Econometrica Supplementary Material

SUPPLEMENT TO "USING INSTRUMENTAL VARIABLES FOR INFERENCE ABOUT POLICY RELEVANT TREATMENT PARAMETERS"

(*Econometrica*, Vol. 86, No. 5, September 2018, 1589–1619)

MAGNE MOGSTAD
Department of Economics, University of Chicago, Statistics Norway, and NBER

ANDRES SANTOS
Department of Economics, University of California at Los Angeles

ALEXANDER TORGOVITSKY
Department of Economics, University of Chicago

This supplement contains: (i) a derivation of the MTR weights for the class of linear IV estimands; (ii) a discussion of the computational advantages of the Bernstein polynomials; (iii) an analysis of consistency for the estimator developed in Section 3.2; and (iv) a discussion of how to construct the preliminary estimators used in Section 3.2.

S1. MTR WEIGHTS FOR LINEAR IV ESTIMANDS

IN THIS APPENDIX, we show that linear IV estimands are a special case of our notion of an IV-like estimand. For the purpose of this discussion, we adopt some of the standard textbook terminology regarding "endogenous variables" and "included" and "excluded" instruments in the context of linear IV models without heterogeneity. Consider a linear IV specification with endogenous variables $\widetilde{X}_1$, included instruments $\widetilde{Z}_1$, and excluded instruments $\widetilde{Z}_2$. We let $\widetilde{X} \equiv [\widetilde{X}_1', \widetilde{Z}_1']'$ and $\widetilde{Z} \equiv [\widetilde{Z}_2', \widetilde{Z}_1']'$. We assume that both $E[\widetilde{Z}\widetilde{Z}']$ and $E[\widetilde{Z}\widetilde{X}']$ have full rank.

The variables in $\widetilde{X}$ and $\widetilde{Z}$ can consist of any (measurable) functions of (D, Z), as long as these two full rank conditions are satisfied. Usually, one would expect that $\widetilde{X}_1$ would include D and possibly some interactions between D and other covariates X. The instruments, $\widetilde{Z}$, would usually consist of functions of the vector Z, which contains X, by notational convention. The included portion of $\widetilde{Z}$, that is, $\widetilde{Z}_1$, would typically also include a constant term as one of its components. However, whether $\widetilde{Z}$ is actually "exogenous" in the usual sense of the linear instrumental variables model is not relevant to our definition of an IV-like estimand or the derivation of the weighting expression (9) given in the main text. In particular, OLS is nested as a linear IV specification through the case in which $\widetilde{Z}_1 = [1, D]'$ and both $\widetilde{X}_1$ and $\widetilde{Z}_2$ are empty vectors.

It may be the case that $\widetilde{Z}$ has dimension larger than $\widetilde{X}$, as in "overidentified" linear models. In such cases, a positive definite weighting matrix Π is used to generate instruments $\Pi\widetilde{Z}$ that have the same dimension as $\widetilde{X}$. A common choice of Π is the two-stage least squares weighting $\Pi_{\text{TSLS}} \equiv E[\widetilde{X}\widetilde{Z}']E[\widetilde{Z}\widetilde{Z}']^{-1}$ which has as its rows the first-stage coefficients corresponding to linear regressions of each component of $\widetilde{X}$ on the entire vector $\widetilde{Z}$. We assume that Π is a known or identified non-stochastic matrix with full rank. This covers Π_{TSLS} and the optimal weighting under heteroscedasticity (optimal GMM) as

Magne Mogstad: magne.mogstad@gmail.com
Andres Santos: andres@econ.ucla.edu
Alexander Torgovitsky: torgovitsky@uchicago.edu

 https://doi.org/10.3982/ECTA15463

particular cases given standard regularity conditions. The instrumental variables estimator that uses $\Pi\widetilde{Z}$ as an instrument for $\widetilde{X}$ in a regression of Y on $\widetilde{X}$ has corresponding estimand

$$\beta_{\mathrm{IV},\Pi} \equiv \big(\Pi E\big[\widetilde{Z}\widetilde{X}'\big]\big)^{-1}\big(\Pi E[\widetilde{Z}Y]\big) = E\big[\big(\Pi E\big[\widetilde{Z}\widetilde{X}'\big]\big)^{-1}\Pi\widetilde{Z}Y\big].$$

Each component of this vector is an IV-like estimand with $s(d, z) \equiv e_j'(\Pi E[\widetilde{Z}\widetilde{X}'])^{-1}\Pi\widetilde{z}$, where e_j is the vector whose jth coordinate equals 1, and all other coordinates equal zero.

S2. BERNSTEIN POLYNOMIALS

The kth Bernstein basis polynomial of degree K is a function $b_k^K : [0, 1] \to \mathbf{R}$ defined as

$$b_k^K(u) \equiv \binom{K}{k} u^k(1-u)^{K-k},$$

for $k = 0, 1, \dots, K$. A degree K Bernstein polynomial B is a linear combination of these $K + 1$ basis polynomials, that is,

$$B(u) \equiv \sum_{k=0}^{K} \theta_k b_k^K(u),$$

for some constants $\theta_0, \theta_1, \dots, \theta_K$. As is well-known, any continuous function on $[0, 1]$ can be uniformly well-approximated by a Bernstein polynomial of sufficiently high order.

