# A Projection Framework for Testing Shape Restrictions That Form Convex Cones*

Zheng Fang
Department of Economics
Texas A&M University
zfang@tamu.edu

Juwon Seo
Department of Economics
National University of Singapore
ecssj@nus.edu.sg

February 10, 2021

**Abstract**

This paper develops a uniformly valid and asymptotically nonconservative test based on projection for a class of shape restrictions. The key insight we exploit is that these restrictions form convex cones, a simple and yet elegant structure that has been barely harnessed in the literature. Based on a monotonicity property afforded by such a geometric structure, we construct a bootstrap procedure that, unlike many studies in nonstandard settings, dispenses with estimation of local parameter spaces, and the critical values are obtained in a way as simple as computing the test statistic. Moreover, by appealing to strong approximations, our framework accommodates nonparametric regression models as well as distributional/density-related and structural settings. Since the test entails a tuning parameter (due to the nonstandard nature of the problem), we propose a data-driven choice and prove its validity. Monte Carlo simulations confirm that our test works well.

KEYWORDS: Nonstandard inference, Shape restrictions, Convex cone, Projection, Strong approximations

*We thank three anonymous referees for valuable suggestions that have helped greatly improve this paper. We are indebted to Andres Santos for extensive discussions and constant support. We thank Xiaohong Chen, Denis Chetverikov, Stefan Hoderlein, Tatiana Komarova, Scott Kostyshak, Simon Lee, Sungwon Lee, Ulrich K. Müller, Pepe Montiel Olea, Alex Poirier, Brandon Reeves, Kevin Song, Yanglei Song, Xun Tang and numerous seminar participants for their useful comments and generous help. We also thank Denis Chetverikov and Simon Lee for sharing their code. Advanced computing resources and consultation provided by Texas A&M High Performance Research Computing and Marinus Pennings are gratefully acknowledged. A previous version of this paper was circulated under the title "A General Framework for Inference on Shape Restrictions."

# 1 Introduction

Shape restrictions play a number of fundamental and prominent roles in economics. For example, they often arise as testable implications of economic theory, and may thus serve as plausible restrictions in specifying economic models (Varian, 1982, 1984). In complementary empirical work, they help achieve point identification, tighten identification bounds, improve estimation precision, and develop powerful tests—see Matzkin (1994) and Chetverikov et al. (2018) for more detailed discussions.

In this paper, we develop a uniformly valid and asymptotically nonconservative test for a class of shape restrictions. To illustrate, consider the regression model:

$$Y = \theta_0(Z) + u \ , \tag{1}$$

where $Z \in [0,1]$, $\theta_0 : [0,1] \to \mathbf{R}$ and $E[u|Z] = 0$. Suppose that we are interested in testing whether $\theta_0$ is nondecreasing. More formally, if $\theta_0 \in \mathbf{H} \equiv \{\theta : [0,1] \to \mathbf{R} : \|\theta\|_{\mathbf{H}} \equiv \{\int_{[0,1]} |\theta(z)|^2 \,\mathrm{d}z\}^{1/2} < \infty\}$ (a mild restriction) and $\Lambda$ is the class of nondecreasing functions in $\mathbf{H}$, then we may formulate the hypotheses as

$$\mathrm{H}_0 : \phi(\theta_0) = 0 \qquad \text{vs.} \qquad \mathrm{H}_1 : \phi(\theta_0) > 0 \ , \tag{2}$$

where $\phi(\theta) \equiv \min_{\lambda\in\Lambda} \|\theta - \lambda\|_{\mathbf{H}}$ is the distance from $\theta$ to $\Lambda$. Thus, given an unconstrained (kernel or sieve) estimator $\hat{\theta}_n$ of $\theta_0$, we may employ a test that rejects $\mathrm{H}_0$ if $r_n\phi(\hat{\theta}_n)$ is "large" for a suitable $r_n \to \infty$. While conceptually intuitive, construction of critical values turns out to be a delicate and challenging matter. In particular, despite the well established results on the rate $r_n$ and pointwise asymptotic normality, $r_n\{\hat{\theta}_n - \theta_0\}$ generically does not converge as a process indexed by $[0,1]$ (Chernozhukov et al., 2013), rendering the Delta method as in Fang and Santos (2019) inapplicable.

As a first step, we note that $\Lambda$ being a (closed) convex cone[1] implies

$$r_n\phi(\hat{\theta}_n) = \|r_n\{\hat{\theta}_n - \theta_0\} + r_n\theta_0 - \Pi_\Lambda(r_n\{\hat{\theta}_n - \theta_0\} + r_n\theta_0)\|_{\mathbf{H}} \ , \tag{3}$$

where $\theta \mapsto \Pi_\Lambda(\theta) \equiv \arg\min_{\lambda\in\Lambda} \|\theta - \lambda\|_{\mathbf{H}}$ is the projection operator, i.e., $\Pi_\Lambda(\theta)$ is the closest (under $\|\cdot\|_{\mathbf{H}}$) nondecreasing function to $\theta$. Hence, (3) reveals that, in estimating the law of $r_n\phi(\hat{\theta}_n)$ in order to obtain critical values, it suffices to quantify the variation of $r_n\{\hat{\theta}_n - \theta_0\}$ and estimate the drift $r_n\theta_0$. Despite the lack of convergence (in $\mathbf{H}$), $r_n\{\hat{\theta}_n - \theta_0\}$ as a process may be approximated in law via strong approximations by a number of methods with only mild computation cost, including the simulation method in Chernozhukov et al. (2013), the weighted bootstrap in Belloni et al. (2015), and the sieve score bootstrap in Chen and Christensen (2018). The treatment of $r_n\theta_0$, on the

[1]By definition, $\Lambda$ is a convex cone if and only if $af + bg \in \Lambda$ whenever $a, b \geq 0$ and $f, g \in \Lambda$.

other hand, consists of the nonstandard step because, as in moment inequality models (Andrews and Soares, 2010), $r_n\theta_0$ cannot be consistently estimated in general. In this regard, the convex cone property (but not convexity alone) implies

$$r_n\phi(\hat{\theta}_n) \leq \|r_n\{\hat{\theta}_n - \theta_0\} + \kappa_n\theta_0 - \Pi_\Lambda(r_n\{\hat{\theta}_n - \theta_0\} + \kappa_n\theta_0)\|_{\mathbf{H}} \ , \tag{4}$$

whenever $0 \leq \kappa_n \leq r_n$. Hence, a valid critical value may be obtained by bootstrapping the upper bound in (4), which is possible because $\kappa_n\theta_0$ may be consistently estimated by $\kappa_n\hat{\theta}_n$ provided $\kappa_n/r_n \to 0$. The very virtue that (4) is an inequality rather than equality may also raise concern for conservativeness of the resulting test. As an extreme case, the choice $\kappa_n = 0$ leads to a least favorable test that may be viewed as assuming $\theta_0 = 0$, suggesting that $\kappa_n$ should not be too small. Along these lines we develop a data-driven choice of $\kappa_n$ that delivers an asymptotically nonconservative test.

While we started with monotonicity in the univariate model (1), the main features of our test are not confined to this special problem. First, the convex cone property is in fact shared by a class of restrictions, e.g., nonnegativity, concavity, Slutsky restriction and supermodularity—see Appendix B. Regrettably, our framework is not directly applicable absent the property, e.g., quasi-concavity, log-concavity and $r$-concavity (Kostyshak, 2017; Komarova and Hidalgo, 2019). Second, our framework is applicable to distributional and structural settings (Pinkse and Schurter, 2019; Bhattacharya, 2020)—see Appendix B. Third, our framework allows for jointly testing multiple restrictions since intersections of convex cones remain convex cones. This is important because it is common that shape restrictions arise simultaneously. Fourth, while some convex-cone restrictions (e.g., monotonicity and concavity) on $\theta_0$ can be characterized as inequalities on derivatives of $\theta_0$, we dispense with derivative estimation because it likely incurs power loss due to a slower convergence rate (Stone, 1982; Chen and Christensen, 2018). Finally, our test does not rely on the least favorable configurations, while being asymptotically nonconservative (thus leading to improved power) and computationally tractable. We stress that, since the norm $\|\cdot\|_{\mathbf{H}}$ is of $L^2$ nature, our test cannot be inverted to obtain uniform confidence bands. In this sense, our paper complements Horowitz and Lee (2017), Freyberger and Reeves (2018) and Chen et al. (2020).

The literature on shape restrictions was initiated in the 1950s (Barlow et al., 1972), with much attention since focused on estimation under solely shape restrictions—see, e.g., Han et al. (2019) and references therein. There are also post-processing methods that enforce restrictions on unconstrained estimators—see Chen et al. (2020) who study a number of shape enforcing operators. The projection method in this paper is of post-processing nature. The convex cone structure was recognized in the 1960s as a device to generalize monotone regression, though the focus is on *analytic* properties of projections (Barlow et al., 1972). For testing, the structure has barely been exploited beyond identifying the least favorable distributions in parametric settings (Wolak, 1987;

Silvapulle and Sen, 2005), deriving minimax bounds in univariate Gaussian white noise models (Juditsky and Nemirovski, 2002), and establishing minimaxity for the likelihood ratio test in Gaussian sequence models (Wei et al., 2019).

Much of the testing literature has been developed by exploiting *particular* structures of restrictions, with concavity and especially monotonicity in univariate models being the primary focuses. Despite a sizable literature, existing tests prevalently rely on the least favorable configurations, including Gijbels et al. (2000), Ghosal et al. (2000), Hall and Heckman (2000), Durot (2003) and Gutknecht (2016) for monotonicity, and Abrevaya and Jiang (2005) for concavity—see also Dümbgen and Spokoiny (2001) and Baraud et al. (2005) who devised tests for both shapes. Chetverikov (2019) developed, as far as we are aware, the first nonparametric uniformly valid tests designed specifically for monotonicity that avoid the least favorable configurations.

There is also a strand of literature motivated by the *common* structure of shape restrictions viewed as inequalities, including Chernozhukov et al. (2015), Lee et al. (2017), Belloni et al. (2019) and Zhu (2020). If the inequalities are based on derivatives, the derivative estimation may then incur power loss. Moreover, all these (uniformly valid) tests, except the least favorable test of Belloni et al. (2019), involve estimation of local parameter spaces. Chernozhukov et al. (2015) and Zhu (2020), unlike our setup, allow for partial identification by working with moment restrictions. By the virtue of their setup, they require estimating the set of minimizers and the use of strong approximations may entail stronger regularity conditions. Similar in spirit to Chernozhukov et al. (2015) and Zhu (2020), Komarova and Hidalgo (2019) propose a moment-based test in the univariate model (1) for shape restrictions that may not form convex cones.

The remainder of the paper is structured as follows. Section 2 introduces the setup and some motivating examples. Section 3 presents our inferential framework, a data-driven choice of the tuning parameter, and an implementation guide. Section 4 conducts simulation studies, while Section 5 concludes. Appendix A contains proofs of the main results. Due to space limitation, we relegate the discussions of the convex cone property, an investigation of the special case when $\hat{\theta}_n$ admits an asymptotic distribution, and presentations of some auxiliary results to the online supplement.

