Inference on Directionally Differentiable Functions

Zheng Fang

Department of Economics

UC San Diego

zfang@ucsd.edu

Andres Santos

Department of Economics

UC San Diego

a2santos@ucsd.edu

February, 2014

Abstract

This paper studies an asymptotic framework for conducting inference on parameters of the form $\phi(\theta_0)$, where ϕ is a known directionally differentiable function and θ_0 is estimated by $\hat{\theta}_n$. In these settings, the asymptotic distribution of the plug-in estimator $\phi(\hat{\theta}_n)$ can be readily derived employing existing extensions to the Delta method. We show, however, that the "standard" bootstrap is only consistent under overly stringent conditions – in particular we establish that differentiability of ϕ is a necessary and sufficient condition for bootstrap consistency whenever the limiting distribution of $\hat{\theta}_n$ is Gaussian. An alternative resampling scheme is proposed which remains consistent when the bootstrap fails, and is shown to provide local size control under restrictions on the directional derivative of ϕ. We illustrate the utility of our results by developing a test of whether a Hilbert space valued parameter belongs to a convex set – a setting that includes moment inequality problems and certain tests of shape restrictions as special cases.

KEYWORDS: Delta method, Bootstrap consistency, Directional differentiability.

1

1 Introduction

The Delta method is a cornerstone of asymptotic analysis, allowing researchers to easily derive asymptotic distributions, compute standard errors, and establish bootstrap consistency.[1] However, an important class of estimation and inference problems in economics fall outside its scope. These problems study parameters of the form $\phi(\theta_0)$, where θ_0 is unknown but estimable and ϕ is a known but potentially non-differentiable function. Such a setting arises frequently in economics, with applications including the construction of parameter confidence regions in moment inequality models (Pakes et al., 2006; Ciliberto and Tamer, 2009), the study of convex partially identified sets (Beresteanu and Molinari, 2008; Bontemps et al., 2012), and the development of tests of superior predictive ability (White, 2000; Hansen, 2005), of stochastic dominance (Linton et al., 2010), and of likelihood ratio ordering (Beare and Moon, 2013).

The aforementioned examples share the common feature of ϕ being directionally differentiable despite full differentiability failing to hold. In this paper, we show that ϕ being directionally differentiable provides enough structure for the development of a unifying asymptotic framework for conducting inference in these problems – much in the same manner the Delta method and its bootstrap counterpart yield a common scheme for analyzing applications in which ϕ is differentiable. Specifically, we let θ_0 be a Banach space valued parameter and require the existence of an estimator $\hat{\theta}_n$ whose asymptotic distribution we denote by $\mathbb{G}_0$ – i.e., for some sequence $r_n \uparrow \infty$, we have that

$$r_n\{\hat{\theta}_n - \theta_0\} \xrightarrow{L} \mathbb{G}_0 \ . \tag{1}$$

Within this framework, we then study the problem of conducting inference on $\phi(\theta_0)$ by employing the estimator $\phi(\hat{\theta}_n)$ – a practice common in, for example, moment inequality (Andrews and Soares, 2010), conditional moment inequality (Andrews and Shi, 2013), and incomplete linear models (Beresteanu and Molinari, 2008).

As has been previously noted in the literature, the traditional Delta method readily generalizes to the case where ϕ is directionally differentiable (Shapiro, 1991; Dümbgen, 1993). In particular, if ϕ is Hadamard directionally differentiable, then

$$r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} \xrightarrow{L} \phi'_{\theta_0}(\mathbb{G}_0) \ , \tag{2}$$

where ϕ'_{θ_0} denotes the directional derivative of ϕ at θ_0. The utility of the asymptotic distribution of $\phi(\hat{\theta}_n)$, however, hinges on our ability to consistently estimate it. While it is tempting in these problems to resort to resampling schemes such as the bootstrap of Efron (1979), we know by way of example that they may be inconsistent even if they are

---

[1]Interestingly, despite its importance, the origins of the Delta method remain obscure. Hoef (2012) recently attributed its invention to the economist Robert Dorfman in his article Dorfman (1938), which was curiously published by the Worcester State Hospital (a public asylum for the insane).

valid for the original estimator $\hat{\theta}_n$ (Bickel et al., 1997; Andrews, 2000). We generalize these examples by providing simple to verify necessary and sufficient conditions for the validity of the bootstrap for $\hat{\theta}_n$ to be inherited by $\phi(\hat{\theta}_n)$. In the ubiquitous case where $\mathbb{G}_0$ is Gaussian, our results imply that full differentiability of ϕ at θ_0 is in fact a necessary and sufficient condition for bootstrap consistency. Thus, we conclude that the failure of "standard" bootstrap approaches is an inherent property of irregular models. Indeed, an immediate corollary of our analysis is that, in this setting, the bootstrap is inconsistent whenever the asymptotic distribution of $\phi(\hat{\theta}_n)$ is not Gaussian.

Intuitively, consistently estimating the asymptotic distribution of $\phi(\hat{\theta}_n)$ requires us to adequately approximate both the law of $\mathbb{G}_0$ and the directional derivative ϕ'_{θ_0} (see (2)). While a consistent bootstrap procedure for $\hat{\theta}_n$ enables us to do the former, the bootstrap fails for $\phi(\hat{\theta}_n)$ due to its inability to properly estimate ϕ'_{θ_0}. These heuristics, however, readily suggests a remedy to the problem – namely to compose a suitable estimator $\hat{\phi}'_n$ for ϕ'_{θ_0} with the bootstrap approximation to the asymptotic distribution of $\hat{\theta}_n$. We formalize this intuition, and provide conditions on $\hat{\phi}'_n$ that ensure the proposed approach yields consistent estimators of the asymptotic distribution of $\phi(\hat{\theta}_n)$ and its quantiles. Moreover, we further show that existing inferential procedures developed in the context of specific applications in fact follow precisely this approach – these include Andrews and Soares (2010) for moment inequalities, Linton et al. (2010) for tests of stochastic dominance, and Kaido (2013) for convex partially identified models.

As argued by Imbens and Manski (2004), pointwise asymptotic approximations may be unreliable, in particular when $\phi(\hat{\theta}_n)$ is not regular. Heuristically, if the asymptotic distribution of $\phi(\hat{\theta}_n)$ is sensitive to local perturbations of the data generating process, then employing (2) as the basis for inference may yield poor size in finite samples. We thus examine the ability of our proposed procedure to provide local size control in the context of employing $\phi(\hat{\theta}_n)$ as a test statistic for the hypothesis

$$H_0 : \phi(\theta_0) \le 0 \qquad\qquad H_1 : \phi(\theta_0) > 0 \ . \qquad\qquad (3)$$

Special cases of (3) include inference in moment inequality models and tests of stochastic dominance – instances in which our framework encompasses procedures that provide local, in fact uniform, size control (Andrews and Soares, 2010; Linton et al., 2010; Andrews and Shi, 2013). We show that the common structure linking these applications is that ϕ'_{θ_0} and $\hat{\theta}_n$ are respectively subadditive and regular. Indeed, we more generally establish that these two properties suffice for guaranteeing the ability of our procedure to locally control size along parametric submodels. As part of this local analysis, we further characterize local power and show that, under mild regularity conditions, the bootstrap is valid for $\phi(\hat{\theta}_n)$ if and only if $\phi(\hat{\theta}_n)$ is regular.

We illustrate the utility of our analysis by developing a test of whether a Hilbert

3

space valued parameter θ_0 belongs to a known convex set Λ – a setting that includes tests of moment inequalities, stochastic dominance, and shape restrictions as special cases. Specifically, we set $\phi(\theta)$ to be the distance between θ and the set Λ, and employ $\phi(\hat{\theta}_n)$ as a test statistic of whether θ_0 belongs to Λ. Exploiting the directional differentiability of projections onto convex sets (Zaranotello, 1971), we show the asymptotic distribution of $\phi(\hat{\theta}_n)$ is given by the distance between $\mathbb{G}_0$ and the tangent cone of Λ at θ_0. While our results imply the bootstrap is inconsistent, we are nonetheless able to obtain valid critical values by constructing a suitable estimator $\hat{\phi}'_n$ which we compose with a bootstrap approximation to the law of $\mathbb{G}_0$. In addition, we establish the directional derivative ϕ'_{θ_0} is always subadditive, and thus conclude that the proposed test is able to locally control size provided $\hat{\theta}_n$ is regular. A brief simulation study confirms our theoretical findings by showing the proposed test possesses good finite sample size control.

In related work, an extensive literature has established the consistency of the bootstrap and its ability to provide a refinement when θ_0 is a vector of means and ϕ is a differentiable function (Hall, 1992; Horowitz, 2001). The setting where ϕ is directionally differentiable was originally examined by Dümbgen (1993), who studied the unconditional distribution of the bootstrap and in this way obtained sufficient, but not necessary, conditions for the bootstrap to fail for $\phi(\hat{\theta}_n)$. In more recent work, applications where ϕ is not fully differentiable have garnered increasing attention due to their preponderance in the analysis of partially identified models (Manski, 2003). Hirano and Porter (2012) and Song (2012), for example, explicitly exploit the directional differentiability of ϕ as well, though their focus is on estimation rather than inference. Other work studying these irregular models, though not explicitly relying on the directional differentiability of ϕ, include Chernozhukov et al. (2007, 2013), Romano and Shaikh (2008, 2010), Bugni (2010), and Canay (2010) among many others.

The remainder of the paper is organized as follows. Section 2 formally introduces the model we study and contains a minor extension of the Delta method for directionally differentiable functions. In Section 3 we characterize necessary and sufficient conditions for bootstrap consistency, develop an alternative method for estimating the asymptotic distribution of $\phi(\hat{\theta}_n)$, and study the local properties of this approach. Section 4 applies these results to develop a test of whether a Hilbert space valued parameter belongs to a closed convex set. All proofs are contained in the Appendix.

2 Setup and Background

In this section, we introduce our notation and review the concepts of Hadamard and directional Hadamard differentiability as well as their implications for the Delta method.

2.1 General Setup

In order to accommodate applications such as conditional moment inequalities and tests of shape restrictions, we must allow for both the parameter θ_0 and the map ϕ to take values in possibly infinite dimensional spaces; see Examples 2.3-2.6 below. We therefore impose the general requirement that $\theta_0 \in \mathbb{D}_\phi$ and $\phi : \mathbb{D}_\phi \subseteq \mathbb{D} \to \mathbb{E}$ for $\mathbb{D}$ and $\mathbb{E}$ Banach spaces with norms $\| \cdot \|_{\mathbb{D}}$ and $\| \cdot \|_{\mathbb{E}}$ respectively, and $\mathbb{D}_\phi$ the domain of ϕ.

The estimator $\hat\theta_n$ is assumed to be a function of a sequence of random variables $\{X_i\}_{i=1}^n$ into the domain of ϕ. The distributional convergence

$$r_n\{\hat\theta_n - \theta_0\} \xrightarrow{L} \mathbb{G}_0 \; , \tag{4}$$

is then understood to be in $\mathbb{D}$ and with respect to the joint law of $\{X_i\}_{i=1}^n$. For instance, if $\{X_i\}_{i=1}^n$ is an i.i.d. sample and each $X_i \in \mathbf{R}^d$ is distributed according to P, then probability statements for $\hat\theta_n : \{X_i\}_{i=1}^n \to \mathbb{D}_\phi$ are understood to be with respect to the product measure $\bigotimes_{i=1}^n P$. We emphasize, however, that with the exception of the local analysis where we assume $\{X_i\}_{i=1}^n$ is i.i.d. for simplicity, our results are applicable to dependent settings as well. In addition, we also note the convergence in distribution in (4) is meant in the Hoffman-Jørgensen sense (van der Vaart and Wellner, 1996). Expectations throughout the text should therefore be interpreted as outer expectations, though we obviate the distinction in the notation. The notation is made explicit in the Appendix whenever differentiating between inner and outer expectations is necessary.

Finally, we introduce notation that is recurrent in the context of our examples. For a set $\mathbf{A}$, we denote the space of bounded functions on $\mathbf{A}$ by

$$\ell^\infty(\mathbf{A}) \equiv \{f : \mathbf{A} \to \mathbf{R} \text{ such that } \|f\|_\infty < \infty\} \qquad \|f\|_\infty \equiv \sup_{a \in \mathbf{A}} |f(a)| \; , \tag{5}$$

and note $\ell^\infty(\mathbf{A})$ is a Banach space under $\| \cdot \|_\infty$. If in addition $\mathbf{A}$ is a compact Hausdorff topological space, then we let $\mathcal{C}(\mathbf{A})$ denote the set of continuous functions on $\mathbf{A}$,

$$\mathcal{C}(\mathbf{A}) \equiv \{f : \mathbf{A} \to \mathbf{R} \text{ such that } f \text{ is continuous } \} \; , \tag{6}$$

which satisfies $\mathcal{C}(\mathbf{A}) \subset \ell^\infty(\mathbf{A})$ and is also a Banach space when endowed with $\| \cdot \|_\infty$.

2.1.1 Examples

In order to fix ideas, we next introduce a series of examples that illustrate the broad applicability of our setting. We return to these examples throughout the paper, and develop a formal treatment of each of them in the Appendix. For ease of exposition, we base our discussion on simplifications of well known models, though we note that our

results apply to the more general problems that motivated them.

Our first example is due to Bickel et al. (1997), and provides an early illustration of the potential failure of the nonparametric bootstrap.

Example 2.1 (Absolute Value of Mean)**.** Let $X \in \mathbf{R}$ be a scalar valued random variable, and suppose we wish to estimate the parameter

$$\phi(\theta_0) = |E[X]| \ . \tag{7}$$

Here, $\theta_0 = E[X]$, $\mathbb{D} = \mathbb{E} = \mathbf{R}$, and $\phi : \mathbf{R} \to \mathbf{R}$ satisfies $\phi(\theta) = |\theta|$ for all $\theta \in \mathbf{R}$. ∎

Our next example is a special case of the intersection bounds model studied in Hirano and Porter (2012), and Chernozhukov et al. (2013) among many others.

Example 2.2 (Intersection Bounds)**.** Let $X = (X^{(1)}, X^{(2)})' \in \mathbf{R}^2$ be a bivariate random variable, and consider the problem of estimating the parameter

$$\phi(\theta_0) = \max\{E[X^{(1)}], E[X^{(2)}]\} \ . \tag{8}$$

In this context, $\theta_0 = (E[X^{(1)}], E[X^{(2)}])'$, $\mathbb{D} = \mathbf{R}^2$, $\mathbb{E} = \mathbf{R}$, and $\phi : \mathbf{R}^2 \to \mathbf{R}$ is given by $\phi(\theta) = \max\{\theta^{(1)}, \theta^{(2)}\}$ for any $(\theta^{(1)}, \theta^{(2)})' = \theta \in \mathbf{R}^2$. Functionals such as (8) are also often employed for inference in moment inequality models; see Chernozhukov et al. (2007), Romano and Shaikh (2008), and Andrews and Soares (2010). ∎

A related example arises in conditional moment inequality models, as studied in Andrews and Shi (2013), Armstrong and Chan (2012), and Chetverikov (2012).

Example 2.3 (Conditional Moment Inequalities)**.** Let $X = (Y, Z')'$ with $Y \in \mathbf{R}$ and $Z \in \mathbf{R}^{d_z}$. For a suitable set of functions $\mathcal{F} \subset \ell^\infty(\mathbf{R}^{d_z})$, Andrews and Shi (2013) propose testing whether $E[Y|Z] \leq 0$ almost surely, by estimating the parameter

$$\phi(\theta_0) = \sup_{f \in \mathcal{F}} E[Yf(Z)] \ . \tag{9}$$

Here, $\theta_0 \in \ell^\infty(\mathcal{F})$ satisfies $\theta_0(f) = E[Yf(Z)]$ for all $f \in \mathcal{F}$, $\mathbb{D} = \ell^\infty(\mathcal{F})$, $\mathbb{E} = \mathbf{R}$, and the map $\phi : \mathbb{D} \to \mathbb{E}$ is given by $\phi(\theta) = \sup_{f \in \mathcal{F}} \theta(f)$. ∎

The following example is an abstract version of an approach pursued in Beresteanu and Molinari (2008) and Bontemps et al. (2012) for studying partially identified models.

Example 2.4 (Convex Identified Sets)**.** Let $\Lambda \subseteq \mathbf{R}^d$ denote a convex and compact set, $\mathbb{S}^d$ be the unit sphere on $\mathbf{R}^d$ and $\mathcal{C}(\mathbb{S}^d)$ denote the space of continuous functions on $\mathbb{S}^d$. For each $p \in \mathbb{S}^d$, the support function $\nu(\cdot, \Lambda) \in \mathcal{C}(\mathbb{S}^d)$ of the set Λ is then

$$\nu(p, \Lambda) \equiv \sup_{\lambda \in \Lambda} \langle p, \lambda \rangle \qquad p \in \mathbb{S}^d \ . \tag{10}$$

As noted by Beresteanu and Molinari (2008) and Bontemps et al. (2012), the functional

$$\phi(\theta_0) = \sup_{p \in \mathbb{S}^d} \{\langle p, \lambda \rangle - \nu(p, \Lambda)\} , \tag{11}$$

can form the basis for a test of whether λ is an element of Λ, since $\lambda \in \Lambda$ if and only if $\phi(\theta_0) \leq 0$. In the context of this example, $\theta_0 = \nu(\cdot, \Lambda)$, $\mathbb{D} = \mathcal{C}(\mathbb{S}^d)$, $\mathbb{E} = \mathbf{R}$, and $\phi(\theta) = \sup_{p \in \mathbb{S}^d} \{\langle p, \lambda \rangle - \theta(p)\}$ for any $\theta \in \mathcal{C}(\mathbb{S}^d)$. ∎

Our next example is based on the Linton et al. (2010) test for stochastic dominance.

Example 2.5 (Stochastic Dominance)**.** Let $X = (X^{(1)}, X^{(2)})' \in \mathbf{R}^2$ be continuously distributed, and define the marginal cdfs $F^{(j)}(u) \equiv P(X^{(j)} \leq u)$ for $j \in \{1, 2\}$. For a positive integrable weighting function $w : \mathbf{R} \to \mathbf{R}_+$, Linton et al. (2010) estimate

$$\phi(\theta_0) = \int_{\mathbf{R}} \max\{F^{(1)}(u) - F^{(2)}(u), 0\} w(u) du , \tag{12}$$

to construct a test of whether $X^{(1)}$ first order stochastically dominates $X^{(2)}$. In this example, we set $\theta_0 = (F^{(1)}, F^{(2)})$, $\mathbb{D} = \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$, $\mathbb{E} = \mathbf{R}$ and $\phi((\theta^{(1)}, \theta^{(2)})) = \int \max\{\theta^{(1)}(u) - \theta^{(2)}(u), 0\} w(u) du$ for any $(\theta^{(1)}, \theta^{(2)}) \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$. ∎

In addition to tests of stochastic dominance, a more recent literature has aimed to examine whether likelihood ratios are monotonic. Our final example is a simplification of a test proposed in Carolan and Tebbs (2005) and Beare and Moon (2013).

Example 2.6 (Likelihood Ratio Ordering)**.** Let $X = (X^{(1)}, X^{(2)})' \in \mathbf{R}^2$ have strictly increasing marginal cdfs $F_j(u) \equiv P(X^{(j)} \leq u)$, and define $G \equiv F_1 \circ F_2^{-1}$. Further let $\mathcal{M} : \ell^\infty([0, 1]) \to \ell^\infty([0, 1])$ be the least concave majorant operator, given by

$$\mathcal{M}f(u) = \inf\{g(u) : g \in \ell^\infty([0, 1]) \text{ is concave and } f(u) \leq g(u) \text{ for all } u \in [0, 1]\} \tag{13}$$

for every $f \in \ell^\infty([0, 1])$. Since the likelihood ratio dF_1/dF_2 is nonincreasing if and only if G is concave on $[0, 1]$ (Carolan and Tebbs, 2005), Beare and Moon (2013) note

$$\phi(\theta_0) = \left\{ \int_0^1 (\mathcal{M}G(u) - G(u))^2 du \right\}^{\frac{1}{2}} \tag{14}$$

characterizes whether dF_1/dF_2 is nonincreasing, since $\phi(\theta_0) = 0$ if and only if G is concave. In this example, $\theta_0 = G$, $\mathbb{D} = \ell^\infty([0, 1])$, $\mathbb{E} = \mathbf{R}$ and $\phi : \mathbb{D} \to \mathbb{E}$ satisfies $\phi(\theta) = \{\int_0^1 (\mathcal{M}\theta(u) - \theta(u))^2 du\}^{\frac{1}{2}}$ for any $\theta \in \ell^\infty([0, 1])$. ∎

2.2 Differentiability Concepts

In all the previous examples, there exist points $\theta \in \mathbb{D}$ at which the map $\phi : \mathbb{D} \to \mathbb{E}$ is not differentiable. Nonetheless, at all such θ at which differentiability is lost, ϕ actually

remains directionally differentiable. This is most easily seen in Examples 2.1 and 2.2, in which the domain of ϕ is a finite dimensional space. In order to address Examples 2.3-2.6, however, a notion of directional differentiability that is suitable for more abstract spaces $\mathbb{D}$ is necessary. Towards this end, we follow Shapiro (1990) and define

Definition 2.1. Let $\mathbb{D}$ and $\mathbb{E}$ be Banach spaces, and $\phi : \mathbb{D}_\phi \subseteq \mathbb{D} \to \mathbb{E}$.

(i) The map ϕ is said to be *Hadamard differentiable* at $\theta \in \mathbb{D}_\phi$ *tangentially* to a set $\mathbb{D}_0 \subseteq \mathbb{D}$, if there is a continuous linear map $\phi'_\theta : \mathbb{D}_0 \to \mathbb{E}$ such that:

$$\lim_{n\to\infty} \| \frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi'_\theta(h) \|_\mathbb{E} = 0 \ , \tag{15}$$

for all sequences $\{h_n\} \subset \mathbb{D}$ and $\{t_n\} \subset \mathbf{R}$ such that $t_n \to 0$, $h_n \to h \in \mathbb{D}_0$ as $n \to \infty$ and $\theta + t_n h_n \in \mathbb{D}_\phi$ for all n.