The shape of B can be constrained by imposing linear restrictions on $\theta_0, \theta_1, \dots, \theta_K$. This computationally appealing property of the Bernstein polynomials has been noted elsewhere by Chak, Madras, and Smith (2005), Chang, Chien, Hsiung, Wen, and Wu (2007), McKay and Ghosh (2011), and Chen, Tamer, and Torgovitsky (2011), among others. The following constraints are especially useful in our framework. Derivations of these properties can be found in Chang et al. (2007) and McKay and Ghosh (2011).

Shape Constraints

S.1 Bounded below by 0: $\theta_k \geq 0$ for all k.
S.2 Bounded above by 1: $\theta_k \leq 1$ for all k.
S.3 Monotonically increasing: $\theta_0 \leq \theta_1 \leq \cdots \leq \theta_K$.
S.4 Concave: $\theta_k - 2\theta_{k+1} + \theta_{k+2} \leq 0$ for $k = 0, \dots, K - 2$.

Each Bernstein basis polynomial is itself an ordinary degree K polynomial. The coefficients on this ordinary polynomial representation (i.e., the power basis representation) can be computed by applying the binomial theorem:

$$b_k^K(u) = \sum_{i=k}^{K} (-1)^{i-k}\binom{K}{i}\binom{i}{k} u^i. \tag{S1}$$

Representation (S1) is useful for computing the terms $\Gamma_d^\star(b_{dk})$ and $\Gamma_{ds}(b_{dk})$ that appear in the finite-dimensional program (22) in the main text. To see this, note for example that, with $d = 1$,

$$\Gamma_{1s}(b_{1k}) \equiv E\left[\int_0^1 b_{1k}(u, Z)\omega_{1s}(u, Z)\,du\right] = E\left[s(1, Z)\int_0^{p(Z)} b_{1k}(u, Z)\,du\right].$$

If $b_{1k}(u, Z) = b_{1k}(u)$ is a Bernstein basis polynomial, then $\int_0^{p(Z)} b_{1k}(u)\,du$ can be computed analytically through elementary calculus using (S1). The result of this integral is a known function of $p(Z)$. The quantity $\Gamma_{1s}(b_{1k})$ is then simply the population average of the product of this known function with $s(1, Z)$, which is also known or identified. This conclusion depends on the form of the weights, and may not hold for all target weights $\omega_d^\star$, although it holds for all of the parameters listed in Table I of the main text. When it does not, one-dimensional numerical integration can be used instead.

S3. CONSISTENCY

We begin with some preliminary notation. In the following, we will view $s \mapsto \beta_s$ as a function of $s \in \mathcal{S}$ and we denote this function as β. We assume that β is uniformly bounded over $\mathcal{S}$, that is, $\beta \in \ell^\infty(\mathcal{S})$. Similarly, for any fixed $m \in \mathcal{M}$, we view $s \mapsto \Gamma_s(m)$ as a function of $s \in \mathcal{S}$, which we denote as $\Gamma(m)$ and assume to also be an element of $\ell^\infty(\mathcal{S})$. We assume that $\mathcal{M}$ is a subset of a Banach space $\mathbf{M}$ and we let $\Gamma : \mathbf{M} \to \ell^\infty(\mathcal{S})$ denote the mapping that returns $\Gamma(m)$ for each $m \in \mathbf{M}$. We assume that we possess estimators $\hat{\beta}$ for β, $\hat{\Gamma}$ for Γ, and $\hat{\Gamma}^\star$ for $\Gamma^\star$. Construction of these estimators is discussed in Appendix S4 of this supplement.

As in the main text, we will limit consideration to the subset of $\mathcal{M}$ that comes within a tolerance of minimizing the estimated counterpart of the observational equivalence condition. In addition, because we allow $\mathcal{M}$ to potentially be infinite-dimensional, for computational reasons we consider finite-dimensional subsets $\mathcal{M}_n \subseteq \mathcal{M}$ that grow "dense" in $\mathcal{M}$. The set $\hat{\mathcal{M}}_{\mathcal{S}}$ is then defined to equal

$$\hat{\mathcal{M}}_{\mathcal{S}} \equiv \Big\{ m \in \mathcal{M}_n : \psi\big(r_n\big[\hat{\Gamma}(m) - \hat{\beta}\big]\big) \leq \inf_{m \in \mathcal{M}_n} \psi\big(r_n\big[\hat{\Gamma}(m) - \hat{\beta}\big]\big) + \kappa_n \Big\}, \tag{S2}$$

where $\psi : \ell^\infty(\mathcal{S}) \to \mathbf{R}_+$ is a convex loss function, $r_n \uparrow \infty$ is the rate of convergence of $\hat{\Gamma}$ and $\hat{\beta}$ (e.g., $\sqrt{n}$), and κ_n is a tolerance parameter that diverges to $+\infty$ more slowly than r_n. Notice that in the main text, $\mathcal{M}_n = \mathcal{M}$, $\psi(b) = \sum_{s \in \mathcal{S}} |b(s)|$, and there is a slight difference in notation in that κ_n in the main text corresponds to κ_n / r_n here, a simplification which was possible because ψ in the main text is homogenous of degree one. This definition generalizes the constraint in (27) of the main text by allowing for more general loss functions, and by allowing $\mathcal{S}$ to have an infinite number of elements. Our estimators $\hat{\underline{\beta}}^\star$ and $\hat{\overline{\beta}}^\star$ are defined as before by minimizing and maximizing $\hat{\Gamma}^\star$ over $\hat{\mathcal{M}}_{\mathcal{S}}$, that is,

$$\hat{\underline{\beta}}^\star \equiv \inf_{m \in \hat{\mathcal{M}}_{\mathcal{S}}} \hat{\Gamma}^\star(m) \quad \text{and} \quad \hat{\overline{\beta}}^\star \equiv \sup_{m \in \hat{\mathcal{M}}_{\mathcal{S}}} \hat{\Gamma}^\star(m). \tag{S3}$$

These estimators of the endpoints provide a set estimator $[\hat{\underline{\beta}}^\star, \hat{\overline{\beta}}^\star]$ of the identified set $[\underline{\beta}^\star, \overline{\beta}^\star]$.