## 2 The Setup and Examples

Throughout, we denote by $\{X_i\}_{i=1}^n$ the sample with each $X_i$ living in some sample space $\mathcal{X}$, and by $P$ the joint law of $\{X_i\}_{i=1}^n$ that belongs to some family $\mathbf{P}$ of distributions on $\mathcal{X}^n$. The dependence of $P$ and $\mathbf{P}$ on $n$ is suppressed for notational simplicity. We stress that, under this configuration, $\{X_i\}_{i=1}^n$ need not be i.i.d. In turn, we let $\theta_0$ be the parameter of interest, and, whenever appropriate, make the dependence of $\theta_0$ on $P$ explicit by instead writing $\theta_P$. In order to accommodate Slutsky restriction, we shall work with an abstract Hilbert space (i.e., a complete inner product space) with inner product $\langle\cdot,\cdot\rangle_{\mathbf{H}}$ and induced norm $\|\cdot\|_{\mathbf{H}}$. Given the sample $\{X_i\}_{i=1}^n$, our objective is then to test whether $\theta_0$ satisfies the shape in question, i.e., the hypotheses in (2). The setup (2) induces two models: $\mathbf{P}_0 \equiv \{P \in \mathbf{P} : \phi(\theta_P) = 0\}$ and $\mathbf{P}_1 \equiv \mathbf{P}\backslash\mathbf{P}_0$.

Before proceeding further, we introduce additional notation. Set $\mathbf{R}_+ \equiv \{x \in \mathbf{R} : x \geq 0\}$ and $L^2(\mathcal{Z}) \equiv \{f : \mathcal{Z} \to \mathbf{R} : \int_{\mathcal{Z}} |f(z)|^2 \,\mathrm{d}z < \infty\}$ for $\mathcal{Z} \subset \mathbf{R}^{d_z}$. Let $\mathbf{M}^{m\times k}$ be the space of $m \times k$ matrices, and, for $A \in \mathbf{M}^{m\times k}$, write its transpose by $A^\intercal$, its trace by $\mathrm{tr}(A)$ if $m = k$, its Moore–Penrose inverse by $A^-$ and its Frobenius norm by $\|A\| \equiv \{\mathrm{tr}(A^\intercal A)\}^{1/2}$. For a sequence $\{h_j\}$ of functions, denote the vector $(h_1, \ldots, h_k)^\intercal$ by $h^k$. For generic families of distributions $\mathbf{P}_n$, a sequence $\{a_n\}$ of positive scalars and a sequence $\{\mathbb{X}_n\}$ of random elements in a normed space $\mathbf{D}$ with norm $\|\cdot\|_{\mathbf{D}}$, write $\mathbb{X}_n = o_p(a_n)$ uniformly in $P \in \mathbf{P}_n$ if $\lim_{n\to\infty} \sup_{P\in\mathbf{P}_n} P(\|\mathbb{X}_n\|_{\mathbf{D}} > a_n\epsilon) = 0$ for any $\epsilon > 0$, and $\mathbb{X}_n = O_p(a_n)$ uniformly in $P \in \mathbf{P}_n$ if $\lim_{M\to\infty} \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_n} P(\|\mathbb{X}_n\|_{\mathbf{D}} > Ma_n) = 0$.

## 2.1 Examples

We now present examples where shape restrictions play important roles. The first example is concerned with nonparametric regression models.

**Example 2.1** (Nonparametric Regression)**.** Let $X \equiv (Y, Z) \in \mathbf{R}^{1+d_z}$ satisfy

$$Y = \theta_0(Z) + u \ , \tag{5}$$

where $\theta_0 : \mathcal{Z} \subset \mathbf{R}^{d_z} \to \mathbf{R}$ with $\mathcal{Z}$ the support of $Z$, and $E[u|Z] = 0$. Here, one may set $\mathbf{H} = L^2(\mathcal{Z})$, let $\Lambda \subset \mathbf{H}$ consist of (say) monotonic functions, and obtain an unconstrained estimator $\hat{\theta}_n$ of $\theta_0$ by kernel methods such as local constant/linear/polynomial regression or sieve methods with various basis functions such as splines (Chernozhukov et al., 2013). The rate $r_n$ equals $(nh_n^{d_z})^{1/2}$ for kernel estimation with bandwidth $h_n$, and $(n/k_n)^{1/2}$ for sieve estimation based on, e.g., B-splines, with $k_n$ (henceforth) the sieve dimension. One may also set up this example as one on the conditional mean $z \mapsto E[Y|Z = z]$. ■

Our second example generalizes (5) to its instrumental variable (IV) analog.

**Example 2.2** (Nonparametric IV Regression)**.** Let $X \equiv (Y, Z, V) \in \mathbf{R}^{1+d_z+d_v}$ satisfy (5) but with $E[u|V] = 0$. Then one may set $\mathbf{H}$ and $\Lambda$ as in Example 2.1, and employ the series two-stage least square (2SLS) estimation with, e.g., B-splines, resulting in a rate $r_n$ equal to $(n/k_n)^{1/2}s_n$, where $s_n$ is the smallest singular value of $E[b^{m_n}(V)h^{k_n}(Z)^\intercal]$ with $h^{k_n}$ and $b^{m_n}$ respectively $k_n \times 1$ and $m_n \times 1$ vectors (with $m_n \geq k_n$) of B-splines for $\theta_0$ and the instrument space (Ai and Chen, 2003; Newey and Powell, 2003; Chen and Christensen, 2018). In practice, $s_n$ is unknown but can be replaced with its sample analog. One may conceivably also employ kernel-type estimators (Hall and Horowitz, 2005; Darolles et al., 2011), though the strong approximation results appear to be lacking. ■

The third example is concerned with nonparametric quantile regression.

**Example 2.3** (Nonparametric Quantile Regression)**.** Let $X \equiv (Y, Z) \in \mathbf{R}^{1+d_z}$ satisfy

$$Y = \theta_0(Z, U) , \tag{6}$$

where $\theta_0 : \mathbf{R}^{d_z} \times [0,1] \to \mathbf{R}$, and $U$ is uniformly distributed on $[0,1]$. If $Z$ and $U$ are independent, then $\theta_0$ may be interpreted as the conditional quantile function of $Y$ given $Z$. Here, one may define $\mathbf{H}$ and $\Lambda$ as previously, and employ the recent sieve estimator of Belloni et al. (2019) with $r_n = (n/k_n)^{1/2}$ for, e.g., B-splines. ■

Our final example concerns a restriction in a possibly endogenous regression model.

**Example 2.4** (Slutsky Restriction)**.** Let $Q \in \mathbf{R}^{d_q}$ (quantities), $P \in \mathbf{R}^{d_q}$ (prices), $Y \in \mathbf{R}$ (income) and $Z \in \mathbf{R}^{d_z}$ (demographics) satisfy the following $d_q$ equations:

$$Q = g_0(P, Y) + \Gamma_0^\intercal Z + U , \tag{7}$$

where $g_0 : \mathbf{R}_+^{d_q+1} \to \mathbf{R}^{d_q}$ is differentiable, $\Gamma_0 \in \mathbf{M}^{d_z \times d_q}$, and $U \in \mathbf{R}^{d_q}$ is the error term. The Slutsky matrix of $g_0$ is the mapping $\theta_0 : \mathbf{R}_+^{d_q+1} \to \mathbf{M}^{d_q \times d_q}$ defined by:

$$\theta_0(p, y) \equiv \frac{\partial g_0(p, y)}{\partial p^\intercal} + \frac{\partial g_0(p, y)}{\partial y} g_0(p, y)^\intercal . \tag{8}$$

The Slutsky restriction refers to $\theta_0(p, y)$ being negative semidefinite (nsd) at each pair $(p, y)$. For this example, we follow Aguiar and Serrano (2017) by letting $\mathbf{H}$ be the space of functions $\theta : \mathbf{R}_+^{d_q+1} \to \mathbf{M}^{d_q \times d_q}$ with inner product $\langle \cdot, \cdot \rangle_{\mathbf{H}}$ defined by

$$\langle \theta_1, \theta_2 \rangle_{\mathbf{H}} \equiv \int \mathrm{tr}(\theta_1(p, y)^\intercal \theta_2(p, y)) \, \mathrm{d}p\mathrm{d}y , \ \forall\, \theta_1, \theta_2 \in \mathbf{H} , \tag{9}$$

and $\Lambda$ the family of mappings $\theta \in \mathbf{H}$ such that $\theta(p, y)$ is nsd and symmetric for all $(p, y)$. In turn, the nonlinear functional $\theta_0$ may be estimated based on the plug-in principle and a sieve estimator of $g_0$ (Donald and Newey, 1994; Ai and Chen, 2003), resulting in a rate $r_n$ as in Example 2.1 (without endogeneity) or 2.2 (with endogeneity). ■

We refer the reader to Appendix B that illustrates how our framework also applies to distributional settings where shape restrictions often take the form of various dominance relations and structural models where shape restrictions may serve as testable implications. For ease of reference, we shall call settings where $\hat{\theta}_n$ admits an asymptotic distribution (in $\mathbf{H}$) *regular* and ones without this property *irregular*.

## 3 The Inferential Framework

### 3.1 The General Framework

We commence with our main assumptions in this paper.

**Assumption 3.1.** *(i) $\Lambda$ is a known nonempty closed convex set in a Hilbert space $\mathbf{H}$ with known inner product $\langle\cdot,\cdot\rangle_{\mathbf{H}}$ and induced norm $\|\cdot\|_{\mathbf{H}}$; (ii) $\Lambda$ is a cone.*

**Assumption 3.2.** *(i) An estimator $\hat{\theta}_n : \{X_i\}_{i=1}^n \to \mathbf{H}$ satisfies $\|r_n\{\hat{\theta}_n - \theta_P\} - \mathbb{Z}_{n,P}\|_{\mathbf{H}} = o_p(c_n)$ uniformly in $P \in \mathbf{P}$ for some $r_n \to \infty$, $\mathbb{Z}_{n,P} \in \mathbf{H}$ and $c_n > 0$ with $c_n = O(1)$; (ii) $\hat{\mathbb{G}}_n \in \mathbf{H}$ is a bootstrap estimator satisfying: $\|\hat{\mathbb{G}}_n - \bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}} = o_p(c_n)$ uniformly in $P \in \mathbf{P}$, for $\bar{\mathbb{Z}}_{n,P}$ a copy of $\mathbb{Z}_{n,P}$ that is independent of $\{X_i\}_{i=1}^n$.*

Assumption 3.1 simply abstracts the convex cone feature, where we single out the conic condition for ease of elucidating the roles it plays in this paper. Assumption 3.2(i) requires that $r_n\{\hat{\theta}_n - \theta_P\}$ be approximated by $\mathbb{Z}_{n,P}$ (uniformly) at a rate $c_n$. In regular settings, it suffices to have $\sqrt{n}\{\hat{\theta}_n - \theta_P\} \xrightarrow{L} \mathbb{G}_P$ uniformly in $P \in \mathbf{P}$, so that $r_n = \sqrt{n}$, and $\mathbb{Z}_{n,P}$ equals $\mathbb{G}_P$ in law—see Appendix C. In irregular settings such as Example 2.1, one may obtain $\hat{\theta}_n$ by kernel or sieve methods. To illustrate, let $\{Y_i, Z_i\}_{i=1}^n$ be a sample generated by (5), and $\{h_k\}_{k=1}^\infty$ be a $k_n \times 1$ vector of B-splines on $\mathcal{Z}$. Then the sieve estimator of $\theta_P$ is given by $\hat{\theta}_n = \hat{\beta}_n^\intercal h^{k_n}$ with $\hat{\beta}_n = [\sum_{i=1}^n h^{k_n}(Z_i)h^{k_n}(Z_i)^\intercal]^- \sum_{i=1}^n h^{k_n}(Z_i)Y_i$. Under regularity conditions, one may obtain the linear expansion in $L^2(\mathcal{Z})$:

