Econometrica Supplementary Material

SUPPLEMENT TO "A PROJECTION FRAMEWORK FOR TESTING SHAPE RESTRICTIONS THAT FORM CONVEX CONES"

(*Econometrica*, Vol. 89, No. 5, September 2021, 2439–2458)

ZHENG FANG
Department of Economics, Texas A&M University

JUWON SEO
Department of Economics, National University of Singapore

THIS SUPPLEMENT is organized as follows. Appendix B discusses particular shape restrictions with the convex cone property, Appendix C specializes our test to the regular case where $r_n\{\hat{\theta}_n - \theta_0\}$ converges, Appendix D collects additional proofs and auxiliary results, and Appendix E presents additional simulation studies and an empirical application. Appendix F verifies the main assumptions for our examples, Appendix G provides proofs for Appendix C, while Appendix H contains simulation results omitted from the main text and Appendix E, all of which are relegated to the arXiv version of this paper (https://arxiv.org/abs/1910.07689) due to space limitation. For ease of reference, we centralize some notation in the table below.

$a \lesssim b$	For some constant M that is universal in the proof, $a \leq Mb$.
$a^{(j)}$	The jth coordinate of a vector $a \in \mathbf{R}^d$.
$a^{(-j)}$	The vector in $\mathbf{R}^{d-1}$ obtained by deleting the jth entry of $a \in \mathbf{R}^d$.
$a \wedge b$	For $a, b \in \mathbf{R}^d$, $a \wedge b \equiv (\min\{a^{(1)}, b^{(1)}\}, \dots, \min\{a^{(d)}, b^{(d)}\})$.
$a \vee b$	For $a, b \in \mathbf{R}^d$, $a \vee b \equiv (\max\{a^{(1)}, b^{(1)}\}, \dots, \max\{a^{(d)}, b^{(d)}\})$.
$a\Lambda$	For a set Λ in a vector space and $a \in \mathbf{R}$, $a\Lambda \equiv \{a\lambda : \lambda \in \Lambda\}$.
$\Lambda + \theta$	For a set Λ and an element θ in a vector space, $\Lambda + \theta \equiv \{\lambda + \theta : \lambda \in \Lambda\}$.
$\overline{\Lambda}$	For a set Λ in a topological space, $\overline{\Lambda}$ is the closure of Λ.
Φ	The standard normal cdf.
$\|f\|_\infty$	For a function $f : T \to \mathbf{M}^{m\times k}$, $\|f\|_\infty \equiv \sup_{t\in T} \sqrt{\mathrm{tr}(f(t)^\intercal f(t))}$.
$\ell^\infty(T)$	For a nonempty set T, $\ell^\infty(T) \equiv \{f : T \to \mathbf{R} : \|f\|_\infty < \infty\}$.

APPENDIX B: SHAPE RESTRICTIONS AS CONVEX CONES

In this section, we discuss the convex cone property for some shape restrictions and provide details in formulating the linearly constrained quadratic program (25), along with additional references omitted from the main text. For ease of exposition, we shall work with $\mathbf{H} = L^2([0, 1]^d)$ except in Example B.3. In turn, we let $\{z_j\}_{j=1}^k$ be a collection of grid points over $[0, 1]^d$, based on which we approximate the $\|\cdot\|_{\mathbf{H}}$-norms via numerical integration; for example, if $d = 2$, then we may take $\{(s/N, t/N) : s = 0, \dots, N, t = 0, \dots, N\}$ with some suitably chosen N. Finally, let $\vartheta \equiv [\theta(z_1), \dots, \theta(z_k)]^\intercal$ and define $D_k \in \mathbf{M}^{(k-1)\times k}$ as the matrix such that $D_k\vartheta = [\theta(z_2) - \theta(z_1), \dots, \theta(z_k) - \theta(z_{k-1})]^\intercal$.

Zheng Fang: zfang@tamu.edu
Juwon Seo: ecssj@nus.edu.sg

 https://doi.org/10.3982/ECTA17764

EXAMPLE B.1—Monotonicity: Let Λ be the class of nondecreasing functions in $\mathbf{H}$. The convex cone property of Λ is well known—see, for example, Theorem 7.1 in Barlow, Bartholomew, Bremner, and Brunk (1972). To compute the projections onto Λ, let $\theta \in \mathbf{H}$. If $d = 1$, then $\Pi_\Lambda \theta$ may be approximated over $\{z_j\}_{j=1}^k$ by h^* that solves

$$\min_{h \in \mathbf{R}^k} \|h - \vartheta\| \quad \text{s.t. } D_k h \geq 0. \tag{B.1}$$

If $d = 2$, then $\Pi_\Lambda \theta$ is approximated by solving the same problem in (B.1) but subject to $Ah \geq 0$, where $A = [A_1^\intercal, A_2^\intercal]^\intercal$ such that $A_1 h \geq 0$ enforces the monotonicity with respect to the first coordinate and $A_2 h \geq 0$ enforces the second. Computations in higher dimensions are analogous though more complicated.

There is a large literature on estimation by imposing solely shape restrictions, mostly based on the maximum likelihood and least squares principles—see, for example, Han, Wang, Chatterjee, and Samworth (2019) and references therein. Alternatively, monotonicity may be enforced by applying certain operators, such as projection (Mammen, Marron, Turlach, and Wand (2001)) and monotone rearrangement (Chernozhukov, Fernández-Val, and Galichon (2010)), to unconstrained estimators. To retain smoothness, smoothed monotone estimators have also been developed—see, for example, Mammen et al. (2001) and Hall and Huang (2001). Finally, as discussed in the Introduction, an overwhelming majority of existing tests, with the notable exception of Chetverikov (2019), are based on least favorite configurations and limited to univariate settings.

EXAMPLE B.2—Concavity/Convexity: Let Λ be the family of concave functions in $\mathbf{H}$, and $\theta \in \mathbf{H}$ be given. Proposition 3 in Lim and Glynn (2012) implies that Λ is a closed convex cone. If $d = 1$ and $\{z_j\}$ are equidistanced, then the projection $\Pi_\Lambda \theta$ may be approximated over $\{z_j\}_{j=1}^k$ by h^* that solves

$$\min_{h \in \mathbf{R}^k} \|h - \vartheta\| \quad \text{s.t. } D_{k-1} D_k h \leq 0. \tag{B.2}$$

Unfortunately, (B.2) is not readily generalizable to multivariate settings. As formalized in Kuosmanen (2008), the projection $\Pi_\Lambda \theta$ may be approximated by the map $z \mapsto \min_{j=1}^k \{a_j^* + z^\intercal b_j^*\}$, where $\{a_j^*, b_j^*\}_{j=1}^k$ solve the problem

$$\min_{a_i \in \mathbf{R}, b_i \in \mathbf{R}^d} \left\{ \sum_{i=1}^k \left[\theta(z_i) - a_i - b_i^\intercal z_i\right]^2 \right\}^{1/2}$$
$$\text{s.t. } a_i + b_i^\intercal z_i \leq a_j + b_j^\intercal z_i \quad \text{for } i, j = 1, 1, \dots, k. \tag{B.3}$$

Note that the number of effective constraints in (B.3) is $k(k-1)$. An attractive feature of the formulation in (B.3) is that the joint test of monotonicity and concavity amounts to the same problem but subject to the *additional* constraints $b_j \geq 0$ for all j.

As with monotonicity, there are three general estimation strategies: estimation under solely convexity/concavity (Han and Wellner (2016)), smoothing (Hall and Huang (2001), Mammen et al. (2001)), and post-processing (Chen, Chernozhukov, Fernández-Val, Kostyshak, and Luo (2020)). The studies on testing are less extensive than monotonicity, and in particular share the features that most of them are conservative and/or limited to univariate settings—see the Introduction for references. Chen and Kato (2019) developed a bootstrap version of Abrevaya and Jiang (2005), which, despite its nonconservativeness, is computationally intensive to implement. Song, Chen, and Kato (2020) proposed an "incomplete" version of Chen and Kato (2019), which, as documented in their simulations, "is consistently on the conservative side."

EXAMPLE B.3—Slutsky Restriction: For simplicity, let us consider the setup of Example 2.4, and note that Λ being a convex cone is well known in linear algebra (see also Aguiar and Serrano (2017), p. 195). The projection $\Pi_\Lambda\theta$ of $\theta \in \mathbf{H}$ onto Λ admits a closed form expression. Specifically, for $\theta_\sigma \equiv (\theta + \theta^\intercal)/2$ the symmetric part of θ, let $\theta_\sigma(t) = \mathbb{U}(t)\mathbb{S}(t)\mathbb{U}(t)^\intercal$ where $\mathbb{S}(t) \equiv \text{diag}(\lambda_1(t), \dots, \lambda_{d_q}(t))$ and $\mathbb{U}$ satisfies $\mathbb{U}(t)\mathbb{U}(t)^\intercal = I_{d_q}$ for all $t \equiv (p, y)$. Here, $\text{diag}(a_1, \dots, a_{d_q}) \in \mathbf{M}^{d_q \times d_q}$ is the diagonal matrix whose diagonal entries are $a_1, \dots, a_{d_q}$. In turn, letting $\mathbb{S}_-(t) \equiv \text{diag}(\lambda_{1,-}(t), \dots, \lambda_{d_q,-}(t))$ with $\lambda_{j,-} \equiv \min\{\lambda_j, 0\}$ for all $j = 1, \dots, d_q$, we have: for all $t \equiv (p, y)$,

$$(\Pi_\Lambda\theta)(t) = \mathbb{U}(t)\mathbb{S}_-(t)\mathbb{U}(t)^\intercal. \tag{B.4}$$

Hoderlein (2011) and Dette, Hoderlein, and Neumeyer (2016) developed tests for fixed (p, y). As theory predicts the restriction for all (p, y), one may employ these tests by discretizing the data. However, discretization entails an extra tuning parameter whose choice may be a delicate matter. Moreover, Dette, Hoderlein, and Neumeyer (2016)'s test, as the authors noted, is in general conservative, while validity of Hoderlein (2011)'s test has not been formally proven—see Chen and Fang (2019) for the challenges involved in a related but different problem.

EXAMPLE B.4—Supermodularity: Let $d \geq 2$ and $\Lambda \subset \mathbf{H}$ be the set of supermodular functions, that is, $f \in \Lambda$ if and only if, for any $y, z \in [0, 1]^d$,

$$f(y) + f(z) \leq f(y \vee z) + f(y \wedge z). \tag{B.5}$$

By Lemma 2.6.1 in Topkis (1998), Λ is a closed convex cone. Consider $d = 2$ first, and pick $\theta \in \mathbf{H}$. For simplicity, let ϑ be the vectorization of the matrix $\Theta^\intercal$ such that the (i, j)th entry of Θ is $\theta(i/n, j/n)$, for $i, j = 0, \dots, N$. Then, following Beresteanu (2007), computing $\Pi_\Lambda(\theta)$ amounts to solving: for $k \equiv N + 1$,

$$\min_{h \in \mathbf{R}^{k^2}} \|h - \vartheta\| \quad \text{s.t. } (D_k \otimes D_k)\vartheta \geq 0, \tag{B.6}$$

where the number of constraints is N^2. If $d \geq 3$, then the equivalence of supermodularity and pairwise supermodularity (Topkis (1998)) implies that each pair of covariates must satisfy the constraint in (B.6). Despite its importance in economics, econometric studies are rather limited. Chetverikov (2019)'s test on monotonicity, as the author noted, may be adapted to handle supermodularity. Interestingly, separability of a function θ_0 in its arguments is equivalent to θ_0 being supermodular and submodular (Topkis (1998), Theorem 2.6.4), and thus also shares the convex cone property.

EXAMPLE B.5—Nonnegativity: Let $\Lambda \subset \mathbf{H}$ be the family of nonnegative functions, and $\theta \in \mathbf{H}$. As is well known (see, e.g., Deutsch (2012), p. 65), Λ is a convex cone and the projection of θ onto Λ is given by: for any $t \in [0, 1]^d$,

$$(\Pi_\Lambda\theta)(t) = \max\{\theta(t), 0\}. \tag{B.7}$$

There are numerous studies on nonnegativity, such as (conditional) moment inequalities characterizing choice probabilities or payoffs (Pakes, Porter, Ho, and Ishii (2015)), (conditional) stochastic dominance for ordering uncertain prospects (Linton, Song, and Whang (2010)), Lorenz dominance for measuring inequality (Sun and Beare (2019)), and inequalities constraining equilibrium bid distributions or winning probabilities in auction models (Guerre, Perrigne, and Vuong (2009)).