We establish consistency of the endpoint estimators (and hence consistency in the Hausdorff metric of the implied set estimator) under the following assumptions.

ASSUMPTION 1:

(i) *$\mathcal{M}_{\mathcal{S}} \equiv \{m \in \mathcal{M} : \Gamma(m) = \beta\}$ is not empty.*

(ii) *$\Gamma^\star : \mathbf{M} \to \mathbf{R}$ and $\Gamma : \mathbf{M} \to \ell^\infty(\mathcal{S})$ are continuous and linear. The set $\mathcal{M} \subset \mathbf{M}$ is compact in the weak topology.*

(iii) $\psi : \ell^\infty(\mathcal{S}) \to \mathbf{R}_+$ *is lower semicontinuous, convex, and satisfies* $\psi(b) = 0$ *if and only if* $b = 0$, *and* $\sup_{b \in B} \psi(b) < \infty$ *for any bounded set* $B \subset \ell^\infty(\mathcal{S})$.

(iv) $(r_n\{\hat{\Gamma} - \Gamma\}, r_n\{\hat{\beta} - \beta\}) \xrightarrow{d} (\mathbb{G}_\Gamma, \mathbb{G}_\beta)$ *in* $\ell^\infty(\mathcal{M} \times \mathcal{S}) \times \ell^\infty(\mathcal{S})$ *as* $r_n \uparrow \infty$ *for some tight stochastic processes* $(\mathbb{G}_\Gamma, \mathbb{G}_\beta)$. *Also*, $\hat{\Gamma}^\star \xrightarrow{p} \Gamma^\star$ *in* $\ell^\infty(\mathcal{M})$.

(v) $\mathcal{M}_n \subseteq \mathcal{M}$ *and there exists a* $\Pi_n : \mathcal{M} \to \mathcal{M}_n$ *such that* $\Gamma^\star \circ \Pi_n = \Gamma^\star + o(1)$ *in* $\ell^\infty(\mathcal{M})$, *and* $r_n\{\hat{\Gamma} \circ \Pi_n - \hat{\Gamma}\} = o_p(1)$ *in* $\ell^\infty(\mathcal{M} \times \mathcal{S})$.

Assumption 1(ii) requires $\mathcal{M}$ to be a compact subset of $\mathbf{M}$, but only in the weak topology. If $\mathbf{M}$ is reflexive, this only requires $\mathcal{M}$ to be closed and bounded. Assumption 1(iii) places some requirements on ψ, including that it is convex. A general specification of ψ allows us to accommodate problems in which $\mathcal{S}$ is infinite, in which case one might choose ψ to be some form of weighted integral. Assumption 1(iv) places some high-level conditions on the preliminary estimators. In verifying Assumption 1(iv), recall that, by Proposition 3, the IV-like estimands need not include parameters such as the LATE that are potentially weakly identified. Typically, $r_n = \sqrt{n}$, even in certain nonparametric applications, but we use r_n since $\sqrt{n}$ plays no special role in our analysis. We note in particular that Assumption 1(iv) allows the stochastic processes $(\mathbb{G}_\Gamma, \mathbb{G}_\beta)$ to be degenerate, so that r_n could be such that $(r_n\{\hat{\Gamma} - \Gamma\}, r_n\{\hat{\beta} - \beta\}) \xrightarrow{d} (0, 0)$. Assumption 1(v) requires the approximation error introduced from employing $\mathcal{M}_n$ instead of $\mathcal{M}$ to vanish sufficiently quickly, that is, the approximating model $\mathcal{M}_n$ should grow sufficiently quickly with the sample size n.

We now state and prove our consistency result. The proof relies on several auxiliary lemmas, which are stated and proven afterwards.

THEOREM 1: *If Assumption 1 holds*, $\kappa_n \uparrow \infty$, *and* $\kappa_n / r_n \to 0$, *then* $\hat{\underline{\beta}}^\star \xrightarrow{p} \underline{\beta}^\star$ *and* $\hat{\overline{\beta}}^\star \xrightarrow{p} \overline{\beta}^\star$.