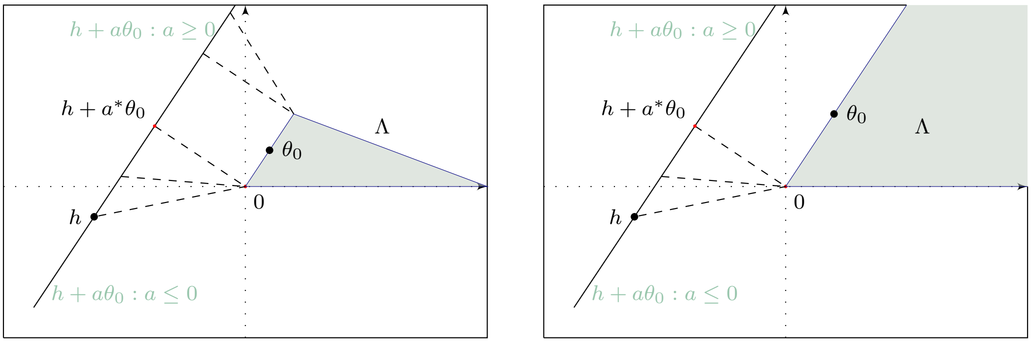
$$r_n\{\hat{\theta}_n - \theta_P\} = k_n^{-1/2}(h^{k_n})^\intercal(E_P[h^{k_n}(Z)h^{k_n}(Z)^\intercal])^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n h^{k_n}(Z_i)u_i + o_p(\frac{1}{\log n}) \ , \tag{10}$$

uniformly in $P \in \mathbf{P}$. Here, undersmoothing is required in order to deliver (10). Following Chernozhukov et al. (2013), one may then verify Assumption 3.2(i) with $r_n = \sqrt{n/k_n}$, $c_n = 1/\log n$ and $\mathbb{Z}_{n,P} = k_n^{-1/2}(h^{k_n})^\intercal(E_P[h^{k_n}(Z)h^{k_n}(Z)^\intercal])^{-1}G_{n,P}$ for some $G_{n,P} \sim N(0, E_P[u^2h^{k_n}(Z)h^{k_n}(Z)^\intercal])$ . The rate $c_n$ serves to cope with potential degeneracy of the test statistic, an issue inherent in the nonstandard nature of the problem.

Assumption 3.2(ii) demands an analogous approximation for the bootstrap. In regular settings, typically $\hat{\mathbb{G}}_n = \sqrt{n}\{\hat{\theta}_n^* - \hat{\theta}_n\}$, where $\hat{\theta}_n^*$ is the same as $\hat{\theta}_n$ but based on samples drawn from the original data. In Example 2.1, the expansion (10) suggests

$$\hat{\mathbb{G}}_n = k_n^{-1/2}(h^{k_n})^\intercal(\frac{1}{n}\sum_{i=1}^n h^{k_n}(Z_i)h^{k_n}(Z_i)^\intercal)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n W_i h^{k_n}(Z_i)\hat{u}_i \ . \tag{11}$$

where $\hat{u}_i \equiv Y_i - \hat{\theta}_n(Z_i)$ and $\{W_i\}_{i=1}^n$ are (scalar) weights (e.g., standard normals). The verification of Assumption 3.2 for Example 2.1 represents a general strategy: establishing an asymptotic linear expansion of $r_n\{\hat{\theta}_n - \theta_P\}$, and then verifying Assumption 3.2




(a) $\Lambda$ is convex but not conic

(b) $\Lambda$ is convex and conic

Figure 1. $\psi_{a,P}(h) \equiv \|h + a\theta_P - \Pi_\Lambda(h + a\theta_P)\|_{\mathbf{H}}$ is weakly decreasing in $a \in [0, \infty)$ if $\Lambda$ is convex and conic as in Figure 1b, but may not be so if it is convex but not conic as in Figure 1a.

based on the linear term—see Chernozhukov et al. (2013), Chernozhukov et al. (2015), Chen and Christensen (2018), Belloni et al. (2019), and Li and Liao (2020) for more illustrations. We stress that Assumption 3.2(ii) leaves the particular form of $\hat{\mathbb{G}}_n$ unspecified, and thus accommodates alternative resampling schemes.

The next lemma lays out a number of important building blocks for our development.

**Lemma 3.1.** *(i) If Assumption 3.1 holds, then any $\hat{\theta}_n \in \mathbf{H}$ and $r_n \in \mathbf{R}_+$ satisfy*

$$r_n \phi(\hat{\theta}_n) = \|r_n\{\hat{\theta}_n - \theta_P\} + r_n\theta_P - \Pi_\Lambda(r_n\{\hat{\theta}_n - \theta_P\} + r_n\theta_P)\|_{\mathbf{H}} \ . \tag{12}$$

*(ii) If Assumption 3.1 holds, $P \in \mathbf{P}_0$ and $\kappa_n \in [0, r_n]$, then it follows that*

$$r_n \phi(\hat{\theta}_n) \leq \|r_n\{\hat{\theta}_n - \theta_P\} + \kappa_n\theta_P - \Pi_\Lambda(r_n\{\hat{\theta}_n - \theta_P\} + \kappa_n\theta_P)\|_{\mathbf{H}} \ . \tag{13}$$

*(iii) If Assumption 3.2(i) holds, $\sup_{P\in\mathbf{P}} E[\|\mathbb{Z}_{n,P}\|_{\mathbf{H}}] < \infty$ uniformly in $n$ and $\kappa_n/r_n = o(c_n)$, then we have that $\kappa_n\hat{\theta}_n - \kappa_n\theta_P = o_p(c_n)$ uniformly in $P \in \mathbf{P}$.*

Lemma 3.1(i) highlights the standard, and critically, the nonstandard features of the problem. Specifically, while $r_n\{\hat{\theta}_n - \theta_P\}$ may be approximated in law by $\hat{\mathbb{G}}_n$ due to Assumption 3.2, $r_n\theta_P$ cannot be consistently estimated in general, i.e., $r_n\hat{\theta}_n - r_n\theta_P \neq o_p(1)$. Lemma 3.1(ii) suggests that we may obtain critical values by instead estimating the law of the upper bound in (13). This is possible because $\kappa_n\theta_P$ can be consistently estimated by $\kappa_n\hat{\theta}_n$ as shown by Lemma 3.1(iii). We stress that (13) is implied by the convex cone property but not convexity alone—see Figure 1.

To formalize the above discussions, we define maps $\psi_{a,P}, \hat{\psi}_{\kappa_n} : \mathbf{H} \to \mathbf{R}$ by

$$\psi_{a,P}(h) \equiv \|h + a\theta_P - \Pi_\Lambda(h + a\theta_P)\|_{\mathbf{H}} \ , \tag{14}$$

$$\hat{\psi}_a(h) \equiv \|h + a\Pi_\Lambda\hat{\theta}_n - \Pi_\Lambda(h + a\Pi_\Lambda\hat{\theta}_n)\|_{\mathbf{H}} \ . \tag{15}$$

Thus, $\psi_{\kappa_n,P}(r_n\{\hat{\theta}_n - \theta_P\})$ is precisely the upper bound in (13), and $\hat{\psi}_{\kappa_n}(\hat{\mathbb{G}}_n)$ is its bootstrap analog in which the null hypothesis is enforced through $\Pi_\Lambda\hat{\theta}_n$ to improve power. Finally, for a significance level $\alpha \in (0,1)$, we define our critical value

$$\hat{c}_{n,1-\alpha} \equiv \inf\{c \in \mathbf{R} : P(\hat{\psi}_{\kappa_n}(\hat{\mathbb{G}}_n) \leq c|\{X_i\}_{i=1}^n) \geq 1-\alpha\} \ . \tag{16}$$

As known in the literature (Chernozhukov et al., 2015), the validity of $\hat{c}_{n,1-\alpha}$, viewed as a mapping from distributions to the real line, additionally demands a suitable continuity condition, in accord with the continuous mapping theorem. To this end, we let $c_{n,P}(1-\alpha)$ be the $(1-\alpha)$-quantile of $\psi_{\kappa_n,P}(\mathbb{Z}_{n,P})$ and impose

**Assumption 3.3.** *(i) $\mathbb{Z}_{n,P}$ is tight and centered Gaussian in $\mathbf{H}$ for each $n \in \mathbf{N}$ and $P \in \mathbf{P}$; (ii) $\sup_{P\in\mathbf{P}} E[\|\mathbb{Z}_{n,P}\|_{\mathbf{H}}] < \infty$ uniformly in $n$; (iii) $c_{n,P}(1-\alpha-\varpi) \geq c_{n,P}(0.5)+\varsigma_n$ for some constant $\varpi, \varsigma_n > 0$, each $n$ and $P \in \mathbf{P}_0$; (iv) $c_n/\varsigma_n^2 = O(1)$ as $n \to \infty$.*

Assumption 3.3(i) formalizes Gaussianity and tightness of each $\mathbb{Z}_{n,P}$, which is fulfilled in Examples 2.1-2.4 as well as most regular settings. Assumption 3.3(ii) is a mild moment condition that, in our examples, is tantamount to uniform boundedness of eigenvalues of some matrices with growing dimensions. Assumptions 3.3(iii)(iv) are tied to the natures of densities of Gaussian functionals. Together with Assumptions 3.1 and 3.3(i)(ii), they in effect amount to the aforementioned continuity condition.

We now state the first main result concerning the size of our test.

**Theorem 3.1.** *Let Assumptions 3.1, 3.2 and 3.3 hold. If $0 \leq \kappa_n/r_n = o(c_n)$, then*

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(r_n\phi(\hat{\theta}_n) > \hat{c}_{n,1-\alpha}) \leq \alpha \ , \tag{17}$$

*and, for* $\bar{\mathbf{P}}_0 \equiv \{P \in \mathbf{P}_0 : \langle \vartheta, \theta_P\rangle_{\mathbf{H}} = 0 \ \forall \vartheta \in \mathbf{H} \text{ s.t. } \sup_{\lambda\in\Lambda}\langle\vartheta,\lambda\rangle_{\mathbf{H}} \leq 0\}$,

$$\limsup_{n\to\infty} \sup_{P\in\bar{\mathbf{P}}_0} |P(r_n\phi(\hat{\theta}_n) > \hat{c}_{n,1-\alpha}) - \alpha| = 0 \ . \tag{18}$$

In addition to size control, Theorem 3.1 shows that the limiting rejection rate equals the nominal level, uniformly over $\bar{\mathbf{P}}_0$. Heuristically, $\bar{\mathbf{P}}_0$ may be viewed as consisting of the least favorable distributions in $\mathbf{P}_0$. Indeed, one can show that (i) $\bar{\mathbf{P}}_0 = \{P \in \mathbf{P}_0 : \theta_P = 0\}$ if $\Lambda = \mathbf{R}_+$, and (ii) $\bar{\mathbf{P}}_0$ contains constant (resp. linear) functions if $\Lambda$ is the family of monotone (resp. concave) functions in $L^2(\mathcal{Z})$. We stress that, the choice $\kappa_n = 0$ leads to a least favorable test that amounts to assuming $\theta_P = 0$ (or $\theta_P \in \bar{\mathbf{P}}_0$ by Lemma D.2). As well documented, least favorable tests, while controlling size uniformly, can be substantially conservative, which has motivated the active development of more powerful tests in nonstandard settings (Andrews and Soares, 2010; Lee et al., 2017). Intuitively, in view of Lemma 3.1(ii), it is desirable to have $\kappa_n \to \infty$ (to match $r_n \to \infty$), a condition recurrent in nonstandard problems for the sake of nonconservativeness—see Fang and Santos (2019) and Appendix C.