(ii) The map ϕ is said to be *Hadamard directionally differentiable* at $\theta \in \mathbb{D}_\phi$ *tangentially* to a set $\mathbb{D}_0 \subseteq \mathbb{D}$, if there is a continuous map $\phi'_\theta : \mathbb{D}_0 \to \mathbb{E}$ such that:

$$\lim_{n\to\infty} \| \frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} - \phi'_\theta(h) \|_\mathbb{E} = 0 \ , \tag{16}$$

for all sequences $\{h_n\} \subset \mathbb{D}$ and $\{t_n\} \subset \mathbf{R}_+$ such that $t_n \downarrow 0$, $h_n \to h \in \mathbb{D}_0$ as $n \to \infty$ and $\theta + t_n h_n \in \mathbb{D}_\phi$ for all n.

As has been extensively noted in the literature, Hadamard differentiability is particularly suited for generalizing the Delta method to metric spaces (van der Vaart and Wellner, 1996). It is therefore natural to employ an analogous approximation requirement when considering an appropriate definition of a directional derivative (compare (15) and (16)). However, despite this similarity, two key differences distinguish Hadamard differentiability from Hadamard directional differentiability. First, in (16) the sequence of scalars $\{t_n\}$ must approach 0 "from the right", heuristically giving the derivative a direction. Second, the map $\phi'_\theta : \mathbb{D}_0 \to \mathbb{E}$ is no longer required to be linear, though it is possible to show (16) implies ϕ'_θ must be homogenous of degree one. It is in fact this latter property that distinguishes the two differentiability concepts.

Proposition 2.1. *Let $\mathbb{D}$, $\mathbb{E}$ be Banach spaces, $\mathbb{D}_0 \subseteq \mathbb{D}$ be a subspace, and $\phi : \mathbb{D}_\phi \subseteq \mathbb{D} \to \mathbb{E}$. Then, ϕ is Hadamard directionally differentiable at $\theta \in \mathbb{D}_\phi$ tangentially to $\mathbb{D}_0$ with linear derivative $\phi'_\theta : \mathbb{D}_0 \to \mathbb{E}$ iff ϕ is Hadamard differentiable at θ tangentially to $\mathbb{D}_0$.*

Thus, while Hadamard differentiability implies Hadamard directional differentiability, Proposition 2.1 shows the converse is true if the directional derivative ϕ'_θ is linear. In what follows, we will show that linearity is in fact not important for the validity of the Delta method, but rather the key requirement is that (16) holds. Linearity, however, will play an instrumental role in determining whether the bootstrap is consistent or not.

Remark 2.1. A more general definition of Hadamard directional differentiability only requires the domain $\mathbb{D}$ to be a Hausdorff topological vector space; see Shapiro (1990). For our purposes, however, it is natural to restrict attention to Banach spaces, and we therefore employ the more specialized Definition 2.1. ∎

Remark 2.2. The condition that the map ϕ'_θ be continuous is automatically satisfied when the topology on $\mathbb{D}$ is metrizable; see Proposition 3.1 in Shapiro (1990). Consequently, when $\mathbb{D}$ is a Banach space, showing (16) holds for some map $\phi'_\theta : \mathbb{D}_0 \to \mathbb{E}$ suffices for establishing the Hadamard directional differentiability of ϕ at θ. ∎

2.2.1 Examples Revisited

We next revisit the examples to illustrate the computation of the directional derivative. The first two examples are straightforward, since the domain of ϕ is finite dimensional.

Example 2.1 (cont.) In this example, simple calculations reveal $\phi'_\theta : \mathbf{R} \to \mathbf{R}$ is

$$\phi'_\theta(h) = \begin{cases} h & \text{if } \theta > 0 \\ |h| & \text{if } \theta = 0 \\ -h & \text{if } \theta < 0 \end{cases} . \tag{17}$$

Note that ϕ is Hadamard differentiable everywhere except at $\theta = 0$, but that it is still Hadamard directionally differentiable at that point. ∎

Example 2.2 (cont.) For $\theta = (\theta^{(1)}, \theta^{(2)})' \in \mathbf{R}^2$, let $j^* = \arg\max_{j \in \{1,2\}} \theta^{(j)}$. For any $h = (h^{(1)}, h^{(2)})' \in \mathbf{R}^2$, it is then straightforward to verify $\phi'_\theta : \mathbf{R}^2 \to \mathbf{R}$ is given by

$$\phi'_\theta(h) = \begin{cases} h^{(j^*)} & \text{if } \theta^{(1)} \neq \theta^{(2)} \\ \max\{h^{(1)}, h^{(2)}\} & \text{if } \theta^{(1)} = \theta^{(2)} \end{cases} . \tag{18}$$

As in (17), ϕ'_θ is nonlinear precisely when Hadamard differentiability is not satisfied. ∎

In the next examples the domain of ϕ is infinite dimensional, and we sometimes need to employ Hadamard directional tangential differentiability – i.e. $\mathbb{D}_0 \neq \mathbb{D}$.

Example 2.3 (cont.) Suppose $E[Y^2] < \infty$ and that $\mathcal{F}$ is compact when endowed with the metric $\|f\|_{L^2(Z)} \equiv \{E[f(Z)^2]\}^{\frac{1}{2}}$. Then, $\theta_0 \in \mathcal{C}(\mathcal{F})$, and Lemma B.1 in the Appendix implies ϕ is Hadamard directionally differentiable tangentially to $\mathcal{C}(\mathcal{F})$ at any $\theta \in \mathcal{C}(\mathcal{F})$. In particular, for $\Psi_\mathcal{F}(\theta) \equiv \arg\max_{f \in \mathcal{F}} \theta(f)$, the directional derivative is

$$\phi'_\theta(h) = \sup_{f \in \Psi_\mathcal{F}(\theta)} h(f) . \tag{19}$$

Interestingly ϕ'_θ is linear at any $\theta \in \mathcal{C}(\mathcal{F})$ for which $\Psi_\mathcal{F}(\theta)$ is a singleton, and hence ϕ is actually Hadamard differentiable at such θ. We note in this example, $\mathbb{D}_0 = \mathcal{C}(\mathcal{F})$. ∎

Example 2.4 (cont.) For any $\theta \in \mathcal{C}(\mathbb{S}^d)$ let $\Psi_{\mathbb{S}^d}(\theta) \equiv \arg\max_{p \in \mathbb{S}^d}\{\langle p, \lambda \rangle - \theta(p)\}$. Lemma B.8 in Kaido (2013) then shows that $\phi'_\theta : \mathcal{C}(\mathbb{S}^d) \to \mathbf{R}$ is given by

$$\phi'_\theta(h) = \sup_{p \in \Psi_{\mathbb{S}^d}(\theta)} -h(p) . \tag{20}$$

As in Example 2.3, $\phi : \mathcal{C}(\mathbb{S}^d) \to \mathbf{R}$ is Hadamard differentiable at any $\theta \in \mathcal{C}(\mathbb{S}^d)$ at which $\Psi_{\mathbb{S}^d}(\theta)$ is a singleton, but is only Hadamard directionally differentiable otherwise. ∎

Example 2.5 (cont.) For any $\theta = (\theta^{(1)}, \theta^{(2)}) \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$ define the sets $B_0(\theta) \equiv \{u \in \mathbf{R} : \theta^{(1)}(u) = \theta^{(2)}(u)\}$ and $B_+(\theta) \equiv \{u \in \mathbf{R} : \theta^{(1)}(u) > \theta^{(2)}(u)\}$. It then follows that ϕ is Hadamard directionally differentiable at any $\theta \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$, and that

$$\phi'_\theta(h) = \int_{B_+(\theta)} (h^{(1)}(u) - h^{(2)}(u))w(u)du + \int_{B_0(\theta)} \max\{h^{(1)}(u) - h^{(2)}(u), 0\}w(u)du \tag{21}$$

for $h = (h^{(1)}, h^{(2)}) \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$ – see Lemma B.2 in the Appendix. In particular, if $B_0(\theta)$ has zero Lebesgue measure, then ϕ is Hadamard differentiable at θ. ∎

Example 2.6 (cont.) Lemma 3.2 in Beare and Moon (2013) establishes the Hadamard directional differentiability of $\mathcal{M} : \ell^\infty([0,1]) \to \ell^\infty([0,1])$ tangentially to $\mathcal{C}([0,1])$ at any concave $\theta \in \ell^\infty([0,1])$. Since norms are directionally differentiable at zero, we have

$$\phi'_\theta(h) = \left\{ \int_0^1 (\mathcal{M}'_\theta(h)(u) - h(u))^2 du \right\}^{\frac{1}{2}} \tag{22}$$

where $\mathcal{M}'_\theta : \mathcal{C}([0,1]) \to \ell^\infty([0,1])$ is the Hadamard directional derivative of $\mathcal{M}$ at θ. ∎

2.3 The Delta Method

While the Delta method for Hadamard differentiable functions has become a standard tool in econometrics (van der Vaart, 1998), the availability of an analogous result for Hadamard directional differentiable maps does not appear to be as well known. To the best of our knowledge, this powerful generalization was independently established in Shapiro (1991) and Dümbgen (1993), but only recently employed in econometrics; see Beare and Moon (2013), Kaido (2013), and Kaido and Santos (2013) for examples.

We next aim to establish a mild extension of the result in Dümbgen (1993) by showing the Delta method also holds in probability – a result we require for our subsequent derivations. Towards this end, we formalize our setup by imposing the following:

Assumption 2.1. (i) $\mathbb{D}$ and $\mathbb{E}$ are Banach spaces with norms $\|\cdot\|_{\mathbb{D}}$ and $\|\cdot\|_{\mathbb{E}}$ respectively; (ii) $\phi : \mathbb{D}_\phi \subseteq \mathbb{D} \to \mathbb{E}$ is Hadamard directionally differentiable at θ_0 tangentially to $\mathbb{D}_0$.

Assumption 2.2. (i) $\theta_0 \in \mathbb{D}_\phi$ and there are $\hat{\theta}_n : \{X_i\}_{i=1}^n \to \mathbb{D}_\phi$ such that, for some $r_n \uparrow \infty$, $r_n\{\hat{\theta}_n - \theta_0\} \overset{L}{\to} \mathbb{G}_0$ in $\mathbb{D}$; (ii) $\mathbb{G}_0$ is tight and its support is included in $\mathbb{D}_0$.

Assumption 2.3. *(i) ϕ'_{θ_0} can be continuously extended to $\mathbb{D}$ (rather than $\mathbb{D}_0 \subseteq \mathbb{D}$); (ii) $\mathbb{D}_0$ is closed under addition – i.e. $h_1 + h_2 \in \mathbb{D}_0$ for all $h_1, h_2 \in \mathbb{D}_0$.*

Assumption 2.1 simply formalizes our previous discussion by requiring that the map $\phi : \mathbb{D}_\phi \to \mathbb{E}$ be Hadamard directionally differentiable at θ_0. In Assumption 2.2(i), we additionally impose the existence of an estimator $\hat{\theta}_n$ for θ_0 that is asymptotically distributed according to $\mathbb{G}_0$ in the Hoffman-Jørgensen sense. The scaling r_n equals $\sqrt{n}$ in Examples 2.1-2.6, but may differ in nonparametric problems. In turn, Assumption 2.2(ii) imposes that the support of the limiting process $\mathbb{G}_0$ be included on the tangential set $\mathbb{D}_0$, and requires the regularity condition that the random variable $\mathbb{G}_0$ be tight. Assumption 2.3(i) allows us to view the map ϕ'_{θ_0} as well defined and continuous on all of $\mathbb{D}$ (rather than just $\mathbb{D}_0$), and is automatically satisfied when $\mathbb{D}_0$ is closed; see Remark 2.3. We emphasize, however, that Assumption 2.3(i) does not demand differentiability of $\phi : \mathbb{D}_\phi \to \mathbb{E}$ tangentially to $\mathbb{D}$ – i.e. the extension of ϕ'_{θ_0} need not satisfy (16) for $h \in \mathbb{D} \setminus \mathbb{D}_0$. For instance, in Example 2.3 ϕ is differentiable tangentially to $\mathbb{D}_0 = \mathcal{C}(\mathcal{F})$, but the map ϕ'_θ in (19) is naturally well defined and continuous on $\mathbb{D} = \ell^\infty(\mathcal{F})$. Finally, Assumption 2.3(ii) imposes that $\mathbb{D}_0$ be closed under addition which, since $\mathbb{D}_0$ is necessarily a cone, is equivalent to demanding that $\mathbb{D}_0$ be convex. This mild requirement is only employed in some of our results and helps ensure that, when multiple extensions of ϕ'_{θ_0} exist, the choice of extension has no impact in our arguments.

Remark 2.3. If $\mathbb{D}_0$ is closed, then the continuity of $\phi'_{\theta_0} : \mathbb{D}_0 \to \mathbb{E}$ and Theorem 4.1 in Dugundji (1951) imply that ϕ'_{θ_0} admits a continuous extension to $\mathbb{D}$ – i.e. there exists a continuous map $\bar{\phi}'_{\theta_0} : \mathbb{D} \to \mathbb{E}$ such that $\bar{\phi}'_{\theta_0}(h) = \phi'_{\theta_0}(h)$ for all $h \in \mathbb{D}_0$. Thus, if $\mathbb{D}_0$ is closed, then Assumption 2.3(i) is automatically satisfied. ∎

Assumptions 2.1 and 2.2 suffice for establishing the validity of the Delta method. The probabilistic version of the Delta method, however, additionally requires Assumption 2.3.

Theorem 2.1. *If Assumptions 2.1 and 2.2 hold, then $r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} \xrightarrow{L} \phi'_{\theta_0}(\mathbb{G}_0)$. If in addition Assumption 2.3(i) is also satisfied, then it follows that*

$$r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} = \phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) + o_p(1) . \tag{23}$$

The intuition behind Theorem 2.1 is the same that motivates the traditional Delta method. Heuristically, the theorem can be obtained from the approximation

$$r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} \approx \phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) , \tag{24}$$

Assumption 2.2(i), and the continuous mapping theorem applied to ϕ'_{θ_0}. Thus, the key requirement is not that ϕ'_{θ_0} be linear, or equivalently that ϕ be Hadamard differentiable, but rather that (24) holds in an appropriate sense – a condition ensured by Hadamard

directional differentiability. Following this insight, Theorem 2.1 can be established using the same arguments as in the proof of the traditional Delta method (van der Vaart and Wellner, 1996). It is worth noting that directional differentiability of ϕ is only assumed at θ_0. In particular, continuity of ϕ'_{θ_0} in θ_0 is not required since such condition is often violated; see Examples 2.1 and 2.2. Strengthening the Delta method to hold in probability further requires Assumption 2.3(i) to ensure $\phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\})$ is well defined.[2]

We conclude this section with a simple Corollary of wide applicability.

Corollary 2.1. *Let $\{X_i\}_{i=1}^n$ be a stationary sequence of random variables with $X_i \in \mathbf{R}^d$ and marginal distribution P. Suppose $\mathcal{F}$ is a collection of measurable functions $f : \mathbf{R}^d \to \mathbf{R}$, and let $\hat{\theta}_n : \mathcal{F} \to \mathbf{R}$ and $\theta_0 : \mathcal{F} \to \mathbf{R}$ be maps pointwise defined by*

$$\hat{\theta}_n(f) \equiv \frac{1}{n}\sum_{i=1}^n f(X_i) \qquad \theta_0(f) \equiv \int f(x)dP(x) \ . \tag{25}$$

Suppose $\sqrt{n}\{\hat{\theta}_n - \theta_0\} \xrightarrow{L} \mathbb{G}_0$ in $\ell^\infty(\mathcal{F})$ for some tight process $\mathbb{G}_0 \in \ell^\infty(\mathcal{F})$, and define

$$\mathcal{C}(\mathcal{F}) \equiv \{g : \mathcal{F} \to \mathbf{R} : g \text{ is continuous under } \|f\|_{\mathbb{G}_0}^2 \equiv E[\mathbb{G}_0(f)^2]\} \ .$$

If for some Banach space $\mathbb{E}$, $\phi : \ell^\infty(\mathcal{F}) \to \mathbb{E}$ is Hadamard directionally differentiable at θ_0 tangentially to $\mathcal{C}(\mathcal{F})$, then $\sqrt{n}\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} \xrightarrow{L} \phi'_{\theta_0}(\mathbb{G}_0)$ in $\mathbb{E}$.

Corollary 2.1 specializes Theorem 2.1 to the case where the parameter of interest $\phi(\theta_0)$ can be expressed as a transformation of a (possibly uncountable) collection of moments. Primitive conditions for the functional central limit theorem to hold can be found, for example, in Dehling and Philipp (2002). As a special case, Corollary 2.1 immediately delivers the relevant asymptotic distributions in Examples 2.1, 2.2, 2.3 and 2.5, but not in Examples 2.4 or 2.6. In the latter two examples $\hat{\theta}_n$ and θ_0 do not take the form in (25), and we therefore need to employ Theorem 2.1 together with the asymptotic distribution of $\sqrt{n}\{\hat{\theta}_n - \theta_0\}$ as available, for example, in Kaido and Santos (2013) for support functions and Beare and Moon (2013) for Example 2.6.

3   The Bootstrap

While Theorem 2.1 enables us to obtain an asymptotic distribution, a suitable method for estimating this limiting law is still required. In this section we will assume that the bootstrap "works" for $\hat{\theta}_n$ and examine how to leverage this result to estimate the asymptotic distribution of $r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\}$. We will show that bootstrap consistency

---

[2]Without Assumption 2.3(i), the domain of ϕ'_{θ_0} must include $\mathbb{D}_0$, but possibly not $\mathbb{D} \setminus \mathbb{D}_0$. Thus, since $r_n\{\hat{\theta}_n - \theta_0\}$ may not belong to $\mathbb{D}_0$, $\phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\})$ may otherwise not be well defined.

is often lost under Hadamard directional differentiable transformations, and propose an alternative resampling scheme which generalizes existing approaches in the literature.

3.1 Bootstrap Setup

We begin by introducing the general setup under which we examine bootstrap consistency. Throughout, we let $\hat{\theta}_n^*$ denote a "bootstrapped version" of $\hat{\theta}_n$, and assume the limiting distribution of $r_n\{\hat{\theta}_n - \theta_0\}$ can be consistently estimated by the law of

$$r_n\{\hat{\theta}_n^* - \hat{\theta}_n\} \tag{26}$$

conditional on the data. In order to formally define $\hat{\theta}_n^*$, while allowing for diverse resampling schemes, we simply impose that $\hat{\theta}_n^*$ be a function mapping the data $\{X_i\}_{i=1}^n$ and random weights $\{W_i\}_{i=1}^n$ that are independent of $\{X_i\}_{i=1}^n$ into $\mathbb{D}_\phi$. This abstract definition suffices for encompassing the nonparametric, Bayesian, block, score, and weighted bootstrap as special cases; see Remark 3.2.

Formalizing the notion of bootstrap consistency further requires us to employ a measure of distance between the limiting distribution $\mathbb{G}_0$ and its bootstrap estimator. Towards this end, we follow van der Vaart and Wellner (1996) and utilize the bounded Lipschitz metric. Specifically, for a metric space $\mathbf{A}$ with norm $\|\cdot\|_{\mathbf{A}}$, denote the set of Lipschitz functionals whose level and Lipschitz constant are bounded by one by

$$\mathrm{BL}_1(\mathbf{A}) \equiv \{f : \mathbf{A} \to \mathbf{R} : \sup_{a \in \mathbf{A}} |f(a)| \le 1 \text{ and } |f(a_1) - f(a_2)| \le \|a_1 - a_2\|_{\mathbf{A}}\} . \tag{27}$$

The bounded Lipschitz distance between two measures L_1 and L_2 on $\mathbf{A}$ then equals the largest discrepancy in the expectation they assign to functions in $\mathrm{BL}_1(\mathbf{A})$, denoted

$$d_{\mathrm{BL}}(L_1, L_2) \equiv \sup_{f \in \mathrm{BL}_1(\mathbf{A})} |\int f(a) dL_1(a) - \int f(a) dL_2(a)| . \tag{28}$$

Given the introduced notation, we can measure the distance between the law of $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$ conditional on $\{X_i\}_{i=1}^n$, and the limiting distribution of $r_n\{\hat{\theta}_n - \theta_0\}$ by[3]

$$\sup_{f \in \mathrm{BL}_1(\mathbb{D})} |E[f(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]| . \tag{29}$$

Employing the distribution of $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$ conditional on the data to approximate the distribution of $\mathbb{G}_0$ is then asymptotically justified if their distance, equivalently (29), converges in probability to zero. This type of consistency can in turn be exploited to

---

[3]More precisely, $E[f(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})|\{X_i\}_{i=1}^n]$ denotes the outer expectation with respect to the joint law of $\{W_i\}_{i=1}^n$, treating the observed data $\{X_i\}_{i=1}^n$ as constant.

validate the use of critical values obtained from the distribution of $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$ conditional on $\{X_i\}_{i=1}^n$ to conduct inference or construct confidence regions; see Remark 3.1.

We formalize the above discussion by imposing the following assumptions on $\hat{\theta}_n^*$:

Assumption 3.1. *(i) $\hat{\theta}_n^* : \{X_i, W_i\}_{i=1}^n \to \mathbb{D}_\phi$ with $\{W_i\}_{i=1}^n$ independent of $\{X_i\}_{i=1}^n$; (ii) $\hat{\theta}_n^*$ satisfies $\sup_{f \in BL_1(\mathbb{D})} |E[f(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]| = o_p(1)$.*

Assumption 3.2. *(i) The sequence $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$ is asymptotically measurable (jointly in $\{X_i, W_i\}_{i=1}^n$); (ii) $f(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\})$ is a measurable function of $\{W_i\}_{i=1}^n$ outer almost surely in $\{X_i\}_{i=1}^n$ for any continuous and bounded $f : \mathbb{D} \to \mathbf{R}$.*

Assumption 3.1(i) defines $\hat{\theta}_n^*$ in accord with our discussion, while Assumption 3.1(ii) imposes the consistency of the law of $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$ conditional on the data for the distribution of $\mathbb{G}_0$ – i.e. the bootstrap "works" for the estimator $\hat{\theta}_n$. In addition, in Assumption 3.2 we further demand mild measurability requirements on $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$. These requirements are automatically satisfied in the context of Corollary 2.1, where $\hat{\theta}_n$ and $\hat{\theta}_n^*$ correspond to the empirical and bootstrapped empirical processes respectively.