PROOF OF THEOREM 1: For any $m \in \mathcal{M}$, define $\hat{\mathbb{G}}_n(m), \hat{\mathbb{G}}_n^{\mathrm{a}}(m) \in \ell^\infty(\mathcal{S})$ as

$$\hat{\mathbb{G}}_n(m) \equiv r_n\big[\big(\hat{\Gamma}(m) - \Gamma(m)\big) - (\hat{\beta} - \beta)\big], \tag{S4}$$

$$\hat{\mathbb{G}}_n^{\mathrm{a}}(m) \equiv r_n\big[\big(\hat{\Gamma}(\Pi_n m) - \Gamma(m)\big) - (\hat{\beta} - \beta)\big]. \tag{S5}$$

Notice that since $\mathcal{M}_\mathcal{S}$ is nonempty by Assumption 1(i) with $\Gamma(m) = \beta$ for all $m \in \mathcal{M}_\mathcal{S}$, one has, by the definition of $\hat{\mathcal{M}}_\mathcal{S}$,

$$\begin{aligned} &P\big(\Pi_n(\mathcal{M}_\mathcal{S}) \subseteq \hat{\mathcal{M}}_\mathcal{S}\big) \\ &\quad = P\Big(\sup_{m \in \mathcal{M}_\mathcal{S}} \psi\big(r_n\big[\hat{\Gamma}(\Pi_n m) - \hat{\beta}\big]\big) \le \inf_{m \in \mathcal{M}_n} \psi\big(r_n\big[\hat{\Gamma}(m) - \hat{\beta}\big]\big) + \kappa_n\Big) \\ &\quad \ge P\Big(\sup_{m \in \mathcal{M}} \psi\big(\hat{\mathbb{G}}_n^{\mathrm{a}}(m)\big) \le \kappa_n\Big), \end{aligned} \tag{S6}$$

where the inequality follows because $\mathcal{M}_\mathcal{S} \subseteq \mathcal{M}$ and $\psi \ge 0$. Together with Lemma 1, (S6) implies that

$$\liminf_{n \to \infty} P\big(\Pi_n(\mathcal{M}_\mathcal{S}) \subseteq \hat{\mathcal{M}}_\mathcal{S}\big) = 1 \tag{S7}$$

since $\kappa_n \uparrow \infty$ by assumption. Using the definitions of $\overline{\beta}^\star$ and $\underline{\beta}^\star$, Assumption 1(v) and (S7) then imply that

$$\overline{\beta}^\star \equiv \sup_{m\in\mathcal{M}_\mathcal{S}} \Gamma^\star(m) = \sup_{m\in\mathcal{M}_\mathcal{S}} \Gamma^\star(\Pi_n m) + o(1) \le \sup_{m\in\hat{\mathcal{M}}_\mathcal{S}} \Gamma^\star(m) + o_p(1) \quad \text{and}$$
$$\underline{\beta}^\star \equiv \inf_{m\in\mathcal{M}_\mathcal{S}} \Gamma^\star(m) = \inf_{m\in\mathcal{M}_\mathcal{S}} \Gamma^\star(\Pi_n m) + o(1) \ge \inf_{m\in\hat{\mathcal{M}}_\mathcal{S}} \Gamma^\star(m) + o_p(1). \tag{S8}$$

In the following, we assume that $r_n \ge 2$, which is without loss of generality since $r_n \uparrow \infty$. Using the convexity of $\psi : \ell^\infty(\mathcal{S}) \to \mathbf{R}$, we obtain

$$\begin{aligned}
\sup_{m\in\hat{\mathcal{M}}_\mathcal{S}} \psi\big(\Gamma(m)-\beta\big) &= \sup_{m\in\hat{\mathcal{M}}_\mathcal{S}} \psi\left(\frac{1}{r_n} r_n\big[\hat{\Gamma}(m)-\hat{\beta}\big] - \frac{(r_n-1)}{r_n}\frac{\hat{\mathbb{G}}_n(m)}{(r_n-1)}\right) \\
&\le \sup_{m\in\hat{\mathcal{M}}_\mathcal{S}} \frac{1}{r_n}\psi\big(r_n\big[\hat{\Gamma}(m)-\hat{\beta}\big]\big) + \frac{(r_n-1)}{r_n}\psi\left(\frac{-\hat{\mathbb{G}}_n(m)}{r_n-1}\right) \\
&\le \frac{1}{r_n}\sup_{m\in\hat{\mathcal{M}}_\mathcal{S}} \psi\big(r_n\big[\hat{\Gamma}(m)-\hat{\beta}\big]\big) + \psi\big(-\hat{\mathbb{G}}_n(m)\big) \\
&\le \frac{\kappa_n}{r_n} + \frac{1}{r_n}\Big(\sup_{m\in\mathcal{M}} \psi\big(\hat{\mathbb{G}}_n^{\mathrm{a}}(m)\big) + \sup_{m\in\mathcal{M}} \psi\big(-\hat{\mathbb{G}}_n(m)\big)\Big),
\end{aligned} \tag{S9}$$

where the second inequality also used the assumption that $\psi(0) = 0$,[1] and the third inequality used the definition of $\hat{\mathcal{M}}_\mathcal{S}$ and $\Pi_n(\mathcal{M}_\mathcal{S}) \subseteq \mathcal{M}_n$. Since $\kappa_n/r_n \to 0$ and $r_n \to \infty$, and the second term of (S9) is $o_p(1)$ by Lemma 1, we conclude that

$$\sup_{m\in\hat{\mathcal{M}}_\mathcal{S}} \psi\big(\Gamma(m)-\beta\big) = o_p(1). \tag{S11}$$

Defining the set $\mathcal{M}_\mathcal{S}(\delta) \equiv \{m \in \mathcal{M} : \psi(\Gamma(m)-\beta) \le \delta\}$ for any $\delta \ge 0$, it follows from (S11) that there exists a sequence of constants $\delta_n \downarrow 0$ such that

$$\liminf_{n\to\infty} P\big(\hat{\mathcal{M}}_\mathcal{S} \subseteq \mathcal{M}_\mathcal{S}(\delta_n)\big) = 1. \tag{S12}$$