EXAMPLE B.6—Joint Restrictions: Shape restrictions often arise simultaneously in economics—see, for example, Aït-Sahalia and Duarte (2003). Existing tests, however, mostly focus on particular restrictions, and a multiple testing based on these tests is generally conservative. In contrast, our framework allows for jointly testing restrictions as intersections of convex cones remain convex cones. For example, letting Λ consist of monotonic and supermodular functions leads to a joint test of monotonicity and supermodularity, for which the constraints in the quadratic program are obtained by vertically stacking the individual A matrices in (25).

We conclude by making a few remarks. First, just as the t-test is inconsistent in testing $\mathrm{H}_0 : \theta_0 < 0$ vs. $\mathrm{H}_1 : \theta_0 \geq 0$ for a mean parameter θ_0, a level α test for a "strict" restriction such as strict concavity is generally inconsistent. Assumption 3.1(i) ensures that "equality" is included under H_0. We note that closedness of Λ (in $\mathbf{H}$) may require identifying shape restrictions through equivalent classes; for example, for monotonicity in $L^2([0, 1])$, we have $\theta \in \Lambda$ if $\theta = \vartheta$ almost everywhere for some monotonic function $\vartheta \in L^2([0, 1])$. Second, the convex cone property depends on a proper choice of the parameter; for example, the range restriction $\Lambda_0 \equiv \{f \in L^2([0, 1]) : f(x) \leq 1 \ \forall x \in [0, 1]\}$ is not a convex cone, but we may consider $\theta_0 \equiv 1 - f_0$ if f_0 is the original parameter, and define $\Lambda \equiv \{g \in L^2([0, 1]) : g(x) \geq 0 \ \forall x \in [0, 1]\}$. It may be necessary to choose a parameter that involves some derivative(s); for example, in Example 2.4, Assumption 3.1 holds for the Slutsky matrix θ_0 (which involves derivatives of g_0) but not for g_0 itself. Third, while Section 2.1 is centered around regression models as a result of their popularity and the space limitation, our framework is also applicable to other settings, such as those concerning densities/distributions, including monotonicity of densities (Fang (2019)), likelihood ratio ordering (Beare and Moon (2015)), and stochastic monotonicity (Lee, Linton, and Whang (2009)). Note that, in the presence of covariates (as controls), some of these results are not directly applicable. Alternatively, one may apply our test in structural models where shape restrictions arise as testable implications—see, for example, Pinkse and Schurter (2019). Finally, in implementing our test, one may be prompted to ignore some features of θ_0 that coexist with the shape restriction but invalidate Assumption 3.1 when incorporated. For example, if $\theta_0 \in L^2([0, 1])$ and $0 \leq \theta_0(x) \leq 1$ with $\theta_0(0) = 0$ and $\theta_0(1) = 1$, then testing monotonicity on θ_0 without the equality constraints may result in power loss—note that Theorem 1.6 in Barlow et al. (1972) implies that projection preserves the range.

APPENDIX C: THE SPECIAL CASE

The aim of this section is twofold. First, we show that, when $\hat{\theta}_n$ admits an asymptotic distribution, Assumptions 3.2 and 3.3 can be simplified to conditions that may be more familiar to practitioners. Second, we expound the point that, even in this special case, our test improves upon existing inferential methods along several dimensions.

We need additional notation and concepts. Specifically, define

$$\mathrm{BL}_1(\mathbf{H}) = \big\{f : \mathbf{H} \to \mathbf{R} : \big|f(x)\big| \leq 1, \big|f(x) - f(y)\big| \leq \|x - y\|_{\mathbf{H}} \text{ for all } x, y \in \mathbf{H}\big\}, \tag{C.1}$$

and denote the tangent cone T_{θ_P} of Λ at $\theta_P \in \Lambda \subset \mathbf{H}$ by $T_{\theta_P} \equiv \overline{\bigcup_{\alpha>0} \alpha\{\Lambda - \theta_P\}}$. In turn, define a map $\phi'_{\theta_P} : \mathbf{H} \to \mathbf{R}$ by $\phi'_{\theta_P}(h) \equiv \|h - \Pi_{T_{\theta_P}} h\|_{\mathbf{H}}$, which is in fact the so-called Hadamard directional derivative of ϕ. Since only the functional form of ϕ'_{θ_P} is relevant to us here, we refer the reader to Fang and Santos (2019) for detailed discussions of this concept.

We next impose an analog of Assumption 3.2 as follows.

ASSUMPTION C.1: (*i*) $\sup_{f\in \mathrm{BL}_1(\mathbf{H})} |E_P[f(r_n\{\hat{\theta}_n - \theta_P\})] - E[f(\mathbb{G}_P)]| = o(1)$ *uniformly in* $P \in \mathbf{P}$ *for an estimator* $\hat{\theta}_n : \{X_i\}_{i=1}^n \to \mathbf{H}$; (*ii*) $\hat{\mathbb{G}}_n \in \mathbf{H}$ *is a bootstrap estimator satisfying* $\sup_{f\in \mathrm{BL}_1(\mathbf{H})} |E[f(\hat{\mathbb{G}}_n)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_P)]| = o_p(1)$ *uniformly in* $P \in \mathbf{P}$.

Assumption C.1 simply requires uniform convergence in distribution and uniform validity of bootstrap, which may be verified by appealing to existing results (Giné and Zinn (1991), Sheehy and Wellner (1992)). Assumption C.1 in fact automatically implies a weak version of Assumption 3.2 obtained by replacing the independence condition in Assumption 3.2(ii) with an asymptotical independence condition characterized as: uniformly in $P \in \mathbf{P}$,

$$\sup_{f\in \mathrm{BL}_1(\mathbf{H})} \big|E\big[f(\bar{\mathbb{Z}}_{n,P})|\{X_i\}_{i=1}^n\big] - E\big[f(\bar{\mathbb{Z}}_{n,P})\big]\big| = o_p(1). \tag{C.2}$$

PROPOSITION C.1: *Let* $\mathbf{H}$ *be a separable Hilbert space. If Assumption* C.1 *holds, then* (*i*) *the above weak version of Assumption* 3.2 *follows, with* $c_n = 1$ *and* $\mathbb{Z}_{n,P}$ *copies of* $\mathbb{G}_P$,[1] *and* (*ii*) $\psi_{\kappa_n,P}(\mathbb{Z}_{n,P}) \xrightarrow{L} \phi'_{\theta_P}(\mathbb{G}_P)$ *for all* $P \in \mathbf{P}_0$, *provided* $\kappa_n \to \infty$.

Since our results in Section 3.1 remain valid under the weak version of Assumption 3.2 by Lemma G.2, Proposition C.1(i) implies that our test is applicable to this special case subject to Assumptions 3.1, C.1, and 3.3. Proposition C.1(ii) further implies that, if $\kappa_n \to \infty$, then the coupling variables $\{\psi_{\kappa_n,P}(\mathbb{Z}_{n,P})\}$ admit a limit in distribution. Therefore, one may replace Assumption 3.3(iii) with $c_P(1-\alpha-\varpi) \geq c_P(0.5) + \varsigma$ for some $\varsigma > 0$ and $c_P(\tau)$ the τ-quantile of $\phi'_{\theta_P}(\mathbb{G}_P)$, which is effectively the same as requiring that $\phi'_{\theta_P}(\mathbb{G}_P)$ be continuous and strictly increasing at $c_P(1-\alpha)$ as imposed in Fang and Santos (2019). In turn, Assumption 3.3(iv) then reduces to $c_n = O(1)$ and so the coupling order $o_p(c_n)$ becomes $o_p(1)$.

We next compare our test to some existing ones. Employing a generalized Delta method, Fang and Santos (2019) obtained that, under Assumptions 3.1(i) and C.1(i),

$$r_n\phi(\hat{\theta}_n) \xrightarrow{L} \phi'_{\theta_P}(\mathbb{G}_P) \equiv \|\mathbb{G}_P - \Pi_{T_{\theta_P}}\mathbb{G}_P\|_{\mathbf{H}}, \tag{C.3}$$

for each $P \in \mathbf{P}_0$. Exploiting the insight that the limit in (C.3) is the composition of ϕ'_{θ_P} and $\mathbb{G}_P$, Fang and Santos (2019) then showed that a general consistent bootstrap of the limit in (C.3) may be obtained by constructing $\hat{\phi}'_n(\hat{\mathbb{G}}_n)$, a composition of a suitably consistent estimator $\hat{\phi}'_n$ of ϕ'_{θ_P} with a consistent bootstrap $\hat{\mathbb{G}}_n$ for $\mathbb{G}_P$.

While the bootstrap $\hat{\mathbb{G}}_n$ is often straightforward to construct as in Section 3.1, obtaining a suitable estimator $\hat{\phi}'_n$ turns out to be nontrivial. The challenge involved may be

[1] We are indebted to Andres Santos for suggesting this result and sketching the proof.

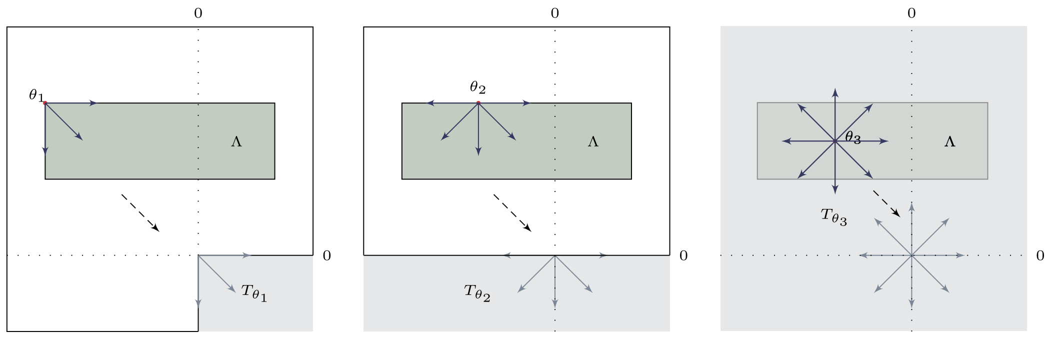

FIGURE C.1.—The tangent cone T_θ depends on θ discontinuously. As θ moves from the corner at θ_1 but still stays on the boundary at θ_2, T_θ changes from the fourth orthant T_{θ_1} to the half plane T_{θ_2}. In turn, as θ moves into the interior at θ_3 from θ_2, T_θ becomes the entire plane T_{θ_3}.

understood in view of the discontinuity of the cone-valued map $\theta \mapsto T_\theta$, as illustrated in Figure C.1. In this regard, Fang and Santos (2019) proposed the following concrete estimator: for any $h \in \mathbf{H}$ and some $\kappa_n \uparrow \infty$,

$$\hat{\phi}'_n(h) = \sup_{\theta\in\Lambda: r_n\|\theta-\Pi_\Lambda\hat{\theta}_n\|_{\mathbf{H}}\leq\kappa_n} \|h - \Pi_{T_\theta} h\|_{\mathbf{H}}. \tag{C.4}$$

Evaluating the supremum in (C.4), however, may be computationally costly as it entails estimation of a local parameter space, that is, T_{θ_P}. Alternatively, one may employ a numerical estimator following Hong and Li (2018), but there are no data-driven procedures to date for selecting the step size (needed to carry out the numerical differentiation). This raises substantive concerns because the resulting bootstrap may be sensitive to the choice of the step size, as documented in Masten and Poirier (2021) and Chen and Fang (2019). One may also appeal to the m out of n bootstrap or subsampling, but the choice of the sub-sample size may be difficult, among other issues—see Remark 3.1 in Chen and Fang (2019).

While our development is undertaken outside the scope of the Delta method, there is an intriguing connection to the general theory of Fang and Santos (2019), as we now flesh out. To this end, recall our bootstrap statistic $\hat{\psi}_{\kappa_n}(\hat{\mathbb{G}}_n)$.

PROPOSITION C.2: *Let Assumptions* 3.1 *and* C.1(*i*) *hold. If* $\kappa_n \to \infty$ *and* $\kappa_n/r_n \to 0$, *then it follows that* $\hat{\psi}_{\kappa_n}(h) \xrightarrow{p} \phi'_{\theta_P}(h)$ *for any* $h \in \mathbf{H}$ *and* $P \in \mathbf{P}_0$.