Remark 3.1. In the special case where $\mathbb{D} = \mathbf{R}^d$, Assumption 3.1(ii) implies that:

$$\sup_{t \in A} |P(r_n\{\hat{\theta}_n^* - \hat{\theta}_n\} \le t|\{X_i\}_{i=1}^n) - P(\mathbb{G}_0 \le t)| = o_p(1) \tag{30}$$

for any closed subset A of the continuity points of the cdf of $\mathbb{G}_0$; see Kosorok (2008). Thus, consistency in the bounded Lipschitz metric implies consistency of the corresponding cdfs. Result (30) then readily yields consistency of the corresponding quantiles at points at which the cdf of $\mathbb{G}_0$ is continuous and strictly increasing. ∎

Remark 3.2. Suppose $\{X_i\}_{i=1}^n$ is an i.i.d. sample, and let the parameter of interest be $\theta_0 = E[X]$ which we estimate by the sample mean $\hat{\theta}_n = \bar{X} \equiv \frac{1}{n}\sum_i X_i$. In this context, the limiting distribution of $\sqrt{n}\{\hat{\theta}_n - \theta_0\}$ can be approximated by law of

$$\sqrt{n}\{\frac{1}{n}\sum_{i=1}^n X_i^* - \bar{X}\} , \tag{31}$$

where the $\{X_i^*\}_{i=1}^n$ are drawn with replacement from the realized sample $\{X_i\}_{i=1}^n$. Equivalently, if $\{W_i\}_{i=1}^n$ is independent of $\{X_i\}_{i=1}^n$ and jointly distributed according to a multinomial distribution over n categories, each with probability $1/n$, then (31) becomes

$$\sqrt{n}\{\frac{1}{n}\sum_{i=1}^n W_i X_i - \bar{X}\} . \tag{32}$$

Thus, by defining $\hat{\theta}_n^* = \frac{1}{n}\sum_i W_i X_i$, we may express (31) in the form $\sqrt{n}\{\hat{\theta}_n^* - \hat{\theta}_n\}$. ∎

3.2 A Necessary and Sufficient Condition

When the transformation $\phi : \mathbb{D}_\phi \to \mathbb{E}$ is Hadamard differentiable at θ_0, the consistency of the bootstrap is inherited by the transformation itself. In other words, if Assumption 3.1(ii) is satisfied, and ϕ is Hadamard differentiable, then the asymptotic distribution of $r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\}$ can be consistently estimated by the law of

$$r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\} \tag{33}$$

conditional on the data (van der Vaart and Wellner, 1996). For conciseness, we refer to the law of (33) conditional on the data as the "standard" bootstrap.

Unfortunately, while the Delta method generalizes to Hadamard directionally differentiable functionals, we know by way of example that the consistency of the standard bootstrap may not (Andrews, 2000). In what follows, we aim to fully characterize the conditions under which the standard bootstrap is consistent when ϕ is Hadamard directionally differentiable. In this regard, a crucial role is played by the following concept:

Definition 3.1. Let $\mathbb{G}_1 \in \mathbb{D}_0$ be independent of $\mathbb{G}_0$ and have the same distribution as $\mathbb{G}_0$. We then say $\phi'_{\theta_0} : \mathbb{D}_0 \to \mathbb{E}$ is $\mathbb{G}_0$-translation invariant if and only if it satisfies

$$\phi'_{\theta_0}(\mathbb{G}_0 + \mathbb{G}_1) - \phi'_{\theta_0}(\mathbb{G}_0) \text{ is independent of } \mathbb{G}_0 . \tag{34}$$

Intuitively, ϕ'_{θ_0} being $\mathbb{G}_0$-translation invariant is equivalent to the distribution of

$$\phi'_{\theta_0}(\mathbb{G}_0 + h) - \phi'_{\theta_0}(h) \tag{35}$$

being constant (invariant) for all h in the support of $\mathbb{G}_0$. For example, if ϕ is Hadamard differentiable at θ_0, then ϕ'_{θ_0} is linear and hence immediately $\mathbb{G}_0$-translation invariant. On the other hand, it is also straightforward to verify that ϕ'_{θ_0} fails to be $\mathbb{G}_0$-translation invariant in Examples 2.1 and 2.2, both instances in which the standard bootstrap is known to fail; see Bickel et al. (1997) and Andrews (2000) respectively. As the following theorem shows, this relationship is not coincidental. The standard bootstrap is in fact consistent if and only if ϕ'_{θ_0} is $\mathbb{G}_0$-translation invariant.

Theorem 3.1. *Let Assumptions 2.1, 2.2, 2.3, 3.1, and 3.2 hold, and suppose that $0 \in \mathbb{D}$ is in the support of $\mathbb{G}_0$. Then, ϕ'_{θ_0} is $\mathbb{G}_0$-translation invariant if and only if*

$$\sup_{f \in BL_1(\mathbb{E})} |E[f(r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\})|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| = o_p(1) . \tag{36}$$

A powerful implication of Theorem 3.1 is that in verifying whether the standard bootstrap is valid at a conjectured θ_0, we need only examine whether ϕ'_{θ_0} is $\mathbb{G}_0$-translation invariant – an often straightforward exercise; see Remark 3.3. The theorem requires

that $0 \in \mathbb{D}$ be in the support of $\mathbb{G}_0$, which is satisfied, for example, whenever $\mathbb{G}_0$ is a centered Gaussian process. This requirement is imposed to establish that ϕ'_{θ_0} being $\mathbb{G}_0$-translation invariant implies the bootstrap is consistent. In particular, without this assumption, it is only possible to show that the bootstrap is consistent for the law in (35), which recall does not depend on h. If in addition $0 \in \mathbb{D}$ is in the support of $\mathbb{G}_0$, then from (35) we can conclude the bootstrap limit is the desired one, since then

$$\phi'_{\theta_0}(\mathbb{G}_0 + h) - \phi'_{\theta_0}(h) \stackrel{d}{=} \phi'_{\theta_0}(\mathbb{G}_0) , \tag{37}$$

where "$\stackrel{d}{=}$" denotes equality in distribution.[4] Relationship (37) is also useful in examining whether ϕ'_{θ_0} is $\mathbb{G}_0$-transaltion invariant. For instance, in the examples we study it is possible to show condition (37) is violated whenever ϕ is not Hadamard differentiable, and hence that the standard bootstrap is inconsistent.

Remark 3.3. In Examples 2.1, 2.2, 2.3, 2.4, 2.5, and 2.6 the map $\phi'_{\theta_0} : \mathbb{D}_0 \to \mathbf{R}$ satisfies

$$\phi'_{\theta_0}(h_1 + h_2) \leq \phi'_{\theta_0}(h_1) + \phi'_{\theta_0}(h_2) \tag{38}$$

for all $h_1, h_2 \in \mathbb{D}_0$. Moreover, since $\mathbb{G}_0$ is Gaussian in these examples, it is possible to verify that whenever ϕ is not Hadamard differentiable there is a $h^\star \in \mathbb{D}_0$ such that

$$P\Big(\phi'_{\theta_0}(\mathbb{G}_0 + h^\star) < \phi'_{\theta_0}(\mathbb{G}_0) + \phi'_{\theta_0}(h^\star)\Big) > 0 . \tag{39}$$

Results (38) and (39) together imply the distribution of $\phi'_{\theta_0}(\mathbb{G}_0 + h^\star) - \phi'_{\theta_0}(h^\star)$ is first order stochastically dominated by that of $\phi'_{\theta_0}(\mathbb{G}_0)$. Therefore, by (37), ϕ'_{θ_0} is not $\mathbb{G}_0$-translation invariant, and from Theorem 3.1 we conclude the bootstrap fails. ∎

3.2.1   Leading Case: Gaussian $\mathbb{G}_0$

As Theorem 3.1 shows, the consistency of the standard bootstrap is equivalent to the map $\phi'_{\theta_0} : \mathbb{D}_0 \to \mathbb{E}$ being $\mathbb{G}_0$-translation invariant – a condition concerning both ϕ'_{θ_0} and $\mathbb{G}_0$. In most applications, however, $\mathbb{G}_0$ is a centered Gaussian measure, and this additional structure has important implications for ϕ'_{θ_0} being $\mathbb{G}_0$-translation invariant. The following theorem establishes that, under Gaussianity of $\mathbb{G}_0$, ϕ'_{θ_0} is in fact $\mathbb{G}_0$-translation invariant if and only if it is linear on the support of $\mathbb{G}_0$.

Theorem 3.2. *If Assumptions 2.1, 2.2(ii) hold, and $\mathbb{G}_0$ is a centered Gaussian measure, then ϕ'_{θ_0} is $\mathbb{G}_0$-translation invariant if and only if it is linear on the support of $\mathbb{G}_0$.*

One direction of the theorem is trivial, since linearity of ϕ'_{θ_0} immediately implies ϕ'_{θ_0} must be $\mathbb{G}_0$-translation invariant (see (34)). The converse, however, is a far subtler result

---

[4]The result is exploiting that $\phi'_{\theta_0}(0) = 0$ implies $\phi'_{\theta_0}(\mathbb{G}_0 + 0) - \phi'_{\theta_0}(0) = \phi'_{\theta_0}(\mathbb{G}_0)$ almost surely; see Lemma A.3 in the Appendix for a formal derivation of (37).

which we establish by relying on insights in van der Vaart (1991) and Hirano and Porter (2012); see Remark 3.4. While perhaps not of independent interest, Theorem 3.2 has important implications when combined with our previous results. First, in conjunction with Theorem 3.1, Theorem 3.2 implies that establishing bootstrap consistency reduces to simply verifying the linearity of ϕ'_{θ_0}. Second, together with Proposition 2.1, these results show that under the maintained assumptions, Hadamard differentiability of ϕ at θ_0 is a necessary and sufficient condition for bootstrap consistency. In particular, we conclude that the bootstrap is inconsistent in all instances for which ϕ is not Hadamard differentiable at θ_0. The failure of the standard bootstrap is therefore an inherent property of these "irregular" models.

A final implication of Theorems 3.1 and 3.2 that merits discussion follows from exploiting that Gaussianity of $\mathbb{G}_0$ and bootstrap consistency together imply linearity of ϕ'_{θ_0}. In particular, whenever ϕ'_{θ_0} is linear and $\mathbb{G}_0$ is Gaussian $\phi'_{\theta_0}(\mathbb{G}_0)$ must also be Gaussian (in $\mathbb{E}$), and thus bootstrap consistency implies Gaussianity of $\phi'_{\theta_0}(\mathbb{G}_0)$. Conversely, we conclude that *the standard bootstrap fails whenever the asymptotic distribution is not Gaussian*. We formalize this conclusion in the following Corollary:

Corollary 3.1. *Let Assumptions 2.1, 2.2, 2.3, 3.1, 3.2 hold, and $\mathbb{G}_0$ be a centered Gaussian measure. If the limiting distribution of $r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\}$ is not Gaussian, then it follows that the standard bootstrap is inconsistent.*

Remark 3.4. If ϕ'_{θ_0} is $\mathbb{G}_0$-translation invariant, then the characteristic functions of $\{\phi'_{\theta_0}(\mathbb{G}_0 + h) - \phi'_{\theta_0}(h)\}$ and $\phi'_{\theta_0}(\mathbb{G}_0)$ must be equal for any h in the support of $\mathbb{G}_0$ (see (37)). The proof of Theorem 3.2 relates these characteristic functions through the Cameron-Martin theorem to show their equality implies ϕ'_{θ_0} must be linear. A similar insight was used in van der Vaart (1991) and Hirano and Porter (2012) who compare characteristic functions in a limit experiment to conclude regular estimability of a functional implies its differentiability. ∎

3.3 An Alternative Approach

Theorems 3.1 and 3.2 together establish that standard bootstrap procedures are inconsistent whenever ϕ is not fully differentiable at θ_0 and $\mathbb{G}_0$ is Gaussian. Thus, given the pervasive failure of the bootstrap in these models, we now proceed to develop a consistent estimator for the limiting distribution in Theorem 2.1 ($\phi'_{\theta_0}(\mathbb{G}_0)$).

Heuristically, the inconsistency of the standard bootstrap arises from its inability to properly estimate the directional derivative ϕ'_{θ_0} whenever it is not $\mathbb{G}_0$-translation invariant. However, the underlying bootstrap process $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$ still provides a consistent estimator for the law of $\mathbb{G}_0$. Intuitively, a consistent estimator for the limiting

distribution in Theorem 2.1 can therefore be constructed employing the law of

$$\hat{\phi}'_n(r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}) \tag{40}$$

conditional on the data for $\hat{\phi}'_n : \mathbb{D} \to \mathbb{E}$ a suitable estimator of the directional derivative $\phi'_{\theta_0} : \mathbb{D}_0 \to \mathbb{E}$. This approach is in fact closely related to the procedure developed in Andrews and Soares (2010) for moment inequality models, and other inferential methods designed for specific examples of $\phi : \mathbb{D} \to \mathbb{E}$; see Section 3.3.1 below.

In order for this approach to be valid, we require $\hat{\phi}'_n$ to satisfy the following condition:

Assumption 3.3. $\hat{\phi}'_n : \mathbb{D} \to \mathbb{E}$ *is a function of* $\{X_i\}_{i=1}^n$, *satisfying for every compact set* $K \subseteq \mathbb{D}_0$, $K^\delta \equiv \{a \in \mathbb{D} : \inf_{b \in K} \|a - b\|_{\mathbb{D}} < \delta\}$, *and every* $\epsilon > 0$, *the property:*

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} P\left(\sup_{h \in K^\delta} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} > \epsilon \right) = 0 . \tag{41}$$

Unfortunately, the requirement in (41) is complicated by the presence of the δ-enlargement of K. Without such enlargement, requirement (41) could just be interpreted as demanding that $\hat{\phi}'_n$ be uniformly consistent for ϕ'_{θ_0} on compact sets $K \subseteq \mathbb{D}_0$. Heuristically, the need to consider K^δ arises from $r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}$ only being guaranteed to lie in $\mathbb{D}$ and not necessarily $\mathbb{D}_0$. However, because $\mathbb{G}_0$ lies in compact subsets of $\mathbb{D}_0$ with arbitrarily high probability, it is possible to conclude that $r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}$ will eventually be "close" to such subsets of $\mathbb{D}_0$. Thus, $\hat{\phi}'_n$ need only be well behaved in arbitrary small neighborhoods of compact sets in $\mathbb{D}_0$, which is the requirement imposed in Assumption 3.3. It is worth noting, however, that in many applications stronger, but simpler, conditions than (41) can be easily verified. For instance, under appropriate additional requirements, the δ factor in (41) may be ignored, and it may even suffice to just verify $\hat{\phi}'_n(h)$ is consistent for $\phi'_{\theta_0}(h)$ for every $h \in \mathbb{D}_0$; see Remarks 3.5 and 3.6.

Remark 3.5. In certain applications, it is sufficient to require $\hat{\phi}'_n : \mathbb{D} \to \mathbb{E}$ to satisfy

$$\sup_{h \in K} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} = o_p(1) , \tag{42}$$

for any compact set $K \subseteq \mathbb{D}$. For instance, if $\mathbb{D} = \mathbf{R}^d$, then the closure of K^δ is compact in $\mathbb{D}$ for any compact $K \subseteq \mathbb{D}_0$, and hence (42) implies (41). Alternatively, if $\mathbb{D}$ is separable, $r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}$ is Borel measurable as a function of $\{X_i, W_i\}_{i=1}^n$ and tight for each n, then $r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}$ is uniformly tight and (42) may be used in place of (41).[5] ∎

Remark 3.6. Assumption 3.3 greatly simplifies whenever the modulus of continuity of $\hat{\phi}'_n : \mathbb{D} \to \mathbb{E}$ can be controlled outer almost surely. For instance, if $\|\hat{\phi}'_n(h_1) - \hat{\phi}'_n(h_2)\|_{\mathbb{E}} \leq$

---

[5]Under uniform tightness, for every $\epsilon > 0$ there is a compact set K such that $\limsup_{n \to \infty} P(r_n\{\hat{\theta}^*_n - \hat{\theta}_n\} \notin K) < \epsilon$. In general, however, we only know $r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}$ to be asymptotically tight, in which case we are only guaranteed $\limsup_{n \to \infty} P(r_n\{\hat{\theta}^*_n - \hat{\theta}_n\} \notin K^\delta) < \epsilon$ for every $\delta > 0$.

$C\|h_1 - h_2\|_{\mathbb{D}}$ for some $C < \infty$ and all $h_1, h_2 \in \mathbb{D}$, then showing that for any $h \in \mathbb{D}_0$

$$\|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} = o_p(1) \tag{43}$$

suffices for establishing (41) holds; see Lemma A.6 in the Appendix. This observation is particularly helpful in the analysis of Examples 2.3 and 2.4; see Section 3.3.1. ∎

Given Assumption 3.3 we can establish the validity of the proposed procedure.

Theorem 3.3. *Under Assumptions 2.1, 2.2, 2.3(i), 3.1, 3.2 and 3.3, it follows that*

$$\sup_{f \in BL_1(\mathbb{E})} |E[f(\hat{\phi}'_n(r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| = o_p(1) . \tag{44}$$

Theorem 3.3 shows that the law of $\hat{\phi}'_n(r_n\{\hat{\theta}^*_n - \hat{\theta}_n\})$ conditional on the data is indeed consistent for the limiting distribution of $r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\}$ derived in Theorem 2.1. In particular, when $\phi(\hat{\theta}_n)$ is a test statistic, and hence scalar valued, Theorem 3.3 enables us to compute critical values for inference by simulating the finite sample distribution of $\hat{\phi}'_n(r_n\{\hat{\theta}^*_n - \hat{\theta}_n\})$ conditional on $\{X_i\}_{i=1}^n$ (but not $\{W_i\}_{i=1}^n$). The following immediate corollary formally establishes this claim.

Corollary 3.2. *Let Assumptions 2.1, 2.2, 2.3(i), 3.1, 3.2 and 3.3 hold, $\mathbb{E} = \mathbf{R}$, and*

$$\hat{c}_{1-\alpha} \equiv \inf\{c : P(\hat{\phi}'_n(r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}) \le c|\{X_i\}_{i=1}^n) \ge 1 - \alpha\} . \tag{45}$$

If the cdf of $\phi'_{\theta_0}(\mathbb{G}_0)$ is strictly increasing at its $1 - \alpha$ quantile $c_{1-\alpha}$, then $\hat{c}_{1-\alpha} \xrightarrow{p} c_{1-\alpha}$.

It is worth noting that ϕ'_{θ_0} being the directional derivative of ϕ at θ_0 is actually never exploited in the proofs of Theorem 3.3 or Corollary 3.2. Therefore, these results can more generally be interpreted as providing a method for approximating distributions of random variables that are of the form $\tau(\mathbb{G}_0)$, where $\mathbb{G}_0 \in \mathbb{D}$ is a tight random variable and $\tau : \mathbb{D} \to \mathbb{E}$ is an unknown continuous map. Finally, it is important to emphasize that due to an appropriate lack of continuity of ϕ'_{θ_0} in θ_0, the "naive" estimator $\hat{\phi}'_n = \phi'_{\hat{\theta}_n}$ often fails to satisfy Assumption 3.3. Nonetheless, alternative estimators are still easily obtained as we next discuss in the context of Examples 2.1-2.6.

3.3.1 Examples Revisited

In order to illustrate the applicability of Theorem 3.3, we now return to Examples 2.1-2.6 and show existing inferential methods may be reinterpreted to fit (40). For conciseness, we group the analysis of examples that share a similar structure.

Examples 2.1 and 2.2 (cont.) In the context of Example 2.2, let $\{X_i\}_{i=1}^n$ be an i.i.d. sample with $X_i = (X_i^{(1)}, X_i^{(2)})' \in \mathbf{R}^2$, and define $\bar{X}^{(j)} \equiv \frac{1}{n}\sum_i X_i^{(j)}$ for $j \in \{1, 2\}$.

Denoting $\hat{j}^* = \arg\max_{j\in\{1,2\}} \bar{X}^{(j)}$ and letting $\kappa_n \uparrow \infty$ satisfy $\kappa_n/\sqrt{n} \downarrow 0$, we then define

$$\hat{\phi}'_n(h) = \begin{cases} h^{(\hat{j}^*)} & \text{if } |\bar{X}^{(1)} - \bar{X}^{(2)}| > \kappa_n \\ \max\{h^{(1)}, h^{(2)}\} & \text{if } |\bar{X}^{(1)} - \bar{X}^{(2)}| \leq \kappa_n \end{cases}, \tag{46}$$

(compare to (18)). Under appropriate moment restrictions, it is then straightforward to verify Assumption 3.3 holds, since $\hat{\phi}'_n : \mathbf{R}^2 \to \mathbf{R}$ in fact satisfies

$$\limsup_{n\to\infty} P\Big(\hat{\phi}'_n(h) = \phi'_{\theta_0}(h) \text{ for all } h \in \mathbf{R}^2 \Big) = 1 . \tag{47}$$

If $\{X_i^*\}_{i=1}^n$ is a sample drawn with replacement from $\{X_i\}_{i=1}^n$, and $\bar{X}^* = \frac{1}{n}\sum_i X_i^*$, then (40) reduces to $\hat{\phi}'_n(\sqrt{n}\{\bar{X}^* - \bar{X}\})$, which was originally studied in Andrews and Soares (2010) and Bugni (2010) for conducting inference in moment inequalities models. Example 2.1 can be studied in a similar manner and we therefore omit its analysis. ∎

Examples 2.3 and 2.4 (cont.) In Example 2.3, recall $\Psi_{\mathcal{F}}(\theta) \equiv \arg\max_{f\in\mathcal{F}} \theta(f)$ and suppose $\hat{\Psi}_{\mathcal{F}}(\theta_0)$ is a Hausdorff consistent estimate of $\Psi_{\mathcal{F}}(\theta_0)$ – i.e. it satisfies[6]

$$d_H(\Psi_{\mathcal{F}}(\theta_0), \hat{\Psi}_{\mathcal{F}}(\theta_0), \|\cdot\|_{L^2(Z)}) = o_p(1) . \tag{48}$$

A natural estimator for ϕ'_{θ_0} is then given by $\hat{\phi}'_n : \ell^\infty(\mathcal{F}) \to \mathbf{R}$ equal to (compare to (19))

$$\hat{\phi}'_n(h) = \sup_{f\in\hat{\Psi}_{\mathcal{F}}(\theta_0)} h(f) , \tag{49}$$

which can easily be shown to satisfy Assumption 3.3; see Lemma B.3 in the Appendix. If the data is i.i.d., $\{(Y_i^*, Z_i^*)\}_{i=1}^n$ is a sample drawn with replacement from $\{(Y_i, Z_i)\}_{i=1}^n$, and $\sqrt{n}\{\hat{\theta}_n^* - \hat{\theta}_n\}$ is the bootstrapped empirical process, then (40) becomes

$$\hat{\phi}'_n(\sqrt{n}\{\hat{\theta}_n^* - \hat{\theta}_n\}) = \sup_{f\in\hat{\Psi}_{\mathcal{F}}(\theta_0)} \frac{1}{\sqrt{n}} \sum_{i=1}^n \{Y_i^* f(Z_i^*) - \frac{1}{n}\sum_{i=1}^n Y_i f(Z_i)\} , \tag{50}$$

which was originally proposed in Andrews and Shi (2013) for conducting inference in conditional moment inequalities models. A similar approach is pursued in Kaido (2013) and Kaido and Santos (2013) in the context of Example 2.4. ∎

Examples 2.5 and 2.6 (cont.) Recall that in Example 2.5, $\theta_0 = (\theta_0^{(1)}, \theta_0^{(2)})$ with $\theta_0^{(j)} \in \ell^\infty(\mathbf{R})$ for $j \in \{1,2\}$, and that $B_0(\theta_0) = \{u \in \mathbf{R} : \theta_0^{(1)}(u) = \theta_0^{(2)}(u)\}$ and $B_+(\theta_0) = \{u \in \mathbf{R} : \theta_0^{(1)}(u) > \theta_0^{(2)}(u)\}$. For $\hat{B}_0(\theta_0)$ and $\hat{B}_+(\theta_0)$ estimators of $B_0(\theta_0)$ and

---

[6]For subsets A, B of a metric space with norm $\|\cdot\|$, the directed Hausdorff distance is $\vec{d}_H(A, B) \equiv \sup_{a\in A}\inf_{b\in B}\|a-b\|$, and the Hausdorff distance is $d_H(A, B, \|\cdot\|) \equiv \max\{\vec{d}_H(A, B, \|\cdot\|), \vec{d}_H(B, A, \|\cdot\|)\}$.