We view $\delta \mapsto \mathcal{M}_\mathcal{S}(\delta)$ as a correspondence, and we denote its graph by

$$\mathrm{Gr}\big(\mathcal{M}_\mathcal{S}(\cdot)\big) \equiv \big\{(m,\delta) \in \mathcal{M} \times \mathbf{R}_+ \text{ s.t. } m \in \mathcal{M}_\mathcal{S}(\delta)\big\}. \tag{S13}$$

In Lemma 2, we show that $\mathrm{Gr}(\mathcal{M}_\mathcal{S}(\cdot))$ is closed in $\mathbf{M} \times \mathbf{R}_+$ when $\mathbf{M}$ and $\mathbf{R}_+$ are endowed with the weak and Euclidean topologies, respectively. Since $\mathcal{M}$ is compact in the weak topology by Assumption 1(ii), and $\mathcal{M}_\mathcal{S}(\delta) \subseteq \mathcal{M}$ for all $\delta \in \mathbf{R}_+$, the correspondence $\delta \mapsto \mathcal{M}_\mathcal{S}(\delta)$ is upper hemicontinuous; see, for example, Theorem 17.11 of Aliprantis

[1] In particular, the second inequality follows because with $r_n \ge 2$, convexity implies

$$\psi\left(\frac{-1}{r_n-1}\hat{\mathbb{G}}_n(m) + \frac{r_n-2}{r_n-1}\times 0\right) \le \left(\frac{1}{r_n-1}\right)\psi\big(-\hat{\mathbb{G}}_n(m)\big) + \left(\frac{r_n-2}{r_n-1}\right)\psi(0). \tag{S10}$$

and Border (2006). As a consequence, the maximum of any continuous linear functional $\tilde{\Gamma}:\mathbf{M}\to\mathbf{R}$ over $\mathcal{M}_{\mathcal{S}}(\delta)$ will be upper semicontinuous as a function of δ, that is,

$$\limsup_{n\to\infty}\sup_{m\in\mathcal{M}_{\mathcal{S}}(\delta_n)}\tilde{\Gamma}(m)\le\sup_{m\in\mathcal{M}_{\mathcal{S}}}\tilde{\Gamma}(m);\tag{S14}$$

see, for example, Lemma 17.30 of Aliprantis and Border (2006). The same function will also be lower semicontinuous in δ because $\mathcal{M}_{\mathcal{S}}\subseteq\mathcal{M}_{\mathcal{S}}(\delta_n)$, so that

$$\sup_{m\in\mathcal{M}_{\mathcal{S}}}\tilde{\Gamma}(m)\le\liminf_{n\to\infty}\sup_{m\in\mathcal{M}_{\mathcal{S}}(\delta_n)}\tilde{\Gamma}(m).\tag{S15}$$

The upper and lower semicontinuity in (S14) and (S15) imply continuity, that is,

$$\lim_{n\to\infty}\sup_{m\in\mathcal{M}_{\mathcal{S}}(\delta_n)}\tilde{\Gamma}(m)=\sup_{m\in\mathcal{M}_{\mathcal{S}}}\tilde{\Gamma}(m).\tag{S16}$$

Applying (S16) with $\tilde{\Gamma}=\Gamma^\star$ and $\tilde{\Gamma}=-\Gamma^\star$ yields

$$\begin{aligned}&\bar{\beta}^\star\equiv\sup_{m\in\mathcal{M}_{\mathcal{S}}}\Gamma^\star(m)=\lim_{n\to\infty}\sup_{m\in\mathcal{M}_{\mathcal{S}}(\delta_n)}\Gamma^\star(m)\quad\text{and}\\&\underline{\beta}^\star\equiv-\sup_{m\in\mathcal{M}_{\mathcal{S}}}-\Gamma^\star(m)=-\lim_{n\to\infty}\sup_{m\in\mathcal{M}_{\mathcal{S}}(\delta_n)}-\Gamma^\star(m)=\lim_{n\to\infty}\inf_{m\in\mathcal{M}_{\mathcal{S}}(\delta_n)}\Gamma^\star(m).\end{aligned}\tag{S17}$$

Combined with (S12), this implies that

$$\begin{aligned}&\sup_{m\in\hat{\mathcal{M}}_{\mathcal{S}}}\Gamma^\star(m)\le\sup_{m\in\mathcal{M}_{\mathcal{S}}(\delta_n)}\Gamma^\star(m)+o_p(1)=\bar{\beta}^\star+o_p(1)\quad\text{and}\\&\inf_{m\in\hat{\mathcal{M}}_{\mathcal{S}}}\Gamma^\star(m)\ge\inf_{m\in\mathcal{M}_{\mathcal{S}}(\delta_n)}\Gamma^\star(m)+o_p(1)=\underline{\beta}^\star+o_p(1).\end{aligned}\tag{S18}$$

To conclude, recall the definitions of $\hat{\underline{\beta}}^\star$, $\hat{\bar{\beta}}^\star$, and observe that $\hat{\mathcal{M}}_{\mathcal{S}}\subseteq\mathcal{M}$ together with Assumption 1(iv) imply

$$\begin{aligned}&\left|\sup_{m\in\hat{\mathcal{M}}_{\mathcal{S}}}\Gamma^\star(m)-\hat{\bar{\beta}}^\star\right|\le\sup_{m\in\mathcal{M}}\left|\hat{\Gamma}^\star(m)-\Gamma^\star(m)\right|=o_p(1),\\&\left|\inf_{m\in\hat{\mathcal{M}}_{\mathcal{S}}}\Gamma^\star(m)-\hat{\underline{\beta}}^\star\right|\le\sup_{m\in\mathcal{M}}\left|\hat{\Gamma}^\star(m)-\Gamma^\star(m)\right|=o_p(1).\end{aligned}\tag{S19}$$

