{g}(U)']$, and

$$\hat{Q}' = \mathbb{E}^{F_*}\big[\hat{k}(U)\hat{g}(U)'\big]\big(\hat{V}^{-1} - \hat{V}^{-1}\hat{G}\big(\hat{G}'\hat{V}^{-1}\hat{G}\big)^{-1}\hat{G}'\hat{V}^{-1}\big) + \hat{J}'\big(\hat{G}'\hat{V}^{-1}\hat{G}\big)^{-1}\hat{G}'\hat{V}^{-1},$$

with $\hat{G} = \frac{\partial}{\partial \theta'}\mathbb{E}^{F_*}[g(U, \theta, \hat{\gamma}, \hat{P}_2)]|_{\theta=\hat{\theta}}$ and $\hat{J} = \frac{\partial}{\partial \theta}\mathbb{E}^{F_*}[k(U, \theta, \hat{\gamma})]|_{\theta=\hat{\theta}}$. Lemma G.12 in Appendix G.9 of Christensen and Connault (2022) shows $\hat{s}$ is consistent. Bounds on counterfactuals as F varies over small neighborhoods of F_* can then be estimated using $\hat{\kappa} \pm \sqrt{\delta \hat{s}}$.

C.2. *Comparison With Other Approaches*

We now compare our approach with Andrews, Gentzkow, and Shapiro (2017, 2020), henceforth AGS, and Bonhomme and Weidner (2022), henceforth BW. To simplify the comparison, we consider models characterized by moments of the form (1b) with $d_2 \geq d_\theta$ and in which there is no γ.

AGS considered a setting in which the moments (1b) are locally misspecified:

$$\mathbb{E}^{F_*}\big[g_2(U, \theta_*)\big] = P_{20} + n^{-1/2}c. \tag{48}$$

Suppose a researcher has a first-stage estimator $\hat{P}_2$, computes an estimator $\hat{\theta}$ by minimizing

$$\big(\mathbb{E}^{F_*}\big[g_2(U, \theta)\big] - \hat{P}_2\big)'\hat{W}\big(\mathbb{E}^{F_*}\big[g_2(U, \theta)\big] - \hat{P}_2\big),$$

[26]See Theorem 4.1 of Csiszár and Shields (2004). The quantity $2\mathbb{E}^{F_*}[\iota(U)^2]$ should be rescaled by a factor of $\phi''(1)$ for other ϕ divergences.

then estimates the counterfactual using $\hat{\kappa} = \mathbb{E}^{F_*}[k(U, \hat{\theta})]$. AGS's measure of *sensitivity* of $\hat{\kappa}$ to $\hat{P}_2$ is $J'(G'WG)^{-1}G'W$, where W is the probability limit of $\hat{W}$. The first-order asymptotic bias of $\hat{\kappa}$ due to local misspecification is therefore $J'(G'WG)^{-1}G'Wc$. AGS's measure of *informativeness* of $\hat{P}_2$ for $\hat{\kappa}$ is 1, meaning that all sampling variation in $\hat{\kappa}$ is explained by sampling variation in $\hat{P}_2$. Our measure s instead characterizes "specification variation" in κ as the researcher varies F subject to the moment condition (1b).

BW considered estimation of a target parameter (κ in our context) using a reference model $\mathcal{M}_R = \{(\theta, F) \in \Theta \times \{F_*\}\}$ when the true (θ_0, F_0) possibly belongs to a larger model $\mathcal{M}_L = \{(\theta, F) \in \Theta \times \mathcal{N}_\delta\}$ with $\delta \downarrow 0$ as the sample size n increases so that $n\delta \to \tau \geq 0$. BW seek estimators of κ under $\mathcal{M}_R$ that minimize worst-case asymptotic bias or MSE over $\mathcal{M}_L$. Consider the one-step estimator

$$\hat{\kappa} = \mathbb{E}^{F_*}\big[k(U, \hat{\theta})\big] + a'\big(\mathbb{E}^{F_*}\big[g_2(U, \hat{\theta})\big] - \hat{P}_2\big),$$

where $\hat{\theta}$ is a $\sqrt{n}$-consistent estimator of θ_* and $a \in \mathbb{R}^{d_2}$ satisfies $J' = -a'G$ so that $\hat{\kappa}$ does not depend asymptotically on $\hat{\theta}$. The true counterfactual is $\kappa_0 = \mathbb{E}^{F_0}[k(U, \theta_0)]$ where $(\theta_0, F_0) \in \mathcal{M}_L$ satisfies $\mathbb{E}^{F_0}[g_2(U, \theta_0)] = P_{20}$. If $\mathcal{M}_R$ is correctly specified so that $\mathbb{E}^{F_*}[g_2(U, \theta_*)] = P_{20}$, then, for any a, the worst-case asymptotic bias of the one-step estimator is

$$\lim_{n\to\infty} \sup_{(\theta_0,F_0)\in\mathcal{M}_L:\mathbb{E}^{F_0}[g_2(U,\theta_0)]=P_{20}} \big|\sqrt{n}(\kappa_* - \kappa_0))\big| = \sqrt{\tau s},$$

where s is our measure of local sensitivity.

If we allow for local misspecification of $\mathcal{M}_R$, so that $\mathbb{E}^{F_*}[g_2(U, \theta_*)] \neq P_{20}$, then the worst-case asymptotic bias of the one-step estimator is

$$\lim_{n\to\infty} \sup_{(\theta_0,F_0)\in\mathcal{M}_L:\mathbb{E}^{F_0}[g_2(U,\theta_0)]=P_{20}} \big|\sqrt{n}\big(\kappa_* - \kappa_0 + a'\big(\mathbb{E}^{F_*}\big[g_2(U, \theta_*)\big] - P_{20}\big)\big)\big| = \sqrt{\tau s_a},$$

where s_a is our local sensitivity measure with k replaced by $k + a'g_2$.

APPENDIX D: ADDITIONAL DETAILS FOR THE EMPIRICAL APPLICATIONS

D.1. *Marital College Premium*

Bootstrap Details. CSs reported in Section 5.1 with $\delta > 0$ are computed using the bootstrap procedure from Section 6.2. To implement the bootstrap, we take 1000 independent draws of $\hat{P}_2^* \sim N(\hat{P}_2, \hat{\Sigma})$ where $\hat{\Sigma}$ is CSW's estimate of the covariance matrix of $\hat{P}_2$. We compute $\hat{P}_2$ and $\hat{\Sigma}$ based on CSW's replication files.

Fixed-θ Bounds. Figure 5(a) plots lower and upper bounds on the "some college" to "college graduate" premium across cohorts when θ is fixed at CSW's estimates (computed under F_*) but F is allowed to vary. These bounds for large δ contain zero across each cohort, and are approximately the same width as the bounds with $\delta = 0.01$ reported in Figure 1(a) where both θ and F are allowed to vary. Imposing exchangeability (Figure 5(b)) is seen to tighten the bounds substantially, producing bounds that span negative values only for early cohorts and positive values only in the latest few cohorts.

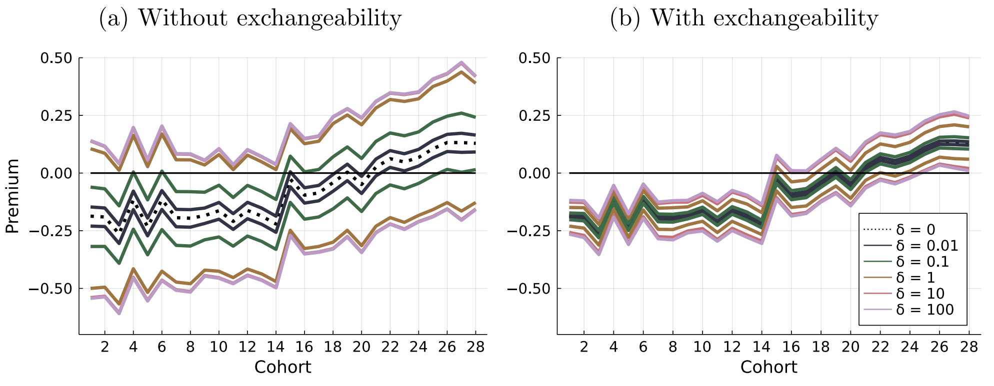

FIGURE 5.—Fixed-θ bounds on the "some college" to "college graduate" premium when structural parameters are held fixed at CSW's estimates.