We emphasize that our test is in general asymptotically nonsimilar, i.e., there may exist a sequence $\{P_n\}$ of distributions from the null such that

$$\liminf_{n\to\infty} P_n(r_n\phi(\hat{\theta}_n) > \hat{c}_{n,1-\alpha}) < \alpha \tag{19}$$

However, this is, in our view, neither an evidence against nonconservativeness nor a deficiency of our test, analogous to the one-sided $t$-test of $\mathrm{H}_0 : \theta_0 \leq 0$ whose rejection rates tend to zero at $\theta_0 < 0$. At a deeper level, (19) is in line with the fact that similarity is not a desirable criterion in nonstandard settings, and can lead to tests with very poor power (Lehmann, 1952; Andrews, 2012). Indeed, many powerful tests in these settings are asymptotically nonsimilar—see, e.g., Andrews and Soares (2010), Lee et al. (2017) and Chetverikov (2019).

Turning to the power of our test, we define $\mathbf{P}_{1,n}^{\Delta} \equiv \{P \in \mathbf{P}_1 : \phi(\theta_P) \geq \Delta/r_n\}$ for $\Delta > 0$. The next theorem shows that our test has nontrivial power against $\mathbf{P}_{1,n}^{\Delta}$.

**Theorem 3.2.** *If Assumptions 3.1, 3.2 and 3.3(ii) hold and $\kappa_n \geq 0$, then*

$$\liminf_{\Delta\to\infty} \liminf_{n\to\infty} \inf_{P\in\mathbf{P}_{1,n}^{\Delta}} P(r_n\phi(\hat{\theta}_n) > \hat{c}_{n,1-\alpha}) = 1 \ . \tag{20}$$

If we employ the kernel estimator in Example (2.1), then Theorem 3.2 predicts that our test is powerful against Pitman drifts of order $(nh_n^{d_z})^{-1/2}$. In comparison, the test of Lee et al. (2017) is powerful against drifts of order $(nh_n^{2\nu})^{-1/2}$ or $(nh_n^{d_z/2+2\nu})^{-1/2}$ depending on the drifts, where $\nu = 1$ for monotonicity and $\nu = 2$ for concavity. Thus, for monotonicity, our convergence rate is faster if $d_z = 1$ but may be slower if $d_z > 4$; for concavity, our rate is faster if $d_z < 4$ but may be slower if $d_z > 8$. Theorem 3.2 also suggests that $\kappa_n$ affects power through higher order terms, and, we reiterate that asymptotic nonconservativeness requires $\kappa_n \to \infty$ subject to $\kappa_n/r_n = o(c_n)$.

### 3.2 Selection of the Tuning Parameter

To motivate, we note that, Lemma 3.1(iii) implies: for any $\epsilon > 0$,

$$P(\frac{\kappa_n}{r_n}\|r_n\{\hat{\theta}_n - \theta_P\}\|_{\mathbf{H}} \leq c_n\epsilon) = P(\|r_n\{\hat{\theta}_n - \theta_P\}\|_{\mathbf{H}} \leq \frac{r_nc_n}{\kappa_n}\epsilon) \to 1 \ . \tag{21}$$

This suggests that we could choose $\kappa_n$ to be such that $r_nc_n/\kappa_n$ is the $(1-\gamma_n)$-quantile of $\|r_n\{\hat{\theta}_n - \theta_P\}\|_{\mathbf{H}}$ with $\gamma_n \downarrow 0$, or the $(1-\gamma_n)$ conditional quantile of $\|\hat{\mathbb{G}}_n\|_{\mathbf{H}}$ (given $\{X_i\}_{i=1}^n$) since $\|r_n\{\hat{\theta}_n - \theta_P\}\|_{\mathbf{H}}$ is unknown. Formally, we let $\hat{\kappa}_n \equiv r_nc_n/\hat{\tau}_{n,1-\gamma_n}$ where

$$\hat{\tau}_{n,1-\gamma_n} \equiv \inf\{c \in \mathbf{R} : P(\|\hat{\mathbb{G}}_n\|_{\mathbf{H}} \leq c|\{X_i\}_{i=1}^n) \geq 1-\gamma_n\} \ . \tag{22}$$

To justify the construction $\hat{\kappa}_n$, we need to introduce our final assumption.

**Assumption 3.4.** $\liminf_{n\to\infty}\inf_{P\in\mathbf{P}_0}\bar{\sigma}^2_{n,P}>0$ *with* $\bar{\sigma}^2_{n,P}\equiv\sup_{h\in\mathbf{H}:\|h\|_{\mathbf{H}}\leq 1}E[\langle h,\mathbb{Z}_{n,P}\rangle^2_{\mathbf{H}}]$.

Heuristically, Assumption 3.4 requires that the coupling variable $\mathbb{Z}_{n,P}$ for $\hat{\theta}_n$ be asymptotically non-degenerate. In turn, we can now verify the validity of $\hat{\kappa}_n$.

**Proposition 3.1.** *Let Assumptions 3.2, 3.3(i)(ii) and 3.4 hold, and set* $\hat{\kappa}_n\equiv r_n c_n/\hat{\tau}_{n,1-\gamma_n}$ *with* $\gamma_n\in(0,1)$ *and* $\hat{\tau}_{n,1-\gamma_n}$ *as in* (22)*. If* $\gamma_n\to 0$*, then* $\hat{\kappa}_n/r_n=o_p(c_n)$ *uniformly in* $P\in\mathbf{P}_0$*. If* $(r_nc_n)^{-2}\log\gamma_n\to 0$*, then* $\hat{\kappa}_n\xrightarrow{p}\infty$ *uniformly in* $P\in\mathbf{P}_0$.

Analogous rate conditions on $\gamma_n$ in parametric settings appear in Chen and Fang (2019). In nonparametric settings, the practice of obtaining data-driven tuning parameters through quantile estimation has appeared in Chernozhukov et al. (2013), Chernozhukov et al. (2015) and Fang and Santos (2019), though a formal theory appears to be lacking. While the choice of $\gamma_n$ remains technically challenging, the situation somewhat improves because $\gamma_n$ is unit/scale-free, and prior studies such as Fang and Santos (2019) and Chen and Fang (2019) have shown that finite sample results are often insensitive to the choice of $\gamma_n$, as also confirmed in our simulations. We recommend $\gamma_n=0.01/\log n$ or $1/n$ for practical implementations.

Finally, the rate $c_n$ (in $\hat{\kappa}_n$) may be ignored in regular settings, and set to be $1/\log n$ in irregular settings without endogeneity. In Example 2.2, we may let $c_n$ be $(\log k_n)^{-\varsigma}$ for some $\varsigma\in[1/2,1]$ with the sieve 2SLS estimation and $k_n$ basis functions for $\theta_0$ (Chen and Christensen, 2018). We note that $\kappa_n$ affects the critical value $\hat{c}_{n,1-\alpha}$ and hence the rejection rates monotonically: a smaller (resp. larger) $\kappa_n$ leads to less (resp. more) rejections—see Lemma D.1. As a result, if one is uncertain about $k_n$ or $\varsigma$, he/she could simply take $c_n=1/\log n$, thereby only making $\hat{c}_{n,1-\alpha}$ larger.

### 3.3 Implementation and Practical Issues

We next provide a guide for implementing our test. Computation of projections (needed in Steps 1 and 2 below) will be discussed in the end.

Step 1: Compute the test statistic $r_n\phi(\hat{\theta}_n)=r_n\|\hat{\theta}_n-\Pi_\Lambda(\hat{\theta}_n)\|_{\mathbf{H}}$.

This step requires an unconstrained estimator $\hat{\theta}_n$ of $\theta_P$, which may be obtained by standard estimation procedures—see Examples 2.1-2.4 for specific estimators and their rates $r_n$. We stress that our framework imposes no additional structures on $\hat{\theta}_n$ as far as implementation is concerned. With $\hat{\theta}_n$ and its projection $\Pi_\Lambda(\hat{\theta}_n)$ in hand, one may approximate $r_n\phi(\hat{\theta}_n)$ by the trapezoid rule in Examples 2.1-2.4 where the $\|\cdot\|_{\mathbf{H}}$-norm takes the form of an integral. Concretely, in Example 2.1 with $\mathcal{Z}=[a,b]$ and $a<b$, we approximate $\phi(\hat{\theta}_n)$ by: for a large $N$ and $\Delta=(b-a)/N$,

$$\{\frac{1}{2}\frac{b-a}{N}[f^2(z_0)+2f^2(z_1)+\cdots+2f^2(z_{N-1})+f^2(z_N)]\}^{1/2}\ , \tag{23}$$

where $f = \hat{\theta}_n - \Pi_\Lambda(\hat{\theta}_n)$ and $z_j = a + j\Delta$.

<u>Step 2:</u> Construct the critical value $\hat{c}_{n,1-\alpha}$ defined in (16).

The construction requires a bootstrap analog $\hat{\mathbb{G}}_n$ of $r_n\{\hat{\theta}_n - \theta_P\}$ and a tuning parameter $\kappa_n$. In regular settings, one often has $\hat{\mathbb{G}}_n = \sqrt{n}\{\hat{\theta}_n^* - \hat{\theta}_n\}$, where $\hat{\theta}_n^*$ is the same as $\hat{\theta}_n$ but based on bootstrap samples. In Examples 2.1-2.4, one may employ various bootstrap schemes in Chernozhukov et al. (2013), Chernozhukov et al. (2015), Chen and Christensen (2018), and Belloni et al. (2019). For $\kappa_n$, we recommend $\hat{\kappa}_n$ proposed in Section 3.2 with a suitable $\gamma_n$ (e.g., $\gamma_n = 0.01/\log n$ or $1/n$). Summarizing, the critical value $\hat{c}_{n,1-\alpha}$ may be then obtained as follows.

(i) Generate $B$ realizations $\{\hat{\mathbb{G}}_{n,b}\}_{b=1}^B$ of $\hat{\mathbb{G}}_n$ (e.g., $B = 200$ or larger); e.g., in (11), generate $\{\{W_{i,b}\}_{i=1}^n\}_{b=1}^B$ that are i.i.d. across both $i$ and $b$, and then obtain each $\hat{\mathbb{G}}_{n,b}$ by evaluating $\hat{\mathbb{G}}_n$ at $\{W_{i,b}\}_{i=1}^n$ which only involves linear calculations.

(ii) Set $\hat{\kappa}_n = r_n c_n/\hat{\tau}_{n,1-\gamma_n}$ where $\hat{\tau}_{n,1-\gamma_n}$ is the $(1-\gamma_n)$-quantile of the $B$ numbers $\|\hat{\mathbb{G}}_{n,1}\|_{\mathbf{H}}, \ldots, \|\hat{\mathbb{G}}_{n,B}\|_{\mathbf{H}}$, and $c_n = 1/\log n$ for Example 2.1. The $\|\cdot\|_{\mathbf{H}}$-norms (here and below) may be computed by the trapezoid rule as in Step 1.

(iii) Approximate $\hat{c}_{n,1-\alpha}$ by the $(1-\alpha)$-quantile of the $B$ numbers

$$\|\hat{\mathbb{G}}_{n,1} + \hat{\kappa}_n\Pi_\Lambda\hat{\theta}_n - \Pi_\Lambda(\hat{\mathbb{G}}_{n,1} + \hat{\kappa}_n\Pi_\Lambda\hat{\theta}_n)\|_{\mathbf{H}}, \\ \ldots, \|\hat{\mathbb{G}}_{n,B} + \hat{\kappa}_n\Pi_\Lambda\hat{\theta}_n - \Pi_\Lambda(\hat{\mathbb{G}}_{n,B} + \hat{\kappa}_n\Pi_\Lambda\hat{\theta}_n)\|_{\mathbf{H}} \; . \tag{24}$$

<u>Step 3:</u> Reject $\mathrm{H}_0$ if and only if $r_n\phi(\hat{\theta}_n) > \hat{c}_{n,1-\alpha}$.