Applying the triangle inequality to (S8), (S18), and (S19) yields the claimed consistency of $\hat{\underline{\beta}}^\star$ and $\hat{\bar{\beta}}^\star$ for $\underline{\beta}^\star$ and $\bar{\beta}^\star$. *Q.E.D.*

LEMMA 1: *Under Assumptions 1(iii), 1(iv), and 1(v), it follows that*

$$\sup_{m\in\mathcal{M}}\psi\big(r_n\big[\big(\hat{\Gamma}(m)-\Gamma(m)\big)-(\hat{\beta}-\beta)\big]\big)=O_p(1),\tag{S20}$$

$$\sup_{m\in\mathcal{M}}\psi\big(r_n\big[\big(\hat{\Gamma}(\Pi_n m)-\Gamma(m)\big)-(\hat{\beta}-\beta)\big]\big)=O_p(1).\tag{S21}$$

PROOF OF LEMMA 1: Fix an arbitrary $\varepsilon > 0$, and note that since $(\mathbb{G}_\Gamma, \mathbb{G}_\beta)$ are tight by Assumption 1(iv), there exist compact sets $K_\Gamma \in \ell^\infty(\mathcal{M} \times \mathcal{S})$ and $K_\beta \in \ell^\infty(\mathcal{S})$ with

$$P\big((\mathbb{G}_\Gamma, \mathbb{G}_\beta) \in K_\Gamma \times K_\beta\big) \geq 1 - \varepsilon. \tag{S22}$$

Since K_Γ and K_β are compact, the set

$$\big\{f \in \ell^\infty(\mathcal{S}) : f = g_\Gamma(m, \cdot) - g_\beta \text{ for some } m \in \mathcal{M} \text{ and } (g_\Gamma, g_\beta) \in K_\Gamma \times K_\beta\big\} \tag{S23}$$

is norm bounded in $\ell^\infty(\mathcal{S})$, and so also

$$\bar{c} \equiv \sup_{(g_\Gamma, g_\beta) \in K_\Gamma \times K_\beta} \sup_{m \in \mathcal{M}} \psi\big(g_\Gamma(m, \cdot) - g_\beta\big) < \infty, \tag{S24}$$

given Assumption 1(iii). Moreover, from (S22) and (S24), we can conclude

$$P\Big(\sup_{m \in \mathcal{M}} \psi\big(\mathbb{G}_\Gamma(m, \cdot) - \mathbb{G}_\beta\big) > \bar{c}\Big) \leq \varepsilon. \tag{S25}$$

For any constant $c \geq 0$, let $\mathcal{G}(c)$ be the subset of $\ell^\infty(\mathcal{M} \times \mathcal{S}) \times \ell^\infty(\mathcal{S})$ defined as

$$\mathcal{G}(c) \equiv \Big\{(g_\Gamma, g_\beta) \in \ell^\infty(\mathcal{M} \times \mathcal{S}) \times \ell^\infty(\mathcal{S}) : \sup_{m \in \mathcal{M}} \psi\big(g_\Gamma(m, \cdot) - g_\beta\big) \geq c\Big\}. \tag{S26}$$

Let c be any value for which $\mathcal{G}(c)$ is not empty, and consider any convergent sequence $(g_{\Gamma,n}, g_{\beta,n}) \in \mathcal{G}(c)$ with limit $(g_{\Gamma,0}, g_{\beta,0}) \in \ell^\infty(\mathcal{M} \times \mathcal{S}) \times \ell^\infty(\mathcal{S})$. Then let $\Delta_n(m) \equiv (g_{\Gamma,n}(m, \cdot) - g_{\beta,n}) - (g_{\Gamma,0}(m, \cdot) - g_{\beta,0})$ and observe that, for some sequence $m_n \in \mathcal{M}$ and any sequence $\lambda_n \in [0, 1]$,

$$\begin{aligned}
c &\leq \liminf_{n\to\infty} \sup_{m \in \mathcal{M}} \psi\big(g_{\Gamma,n}(m, \cdot) - g_{\beta,n}\big) \\
&= \liminf_{n\to\infty} \psi\big(g_{\Gamma,n}(m_n, \cdot) - g_{\beta,n}\big) \\
&= \liminf_{n\to\infty} \psi\bigg((1 - \lambda_n)\big(g_{\Gamma,0}(m_n, \cdot) - g_{\beta,0}\big) \\
&\quad + \mathbb{1}[\lambda_n > 0]\lambda_n\bigg[g_{\Gamma,0}(m_n, \cdot) - g_{\beta,0} + \frac{\Delta_n(m_n)}{\lambda_n}\bigg] + \mathbb{1}[\lambda_n = 0]\Delta_n(m_n)\bigg).
\end{aligned} \tag{S27}$$