Since $h \mapsto \hat{\psi}_{\kappa_n}(h)$ is Lipschitz continuous, Proposition C.2 implies that $\hat{\psi}_{\kappa_n}$ is consistent in estimating ϕ'_{θ_P} in the sense of Fang and Santos (2019)—see their Remark 3.4. Therefore, when $r_n\{\hat{\theta}_n - \theta_P\}$ converges in distribution, our test is effectively the test of Fang and Santos (2019) (with respect to their general theory), but based on a derivative estimator that is new and simpler relative to (C.4). We stress that the computational advantage hinges on the convex cone property but not convexity alone. In accord with previous discussions, Proposition C.2 also shows that, by letting $\kappa_n \to \infty$ (in addition to $\kappa_n/r_n \to 0$), our test is not conservative in the sense that it is pointwise (in P) asymptotically exact as in Fang and Santos (2019).

APPENDIX D: MORE PROOFS AND AUXILIARY RESULTS

PROOF OF PROPOSITION 3.1: Let $\hat{F}_n$ be the conditional cdf of $\|\hat{\mathbb{G}}_n\|_{\mathbf{H}}$ given $\{X_i\}_{i=1}^n$, and let $F_{n,P}$ be the cdf of $\|\mathbb{Z}_{n,P}\|_{\mathbf{H}}$. Note that $F_{n,P}$ is also the cdf of $\|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}}$ since $\bar{\mathbb{Z}}_{n,P}$ is a copy of $\mathbb{Z}_{n,P}$ by Assumption 3.2(ii). As a first step, we show that $\hat{F}_n$ and $F_{n,P}$ are suitably close in probability. By Assumptions 3.2(ii), we obtain

$$P\big(\|\hat{\mathbb{G}}_n - \bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}} > \delta_n\big) = o(1), \tag{D.1}$$

for some $\delta_n = o(c_n)$, uniformly in $P \in \mathbf{P}$. Fix $\eta > 0$. By Markov's inequality, Fubini's theorem, and result (D.1), we may in turn have that, uniformly in $P \in \mathbf{P}$,

$$\begin{aligned} &P\big(P\big(\big|\|\hat{\mathbb{G}}_n\|_{\mathbf{H}} - \|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}}\big| > \delta_n|\{X_i\}_{i=1}^n\big) > \eta\big) \\ &\quad\le \frac{1}{\eta}P\big(\big|\|\hat{\mathbb{G}}_n\|_{\mathbf{H}} - \|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}}\big| > \delta_n\big) \le \frac{1}{\eta}P\big(\|\hat{\mathbb{G}}_n - \bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}} > \delta_n\big) = o(1). \end{aligned} \tag{D.2}$$

Since $\eta > 0$ is arbitrary, we may therefore conclude from (D.2) that

$$P\big(\big|\|\hat{\mathbb{G}}_n\|_{\mathbf{H}} - \|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}}\big| > \delta_n|\{X_i\}_{i=1}^n\big) = o_p(1). \tag{D.3}$$

By simple manipulations, we then have: for all $t \in \mathbf{R}$,

$$\begin{aligned} \hat{F}_n(t) - F_{n,P}(t) &= P\big(\|\hat{\mathbb{G}}_n\|_{\mathbf{H}} \le t|\{X_i\}_{i=1}^n\big) - P\big(\|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}} \le t\big) \\ &\le P\big(\|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}} \le t + \delta_n|\{X_i\}_{i=1}^n\big) - P\big(\|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}} \le t\big) \\ &\quad + P\big(\big|\|\hat{\mathbb{G}}_n\|_{\mathbf{H}} - \|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}}\big| > \delta_n|\{X_i\}_{i=1}^n\big) \\ &\le P\big(\big|\|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}} - t\big| \le \delta_n\big) + o_p(1), \end{aligned} \tag{D.4}$$

uniformly in $P \in \mathbf{P}$, where the second inequality follows by $\bar{\mathbb{Z}}_{n,P}$ being independent of $\{X_i\}_{i=1}^n$ (so that $P(\|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}} \le t) = P(\|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}} \le t|\{X_i\}_{i=1}^n)$) and result (D.3). By analogous arguments, we also have: for all $t \in \mathbf{R}$,

$$F_{n,P}(t) - \hat{F}_n(t) \le P\big(\big|\|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}} - t\big| \le \delta_n\big) + o_p(1), \tag{D.5}$$

uniformly in $P \in \mathbf{P}$. Combining results (D.4) and (D.5), we arrive at:

$$\big|\hat{F}_n(t) - F_{n,P}(t)\big| \le P\big(\big|\|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}} - t\big| \le \delta_n\big) + o_p(1), \tag{D.6}$$

for all $t \in \mathbf{R}$, uniformly in $P \in \mathbf{P}$, where the $o_p(1)$ term does not involve t.

Let $m_{n,P}$ be the median of $F_{n,P}$. By Assumptions 3.3(i) and 3.4, we may apply Lemma D.3 to conclude that, for any $t > m_{n,P} + \delta_n$,

$$\begin{aligned} P\big(\big|\|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}} - t\big| \le \delta_n\big) &= \int_{t-\delta_n}^{t+\delta_n} F'_{n,P}(r)\,\mathrm{d}r \\ &\le \int_{t-\delta_n}^{t+\delta_n} \frac{2r - m_{n,P}}{(r - m_{n,P})^2}\,\mathrm{d}r \le 2\delta_n \frac{2(t-\delta_n) - m_{n,P}}{(t - \delta_n - m_{n,P})^2}, \end{aligned} \tag{D.7}$$

where the second inequality (in the second line) follows by $r \mapsto (2r - m_{n,P})/(r - m_{n,P})^2$ being decreasing on $(m_{n,P}, \infty)$. Since $\bar{\mathbb{Z}}_{n,P}$ is a copy of $\mathbb{Z}_{n,P}$, by Kwapień (1994) and Assumption 3.3(ii), we then have: for some constant $\zeta > 0$,

$$\sup_{P \in \mathbf{P}} m_{n,P} \leq \sup_{P \in \mathbf{P}} E_P\big[\|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}}\big] \leq \zeta < \infty. \tag{D.8}$$

Since $\delta_n = o(c_n) = o(1)$ due to $c_n = O(1)$, we obtain from results (D.7) and (D.8) that, for all n large so that $\delta_n \leq 1$ and for all $t \geq \zeta + 2$,

$$\begin{aligned} &\sup_{P \in \mathbf{P}_0} P\big(\big|\|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}} - t\big| \leq \delta_n\big) \\ &\quad \leq 2\delta_n \left\{ \frac{2}{t - \delta_n - m_{n,P}} + \frac{m_{n,P}}{(t - \delta_n - m_{n,P})^2} \right\} \leq 2\delta_n(2 + \zeta). \end{aligned} \tag{D.9}$$

Exploiting $\delta_n = o(1)$ again, we may combine (D.6) and (D.9) to conclude

$$\big|\hat{F}_n(t) - F_{n,P}(t)\big| = o_p(1), \tag{D.10}$$

uniformly in $P \in \mathbf{P}_0$ and $t \in [\zeta + 2, \infty)$.

Next, we aim to prove the first claim of the proposition. Let $M > \zeta + 2$ be any large constant. By Lemma 6.10 in Aliprantis and Border (2006), we have

$$\|\mathbb{Z}_{n,P}\|_{\mathbf{H}} = \sup_{h \in \mathbf{H}_1} \langle h, \mathbb{Z}_{n,P} \rangle_{\mathbf{H}}, \tag{D.11}$$

where $\mathbf{H}_1 \equiv \{h \in \mathbf{H} : \|h\|_{\mathbf{H}} \leq 1\}$. In turn, it follows from result (D.11) that

$$\begin{aligned} F_{n,P}(M) &= P\Big(\sup_{h' \in \mathbf{H}_1} \langle h', \mathbb{Z}_{n,P} \rangle_{\mathbf{H}} \leq M\Big) \\ &\leq P\big(\langle h, \mathbb{Z}_{n,P} \rangle_{\mathbf{H}} \leq M\big) = \Phi\bigg(\frac{M}{\sigma_{n,P}(h)}\bigg), \end{aligned} \tag{D.12}$$

for all $h \in \mathbf{H}_1$, where $\sigma^2_{n,P}(h) \equiv E[\langle h, \mathbb{Z}_{n,P} \rangle^2_{\mathbf{H}}]$. By the definition of $\bar{\sigma}^2_{n,P}$, we may then select a sequence $\{h_j\}$ in $\mathbf{H}_1$ such that $\sigma^2_{n,P}(h_j) \to \bar{\sigma}^2_{n,P}$ as $j \to \infty$. By continuity of $\sigma \mapsto \Phi(M/\sigma)$, we thus obtain from (D.12) that

$$F_{n,P}(M) \leq \Phi\bigg(\frac{M}{\bar{\sigma}_{n,P}}\bigg), \tag{D.13}$$

for any $P \in \mathbf{P}_0$ and n. By Assumption 3.4, we may select some constant $\underline{\sigma} > 0$ such that $\inf_{P \in \mathbf{P}_0} \bar{\sigma}_{n,P} > \underline{\sigma}$ for large n. By result (D.13), we then must have

$$F_{n,P}(M) \leq \Phi\bigg(\frac{M}{\underline{\sigma}}\bigg) < 1, \tag{D.14}$$

for any $P \in \mathbf{P}_0$ and n. Now, by the definition of $\hat{\tau}_{n,1-\gamma_n}$, we note that

$$\begin{aligned} P(\hat{\tau}_{n,1-\gamma_n} \leq M) &\leq P\big(\hat{F}_n(M) \geq 1 - \gamma_n\big) \\ &= P\big(o_p(1) + F_{n,P}(M) \geq 1 - \gamma_n\big), \end{aligned} \tag{D.15}$$

uniformly in $P \in \mathbf{P}_0$, where the second equality follows by result (D.10) since $M \geq \zeta + 2$ by choice. Combining results (D.14) and (D.15), we therefore conclude that

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\hat{\tau}_{n,1-\gamma_n} \leq M) = 0, \tag{D.16}$$

whenever $\gamma_n \to 0$. Since M is arbitrary, $\hat{\tau}_{n,1-\gamma_n} \xrightarrow{p} \infty$ uniformly in $P \in \mathbf{P}_0$ and so the first claim of the proposition follows.

For the second claim, define $\psi : \mathbf{H} \to \ell^\infty(\mathbf{H}_1)$ by: for each $h \in \mathbf{H}$ and $t \in \mathbf{H}_1$,

$$\psi(h)(t) \equiv \langle t, h\rangle_{\mathbf{H}}. \tag{D.17}$$

By Corollary 6.55 (the Riesz representation theorem) and Lemma 6.10 in Aliprantis and Border (2006), $\sup_{t\in\mathbf{H}_1} |\psi(\bar{\mathbb{Z}}_{n,P})(t)| = \|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}}$. Clearly, ψ is linear and continuous. In turn, by Assumption 3.3(i), $\psi(\bar{\mathbb{Z}}_{n,P})$ is tight and centered Gaussian in $\ell^\infty(\mathbf{H}_1)$ by Lemma 2.2.2 in Bogachev (1998). By Example 1.5.10 in van der Vaart and Wellner (1996) and Proposition 2.1.12 in Giné and Nickl (2016), $\{\psi(\bar{\mathbb{Z}}_{n,P})(t) : t \in \mathbf{H}_1\}$ is separable as a process; it also has finite median by (D.8). By Proposition A.2.4 in van der Vaart and Wellner (1996) and (D.8), we have: for some absolute constant $C > 0$,

$$E\big[\|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}}^2\big] \leq C\big(E\big[\|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}}\big]\big)^2 \leq C\zeta^2. \tag{D.18}$$

By Proposition A.2.1 in van der Vaart and Wellner (1996) and result (D.18), we may thus conclude that, for all $x > 0$, all n and all $P \in \mathbf{P}_0$,

$$P\big(\|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}} > x\big) \leq 2\exp\left\{-\frac{x^2}{8E\big[\|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}}^2\big]}\right\} \leq 2\exp\left\{-\frac{x^2}{8C\zeta^2}\right\}. \tag{D.19}$$

By the definition of $\hat{\tau}_{n,1-\gamma_n}$ and the triangle inequality, we have

$$\begin{aligned}\gamma_n &< P\big(\|\hat{\mathbb{G}}_n\|_{\mathbf{H}} > \hat{\tau}_{n,1-\gamma_n} - \delta_n | \{X_i\}_{i=1}^n\big) \\ &\leq P\big(\|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}} > \hat{\tau}_{n,1-\gamma_n} - \delta_n - e_{n,P} | \{X_i\}_{i=1}^n\big),\end{aligned} \tag{D.20}$$

where $e_{n,P} \equiv \|\hat{\mathbb{G}}_n - \bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}} = o_p(c_n)$ uniformly in $P \in \mathbf{P}_0$ (by Assumption 3.2(ii)). By result (D.16) and $c_n = O(1)$ by Assumption 3.2(i), we note that