$B_+(\theta)$ respectively, it is then natural for any $h = (h^{(1)}, h^{(2)}) \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$ to define

$$\hat{\phi}'_n(h) = \int_{\hat{B}_+(\theta_0)} (h^{(1)}(u) - h^{(2)}(u))w(u)du + \int_{\hat{B}_0(\theta_0)} \max\{h^{(1)}(u) - h^{(2)}(u), 0\}w(u)du \quad (51)$$

(compare to (21)). For $A \triangle B$ the symmetric set difference between sets A and B, it is then straightforward to verify Assumption 3.3 is satisfied provided the Lebesgue measure of $B_0(\theta_0) \triangle \hat{B}_0(\theta_0)$ and $B_+(\theta_0) \triangle \hat{B}_+(\theta_0)$ converges in probability to zero. When $\sqrt{n}\{\hat{\theta}^*_n - \hat{\theta}\}$ is given by the bootstrap empirical process, $\hat{\phi}'_n(\sqrt{n}\{\hat{\theta}^*_n - \hat{\theta}_n\})$ reduces to the procedure studied in Linton et al. (2010) for testing stochastic dominance. For a related analysis of Example 2.6 we refer the reader to Beare et al. (2013). ∎

3.4 Local Analysis

As evidenced in Examples 2.1-2.6, $\phi(\hat{\theta}_n)$ is not a regular estimator for $\phi(\theta_0)$ whenever ϕ is not Hadamard differentiable at θ_0. In order to evaluate the usefulness of Theorems 2.1 and 3.3 for conducting inference, it is therefore crucial to complement these results by studying the asymptotic behavior of $r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\}$ under local perturbations to the underlying distribution of the data. In this section, we first develop such a local analysis and then proceed to examine its implications for inference.

For simplicity, we specialize to the i.i.d. setting where each X_i is distributed according to $P \in \mathbf{P}$. Here, $\mathbf{P}$ denotes the set of possible distributions for X_i and may be parametric or nonparametric in particular applications. To explicitly allow θ_0 to depend on P, we let θ_0 be the value a known map $\theta : \mathbf{P} \to \mathbb{D}_\phi$ takes at the unknown value P – e.g. $\theta_0 \equiv \theta(P)$.[7] The following Assumption formally imposes these requirements.

Assumption 3.4. *(i) $\{X_i\}_{i=1}^n$ is an i.i.d. sequence with each $X_i \in \mathbf{R}^d$ distributed according to $P \in \mathbf{P}$; (ii) $\theta_0 \equiv \theta(P)$ for some known map $\theta : \mathbf{P} \to \mathbb{D}_\phi$.*

We examine the effect of locally perturbing the distribution P through the framework of local asymptotic normality. Heuristically, we aim to conduct an asymptotic analysis in which the distribution of X_i depends on the sample size n and converges smoothly to a distribution $P \in \mathbf{P}$. In order to formalize this approach, we define a "curve in $\mathbf{P}$" by:

Definition 3.2. A function $t \mapsto \wp_t$ mapping a neighborhood $N \subseteq \mathbf{R}$ of zero into $\mathbf{P}$ is a "curve in $\mathbf{P}$" if $\wp_0 = P$ and for some $\wp'_0 : \mathbf{R}^d \to \mathbf{R}$ and dominating measure μ

$$\lim_{t \to 0} \int \frac{1}{t^2} \left(\frac{d\wp_t^{\frac{1}{2}}}{d\mu}(x) - \frac{dP^{\frac{1}{2}}}{d\mu}(x) - t\wp'_0(x) \right)^2 d\mu(x) = 0 \ . \quad (52)$$

---

[7]For instance, in Examples 2.1 and 2.2 the known map $P \mapsto \theta(P)$ is given by $\theta(P) \equiv \int x dP(x)$.

Thus, a curve in $\mathbf{P}$ is simply a parametric submodel that is smooth in the sense of being differentiable in quadratic mean. Following the literature on limiting experiments (LeCam, 1986), we consider a local analysis in which at sample size n, X_i is distributed according to $\wp_{\eta/\sqrt{n}}$ where $\wp$ is an arbitrary curve in $\mathbf{P}$ and η is an arbitrary scalar. Intuitively, as in the literature on semiparametric efficiency, such analysis enables us to characterize the local asymptotic behavior along arbitrarily rich parametric submodels of the possibly nonparametric set $\mathbf{P}$. To proceed, however, we must first specify how the original estimator $\hat\theta_n$ is affected by these local perturbations, and to this end we impose:

Assumption 3.5. *(i) $\hat\theta_n$ is a regular estimator for $\theta(P)$;[8] (ii) For every curve $\wp$ in $\mathbf{P}$ there is a $\theta'(\wp) \in \mathbb{D}_0$ such that $\|\theta(\wp_t) - \theta(P) - t\theta'(\wp)\|_{\mathbb{D}} = o(t)$ (as $t \to 0$).*

Assumption 3.5(i) demands that the distributional convergence of $\hat\theta_n$ be robust to local perturbations of P, while Assumption 3.5(ii) imposes that the parameter $P \mapsto \theta(P)$ be smooth in P. As shown in van der Vaart (1991), these requirements are closely related, whereby Assumption 3.5(i) and mild regularity conditions on $\hat\theta_n$ and the tangent space actually imply Assumption 3.5(ii). Assumption 3.5 is immediately satisfied, for instance, when $\theta(P)$ is a (possible uncountable) collection of moments, as in Examples 2.1, 2.2, 2.3 and 2.5. We also note that our results can still be applied in instances where $\theta(P)$ does not admit for a regular estimator, but can be expressed as a Hadamard directionally transformation of a regular parameter; see Remark 3.7.

Remark 3.7. Suppose $\theta(P)$ is not a regular parameter, but that $\theta(P) = \psi(\vartheta(P))$ for some parameter $\vartheta(P)$ admitting a regular estimator $\hat\vartheta_n$, and a Hadamard directionally differentiable map ψ. By the chain rule for Hadamard directionally differentiable maps (Shapiro, 1990), our results may then be applied with $\tilde\phi \equiv \phi \circ \psi$, $\tilde\theta(P) \equiv \vartheta(P)$, and $\hat\vartheta_n$ in place of ϕ, $\theta(P)$ and $\hat\theta_n$ respectively. ∎

Given the stated assumptions, we can now establish the following Lemma.

Lemma 3.1. *For an arbitrary curve $\wp$ in $\mathbf{P}$ and $\eta \in \mathbf{R}$ let $P_n = \wp_{\eta/\sqrt{n}}$ and L_n denote the law under $\bigotimes_{i=1}^n P_n$. If Assumptions 2.1, 2.2, 2.3, 3.4, and 3.5 hold, then*

$$\sqrt{n}\{\phi(\hat\theta_n) - \phi(\theta(P_n))\} \overset{L_n}{\to} \phi'_{\theta_0}(\mathbb{G}_0 + \eta\theta'(\wp)) - \phi'_{\theta_0}(\eta\theta'(\wp)) . \tag{53}$$

Lemma 3.1 characterizes the asymptotic distribution of $\phi(\hat\theta_n)$ under a sequence of local perturbations to P. As expected, the asymptotic limit in (53) need not equal the pointwise asymptotic distribution derived in Theorem 2.1. Intuitively, the asymptotic approximation in (53) reflects the importance of local parameters and for this reason can be expected to provide a better approximation to finite sample distributions – a point

---

[8]Formally, $\hat\theta_n$ is a regular estimator if for every curve $\wp$ in $\mathbf{P}$ and every $\eta \in \mathbf{R}$ we have $\sqrt{n}\{\hat\theta_n - \theta(P_n)\} \overset{L_n}{\to} \mathbb{G}_0$, where $P_n \equiv \wp_{\eta/\sqrt{n}}$ and L_n denotes the law under $\bigotimes_{i=1}^n P_n$.

forcefully argued in the study of moment inequality models by Andrews and Soares (2010) and Andrews and Shi (2013); see Remark 3.8 below.

Remark 3.8. In the context of Example 2.2, let $\{X_i\}_{i=1}^n$ be an i.i.d. sample with $X_i \sim P$, $\hat{\theta}_n = \frac{1}{n}\sum X_i$ and $\theta(P) \equiv \int x dP(x)$. By Theorem 2.1 we then obtain

$$\sqrt{n}\{\phi(\hat{\theta}_n) - \phi(\theta(P))\} \overset{L}{\to} \begin{cases} \mathbb{G}_0^{(j^*)} & \text{if } \theta^{(1)}(P) \neq \theta^{(2)}(P) \\ \max\{\mathbb{G}_0^{(1)}, \mathbb{G}_0^{(2)}\} & \text{if } \theta^{(1)}(P) = \theta^{(2)}(P) \end{cases}, \qquad (54)$$

where $\mathbb{G}_0 = (\mathbb{G}_0^{(1)}, \mathbb{G}_0^{(2)})'$ is a normal vector, and $j^* = \arg\max_{j \in \{1,2\}} \theta^{(j)}(P)$ (see (18)). As argued in Andrews and Soares (2010), the discontinuity of the pointwise asymptotic distribution in (54) can be a poor approximation for the finite sample distribution which depends continuously on $\theta^{(1)}(P) - \theta^{(2)}(P)$. An asymptotic analysis local to a P such that $\theta^{(1)}(P) = \theta^{(2)}(P)$, however, lets us address this problem. Specifically, for a submodel $\wp$ with $\theta(\wp_t) = \theta(P) + th$ for $t \in \mathbf{R}$ and $h = (h^{(1)}, h^{(2)})' \in \mathbf{R}^2$, Lemma 3.1 yields

$$\sqrt{n}\{\phi(\hat{\theta}_n) - \phi(\theta(P_n))\} \overset{L_n}{\to} \max\{\mathbb{G}_0^{(1)} + h^{(1)}, \mathbb{G}_0^{(2)} + h^{(2)}\} - \max\{h^{(1)}, h^{(2)}\}. \qquad (55)$$

Thus, by reflecting the importance of the "slackness" parameter h, (55) provides a better framework with which to evaluate the performance of our proposed procedure. ∎

It is interesting to note that by setting $\eta = 0$ in (53) we can conclude from Lemma 3.1 that $\phi(\hat{\theta}_n)$ is a regular estimator for $\phi(\theta(P))$ if and only if

$$\phi'_{\theta_0}(\mathbb{G}_0 + \eta\theta'(\wp)) - \phi'_{\theta_0}(\eta\theta'(\wp)) \overset{d}{=} \phi'_{\theta_0}(\mathbb{G}_0) \qquad (56)$$

for all curves $\wp$ in $\mathbf{P}$ and all scalars $\eta \in \mathbf{R}$. Therefore, we immediately obtain from Lemma 3.1 that $\phi(\hat{\theta}_n)$ is a regular estimator for $\phi(\theta(P))$ whenever ϕ'_{θ_0} is linear, or equivalently, whenever ϕ is Hadamard differentiable at $\theta_0 = \theta(P)$. More generally, however, Lemma 3.1 implies $\phi(\hat{\theta}_n)$ will often not be regular when ϕ is directionally, but not fully, Hadamard differentiable at θ_0. Condition (56) in fact closely resembles the requirement that ϕ'_{θ_0} be $\mathbb{G}_0$-translation invariant (compare to (37)). In order to formalize this connection, we let $\overline{\bigcup_\wp \theta'(\wp)}$ denote the closure under $\|\cdot\|_{\mathbb{D}}$ of the collection of $\theta'(\wp)$ generated by all curves $\wp \in \mathbf{P}$. The following Corollary shows that, under the requirement that the support of $\mathbb{G}_0$ be equal to $\overline{\bigcup_\wp \theta'(\wp)}$ (see Remark 3.9), $\phi(\hat{\theta}_n)$ is indeed a regular estimator if and only if ϕ'_{θ_0} is $\mathbb{G}_0$-translation invariant.

Corollary 3.3. *If Assumptions 2.1, 2.2, 2.3, 3.4, 3.5 hold, and the support of $\mathbb{G}_0$ equals $\overline{\bigcup_\wp \theta'(\wp)}$, then $\phi(\hat{\theta}_n)$ is a regular estimator if and only if ϕ'_{θ_0} is $\mathbb{G}_0$-translation invariant.*

Perhaps the most interesting implication of Corollary 3.3 arises from combining it with Theorem 3.1. Together, these results imply that the standard bootstrap is consistent if and only if $\phi(\hat{\theta}_n)$ is a regular estimator for $\phi(\theta(P))$. Thus, we can conclude

from Corollary 3.3 that the failure of the bootstrap is an innate characteristic of irregular models. A similar relationship between regularity and bootstrap consistency had been found by Beran (1997), who showed that in finite dimensional likelihood models the parametric bootstrap is consistent if and only if the estimator is regular.

Remark 3.9. Since $\hat{\theta}_n$ is a regular estimator, the Convolution Theorem implies that

$$\mathbb{G}_0 \stackrel{d}{=} \Delta_0 + \Delta_1 \ ,$$

where: (i) Δ_0 is centered Gaussian, (ii) Δ_0 and Δ_1 are independent, and (iii) the support of Δ_0 equals $\overline{\bigcup_\wp \theta'(\wp)}$; see, for example, Theorem 3.11.2 in van der Vaart and Wellner (1996). Hence, since the support of Δ_0 is a vector space, we conclude that the requirement that the support of $\mathbb{G}_0$ be equal to $\overline{\bigcup_\wp \theta'(\wp)}$ is satisfied whenever the support of Δ_1 is included in that of Δ_0 – for example, whenever $\hat{\theta}_n$ is efficient. ∎

3.4.1 Implications for Testing

As has been emphasized in the moment inequalities literature, the lack of regularity of $\phi(\hat{\theta}_n)$ can render pointwise (in P) asymptotic approximations unreliable (Imbens and Manski, 2004). However, since in Examples 2.2, 2.3, and 2.5 our results encompass procedures that are valid uniformly in P, we also know that irregularity of $\phi(\hat{\theta}_n)$ does not preclude our approach from remaining valid (Andrews and Soares, 2010; Linton et al., 2010; Andrews and Shi, 2013). In what follows, we note that the aforementioned examples are linked by the common structure of ϕ'_{θ_0} being subadditive. More generally, we exploit Lemma 3.1 to show that whenever such property holds, the bootstrap procedure of Theorem 3.3 can control size locally to P along arbitrary submodels.

We consider hypothesis testing problems in which ϕ is scalar valued ($\mathbb{E} = \mathbf{R}$), and we are concerned with evaluating whether $P \in \mathbf{P}$ satisfies

$$H_0 : \phi(\theta(P)) \leq 0 \qquad\qquad H_1 : \phi(\theta(P)) > 0 \ . \qquad\qquad (57)$$

A natural test statistic for this problem is then $\sqrt{n}\phi(\hat{\theta}_n)$, while Theorem 2.1 suggests

$$c_{1-\alpha} \equiv \inf\{c : P(\phi'_{\theta_0}(\mathbb{G}_0) \leq c) \geq 1 - \alpha\}$$

is an appropriate unfeasible critical value for a $1 - \alpha$ level test.[9] For $\hat{c}_{1-\alpha}$ the developed boootstrap estimator for $c_{1-\alpha}$ (see (45)), Theorem 2.1 and Corollary 3.2 then establish the (pointwise in P) validity of rejecting H_0 whenever $\sqrt{n}\phi(\hat{\theta}_n) > \hat{c}_{1-\alpha}$.

In order to evaluate both the local size control and local power of the proposed test,

---

[9]Note that $c_{1-\alpha}$ is the $1 - \alpha$ quantile of the asymptotic distribution of $\sqrt{n}\phi(\hat{\theta}_n)$ when $\phi(\theta(P)) = 0$.

we assume $\phi(\theta(P)) = 0$ and consider curves $\wp$ in $\mathbf{P}$ that also belong to the set

$$\mathbf{H} \equiv \{\wp : \text{ (i) } \phi(\theta(\wp_t)) \leq 0 \text{ if } t \leq 0, \text{and (ii) } \phi(\theta(\wp_t)) > 0 \text{ if } t > 0\} \ .$$

Thus, a curve $\wp \in \mathbf{H}$ is such that $\wp_t$ satisfies the null hypothesis whenever $t \leq 0$, but switches to satisfying the alternative hypothesis at all $t > 0$. As in Lemma 3.1, for a curve $\wp \in \mathbf{H}$ and scalar η we let $P_n^n \equiv \bigotimes_{i=1}^n \wp_{\eta/\sqrt{n}}$, and we denote the power at sample size n for the test that rejects whenever $\sqrt{n}\phi(\hat{\theta}_n) > \hat{c}_{1-\alpha}$ by

$$\pi_n(\wp_{\eta/\sqrt{n}}) \equiv P_n^n(\sqrt{n}\phi(\hat{\theta}_n) > \hat{c}_{1-\alpha}) \ .$$

To conduct the local analysis, we further require the following Assumption.

Assumption 3.6. *(i) $\mathbb{E} = \mathbf{R}$; (ii) The cdf of $\phi'_{\theta_0}(\mathbb{G}_0)$ is continuous and strictly increasing at $c_{1-\alpha}$; (iii) $\phi'_{\theta_0}(h_1 + h_2) \leq \phi'_{\theta_0}(h_1) + \phi'_{\theta_0}(h_2)$ for all $h_1, h_2 \in \mathbb{D}_0$.*

Assumption 3.6(i) formalizes the requirement that ϕ be scalar valued. In turn, in Assumption 3.6(ii) we impose that the cdf of $\phi'_{\theta_0}(\mathbb{G}_0)$ be strictly increasing and continuous. Strict monotonicity is required to establish the consistency of $\hat{c}_{1-\alpha}$, while continuity ensures the test controls size at least pointwise in P. Assumption 3.6(iii) demands that ϕ'_{θ_0} be subadditive, which represents the key condition that ensures local size control. Since ϕ'_{θ_0} is also positively homogenous of degree one, Assumption 3.6(iii) is in fact equivalent to demanding that ϕ'_{θ_0} be convex, which greatly simplifies verifying Assumption 3.6(ii) when $\mathbb{G}_0$ is Gaussian; see Remark 3.11. We further note that Assumption 3.6 is trivially satisfied when ϕ'_{θ_0} is linear, which by Lemma 3.1 also implies $\phi(\hat{\theta}_n)$ is regular. However, we emphasize that Assumption 3.6 can also hold at points $\theta(P)$ at which ϕ is not Hadamard differentiable, as is easily verified in Examples 2.1-2.6.

The following Theorem derives the asymptotic limit of the power $\pi_n(\wp_{\eta/\sqrt{n}})$.

Theorem 3.4. *Let Assumptions 2.1, 2.2, 2.3, 3.1, 3.2, 3.3, 3.4, 3.5, and 3.6(i)-(ii) hold. It then follows that for any curve $\wp$ in $\mathbf{H}$, and every $\eta \in \mathbf{R}$ we have*

$$\liminf_{n \to \infty} \pi_n(\wp_{\eta/\sqrt{n}}) \geq P(\phi'_{\theta_0}(\mathbb{G}_0 + \eta\theta'(\wp)) > c_{1-\alpha}) \ . \tag{58}$$

If in addition Assumption 3.6(iii) also holds, then we can conclude that for any $\eta \leq 0$

$$\limsup_{n \to \infty} \pi_n(\wp_{\eta/\sqrt{n}}) \leq \alpha \ . \tag{59}$$

The first claim of the Theorem derives a lower bound on the power against local alternatives, with (58) holding with equality whenever $c_{1-\alpha}$ is a continuity point of the cdf of $\phi'_{\theta_0}(\mathbb{G}_0 + \eta\theta'(\wp))$. In turn, provided ϕ'_{θ_0} is subadditive, the second claim of Theorem

3.4 establishes the ability of the test to locally control size along parametric submodels. Heuristically, the role of subadditivity can be seen from (58) and the inequalities

$$P(\phi'_{\theta_0}(\mathbb{G}_0 + \eta\theta'(\wp)) > c_{1-\alpha}) \le P(\phi'_{\theta_0}(\mathbb{G}_0) + \phi'_{\theta_0}(\eta\theta'(\wp)) > c_{1-\alpha}) \le \alpha \ ,$$

where the final inequality results from $\phi'_{\theta_0}(\eta\theta'(\wp)) \le 0$ due to $\phi(\theta(P_n)) - \phi(\theta(P)) \le 0$.[10] Thus, ϕ'_{θ_0} being subadditive implies $\eta = 0$ is the "least favorable" point in the null, which in turn delivers local size control as in (59). We note a similar logic can be employed to evaluate confidence regions built using Theorems 2.1 and 3.3; see Remark 3.10.