Next, we illustrate the computation of projections. As described in Appendix B, when closed form expressions do not exist, the projection $\Pi_\Lambda(\theta)$ can be computed by solving a linearly constrained quadratic program: for some $A \in \mathbf{M}^{m\times k}$,

$$\min_{h\in\mathbf{R}^k} \|h - \vartheta\| \qquad \text{s.t.} \quad Ah \geq 0 \; , \tag{25}$$

where $\vartheta \in \mathbf{R}^k$ is the vector of values $\theta$ takes at the grid points, and $Ah \geq 0$ is the discretized version of the restriction in question. As an extensively studied problem, (25) admits polynomial-time algorithms; e.g., the iteration complexity of the interior point method is $O(\sqrt{m+k}\log(1/\epsilon))$ for an $\epsilon$-accurate solution.

Finally, since (25) is inherently more complicated to compute in higher dimensions, we note a number of strategies for ameliorating the situation. First, the recently developed open-source solver OSQP (`https://osqp.org`) is very robust in solving large scale quadratic programming problems. Second, while projection under convexity/concavity is computationally demanding in multivariate settings, the representation result in Kuosmanen (2008), when coupled with the OSQP solver, can help significantly reduce the computation cost. Lastly, with regressors more than two or three, one may consider employing semiparametric models, instead of fully nonparametric ones, to alleviate the curse of dimensionality.

# 4 Simulation Studies

This section examines the finite sample performance of our test. Due to space limitation, we focus on monotonicity, and defer concavity, monotonicity jointly with concavity, and Slutsky restriction to the online appendix. Throughout, the significance level is 5%, and, unless otherwise specified, the number of Monte Carlo replications is 3000 while the number of bootstrap repetitions for each replication is 200. All integrals are approximated by the trapezoid rule, and quadratic programs are solved by the OSQP solver. Our test is implemented based on the data-driven choice $\hat{\kappa}_n$ with $\gamma_n \in \{n^{-1/2}, n^{-3/4}, 1/n, 0.1/\log n, 0.05/\log n, 0.01/\log n, 0.1, 0.05, 0.01\}$.

## 4.1 Simulation Designs

We aim to test whether $\theta_0$ is nondecreasing in Example 2.1 with $d_z \in \{1, 2\}$. For $d_z = 1$, the regression function $\theta_0 : [-1, 1] \to \mathbf{R}$ is specified as: for some $(\mathsf{a}, \mathsf{b}, \mathsf{c}) \in \mathbf{R}^3$,

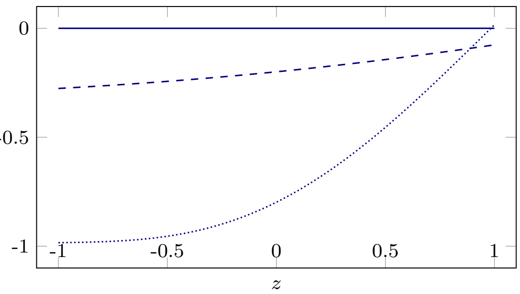
$$\theta_0(z) = \mathsf{a}z - \mathsf{b}\varphi(\mathsf{c}z) \ , \tag{26}$$

where $\varphi$ is the standard normal pdf. We consider three choices for $(\mathsf{a}, \mathsf{b}, \mathsf{c})$ under the null hypothesis, namely $(0, 0, 0)$, $(0.1, 0.5, 0.5)$ and $(0.5, 2, 1)$ that are labeled D1, D2 and D3 respectively, and the collection $\{(\mathsf{a}, \mathsf{b}, \mathsf{c}) : \mathsf{a} = 0, \mathsf{b} = 0.2\delta, \mathsf{c} = 5 + 0.1\delta, \delta = 1, \ldots, 10\}$ under the alternative—see Figure 2. We then draw i.i.d. samples $\{Z_i^*, u_i\}_{i=1}^n$ with $n \in \{500, 750, 1000\}$ from the standard normal distribution in $\mathbf{R}^2$, and set $Z_i = -1 + 2\Phi(Z_i^*) \in [-1, 1]$ with $\Phi$ the standard normal cdf. For $d_z = 2$, the regression function $\theta_0 : [0, 1]^2 \to \mathbf{R}$ is of the form: for some $(\mathsf{a}, \mathsf{b}, \mathsf{c}) \in \mathbf{R}^3$,

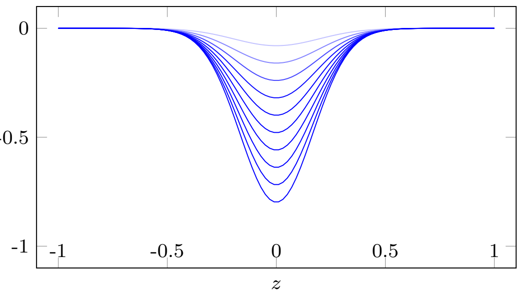
$$\theta_0(z_1, z_2) = \mathsf{a}\Big(\frac{1}{2}z_1^{\mathsf{b}} + \frac{1}{2}z_2^{\mathsf{b}}\Big)^{1/\mathsf{b}} + \mathsf{c}\log(1 + z_1 + z_2) \ . \tag{27}$$

We consider three choices of $(\mathsf{a}, \mathsf{b}, \mathsf{c})$ for the null, namely $(0, 0, 0)$, $(0.2, 1, 0)$ and $(0.5, 0, 0.5)$ that are labeled D1, D2 and D3 respectively, and, for the alternative, set $\mathsf{b} = 0$ and $\mathsf{a} = \mathsf{c} = -\Delta\delta$ with $\Delta = 0.05$ and $\delta = 1, \ldots, 10$. Note that the first term on the right hand side of (27) collapses to $\mathsf{a}\sqrt{z_1 z_2}$ whenever $\mathsf{b} = 0$. In turn, we draw i.i.d. samples $\{Z_{1i}^*, Z_{2i}^*, u_i\}_{i=1}^n$ with $n \in \{500, 750, 1000\}$ from the standard normal distribution in $\mathbf{R}^3$, and set $Z_i \equiv (Z_{1i}, Z_{2i})$ with $Z_{ji} = \Phi(Z_{ji}^*) \in [0, 1]$ for all $i$ and $j = 1, 2$.

To implement our test for (26), we obtain $\hat{\theta}_n$ by sieve estimation with cubic B-splines with 3, 5 or 7 interior knots at the equispaced quantiles of $\{Z_i\}_{i=1}^n$, so that the sieve




(a) $\mathrm{H}_0$: D1 (solid), D2 (dashed) and D3 (dotted)




(b) $\mathrm{H}_1$: $\delta = 1, 2, \ldots, 10$ (from top to bottom)

Figure 2. The function $\theta_0$ in (26) where in Figure 2b, $\mathsf{a} = 0$, $\mathsf{b} = 0.2\delta$ and $\mathsf{c} = 5 + 0.1\delta$. Note that the standard deviation of the error is designed to be no smaller than the range of $\theta_0$.

dimension $k_n$ equals 7, 9 or 11 respectively. In multivariate settings, we construct the series functions via tensor product of univariate B-splines. Since the sieve dimension grows quickly as $d_z$ increases (e.g., $k_n = 49$ with cubic B-splines, 3 knots and $d_z = 2$), we employ univariate quadratic as well as cubic B-splines with one or zero knots along each dimension for (27). Thus, for example, $k_n = 9$ with quadratic B-splines and zero knots. In both designs, we compute $\hat{\mathbb{G}}_{n,b}$ as in (11) by drawing i.i.d. weights from the standard normal distribution, and let the coupling rate $c_n = 1/\log n$. To alleviate the boundary effects, we evaluate the $L^2$-norms (here and below) over $[-0.9, 0.9]$ for (26) and $[0.1, 0.9]^2$ for (27), with step size 0.05. For ease of reference, we label our test with quadratic B-splines and $j$ knots as FS-Q$j$; similarly, FS-C$j$ is the implementation with cubic B-splines and $j$ knots.

To compare, we implement two alternative nonconservative tests, Lee et al. (2017) and Chetverikov (2019). The latter also compares with some prior tests in simulations, which show marked power superiority of the author's three tests so that we take them as important benchmarks. For brevity, however, we only present the one-step test, labeled C-OS, and note that the results for the other two tests are very similar to those produced by C-OS, as also observed in Chetverikov (2019). The details for implementing C-OS in the univariate case are clearly laid out in Chetverikov (2019, p.749). For the bivariate case, we compute the test statistic as on p.27 in the arXiv version of Chetverikov (2019), adopt equispaced empirical quantiles $0.1, 0.15, ..., 0.9$ of each covariate as locations for the weighting function to make the computation feasible, and otherwise follow the implementation in the univariate case.

Lee et al. (2017) consider $L^p$-type statistics. For the sake of comparison, we focus on $p = 2$, since different choices of $p$ implicitly aim power at different alternatives. We estimate the first derivatives of $\theta_0$ by local quadratic regression (Fan and Gijbels, 1996, p.59), with a kernel $z \mapsto 1.5 \max\{1 - (2z)^2, 0\}$ for (26) and $z \mapsto 0.75 \max\{1 - z^2, 0\}$ for (27). We choose two bandwidths: a "large" one $2s_n n^{-1/(d_z+2q+2)}$ (in the spirit of Lee et al. (2017)) and a "small" one $2s_n n^{-1/(d_z+2(q-1)+2)}$ with $q = 2$ and $s_n$ the standard derivation of $\{Z_i\}_{i=1}^n$, resulting in two tests labeled LSW-L and LSW-S respectively.

As studentization can be crucial for power (based on unreported simulations), we estimate the standard errors following Fan and Gijbels (1996, p.115), with the variance of the error estimated by a local polynomial regression of order $q+2$ with bandwidth $2s_n n^{-1/(d_z+2(q+2)+2)}$. Next, we construct critical values based on the empirical bootstrap (as in Lee et al. (2017)) and the tuning parameter $\hat{c}_n$ in their Section 5.1 with $C_{\mathrm{cs}} = 0.4$ (since their results are quite insensitive to other choices of $C_{\mathrm{cs}}$ there). Finally, to ease computation for the bivariate designs, the number of simulation replications is decreased to be 1000 (for LSW only).