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}_0} P(\hat{\tau}_{n,1-\gamma_n} - \delta_n - e_{n,P} > 0) = 1. \tag{D.21}$$

Since $\bar{\mathbb{Z}}_{n,P}$ is independent of $\{X_i\}_{i=1}^n$, we may conclude from results (D.19), (D.20), and (D.21) that, with probability approaching 1 and uniformly in $P \in \mathbf{P}_0$,

$$\gamma_n \leq 2\exp\left\{-\frac{(\hat{\tau}_{n,1-\gamma_n} - \delta_n - e_{n,P})^2}{8C\zeta^2}\right\}. \tag{D.22}$$

Taking natural logarithms on both sides of (D.22) plus simple algebra yield

$$\frac{1}{8C\zeta^2}\left(\frac{\hat{\tau}_{n,1-\gamma_n}}{r_n c_n} - \frac{\delta_n}{r_n c_n} - \frac{e_{n,P}}{r_n c_n}\right)^2 \leq -\frac{\log\gamma_n}{r_n^2 c_n^2} + \frac{\log 2}{r_n^2 c_n^2}. \tag{D.23}$$

Suppose $(r_n c_n)^{-2}\log\gamma_n \to 0$. Then we must have $r_n c_n \to \infty$ since $\gamma_n \to 0$ (and so $\log\gamma_n \to -\infty$). Since also $\delta_n = o(c_n)$ and $e_{n,P} = o_p(c_n)$ uniformly in $P \in \mathbf{P}_0$, we obtain from (D.23) that $\hat{\tau}_{n,1-\gamma_n}/(r_n c_n) \xrightarrow{p} 0$ and hence $\kappa_n \equiv r_n c_n/\hat{\tau}_{n,1-\gamma_n} \xrightarrow{p} \infty$ uniformly in $P \in \mathbf{P}_0$. This completes the proof of the second claim of the proposition. *Q.E.D.*

LEMMA D.1: *Let Assumption* 3.1 *hold and* $\theta_0 \in \Lambda$. *Define* $\psi_a(h) \equiv \|h + a\theta_0 - \Pi_\Lambda(h + a\theta_0)\|_{\mathbf{H}}$ *for* $h \in \mathbf{H}$ *and* $a \geq 0$. *Then* $a \mapsto \psi_a(h)$ *is weakly decreasing on* $[0,\infty)$.

PROOF: The lemma immediately follows if we can show that

$$\psi_a(h) = \min_{|a'|\leq a}\big\|h + a'\theta_0 - \Pi_\Lambda\big(h + a'\theta_0\big)\big\|_{\mathbf{H}}. \tag{D.24}$$

Let $\Lambda_1^\circ \equiv \{h^* \in \mathbf{H} : \langle h^*, \lambda\rangle_{\mathbf{H}} \leq 0 \text{ for all } \lambda \in \Lambda, \|h^*\|_{\mathbf{H}} \leq 1\}$. By Assumption 3.1 and Deutsch (2012), pp. 125–127, we then have: for all $h \in \mathbf{H}$,

$$\begin{aligned}
&\min_{|a'|\leq a}\big\|h + a'\theta_0 - \Pi_\Lambda\big(h + a'\theta_0\big)\big\|_{\mathbf{H}} \\
&\quad = \min_{|a'|\leq a}\max_{h^*\in\Lambda_1^\circ}\big\langle h^*, h + a'\theta_0\big\rangle_{\mathbf{H}} \\
&\quad = \min_{|a'|\leq a}\max_{h^*\in\Lambda_1^\circ}\big\{\big\langle h^*, h\big\rangle_{\mathbf{H}} + a'\big\langle h^*, \theta_0\big\rangle_{\mathbf{H}}\big\}.
\end{aligned} \tag{D.25}$$

In turn, by Theorems 49.A and 49.B in Zeidler (1985), we obtain

$$\begin{aligned}
&\min_{|a'|\leq a}\max_{h^*\in\Lambda_1^\circ}\big\{\big\langle h^*, h\big\rangle_{\mathbf{H}} + a'\big\langle h^*, \theta_0\big\rangle_{\mathbf{H}}\big\} \\
&\quad = \max_{h^*\in\Lambda_1^\circ}\min_{|a'|\leq a}\big\{\big\langle h^*, h\big\rangle_{\mathbf{H}} + a'\big\langle h^*, \theta_0\big\rangle_{\mathbf{H}}\big\}.
\end{aligned} \tag{D.26}$$

Since $\langle h^*, \theta_0\rangle_{\mathbf{H}} \leq 0$ for all $h^* \in \Lambda_1^\circ$, it follows from result (D.25) that

$$\begin{aligned}
&\max_{h^*\in\Lambda_1^\circ}\min_{|a'|\leq a}\big\{\big\langle h^*, h\big\rangle_{\mathbf{H}} + a'\big\langle h^*, \theta_0\big\rangle_{\mathbf{H}}\big\} \\
&\quad = \max_{h^*\in\Lambda_1^\circ}\big\{\big\langle h^*, h\big\rangle_{\mathbf{H}} + a\big\langle h^*, \theta_0\big\rangle_{\mathbf{H}}\big\} \\
&\quad = \max_{h^*\in\Lambda_1^\circ}\big\{\big\langle h^*, h + a\theta_0\big\rangle_{\mathbf{H}}\big\} = \big\|(h + a\theta_0) - \Pi_\Lambda(h + a\theta_0)\big\|_{\mathbf{H}},
\end{aligned} \tag{D.27}$$

where the last step is by Deutsch (2012), pp. 125–127. The equality in (D.24) then follows from combining (D.25), (D.26), and (D.27). *Q.E.D.*

LEMMA D.2: *Let Assumption* 3.1 *hold and* $\bar{\mathbf{P}}_0$ *be as in Theorem* 3.1. *Then it follows that, for any* $h \in \mathbf{H}$, $a \in \mathbf{R}_+$ *and* $P \in \bar{\mathbf{P}}_0$,

$$\Pi_\Lambda(h + a\theta_P) = \Pi_\Lambda(h) + a\theta_P. \tag{D.28}$$

PROOF: Let $\Lambda^\circ \equiv \{\vartheta \in \mathbf{H} : \sup_{\lambda\in\Lambda}\langle\vartheta, \lambda\rangle_{\mathbf{H}} \leq 0\}$. Fix any $h \in \mathbf{H}$, $a \in \mathbf{R}_+$, and $P \in \bar{\mathbf{P}}_0$. By Assumption 3.1, $\Pi_\Lambda(h) + a\theta_P \in \Lambda$. First, note that, for any $\lambda \in \Lambda$,

$$\big\langle h + a\theta_P - \big\{\Pi_\Lambda(h) + a\theta_P\big\}, \lambda\big\rangle_{\mathbf{H}} = \big\langle h - \Pi_\Lambda(h), \lambda\big\rangle_{\mathbf{H}} \leq 0, \tag{D.29}$$

where the inequality follows by Assumption 3.1 and Theorem 4.7 in Deutsch (2012). Next, for $\lambda_0 \equiv \Pi_\Lambda(h) + a\theta_P \in \Lambda$, we have

$$\begin{aligned}&\big\langle h + a\theta_P - \big\{\Pi_\Lambda(h) + a\theta_P\big\}, \lambda_0\big\rangle_{\mathbf{H}} \\ &\quad = \big\langle h - \Pi_\Lambda(h), \Pi_\Lambda(h)\big\rangle_{\mathbf{H}} + a\big\langle h - \Pi_\Lambda(h), \theta_P\big\rangle_{\mathbf{H}} = 0,\end{aligned} \tag{D.30}$$

where the second equality is due to $\langle h - \Pi_\Lambda(h), \Pi_\Lambda(h)\rangle_{\mathbf{H}} = 0$ by Assumption 3.1 and Theorem 4.7 in Deutsch (2012), $h - \Pi_\Lambda(h) \in \Lambda^\circ$ by Assumption 3.1 and Theorem 5.6 in Deutsch (2012), and the definition of $\bar{\mathbf{P}}_0$. The conclusion of the lemma then follows from applying Theorem 4.7 in Deutsch (2012) to (D.29) and (D.30). *Q.E.D.*

PROPOSITION D.1: *Let Assumptions* 3.1 *and* 3.3 *hold, and* $\psi_{a,P}$ *be defined as in* (14). *Then for any sequence* $\{\epsilon_n\}$ *of positive scalars satisfying* $\epsilon_n = o(c_n)$,

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} \sup_{x\in[c_{n,P}(0.5)+\varsigma_n,\infty)} P\big(\big|\psi_{\kappa_n,P}(\mathbb{Z}_{n,P}) - x\big| \le \epsilon_n\big) = 0. \tag{D.31}$$

PROOF: Let $\{\epsilon_n\}$ be an arbitrary sequence of positive scalars satisfying $\epsilon_n = o(c_n)$ as $n \to \infty$. Fix $n \in \mathbf{N}$ and $P \in \mathbf{P}_0$ for the moment. Let $\Lambda_1^\circ \equiv \{t \in \mathbf{H} : \langle t, \lambda\rangle_{\mathbf{H}} \le 0 \text{ for all } \lambda \in \Lambda, \|t\|_{\mathbf{H}} \le 1\}$. By Assumption 3.1 and Deutsch (2012), pp. 125–127, we may then write: for $e_t(n, P) \equiv \kappa_n\langle t, \theta_P\rangle_{\mathbf{H}}$,

$$\psi_{\kappa_n,P}(\mathbb{Z}_{n,P}) = \max_{t\in\Lambda_1^\circ}\big\{\langle t, \mathbb{Z}_{n,P}\rangle_{\mathbf{H}} + e_t(n, P)\big\}. \tag{D.32}$$

Since $0 \in \Lambda_1^\circ$ and $\langle t, \mathbb{Z}_{n,P}\rangle_{\mathbf{H}} + e_t(n, P) = 0$ at $t = 0$, the maximum in (D.32) must be attained at $t \in \Lambda_1^\circ$ such that $\langle t, \mathbb{Z}_{n,P}\rangle_{\mathbf{H}} + e_t(n, P) \ge 0$. Moreover, $\langle t, \mathbb{Z}_{n,P}\rangle_{\mathbf{H}} \le \|\mathbb{Z}_{n,P}\|_{\mathbf{H}}$ for all $t \in \Lambda_1^\circ$ by the Cauchy–Schwarz inequality. Therefore, whenever $\|\mathbb{Z}_{n,P}\|_{\mathbf{H}} \le M$ with $M > 0$, the maximum in (D.32) must be attained at some $t \in \Lambda_1^\circ$ with $e_t(n, P) \ge -M$. It follows that, whenever $\|\mathbb{Z}_{n,P}\|_{\mathbf{H}} \le M$,

$$\psi_{\kappa_n,P}(\mathbb{Z}_{n,P}) = \max_{t\in\Lambda_{1,M}^\circ(n,P)}\big\{\langle t, \mathbb{Z}_{n,P}\rangle_{\mathbf{H}} + e_t(n, P)\big\}, \tag{D.33}$$

where $\Lambda_{1,M}^\circ(n, P) \equiv \{t \in \Lambda_1^\circ : e_t(n, P) \ge -M\}$. Hence, for any $x \in \mathbf{R}$,

$$\begin{aligned}&P\big(\big|\psi_{\kappa_n,P}(\mathbb{Z}_{n,P}) - x\big| \le \epsilon_n\big) \\ &\quad \le P\Big(\Big|\max_{t\in\Lambda_{1,M}^\circ(n,P)}\big\{\langle t, \mathbb{Z}_{n,P}\rangle_{\mathbf{H}} + e_t(n, P)\big\} - x\Big| \le \epsilon_n\Big) + P\big(\|\mathbb{Z}_{n,P}\|_{\mathbf{H}} > M\big) \\ &\quad \le P\Big(\Big|\max_{t\in\Lambda_{1,M}^\circ(n,P)}\big\{\langle t, \mathbb{Z}_{n,P}\rangle_{\mathbf{H}} + e_t(n, P)\big\} - x\Big| \le \epsilon_n\Big) + \frac{\zeta}{M},\end{aligned} \tag{D.34}$$

for some constant $\zeta > 0$ satisfying $\sup_{P\in\mathbf{P}} E[\|\mathbb{Z}_{n,P}\|_{\mathbf{H}}] < \zeta$, where the existence of ζ is guaranteed by Markov's inequality and Assumption 3.3(ii).