Since the results of Theorem 3.4 are local to P in nature, their relevance is contingent to them applying to all $P \in \mathbf{P}$ that are deemed possible distributions of the data. We emphasize that the three key requirements in this regard are Assumptions 3.5(i), 3.6(ii), and 3.6(iii) – i.e. that $\hat\theta_n$ be regular, the cdf of $\phi'_{\theta_0}(\mathbb{G}_0)$ be continuous and strictly increasing at $c_{1-\alpha}$, and that ϕ'_{θ_0} be subadditive. We view Assumption 3.6(ii) as mainly a technical requirement that can be dispensed with following insights in Andrews and Shi (2013); see Remark 3.12. Regularity of $\hat\theta_n$ and subadditivity of ϕ'_{θ_0}, however, are instrumental in establishing the validity of our proposed procedure. In certain applications, such as in Examples 2.1, 2.2, 2.3, and 2.5, both these requirements are seen to be easily satisfied for a large class of possible P. However, in other instances, such as in Example 2.4 applied to estimator in Kaido and Santos (2013), ϕ'_{θ_0} is always subadditive, but the regularity of $\hat\theta_n$ can fail to hold for an important class of P.

Remark 3.10. As usual, we can obtain confidence regions for $\phi(\theta(P))$ by test inverting

$$H_0 : \phi(\theta(P)) = c_0 \qquad\qquad H_1 : \phi(\theta(P)) \ne c_0 \ , \tag{60}$$

for different $c_0 \in \mathbb{E}$. Defining $\bar\phi : \mathbb{D}_\phi \subseteq \mathbb{D} \to \mathbf{R}$ pointwise by $\bar\phi(\theta) \equiv \|\phi(\theta) - c_0\|_\mathbb{E}$, it is then straightforward to see (60) can be expressed as in (57) with $\bar\phi$ in place of ϕ. In particular, the chain rule implies $\bar\phi'_{\theta_0}(\cdot) = \|\phi'_{\theta_0}(\cdot)\|_\mathbb{E}$, and hence the subadditivity of $\|\phi'_{\theta_0}(\cdot)\|_\mathbb{E}$ suffices for establishing local size control. ∎

Remark 3.11. Under Assumptions 2.1 and 3.6(iii), it follows that $\phi'_{\theta_0} : \mathbb{D}_0 \to \mathbf{R}$ is a continuous convex functional. Therefore, if $\mathbb{G}_0$ is in addition Gaussian, then Theorem 11.1 in Davydov et al. (1998) implies that the cdf of $\phi'_{\theta_0}(\mathbb{G}_0)$ is continuous and strictly increasing at all points in the interior of its support (relative to $\mathbf{R}$). ∎

Remark 3.12. In certain applications, such as in Examples 2.3 and 2.5, Assumption 3.6(ii) may be violated at distributions P of interest. To address this problem, Andrews and Shi (2013) propose employing the critical value $\hat c_{1-\alpha} + \delta$ for an arbitrarily small $\delta > 0$. It is then possible to show that, even if Assumption 3.6(ii) fails, we still have

$$\liminf_{n\to\infty} P(\hat c_{1-\alpha} + \delta \ge c_{1-\alpha}) = 1 \ . \tag{61}$$

---

[10]More precisely, we are exploiting that $\phi'_{\theta_0}(\eta\theta'(\wp)) = \lim_{n\to\infty} \sqrt{n}\{\phi(\theta(P_n)) - \phi(\theta(P))\} \le 0$.

Therefore, by contiguity it follows that the local size control established in (59) holds without Assumption 3.6(ii) if we employ $\hat{c}_{1-\alpha} + \delta$ instead of $\hat{c}_{1-\alpha}$. $\blacksquare$

4   Convex Set Projections

In this section, we demonstrate the usefulness of the developed asymptotic framework by constructing a test of whether a Hilbert space valued parameter belongs to a known closed convex set. Despite the generality of the problem, we show that its geometry and our previous results make its analysis transparent and straightforward.

4.1   Projection Setup

In what follows, we let $\mathbb{H}$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ and norm $\| \cdot \|_{\mathbb{H}}$. For a known closed convex set $\Lambda \subseteq \mathbb{H}$, we then consider the hypothesis testing problem

$$H_0 : \theta_0 \in \Lambda \qquad H_1 : \theta_0 \notin \Lambda , \qquad (62)$$

where the parameter $\theta_0 \in \mathbb{H}$ is unknown, but for which we possess an estimator $\hat{\theta}_n$. Special cases of this problem have been widely studied in the setting where $\mathbb{H} = \mathbf{R}^d$, and to a lesser extent when $\mathbb{H}$ is infinite dimensional; see Examples 4.1-4.3 below.

We formalize the introduced structure through the following assumption.

Assumption 4.1. *(i) $\mathbb{D} = \mathbb{H}$ where $\mathbb{H}$ is Hilbert Space with inner product $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ and corresponding norm $\| \cdot \|_{\mathbb{H}}$; (ii) $\Lambda \subseteq \mathbb{H}$ is a known closed and convex set.*

Since projections onto closed convex sets in Hilbert spaces are attained and unique, we may define the projection operator $\Pi_\Lambda : \mathbb{H} \to \Lambda$, which for each $\theta \in \mathbb{H}$ satisfies

$$\|\theta - \Pi_\Lambda \theta\|_{\mathbb{H}} = \inf_{h \in \Lambda} \|\theta - h\|_{\mathbb{H}} . \qquad (63)$$

Thus, the hypothesis testing problem in (62) can be rewritten in terms of the distance between θ_0 and Λ, or equivalently between θ_0 and its projection $\Pi_\Lambda \theta_0$ – i.e.

$$H_0 : \|\theta_0 - \Pi_\Lambda \theta_0\|_{\mathbb{H}} = 0 \qquad H_1 : \|\theta_0 - \Pi_\Lambda \theta_0\|_{\mathbb{H}} > 0 . \qquad (64)$$

Interpreted in this manner, it is clear that (64) is a special case of (57), with $\mathbb{D} = \mathbb{H}$, $\mathbb{E} = \mathbf{R}$, and $\phi : \mathbb{H} \to \mathbf{R}$ given by $\phi(\theta) \equiv \|\theta - \Pi_\Lambda \theta\|_{\mathbb{H}}$ for any $\theta \in \mathbb{H}$. The corresponding test statistic $r_n \phi(\hat{\theta}_n)$ is then simply the scaled distance between the estimator $\hat{\theta}_n$ and the known convex set Λ – i.e. $r_n \phi(\hat{\theta}_n) = r_n \|\hat{\theta}_n - \Pi_\Lambda \hat{\theta}_n\|_{\mathbb{H}}$.

Figure 1: Illustrations of Tangent Cones.
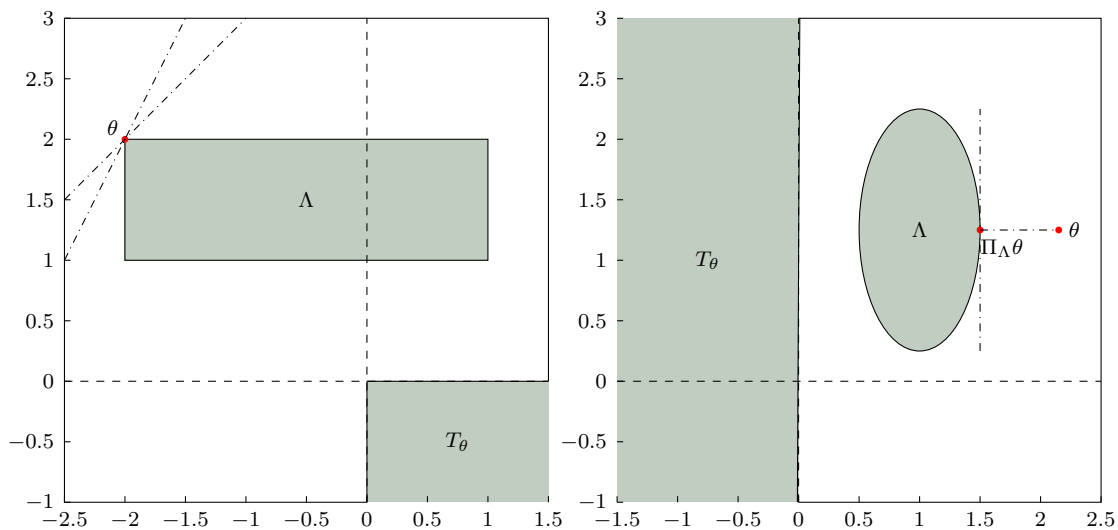
As a final piece of notation, we need to introduce the tangent cone of Λ at a $\theta \in \mathbb{H}$, which plays a fundamental role in our analysis. To this end, for any set $A \subseteq \mathbb{H}$ let $\overline{A}$ denote its closure under $\|\cdot\|_{\mathbb{H}}$, and define the tangent cone of Λ at $\theta \in \mathbb{H}$ by

$$T_\theta \equiv \overline{\bigcup_{\alpha \geq 0} \alpha\{\Lambda - \Pi_\Lambda \theta\}} \; , \tag{65}$$

which is convex by convexity of Λ. Heuristically, T_θ represents the directions from which the projection $\Pi_\Lambda \theta \in \Lambda$ can be approached from within the set Λ. As such, T_θ can be seen as a local approximation to the set Λ at $\Pi_\Lambda \theta$ and employed to study the differentiability properties of the projection operator Π_Λ. Figure 4.1 illustrates the tangent cone in two separate cases: one in which $\theta \in \Lambda$, and a second in which $\theta \notin \Lambda$.

4.1.1   Examples

In order to aid exposition and illustrate the generality of (62), we next provide examples of both well studied and new problems that fit our framework.

Example 4.1. Suppose $X \in \mathbf{R}^d$ and that we aim to test the moment inequalities

$$H_0 : E[X] \leq 0 \qquad H_1 : E[X] \nleq 0 \; , \tag{66}$$

where the null is meant to hold at all coordinates, and the alternative indicates at least one coordinate of $E[X]$ is strictly positive. In this instance, $\mathbb{H} = \mathbf{R}^d$, Λ is the negative

orthant in $\mathbf{R}^d$ ($\Lambda \equiv \{h \in \mathbf{R}^d : h \leq 0\}$), and the distance of θ to Λ is equal to

$$\phi(\theta) = \|\Pi_\Lambda \theta - \theta\|_{\mathbb{H}} = \Big\{ \sum_{i=1}^d (E[X^{(i)}])_+^2 \Big\}^{\frac{1}{2}} \ , \tag{67}$$

where $(a)_+ = \max\{a, 0\}$ and $X^{(i)}$ denotes the ith coordinate of X. More generally, this example applies to any regular parameter in $\mathbf{R}^d$ such as testing for moment inequalities on regression coefficients (Wolak, 1988). Analogously, conditional moment inequalities as in Example 2.3 can be encompassed by employing a weight function on $\mathcal{F}$ – this approach leads to the Cramer-von-Mises statistic studied in Andrews and Shi (2013). $\blacksquare$

The next example concerns quantile models, as employed by Buchinsky (1994) to characterize the U.S. wage structure conditional on levels of education, or by Abadie et al. (2002) to estimate the effect of subsidized training on earnings.

Example 4.2. Let $(Y, D, X) \in \mathbf{R} \times \mathbf{R} \times \mathbf{R}^{d_z}$ and consider the quantile regression

$$(\theta_0(\tau), \beta(\tau)) \equiv \arg \min_{\theta \in \mathbf{R}, \beta \in \mathbf{R}^{d_z}} E[\rho_\tau(Y - D\theta - Z'\beta)] \tag{68}$$

where $\rho_\tau(u) = (\tau - 1\{u \leq 0\})u$ and $\tau \in (0, 1)$. Under appropriate restrictions, the estimator $\hat{\theta}_n$ for θ_0 converges in distribution in $\ell^\infty([\epsilon, 1 - \epsilon])$ for any $\epsilon > 0$ (Angrist et al., 2006).[11] Hence, we may test shape restrictions on θ_0 by letting

$$\mathbb{H} \equiv \{\theta : [\epsilon, 1 - \epsilon] \to \mathbf{R} : \langle \theta, \theta \rangle_{\mathbb{H}} < \infty\} \qquad \langle \theta_1, \theta_2 \rangle_{\mathbb{H}} \equiv \int_\epsilon^{1-\epsilon} \theta_1(\tau) \theta_2(\tau) d\tau \ , \tag{69}$$

and considering appropriate convex sets $\Lambda \subseteq \mathbb{H}$. For instance, in randomized experiments where D is a dummy for treatment, $\theta_0(\tau)$ is the quantile treatment effect and we may test for its constancy or monotonicity; see Muralidharan and Sundararaman (2011) for an examination of these features in the evaluation of teacher performance pay. A similar approach may also be employed to test whether the pricing kernel satisfies theoretically predicted restrictions such as a monotonicity (Jackwerth, 2000). $\blacksquare$

Our final example may be interpreted as a generalization of Example 4.2.

Example 4.3. Let $Z \in \mathbf{R}^{d_z}$, $\Theta \subseteq \mathbf{R}^{d_\theta}$, and $\mathcal{T} \subseteq \mathbf{R}^{d_\tau}$. Suppose there exists a function $\rho : \mathbf{R}^{d_z} \times \Theta \times \mathcal{T} \to \mathbf{R}^{d_\rho}$ such that for each $\tau \in \mathcal{T}$ there is a unique $\theta_0(\tau) \in \Theta$ satisfying

$$E[\rho(Z, \theta_0(\tau), \tau)] = 0 \ . \tag{70}$$

Such a setting arises, for instance, in sensitivity analysis (Chen et al., 2011), and in partially identified models where the identified set is a curve (Arellano et al., 2012)

---

[11]This result also holds for the instrumental variables estimator of Chernozhukov and Hansen (2005).

or can be described by a functional lower and upper bound (Kline and Santos, 2013; Chandrasekarh et al., 2013). Escanciano and Zhu (2013) derives an estimator $\hat{\theta}_n$ which converges in distribution in $\bigotimes_{i=1}^{d_\theta} \ell^\infty(\mathcal{T})$, and hence for an integrable function w also in

$$\mathbb{H} \equiv \{\theta : \mathcal{T} \to \mathbf{R}^{d_\theta} : \langle \theta, \theta \rangle_\mathbb{H} < \infty\} \qquad \langle \theta_1, \theta_2 \rangle_\mathbb{H} \equiv \int_\mathcal{T} \theta_1(\tau)' \theta_2(\tau) w(\tau) d\tau . \qquad (71)$$

Appropriate choices of Λ then enable us to test, for example, whether the model is identified in Arellano et al. (2012), or whether the identified set in Kline and Santos (2013) is consistent with increasing returns to education across quantiles. ∎

4.2 Theoretical Results

4.2.1 Asymptotic Distribution

Our analysis crucially relies on the seminal work of Zaranotello (1971), who established the Hadamard directional differentiability of metric projections onto convex sets in Hilbert spaces. Specifically, Zaranotello (1971) showed $\Pi_\Lambda : \mathbb{H} \to \Lambda$ is Hadamard directionally differentiable at any $\theta \in \Lambda$, and its directional derivative is equal to the projection operator onto the tangent cone of Λ at θ, which we denote by $\Pi_{T_\theta} : \mathbb{H} \to T_\theta$. Figure 2 illustrates a simple example in which the derivative approximation

$$\Pi_\Lambda \theta_1 - \Pi_\Lambda \theta_0 \approx \Pi_{T_{\theta_0}}(\theta_1 - \theta_0) \qquad (72)$$

actually holds with equality.[12] We note that it is also immediate from Figure 2 that the directional derivative $\Pi_{T_{\theta_0}}$ is not linear, and hence Π_Λ is not fully differentiable.

Given the result in Zaranotello (1971), the asymptotic distribution of $r_n \phi(\hat{\theta}_n)$ can then be obtained as an immediate consequence of Theorem 2.1.

Proposition 4.1. *Let Assumption 2.2 and 4.1 hold. If $\theta_0 \in \Lambda$, then it follows that*

$$r_n \|\hat{\theta}_n - \Pi_\Lambda \hat{\theta}_n\|_\mathbb{H} \xrightarrow{L} \|\mathbb{G}_0 - \Pi_{T_{\theta_0}} \mathbb{G}_0\|_\mathbb{H} . \qquad (73)$$

In particular, Proposition 4.1 follows from norms being directionally differentiable at zero, and hence by the chain rule the directional derivative $\phi'_{\theta_0} : \mathbb{H} \to \mathbf{R}$ satisfies

$$\phi'_{\theta_0}(h) = \|h - \Pi_{T_{\theta_0}} h\|_\mathbb{H} . \qquad (74)$$

It is interesting to note that $\Lambda \subseteq T_{\theta_0}$ whenever Λ is a cone, and hence $\|h - \Pi_{T_{\theta_0}} h\|_\mathbb{H} \le \|h - \Pi_\Lambda h\|_\mathbb{H}$ for all $h \in \mathbb{H}$. Therefore, the distribution of $\|\mathbb{G}_0 - \Pi_\Lambda \mathbb{G}_0\|_\mathbb{H}$ first order

---

[12]Note that in Figure 2 we are exploiting that $\Pi_\Lambda \theta_0 = \theta_0$ if $\theta_0 \in \Lambda$.

Figure 2: Directional Differentiability
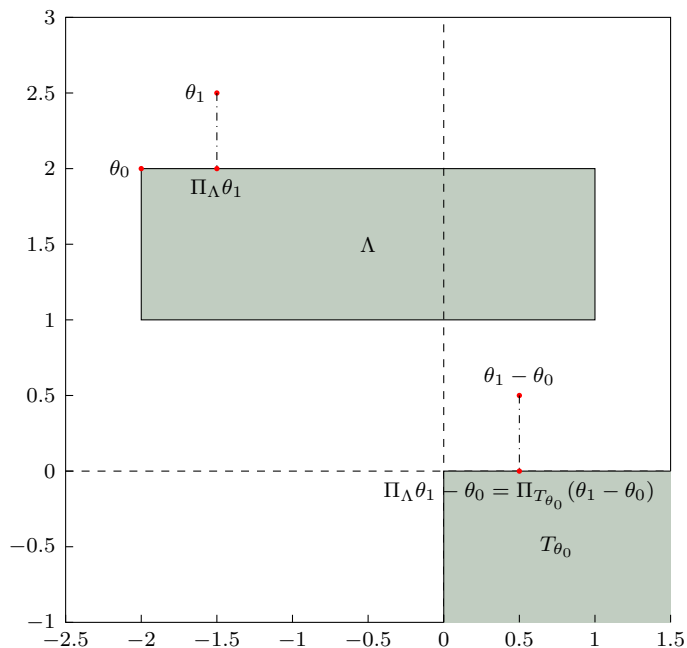
stochastically dominates that of $\|\mathbb{G}_0 - \Pi_{T_{\theta_0}}\mathbb{G}_0\|_{\mathbb{H}}$, and by Proposition 4.1 its quantiles may be employed for potentially conservative inference – an approach that may be viewed as a generalization of assuming all moments are binding in moment inequalities models. It is also worth noting that Proposition 4.1 can be readily extended to study the projection itself rather than its norm, allow for nonconvex sets Λ, and incorporate weight functions into the test statistic; see Remarks 4.1 and 4.2.

Remark 4.1. Zaranotello (1971) and Theorem 2.1 can be employed to derive the asymptotic distribution of the projection $r_n\{\Pi_\Lambda\hat{\theta}_n - \Pi_\Lambda\theta_0\}$ itself. However, when studying the projection, it is perhaps natural to aim to relax the requirement that $\theta_0 \in \Lambda$. Such an extension, as well as considering non-convex Λ, is possible under appropriate regularity conditions – see Shapiro (1994) for the relevant directional differentiability results. ∎

Remark 4.2. While we do not consider it for simplicity, it is straightforward to incorporate weight functions into the test statistic.[13] Formally, a weight function may be seen as a linear operator $A : \mathbb{H} \to \mathbb{H}$, and for any estimator $\hat{A}_n$ such that $\|\hat{A}_n - A\|_o = o_p(1)$ for $\|\cdot\|_o$ the operator norm, we obtain by asymptotic tightness of $r_n\{\hat{\theta}_n - \Pi_\Lambda\hat{\theta}_n\}$ that

$$r_n\|\hat{A}_n\{\hat{\theta}_n - \Pi_\Lambda\hat{\theta}_n\}\|_{\mathbb{H}} \xrightarrow{L} \|A\{\mathbb{G}_0 - \Pi_{T_{\theta_0}}\mathbb{G}_0\}\|_{\mathbb{H}} . \tag{75}$$

Thus, estimating weights has no first order effect on the asymptotic distribution. ∎

---

[13]For instance in (67) we may wish to consider $\{\sum_{i=1}^{d}(E[X^{(i)}])_+^2/Var(X^{(i)})\}^{\frac{1}{2}}$ instead.

4.2.2 Critical Values

In order to construct critical values to conduct inference, we next aim to employ Theorem 3.3, which requires the availability of a suitable estimator $\hat\phi'_n$ for the directional derivative ϕ'_{θ_0}. To this end, we develop an estimator $\hat\phi'_n$ which, despite being computationally intensive, is guaranteed to satisfy Assumption 3.3 under no additional requirements.

Specifically, for an appropriate $\epsilon_n \downarrow 0$, we define $\hat\phi'_n : \mathbb{H} \to \mathbf{R}$ pointwise in $h \in \mathbb{H}$ by

$$\hat\phi'_n(h) \equiv \sup_{\theta \in \Lambda : \|\theta - \Pi_\Lambda \hat\theta_n\|_{\mathbb{H}} \leq \epsilon_n} \|h - \Pi_{T_\theta} h\|_{\mathbb{H}} . \tag{76}$$

Heuristically, we estimate $\phi'_{\theta_0}(h) = \|h - \Pi_{T_{\theta_0}} h\|_{\mathbb{H}}$ by the distance between h and the "least favorable" tangent cone T_θ that can be generated by the $\theta \in \Lambda$ that are in a neighborhood of $\Pi_\Lambda \hat\theta_n$. It is evident from this construction that provided $\epsilon_n \downarrow 0$ at an appropriate rate, the shrinking neighborhood of $\Pi_\Lambda \hat\theta_n$ will include θ_0 with probability tending to one and as a result $\hat\phi'_n(h)$ will provide a potentially conservative estimate of $\phi'_{\theta_0}(h)$. As the following Proposition shows, however, $\hat\phi'_n(h)$ is in fact not conservative, and $\hat\phi'_n$ provides a suitable estimator for ϕ'_{θ_0} in the sense required by Theorem 3.3.