Table I. Empirical Size of Monotonicity Tests for $\theta_0$ in (26) at $\alpha = 5\%$

| $n$ | $\gamma_n$ | FS-C3: $k_n = 7$ | | | FS-C5: $k_n = 9$ | | | FS-C7: $k_n = 11$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | D1 | D2 | D3 | D1 | D2 | D3 | D1 | D2 | D3 |
| 500 | $1/n$ | 0.053 | 0.016 | 0.003 | 0.058 | 0.021 | 0.003 | 0.058 | 0.019 | 0.003 |
| | $0.01/\log n$ | 0.053 | 0.016 | 0.003 | 0.058 | 0.021 | 0.003 | 0.058 | 0.019 | 0.003 |
| | 0.01 | 0.053 | 0.016 | 0.003 | 0.058 | 0.021 | 0.003 | 0.058 | 0.019 | 0.003 |
| 750 | $1/n$ | 0.052 | 0.010 | 0.001 | 0.056 | 0.014 | 0.002 | 0.059 | 0.017 | 0.003 |
| | $0.01/\log n$ | 0.052 | 0.010 | 0.001 | 0.056 | 0.014 | 0.002 | 0.059 | 0.017 | 0.003 |
| | 0.01 | 0.052 | 0.010 | 0.001 | 0.056 | 0.014 | 0.002 | 0.059 | 0.017 | 0.003 |
| 1000 | $1/n$ | 0.056 | 0.011 | 0.001 | 0.056 | 0.011 | 0.000 | 0.056 | 0.013 | 0.001 |
| | $0.01/\log n$ | 0.056 | 0.011 | 0.001 | 0.056 | 0.011 | 0.000 | 0.056 | 0.013 | 0.001 |
| | 0.01 | 0.056 | 0.011 | 0.001 | 0.056 | 0.012 | 0.000 | 0.056 | 0.013 | 0.001 |
| | $n$ | LSW-S | | | LSW-L | | | C-OS | | |
| | | D1 | D2 | D3 | D1 | D2 | D3 | D1 | D2 | D3 |
| | 500 | 0.060 | 0.041 | 0.008 | 0.066 | 0.035 | 0.004 | 0.060 | 0.041 | 0.012 |
| | 750 | 0.057 | 0.036 | 0.005 | 0.059 | 0.030 | 0.006 | 0.054 | 0.034 | 0.008 |
| | 1000 | 0.061 | 0.035 | 0.005 | 0.065 | 0.035 | 0.003 | 0.049 | 0.036 | 0.009 |

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

## 4.2 Simulation Results

Tables I-II summarize the empirical sizes. Due to space limitation, we only present our tests with $\gamma_n \in \{1/n, 0.01/\log n, 0.01\}$, and relegate to Tables H.1-H.4 in Appendix H the complete set of results. Together, these tables show that our tests are insensitive to the choice of $\gamma_n$. For the univariate design, all tests control size well, though, relatively speaking, the two LSW tests especially LSW-L tend to over-reject under D1, while our tests tend to under-reject under D2 and D3. Note, however, that D1 is a least favorable case, D2 is in the "interior" (not in the topological sense), and D3 is further into the "interior." Thus, the empirical sizes under D2 and D3 are expected to be smaller than 5%. For the bivariate design, our tests are over-sized under D1 in small samples, though this feature is also shared by LSW-L and C-OS (to an overall lesser extent). The over-rejection of our tests, in particular FS-C1 (in which case $k_n = 25$), is likely because the number of series functions are so "large" that the Gaussian approximation is somewhat inaccurate—note that the overall situation improves as $n$ increases. Thus, while undersmoothing requires a "large" $k_n$, there lies the tension that it should not be "too large" for the sake of distributional approximations.

Table II. Empirical Size of Monotonicity Tests for $\theta_0$ in (27) at $\alpha = 5\%$

| $n$ | $\gamma_n$ | FS-Q0: $k_n = 9$ | | | FS-Q1: $k_n = 16$ | | | FS-C0: $k_n = 16$ | | | FS-C1: $k_n = 25$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | D1 | D2 | D3 | D1 | D2 | D3 | D1 | D2 | D3 | D1 | D2 | D3 |
| 500 | $1/n$ | 0.061 | 0.018 | 0.000 | 0.068 | 0.030 | 0.002 | 0.067 | 0.030 | 0.001 | 0.083 | 0.044 | 0.004 |
| | $0.01/\log n$ | 0.061 | 0.018 | 0.000 | 0.068 | 0.030 | 0.002 | 0.067 | 0.030 | 0.001 | 0.083 | 0.044 | 0.004 |
| | 0.01 | 0.063 | 0.019 | 0.000 | 0.069 | 0.030 | 0.002 | 0.067 | 0.031 | 0.001 | 0.084 | 0.045 | 0.004 |
| 750 | $1/n$ | 0.058 | 0.010 | 0.000 | 0.065 | 0.025 | 0.000 | 0.064 | 0.022 | 0.000 | 0.074 | 0.032 | 0.000 |
| | $0.01/\log n$ | 0.058 | 0.010 | 0.000 | 0.065 | 0.025 | 0.000 | 0.064 | 0.022 | 0.000 | 0.074 | 0.032 | 0.000 |
| | 0.01 | 0.060 | 0.011 | 0.000 | 0.066 | 0.025 | 0.000 | 0.065 | 0.022 | 0.000 | 0.074 | 0.033 | 0.001 |
| 1000 | $1/n$ | 0.050 | 0.011 | 0.000 | 0.055 | 0.021 | 0.000 | 0.053 | 0.020 | 0.000 | 0.059 | 0.023 | 0.000 |
| | $0.01/\log n$ | 0.050 | 0.011 | 0.000 | 0.055 | 0.021 | 0.000 | 0.053 | 0.020 | 0.000 | 0.059 | 0.023 | 0.000 |
| | 0.01 | 0.050 | 0.011 | 0.000 | 0.056 | 0.021 | 0.000 | 0.053 | 0.020 | 0.000 | 0.059 | 0.023 | 0.000 |

| $n$ | LSW-S | | | LSW-L | | | C-OS | | |
|---|---|---|---|---|---|---|---|---|---|
| | D1 | D2 | D3 | D1 | D2 | D3 | D1 | D2 | D3 |
| 500 | 0.043 | 0.020 | 0.000 | 0.061 | 0.026 | 0.002 | 0.068 | 0.061 | 0.037 |
| 750 | 0.054 | 0.017 | 0.000 | 0.066 | 0.031 | 0.000 | 0.057 | 0.045 | 0.023 |
| 1000 | 0.044 | 0.021 | 0.000 | 0.054 | 0.019 | 0.000 | 0.055 | 0.042 | 0.019 |

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

Figure 3 depicts the power curves, where we only show our tests with $\gamma_n = 0.01/\log n$ for brevity and the fact that other choices of $\gamma_n$ lead to very similar curves—see Figure H.7 in Appendix H. For $d_z = 1$, our tests are moderately more powerful than C-OS, across sample sizes and the number of interior knots, and they are all considerably more powerful than LSW-L and in particular LSW-S. For $d_z = 2$, our tests remain competitive in terms of power, though LSW-L is more powerful than FS-C1 (the least powerful among our tests). Interestingly, C-OS is the least powerful of all tests, which may be explained by the fact that only part of the discordance between regressors and the outcome is being picked up through the indicator function in the test function $b(s)$—see p.27 in the arXiv version of Chetverikov (2019). We note that the power of our tests is overall decreasing in $k_n$, which is consistent with Theorem 3.2 since $r_n = \sqrt{n/k_n}$.

Table III. Run-times (in Seconds) of Monotonicity Tests

| $n$ | Design (26) | | | | | Design (27) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | FS-C3 | FS-C7 | LSW-L | LSW-S | C-OS | FS-Q0 | FS-C1 | LSW-L | LSW-S | C-OS |
| 500 | 0.12 | 0.15 | 22.29 | 22.22 | 0.08 | 0.40 | 0.45 | 38.89 | 39.25 | 7.19 |
| 750 | 0.13 | 0.13 | 56.86 | 56.62 | 0.16 | 0.41 | 0.50 | 128.85 | 128.14 | 17.39 |
| 1000 | 0.14 | 0.16 | 103.58 | 102.19 | 0.31 | 0.39 | 0.42 | 219.24 | 222.79 | 29.49 |

Finally, we compare the run-times of a single replication based on the design D1 in both univariate and bivariate cases. For brevity we only report of our tests with the smallest and the largest $k_n$, based on $\gamma_n = 0.01/\log n$. All numbers in Table III are obtained by running MATLAB R2019b in a Windows 10 PC with 16 GB RAM and an Intel® Core™ i7-7700 processor having 4 cores and 3.60 GHz base speed. Overall, Table III shows that our tests are relatively simple to implement, in both univariate and bivariate settings, and their computation cost increases only modestly as we move from

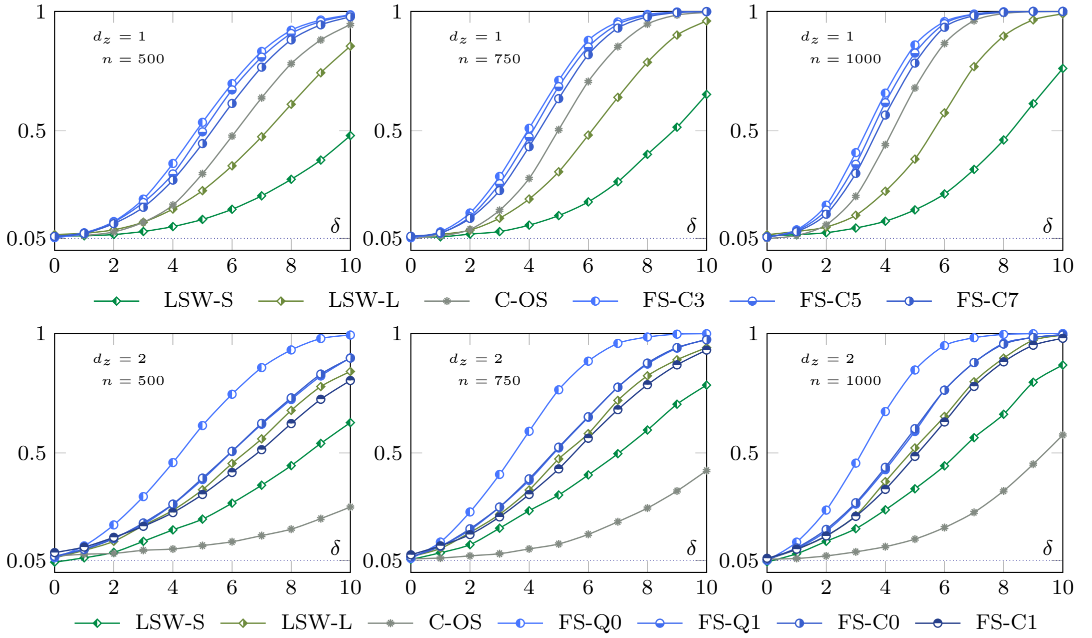



Figure 3. Empirical power of our test (with $\gamma_n = 0.01/\log n$), LSW-L, LSW-S and C-OS for the designs (26) and (27), where corresponding to $\delta = 0$ are the empirical sizes under D1. Note that FS-Q1 and FS-C1 nearly overlap each other in the second row.

$d_z = 1$ to $d_z = 2$. We stress that the computational complexity of quadratic programs involved in our tests depends on the fineness of discretization, not the sample size.

## 5 Conclusion

In this paper, we have developed a uniformly valid and asymptotically nonconservative test for a class of shape restrictions, which is applicable to nonparametric regression models as well as parametric, distributional and structural settings. The key insight we exploit is that these restrictions form convex cones in Hilbert spaces, a structure that enables us to employ a projection-based test whose properties may be analyzed in an elegant, transparent and unifying way. In particular, while the problem is inherently nonstandard, we are able to develop a bootstrap procedure that may be implemented in a way as simple as computing the test statistic.