In particular, we set $\lambda_n = \min\{\|\Delta_n(m_n)\|_\infty, 1\}$, which tends to 0, since $\|\Delta_n(m_n)\|_\infty \to 0$. Applying the convexity of ψ in (S27), we have

$$\begin{aligned}
c &\leq \liminf_{n\to\infty} \bigg[(1 - \lambda_n)\psi\big(g_{\Gamma,0}(m_n, \cdot) - g_{\beta,0}\big) \\
&\quad + \mathbb{1}[\lambda_n > 0]\lambda_n\psi\bigg(g_{\Gamma,0}(m_n, \cdot) - g_{\beta,0} + \frac{\Delta_n(m_n)}{\lambda_n}\bigg)\bigg] \\
&\leq \sup_{m \in \mathcal{M}} \psi\big(g_{\Gamma,0}(m, \cdot) - g_{\beta,0}\big) + \liminf_{n\to\infty} \lambda_n \sup_{\substack{b \in \ell^\infty(\mathcal{S}) \text{ s.t.} \\ \|b\|_\infty \leq \|g_{\Gamma,0}\|_\infty + \|g_{\beta,0}\|_\infty + 1}} \psi(b),
\end{aligned} \tag{S28}$$

where the second inequality follows because $\psi \geq 0$, $m_n \in \mathcal{M}$, $\lambda_n \geq 0$, and $\|\Delta_n(m_n)\|_\infty = \lambda_n$ for n sufficiently large. Since $(g_{\Gamma,0}, g_{\beta,0}) \in \ell^\infty(\mathcal{M} \times \mathcal{S}) \times \ell^\infty(\mathcal{S})$, it follows that $\|g_{\Gamma,0}\|_\infty +$

$\|g_{\beta,0}\|_\infty < \infty$ and hence the supremum in the second term of (S28) is finite by Assumption 1(iii). Therefore, given that $\lambda_n \to 0$, we can conclude that

$$c \leq \sup_{m \in \mathcal{M}} \psi\big(g_{\Gamma,0}(m, \cdot) - g_{\beta,0}\big), \tag{S29}$$

which in turn implies that $\mathcal{G}(c)$ is closed for any $c \geq 0$ for which it is not empty. In particular, since $\mathcal{G}(2\overline{c})$ is either the empty set, or nonempty and closed, Assumption 1(iv) together with the Portmanteau Theorem (e.g., Theorem 1.3.4(iii) of van der Vaart and Wellner (1996)) imply that

$$\begin{aligned}
&\limsup_{n\to\infty} P\Big(\sup_{m\in\mathcal{M}} \psi\big(r_n\big[(\hat{\Gamma} - \Gamma)(m) - (\hat{\beta} - \beta)\big]\big) \geq 2\overline{c}\Big) \\
&\quad \leq P\Big(\sup_{m\in\mathcal{M}} \psi\big(\mathbb{G}_\Gamma(m) - \mathbb{G}_\beta\big) \geq 2\overline{c}\Big) \\
&\quad \leq P\Big(\sup_{m\in\mathcal{M}} \psi\big(\mathbb{G}_\Gamma(m) - \mathbb{G}_\beta\big) > \overline{c}\Big) \leq \varepsilon,
\end{aligned} \tag{S30}$$

where the final inequality was shown in (S25). Since $\varepsilon > 0$ was arbitrary, result (S30) establishes (S20). Result (S21) follows from identical arguments after observing that Assumptions 1(iv) and 1(v) together imply $(r_n\{\hat{\Gamma} \circ \Pi_n - \Gamma\}, r_n\{\hat{\beta} - \beta\}) \xrightarrow{d} (\mathbb{G}_\Gamma, \mathbb{G}_\beta)$ in $\ell^\infty(\mathcal{M} \times \mathcal{S}) \times \ell^\infty(\mathcal{S})$. *Q.E.D.*

LEMMA 2: *Under Assumptions 1(ii) and 1(iii), the set*

$$\mathcal{A} \equiv \big\{(m, \delta) \in \mathcal{M} \times \mathbf{R}_+ : \psi\big(\Gamma(m) - \beta\big) \leq \delta\big\}$$

is closed in $\mathbf{M} \times \mathbf{R}_+$ with respect to the product of the weak and Euclidean topologies.

PROOF OF LEMMA 2: Our proof will show that $\mathbf{M} \times \mathbf{R}_+ \setminus \mathcal{A}$ is open. To this end, take any $(m_0, \delta_0) \in \mathbf{M} \times \mathbf{R}_+$ such that $(m_0, \delta_0) \notin \mathcal{A}$, and consider two mutually exclusive and exhaustive cases.

Case I: Suppose $m_0 \notin \mathcal{M}$. Since $\mathcal{M}$ is assumed to be compact in the weak topology under Assumption 1(ii), Lemma 2.32 in Aliprantis and Border (2006) implies that $\mathbf{M} \setminus \mathcal{M}$ is open in the weak topology. As a consequence, there exists an open neighborhood $\mathcal{N}$ of (m_0, δ_0) in $\mathbf{M} \times \mathbf{R}_+$ such that $m \notin \mathcal{M}$—and hence $(m, \delta) \notin \mathcal{A}$—for every $(m, \delta) \in \mathcal{N}$. This establishes the claim for this case.