We next aim to control the first term on the right-hand side of (D.34) by bounding the density of $\max_{t\in\Lambda_{1,M}^\circ(n,P)}\{\langle t, \mathbb{Z}_{n,P}\rangle_{\mathbf{H}} + e_t(n, P)\}$. To this end, let $F_{n,P,M}$ be the cdf of $\max_{t\in\Lambda_{1,M}^\circ(n,P)}\{\langle t, \mathbb{Z}_{n,P}\rangle_{\mathbf{H}} + e_t(n, P)\}$. We proceed with some useful facts. First, by Assumption 3.3(i), Lemma 1.3.2 in van der Vaart and Wellner (1996), and the corollary to Theorem I.3.1 in Vakhania, Tarieladze, and Chobanyan (1987), $\mathbb{Z}_{n,P}$ is a centered Radon

Gaussian variable in $\mathbf{H}$. Second, for $\underline{r}_M(n,P) \equiv \inf\{r \in \mathbf{R} : F_{n,P,M}(r) > 0\}$, Theorem 11.1 in Davydov, Lifshits, and Smorodina (1998) in turn implies that $F_{n,P,M}$ is absolutely continuous on $(\underline{r}_M(n,P), \infty)$ so that it admits a density on $(\underline{r}_M(n,P), \infty)$ which we denote by $f_{n,P,M}$. Third, by Proposition 11.2 in Davydov, Lifshits, and Smorodina (1998), we may assume without loss of generality that $\Lambda^\circ_{1,M}(n,P)$ is countable. Fourth, since $e_t(n,P) \le 0$ for any $t \in \Lambda^\circ_1$ and $P \in \mathbf{P}_0$, we have $e_{t,M}(n,P) \equiv e_t(n,P) + M \le M$, which, together with the Cauchy–Schwarz inequality and $\mathbb{Z}_{n,P} \in \mathbf{H}$ (by Assumption 3.2(i)), implies

$$\max_{t\in\Lambda^\circ_{1,M}(n,P)} \big\{\langle t, \mathbb{Z}_{n,P}\rangle_{\mathbf{H}} + e_{t,M}(n,P)\big\} \le \|\mathbb{Z}_{n,P}\|_{\mathbf{H}} + M < \infty, \tag{D.35}$$

almost surely. Fifth, for $\bar{\sigma}^2_{n,P,M} \equiv \sup_{t\in\Lambda^\circ_{1,M}(n,P)} E[\langle t, \mathbb{Z}_{n,P}\rangle^2_{\mathbf{H}}]$, we shall show towards the end of the proof that, for all large $M > 0$,

$$\bar{\sigma}^2_{n,P,M} > 0. \tag{D.36}$$

In what follows, we fix any such large M. Sixth, for any $r > \underline{r}_M(n,P)$, we note

$$\begin{aligned} F_{n,P,M}(r) &\equiv P\Big(\max_{t\in\Lambda^\circ_{1,M}(n,P)} \{\langle t, \mathbb{Z}_{n,P}\rangle_{\mathbf{H}} + e_t(n,P) \le r\Big) \\ &= P\Big(\max_{t\in\Lambda^\circ_{1,M}(n,P)} \big\{\langle t, \mathbb{Z}_{n,P}\rangle_{\mathbf{H}} + e_{t,M}(n,P)\big\} \le r + M\Big). \end{aligned} \tag{D.37}$$

Seventh, for $m_{n,P,M}$ the median of $F_{n,P,M}$, we have by the quantile equivariance that the median of $\max_{t\in\Lambda^\circ_{1,M}(n,P)}\{\langle t, \mathbb{Z}_{n,P}\rangle_{\mathbf{H}} + e_{t,M}(n,P)\}$ is $m_{n,P,M} + M$. Note that $m_{n,P,M} \ge \underline{r}_M(n,P) \ge 0$ because $\max_{t\in\Lambda^\circ_{1,M}(n,P)}\{\langle t, \mathbb{Z}_{n,P}\rangle_{\mathbf{H}} + e_t(n,P)\} \ge 0$.

With the above preparations, we may apply Theorem 2.2.2 in Yurinsky (1995) with $b = m_{n,P,M} + M$ and $u = r + M$ to conclude:

$$\begin{aligned} f_{n,P,M}(r) = F'_{n,P,M}(r) &\le \frac{2(r+M) - (m_{n,P,M} + M)}{\big[(r+M) - (m_{n,P,M} + M)\big]^2} \\ &= \frac{2r - m_{n,P,M} + M}{(r - m_{n,P,M})^2} \end{aligned} \tag{D.38}$$

for all $r > m_{n,P,M}$. By the choice of ϵ_n and $c_n = O(1)$, we note that

$$\epsilon_n = o(c_n) = o\Big(\sqrt{c_n/\varsigma_n^2}\sqrt{c_n}\varsigma_n\Big) = o(\varsigma_n), \tag{D.39}$$

as $n \to \infty$. Therefore, we have $\epsilon_n \le \varsigma_n/2$ for all n sufficiently large, so that

$$x - \epsilon_n - m_{n,P,M} \ge m_{n,P,M} + \varsigma_n - \epsilon_n - m_{n,P,M} \ge \frac{\varsigma_n}{2} \tag{D.40}$$

whenever $x \ge m_{n,P,M} + \varsigma_n$. Since $r \mapsto (2r - m_{n,P,M} + M)/(r - m_{n,P,M})^2$ is decreasing on $(m_{n,P,M}, \infty)$, we may thus conclude by the fundamental theorem of calculus and results (D.37) and (D.38) that, for all $x \ge m_{n,P,M} + \varsigma_n$ and n large,

$$P\Big(\Big|\max_{t\in\Lambda^\circ_{1,M}(n,P)} \big\{\langle t, \mathbb{Z}_{n,P}\rangle_{\mathbf{H}} + e_t(n,P)\big\} - x\Big| \le \epsilon_n\Big)$$

$$= \int_{x-\epsilon_n}^{x+\epsilon_n} f_{n,P,M}(r)\,\mathrm{d}r \le 2\epsilon_n \frac{2(m_{n,P,M} + \varsigma_n/2) - m_{n,P,M} + M}{(\varsigma_n/2)^2}. \tag{D.41}$$

Since $\Lambda^\circ_{1,M}(n, P) \subset \Lambda^\circ_1$, we obtain in view of (D.32) and Lemma D.1 that

$$\max_{t\in\Lambda^\circ_{1,M}(n,P)} \big\{\langle t, \mathbb{Z}_{n,P}\rangle_{\mathbf{H}} + e_t(n, P)\big\} \le \psi_{\kappa_n,P}(\mathbb{Z}_{n,P}) \le \psi_{0,P}(\mathbb{Z}_{n,P})$$
$$= \|\mathbb{Z}_{n,P}\|_{\mathbf{H}}. \tag{D.42}$$

By result (D.42), Kwapień (1994), and Assumption 3.3(ii), we note

$$m_{n,P,M} \le m_{n,P} \equiv c_{n,P}(0.5) \le E\big[\|\mathbb{Z}_{n,P}\|_{\mathbf{H}}\big] \le \zeta, \tag{D.43}$$

where we remind the reader our choice of ζ from (D.34). Combining results (D.34), (D.41), and (D.42), we thus obtain that

$$\sup_{P\in\mathbf{P}_0}\ \sup_{x\in[c_{n,P}(0.5)+\varsigma_n)} P\big(\big|\psi_{\kappa_n,P}(\mathbb{Z}_{n,P}) - x\big| \le \epsilon_n\big) \lesssim \epsilon_n \frac{\zeta + \varsigma_n + M}{\varsigma_n^2} + \frac{\zeta}{M}. \tag{D.44}$$

Since $\epsilon_n = o(c_n)$, we may select a sequence $a_n \downarrow 0$ (sufficiently slow) such that $\epsilon_n = o(a_n c_n)$. In turn, by setting $M \equiv M_n = a_n^{-1}$ which diverges to infinity, we may then conclude by Assumption 3.3(iv) and results (D.39) and (D.44) that

$$\sup_{P\in\mathbf{P}_0}\ \sup_{x\in[c_{n,P}(0.5)+\varsigma_n)} P\big(\big|\psi_{\kappa_n,P}(\mathbb{Z}_{n,P}) - x\big| \le \epsilon_n\big) \to 0. \tag{D.45}$$

It remains to prove (D.36). For this, we fix n and $P \in \mathbf{P}_0$ in what follows. Let $\bar{\sigma}^2_{n,P} \equiv \sup_{t\in\Lambda^\circ_1} E[\langle t, \mathbb{Z}_{n,P}\rangle^2_{\mathbf{H}}]$. Then we must have $\bar{\sigma}^2_{n,P} > 0$. Indeed, suppose by way of contradiction that $\bar{\sigma}^2_{n,P} = 0$. This implies $\langle t, \mathbb{Z}_{n,P}\rangle_{\mathbf{H}} = 0$ almost surely for all $t \in \Lambda^\circ_1$. By result (D.32) and Proposition 11.2 in Davydov, Lifshits, and Smorodina (1998), we have $\psi_{\kappa_n}(\mathbb{Z}_{n,P}) = 0$ almost surely. Then all quantiles of $\psi_{\kappa_n}(\mathbb{Z}_{n,P})$ are equal to zero, contradicting Assumption 3.3(iii). Next, fix $\eta > 0$. Then we may select some $t_{n,P} \in \Lambda^\circ_1$ such that

$$\bar{\sigma}^2_{n,P} \le E\big[\langle t_{n,P}, \mathbb{Z}_{n,P}\rangle^2_{\mathbf{H}}\big] + \eta. \tag{D.46}$$

Moreover, by choosing $M \ge \kappa_n\|\theta_P\|_{\mathbf{H}}$, we may employ the Cauchy–Schwarz inequality and $\|t_{n,P}\|_{\mathbf{H}} \le 1$ (due to $t_{n,P} \in \Lambda^\circ_1$) to obtain that

$$\big|e_{t_{n,P}}(n, P)\big| \equiv \big|\kappa_n\langle t_{n,P}, \theta_P\rangle_{\mathbf{H}}\big| \le \kappa_n\|\theta_P\|_{\mathbf{H}} \le M. \tag{D.47}$$

In turn, it follows from result (D.47) that $t_{n,P} \in \Lambda^\circ_{1,M}(n, P)$ so that

$$E\big[\langle t_{n,P}, \mathbb{Z}_{n,P}\rangle^2_{\mathbf{H}}\big] \le \sup_{t\in\Lambda^\circ_{1,M}(n,P)} E\big[\langle t, \mathbb{Z}_{n,P}\rangle^2_{\mathbf{H}}\big] = \bar{\sigma}^2_{n,P,M}. \tag{D.48}$$

Combining results (D.46) and (D.48), we may then conclude that

$$\bar{\sigma}^2_{n,P} \le \bar{\sigma}^2_{n,P,M} + \eta \le \bar{\sigma}^2_{n,P} + \eta \tag{D.49}$$

whenever $M \ge \kappa_n\|\theta_P\|_{\mathbf{H}}$. Since η is arbitrary, result (D.49) implies that $\bar{\sigma}^2_{n,P,M} \to \bar{\sigma}^2_{n,P}$ as $M \to \infty$. This, together with $\bar{\sigma}^2_{n,P} > 0$, implies (D.36). *Q.E.D.*