Projection CSs. Figure 6(a) reports projection CSs computed using the procedure in Section 6.3. We formed 95% rectangular CSs for each cohort's P_{20} as described in Section 6.3 using CSW's estimates for $\hat{P}_2$ and their asymptotic variance estimates for $\hat{\Sigma}$. These CSs are significantly wider than the bootstrap CSs reported in Figure 1. Some conservativeness is to be expected, as these CSs project a 95% CS for P_{20} down to one dimension. The relative inefficiency is especially pronounced for the earlier cohorts. Note also from Figure 6(b) that the projection CSs with $\delta = 0.01$ span zero across each cohort, whereas the bootstrap CSs with $\delta = 0.01$ in Figure 1(b) contain negative values only in some early cohorts and positive values only in later cohorts.

Computation Times. Table V reports times for solving the inner problem for maximizing the premium in cohort 1. This optimization problem defines the criterion function $\overline{K}_\delta(\theta; \hat{P})$. As times vary with θ, we report times at CSW's estimates. Times increase somewhat with δ, but are all under 0.6 seconds. The outer optimization times varied with cohort, δ, and implementation, but were typically solved in at most a few minutes (often under 90 seconds).

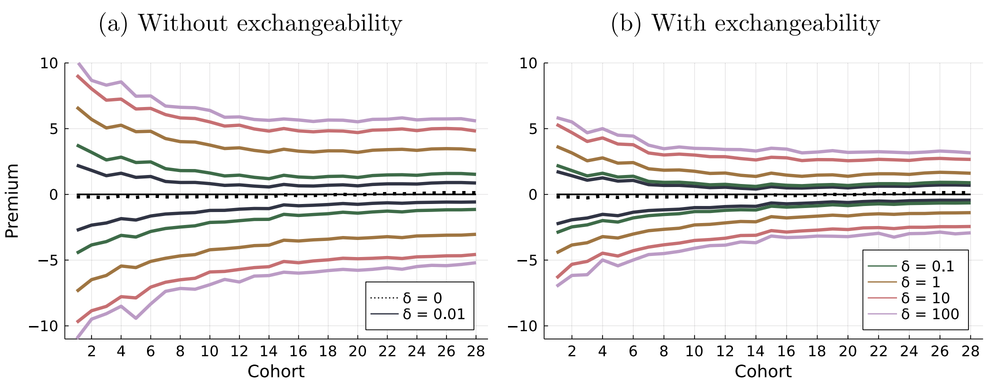

FIGURE 6.—Projection 95% CSs for bounds on the "some college" to "college graduate" premium across cohorts.

TABLE V

COMPUTATION TIMES FOR THE INNER PROBLEM IN THE MATCHING APPLICATION.

Implementation	δ				
	0.01	0.1	1	10	100
Without exchangeability	0.074	0.056	0.076	0.579	0.188
With exchangeability	0.146	0.184	0.350	0.311	0.488

Note: Times (in seconds) for solving the inner optimization problem for maximizing the premium in cohort 1 at CSW's parameter estimate θ. We use 50,000 Monte Carlo draws without exchangeability and 120,000 draws with exchangeability. All computations are performed in Julia version 1.6.4 and Knitro 12.4.0 on a 2.7 GHz MacBook Pro with 16 GB memory.

Sensitivity to ϕ. Using χ^2 and L^4 divergences produced near identical bounds for $\delta = 0.01$ and 0.1. The χ^2 bounds with $\delta = 1$ and 10 were at most 10% narrower than the hybrid bounds. The L^4 bounds were 60%–70% of the width of the hybrid bounds for $\delta = 1$, 10, and 100 across cohorts (L^4 divergence is stronger than χ^2 and hybrid divergence). The shapes of the sets were also similar to those reported for hybrid divergence. Overall, these results show that the conclusions we draw from our analysis are not sensitive to the choice of ϕ.

D.2. *Welfare Analysis in a Rust Model*

Bootstrap Details. Bootstrap CSs reported in Section 5.2 with $\delta > 0$ are computed using the procedure from Section 6.2. We take 1000 independent draws of $\hat{\theta}^*_\pi \sim N(\hat{\theta}_\pi, \hat{\Sigma})$ where $\hat{\theta}_\pi$ is the MLE of (RC, MC) under the i.i.d. Gumbel assumption and $\hat{\Sigma}$ is an estimate of the inverse information matrix. We then set $\hat{P}^*_2$ to be the model-implied CCPs at $\hat{\theta}^*_\pi$ under the i.i.d. Gumbel assumption.

As k depends only implicitly on u through θ, we compute $\hat{\underline{\kappa}}_\delta$ and $\hat{\overline{\kappa}}_\delta$ using the criterion functions in display (17), which is more computationally efficient than using criterions (13) and (14). The λ multipliers on the minimum divergence problem $\Delta(\theta; P)$ in (17) differ from λ in criterions (13) and (14) by the factor η (see the discussion in Section 2.3). As our bootstrap methods are derived based on criterions (13) and (14), when implementing the bootstrap we rescale the multiplier λ solving (16) by the multiplier η on the constraint $\Delta(\theta; \hat{P}) \leq \delta$ in the outer optimization.[27] As η and λ are computed separately in the outer and inner optimizations, respectively, it is computationally most convenient to implement our bootstrap CSs with $\hat{\nu} = 0$. As discussed in Section 6.2, this choice is valid but possi-

[27]This rescaling is also justified as follows. Let $\overline{b}_\delta(P) = \sup_{\theta \in \Theta : \Delta(\theta; P) \leq \delta} k(\theta)$ and note $\overline{\kappa}_\delta = \overline{b}_\delta(P_0)$ and $\hat{\overline{\kappa}}_\delta = \overline{b}_\delta(\hat{P})$. By similar arguments to Corollary 5 of Milgrom and Segal (2002), one may deduce that the directional derivative of $\overline{b}_\delta(P)$ at P_0 involves multiplying the directional derivative of $P \mapsto \Delta(\theta; P)$ at P_0 by the multiplier for $\Delta(\theta; P) \leq \delta$. The directional derivative of $P \mapsto \Delta(\theta; P)$ at P_0 may be shown to be

$$\lim_{n \to \infty} t_n^{-1}\big(\Delta(\theta; P_0 + t_n h_n) - \Delta(\theta; P_0)\big) = \sup_{\lambda_{12} \in \underline{\Delta}(\theta; P_0)} -\lambda'_{12} h,$$

where $\underline{\Delta}(\theta; P_0)$ is constructed analogously to $\underline{\Lambda}_\delta(\theta; P)$ in Section 6.2 using the set of multipliers that solve the minimum divergence problem (16).

TABLE VI

COMPUTATION TIMES FOR THE INNER PROBLEM IN THE DDC APPLICATION.

	δ				
	0.01	0.1	1	10	100
Lower bound	0.124	0.144	0.164	0.285	0.265
Upper bound	0.101	0.119	0.142	0.266	1.039

Note: Times (in seconds) for solving the inner optimization problem at the parameter values at which $\hat{\underline{\kappa}}_\delta$ and $\hat{\overline{\kappa}}_\delta$ are attained. All computations are performed in Julia version 1.6.4 and Knitro 12.4.0 on a 2.7 GHz MacBook Pro with 16 GB memory.

bly conservative. Despite this potentially conservative choice, the bootstrap CSs are not materially wider than the bootstrap CSs under the i.i.d. Gumbel assumption.