## Appendix A Proofs of Main Results

Proof of Lemma 3.1: Part (i) is immediate because $h \mapsto \Pi_\Lambda(h)$ is positively homogeneous by Assumption 3.1 and Theorem 5.6-(7) in Deutsch (2012), while part (ii) follows by Lemma D.1, part (i), $\theta_P \in \Lambda$ and $\kappa_n \in [0, r_n]$. For part (iii), note that $\kappa_n\hat{\theta}_n - \kappa_n\theta_P = r_n\{\hat{\theta}_n - \theta_P\} \cdot (\kappa_n/r_n)$. By Assumption 3.2(i), $\sup_{P\in\mathbf{P}} E[\|\mathbb{Z}_{n,P}\|_{\mathbf{H}}] < \infty$

uniformly in $n$ and Markov's inequality, we have: uniformly in $P \in \mathbf{P}$,

$$\|r_n\{\hat{\theta}_n - \theta_P\}\|_{\mathbf{H}} \le \|r_n\{\hat{\theta}_n - \theta_P\} - \mathbb{Z}_{n,P}\|_{\mathbf{H}} + \|\mathbb{Z}_{n,P}\|_{\mathbf{H}} = O_p(1) \ . \tag{A.1}$$

Part (iii) now follows from combining (A.1) and $\kappa_n/r_n = o(c_n)$. ■

Proof of Theorem 3.1: We structure our proof in four steps.

Step 1: Build up a strong approximation for $r_n\phi(\hat{\theta}_n)$ that is valid uniformly in $P \in \mathbf{P}$.

We make use of the map $\psi_{a,P}$ defined in (14), and let $\mathbb{G}_{n,P} \equiv r_n\{\hat{\theta}_n - \theta_P\}$. By Lemma 3.1 and the definition of $\psi_{a,P}$, we may rewrite the test statistic:

$$r_n\phi(\hat{\theta}_n) = \|\mathbb{G}_{n,P} + r_n\theta_P - \Pi_\Lambda(\mathbb{G}_{n,P} + r_n\theta_P)\|_{\mathbf{H}} = \psi_{r_n,P}(\mathbb{G}_{n,P}) \ . \tag{A.2}$$

By Theorem 3.16 in Aliprantis and Border (2006) and Assumption 3.2(i), we have:

$$|\psi_{r_n,P}(\mathbb{G}_{n,P}) - \psi_{r_n,P}(\mathbb{Z}_{n,P})| \le \|\mathbb{G}_{n,P} - \mathbb{Z}_{n,P}\|_{\mathbf{H}} = o_p(c_n) \ , \tag{A.3}$$

uniformly in $P \in \mathbf{P}$. Combining results (A.2) and (A.3), we thus obtain the strong approximation for our test statistic: uniformly in $P \in \mathbf{P}$,

$$r_n\phi(\hat{\theta}_n) = \psi_{r_n,P}(\mathbb{Z}_{n,P}) + o_p(c_n) \ . \tag{A.4}$$

Step 2: Build up a strong approximation of $\hat{\psi}_{\kappa_n}(\hat{\mathbb{G}}_n)$ that is valid uniformly in $P \in \mathbf{P}_0$.

First, Theorem 3.16 in Aliprantis and Border (2006) implies: for each $P \in \mathbf{P}_0$,

$$|\hat{\psi}_{\kappa_n}(\hat{\mathbb{G}}_n) - \psi_{\kappa_n,P}(\hat{\mathbb{G}}_n)| \le \kappa_n\|\Pi_\Lambda\hat{\theta}_n - \theta_P\|_{\mathbf{H}} \le \|\kappa_n\{\hat{\theta}_n - \theta_P\}\|_{\mathbf{H}} = o_p(c_n) \ , \tag{A.5}$$

where the second inequality is due to $\theta_P = \Pi_\Lambda(\theta_P)$ for all $P \in \mathbf{P}_0$, Assumption 3.1(i), Lemma 6.54-d in Aliprantis and Border (2006), and the final step is due to Lemma 3.1(iii). Again by Theorem 3.16 in Aliprantis and Border (2006), we have

$$|\psi_{\kappa_n,P}(\hat{\mathbb{G}}_n) - \psi_{\kappa_n,P}(\bar{\mathbb{Z}}_{n,P})| \le \|\hat{\mathbb{G}}_n - \bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}} = o_p(c_n) \ , \tag{A.6}$$

uniformly in $P \in \mathbf{P}$, where the equality is due to Assumption 3.2(ii). We thus obtain by (A.5) and (A.6) and the triangle inequality that: uniformly in $P \in \mathbf{P}_0$,

$$\hat{\psi}_{\kappa_n}(\hat{\mathbb{G}}_n) = \psi_{\kappa_n,P}(\bar{\mathbb{Z}}_{n,P}) + o_p(c_n) \ . \tag{A.7}$$

Step 3: Control the estimation error of $\hat{c}_{n,1-\alpha}$. By results (A.4) and (A.7), we may select a sequence of positive scalars $\epsilon_n = o(c_n)$ (sufficiently slow) such that, as $n \to \infty$,

$$\sup_{P\in\mathbf{P}_0} P(|r_n\phi(\hat{\theta}_n) - \psi_{r_n,P}(\mathbb{Z}_{n,P})| > \epsilon_n) = o(1) \ , \tag{A.8}$$

$$\sup_{P\in\mathbf{P}_0} P(|\hat{\psi}_{\kappa_n}(\hat{\mathbb{G}}_n) - \psi_{\kappa_n,P}(\bar{\mathbb{Z}}_{n,P})| > \epsilon_n) = o(1) \ . \tag{A.9}$$

By Markov's inequality, Fubini's theorem and (A.9), we have: for each $\eta > 0$,

$$\sup_{P\in\mathbf{P}_0} P(P(|\hat{\psi}_{\kappa_n}(\hat{\mathbb{G}}_n) - \psi_{\kappa_n,P}(\bar{\mathbb{Z}}_{n,P})| > \epsilon_n | \{X_i\}_{i=1}^n) > \eta)$$
$$\leq \sup_{P\in\mathbf{P}_0} \frac{1}{\eta} P(|\hat{\psi}_{\kappa_n}(\hat{\mathbb{G}}_n) - \psi_{\kappa_n,P}(\bar{\mathbb{Z}}_{n,P})| > \epsilon_n) = o(1) \ . \tag{A.10}$$

Thus, we may select a sequence of positive scalars $\eta_n \downarrow 0$ such that

$$P(|\hat{\psi}_{\kappa_n}(\hat{\mathbb{G}}_n) - \psi_{\kappa_n,P}(\bar{\mathbb{Z}}_{n,P})| > \epsilon_n | \{X_i\}_{i=1}^n) = o_p(\eta_n) \ , \tag{A.11}$$

uniformly in $P \in \mathbf{P}_0$. Since $\bar{\mathbb{Z}}_{n,P}$ is independent of $\{X_i\}_{i=1}^n$ by Assumption 3.2(ii), the conditional cdf of $\psi_{\kappa_n,P}(\bar{\mathbb{Z}}_{n,P})$ given $\{X_i\}_{i=1}^n$ is precisely its unconditional analog. Thus, we may conclude by Lemma 11 in Chernozhukov et al. (2013) and result (A.11) that

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}_0} P(\hat{c}_{n,1-\alpha} + \epsilon_n \geq c_{n,P}(1-\alpha-\eta_n)) = 1 \ . \tag{A.12}$$

<u>Step 4</u>: Conclude with the help of a partial anti-concentration inequality.

To begin with, note that by results (A.8) and (A.12), we have:

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(r_n\phi(\hat{\theta}_n) > \hat{c}_{n,1-\alpha})$$
$$\leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(r_n\phi(\hat{\theta}_n) > \hat{c}_{n,1-\alpha}, |r_n\phi(\hat{\theta}_n) - \psi_{r_n,P}(\mathbb{Z}_{n,P})| \leq \epsilon_n)$$
$$\leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\psi_{r_n,P}(\mathbb{Z}_{n,P}) > c_{n,P}(1-\alpha-\eta_n) - 2\epsilon_n)$$
$$\leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\psi_{\kappa_n,P}(\mathbb{Z}_{n,P}) > c_{n,P}(1-\alpha-\eta_n) - 2\epsilon_n) \ , \tag{A.13}$$

where the final step follows by Lemma D.1 and $0 \leq \kappa_n \leq r_n$ for all large $n$ (due to $\kappa_n/r_n = o(c_n)$ and $c_n = O(1)$). In turn, we note that, since $\eta_n \downarrow 0$ and $\epsilon_n = o(c_n)$, it follows by Assumption 3.3(iii), Proposition D.1 and result (A.13) that

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(r_n\phi(\hat{\theta}_n) > \hat{c}_{n,1-\alpha}) \leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} P(\psi_{\kappa_n,P}(\mathbb{Z}_{n,P}) > c_{n,P}(1-\alpha-\eta_n))$$
$$\leq \limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} \{\alpha + \eta_n\} = \alpha \ , \tag{A.14}$$

as desired for the first claim. For the second claim, it thus suffices to show

$$\liminf_{n\to\infty} \inf_{P\in\bar{\mathbf{P}}_0} P(r_n\phi(\hat{\theta}_n) > \hat{c}_{n,1-\alpha}) \geq \alpha \ . \tag{A.15}$$

For this, we note that, by simple manipulations,

$$\begin{aligned}
&\liminf_{n\to\infty} \inf_{P\in\bar{\mathbf{P}}_0} P(r_n\phi(\hat{\theta}_n) > \hat{c}_{n,1-\alpha}) \\
&\qquad \geq \liminf_{n\to\infty} \inf_{P\in\bar{\mathbf{P}}_0} P(r_n\phi(\hat{\theta}_n) > \hat{c}_{n,1-\alpha}, |r_n\phi(\hat{\theta}_n) - \psi_{r_n,P}(\mathbb{Z}_{n,P})| \leq \epsilon_n) \\
&\qquad = \liminf_{n\to\infty} \inf_{P\in\bar{\mathbf{P}}_0} P(\psi_{r_n,P}(\mathbb{Z}_{n,P}) - \epsilon_n > \hat{c}_{n,1-\alpha}) \ , \qquad (A.16)
\end{aligned}$$

where the last step is due to result (A.8) and $\bar{\mathbf{P}}_0 \subset \mathbf{P}_0$. Moreover, another application of Lemma 11 in Chernozhukov et al. (2013) to (A.11) yields

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}_0} P(\hat{c}_{n,1-\alpha} \leq c_{n,P}(1-\alpha+\eta_n) + \epsilon_n) = 1 \ . \tag{A.17}$$

By the definition of $\bar{\mathbf{P}}_0$ and Lemma D.2, we also note that, for all $n$ and $P \in \bar{\mathbf{P}}_0$,

$$\psi_{r_n,P}(\mathbb{Z}_{n,P}) = \psi_{\kappa_n,P}(\mathbb{Z}_{n,P}) = \|\mathbb{Z}_{n,P} - \Pi_\Lambda(\mathbb{Z}_{n,P})\|_{\mathbf{H}} \ . \tag{A.18}$$

Combining results (A.16), (A.17) and (A.18) with Proposition D.1 yields

$$\begin{aligned}
&\liminf_{n\to\infty} \inf_{P\in\bar{\mathbf{P}}_0} P(r_n\phi(\hat{\theta}_n) > \hat{c}_{n,1-\alpha}) \\
&\qquad \geq \liminf_{n\to\infty} \inf_{P\in\bar{\mathbf{P}}_0} P(\psi_{r_n,P}(\mathbb{Z}_{n,P}) - \epsilon_n > c_{n,P}(1-\alpha+\eta_n) + \epsilon_n) \\
&\qquad = \liminf_{n\to\infty} \inf_{P\in\bar{\mathbf{P}}_0} P(\psi_{\kappa_n,P}(\mathbb{Z}_{n,P}) > c_{n,P}(1-\alpha+\eta_n) + 2\epsilon_n) \\
&\qquad = \liminf_{n\to\infty} \inf_{P\in\bar{\mathbf{P}}_0} P(\psi_{\kappa_n,P}(\mathbb{Z}_{n,P}) > c_{n,P}(1-\alpha+\eta_n) - 2\epsilon_n) \ . \qquad (A.19)
\end{aligned}$$