LEMMA D.3: *Let* $\mathbf{D}$ *be a Banach space with norm* $\|\cdot\|_{\mathbf{D}}$ *and* $\mathbf{D}_1^* \equiv \{x^* \in \mathbf{D}^* : \sup_{\|x\|_{\mathbf{D}}\le 1} |x^*(x)| \le 1\}$, *the unit ball in the topological dual* $\mathbf{D}^*$ *of* $\mathbf{D}$. *If* $\mathbb{G} \in \mathbf{D}$ *is a tight centered Gaussian variable such that* $\sup_{x^*\in\mathbf{D}_1^*} E[x^*(\mathbb{G})^2] > 0$, *then the cdf* F *of* $\|\mathbb{G}\|_{\mathbf{D}}$ *is absolutely continuous on* $(0,\infty)$, *and, for any* $r > m_F$ *with* m_F *the median of* F,

$$F'(r) \le \frac{2r - m_F}{(r-m_F)^2}. \tag{D.50}$$

PROOF: Since $\mathbb{G}$ is tight and $\mathbf{D}$ is Banach, Lemma 1.3.2 in van der Vaart and Wellner (1996) and the corollary to Theorem I.3.1 in Vakhania, Tarieladze, and Chobanyan (1987) imply that $\mathbb{G}$ is Radon. Hence, since $\mathbb{G}$ is centered Gaussian, we know by the remark following Proposition 7.4 in Davydov, Lifshits, and Smorodina (1998) that the support $\mathbf{D}_0$ of $\mathbb{G}$ is a closed separable subspace of $\mathbf{D}$ and hence a separable Banach space under $\|\cdot\|_{\mathbf{D}}$. Therefore, by Proposition 1.12.17 in Bogachev and Smolyanov (2017), it follows that, for all $x \in \mathbf{D}_0$,

$$\|x\|_{\mathbf{D}} = \sup_{n=1}^{\infty} x_n^*(x), \tag{D.51}$$

where $\{x_n^*\}_{n=1}^\infty$ live in $\mathbf{D}_{0,1}^*$, the unit ball of the topological dual space $\mathbf{D}_0^*$ of $\mathbf{D}_0$. By the Hahn–Banach extension theorem (see, e.g., Theorem 5.53 in Aliprantis and Border (2006)), each x_n^* admits an extension that belongs to $\mathbf{D}_1^*$, which we continue to denote by x_n^* with some abuse of notation. In other words, (D.52) holds with $\{x_n^*\}_{n=1}^\infty$ living in $\mathbf{D}_1^*$. Since $P(\mathbb{G} \in \mathbf{D}_0) = 1$, we then obtain that, almost surely,

$$\|\mathbb{G}\|_{\mathbf{D}} = \sup_{n=1}^{\infty} x_n^*(\mathbb{G}). \tag{D.52}$$

For each n, we have $E[x_n^*(\mathbb{G})] = 0$ due to $\mathbb{G}$ being centered. Moreover, the supremum in (D.52) is finite almost surely. Since $\sup_{x^*\in\mathbf{D}_1^*} E[x^*(\mathbb{G})^2] > 0$ by assumption, Theorem 2.2.1 in Yurinsky (1995) implies that F is absolutely continuous on (r_0,∞) with $r_0 \equiv \inf\{r \in \mathbf{R} : F(r) > 0\}$. Since the support $\mathbf{D}_0$ of $\mathbb{G}$ as a subspace includes 0 (in $\mathbf{D}$), we have by Problem 11.3 in Davydov, Lifshits, and Smorodina (1998) that $r_0 = 0$. This proves the first claim. The second claim follows immediately by applying Theorem 2.2.2-(a) in Yurinsky (1995) with $b = m_F$ and noting that $t \equiv \Phi^{-1}(F(m_F)) \ge \Phi^{-1}(0.5) = 0$. *Q.E.D.*

APPENDIX E: MORE SIMULATION STUDIES AND EMPIRICAL APPLICATION

E.1. *More Simulation Studies*

This section conducts more simulation studies for three restrictions: concavity/convexity, monotonicity jointly with convexity, and Slutsky restriction. For the first two, we shall compare to the test by Lee, Song, and Whang (2017) which is asymptotically nonconservative and meanwhile computationally manageable—see the discussions of other existing tests in Example B.2. For the Slutsky restriction, one may also adopt the nonconservative test by Chernozhukov, Newey, and Santos (2015). However, its implementation requires nonlinearly constrained optimization (in addition to optimization over the estimated set of minimizers) in each bootstrap repetition, and the computation cost grows quickly with the relevant dimension (Zhu (2020), p. 617). By restricting to linear (in g_0 in the context of Example 2.4) constraints, Zhu (2020) developed a computationally simpler inferential framework, which unfortunately excludes the Slutsky restriction. For these reasons, we shall only implement our test for the Slutsky restriction. We stress, however, that Chernozhukov, Newey, and Santos (2015) accommodated partial identification while we cannot.

The first set of simulations makes use of exactly the same univariate design (26) in Section 4, and we aim to test whether θ_0 is convex, and whether θ_0 is nondecreasing *and* convex. The implementation of our tests remains unchanged other than adjusting linear constraints in quadratic programs accordingly. Following Fan and Gijbels (1996), p. 59, the LSW tests are implemented similarly as before but now based on local polynomial regression of order $q = 3$ for both restrictions (so that the bandwidths are evaluated at $q = 3$). Note in particular that, for the joint test of monotonicity and convexity, we estimate the first and second derivatives of θ_0 in a single local polynomial regression of order 3, instead of two separate regressions, for ease of computation. Thus, in assessing that "additional restrictions help improve power," one should compare the resulting power curves to those for convexity, rather than those for monotonicity in Section 4 which are associated with a different convergence rate r_n (through its dependence on q).

The second set of simulations are based on the same design for (27) except

$$\theta_0(z_1, z_2) = \mathsf{a}\left(\frac{1}{2}z_1^{\mathsf{b}} + \frac{1}{2}z_2^{\mathsf{b}}\right)^{1/\mathsf{b}} + \mathsf{c}\log\bigl(1 + 5(z_1 + z_2)\bigr), \tag{E.1}$$

where we adopt the same set of choices for $(\mathsf{a}, \mathsf{b}, \mathsf{c})$ but with $\Delta = 0.05$ replaced by $\Delta = 0.2$, so that the power of the implemented tests is close to 1 as δ increases from 1 to 10. We then aim to test concavity of θ_0. To ease computation, the L^2-integrals for our test are evaluated over $[0.1, 0.9]^2$ but now with marginal step size 0.1. The LSW tests are based on the Hessian matrix $z \mapsto \Theta_0(z)$ of θ_0 so that, in the notation of LSW, $J = 1$ and $v_{\tau,1}(z) = a_\tau^\intercal \Theta_0(z) a_\tau$ with $a_\tau \equiv [\cos(\tau), \sin(\tau)]^\intercal$. To reduce computation cost, we approximate the resulting triple integrals over $z \in [0.1, 0.9]^2$ with marginal step size 0.1 and over τ based on 500 draws from the uniform distribution on $[0, 2\pi]$. As with the LSW tests for (27), the number of Monte Carlo simulation replications for the LSW tests in the bivariate design (E.1) is decreased to be 1000.

Tables E.I–E.II summarize the empirical sizes with $\gamma_n \in \{1/n, 0.01/\log n, 0.01\}$—see also Tables H.II–H.III in Appendix H. Once again, our tests are insensitive to the choice of γ_n. In the univariate case, our tests control sizes well across shapes, sample sizes, and the number of knots, while LSW's tests for monotonicity jointly with convexity are slightly over-sized. In the bivariate case, our tests, especially FS-C1 (in which case the sieve dimension is 25), tend to over-reject, though to an overall lesser extent as n increases. The size distortions in small samples may be explained by the fact that the Gaussian approximation is inaccurate due to a "large" number of regressors being used in the sieve estimation. On the other hand, LSW-L and in particular LSW-S exhibit overall less size distortions compared to our tests except FS-Q0.

In turn, Figures E.1–E.2 depict the power curves, where we only show our tests with $\gamma_n = 0.01/\log n$ due to space limitation and the fact that other choices of γ_n enjoy very similar curves—see also Figures H.2–H.3 and H.5 in Appendix H. Overall, our tests appear to be significantly more powerful than the LSW tests across shapes, sample sizes, and the number of interior knots, in both univariate and bivariate designs. The power of the LSW tests in the bivariate case is less than 25% across sample sizes. The substantial power gaps are in line with the fact that the LSW tests entail estimation of the second derivatives of θ_0, which admit slower rates of convergence. We note, however, that our

TABLE E.I

EMPIRICAL SIZE OF SHAPE TESTS FOR θ_0 IN (26) AT $\alpha = 5\%$[a]

Shape	n	γ_n	FS-C3: $k_n = 7$ D1	FS-C3: $k_n = 7$ D2	FS-C3: $k_n = 7$ D3	FS-C5: $k_n = 9$ D1	FS-C5: $k_n = 9$ D2	FS-C5: $k_n = 9$ D3	FS-C7: $k_n = 11$ D1	FS-C7: $k_n = 11$ D2	FS-C7: $k_n = 11$ D3
Con	500	$1/n$	0.048	0.042	0.009	0.056	0.047	0.016	0.053	0.045	0.017
		$0.01/\log n$	0.048	0.042	0.009	0.056	0.047	0.016	0.053	0.045	0.017
		0.01	0.049	0.042	0.009	0.056	0.048	0.016	0.053	0.045	0.017
	750	$1/n$	0.058	0.046	0.007	0.062	0.055	0.011	0.059	0.055	0.019
		$0.01/\log n$	0.058	0.046	0.007	0.062	0.055	0.011	0.059	0.055	0.019
		0.01	0.058	0.046	0.008	0.062	0.055	0.011	0.059	0.056	0.020
	1000	$1/n$	0.052	0.044	0.005	0.055	0.047	0.010	0.054	0.044	0.013
		$0.01/\log n$	0.052	0.044	0.005	0.055	0.047	0.010	0.054	0.044	0.013
		0.01	0.052	0.044	0.005	0.055	0.047	0.010	0.054	0.045	0.013
Mon-Con	500	$1/n$	0.050	0.026	0.007	0.054	0.032	0.011	0.054	0.032	0.013
		$0.01/\log n$	0.050	0.026	0.007	0.054	0.032	0.011	0.054	0.032	0.013
		0.01	0.050	0.026	0.007	0.054	0.033	0.011	0.054	0.032	0.014
	750	$1/n$	0.056	0.026	0.005	0.059	0.034	0.008	0.057	0.034	0.017
		$0.01/\log n$	0.056	0.026	0.005	0.059	0.034	0.008	0.057	0.034	0.017
		0.01	0.056	0.026	0.005	0.059	0.035	0.008	0.057	0.034	0.018
	1000	$1/n$	0.055	0.022	0.004	0.055	0.029	0.006	0.053	0.030	0.010
		$0.01/\log n$	0.055	0.022	0.004	0.055	0.029	0.006	0.053	0.030	0.010
		0.01	0.055	0.023	0.004	0.056	0.029	0.006	0.053	0.030	0.010

Shape	Tests	$n = 500$ D1	$n = 500$ D2	$n = 500$ D3	$n = 750$ D1	$n = 750$ D2	$n = 750$ D3	$n = 1000$ D1	$n = 1000$ D2	$n = 1000$ D3
Con	LSW-S	0.059	0.058	0.048	0.063	0.058	0.049	0.057	0.055	0.046
	LSW-L	0.063	0.066	0.050	0.064	0.064	0.047	0.058	0.058	0.046
Mon-Con	LSW-S	0.065	0.057	0.030	0.065	0.052	0.032	0.060	0.048	0.026
	LSW-L	0.068	0.057	0.030	0.069	0.053	0.031	0.065	0.054	0.026

[a]*Note*: "Con" refers to "Convexity," and "Mon-Con" refers to "Monotonicity and Convexity." The parameter γ_n determines $\hat{\kappa}_n$ proposed in Section 3.2 with $c_n = 1/\log n$ and $r_n = (n/k_n)^{1/2}$.

test of convexity in the design (26) has power slightly below 5% when $\delta = 1$. This is a setting where θ_0 is visually close to being convex. By further imposing monotonicity, the power discrepancies at $\delta = 1$ then vanish—see the second row in Figure E.1.

Our final set of Monte Carlo simulations concerns Slutsky restriction based on Example 2.4 with $d_q = 2$. Concretely, we draw i.i.d. samples $\{P_{1i}^*, P_{2i}^*, Y_i^*, Z_i^*, U_{1i}, U_{2i}\}_{i=1}^n$ from the standard normal distribution in $\mathbf{R}^6$ and set $P_i = [P_{1i}, P_{2i}]^\intercal$ with $P_{ji} = 1 + \Phi(P_{ji}^*)$, $Y_i = \Phi(Y_i^*)$, $Z_i = \Phi(Z_i^*)$, and $U_i = [U_{1i}, U_{2i}]^\intercal$ for all i and $j = 1, 2$. In turn, we let $\Gamma_0 = [1, 1]^\intercal$ and consider three specifications for $g_0 \equiv [g_{10}, g_{20}]^\intercal$ under the null:

$$g_{j0}(p_1, p_2, y) = \mathsf{a} p_j^{\frac{1}{\mathsf{b}-1}} \frac{y}{p_1^{\mathsf{b}/(\mathsf{b}-1)} + p_2^{\mathsf{b}/(\mathsf{b}-1)}} + \mathsf{c}, \quad j = 1, 2, \tag{E.2}$$

with $(\mathsf{a}, \mathsf{b}, \mathsf{c}) = (0, 0.5, 0.5), (0.5, 0, 0)$, and $(1, 0.5, 0)$, labeled D1, D2, and D3, respectively. Note that D1 is a least favorable case, while D2 and D3 may be respectively rationalized by a Cobb–Douglas and a CES (constant elasticity of substitution) utility function.