To construct the projection CSs, we form a 95% rectangular CS for P_{20} as described in Section 6.3. For each draw of $\hat{\theta}^*_\pi$, we compute the model-implied CCPs $\hat{P}^*_2$ under the i.i.d. Gumbel assumption. We construct t-statistics for each CCP by centering $\hat{P}^*_2$ at $\hat{P}_2$ and studentizing by its standard deviation across draws. For each draw, we compute the maximum of the absolute value of the t-statistics. We then take the critical value $\hat{c}_{2,1-\alpha}$ to be the $1-\alpha$ quantile of the maximum statistic across draws.

Computation Times. Table VI reports computation times for the inner optimization for evaluating the criterion functions $\hat{\underline{K}}_\delta(\theta; \hat{\gamma}, \hat{P})$ and $\hat{\overline{K}}_\delta(\theta; \hat{\gamma}, \hat{P})$ at the parameter values at which $\hat{\underline{\kappa}}_\delta$ and $\hat{\overline{\kappa}}_\delta$ are attained. The computation times correspond to solving the minimum divergence problem $\Delta(\theta; \hat{\gamma}, \hat{P})$ because k does not depend on u (cf. display (17)). The outer optimizations were typically solved in a few minutes in an 8-core environment with 64GB memory.

Sensitivity to ϕ. Bounds with χ^2-divergence were between 4% narrower and 1% wider than the bounds for hybrid divergence for all values of δ. Repeating the analysis with L^4-divergence, which is stronger than χ^2 and hybrid divergence, produced bounds that were 10%–30% narrower than the hybrid divergence bounds up to $\delta = 1$ and at most 5% narrower than the hybrid divergence bounds for larger values of δ. As with the matching application, these results again show that the conclusions we draw from our analysis are not sensitive to the choice of ϕ function.

APPENDIX E: PROOFS OF MAIN RESULTS

Throughout the proofs, we abbreviate upper-semicontinuous and upper-semicontinuity to u.s.c. and lower-semicontinuous and lower-semicontinuity to l.s.c.

E.1. *Proofs for Section 2*

PROOF OF PROPOSITION 2.1: Immediate from Proposition G.1 in Appendix G.2 of Christensen and Connault (2022). *Q.E.D.*

Recall Condition S from Definition 2.1 and Condition S_{np} from Definition B.1.

LEMMA E.1: *Suppose that Assumption* Φ *holds and* μ *and* F_* *are mutually absolutely continuous. Then Condition* S *holds at* (θ, γ, P) *if and only if Condition* S_{np} *holds at* (θ, γ, P).

PROOF OF LEMMA E.1: In view of Hölder's inequality for Orlicz classes (see (A.1) in Appendix F of Christensen and Connault (2022)), Assumption Φ implies $\mathcal{N}_\infty = \{F : D_\phi(F\|F_*) < \infty\} \subseteq \mathcal{F}_\theta$. Therefore,

$$\mathcal{G}_\infty := \{\mathbb{E}^F[g(U, \theta, \gamma)] : F \in \mathcal{N}_\infty\} \subseteq \{\mathbb{E}^F[g(U, \theta, \gamma)] : F \in \mathcal{F}_\theta\} =: \mathcal{G}_\theta.$$

By Corollary 6.6.2 of Rockafellar (1970), it suffices to show $\mathrm{ri}(\mathcal{G}_\infty) = \mathrm{ri}(\mathcal{G}_\theta)$. As $\mathrm{ri}(\mathcal{G}_\infty) \subseteq \mathcal{G}_\infty \subseteq \mathcal{G}_\theta$, it suffices to show $\mathcal{G}_\theta \subseteq \mathrm{cl}(\mathcal{G}_\infty)$ (Hiriart-Urruty and Lemaréchal, 2001, Remark 2.1.9). For any $x \in \mathcal{G}_\theta$, we have $x = \mathbb{E}^F[g(U, \theta, \gamma)]$ for some $F \in \mathcal{F}_\theta$. As $F \ll \mu$ and F_* and μ are mutually absolutely continuous, F has a density, say m, with respect to F_*. For each $n \geq 1$, let $m(u) \wedge n = \min\{m(u), n\}$ and define

$$m_n(u) = \frac{m(u) \wedge n}{\int (m(u) \wedge n)\, \mathrm{d}F_*(u)}.$$

Each F_n defined by $\mathrm{d}F_n = m_n\, \mathrm{d}F_*$ belongs to $\mathcal{N}_\infty$. It follows that $\mathbb{E}^{F_n}[g(U, \theta, \gamma)] \in \mathcal{G}_\infty$. By monotone convergence, we have $\mathbb{E}^{F_n}[g(U, \theta, \gamma)] \to x$. Therefore, $x \in \mathrm{cl}(\mathcal{G}_\infty)$. *Q.E.D.*

PROOF OF THEOREM 2.1: We prove only the result for $\inf \mathcal{K}$; the result for $\sup \mathcal{K}$ follows similarly. Note

$$\inf \mathcal{K} = \inf_{\theta \in \Theta} \underline{K}_{\mathrm{np}}(\theta; \gamma_0, P_0) = \inf_{\theta \in \Theta_I} \underline{K}_{\mathrm{np}}(\theta; \gamma_0, P_0),$$

where the first equality is by definition and the second equality holds because, if $\theta \notin \Theta_I$, then there does not exist a distribution $F \in \mathcal{F}_\theta$ under which the moment conditions hold at (θ, γ_0, P_0) and consequently $\underline{K}_{\mathrm{np}}(\theta; \gamma_0, P_0) = +\infty$. If $\theta \notin \Theta_I$, then there does not exist $F \in \mathcal{N}_\infty$ under which the moment conditions hold at (θ, γ_0, P_0) either because $\mathcal{N}_\infty \subseteq \mathcal{F}_\theta$ for all θ under Assumption Φ. Therefore, $\underline{K}_\infty(\theta; \gamma_0, P_0) = +\infty$ in that case, too. We therefore have

$$\inf \mathcal{K}_\infty = \inf_{\theta \in \Theta_I} \underline{K}_\infty(\theta; \gamma_0, P_0).$$

In view of Lemma B.1, it suffices to show $\inf \mathcal{K} = \inf \mathcal{K}_\infty$. Note that $\inf \mathcal{K} \leq \inf \mathcal{K}_\infty$ holds by virtue of the inclusion $\mathcal{N}_\infty \subseteq \mathcal{F}_\theta$ for all θ. For the reverse inequality, choose any $\epsilon > 0$. By S-regularity of Θ_I, there exists $\underline{\theta} \in \Theta_I$ for which Condition S holds at $(\underline{\theta}, \gamma_0, P_0)$ and for which $\underline{K}_{\mathrm{np}}(\underline{\theta}; \gamma_0, P_0) \leq \inf \mathcal{K} + \epsilon$. As Condition S holds at $(\underline{\theta}, \gamma_0, P_0)$ and $\mu \ll F_* \ll \mu$, Lemma E.1 implies that Condition S_{np} must also hold at $(\underline{\theta}, \gamma_0, P_0)$. Moreover, the μ-essential infimum and F_*-essential infimum of any function are equal because $\mu \ll F_* \ll \mu$. Therefore, by Lemmas B.2 and B.3, we have $\underline{K}_\infty(\underline{\theta}; \gamma_0, P_0) = \underline{K}_{\mathrm{np}}(\underline{\theta}; \gamma_0, P_0)$. It follows by definition of $\inf \mathcal{K}_\infty$ that $\inf \mathcal{K}_\infty \leq \underline{K}_\infty(\underline{\theta}; \gamma_0, P_0) = \underline{K}_{\mathrm{np}}(\underline{\theta}; \gamma_0, P_0) \leq \inf \mathcal{K} + \epsilon$. Therefore, $\inf \mathcal{K}_\infty \leq \inf \mathcal{K}$. *Q.E.D.*

E.2. *Proofs for Section 3*

PROOF OF PROPOSITION 3.1: We prove the result for $\underline{K}_\delta$; the proof for $\overline{K}_\delta$ follows similarly. Consider

$$v^A := \inf_{\theta \in \Theta, F \in \mathcal{N}_\delta} \mathbb{E}^F[k(U, \theta, \gamma)] \quad \text{subject to (1a)–(1d) holding at } (\theta, \gamma, P), \quad \text{(Program A)}$$

$$v^B := \inf_{\theta\in\Theta} \mathbb{E}^{\underline{F}_{\delta,\theta}}\big[k(U,\theta,\gamma)\big] \quad \text{subject to } \mathbb{E}^{\underline{F}_{\delta,\theta}}\big[g_{4e}(U,\theta,\gamma)\big] = 0, \qquad \text{(Program B)}$$

where $\underline{F}_{\delta,\theta}$ solves

$$\inf_{F\in\mathcal{N}_\delta} \mathbb{E}^F\big[k(U,\theta,\gamma)\big] \quad \text{subject to (19) holding at } (\theta,\gamma,P),$$

and $v^B = +\infty$ if there is no solution to this problem. Program A is the approach described in Section 2 whereas Program B is equivalent to our MPEC implementation.