Since $c_{n,P}(1-\alpha+\eta_n) - \epsilon_n < c_{n,P}(1-\alpha+\eta_n)$, we thus obtain by result, (A.19), the definition of quantiles and $\eta_n = o(1)$ that

$$\liminf_{n\to\infty} \inf_{P\in\bar{\mathbf{P}}_0} P(r_n\phi(\hat{\theta}_n) > \hat{c}_{n,1-\alpha}) \geq \liminf_{n\to\infty}\{\alpha - \eta_n\} = \alpha \ , \tag{A.20}$$

which, together with the first claim, establishes the second claim of the theorem. ■

Proof of Theorem 3.2: First, by Assumption 3.1 and Lemma D.1, we have

$$\begin{aligned}
\|\hat{\mathbb{G}}_n + \kappa_n\Pi_\Lambda\hat{\theta}_n - \Pi_\Lambda(\hat{\mathbb{G}}_n + \kappa_n\Pi_\Lambda\hat{\theta}_n)\|_{\mathbf{H}} &\leq \|\hat{\mathbb{G}}_n - \Pi_\Lambda\hat{\mathbb{G}}_n\|_{\mathbf{H}} \\
&\leq \|\hat{\mathbb{G}}_n\|_{\mathbf{H}} \leq \|\hat{\mathbb{G}}_n - \bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}} + \|\bar{\mathbb{Z}}_{n,P}\|_{\mathbf{H}} \ , \qquad (A.21)
\end{aligned}$$

where the second inequality follows from Assumption 3.1 and Theorem 5.6(5) in Deutsch

(2012), and the third inequality is due to the triangle inequality. By Assumptions 3.2 and 3.3(ii), we in turn have from (A.21) that, uniformly in $P \in \mathbf{P}$,

$$\|\hat{\mathbb{G}}_n + \kappa_n \Pi_\Lambda \hat{\theta}_n - \Pi_\Lambda(\hat{\mathbb{G}}_n + \kappa_n \Pi_\Lambda \hat{\theta}_n)\|_{\mathbf{H}} = o_p(c_n) + O_p(1) = O_p(1) \ . \tag{A.22}$$

By the definition of $\hat{c}_{n,1-\alpha}$, we note that, for $M > 0$ and uniformly in $P \in \mathbf{P}$,

$$\begin{aligned} P(\hat{c}_{n,1-\alpha} > M) &\leq P(P(\|\hat{\mathbb{G}}_n + \kappa_n \Pi_\Lambda \hat{\theta}_n - \Pi_\Lambda(\hat{\mathbb{G}}_n + \kappa_n \Pi_\Lambda \hat{\theta}_n)\|_{\mathbf{H}} > M | \{X_i\}_{i=1}^n) > \alpha) \\ &\leq \frac{1}{\alpha} P(\|\hat{\mathbb{G}}_n + \kappa_n \Pi_\Lambda \hat{\theta}_n - \Pi_\Lambda(\hat{\mathbb{G}}_n + \kappa_n \Pi_\Lambda \hat{\theta}_n)\|_{\mathbf{H}} > M) \ , \end{aligned} \tag{A.23}$$

where the second inequality holds by Markov's inequality and Fubini's theorem. It follows from results (A.22) and (A.23) that $\hat{c}_{n,1-\alpha} = O_p(1)$ uniformly in $P \in \mathbf{P}$.

Next, we bound $r_n \phi(\hat{\theta}_n)$ from below. By Theorem 3.16 in Aliprantis and Border (2006) and the triangle inequality, we have: uniformly in $P \in \mathbf{P}$,

$$\begin{aligned} |r_n \phi(\hat{\theta}_n) - r_n \phi(\theta_P)| &\leq \|r_n\{\hat{\theta}_n - \theta_P\}\|_{\mathbf{H}} \\ &\leq \|r_n\{\hat{\theta}_n - \theta_P\} - \mathbb{Z}_{n,P}\|_{\mathbf{H}} + \|\mathbb{Z}_{n,P}\|_{\mathbf{H}} \leq o_p(c_n) + O_p(1) = O_p(1) \ , \end{aligned} \tag{A.24}$$

where the third inequality follows by Assumptions 3.2(i) and 3.3(ii), and the last step is due to $c_n = O(1)$. It follows from result (A.24) and the definition of $\mathbf{P}_{1,n}^{\Delta}$ that

$$r_n \phi(\hat{\theta}_n) = r_n \phi(\theta_P) + r_n \phi(\hat{\theta}_n) - r_n \phi(\theta_P) \geq \Delta + O_p(1) \ , \tag{A.25}$$

uniformly in $P \in \mathbf{P}_{1,n}^{\Delta}$. The theorem thus follows from combining result (A.25) and the order $\hat{c}_{n,1-\alpha} = O_p(1)$ uniformly in $P \in \mathbf{P}$ that we have established . ■

# References


Abrevaya, J. and W. Jiang (2005): "A nonparametric approach to measuring and testing curvature," *Journal of Business & Economic Statistics*, 23, 1–19.

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.

Ai, C. and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843.

Aliprantis, C. D. and K. Border (2006): *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Berlin: Springer Verlag, third ed.

Andrews, D. W. K. (2012): "Similar-on-the-Boundary Tests for Moment Inequalities Exist, But Have Poor Power," Tech. rep., Yale University.

Andrews, D. W. K. and G. Soares (2010): "Inference for Parameters Defined by

Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–157.

Baraud, Y., S. Huet, and B. Laurent (2005): "Testing convex hypotheses on the mean of a Gaussian vector. Application to testing qualitative hypotheses on a regression function," *The Annals of Statistics*, 33, 214–257.

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.

Belloni, A., V. Chernozhukov, D. Chetverikov, and I. Fernández-Val (2019): "Conditional quantile processes based on series or many regressors," *Journal of Econometrics*, 213, 4 – 29.

Belloni, A., V. Chernozhukov, D. Chetverikov, and K. Kato (2015): "Some New Asymptotic Theory for Least Squares Series: Pointwise and Uniform Results," *Journal of Econometrics*, 186, 345 – 366.

Bhattacharya, D. (2020): "The Empirical Content of Binary Choice Models," *arXiv preprint arXiv:1902.11012*.

Chen, Q. and Z. Fang (2019): "Improved Inference on the Rank of a Matrix," *Quantitative Economics*, 10, 1787–1824.

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*.

Chen, X. and T. M. Christensen (2018): "Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression," *Quantitative Economics*, 9, 39–84.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81, 667–737.

Chernozhukov, V., W. K. Newey, and A. Santos (2015): "Constrained Conditional Moment Restriction Models," *arXiv preprint arXiv:1509.06311*.

Chetverikov, D. (2019): "Testing Regression Monotonicity in Econometric Models," *Econometric Theory*, 35, 729–776.

Chetverikov, D., A. Santos, and A. M. Shaikh (2018): "The Econometrics of Shape Restrictions," *Annual Review of Economics*, 10, 31–63.

Darolles, S., Y. Fan, J. P. Florens, and E. Renault (2011): "Nonparametric Instrumental Regression," *Econometrica*, 79, 1541–1565.

Deutsch, F. R. (2012): *Best Approximation in Inner Product Spaces*, Springer.

Donald, S. and W. Newey (1994): "Series Estimation of Semilinear Models," *Journal of Multivariate Analysis*, 50, 30 – 40.

Dümbgen, L. and V. G. Spokoiny (2001): "Multiscale Testing of Qualitative Hypotheses," *The Annals of Statistics*, 29, 124–152.

Durot, C. (2003): "A Kolmogorov-Type Test for Monotonicity of Regression," *Statis-*

*tics & Probability Letters*, 63, 425 – 433.

Fan, J. and I. Gijbels (1996): *Local Polynomial Modelling and Its Applications*, CRC Press.

Fang, Z. and A. Santos (2019): "Inference on directionally differentiable functions," *The Review of Economic Studies*, 86, 377–412.

Freyberger, J. and B. Reeves (2018): "Inference under shape restrictions," *Available at SSRN 3011474*.

Ghosal, S., A. Sen, and A. W. van der Vaart (2000): "Testing monotonicity of regression," *The Annals of Statistics*, 28, 1054–1082.

Gijbels, I., P. Hall, M. Jones, and I. Koch (2000): "Tests for monotonicity of a regression mean with guaranteed level," *Biometrika*, 87, 663–673.

Gutknecht, D. (2016): "Testing for monotonicity under endogeneity: An application to the reservation wage function," *Journal of Econometrics*, 190, 100–114.

Hall, P. and N. E. Heckman (2000): "Testing for monotonicity of a regression mean by calibrating for linear functions," *The Annals of Statistics*, 28, 20–39.

Hall, P. and J. L. Horowitz (2005): "Nonparametric methods for inference in the presence of instrumental variables," *The Annals of Statistics*, 33, 2904–2929.

Han, Q., T. Wang, S. Chatterjee, and R. J. Samworth (2019): "Isotonic regression in general dimensions," *The Annals of Statistics*, 47, 2440–2471.

Horowitz, J. L. and S. Lee (2017): "Nonparametric estimation and inference under shape restrictions," *Journal of Econometrics*, 201, 108–126.

Juditsky, A. and A. Nemirovski (2002): "On nonparametric tests of positivity/monotonicity/convexity," *The Annals of Statistics*, 30, 498–527.

Komarova, T. and J. Hidalgo (2019): "Testing nonparametric shape restrictions," *arXiv preprint arXiv:1909.01675*.

Kostyshak, S. (2017): "Non-Parametric Testing of U-Shaped Relationships," *Available at SSRN 2905833*.

Kuosmanen, T. (2008): "Representation theorem for convex nonparametric least squares," *The Econometrics Journal*, 11, 308–325.

Lee, S., K. Song, and Y.-J. Whang (2017): "Testing for a general class of functional inequalities," *Econometric Theory*, 1–47.

Lehmann, E. L. (1952): "Testing Multiparameter Hypotheses," *Annals of Mathematical Statistics*, 23, 541–552.

Li, J. and Z. Liao (2020): "Uniform nonparametric inference for time series," *Journal of Econometrics*, 219, 38 – 51.

Matzkin, R. L. (1994): "Restrictions of economic theory in nonparametric methods," in *Handbook of Econometrics*, Elsevier, vol. 4, 2523 – 2558.

Newey, W. K. and J. L. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

Pinkse, J. and K. Schurter (2019): "Estimation of Auction Models with Shape

Restrictions," *arXiv preprint arXiv:1912.07466*.

Silvapulle, M. J. and P. K. Sen (2005): *Constrained Statistical Inference*, Wiley.

Stone, C. J. (1982): "Optimal Global Rates of Convergence for Nonparametric Regression," *The Annals of Statistics*, 10, 1040–1053.

Varian, H. R. (1982): "The Nonparametric Approach to Demand Analysis," *Econometrica*, 50, 945–973.

——— (1984): "The Nonparametric Approach to Production Analysis," *Econometrica*, 52, 579–597.

Wei, Y., M. J. Wainwright, and A. Guntuboyina (2019): "The geometry of hypothesis testing over convex cones: Generalized likelihood ratio tests and minimax radii," *The Annals of Statistics*, 47, 994–1024.

Wolak, F. A. (1987): "An exact test for multiple inequality and equality constraints in the linear regression model," *Journal of the American Statistical Association*, 82, 782–793.

Zhu, Y. (2020): "Inference in nonparametric/semiparametric moment equality models with shape restrictions," *Quantitative Economics*, 11, 609–636.