Proposition 4.2. *Let Assumptions 2.2, 4.1 hold, and $\phi'_{\theta_0}(h) \equiv \|h - \Pi_{T_{\theta_0}} h\|_{\mathbb{H}}$. Then,*

(i) *If $\epsilon_n \downarrow 0$ and $\epsilon_n r_n \uparrow \infty$, then $\hat\phi'_n$ as defined in (76) satisfies Assumption 3.3.*

(ii) *$\phi'_{\theta_0} : \mathbb{H} \to \mathbf{R}$ satisfies $\phi'_{\theta_0}(h_1 + h_2) \leq \phi'_{\theta_0}(h_1) + \phi'_{\theta_0}(h_2)$ for all $h_1, h_2 \in \mathbb{H}$.*

The first claim of the Proposition shows that $\hat\phi'_n$ satisfies Assumption 3.3. Therefore, provided the bootstrap is consistent for the asymptotic distribution of $r_n\{\hat\theta_n - \theta_0\}$, Theorem 3.3 implies $\hat\phi'_n$ can be employed to construct critical values. We note that Proposition 4.2(i) holds irrespective of whether the null hypothesis is satisfied, which readily implies the consistency of the corresponding test.[14] In turn, Proposition 4.2(ii) exploits the properties of closed convex cones to show the directional derivative ϕ'_{θ_0} is always subadditive. Thus, one of the key requirement of Theorem 3.4 is satisfied, and we can conclude the proposed test is able to locally control size whenever $\hat\theta_n$ is regular. This latter conclusion of course continues to hold if an alternative estimator to (76) is employed to construct critical values. Hence, we emphasize that while $\hat\phi'_n$ as defined in (76) is appealing due to its general applicability, its use may not be advisable in instances where simpler estimators of ϕ'_{θ_0} are available; see Remark 4.3.

Remark 4.3. In certain applications, the tangent cone T_{θ_0} can be easily estimated and as a result so can ϕ'_{θ_0}. For instance, in the moment inequalities model of Example 4.1,

$$T_{\theta_0} = \{h \in \mathbf{R}^d : h^{(i)} \leq 0 \text{ for all } i \text{ such that } E[X^{(i)}] = 0\} . \tag{77}$$

---

[14]Formally, the law of $\hat\phi'_n(r_n\{\hat\theta_n^* - \hat\theta_n\})$ conditional on the data converges in probability to the law of $\|\mathbb{G}_0 - \Pi_{T_{\theta_0}}\mathbb{G}_0\|_{\mathbb{H}}$ regardless of whether $\theta_0 \in \Lambda$.

For $\bar{X}$ the mean of an i.i.d. sample $\{X_i\}_{i=1}^n$, a natural estimator for T_{θ_0} is then given by

$$\hat{T}_n = \{h \in \mathbf{R}^d : h^{(i)} \leq 0 \text{ for all } i \text{ such that } \bar{X}^{(i)} \geq -\epsilon_n\} \tag{78}$$

for some sequence $\epsilon_n \downarrow 0$ and satisfying $\epsilon_n\sqrt{n} \uparrow \infty$. It is then straightforward to verify that $\hat{\phi}'_n(h) = \|h - \Pi_{\hat{T}_n} h\|_{\mathbb{H}}$ satisfies Assumption 3.3 (compare to (74)) and, more interestingly, that the bootstrap procedure of Theorem 3.3 then reduces to the generalized moment selection approach of Andrews and Soares (2010). ∎

4.3 Simulation Evidence

In order to examine the finite sample performance of the proposed test and illustrate its implementation, we next conduct a limited Monte Carlo study based on Example 4.2. Specifically, we consider a quantile treatment effect model in which the treatment dummy $D \in \{0, 1\}$ satisfies $P(D = 1) = 1/2$, the covariates $Z = (1, Z^{(1)}, Z^{(2)})' \in \mathbf{R}^3$ satisfy $(Z^{(1)}, Z^{(2)})' \sim N(0, I)$ for I the identity matrix, and Y is related by

$$Y = \frac{\Delta}{\sqrt{n}} D \times U + Z'\beta + U , \tag{79}$$

where $\beta = (0, 1/\sqrt{2}, 1/\sqrt{2})'$ and U is unobserved, uniformly distributed on $[0, 1]$, and independent of (D, Z). It is then straightforward to verify that (Y, D, Z) satisfy

$$P(Y \leq D\theta_0(\tau) + Z'\beta(\tau)|D, Z) = \tau , \tag{80}$$

for $\theta_0(\tau) \equiv \tau\Delta/\sqrt{n}$ and $\beta(\tau) \equiv (\tau, 1/\sqrt{2}, 1/\sqrt{2})'$. Hence, in this context the quantile treatment effect has been set local to zero at all τ, which enables us to evaluate the local power and local size control of the proposed test.

We employ the developed framework to study whether the quantile treatment effect $\theta_0(\tau)$ is monotonically increasing in τ, which corresponds to the special case of (62) in which Λ equals the set of monotonically increasing functions. For ease of computation, we obtain quantile regression estimates $\hat{\theta}_n(\tau)$ on a grid $\{0.2, 0.225, \ldots, 0.775, 0.8\}$ and compute the distance of $\hat{\theta}_n$ to the set of monotone functions on this grid as our test statistic. In turn, critical values for this test statistic are obtained by computing two hundred bootstrapped quantile regression coefficients $\hat{\theta}_n^*(\tau)$ at all $\tau \in \{0.2, 0.225, \ldots, 0.775, 0.8\}$, and using the $1-\alpha$ quantile across bootstrap replications of the statistic $\hat{\phi}'_n(\sqrt{n}\{\hat{\theta}_n^* - \hat{\theta}_n\})$, where $\hat{\phi}'_n$ is computed according to (76) with $\epsilon_n = Cn^\kappa$ for different choices of C and κ. All reported results are based on five thousand Monte Carlo replications.

Table 1 reports the empirical rejection rates for different values of the local parameter $\Delta \in \{0, 1, 2\}$ – recall that since $\theta_0(\tau) = \tau\Delta/\sqrt{n}$, the null hypothesis that θ_0 is mono-

Table 1: Empirical Size

			$n = 200$								
Bandwidth		$\alpha = 0.1$			$\alpha = 0.05$			$\alpha = 0.01$			
C	κ	$\Delta = 0$	$\Delta = 1$	$\Delta = 2$	$\Delta = 0$	$\Delta = 1$	$\Delta = 2$	$\Delta = 0$	$\Delta = 1$	$\Delta = 2$	
1	$1/4$	0.042	0.017	0.006	0.020	0.008	0.002	0.005	0.001	0.000	
1	$1/3$	0.042	0.017	0.006	0.020	0.008	0.002	0.005	0.001	0.000	
0.01	$1/4$	0.082	0.053	0.035	0.035	0.023	0.013	0.007	0.002	0.001	
0.01	$1/3$	0.087	0.059	0.042	0.038	0.025	0.015	0.007	0.002	0.001	
Theoretical		0.100	0.042	0.015	0.050	0.018	0.006	0.010	0.003	0.001	

			$n = 500$								
Bandwidth		$\alpha = 0.1$			$\alpha = 0.05$			$\alpha = 0.01$			
C	κ	$\Delta = 0$	$\Delta = 1$	$\Delta = 2$	$\Delta = 0$	$\Delta = 1$	$\Delta = 2$	$\Delta = 0$	$\Delta = 1$	$\Delta = 2$	
1	$1/4$	0.051	0.020	0.007	0.026	0.011	0.002	0.005	0.001	0.000	
1	$1/3$	0.051	0.020	0.007	0.026	0.011	0.002	0.005	0.001	0.000	
0.01	$1/4$	0.096	0.058	0.038	0.047	0.025	0.015	0.009	0.005	0.001	
0.01	$1/3$	0.103	0.065	0.045	0.049	0.030	0.017	0.009	0.005	0.001	
Theoretical		0.100	0.042	0.015	0.050	0.018	0.006	0.010	0.003	0.001	

tonically increasing is satisfied for all such Δ. The bandwidth parameter ϵ_n employed in the construction of the estimator $\hat{\phi}'_n$ is set according $\epsilon_n = Cn^\kappa$ for $C \in \{0.01, 1\}$ and $\kappa \in \{1/4, 1/3\}$. For the explored sample sizes of two and five hundred observations, we observe little sensitivity to the value of κ but a more significant effect of the choice of C. In addition, the row labeled "Theoretical" reports the rejection rates we should expect according to the local asymptotic approximation of Theorem 3.4. Throughout the specifications, we see that the test effectively controls size, and Theorem 3.4 provides an adequate approximation often in between the rejection probabilities obtained from employing $C = 1$ and those corresponding to the more aggressive selection of $C = 0.01$.

In Table 2, we examine the local power of a 5% level test by considering values of $\Delta \in \{-1, \ldots, -8\}$. For such choices of the local parameter, the null hypothesis is violated since $\theta_0(\tau) = \tau\Delta/\sqrt{n}$ is in fact monotonically decreasing in τ (rather than increasing). In this context, we see that the theoretical local power is slightly above the empirical rejection rates, in particular for small values of Δ. These distortions are most severe for n equal to two hundred, though we note a quick improvement in the approximation error when n is set to equal five hundred. Overall, we find the results of the Monte Carlo study encouraging, though certainly limited in their scope.

5   Conclusion

In this paper, we have developed a general asymptotic framework for conducting inference in an important class of irregular models. In analogy with the Delta method, we have shown crucial features of these problems can be understood simply in terms of

Table 2: Local Power of 0.05 Level Test

Bandwidth		$n = 200$							
C	κ	$\Delta = -1$	$\Delta = -2$	$\Delta = -3$	$\Delta = -4$	$\Delta = -5$	$\Delta = -6$	$\Delta = -7$	$\Delta = -8$
1	1/4	0.061	0.155	0.321	0.555	0.782	0.934	0.989	1.000
1	1/3	0.061	0.155	0.321	0.555	0.782	0.934	0.989	1.000
0.01	1/4	0.078	0.172	0.330	0.558	0.783	0.934	0.989	1.000
0.01	1/3	0.081	0.174	0.331	0.559	0.783	0.934	0.989	1.000
Theoretical		0.120	0.245	0.423	0.623	0.796	0.911	0.970	0.992

Bandwidth		$n = 500$							
C	κ	$\Delta = -1$	$\Delta = -2$	$\Delta = -3$	$\Delta = -4$	$\Delta = -5$	$\Delta = -6$	$\Delta = -7$	$\Delta = -8$
1	1/4	0.071	0.181	0.355	0.576	0.789	0.925	0.981	0.997
1	1/3	0.071	0.181	0.355	0.576	0.789	0.925	0.981	0.997
0.01	1/4	0.094	0.201	0.370	0.583	0.791	0.925	0.981	0.997
0.01	1/3	0.098	0.204	0.371	0.585	0.791	0.925	0.981	0.997
Theoretical		0.120	0.245	0.423	0.623	0.796	0.911	0.970	0.992

the asymptotic distribution $\mathbb{G}_0$ and the directional derivative ϕ'_{θ_0}. The utility of these insights were demonstrated by both unifying diverse existing results, and easily studying the otherwise challenging problem of testing for convex set membership. We hope these are just the first applications of this framework, which should be of use to theorists and empirical researchers alike in determining statistical properties such as asymptotic distributions, bootstrap validity, and ability of tests to locally control size.

Appendix A - Proof of Main Results

The following list includes notation and definitions that will be used in the appendix.

$a \lesssim b$	$a \leq Mb$ for some constant M that is universal in the proof.
$\|\cdot\|_{L^q(W)}$	For a random variable W and function f, $\|f\|_{L^q(W)} \equiv \{E[\|f(W)\|^q]\}^{\frac{1}{q}}$.
$\mathcal{C}(\mathbf{A})$	For a set $\mathbf{A}$, $\mathcal{C}(\mathbf{A}) \equiv \{f : \mathbf{A} \to \mathbf{R} : \sup_{a \in A} |f(a)| < \infty$ and f is continuous$\}$.
$\vec{d}_H(\cdot, \cdot, \|\cdot\|)$	For sets A, B, $\vec{d}_H(A, B, \|\cdot\|) \equiv \sup_{a \in A} \inf_{b \in B} \|a - b\|$.
$d_H(\cdot, \cdot, \|\cdot\|)$	For sets A, B, $d_H(A, B, \|\cdot\|) \equiv \max\{\vec{d}_H(A, B, \|\cdot\|), \vec{d}_H(B, A, \|\cdot\|)\}$.
$\ell^\infty(\mathbf{A})$	For a set $\mathbf{A}$, $\ell^\infty(\mathbf{A}) \equiv \{f : \mathbf{A} \to \mathbf{R} : \sup_{a \in A} |f(a)| < \infty\}$.

PROOF OF PROPOSITION 2.1: One direction is clear since, by definition, ϕ being Hadamard differentiable implies that its Hadamard directional derivative exists, equals the Hadamard derivative of ϕ, and hence must be linear.

Conversely suppose the Hadamard directional derivative $\phi'_\theta : \mathbb{D}_0 \to \mathbb{E}$ exists and is linear. Let $\{h_n\}$ and $\{t_n\}$ be sequences such that $h_n \to h \in \mathbb{D}_0$, $t_n \to 0$ and $\theta + t_n h_n \in \mathbb{D}_\phi$ for all n. Then note that from any subsequence $\{t_{n_k}\}$ we can extract a further subsequence $\{t_{n_{k_j}}\}$, such that either: (i) $t_{n_{k_j}} > 0$ for all j or (ii) $t_{n_{k_j}} < 0$ for all j. When (i) holds, ϕ being Hadamard directional differentiable, then immediately yields that:

$$\lim_{j \to \infty} \frac{\phi(\theta + t_{n_{k_j}} h_{n_{k_j}}) - \phi(\theta)}{t_{n_{k_j}}} = \phi'_\theta(h) . \tag{A.1}$$

On the other hand, if (ii) holds, then $h \in \mathbb{D}_0$ and $\mathbb{D}_0$ being a subspace implies $-h \in \mathbb{D}_0$. Therefore, by Hadamard directional differentiability of ϕ and $-t_{n_{k_j}} > 0$ for all j:

$$\lim_{j \to \infty} \frac{\phi(\theta + t_{n_{k_j}} h_{n_{k_j}}) - \phi(\theta)}{t_{n_{k_j}}}$$
$$= -\lim_{j \to \infty} \frac{\phi(\theta + (-t_{n_{k_j}})(-h_{n_{k_j}})) - \phi(\theta)}{-t_{n_{k_j}}} = -\phi'_\theta(-h) = \phi'_\theta(h) , \quad \text{(A.2)}$$

where the final equality holds by the assumed linearity of ϕ'_θ. Thus, results (A.1) and (A.2) imply that every subsequence $\{t_{n_k}, h_{n_k}\}$ has a further subsequence along which

$$\lim_{j \to \infty} \frac{\phi(\theta + t_{n_{k_j}} h_{n_{k_j}}) - \phi(\theta)}{t_{n_{k_j}}} = \phi'_\theta(h) . \tag{A.3}$$

Since the subsequence $\{t_{n_k}, h_{n_k}\}$ is arbitrary, it follows that (A.3) must hold along the original sequence $\{t_n, h_n\}$ and hence ϕ is Hadamard differentiable tangentially to $\mathbb{D}_0$. ∎

PROOF OF THEOREM 2.1: The proof closely follows the proof of Theorem 3.9.4 in van der Vaart and Wellner (1996), and we include it here only for completeness. First,

36

let $\mathbb{D}_n \equiv \{h \in \mathbb{D} : \theta_0 + h/r_n \in \mathbb{D}_\phi\}$ and define $g_n : \mathbb{D}_n \to \mathbb{E}$ to be given by

$$g_n(h_n) \equiv r_n\{\phi(\theta_0 + \frac{h_n}{r_n}) - \phi(\theta_0)\} \tag{A.4}$$

for any $h_n \in \mathbb{D}_n$. Then note that for every sequence $\{h_n\}$ with $h_n \in \mathbb{D}_n$ satisfying $\|h_n - h\|_{\mathbb{D}} = o(1)$ with $h \in \mathbb{D}_0$, it follows from Assumption 2.1(ii) that $\|g_n(h_n) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} = o(1)$. Therefore, the first claim follows by Theorem 1.11.1 in van der Vaart and Wellner (1996) and $\mathbb{G}_0$ being tight implying that it is also separable by Lemma 1.3.2 in van der Vaart and Wellner (1996).

For the second claim of the Theorem, we define $f_n : \mathbb{D}_n \times \mathbb{D} \to \mathbb{E} \times \mathbb{E}$ by:

$$f_n(h_n, h) = (g_n(h_n), \phi'_{\theta_0}(h)) , \tag{A.5}$$

for any $(h_n, h) \in \mathbb{D}_n \times \mathbb{D}$. It then follows by applying Theorem 1.11.1 in van der Vaart and Wellner (1996) again, that as processes in $\mathbb{E} \times \mathbb{E}$ we have:

$$\begin{bmatrix} r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} \\ \phi'_{\theta_0}(r_n\{\hat{\theta}_n - \theta_0\}) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \phi'_{\theta_0}(\mathbb{G}_0) \\ \phi'_{\theta_0}(\mathbb{G}_0) \end{bmatrix} . \tag{A.6}$$

In particular, result (A.6) and the continuous mapping theorem allow us to conclude:

$$r_n\{\phi(\hat{\theta}_n) - \phi(\theta_0)\} - \phi'_{\theta_0}(r_n(\hat{\theta}_n - \theta_0)) \xrightarrow{L} 0 . \tag{A.7}$$

The second claim then follows from (A.7) and Lemma 1.10.2(iii) in van der Vaart and Wellner (1996). ■

PROOF OF COROLLARY 2.1: Follows immediately from Theorem 2.1 applied with $r_n = \sqrt{n}$, $\mathbb{D} = \ell^\infty(\mathcal{F})$ and $\mathbb{D}_0 = \mathcal{C}(\mathcal{F})$, and by noting that $P(\mathbb{G}_0 \in \mathcal{C}(\mathcal{F})) = 1$ by Example 1.5.10 in van der Vaart and Wellner (1996). ■

PROOF OF THEOREM 3.1: In these arguments we need to distinguish between outer and inner expectations, and we therefore employ the notation E^* and E_* respectively. In addition, for notational convenience we let $\mathbb{G}_n \equiv r_n\{\hat{\theta}_n - \theta_0\}$ and $\mathbb{G}_n^* \equiv r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$. To begin, note that Lemma A.2 and the continuous mapping theorem imply that:

$$(r_n\{\hat{\theta}_n^* - \theta_0\}, r_n\{\hat{\theta}_n - \theta_0\})$$
$$= (r_n\{\hat{\theta}_n^* - \hat{\theta}_n\} + r_n\{\hat{\theta}_n - \theta_0\}, r_n\{\hat{\theta}_n - \theta_0\}) \xrightarrow{L} (\mathbb{G}_1 + \mathbb{G}_2, \mathbb{G}_2) \tag{A.8}$$

on $\mathbb{D} \times \mathbb{D}$, where $\mathbb{G}_1$ and $\mathbb{G}_2$ are independent copies of $\mathbb{G}_0$. Further let $\Phi : \mathbb{D}_\phi \times \mathbb{D}_\phi \to \mathbb{E}$ be given by $\Phi(\theta_1, \theta_2) = \phi(\theta_1) - \phi(\theta_2)$ for any $\theta_1, \theta_2 \in \mathbb{D}_\phi \times \mathbb{D}_\phi$. Then observe that Assumption 2.1(ii) implies Φ is Hadamard directionally differentiable at (θ_0, θ_0)

tangentially to $\mathbb{D}_0 \times \mathbb{D}_0$ with derivative $\Phi'_{\theta_0} : \mathbb{D}_0 \times \mathbb{D}_0 \to \mathbb{E}$ given by

$$\Phi'_{\theta_0}(h_1, h_2) = \phi'_{\theta_0}(h_1) - \phi'_{\theta_0}(h_2) \tag{A.9}$$

for any $(h_1, h_2) \in \mathbb{D}_0 \times \mathbb{D}_0$. Thus, by Assumptions 2.2(ii) and 2.3(ii), Theorem 2.1, result (A.8), and $r_n\{\hat{\theta}_n^* - \theta_0\} = \mathbb{G}_n^* + \mathbb{G}_n$ we can conclude that

$$\begin{aligned} r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\} &= r_n\{\Phi(\hat{\theta}_n^*, \hat{\theta}_n) - \Phi(\theta_0, \theta_0)\} \\ &= \Phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n, \mathbb{G}_n) + o_p(1) = \phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n) + o_p(1) \ . \end{aligned} \tag{A.10}$$

Further observe that for any $\epsilon > 0$, it follows from the definition of $\mathrm{BL}_1(\mathbb{E})$ that:

$$\sup_{h \in \mathrm{BL}_1(\mathbb{E})} |E^*[h(r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\}) - h(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n]|$$

$$\leq \epsilon + 2P^*(\|r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\} - \{\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n)\}\|_{\mathbb{E}} > \epsilon|\{X_i\}_{i=1}^n) \tag{A.11}$$

Moreover, Lemma 1.2.6 in van der Vaart and Wellner (1996) and result (A.10) also yield:

$$E^*[P^*(\|r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\} - \{\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n)\}\|_{\mathbb{E}} > \epsilon|\{X_i\}_{i=1}^n)]$$

$$\leq P^*(\|r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\} - \{\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n)\}\|_{\mathbb{E}} > \epsilon) = o(1) \ . \tag{A.12}$$

Therefore, since $\epsilon > 0$ was arbitrary, we obtain from results (A.11) and (A.12) that:

$$\sup_{h \in \mathrm{BL}_1(\mathbb{E})} |E^*[h(r_n\{\phi(\hat{\theta}_n^*) - \phi(\hat{\theta}_n)\})|\{X_i\}_{i=1}^n] - E[h(\phi'_{\theta_0}(\mathbb{G}_0))]|$$

$$= \sup_{h \in \mathrm{BL}_1(\mathbb{E})} |E^*[h(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n] - E[h(\phi'_{\theta_0}(\mathbb{G}_0))]| + o_p(1) \tag{A.13}$$

Thus, in establishing the Theorem, it suffices to study the right hand side of (A.13).