The inequality $v^A \le v^B$ is trivial if $v^B = +\infty$. If v^B is finite, for any $\varepsilon > 0$ there exists $\theta^B_\varepsilon \in \Theta$ for which $\mathbb{E}^{\underline{F}_{\delta,\theta^B_\varepsilon}}[k(U,\theta^B_\varepsilon,\gamma)] \le v^B + \varepsilon$ and $\mathbb{E}^{\underline{F}_{\delta,\theta^B_\varepsilon}}[g_{4e}(U,\theta^B_\varepsilon,\gamma)] = 0$ where $\underline{F}_{\delta,\theta^B_\varepsilon}$ is well defined by Lemma G.2(ii) of Christensen and Connault (2022). As $(\theta^B_\varepsilon, \underline{F}_{\delta,\theta^B_\varepsilon})$ are feasible for Program A, we have $v^A \le v^B + \varepsilon$. As ε is arbitrary, we have $v^A \le v^B$ whenever $v^B > -\infty$.

A similar argument applies when $v^B = -\infty$: for any $n \in \mathbb{N}$, there exists $\theta^B_n \in \Theta$ for which $\mathbb{E}^{\underline{F}_{\delta,\theta^B_n}}[k(U,\theta^B_\varepsilon,\gamma)] \le -n$ and $\mathbb{E}^{\underline{F}_{\delta,\theta^B_n}}[g_{4e}(U,\theta^B_\varepsilon,\gamma)] = 0$, where the distribution $\underline{F}_{\delta,\theta^B_n}$ is well defined by Lemma G.2(ii) of Christensen and Connault (2022). As $(\theta^B_n, \underline{F}_{\delta,\theta^B_n})$ are feasible for Program A, we have $v^A \le -n$. Hence, $v^A = v^B = -\infty$.

Note $v^B \le v^A$ holds trivially if $v^A = +\infty$. If v^A is finite, rewrite Program B as

$$\inf_{\kappa\in\mathbb{R},\theta\in\Theta} \kappa \quad \text{subject to } \mathbb{E}^{\underline{F}_{\delta,\theta,\kappa}}\big[g_{4e}(U,\theta,\gamma)\big] = 0,$$

where $\underline{F}_{\delta,\theta,\kappa}$ solves the feasibility program

$$\inf_{F\in\mathcal{N}_\delta} 0 \quad \text{subject to (19) and } \mathbb{E}^F\big[k(U,\theta,\gamma)\big] = \kappa \text{ holding at } (\theta,\gamma,P). \qquad (49)$$

For any $\varepsilon > 0$, there exists $\theta^A_\varepsilon \in \Theta$ and $F^A_\varepsilon \in \mathcal{N}_\delta$ such that the constraints in Program A are satisfied, that is, $\mathbb{E}^{F^A_\varepsilon}[g_1(U,\theta^A_\varepsilon,\gamma)] \le P_1, \ldots, \mathbb{E}^{F^A_\varepsilon}[g_4(U,\theta^A_\varepsilon,\gamma)] = 0$, and

$$\mathbb{E}^{F^A_\varepsilon}\big[k\big(U,\theta^A_\varepsilon,\gamma\big)\big] \le v^A + \varepsilon.$$

Then $\underline{F}^A_\varepsilon$ solves the feasibility program (49) with $\theta = \theta^A_\varepsilon$ and $\kappa = \kappa^A_\varepsilon := \mathbb{E}^{F^A_\varepsilon}[k(U,\theta^A_\varepsilon,\gamma)]$. Note that $\mathbb{E}^{F^A_\varepsilon}[g_{4e}(U,\theta^A_\varepsilon,\gamma)] = 0$ also holds by construction. Therefore, $(\kappa^A_\varepsilon, \theta^A_\varepsilon)$ are feasible for the augmented form of Program B. It follows that $v^B \le \kappa^A_\varepsilon \le v^A + \varepsilon$ holds for each $\varepsilon > 0$. As $\varepsilon > 0$ is arbitrary, we have $v^B \le v^A$ whenever $v^A > -\infty$.

A similar argument applies if $v^A = -\infty$: for any $n \in \mathbb{N}$, we may choose $\theta^A_n \in \Theta$ and $F^A_n \in \mathcal{N}_\delta$ such that the constraints in Program A are satisfied and $\mathbb{E}^{F^A_n}[k(U,\theta^A_n,\gamma)] \le -n$. It follows that $v^B \le -n$. Hence, $v^B = v^A = -\infty$. *Q.E.D.*

PROOF OF PROPOSITION 3.2: We prove the result for $\underline{F}_{\delta,\theta}$; the result for $\overline{F}_{\delta,\theta}$ follows similarly. We drop dependence of k and g on (θ,γ) to simplify notation in what follows.

First, suppose k depends on u. The dual formulation is justified by Proposition 2.1. A dual solution $(\underline{\eta}, \underline{\zeta}, \underline{\lambda})$ exists by Proposition G.1(iii) of Christensen and Connault (2022).

Suppose $\underline{\eta} > 0$. We wish to show that the change of measure $\underline{m}_{\delta,\theta}(u) = \dot{\phi}^\star(-\underline{\eta}^{-1}(k(u) + \underline{\zeta} + \underline{\lambda}' g_s(u)))$ induces a distribution that solves the primal problem (20) at θ. Differentiability of the objective function in (η,ζ,λ) is guaranteed by Assumption Φ. Also note that

Assumption Φ(i) ensures $\dot{\phi}^\star \geq 0$. The first-order condition (FOC) for $\underline{\zeta}$ is

$$0 = \mathbb{E}^{F_*}\big[\dot{\phi}^\star\big(-\underline{\eta}^{-1}\big(k(U) + \underline{\zeta} + \underline{\lambda}' g_s(U)\big)\big)\big] - 1,$$

which implies $\mathbb{E}^{F_*}[\underline{m}_{\delta,\theta}] = 1$ and hence that $\underline{F}_{\delta,\theta}$ is a probability measure. The FOC for $\underline{\lambda}$ is

$$\begin{aligned}
0 &\geq \mathbb{E}^{F_*}\big[\dot{\phi}^\star\big(-\underline{\eta}^{-1}\big(k(U) + \underline{\zeta} + \underline{\lambda}' g_s(U)\big)\big)g_1(U)\big] - P_1,\\
0 &= \mathbb{E}^{F_*}\big[\dot{\phi}^\star\big(-\underline{\eta}^{-1}\big(k(U) + \underline{\zeta} + \underline{\lambda}' g_s(U)\big)\big)g_2(U)\big] - P_2,\\
0 &\geq \mathbb{E}^{F_*}\big[\dot{\phi}^\star\big(-\underline{\eta}^{-1}\big(k(U) + \underline{\zeta} + \underline{\lambda}' g_s(U)\big)\big)g_3(U)\big],\\
0 &= \mathbb{E}^{F_*}\big[\dot{\phi}^\star\big(-\underline{\eta}^{-1}\big(k(U) + \underline{\zeta} + \underline{\lambda}' g_s(U)\big)\big)g_{4s}(U)\big],
\end{aligned}$$

hence (1a)–(1c) and $\mathbb{E}^F[g_{4s}(U, \theta, \gamma)] = 0$ hold at (θ, γ, P) under $\underline{F}_{\delta,\theta}$. The FOC for $\underline{\eta} > 0$ is

$$\begin{aligned}
0 = {} & \mathbb{E}^{F_*}\big[\dot{\phi}^\star\big(-\underline{\eta}^{-1}\big(k(U) + \underline{\zeta} + \underline{\lambda}' g_s(U)\big)\big)\big(-\underline{\eta}^{-1}\big(k(U) + \underline{\zeta} + \underline{\lambda}' g_s(U)\big)\big)\big]\\
& - \mathbb{E}^{F_*}\big[\phi^\star\big(-\underline{\eta}^{-1}\big(k(U) + \underline{\zeta} + \underline{\lambda}' g_s(U)\big)\big)\big] - \delta.
\end{aligned}$$