Case II: Suppose $m_0 \in \mathcal{M}$. If $(m_0, \delta_0) \notin \mathcal{A}$, then it must be the case that

$$\psi\big(\Gamma(m_0) - \beta\big) > \delta_0. \tag{S31}$$

Let $\eta > 0$ be sufficiently small so that $\psi(\Gamma(m_0) - \beta) > \delta_0 + \eta$, and define the set

$$\mathcal{N}(m_0) \equiv \big\{m \in \mathbf{M} : \psi\big(\Gamma(m) - \beta\big) > \delta_0 + \eta\big\}. \tag{S32}$$

We first establish that $\mathcal{N}(m_0)$ is open. To see this, first note that under the conditions on ψ in Assumption 1(iii), the set $\{b \in \ell^\infty(\mathcal{S}) : \psi(b) \leq \delta_0 + \eta\}$ is norm closed and convex, and therefore also closed in the weak topology (see, e.g., Zeidler (1986, p. 782)). The set $\{b \in \ell^\infty(\mathcal{S}) : \psi(b) > \delta_0 + \eta\}$ is its complement, and so is open in the weak topology of $\ell^\infty(\mathcal{S})$. Also, since $\Gamma : \mathbf{M} \to \ell^\infty(\mathcal{S})$ is linear and norm continuous, it follows that Γ

remains continuous when $\mathbf{M}$ and $\ell^\infty(\mathcal{S})$ are endowed with their weak topologies instead. Since Γ is weakly continuous, these two observations together imply that $\mathcal{N}(m_0)$ is weakly open. Now let

$$\mathcal{N} \equiv \mathcal{N}(m_0) \times \big((\delta_0 - \eta, \delta_0 + \eta) \cap \mathbf{R}_+\big),$$

which is an open set. Moreover, $\psi(\Gamma(m) - \beta) > \delta_0 + \eta > \delta$ for any $(m, \delta) \in \mathcal{N}$, which implies that $\mathcal{N} \subseteq \mathbf{M} \times \mathbf{R}_+ \setminus \mathcal{A}$. This establishes the claim for this case. *Q.E.D.*

S4. FIRST STEP ESTIMATORS

We consider the finite-dimensional problem as in (22) of the main text. Suppose that we have a sample of data denoted as $\{Y_i, D_i, Z_i\}_{i=1}^n$ with $Z_i = [X_i', Z_{0i}']'$. Let $\hat{\Gamma}_d^\star(b_{dk})$ and $\hat{\Gamma}_{ds}(b_{dk})$ denote estimators of $\Gamma_d^\star(b_{dk})$ and $\Gamma_{ds}(b_{dk})$, respectively. For the latter terms, natural estimators are

$$\begin{aligned} \hat{\Gamma}_{ds}(b_{dk}) &\equiv \frac{1}{n}\sum_{i=1}^n \int_0^1 b_{dk}(u, X_i)\hat{\omega}_{ds}(u, Z_i)\, d\mu^\star(u), \\ &\text{where } \hat{\omega}_{0s}(u, z) \equiv \hat{s}(0, z)\mathbb{1}\big[u > \hat{p}(z)\big] \\ &\text{and } \hat{\omega}_{1s}(u, z) \equiv \hat{s}(1, z)\mathbb{1}\big[u \leq \hat{p}(z)\big], \end{aligned} \tag{S33}$$

where $\hat{s}$ is an estimator of s, and $\hat{p}$ is an estimator of the propensity score. An estimator of $\hat{\Gamma}_d^\star(b_{dk})$ can be constructed similarly as

$$\hat{\Gamma}^\star(b_{dk}) \equiv \frac{1}{n}\sum_{i=1}^n \int_0^1 b_{dk}(u, X_i)\hat{\omega}_d^\star(u, Z_i)\, d\mu^\star(u), \tag{S34}$$

where $\hat{\omega}_d^\star$ is an estimator of $\omega_d^\star$, the form of which will depend on the form of the target parameter. Depending on the choices of basis and target parameter, the integrals in (S33) and (S34) can often be evaluated analytically, as discussed in Section S2. Estimators of β_s can be formed from $\hat{s}$ as

$$\hat{\beta}_s \equiv \frac{1}{n}\sum_{i=1}^n \hat{s}(D_i, Z_i)Y_i.$$

REFERENCES

ALIPRANTIS, C. D., AND K. C. BORDER (2006): *Infinite Dimensional Analysis—A Hitchhiker's Guide*. Berlin: Springer-Verlag. [5,6,8]

CHAK, P. M., N. MADRAS, AND B. SMITH (2005): "Semi-Nonparametric Estimation With Bernstein Polynomials," *Economics Letters*, 89 (2), 153–156. [2]

CHANG, I., L.-C. CHIEN, C. A. HSIUNG, C.-C. WEN, AND Y.-J. WU (2007): "Shape Restricted Regression With Random Bernstein Polynomials," in *Lecture Notes—Monograph Series*, Vol. 54, ed. by R. Liu, W. Strawderman, and C.-H. Zhang. Beachwood, Ohio, USA: Institute of Mathematical Statistics, 187–202. [2]

CHEN, X., E. TAMER, AND A. TORGOVITSKY (2011): "Sensitivity Analysis in Semiparametric Likelihood Models," Cowles Foundation Discussion Paper 1836. [2]

MCKAY, C. S., AND S. K. GHOSH (2011): "A Variable Selection Approach to Monotonic Regression With Bernstein Polynomials," *Journal of Applied Statistics*, 38 (5), 961–976. [2]

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. Springer Verlag. [8]

ZEIDLER, E. (1986): “Nonlinear Functional Analysis vol. 1: Fixed-Point Theorems.” [8]

Co-editor Liran Einav handled this manuscript.

Manuscript received 25 June, 2017; final version accepted 31 May, 2018; available online 1 June, 2018.