<u>First Claim</u>: We aim to establish that if the bootstrap is consistent, then $\phi'_{\theta_0} : \mathbb{D}_0 \to \mathbb{E}$ must be $\mathbb{G}_0$-translation invariant. Towards this end, note that Lemma A.2 implies

$$(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n), \mathbb{G}_n) \xrightarrow{L} (\phi'_{\theta_0}(\mathbb{G}_1 + \mathbb{G}_2) - \phi'_{\theta_0}(\mathbb{G}_2), \mathbb{G}_2) \tag{A.14}$$

on $\mathbb{E} \times \mathbb{D}$ by the continuous mapping theorem. Let $f \in \mathrm{BL}_1(\mathbb{E})$ and $g \in \mathrm{BL}_1(\mathbb{D})$ satisfy $f(h_1) \geq 0$ and $g(h_2) \geq 0$ for any $h_1 \in \mathbb{E}$ and $h_2 \in \mathbb{D}$. By (A.14) we then have:

$$\lim_{n \to \infty} E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))g(\mathbb{G}_n)] = E[f(\phi'_{\theta_0}(\mathbb{G}_1 + \mathbb{G}_2) - \phi'_{\theta_0}(\mathbb{G}_2))g(\mathbb{G}_2)] \tag{A.15}$$

On the other hand, also note that if the bootstrap is consistent, then result (A.13) yields

$$\sup_{h \in \mathrm{BL}_1(\mathbb{E})} |E^*[h(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n] - E[h(\phi'_{\theta_0}(\mathbb{G}_0))]| = o_p(1) \ . \tag{A.16}$$

Moreover, since $\|g\|_\infty \le 1$ and $\|f\|_\infty \le 1$, it also follows that for any $\epsilon > 0$ we have:

$$\lim_{n\to\infty} E^*[|E^*[f(\phi'_{\theta_0}(\mathbb{G}^*_n + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]|g(\mathbb{G}_n)]$$

$$\le \lim_{n\to\infty} E^*[|E^*[f(\phi'_{\theta_0}(\mathbb{G}^*_n + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]|]$$

$$\le \lim_{n\to\infty} 2P^*(|E^*[f(\phi'_{\theta_0}(\mathbb{G}^*_n + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| > \epsilon) + \epsilon .$$

$$(A.17)$$

Thus, result (A.16), ϵ being arbitrary in (A.17), Lemma A.5(v), $g(h) \ge 0$ for all $h \in \mathbb{D}$, and $\mathbb{G}_n \xrightarrow{L} \mathbb{G}_2$ by result (A.14) allow us to conclude that:

$$\lim_{n\to\infty} E^*[E^*[f(\phi'_{\theta_0}(\mathbb{G}^*_n + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n]g(\mathbb{G}_n)]$$

$$= \lim_{n\to\infty} E^*[E[f(\phi'_{\theta_0}(\mathbb{G}_0))]g(\mathbb{G}_n)] = E[f(\phi'_{\theta_0}(\mathbb{G}_0))]E[g(\mathbb{G}_2)] . \quad (A.18)$$

In addition, we also note that by Lemma 1.2.6 in van der Vaart and Wellner (1996):

$$\lim_{n\to\infty} E_*[f(\phi'_{\theta_0}(\mathbb{G}^*_n + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))g(\mathbb{G}_n)]$$

$$\le \lim_{n\to\infty} E^*[E^*[f(\phi'_{\theta_0}(\mathbb{G}^*_n + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n]g(\mathbb{G}_n)]$$

$$\le \lim_{n\to\infty} E^*[f(\phi'_{\theta_0}(\mathbb{G}^*_n + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))g(\mathbb{G}_n)] \quad (A.19)$$

since $\mathbb{G}_n$ is a function of $\{X_i\}_{i=1}^n$ only and $g(\mathbb{G}_n) \ge 0$. However, by (A.14) and Lemma 1.3.8 in van der Vaart and Wellner (1996), $(\phi'_{\theta_0}(\mathbb{G}^*_n + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n), \mathbb{G}_n)$ is asymptotically measurable, and thus combining results (A.18) and (A.19) we can conclude:

$$\lim_{n\to\infty} E^*[f(\phi'_{\theta_0}(\mathbb{G}^*_n + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))g(\mathbb{G}_n)] = E[f(\phi'_{\theta_0}(\mathbb{G}_0))]E[g(\mathbb{G}_2)] . \quad (A.20)$$

Hence, comparing (A.15) and (A.20) with $g \in \mathrm{BL}_1(\mathbb{D})$ given by $g(a) = 1$ for all $a \in \mathbb{D}$,

$$E[f(\phi'_{\theta_0}(\mathbb{G}_0))]E[g(\mathbb{G}_2)] = E[f(\phi'_{\theta_0}(\mathbb{G}_1 + \mathbb{G}_2) - \phi'_{\theta_0}(\mathbb{G}_2))]E[g(\mathbb{G}_2)]$$

$$= E[f(\phi'_{\theta_0}(\mathbb{G}_1 + \mathbb{G}_2) - \phi'_{\theta_0}(\mathbb{G}_2))g(\mathbb{G}_2)] , \quad (A.21)$$

where the second equality follows again by (A.15) and (A.20). Since (A.21) must hold for any $f \in \mathrm{BL}_1(\mathbb{E})$ and $g \in \mathrm{BL}_1(\mathbb{D})$ with $f(h_1) \ge 0$ and $g(h_2) \ge 0$ for any $h_1 \in \mathbb{E}$ and $h_2 \in \mathbb{D}$, Lemma 1.4.2 in van der Vaart and Wellner (1996) implies $\phi'_{\theta_0}(\mathbb{G}_1 + \mathbb{G}_2) - \phi'_{\theta_0}(\mathbb{G}_2)$ must be independent of $\mathbb{G}_2$, or equivalently, that ϕ'_{θ_0} is $\mathbb{G}_0$-translation invariant.

<u>Second Claim</u>: To conclude, we show that if $\phi'_{\theta_0} : \mathbb{D}_0 \to \mathbb{E}$ is $\mathbb{G}_0$-translation invariant, then the bootstrap is consistent. Fix $\epsilon > 0$, and note that by Assumption 2.2, Lemma A.1, and Lemma 1.3.8 in van der Vaart and Wellner (1996), $\mathbb{G}_n$ and $\mathbb{G}^*_n$ are asymptoti-

cally tight. Therefore, there exists a compact set $K \subset \mathbb{D}$ such that for any $\delta > 0$:

$$\liminf_{n \to \infty} P_*(\mathbb{G}_n^* \in K^\delta) \geq 1 - \epsilon \qquad \liminf_{n \to \infty} P_*(\mathbb{G}_n \in K^\delta) \geq 1 - \epsilon , \qquad \text{(A.22)}$$

where $K^\delta \equiv \{a \in \mathbb{D} : \inf_{b \in K} \|a - b\|_{\mathbb{D}} < \delta\}$. Furthermore, by the Portmanteau Theorem we may assume without loss of generality that K is a subset of the support of $\mathbb{G}_0$ and that $0 \in K$. Next, let $K + K \equiv \{a \in \mathbb{D} : a = b + c \text{ for some } b, c \in K\}$ and note that the compactness of K implies $K + K$ is also compact. Thus, by Lemma A.4 and continuity of $\phi'_{\theta_0} : \mathbb{D} \to \mathbb{E}$, there exist scalars $\delta_0 > 0$ and $\eta_0 > 0$ such that:

$$\sup_{a,b \in (K+K)^{\delta_0} : \|a-b\|_{\mathbb{D}} < \eta_0} \|\phi'_{\theta_0}(a) - \phi'_{\theta_0}(b)\|_{\mathbb{E}} < \epsilon . \qquad \text{(A.23)}$$

Next, for each $a \in K$, let $B_{\eta_0/2}(a) \equiv \{b \in \mathbb{D} : \|a - b\|_{\mathbb{D}} < \eta_0/2\}$. Since $\{B_{\eta_0/2}(a)\}_{a \in K}$ is an open cover of K, there exists a finite collection $\{B_{\eta_0/2}(a_j)\}_{j=1}^J$ also covering K. Therefore, since for any $b \in K^{\frac{\eta_0}{2}}$ there is a $\Pi b \in K$ such that $\|b - \Pi b\|_{\mathbb{D}} < \eta_0/2$, it follows that for every $b \in K^{\frac{\eta_0}{2}}$ there is a $1 \leq j \leq J$ such that $\|b - a_j\|_{\mathbb{D}} < \eta_0$. Setting $\delta_1 \equiv \min\{\delta_0, \eta_0\}/2$, we obtain that if $a \in K^{\delta_1}$ and $b \in K^{\delta_1}$, then: (i) $a + b \in (K + K)^{\delta_0}$ since $K^{\frac{\delta_0}{2}} + K^{\frac{\delta_0}{2}} \subseteq (K + K)^{\delta_0}$, (ii) there is a $1 \leq j \leq J$ such that $\|b - a_j\|_{\mathbb{D}} < \eta_0$, and (iii) $(a + a_j) \in (K + K)^{\delta_0}$ since $a_j \in K$ and $a \in K^{\frac{\delta_0}{2}}$. Therefore, since $0 \in K$, we can conclude from (A.23) that for every $b \in K^{\delta_1}$ there exists a $1 \leq j(b) \leq J$ such that

$$\sup_{a \in K^{\delta_1}} \|\{\phi'_{\theta_0}(a + b) - \phi'_{\theta_0}(b)\} - \{\phi'_{\theta_0}(a + a_{j(b)}) - \phi'_{\theta_0}(a_{j(b)})\}\|_{\mathbb{E}}$$

$$\leq \sup_{a,b \in (K+K)^{\delta_0} : \|a-b\|_{\mathbb{D}} < \eta_0} 2\|\phi'_{\theta_0}(a) - \phi'_{\theta_0}(b)\|_{\mathbb{E}} < 2\epsilon . \qquad \text{(A.24)}$$

In particular, if we define the set $\Delta_n \equiv \{\mathbb{G}_n^* \in K^{\delta_1}, \mathbb{G}_n \in K^{\delta_1}\}$, then (A.24) implies that for every realization of $\mathbb{G}_n$ there is an a_j independent of $\mathbb{G}_n^*$ such that:

$$\sup_{f \in \mathrm{BL}_1(\mathbb{E})} |(f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n)) - f(\phi'_{\theta_0}(\mathbb{G}_n^* + a_j) - \phi'_{\theta_0}(a_j)))1\{\Delta_n\}| < 2\epsilon .$$
$$\text{(A.25)}$$

Letting Δ_n^c denote the complement of Δ_n, result (A.25) then allows us to conclude

$$\sup_{f \in \mathrm{BL}_1(\mathbb{E})} |E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + \mathbb{G}_n) - \phi'_{\theta_0}(\mathbb{G}_n))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| \leq 2P^*(\Delta_n^c|\{X_i\}_{i=1}^n)$$

$$+ \max_{1 \leq j \leq J} \sup_{f \in \mathrm{BL}_1(\mathbb{E})} |E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + a_j) - \phi'_{\theta_0}(a_j))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| + 2\epsilon \qquad \text{(A.26)}$$

since $\|f\|_\infty \leq 1$ for all $f \in \mathrm{BL}_1(\mathbb{E})$. However, by Assumptions 3.1(i)-(ii) and 3.2(ii), and Theorem 10.8 in Kosorok (2008) it follows that for any $1 \leq j \leq J$:

$$\sup_{f \in \mathrm{BL}_1(\mathbb{E})} |E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + a_j) - \phi'_{\theta_0}(a_j))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_j) - \phi'_{\theta_0}(a_j))]| = o_p(1) .$$
(A.27)

Thus, since K is a subset of the support of $\mathbb{G}_0$, Lemma A.3, result (A.27), the continuous mapping theorem, and $J < \infty$ allow us to conclude that:

$$\max_{1 \leq j \leq J} \sup_{f \in \mathrm{BL}_1(\mathbb{E})} |E^*[f(\phi'_{\theta_0}(\mathbb{G}_n^* + a_j) - \phi'_{\theta_0}(a_j))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| = o_p(1) . \quad \text{(A.28)}$$

Moreover, for any $\epsilon \in (0,1)$ we also have by Markov's inequality, Lemma 1.2.6 in van der Vaart and Wellner (1996), $1\{\Delta_n^c\} \leq 1\{\mathbb{G}_n^* \notin K^{\delta_1}\} + 1\{\mathbb{G}_n \notin K^{\delta_1}\}$, and (A.22) that:

$$\limsup_{n \to \infty} P^*(2P^*(\Delta_n^c|\{X_i\}_{i=1}^n) + 2\epsilon > 6\sqrt{\epsilon}) \leq \limsup_{n \to \infty} P^*(P^*(\Delta_n^c|\{X_i\}_{i=1}^n) > 2\sqrt{\epsilon})$$

$$\leq \frac{1}{2\sqrt{\epsilon}} \times \limsup_{n \to \infty} \{P^*(\mathbb{G}_n \notin K^{\delta_1}) + P^*(\mathbb{G}_n^* \notin K^{\delta_1})\} \leq \sqrt{\epsilon} . \quad \text{(A.29)}$$

Since $\epsilon > 0$ was arbitrary, combining (A.13), (A.26), (A.28), and (A.29) imply (36) holds, or equivalently, that ϕ'_{θ_0} being $\mathbb{G}_0$-translation invariant implies bootstrap consistency. $\blacksquare$

PROOF OF THEOREM 3.2: Let P denote the distribution of $\mathbb{G}_0$ on $\mathbb{D}_0$, and note that by Assumption 2.2(ii) and Lemma A.7 we may assume without loss of generality that the support of $\mathbb{G}_0$ equals $\mathbb{D}$ and that $\mathbb{D}$ is separable. Further note that if $\mathbb{G}_1$ is an independent copy of $\mathbb{G}_0$ and $\phi'_{\theta_0} : \mathbb{D} \to \mathbb{E}$ is linear, then we immediately obtain that:

$$\phi'_{\theta_0}(\mathbb{G}_1 + \mathbb{G}_0) - \phi'_{\theta_0}(\mathbb{G}_0) = \phi'_{\theta_0}(\mathbb{G}_1) , \quad \text{(A.30)}$$

which is independent of $\mathbb{G}_0$, and hence ϕ'_{θ_0} is trivially $\mathbb{G}_0$-translation invariant.

The opposite direction is more challenging and requires us to introduce additional notation which closely follows Chapter 7 in Davydov et al. (1998). First, let $\mathbb{D}^*$ denote the dual space of $\mathbb{D}$, and $\langle d, d^* \rangle_{\mathbb{D}} = d^*(d)$ for any $d \in \mathbb{D}$ and $d^* \in \mathbb{D}^*$. Similarly denote the dual space of $\mathbb{E}$ by $\mathbb{E}^*$ and corresponding bilinear form by $\langle \cdot, \cdot \rangle_{\mathbb{E}}$. Further let

$$\mathbb{D}'_P \equiv \left\{ d' : \mathbb{D} \to \mathbf{R} : d' \text{ is linear, Borel-measurable, and } \int_{\mathbb{D}} (d'(d))^2 dP(d) < \infty \right\} ,$$
(A.31)

and with some abuse of notation also write $d'(d) = \langle d', d \rangle_{\mathbb{D}}$ for any $d' \in \mathbb{D}'_P$ and $d \in \mathbb{D}$. Finally, for each $h \in \mathbb{D}$ we let P^h denote the law of $\mathbb{G}_0 + h$, write $P^h \ll P$ whenever P^h is absolutely continuous with respect to P, and define the set:

$$\mathbb{H}_P \equiv \{h \in \mathbb{D} : P^{rh} \ll P \text{ for all } r \in \mathbf{R}\} . \quad \text{(A.32)}$$

To proceed, note that since $\mathbb{D}$ is separable, the Borel σ-algebra, the σ-algebra generated by the weak topology, and the cylindrical σ-algebra all coincide (Ledoux and Talagrand, 1991, p. 38). Furthermore, by Theorem 7.1.7 in Bogachev (2007), P is Radon with respect to the Borel σ-algebra, and hence also with respect to the cylindrical σ-algebra. Hence, by Theorem 7.1 in Davydov et al. (1998), it follows that there exists a linear map $I : \mathbb{H}_P \to \mathbb{D}'_P$ satisfying for every $h \in \mathbb{H}_P$:

$$\frac{dP^h}{dP}(d) = \exp\left\{\langle d, Ih\rangle_{\mathbb{D}} - \frac{1}{2}\sigma^2(h)\right\} \qquad \sigma^2(h) \equiv \int_{\mathbb{D}} \langle d, Ih\rangle_{\mathbb{D}}^2 dP(d) . \qquad \text{(A.33)}$$

Next, fix an arbitrary $e^* \in \mathbb{E}^*$ and $h \in \mathbb{H}_P$. Then note that Lemma A.3 and Lemma 1.3.12 in van der Vaart and Wellner (1996) imply $\langle e^*, \phi'_{\theta_0}(\mathbb{G}_0 + rh) - \phi'_{\theta_0}(rh)\rangle_{\mathbb{E}}$ and $\langle e^*, \phi'_{\theta_0}(\mathbb{G}_0)\rangle_{\mathbb{E}}$ must be equal in distribution for all $r \in \mathbf{R}$. In particular, their characteristic functions must equal each other, and hence for all $r \geq 0$ and $t \in \mathbf{R}$:

$$E[\exp\{it\langle e^*, \phi'_{\theta_0}(\mathbb{G}_0)\rangle_{\mathbb{E}}\}] = E[\exp\{it\{\langle e^*, \phi'_{\theta_0}(\mathbb{G}_0 + rh) - \phi'_{\theta_0}(rh)\rangle_{\mathbb{E}}\}\}]$$
$$= \exp\{-itr\langle e^*, \phi'_{\theta_0}(h)\rangle_{\mathbb{E}}\}E[\exp\{it\langle e^*, \phi'_{\theta_0}(\mathbb{G}_0 + rh)\rangle_{\mathbb{E}}\}] , \quad \text{(A.34)}$$

where in the second equality we have exploited that $\phi'_{\theta_0}(rh) = r\phi'_{\theta_0}(h)$ due to ϕ'_{θ_0} being positively homogenous of degree one. Setting $C(t) \equiv E[\exp\{it\langle e^*, \phi'_{\theta_0}(\mathbb{G}_0)\rangle_{\mathbb{E}}\}]$ and exploiting result (A.34) we can then obtain by direct calculation that for all $t \in \mathbf{R}$

$$itC(t) \times \langle e^*, \phi'_{\theta_0}(h)\rangle_{\mathbb{E}} = \lim_{r\downarrow 0} \frac{1}{r}\{E[\exp\{it\langle e^*, \phi'_{\theta_0}(\mathbb{G}_0 + rh)\rangle_{\mathbb{E}}\}] - C(t)\}$$
$$= \lim_{r\downarrow 0} \frac{1}{r} \int_{\mathbb{D}} \left\{\exp\left\{it\langle e^*, \phi'_{\theta_0}(d)\rangle_{\mathbb{E}} + r\langle d, Ih\rangle_{\mathbb{D}} - \frac{r^2}{2}\sigma^2(h)\right\} - C(t)\right\}dP(d) , \quad \text{(A.35)}$$

where in the second equality we exploited result (A.33), linearity of $I : \mathbb{H}_P \to \mathbb{D}'_P$ and that $h \in \mathbb{H}_P$ implies $rh \in \mathbb{H}_P$ for all $r \in \mathbf{R}$. Furthermore, by the mean value theorem

$$\sup_{r\in(0,1]}\frac{1}{r}\left|\exp\left\{it\langle e^*, \phi'_{\theta_0}(d)\rangle_{\mathbb{E}} + r\langle d, Ih\rangle_{\mathbb{D}} - \frac{r^2}{2}\sigma^2(h)\right\} - \exp\{it\langle e^*, \phi'_{\theta_0}(d)\rangle_{\mathbb{E}}\}\right|$$
$$\leq \sup_{r\in(0,1]}\left|\exp\left\{it\langle e^*, \phi'_{\theta_0}(d)\rangle_{\mathbb{E}} + r\langle d, Ih\rangle_{\mathbb{D}} - \frac{r^2}{2}\sigma^2(h)\right\} \times \{\langle d, Ih\rangle_{\mathbb{D}} - r\sigma^2(h)\}\right|$$
$$\leq \exp\{|\langle d, Ih\rangle_{\mathbb{D}}|\} \times \{|\langle d, Ih\rangle_{\mathbb{D}}| + \sigma^2(h)\} , \qquad \text{(A.36)}$$

where the final inequality follows from $\sigma^2(h) > 0$ and $|\exp\{it\langle e^*, \phi'_{\theta_0}(d)\rangle_{\mathbb{E}}\}| \leq 1$. Moreover, by Proposition 2.10.3 in Bogachev (1998) and $Ih \in \mathbb{D}'_P$, it follows that

$\langle \mathbb{G}_0, Ih \rangle_{\mathbb{D}} \sim N(0, \sigma^2(h))$. Thus, we can obtain by direct calculation:

$$\int_{\mathbb{D}} \exp\{|\langle d, Ih \rangle_{\mathbb{D}}|\} \times \{|\langle d, Ih \rangle_{\mathbb{D}}| + \sigma^2(h)\} dP(d)$$

$$= \int_{\mathbf{R}} \frac{\{|u| + \sigma^2(h)\}}{\sigma(h)\sqrt{2\pi}} \times \exp\left\{|u| - \frac{u^2}{2\sigma^2(h)}\right\} du < \infty . \quad \text{(A.37)}$$

Hence, results (A.36) and (A.37) justify the use of the dominated convergence theorem in (A.35). Also note that $t \mapsto C(t)$ is the characteristic function of $\langle e^*, \phi'_{\theta_0}(\mathbb{G}_0) \rangle_{\mathbb{E}}$ and hence it is continuous. Thus, since $C(0) = 1$ there exists a $t_0 > 0$ such that $C(t_0)t_0 \neq 0$. For such t_0 we then finally conclude from the above results that

$$\langle e^*, \phi'_{\theta_0}(h) \rangle_{\mathbb{E}} = -\frac{iE[\exp\{it\langle e^*, \phi'_{\theta_0}(\mathbb{G}_0) \rangle_{\mathbb{E}}\} \langle \mathbb{G}_0, Ih \rangle_{\mathbb{D}}]}{t_0 C(t_0)} . \quad \text{(A.38)}$$

To conclude note that $\mathbb{H}_P$ being a vector space (Davydov et al., 1998, p. 38) and $I : \mathbb{D} \to \mathbb{D}'_P$ being linear imply together with result (A.38) that $h \mapsto \langle e^*, \phi'_{\theta_0}(h) \rangle_{\mathbb{E}}$ is linear on $\mathbb{H}_P$. Moreover, note that $h \mapsto \langle e^*, \phi'_{\theta_0}(h) \rangle_{\mathbb{E}}$ is also continuous on $\mathbb{D}$ due to continuity of ϕ'_{θ_0} and having $e^* \in \mathbb{E}^*$. Hence, since $\mathbb{H}_P$ is dense in $\mathbb{D}$ by Proposition 7.4(ii) in Davydov et al. (1998) we can conclude that $\langle e^*, \phi'_{\theta_0}(\cdot) \rangle_{\mathbb{E}} : \mathbb{D} \to \mathbf{R}$ is linear and continuous. Since this result holds for all $e^* \in \mathbb{E}^*$, Lemma A.2 in van der Vaart (1991) implies $\phi'_{\theta_0} : \mathbb{D} \to \mathbb{E}$ must be linear and continuous, which establishes the Theorem. ∎