By Assumption Φ(i), we may write the convex conjugate $\phi^{\star\star}$ of $\phi^\star$ using its Legendre transform:

$$\phi^{\star\star}(x^\star) = x^\star\big(\dot{\phi}^\star\big)^{-1}(x^\star) - \phi^\star\big(\big(\dot{\phi}^\star\big)^{-1}(x^\star)\big)$$

for any $x^\star$ in the range of $\dot{\phi}^\star$ Rockafellar (1970, Theorem 26.4). Setting $x^\star = \dot{\phi}^\star(x)$ and noting that $\phi^{\star\star} = \phi$ holds by the Fenchel–Moreau theorem, we obtain

$$\phi\big(\dot{\phi}^\star(x)\big) = x\dot{\phi}^\star(x) - \phi^\star(x).$$

It follows that we may rewrite the FOC for $\underline{\eta}$ as $\delta = \mathbb{E}^{F^*}[\phi(\underline{m}_{\delta,\theta}(U))]$ and so $\underline{F}_{\delta,\theta} \in \mathcal{N}_\delta$.

Now suppose $\underline{\eta} = 0$. Here we wish to show that $\underline{m}_{\delta,\theta}(u) = \mathbb{1}\{u \in \underline{A}_{\delta,\theta}\}/F_*(\underline{A}_{\delta,\theta})$ induces a distribution that solves the primal problem (20) at θ. As the neighborhood constraint $F \in \mathcal{N}_\delta$ is not binding, the value of the objective must be the same as the optimal value when $\delta = \infty$. In view of Lemma B.2, the value is $F_*\text{-ess}\inf(k(\cdot) + \underline{\lambda}' g_s(\cdot)) - \underline{\lambda}_{12}' P$. We can write problem (22) as a nested optimization:

$$\sup_{\lambda \in \Lambda_s}\left(\sup_{\eta > 0, \zeta \in \mathbb{R}} -\eta \mathbb{E}^{F_*}\left[\phi^\star\left(\frac{k(U) + \zeta + \lambda' g_s(U)}{-\eta}\right)\right] - \eta\delta - \zeta - \lambda_{12}' P\right).$$

At $\lambda = \underline{\lambda}$, the inner problem is the dual of $\inf_{F \in \mathcal{N}_\delta} \mathbb{E}^F[k(U) + \underline{\lambda}' g_s(U) - \underline{\lambda}_{12}' P]$. As $\underline{\eta} = 0$, the constraint $F \in \mathcal{N}_\delta$ is not binding and so the minimizing distribution must be supported on $\underline{A}_{\delta,\theta}$. Finally, by convexity of ϕ, the distribution induced by $\underline{m}_{\delta,\theta}$ minimizes $D_\phi(\cdot \| F_*)$ among all distributions with support $\underline{A}_{\delta,\theta}$.

Now suppose k does not depend on u. By Proposition G.2 of Christensen and Connault (2022), the primal and dual values of (15) are equal and a dual solution exists. By similar arguments to above, $\mathbb{E}^{F_*}[\underline{m}_{\delta,\theta}(U)] = 1$, and (1a)–(1c) and $\mathbb{E}^F[g_{4s}(U, \theta, \gamma)] = 0$ hold at (θ, γ, P) under $\underline{F}_{\delta,\theta}$. Finally, as there exists $F \in \mathcal{N}_\delta$ under which the moment conditions (1a)–(1c) and $\mathbb{E}^F[g_{4s}(U, \theta, \gamma)] = 0$ hold at (θ, γ, P), we must have $D(\underline{F}_{\delta,\theta} \| F_*) \leq D(F \| F_*) \leq \delta$, as required. *Q.E.D.*

E.3. *Proofs for Section 4*

PROOF OF PROPOSITION 4.1: As $\phi_1(x) \leq \bar{a}\phi_2(x)$ for all $x > 0$, we have $D_{\phi_1}(F\|F_*) \leq \bar{a}D_{\phi_2}(F\|F_*)$. Hence, $\mathcal{N}_{\delta,2} \subseteq \mathcal{N}_{\bar{a}\delta,1}$ for each $\delta > 0$. The result follows from this inclusion, noting that $\underline{\kappa}_{\bar{a}\delta,1}$ and $\underline{\kappa}_{\bar{a}\delta,1}$ are both finite because Assumption Φ holds for ϕ_1. *Q.E.D.*

E.4. *Proofs for Section 6*

We first present some preliminary lemmas.

LEMMA E.2: *Suppose that Assumptions* Φ *and* M(i), (v) *hold. Let* $\{(F_n, \theta_n, \gamma_n, P_n)\} \subseteq \mathcal{N}_\delta \times \Theta \times \Gamma \times \mathcal{P}$ *with* $(\gamma_n, P_n) \to (\tilde{\gamma}, \tilde{P}) \in \Gamma \times \mathcal{P}$ *and with* (1a)–(1d) *holding under* F_n *at* $(\theta_n, \gamma_n, P_n)$. *Then: there exists a convergent subsequence* $(F_{n_l}, \theta_{n_l}, \gamma_{n_l}, P_{n_l}) \to (\tilde{F}, \tilde{\theta}, \tilde{\gamma}, \tilde{P}) \in \mathcal{N}_\delta \times \Theta \times \Gamma \times \mathcal{P}$ *along which* $\lim_{l\to\infty} \mathbb{E}^{F_{n_l}}[k(U, \theta_{n_l}, \gamma_{n_l})] = \mathbb{E}^{\tilde{F}}[k(U, \tilde{\theta}, \tilde{\gamma})]$ *and similarly for each entry of* $g_1, \dots, g_4$, *and* (1a)–(1d) *holds under* $\tilde{F}$ *at* $(\tilde{\theta}, \tilde{\gamma}, \tilde{P})$.

PROOF OF LEMMA E.2: Let $m_n = \frac{\mathrm{d}F_n}{\mathrm{d}F_*}$. By Assumption M(v), $\{\theta_n\}$ has a convergent subsequence $\{\theta_{n_l}\}$. As $\{m_{n_l}\}$ is $\|\cdot\|_\phi$-norm bounded (Lemma F.1(ii) of Christensen and Connault (2022)), taking a further subsequence if necessary, we may assume $\{m_{n_l}\}$ is $\mathcal{E}$-weakly convergent to $\tilde{m} \in \mathcal{L}$ (see Appendix F of Christensen and Connault (2022)). By the triangle inequality, Hölder's inequality for Orlicz classes (see (A.1) in Appendix F of Christensen and Connault (2022)), $\mathcal{E}$-weak convergence, and Assumption M(i), we have

$$
\begin{aligned}
&\left|\mathbb{E}^{F_{n_l}}\left[m_{n_l}(U)k(U, \theta_{n_l}, \gamma_{n_l})\right] - \mathbb{E}^{F_*}\left[\tilde{m}(U)k(U, \tilde{\theta}, \tilde{\gamma})\right]\right| \\
&\quad \leq \left|\mathbb{E}^{F_*}\left[\left(m_{n_l}(U) - \tilde{m}(U)\right)k(U, \tilde{\theta}, \tilde{\gamma})\right]\right| + \|m_{n_l}\|_\phi \left\|k(\cdot, \theta_{n_l}, \gamma_{n_l}) - k(\cdot, \tilde{\theta}, \tilde{\gamma})\right\|_\psi \to 0.
\end{aligned}
$$