TABLE E.II

EMPIRICAL SIZE OF CONCAVITY TESTS FOR θ_0 IN (E.1) AT $\alpha = 5\%$[a]

		FS-Q0: $k_n = 9$			FS-Q1: $k_n = 16$			FS-C0: $k_n = 16$			FS-C1: $k_n = 25$		
n	γ_n	D1	D2	D3	D1	D2	D3	D1	D2	D3	D1	D2	D3
500	$1/n$	0.062	0.061	0.015	0.069	0.067	0.029	0.070	0.067	0.028	0.083	0.081	0.044
	$0.01/\log n$	0.062	0.061	0.015	0.069	0.067	0.029	0.070	0.067	0.028	0.083	0.081	0.044
	0.01	0.063	0.061	0.015	0.069	0.067	0.029	0.071	0.068	0.029	0.084	0.082	0.045
750	$1/n$	0.064	0.063	0.011	0.073	0.073	0.027	0.072	0.073	0.024	0.069	0.071	0.035
	$0.01/\log n$	0.064	0.063	0.011	0.073	0.073	0.027	0.072	0.073	0.024	0.069	0.071	0.035
	0.01	0.065	0.063	0.011	0.074	0.074	0.028	0.074	0.074	0.024	0.069	0.071	0.036
1000	$1/n$	0.057	0.059	0.004	0.067	0.066	0.018	0.069	0.067	0.014	0.066	0.065	0.027
	$0.01/\log n$	0.057	0.059	0.004	0.067	0.066	0.018	0.069	0.067	0.014	0.066	0.065	0.027
	0.01	0.057	0.059	0.004	0.067	0.067	0.018	0.070	0.068	0.014	0.067	0.065	0.027

	LSW-S			LSW-L		
n	D1	D2	D3	D1	D2	D3
500	0.046	0.049	0.043	0.059	0.055	0.049
750	0.068	0.056	0.049	0.071	0.074	0.048
1000	0.053	0.053	0.043	0.062	0.051	0.037

[a]*Note*: The parameter γ_n determines $\hat{\kappa}_n$ proposed in Section 3.2 with $c_n = 1/\log n$ and $r_n = (n/k_n)^{1/2}$.

For specifications under the alternative, we choose

$$\begin{aligned}&\big[g_{10}(p_1, p_2, y), g_{20}(p_1, p_2, y)\big]\\&\quad = \big[\exp\big\{(p_1 - 1.5)0.1\delta\big\}, \exp\big\{-(p_2 - 1.5)0.1\delta\big\}\big],\end{aligned} \tag{E.3}$$

where $\delta = 1, \dots, 10$. The resulting Slutsky matrix $\theta_0(p_1, p_2, y)$ at each (p_1, p_2, y) (as defined in (8)) has one of its eigenvalues positive and the other negative.

To implement our test, we construct a vector h^{k_n} of series functions via tensor product of univariate B-splines, obtain $\hat{g}_n$ by regressing $\{Q_i\}_{i=1}^n$ on $\{h^{k_n}(P_i, Y_i), Z_i\}_{i=1}^n$, and then derive $\hat{\theta}_n$ by differentiating $\hat{g}_n$. The whole procedure can be streamlined by the commands `spmak`, `fnval`, and `fnder` provided by the Curve Fitting Toolbox in MATLAB. A practical issue of grave concern is, however, that estimation of θ_0 now involves trivariate nonparametric functions, resulting in potentially too large a sieve dimension k_n (e.g., $k_n = 125$ for FS-C1). For this reason, we employ the same set of B-splines as in the bivariate design, but experiment with $n \in \{1000, 3000, 5000\}$. In turn, we evaluate the integrals (see (9)) over $[1.1, 1.9]^2 \times [0.1, 0.9]$ with marginal step size 0.05. Finally, we construct the critical values based on the sieve score bootstrap with i.i.d. standard normals as weights—see Appendix F.3 (note that our designs are configured without endogeneity for simplicity).

Table E.III and Figure E.3 report partial results of our simulations—see Table H.VI and Figure H.6 in Appendix H for the full set of results. Not surprisingly, our tests exhibit marked size distortions when the sieve dimension is "large" relative to the sample size, but otherwise control size reasonably well. As emphasized previously, Gaussian approximation may be inaccurate if k_n is "too large." On the other hand, the power performance is influenced by k_n through two channels: accuracy of the Gaussian approximation and the rate $r_n = \sqrt{n/k_n}$. This may explain the relative low power of our tests when $n = 1000$, though all power curves improve as n increases.

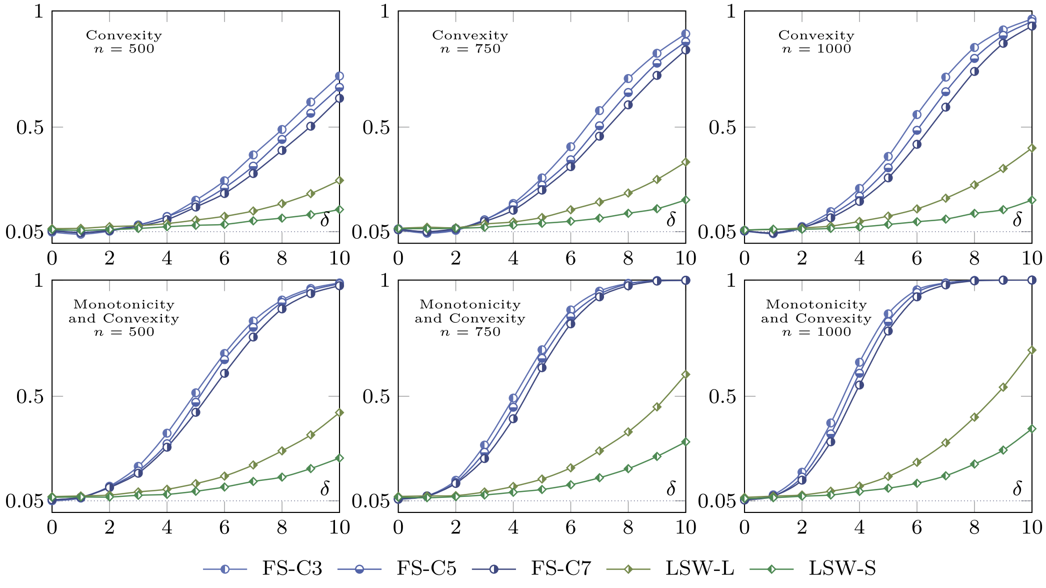

FIGURE E.1.—Empirical power of shape tests for (26) where our tests are implemented with $\gamma_n = 0.01/\log n$ and corresponding to $\delta = 0$ are the empirical sizes under D1.

To conclude, we report the run-times of a single replication based on designs D1 in the computing environment of Section 4. As before, we only report our tests with the smallest and the largest k_n, based on $\gamma_n = 0.01/\log n$. Overall, Table E.IV supports our previous claim on the relative computational simplicity of our tests—when comparing run-times across shapes and the dimensions of covariates, keep in mind that the fineness of discretization varies. When working with real data sets, one may increase the number of grid points and the number of bootstrap repetitions, as the computational cost is no more than one Monte Carlo simulation replication.

E.2. *Empirical Application*

To further illustrate the implementation of our test, we revisit the problem of option pricing functions under shape restrictions in financial economics. As forcefully argued

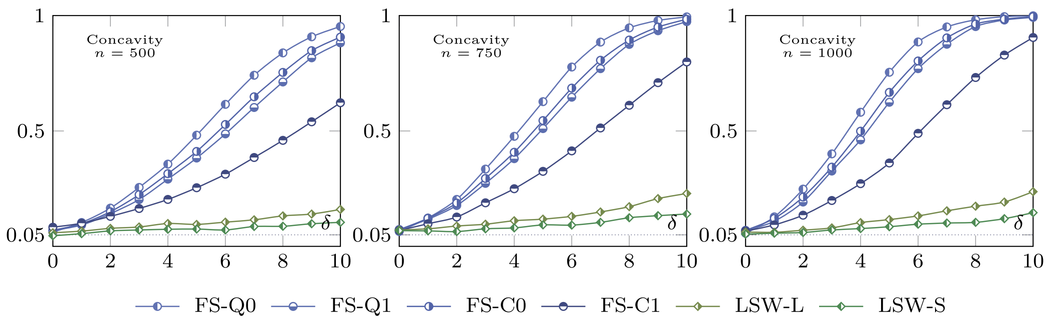

FIGURE E.2.—Empirical power of concavity tests for (E.1) where our tests are implemented with $\gamma_n = 0.01/\log n$ and corresponding to $\delta = 0$ are the empirical sizes under D1.

TABLE E.III

EMPIRICAL SIZE OF TESTING SLUTSKY RESTRICTION ON g_0 IN (E.2) AT $\alpha = 5\%$[a]

n	γ_n	FS-Q0: $k_n = 27$ D1	D2	D3	FS-Q1: $k_n = 64$ D1	D2	D3	FS-C0: $k_n = 64$ D1	D2	D3	FS-C1: $k_n = 125$ D1	D2	D3
1000	$1/n$	0.072	0.037	0.022	0.092	0.056	0.037	0.098	0.058	0.039	0.153	0.101	0.072
	$0.01/\log n$	0.072	0.037	0.022	0.092	0.056	0.037	0.098	0.058	0.039	0.153	0.101	0.072
	0.01	0.073	0.037	0.022	0.092	0.057	0.037	0.098	0.058	0.039	0.155	0.101	0.072
3000	$1/n$	0.054	0.019	0.009	0.065	0.026	0.014	0.065	0.024	0.012	0.083	0.035	0.016
	$0.01/\log n$	0.054	0.019	0.009	0.065	0.026	0.014	0.065	0.024	0.012	0.083	0.035	0.016
	0.01	0.054	0.019	0.009	0.065	0.026	0.014	0.065	0.024	0.012	0.084	0.036	0.017
5000	$1/n$	0.056	0.020	0.008	0.067	0.022	0.009	0.066	0.021	0.008	0.067	0.023	0.009
	$0.01/\log n$	0.056	0.020	0.008	0.067	0.022	0.009	0.066	0.021	0.008	0.067	0.023	0.009
	0.01	0.056	0.020	0.008	0.067	0.022	0.009	0.066	0.021	0.008	0.067	0.023	0.009

[a] *Note*: The parameter γ_n determines $\hat{\kappa}_n$ proposed in Section 3.2 with $c_n = 1/\log n$ and $r_n = (n/k_n)^{1/2}$.

in the literature, parametric models are barely grounded in financial theory and may be inadequate in capturing key aspects of the relationship under consideration. This has spurred a line of research on nonparametric estimation of option pricing functions under shape restrictions (Aït-Sahalia and Duarte (2003), Birke and Pilz (2009)). In particular, completeness of the market and absence of arbitrage opportunities imply two prominent restrictions: monotonicity and convexity of the call/put option price with respect to the strike price of the option, at a specific valuation date and for the same time-to-expiration. Below, we complement the literature by testing the validity of these restrictions.

We approach the problem in the setup of Example 2.1 following the aforementioned studies, where Y denotes the option price and Z the corresponding strike price. We aim to test three shape restrictions on θ_0, that is, monotonicity, convexity, and monotonicity jointly with convexity. While θ_0 should be convex for both call and put options, θ_0 should be nonincreasing for the former and nondecreasing for the latter. We make use of the data set analyzed in Beare and Schmidt (2016), which consists of prices for European call and put options written on the S&P 500 index—see Section 4 in Beare and Schmidt (2016) for detailed descriptions of the data set. We select two dates for our test problems: October 22, 2008 which has the maximal number of call options ($n = 93$), and October 19, 2011 which has the maximal number of put options ($n = 143$). Such small sample sizes,

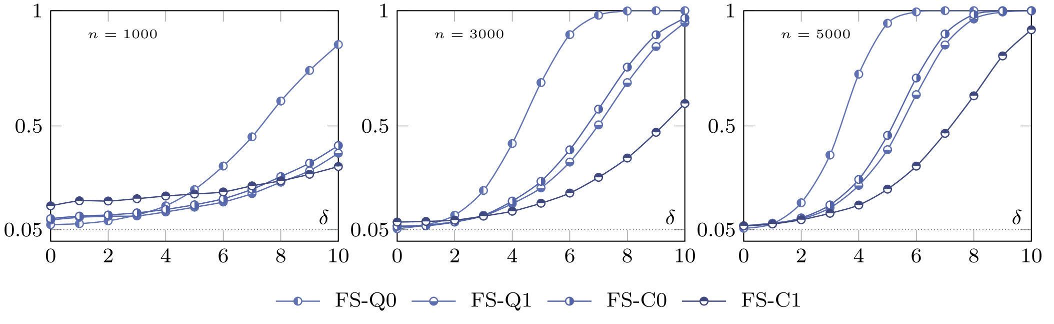

FIGURE E.3.—Empirical power of testing Slutsky restriction on g_0 in (E.3) with $\gamma_n = 0.01/\log n$, where corresponding to $\delta = 0$ are the empirical sizes under D1.