It follows by similar arguments that

$$
\begin{aligned}
\mathbb{E}^{F_*}\left[\tilde{m}(U)\right] &= 1, \qquad \mathbb{E}^{F_*}\left[\tilde{m}(U)g_1(U, \tilde{\theta}, \tilde{\gamma})\right] \leq \tilde{P}_1, \\
\mathbb{E}^{F_*}\left[\tilde{m}(U)g_2(U, \tilde{\theta}, \tilde{\gamma})\right] &= \tilde{P}_2, \qquad \mathbb{E}^{F_*}\left[\tilde{m}(U)g_3(U, \tilde{\theta}, \tilde{\gamma})\right] \leq 0, \\
\mathbb{E}^{F_*}\left[\tilde{m}(U)g_4(U, \tilde{\theta}, \tilde{\gamma})\right] &= 0.
\end{aligned}
$$

Finally, by Lemma F.1(i) of Christensen and Connault (2022), we have the inequality $\delta \geq \liminf_{l\to\infty} \mathbb{E}^{F_*}[\phi(m_{n_l}(U))] \geq \mathbb{E}^{F_*}[\phi(\tilde{m}(U))]$. *Q.E.D.*

LEMMA E.3: *Suppose that Assumptions* Φ *and* M(i), (iii)–(v) *hold. Then* $\underline{\kappa}_\delta$ *and* $\overline{\kappa}_\delta$ *are finite, and*

$$
\underline{\kappa}_\delta = \inf_{\theta \in \Theta_\delta(\gamma_0, P_0)} \underline{K}_\delta(\theta; \gamma_0, P_0), \qquad \overline{\kappa}_\delta = \sup_{\theta \in \Theta_\delta(\gamma_0, P_0)} \overline{K}_\delta(\theta; \gamma_0, P_0).
$$

PROOF OF LEMMA E.3: We prove the result only for $\underline{\kappa}_\delta$; the result for $\overline{\kappa}_\delta$ follows similarly.

Finiteness of $\underline{\kappa}_\delta$ follows by Assumptions Φ and M(i), (v) and Hölder's inequality for Orlicz classes (see (A.1) in Appendix F of Christensen and Connault (2022)). To simplify notation, we suppress dependence of $\Theta_\delta(\gamma_0, P_0)$ on (γ_0, P_0) in what follows. Suppose there is $\underline{\theta} \notin \Theta_\delta$ with $\underline{K}_\delta(\underline{\theta}; \gamma_0, P_0) < \inf_{\theta\in\Theta_\delta} \underline{K}_\delta(\theta; \gamma_0, P_0)$. Then there must exist $F_{\underline{\theta}} \in \mathcal{N}_\delta$

satisfying (1a)–(1d) at $(\underline{\theta}, \gamma_0, P_0)$. As $\Delta(\underline{\theta}; \gamma_0, P_0) = \delta$, it follows by convexity of ϕ that $F_{\underline{\theta}}$ must be the unique such distribution in $\mathcal{N}_\delta$. Therefore,

$$\mathbb{E}^{F_{\underline{\theta}}}\big[k(U, \underline{\theta}, \gamma_0)\big] = \underline{K}_\delta(\underline{\theta}; \gamma_0, P_0) < \inf_{\theta \in \Theta_\delta} \underline{K}_\delta(\theta; \gamma_0, P_0) \le \inf_{\theta \in \Theta_\delta} \mathbb{E}^{F_\theta}\big[k(U, \theta_0, \gamma_0)\big], \tag{50}$$

where, for each $\theta \in \Theta_\delta$, the distribution F_θ solves $\inf_F D_\phi(F \| F_*)$ subject to (1a)–(1d). Existence of F_θ follows by Lemma G.5(iii) of Christensen and Connault (2022); uniqueness follows by strict convexity of ϕ.

Choose $\{\theta_n\} \subset \Theta_\delta$ with $\theta_n \to \underline{\theta}$ (we may choose such a sequence by Assumption M(iv)). By Lemma E.2, there is a subsequence $\{(\theta_{n_l}, F_{\theta_{n_l}}, \gamma_0, P_0)\}$ with $(\theta_{n_l}, F_{\theta_{n_l}}) \to (\underline{\theta}, \underline{F})$ for some $\underline{F} \in \mathcal{N}_\delta$ for which (1a)–(1d) holds under $\underline{F}$ at $(\underline{\theta}, \gamma_0, P_0)$. It follows by uniqueness of $F_{\underline{\theta}}$ that $\underline{F} = F_{\underline{\theta}}$. By Lemma E.2, we therefore have

$$\inf_{\theta \in \Theta_\delta} \mathbb{E}^{F_\theta}\big[k(U, \theta, \gamma_0)\big] \le \lim_{l \to \infty} \mathbb{E}^{F_{\theta_{n_l}}}\big[k(U, \theta_{n_l}, \gamma_0)\big] = \mathbb{E}^{F_{\underline{\theta}}}\big[k(U, \underline{\theta}, \gamma_0)\big],$$

which contradicts (50). *Q.E.D.*

Define

$$\underline{b}_\delta(\gamma, P) = \inf_{\theta \in \Theta_\delta(\gamma, P)} \underline{K}_\delta(\theta; \gamma, P), \qquad \overline{b}_\delta(\gamma, P) = \inf_{\theta \in \Theta_\delta(\gamma, P)} \overline{K}_\delta(\theta; \gamma, P).$$

LEMMA E.4: *Suppose that Assumptions* Φ *and* M(i)–(v) *hold. Then* $\underline{b}_\delta(\gamma, P)$ *and* $\overline{b}_\delta(\gamma, P)$ *are continuous at* (γ_0, P_0).

PROOF OF LEMMA E.4: We prove the result only for $\underline{b}_\delta$; the result for $\overline{b}_\delta$ follows similarly.

Fix $\varepsilon > 0$. By Lemma E.3, we may choose $\theta_\varepsilon \in \Theta_\delta(\gamma_0, P_0)$ such that $\underline{K}_\delta(\theta_\varepsilon; \gamma_0, P_0) < \underline{b}_\delta(\gamma_0, P_0) + \varepsilon$. By Lemma G.8 of Christensen and Connault (2022) and Assumption M(ii), we have $\Delta(\theta_\varepsilon; \gamma, P) < \delta$ on a neighborhood N of (γ_0, P_0). Moreover, by Lemma G.9(i) of Christensen and Connault (2022) and Assumption M(i)–(iii), we have

$$\underline{K}_\delta(\theta_\varepsilon; \gamma, P) < \underline{K}_\delta(\theta_\varepsilon; \gamma_0, P_0) + \varepsilon$$

on a neighborhood N' of (γ_0, P_0). On $N \cap N'$ we therefore have

$$\underline{b}_\delta(\gamma, P) \le \underline{K}_\delta(\theta_\varepsilon; \gamma, P) < \underline{K}_\delta(\theta_\varepsilon; \gamma_0, P_0) + \varepsilon < \underline{b}_\delta(\gamma_0, P_0) + 2\varepsilon,$$

establishing u.s.c. of $\underline{b}_\delta(\gamma, P)$ at (γ_0, P_0).

To establish l.s.c., suppose there are $\varepsilon > 0$ and $(\gamma_n, P_n) \to (\gamma_0, P_0)$ along which

$$\underline{b}_\delta(\gamma_n, P_n) \le \underline{b}_\delta(\gamma_0, P_0) - 2\varepsilon. \tag{51}$$

Note $\Theta_\delta(\gamma_n, P_n)$ is nonempty for n sufficiently large by Lemma G.8 of Christensen and Connault (2022) and Assumption M(ii), (iii). For each n sufficiently large, choose $\theta_n \in \Theta_\delta(\gamma_n, P_n)$ and $F_n \in \mathcal{N}_\delta$ for which

$$\mathbb{E}^{F_n}\big[k(U, \theta_n, \gamma_n)\big] < \underline{b}_\delta(\gamma_n, P_n) + \varepsilon. \tag{52}$$

By Lemma E.2, there is a subsequence $(F_{n_l}, \theta_{n_l}, \gamma_{n_l}, P_{n_l}) \to (\underline{F}, \underline{\theta}, \gamma_0, P_0)$ for some $\underline{F} \in \mathcal{N}_\delta$ and $\underline{\theta} \in \Theta$, such that (1a)–(1d) holds under $\underline{F}$ at $(\underline{\theta}, \gamma_0, P_0)$, and for which

$$\lim_{l\to\infty} \mathbb{E}^{F_{n_l}}\big[k(U, \theta_{n_l}, \gamma_{n_l})\big] = \mathbb{E}^{\underline{F}}\big[k(U, \underline{\theta}, \gamma_0)\big] \geq \underline{K}_\delta(\underline{\theta}; \gamma_0, P_0).$$

In view of (51) and (52) and Lemma E.3, this implies $\underline{K}_\delta(\underline{\theta}; \gamma_0, P_0) \leq \underline{b}_\delta(\gamma_0, P_0) - \varepsilon = \underline{\kappa}_\delta - \varepsilon$, contradicting the definition of $\underline{\kappa}_\delta$. *Q.E.D.*

PROOF OF THEOREM 6.1: Note that $\underline{\kappa}_\delta = \underline{b}_\delta(\gamma_0, P_0)$ and $\overline{\kappa}_\delta = \overline{b}_\delta(\gamma_0, P_0)$ by Lemma E.3 and $\hat{\underline{\kappa}}_\delta = \underline{b}_\delta(\hat{\gamma}, \hat{P})$ and $\hat{\overline{\kappa}}_\delta = \overline{b}_\delta(\hat{\gamma}, \hat{P})$ by definition. The result now follows by Lemma E.4 and Slutsky's theorem. *Q.E.D.*

LEMMA E.5: *Suppose that Assumptions* Φ *and* M(i), (ii) *hold*, *Condition* S′ *holds at* (θ, γ, P), *and* $\Delta(\theta; \gamma, P) < \delta$. *Then there is a neighborhood* N *of* (θ, γ, P) *such that Condition* S′ *holds at* $(\tilde{\theta}, \tilde{\gamma}, \tilde{P})$ *and* $\Delta(\tilde{\theta}; \tilde{\gamma}, \tilde{P}) < \delta$ *for all* $(\tilde{\theta}, \tilde{\gamma}, \tilde{P}) \in N$.

PROOF OF LEMMA E.5: By Lemma G.7 of Christensen and Connault (2022), Condition S′ holds at all $(\tilde{\theta}, \tilde{\gamma}, \tilde{P})$ in a neighborhood N' of (θ, γ, P). Moreover, $\Delta(\tilde{\theta}; \tilde{\gamma}, \tilde{P}) < \delta$ holds at all $(\tilde{\theta}, \tilde{\gamma}, \tilde{P})$ in a neighborhood N'' of (θ, γ, P) by Lemma G.8 of Christensen and Connault (2022). Set $N = N' \cap N''$. *Q.E.D.*

In the remainder of this subsection, we drop dependence of all quantities on γ.

PROOF OF THEOREM 6.2: We prove the result only for $\underline{b}_\delta$; the result for $\overline{b}_\delta$ follows similarly.

Step 1: We first show $\underline{\Theta}_\delta(P_0)$ is nonempty and compact. For nonemptiness, choose $\{\theta_n\}$ such that $\underline{K}_\delta(\theta_n; P_0) \downarrow \underline{\kappa}_\delta$. Let F_n solve the primal problem for θ_n. By Lemma E.2, there is a subsequence $(F_{n_l}, \theta_{n_l}) \to (\underline{F}, \underline{\theta})$ with $\underline{F} \in \mathcal{N}_\delta$ and $\underline{\theta} \in \Theta$ such that (1a)–(1d) holds under $\underline{F}$ at $(\underline{\theta}, P_0)$ and for which

$$\underline{\kappa}_\delta = \lim_{l\to\infty} \mathbb{E}^{F_{n_l}}\big[k(U, \theta_{n_l})\big] = \mathbb{E}^{\underline{F}}\big[k(U, \underline{\theta})\big].$$

Therefore, $\Theta_\delta(P_0)$ is nonempty. We may deduce by similar arguments that $\underline{\Theta}_\delta(P_0)$ is closed. Compactness now follows by Assumption M(v).

Step 2: We now prove directional differentiability. Let $P_n = P_0 + t_n h_n$ with $t_n \downarrow 0$ and $h_n \to h$. Choose $\underline{\theta} \in \underline{\Theta}_\delta(P_0)$. By Lemma E.5 and Assumption M(iii), (vi), Condition S′ holds at $(\underline{\theta}, P_n)$ and $\Delta(\underline{\theta}; P_n) < \delta$ for n sufficiently large, so by Proposition G.1(iv) of Christensen and Connault (2022), the set $\underline{\Lambda}_\delta(\underline{\theta}; P_n)$ is nonempty and compact for n sufficiently large. It now follows by definition of the objective (13) that

$$\underline{b}_\delta(P_n) - \underline{b}_\delta(P_0) \leq \underline{K}_\delta(\underline{\theta}; P_n) - \underline{K}_\delta(\underline{\theta}; P_0) \leq t_n \times -\underline{\lambda}_{12}' h_n,$$

for all $\underline{\lambda}_{12} \in \underline{\Lambda}_\delta(\underline{\theta}; P_n)$. Finally, by Lemma G.9(ii) of Christensen and Connault (2022), we obtain

$$\limsup_{n\to\infty} \frac{\underline{b}_\delta(P_n) - \underline{b}_\delta(P_0)}{t_n} \leq \max_{\underline{\lambda}_{12} \in \underline{\Lambda}_\delta(\underline{\theta}; P_0)} -\underline{\lambda}_{12}' h.$$

Taking the infimum of both sides over $\underline{\theta} \in \underline{\Theta}_\delta$ yields

$$\limsup_{n\to\infty} \frac{\underline{b}_\delta(P_n) - \underline{b}_\delta(P_0)}{t_n} \leq \inf_{\theta\in\underline{\Theta}_\delta} \max_{\underline{\lambda}_{12}\in\underline{\Lambda}_\delta(\theta;P_0)} -\underline{\lambda}_{12}' h. \tag{53}$$

For the lower bound, choose $\theta_n \in \Theta_\delta(P_n)$ with $\underline{K}_\delta(\theta_n; P_n) \leq \underline{b}_\delta(P_n) + t_n^2$ for all n sufficiently large. Take a subsequence $\{\theta_{n_l}\}$. By Assumption M(v) (taking a further subsequence if necessary), we have $\theta_{n_l} \to \underline{\theta} \in \Theta$. By similar arguments to step 1, we may in fact deduce that $\underline{\theta} \in \underline{\Theta}_\delta$. Reasoning as above, for l sufficiently large we have

$$\underline{b}_\delta(P_{n_l}) - \underline{b}_\delta(P_0) \geq \underline{K}_\delta(\theta_{n_l}; P_{n_l}) - \underline{K}_\delta(\theta_{n_l}; P_0) - t_{n_l}^2 \geq t_{n_l} \times -\underline{\lambda}_{12}' h_{n_l} - t_{n_l}^2,$$

where the final inequality holds for any $\underline{\lambda}_{12} \in \underline{\Lambda}_\delta(\theta_{n_l}; P_0)$. By Assumption M(vii), we may choose $\underline{\lambda}_{12,n_l} \in \underline{\Lambda}_\delta(\theta_{n_l}; P_0)$ for which $-\underline{\lambda}_{12,n_l}' h \to \max_{\underline{\lambda}_{12}\in\underline{\Lambda}_\delta(\underline{\theta};P_0)} -\underline{\lambda}_{12}' h$ as $l \to \infty$. Therefore,

$$\liminf_{l\to\infty} \frac{\underline{b}_\delta(P_{n_l}) - \underline{b}_\delta(P_0)}{t_{n_l}} \geq \max_{\underline{\lambda}_{12}\in\underline{\Lambda}_\delta(\underline{\theta};P_0)} -\underline{\lambda}_{12}' h \geq \inf_{\theta\in\underline{\Theta}_\delta} \max_{\underline{\lambda}_{12}\in\underline{\Lambda}_\delta(\theta;P_0)} -\underline{\lambda}_{12}' h.$$