TABLE E.IV

RUN-TIMES (IN SECONDS) OF SHAPE TESTS[a]

	Convexity: (26)				Mon-Con: (26)				Concavity: (E.1)				Slutsky: (E.2)	
	FS		LSW		FS		LSW		FS		LSW		FS	
n	C3	C7	L	S	Q0	C1	L	S	Q0	C1	L	S	Q0	C1
500	0.24	0.24	23.02	22.32	0.24	0.27	23.61	23.25	16.05	17.09	13.16	13.29	9.59	16.85
750	0.25	0.26	56.42	57.43	0.26	0.29	57.56	57.59	14.96	16.53	38.76	38.78	9.80	17.23
1000	0.25	0.25	102.21	101.64	0.26	0.26	101.41	102.78	16.12	17.12	68.08	69.11	10.45	19.00

[a]*Note*: The sample sizes for the Slutsky restriction from top to bottom should be 1000, 3000, and 5000.

while not uncommon in practice, may raise concerns on the performance of our test. In unreported simulations based on the univariate designs in Section 4, we found that, with the sample size equal to $n = 100$ and $\gamma_n \in \{0.01/\log n, 1/n\}$, series estimation based on quadratic B-splines with two interior knots (labeled Q2) and cubic B-splines with one knot (labeled C1) delivers null rejection rates no larger than 0.068 (at 5% nominal level) and reasonable power (over 0.5 at $\delta = 10$). Thus, our implementation below will be based on these choices of splines and knots.

The remaining details of the implementation are the same as those in Sections 4 and E.1 (for the univariate designs) beyond the following changes. First, the strike prices are converted via the affine transformation $z \mapsto 2(z-a)/(b-a)-1$, with a and b respectively the minimal and maximal strike prices in the data. As a result, the converted strike prices fall within the range $[-1, 1]$ (to be consistent with Section 4) without changing the shape restrictions under consideration. Second, the number of bootstrap repetitions is increased to 1000, while the step size for numerical integration is decreased to 0.01. These changes echo our previous remark that, in applications, "one may increase the number of grid points and the number of bootstrap repetitions, as the computational cost is no more than one Monte Carlo simulation replication."

Table E.V reports the p-values of our test (with $\gamma_n = 0.01/\log n$). We fail to reject the three null hypotheses for both call and put options, at all conventional significance levels. In some cases, there are sizable discrepancies in the p-values across Q2 and C1 (for the same shape). This may be explained by the small sample issue, which is also in line with our simulation results for the Slutsky restrictions (those with high ratios of k_n/n). Overall, though, our findings point to strong evidences of the presence of the three shape restrictions (in the present rather restrictive setting).

TABLE E.V

TESTING SHAPE RESTRICTIONS OF OPTION PRICING FUNCTIONS: p-VALUES

Call options						Put options					
Monotonicity		Convexity		Mon-Con		Monotonicity		Convexity		Mon-Con	
Q2	C1	Q2	C1	Q2	C1	Q2	C1	Q2	C1	Q2	C1
0.70	0.22	0.55	0.14	0.61	0.30	0.57	0.76	0.36	0.89	0.72	0.86

REFERENCES

ABREVAYA, J., AND W. JIANG (2005): "A Nonparametric Approach to Measuring and Testing Curvature," *Journal of Business & Economic Statistics*, 23, 1–19. [2]

AGUIAR, V. H., AND R. SERRANO (2017): "Slutsky Matrix Norms: The Size, Classification, and Comparative Statics of Bounded Rationality," *Journal of Economic Theory*, 172, 163–201. [3]

AÏT-SAHALIA, Y., AND J. DUARTE (2003): "Nonparametric Option Pricing Under Shape Restrictions," *Journal of Econometrics*, 116, 9–47. [4,19]

ALIPRANTIS, C. D., AND K. BORDER (2006): *Infinite Dimensional Analysis: A Hitchhiker's Guide* (Third Ed.). Berlin: Springer Verlag. [8,9,14]

BARLOW, R. E., D. J. BARTHOLOMEW, J. M. BREMNER, AND H. D. BRUNK (1972): *Statistical Inference Under Order Restrictions: The Theory and Application of Isotonic Regression*. Wiley. [2,4]

BEARE, B. K., AND J.-M. MOON (2015): "Nonparametric Tests of Density Ratio Ordering," *Econometric Theory*, 31, 471–492. [4]

BEARE, B. K., AND L. D. W. SCHMIDT (2016): "An Empirical Test of Pricing Kernel Monotonicity," *Journal of Applied Econometrics*, 31, 338–356. [19]

BERESTEANU, A. (2007): "Nonparametric Estimation of Regression Functions Under Restrictions on Partial Derivatives," Duke Economics Working Paper #04-06. [3]

BIRKE, M., AND K. F. PILZ (2009): "Nonparametric Option Pricing With No-Arbitrage Constraints," *Journal of Financial Econometrics*, 7, 53–76. [19]

BOGACHEV, V. I. (1998): *Gaussian Measures*. Providence: American Mathematical Society. [9]

BOGACHEV, V. I., AND O. G. SMOLYANOV (2017): *Topological Vector Spaces and Their Applications*. Springer. [14]

CHEN, Q., AND Z. FANG (2019): "Improved Inference on the Rank of a Matrix," *Quantitative Economics*, 10, 1787–1824. [3,6]

CHEN, X., AND K. KATO (2019): "Jackknife Multiplier Bootstrap: Finite Sample Approximations to the *U*-Process Supremum With Applications," *Probability Theory and Related Fields*, 1–67. [2,3]

CHEN, X., V. CHERNOZHUKOV, I. FERNÁNDEZ-VAL, S. KOSTYSHAK, AND Y. LUO (2020): "Shape-Enforcing Operators for Point and Interval Estimators," arXiv preprint arXiv:1809.01038. [2]

CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, AND A. GALICHON (2010): "Quantile and Probability Curves Without Crossing," *Econometrica*, 78, 1093–1125. [2]

CHERNOZHUKOV, V., W. K. NEWEY, AND A. SANTOS (2015): "Constrained Conditional Moment Restriction Models," arXiv preprint arXiv:1509.06311. [14,15]

CHETVERIKOV, D. (2019): "Testing Regression Monotonicity in Econometric Models," *Econometric Theory*, 35, 729–776. [2,3]

DAVYDOV, Y. A., M. A. LIFSHITS, AND N. V. SMORODINA (1998): *Local Properties of Distributions of Stochastic Functionals*. American Mathematical Society. [12-14]

DETTE, H., S. HODERLEIN, AND N. NEUMEYER (2016): "Testing Multivariate Economic Restrictions Using Quantiles: The Example of Slutsky Negative Semidefiniteness," *Journal of Econometrics*, 191, 129–144. [3]

DEUTSCH, F. R. (2012): *Best Approximation in Inner Product Spaces*. Springer. [3,10,11]

FAN, J., AND I. GIJBELS (1996): *Local Polynomial Modelling and Its Applications*. CRC Press. [15]

FANG, Z. (2019): "Refinements of the Kiefer–Wolfowitz Theorem and a Test of Concavity," *Electronic Journal of Statistics*, 13, 4596–4645. [4]

FANG, Z., AND A. SANTOS (2019): "Inference on Directionally Differentiable Functions," *The Review of Economic Studies*, 86, 377–412. [5,6]

GINÉ, E., AND R. NICKL (2016): *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press. [9]

GINÉ, E., AND J. ZINN (1991): "Gaussian Characterization of Uniform Donsker Classes of Functions," *The Annals of Probability*, 19, 758–782. [5]

GUERRE, E., I. PERRIGNE, AND Q. VUONG (2009): "Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions," *Econometrica*, 77, 1193–1227. [4]

HALL, P., AND L.-S. HUANG (2001): "Nonparametric Kernel Regression Subject to Monotonicity Constraints," *The Annals of Statistics*, 29, 624–647. [2]

HAN, Q., AND J. A. WELLNER (2016): "Multivariate Convex Regression: Global Risk Bounds and Adaptation," arXiv preprint arXiv:1601.06844. [2]

HAN, Q., T. WANG, S. CHATTERJEE, AND R. J. SAMWORTH (2019): "Isotonic Regression in General Dimensions," *The Annals of Statistics*, 47, 2440–2471. [2]

HODERLEIN, S. (2011): "How Many Consumers Are Rational?" *Journal of Econometrics*, 164, 294–309. [3]

HONG, H., AND J. LI (2018): "The Numerical Delta Method," *Journal of Econometrics*, 206, 379–394. [6]

KUOSMANEN, T. (2008): "Representation Theorem for Convex Nonparametric Least Squares," *The Econometrics Journal*, 11, 308–325. [2]
KWAPIEŃ, S. (1994): "A Remark on the Median and the Expectation of Convex Functions of Gaussian Vectors," in *Probability in Banach Spaces, 9*, ed. by J. Hoffmann-Jørgensen, J. Kuelbs, and M. B. Marcus. Boston, MA: Birkhäuser, 271–272. [8,13]
LEE, S., O. LINTON, AND Y.-J. WHANG (2009): "Testing for Stochastic Monotonicity," *Econometrica*, 77, 585–602. [4]
LEE, S., K. SONG, AND Y.-J. WHANG (2017): "Testing for a General Class of Functional Inequalities," *Econometric Theory*, 1–47. [14]
LIM, E., AND P. W. GLYNN (2012): "Consistency of Multidimensional Convex Regression," *Operations Research*, 60, 196–208. [2]
LINTON, O., K. E. SONG, AND Y.-J. WHANG (2010): "An Improved Bootstrap Test of Stochastic Dominance," *Journal of Econometrics*, 154, 186–202. [4]
MAMMEN, E., J. S. MARRON, B. A. TURLACH, AND M. P. WAND (2001): "A General Projection Framework for Constrained Smoothing," *Statistical Science*, 16, 232–248. [2]
MASTEN, M. A., AND A. POIRIER (2021): "Inference on Breakdown Frontiers," *Quantitative Economics*, 11, 41–111. [6]
PAKES, A., J. PORTER, K. HO, AND J. ISHII (2015): "Moment Inequalities and Their Application," *Econometrica*, 83, 315–334. [3]
PINKSE, J., AND K. SCHURTER (2019): "Estimation of Auction Models With Shape Restrictions," arXiv preprint arXiv:1912.07466. [4]
SHEEHY, A., AND J. A. WELLNER (1992): "Uniform Donsker Classes of Functions," *The Annals of Probability*, 20, 1983–2030. [5]
SONG, Y., X. CHEN, AND K. KATO (2020): "Stratified Incomplete Local Simplex Tests for Curvature of Nonparametric Multiple Regression," arXiv preprint arXiv:2003.09091. [3]
SUN, Z., AND B. K. BEARE (2019): "Improved Nonparametric Bootstrap Tests of Lorenz Dominance," *Journal of Business & Economic Statistics*, 1–11. [4]
TOPKIS, D. M. (1998): *Supermodularity and Complementarity*. Princeton University Press. [3]
VAKHANIA, N., V. TARIELADZE, AND S. CHOBANYAN (1987): *Probability Distributions on Banach Spaces*. Springer. [11,14]
VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. New York: Springer Verlag. [9,11,14]
YURINSKY, V. V. (1995): *Sums and Gaussian Vectors*. Berlin: Springer-Verlag. [12,14]
ZEIDLER, E. (1985): *Nonlinear Functional Analysis and Its Applications, III*. Springer. [10]
ZHU, Y. (2020): "Inference in Nonparametric/Semiparametric Moment Equality Models With Shape Restrictions," *Quantitative Economics*, 11, 609–636. [14]

Co-editor Guido Imbens handled this manuscript.

Manuscript received 16 October, 2019; final version accepted 15 March, 2021; available online 19 March, 2021.