As the lower bound does not depend on the subsequence $\{\theta_{n_l}\}$, we have

$$\liminf_{n\to\infty} \frac{\underline{b}_\delta(P_n) - \underline{b}_\delta(P_0)}{t_n} \geq \inf_{\theta\in\underline{\Theta}_\delta} \max_{\underline{\lambda}_{12}\in\underline{\Lambda}_\delta(\theta;P_0)} -\underline{\lambda}_{12}' h, \tag{54}$$

proving directional differentiability. Finally, Assumption M(vii) and Lemma G.9(ii) of Christensen and Connault (2022) imply $\theta \mapsto \underline{\Lambda}_\delta(\theta; P_0)$ is continuous at each $\theta \in \underline{\Theta}_\delta$. The set $\underline{\Lambda}_\delta(\theta; P_0)$ is also compact for each $\theta \in \underline{\Theta}_\delta$ by Proposition G.1(iv) of Christensen and Connault (2022). It follows by the maximum theorem that the infima in (53) and (54) can be replaced by minima.

Step 3: In view of step 2, the asymptotic distribution follows by Theorem 2.1 of Shapiro (1991) and the fact that $\sqrt{n}(\hat{P} - P) \to_d N(0, \Sigma)$. *Q.E.D.*

PROOF OF THEOREM 6.3: We verify the conditions of Theorem 3.2 of Fang and Santos (2019). Their Assumptions 1 and 2 hold by Theorem 6.2 and because $\sqrt{n}(\hat{P} - P_0) \to_d N(0, \Sigma)$ with Σ finite, respectively. Their Assumption 3 is assumed directly. Finally, Lemma G.11 of Christensen and Connault (2022) shows that $\widehat{d\underline{b}}_{\delta,P_0}$ and $\widehat{d\overline{b}}_{\delta,P_0}$ satisfy the sufficient conditions for Assumption 4 of Fang and Santos (2019), which is presented in their Remark 3.4. This proves consistency. Coverage of $\mathrm{CS}_{\delta,L}^{1-\alpha}$ and $\mathrm{CS}_{\delta,U}^{1-\alpha}$ follows by continuity of the distribution functions. Coverage of $\mathrm{CS}_\delta^{1-\alpha}$ follows by the Bonferroni inequality. *Q.E.D.*

PROOF OF THEOREM 6.4: We prove the result only for $\mathrm{CS}_\delta^{1-\alpha}$; the results for the other CSs follow similarly. Say that $P_0 \in \mathrm{CS}_{P_0}^{1-\alpha}$ if $P_{10} \leq \hat{P}_{1,U}^{1-\alpha}$ and $P_{20} \in [\hat{P}_{2,L}^{1-\alpha}, \hat{P}_{2,U}^{1-\alpha}]$ both hold. By Lemma E.3, for each $\varepsilon > 0$, we may choose $\underline{\theta}_\varepsilon, \overline{\theta}_\varepsilon \in \Theta_\delta(P_0)$ such that $\underline{K}_\delta(\underline{\theta}_\varepsilon; P_0) < \underline{\kappa}_\delta + \varepsilon$ and $\overline{K}_\delta(\overline{\theta}_\varepsilon; P_0) > \overline{\kappa}_\delta - \varepsilon$. Let $\underline{F}_{\underline{\theta}_\varepsilon}$ and $\overline{F}_{\overline{\theta}_\varepsilon}$ solve problem (15) at $(\underline{\theta}_\varepsilon; P_0)$ and $(\overline{\theta}_\varepsilon; P_0)$, respectively. Whenever $P_0 \in \mathrm{CS}_{P_0}^{1-\alpha}$ holds, $\underline{F}_{\underline{\theta}_\varepsilon}$ and $\overline{F}_{\overline{\theta}_\varepsilon}$ must also satisfy the "relaxed" moment conditions used for computing $\hat{\underline{\kappa}}_{\delta,1-\alpha}$ and $\hat{\overline{\kappa}}_{\delta,1-\alpha}$, so it follows that $\Delta_{\mathrm{cs}}(\underline{\theta}_\varepsilon; \hat{P}_{1-\alpha}) < \delta$

and $\Delta_{\text{cs}}(\overline{\theta}_\varepsilon; \hat{P}_{1-\alpha}) < \delta$. Moreover, as the primal solutions for $\underline{K}_\delta(\underline{\theta}_\varepsilon; P_0)$ and $\overline{K}_\delta(\overline{\theta}_\varepsilon; P_0)$ are feasible for the relaxed problem whenever $P_0 \in \text{CS}^{1-\alpha}_{P_0}$, we have

$$\hat{\underline{\kappa}}_{\delta,1-\alpha} \leq \underline{K}_{\delta,\text{cs}}(\underline{\theta}_\varepsilon; \hat{P}_{1-\alpha}) \leq \underline{K}_\delta(\underline{\theta}_\varepsilon; P_0) < \underline{\kappa}_\delta + \varepsilon,$$

and similarly $\hat{\overline{\kappa}}_{\delta,1-\alpha} > \overline{\kappa}_\delta - \varepsilon$. As ε is arbitrary, we have that $\underline{\kappa}_\delta \geq \hat{\underline{\kappa}}_{\delta,1-\alpha}$ and $\overline{\kappa}_\delta \leq \hat{\overline{\kappa}}_{\delta,1-\alpha}$ hold whenever $P_0 \in \text{CS}^{1-\alpha}_{P_0}$. The desired coverage now follows by (30). *Q.E.D.*

REFERENCES

ANDREWS, ISAIAH, MATTHEW GENTZKOW, AND JESSE M. SHAPIRO (2017): "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *The Quarterly Journal of Economics*, 132 (4), 1553–1592. [8]

——— (2020): "On the Informativeness of Descriptive Statistics for Structural Estimates," *Econometrica*, 6, 2231–2258. [8]

BONHOMME, STEPHANE, AND MARTIN WEIDNER (2022): "Minimizing Sensitivity to Model Misspecification," *Quantitative Economics*, 13 (3), 907–954. [8]

CHRISTENSEN, TIMOTHY, AND BENJAMIN CONNAULT (2022): "Counterfactual Sensitivity and Robustness," arxiv preprint, arXiv:1904.00989 [econ.em]. [1,6,8,12-19]

CSISZÁR, IMRE, AND PAUL C. SHIELDS (2004): "Information Theory and Statistics: A Tutorial," *Communications and Information Theory*, 1 (4), 417–528. [8]

FANG, ZHENG, AND ANDRES SANTOS (2019): "Inference on Directionally Differentiable Functions," *Review of Economic Studies*, 86 (1), 377–412. [19]

HIRIART-URRUTY, JEAN-BAPTISTE, AND CLAUDE LEMARÉCHAL (2001): *Fundamentals of Convex Analysis*. Springer. [13]

LEHMANN, ERICH L., AND GEORGE CASELLA (1998): *Theory of Point Estimation* (2nd Ed.). New York: Springer-Verlag. [1]

MILGROM, PAUL, AND ILYA SEGAL (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70 (2), 583–601. [11]

ROCKAFELLAR, R. TYRRELL (1970): *Convex Analysis*. Princeton University Press. [13,15]

SHAPIRO, ALEXANDER (1991): "Asymptotic Analysis of Stochastic Programs," *Annals of Operations Research*, 30 (1), 169–186. [19]

Co-editor Aviv Nevo handled this manuscript.

Manuscript received 15 April, 2019; final version accepted 18 July, 2022; available online 14 September, 2022.