PROOF OF COROLLARY 3.1: By Theorems 3.1 and 3.2 the bootstrap is consistent if and only if ϕ'_{θ_0} is linear. However, since $\mathbb{G}_0$ is Gaussian and $\phi'_{\theta_0} : \mathbb{D}_0 \to \mathbb{E}$ is continuous, Lemma 2.2.2 in Bogachev (1998) implies $\phi'_{\theta_0}(\mathbb{G}_0)$ must be Gaussian (on $\mathbb{E}$) whenever ϕ'_{θ_0} is linear, and hence the claim of the Corollary follows. ∎

PROOF OF THEOREM 3.3: Fix arbitrary $\epsilon > 0$, $\eta > 0$ and for notational convenience let $\mathbb{G}_n^* \equiv r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$. By Assumption 2.2(ii) there is a compact set $K_0 \subseteq \mathbb{D}_0$ such that

$$P(\mathbb{G}_0 \notin K_0) < \frac{\epsilon\eta}{2} . \quad \text{(A.39)}$$

Thus, by Lemma A.1 and the Portmanteau Theorem, we conclude that for any $\delta > 0$

$$\limsup_{n \to \infty} P(\mathbb{G}_n^* \notin K_0^\delta) \leq P(\mathbb{G}_0 \notin K_0^\delta) \leq P(\mathbb{G}_0 \notin K_0) < \frac{\epsilon\eta}{2} . \quad \text{(A.40)}$$

On the other hand, since K_0 is compact, Assumption 3.3 yields that for some $\delta_0 > 0$:

$$\limsup_{n \to \infty} P(\sup_{h \in K_0^{\delta_0}} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} > \epsilon) < \eta . \quad \text{(A.41)}$$

Next, note that Lemma 1.2.2(iii) in van der Vaart and Wellner (1996), $h \in \mathrm{BL}_1(\mathbb{E})$ being bounded by one and satisfying $|h(e_1) - h(e_2)| \leq \|e_1 - e_2\|_{\mathbb{E}}$ for all $e_1, e_2 \in \mathbb{E}$, imply:

$$\sup_{f \in \mathrm{BL}_1(\mathbb{E})} |E[f(\hat{\phi}'_n(\mathbb{G}^*_n))|\{X_i\}] - E[f(\phi'_{\theta_0}(\mathbb{G}^*_n))|\{X_i\}]|$$

$$\leq \sup_{f \in \mathrm{BL}_1(\mathbb{E})} E[|f(\hat{\phi}'_n(\mathbb{G}^*_n)) - f(\phi'_{\theta_0}(\mathbb{G}^*_n))||\{X_i\}]$$

$$\leq E[2 \times 1\{\mathbb{G}^*_n \notin K_0^{\delta_0}\} + \sup_{f \in K_0^{\delta_0}} \|\hat{\phi}'_n(f) - \phi'_{\theta_0}(f)\|_{\mathbb{E}}|\{X_i\}]$$

$$\leq 2P(\mathbb{G}^*_n \notin K_0^{\delta_0}|\{X_i\}_{i=1}^n) + \sup_{f \in K_0^{\delta_0}} \|\hat{\phi}'_n(f) - \phi'_{\theta_0}(f)\|_{\mathbb{E}} , \quad \text{(A.42)}$$

where in the final inequality we exploited Lemma 1.2.2(i) in van der Vaart and Wellner (1996) and $\hat{\phi}'_n : \mathbb{D} \to \mathbb{E}$ depending only on $\{X_i\}_{i=1}^n$. Furthermore, Markov's inequality, Lemma 1.2.7 in van der Vaart and Wellner (1996), and result (A.40) yield:

$$\limsup_{n \to \infty} P(P(\mathbb{G}^*_n \notin K_0^{\delta_0}|\{X_i\}_{i=1}^n) > \epsilon) \leq \limsup_{n \to \infty} P(\mathbb{G}^*_n \notin K_0^{\delta_0}) < \eta . \quad \text{(A.43)}$$

Next, also note that Assumption 3.1(i) and Theorem 10.8 in Kosorok (2008) imply that:

$$\sup_{f \in \mathrm{BL}_1(\mathbb{E})} |E[f(\phi'_{\theta_0}(\mathbb{G}^*_n))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| = o_p(1) . \quad \text{(A.44)}$$

Thus, by combining results (A.41), (A.42), (A.43) and (A.44) we can finally conclude:

$$\limsup_{n \to \infty} P(\sup_{f \in \mathrm{BL}_1(\mathbb{E})} |E[f(\hat{\phi}'_n(\mathbb{G}^*_n))|\{X_i\}_{i=1}^n] - E[f(\phi'_{\theta_0}(\mathbb{G}_0))]| > 3\epsilon) < 3\eta . \quad \text{(A.45)}$$

Since ϵ and η were arbitrary, the claim of the Theorem then follows from (A.45). ∎

PROOF OF COROLLARY 3.2: Let F denote the cdf of $\phi'_{\theta_0}(\mathbb{G}_0)$, and similarly define:

$$\hat{F}_n(c) \equiv P(\hat{\phi}'_n(r_n\{\hat{\theta}^*_n - \hat{\theta}_n\}) \leq c|\{X_i\}_{i=1}^n) . \quad \text{(A.46)}$$

Next, observe that Theorem 3.3 and Lemma 10.11 in Kosorok (2008) imply that:

$$\hat{F}_n(c) = F(c) + o_p(1) , \quad \text{(A.47)}$$

for all $c \in \mathbf{R}$ that are continuity points of F. Fix $\epsilon > 0$, and note that since F is strictly increasing at $c_{1-\alpha}$ and the set of continuity of points of F is dense in $\mathbf{R}$, it follows that there exist points $c_1, c_2 \in \mathbf{R}$ such that: (i) $c_1 < c_{1-\alpha} < c_2$, (ii) $|c_1 - c_{1-\alpha}| < \epsilon$ and $|c_2 - c_{1-\alpha}| < \epsilon$, (iii) c_1 and c_2 are continuity points of F, and (iv) $F(c_1) + \delta < 1 - \alpha <$

$F(c_2) - \delta$ for some $\delta > 0$. We can then conclude that:

$$\limsup_{n \to \infty} P(|\hat{c}_{1-\alpha} - c_{1-\alpha}| > \epsilon)$$
$$\leq \limsup_{n \to \infty} \{P(|\hat{F}_n(c_1) - F(c_1)| > \delta) + P(|\hat{F}_n(c_2) - F(c_2)| > \delta)\} = 0 , \quad (A.48)$$

due to (A.47). Since $\epsilon > 0$ was arbitrary, the Corollary then follows. $\blacksquare$

PROOF OF LEMMA 3.1: First note that by Assumption 3.5(ii) we can conclude:

$$\lim_{n \to \infty} \|\sqrt{n}\{\theta(P_n) - \theta(P)\} - \eta\theta'(\wp)\|_{\mathbb{D}} = 0 . \quad (A.49)$$

Similarly, letting $t_n \equiv n^{-\frac{1}{2}}$, $h_n \equiv \sqrt{n}\{\theta(P_n) - \theta(P)\}$ we note $\theta(P) + t_n h_n = \theta(P_n) \in \mathbb{D}_\phi$, and by (A.49) that $\|h_n - h\|_{\mathbb{D}} = o(1)$ as $n \to \infty$, for $h \equiv \eta\theta'(\wp)$. Further note that $\eta\theta'(\wp) \in \mathbb{D}_0$ by Assumption 3.5(ii) since $\eta\theta'(\wp) = \theta'(\tilde{\wp})$ for the curve $t \mapsto \tilde{\wp}_t \equiv t \mapsto \wp_{\eta t}$. Thus, from Assumption 2.1(ii) and Definition 2.1 we can conclude that

$$\lim_{n \to \infty} \|\sqrt{n}\{\phi(\theta(P_n)) - \phi(\theta(P))\} - \phi'_{\theta_0}(\eta\theta'(\wp))\|_{\mathbb{E}}$$
$$= \lim_{n \to \infty} \|\frac{\phi(\theta(P) + t_n h_n) - \phi(\theta(P))}{t_n} - \phi'_{\theta_0}(h)\|_{\mathbb{E}} = 0 . \quad (A.50)$$

Next, let $P^n \equiv \bigotimes_{i=1}^n P$ and $P_n^n \equiv \bigotimes_{i=1}^n P_n$. By Theorem 2.1 we then have that:

$$\sqrt{n}\{\phi(\hat{\theta}_n) - \phi(\theta(P))\} = \phi'_{\theta_0}(\sqrt{n}\{\hat{\theta}_n - \theta(P)\}) + o_p(1) \quad (A.51)$$

under P^n. However, by Theorem 12.2.3 and Corollary 12.3.1 in Lehmann and Romano (2005), P_n^n and P^n are mutually contiguous. Hence, from (A.50) and (A.51) we obtain

$$\sqrt{n}\{\phi(\hat{\theta}_n) - \phi(\theta(P_n))\} = \sqrt{n}\{\phi(\hat{\theta}_n) - \phi(\theta(P))\} - \sqrt{n}\{\phi(\theta(P_n)) - \phi(\theta(P))\}$$
$$= \phi'_{\theta_0}(\sqrt{n}\{\hat{\theta}_n - \theta(P)\}) - \phi'_{\theta_0}(\eta\theta'(\wp)) + o_p(1) . \quad (A.52)$$

under P_n^n. Furthermore, by regularity of $\hat{\theta}_n$ and result (A.49) we also have that:

$$\sqrt{n}\{\hat{\theta}_n - \theta(P)\} = \sqrt{n}\{\hat{\theta}_n - \theta(P_n)\} + \sqrt{n}\{\theta(P_n) - \theta(P)\} \xrightarrow{L_n} \mathbb{G}_0 + \eta\theta'(\wp) . \quad (A.53)$$

Thus, the Lemma follows from (A.52), (A.53) and the continuous mapping theorem. $\blacksquare$

PROOF OF COROLLARY 3.3: Let $\mathbb{D}_L$ denote the support of $\mathbb{G}_0$, and note that if $\tilde{\wp}_t = P$ for all t, then $\tilde{\wp}$ is trivially a curve in $\mathbf{P}$ with $\tilde{\wp}' = 0 \in \mathbb{D}$, and hence $0 \in \overline{\bigcup_\wp \theta'(\wp)} = \mathbb{D}_L$. We first show that $\mathbb{G}_0$-translation invariance implies $\phi(\hat{\theta}_n)$ is regular. To this end, note $0 \in \mathbb{D}_L$, Lemma A.3, and $\overline{\bigcup_\wp \theta'(\wp)} = \mathbb{D}_L$ implies for any $\eta \in \mathbf{R}$ and curve $\wp$ in $\mathbf{P}$:

$$E[h(\phi'_{\theta_0}(\mathbb{G}_0 + \eta\theta'(\wp)) - \phi'_{\theta_0}(\eta\theta'(\wp)))] = E[h(\phi'_{\theta_0}(\mathbb{G}_0))] \quad (A.54)$$

for all bounded and continuous $h : \mathbb{E} \to \mathbf{R}$. Letting "$\overset{d}{=}$" denote equality in distribution, we then conclude from (A.54) and Lemma 1.3.12 in van der Vaart and Wellner (1996):

$$\phi'_{\theta_0}(\mathbb{G}_0 + \eta\theta'(\wp)) - \phi'_{\theta_0}(\eta\theta'(\wp)) \overset{d}{=} \phi'_{\theta_0}(\mathbb{G}_0) \tag{A.55}$$

for any $\eta \in \mathbf{R}$ and curve $\wp$ in $\mathbf{P}$. Thus, result (A.55) and Lemma 3.1 imply $\phi(\hat{\theta}_n)$ is a regular estimator for $\phi(\theta_0(P))$ establishing the first direction of the Corollary.

For the opposite direction, suppose now that $\phi(\hat{\theta}_n)$ is a regular estimator of $\phi(\theta(P))$. For notational simplicity, further let $\Phi : \mathbb{D} \times \mathbb{D} \to \mathbb{E}$ be given by:

$$\Phi(h_0, h_1) \equiv \phi'_{\theta_0}(h_0 + h_1) - \phi'_{\theta_0}(h_0) . \tag{A.56}$$

Next, fix arbitrary continuous and bounded functions $f : \mathbb{E} \to \mathbf{R}$ and $g : \mathbb{D} \to \mathbf{R}$, and let $\mathbb{G}_1$ be an independent copy of $\mathbb{G}_0$. Then note that: (i) Continuity of $\phi'_{\theta_0} : \mathbb{D} \to \mathbb{E}$ implies $h_0 \mapsto \Phi(h_0, h_1)$ is continuous for any $h_1 \in \mathbb{D}$, and (ii) $\bigcup_\wp \theta'(\wp)$ being dense in $\mathbb{D}_L$ implies that for any $h_0 \in \mathbb{D}_L$ there is a sequence $h_{0,n} \in \bigcup_\wp \theta'(\wp)$ such that $\|h_0 - h_{0,n}\|_{\mathbb{D}} = o(1)$ as $n \to \infty$. Therefore, the dominated convergence theorem yields

$$E[f(\Phi(h_0, \mathbb{G}_1))] = \lim_{n \to \infty} E[f(\Phi(h_{0,n}, \mathbb{G}_1))] = E[f(\Phi(0, \mathbb{G}_1))] = E[f(\phi'_{\theta_0}(\mathbb{G}_1))] , \tag{A.57}$$

where the second equality follows from $0 \in \bigcup_\wp \theta'(\wp)$ together with Lemma 3.1 and $\phi(\hat{\theta}_n)$ being regular implying the distribution of $\Phi(h_0, \mathbb{G}_1)$ is constant in $h_0 \in \bigcup_\wp \theta'(\wp)$, while the last equality results from (A.56) and $\phi'_{\theta_0}(0) = 0$. Hence, Fubini's theorem, result (A.57) and $\mathbb{G}_0$ and $\mathbb{G}_1$ being independent allow us to conclude that:

$$E[f(\phi'_{\theta_0}(\mathbb{G}_0 + \mathbb{G}_1) - \phi'_{\theta_0}(\mathbb{G}_0))g(\mathbb{G}_0)]$$
$$= \int_{\mathbb{D}_L} E[f(\Phi(h_0, \mathbb{G}_1))]g(h_0)dP(h_0) = E[f(\phi'_{\theta_0}(\mathbb{G}_1))]E[g(\mathbb{G}_0)] \tag{A.58}$$

where with some abuse of notation we let P also denote the distribution of $\mathbb{G}_0$ on $\mathbb{D}$. Since (A.58) holds for any bounded and continuous $f : \mathbb{E} \to \mathbf{R}$ and $g : \mathbb{D} \to \mathbf{R}$, Lemma 1.4.2 in van der Vaart and Wellner (1996) implies $\phi'_{\theta_0}(\mathbb{G}_0 + \mathbb{G}_1) - \phi'_{\theta_0}(\mathbb{G}_0)$ and $\mathbb{G}_1$ are independent, or equivalently, that ϕ'_{θ_0} is $\mathbb{G}_0$-translation invariant. ∎

Proof of Theorem 3.4: Recall that we have set $P_n \equiv \wp_{\eta/\sqrt{n}}$, $P_n^n \equiv \bigotimes_{i=1}^n P_n$, and similarly define $P^n \equiv \bigotimes_{i=1}^n P$. Then note that by Theorem 12.2.3 and Corollary 12.3.1 in Lehmann and Romano (2005), P_n^n and P^n are mutually contiguous. Therefore, since by Corollary 3.2 $\hat{c}_{1-\alpha} \overset{p}{\to} c_{1-\alpha}$ under P^n, it follows that we also have:

$$\hat{c}_{1-\alpha} = c_{1-\alpha} + o_p(1) \quad \text{under } P_n^n . \tag{A.59}$$

Moreover, since $\phi(\theta(P)) = 0$, we also obtain from result (A.51) that under P_n^n we have:

$$
\begin{aligned}
\sqrt{n}\phi(\hat{\theta}_n) &= \sqrt{n}\{\phi(\hat{\theta}_n) - \phi(\theta(P))\} \\
&= \phi'_{\theta_0}(\sqrt{n}\{\hat{\theta}_n - \theta(P)\}) + o_p(1) \overset{L_n}{\to} \phi'_{\theta_0}(\mathbb{G}_0 + \eta\theta'(\wp)) ,\quad \text{(A.60)}
\end{aligned}
$$

where the final result holds for L_n denoting law under P_n^n by result (A.53) and the continuous mapping theorem. Thus, (58) holds by (A.60) and the Portmanteau Theorem.

In order to establish (59) holds whenever $\eta \leq 0$, first note that (A.50) implies

$$
0 \geq \lim_{n \to \infty} \sqrt{n}\{\phi(\theta(P_n)) - \phi(\theta(P))\} = \phi'_{\theta_0}(\eta\theta'(\wp)) ,\quad \text{(A.61)}
$$

where we have exploited that $\phi(\theta(P)) = 0$ and $\phi(\theta(P_n)) \leq 0$ for all $\eta \leq 0$. Therefore, result (A.59) together with the second equality in (A.60) allow us to conclude

$$
\begin{aligned}
\limsup_{n \to \infty} P_n^n(\sqrt{n}\phi(\hat{\theta}_n) &> \hat{c}_{1-\alpha}) \\
&\leq \limsup_{n \to \infty} P_n^n(\phi'_{\theta_0}(\sqrt{n}\{\hat{\theta}_n - \theta(P)\}) \geq c_{1-\alpha}) \\
&\leq \limsup_{n \to \infty} P_n^n(\phi'_{\theta_0}(\sqrt{n}\{\hat{\theta}_n - \theta(P_n)\}) + \phi'_{\theta_0}(\sqrt{n}\{\theta(P_n) - \theta(P)\}) \geq c_{1-\alpha}) \\
&\leq \limsup_{n \to \infty} P_n^n(\phi'_{\theta_0}(\sqrt{n}\{\hat{\theta}_n - \theta(P_n)\}) \geq c_{1-\alpha}) \\
&= P_n^n(\phi'_{\theta_0}(\mathbb{G}_0) \geq c_{1-\alpha}) ,\quad \text{(A.62)}
\end{aligned}
$$

where the second inequality follows from subadditivity of ϕ'_{θ_0}, the third inequality is implied by (A.61), and the final result follows from $\sqrt{n}\{\hat{\theta}_n - \theta(P_n)\} \overset{L_n}{\to} \mathbb{G}_0$ by Assumption 3.5(i), the continuous mapping theorem, and $c_{1-\alpha}$ being a continuity point of the cdf of $\phi'_{\theta_0}(\mathbb{G}_0)$. Since $P(\phi'_{\theta_0}(\mathbb{G}_0) \geq c_{1-\alpha}) = \alpha$ by construction, result (59) follows. $\blacksquare$

Lemma A.1. *If Assumptions 2.1(i), 2.2(ii), 3.1, 3.2(i) hold, then $r_n\{\hat{\theta}_n^* - \hat{\theta}_n\} \overset{L}{\to} \mathbb{G}_0$.*

PROOF: In these arguments we need to distinguish between outer and inner expectations, and we therefore employ the notation E^* and E_* respectively. For notational simplicity also let $\mathbb{G}_n^* \equiv r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$. First, let $f \in \mathrm{BL}_1(\mathbb{D})$, and then note that by Lemma A.5(i) and Lemma 1.2.6 in van der Vaart and Wellner (1996) we have that:

$$
\begin{aligned}
E^*[f(\mathbb{G}_n^*)] - E[f(\mathbb{G}_0)] &\geq E^*[E^*[f(\mathbb{G}_n^*)|\{X_i\}_{i=1}^n]] - E[f(\mathbb{G}_0)] \\
&\geq -E^*[|E^*[f(\mathbb{G}_n^*)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|] \\
&\geq -E^*[\sup_{f \in \mathrm{BL}_1(\mathbb{D})} |E^*[f(\mathbb{G}_n^*)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|] .\quad \text{(A.63)}
\end{aligned}
$$

Similarly, applying Lemma 1.2.6 in van der Vaart and Wellner (1996) once again together with Lemma A.5(ii), and exploiting that $f \in \mathrm{BL}_1(\mathbb{D})$ we can conclude that:

$$
\begin{aligned}
E_*[f(\mathbb{G}_n^*)] - E[f(\mathbb{G}_0)] &\leq E_*[E^*[f(\mathbb{G}_n^*)|\{X_i\}_{i=1}^n]] - E[f(\mathbb{G}_0)] \\
&\leq E^*[|E^*[f(\mathbb{G}_n^*)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|] \\
&\leq E^*[\sup_{f \in \mathrm{BL}_1(\mathbb{D})} |E^*[f(\mathbb{G}_n^*)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|] . \quad (A.64)
\end{aligned}
$$

However, since $\|f\|_\infty \leq 1$ for all $f \in \mathrm{BL}_1(\mathbb{D})$, it also follows that for any $\eta > 0$ we have:

$$
\begin{aligned}
E^*[\sup_{f \in \mathrm{BL}_1(\mathbb{D})} |E^*[f(\mathbb{G}_n^*)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|] \\
\leq 2P^*(\sup_{f \in \mathrm{BL}_1(\mathbb{D})} |E^*[f(\mathbb{G}_n^*)|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]| > \eta) + \eta . \quad (A.65)
\end{aligned}
$$

Moreover, by Assumption 3.2(i), $E^*[f(\mathbb{G}_n^*)] = E_*[f(\mathbb{G}_n^*)] + o(1)$. Thus, Assumption 3.1(ii), η being arbitrary, and results (A.63) and (A.64) together imply that:

$$
\lim_{n \to \infty} E^*[f(\mathbb{G}_n^*)] = E[f(\mathbb{G}_0)] \quad (A.66)
$$

for any $f \in \mathrm{BL}_1(\mathbb{D})$. Further note that since $\mathbb{G}_0$ is tight by Assumption 2.2(ii) and $\mathbb{D}$ is a Banach space by Assumption 2.1(i), Lemma 1.3.2 in van der Vaart and Wellner (1996) implies $\mathbb{G}_0$ is separable. Therefore, the claim of the Lemma follows from (A.66), Theorem 1.12.2 and Addendum 1.12.3 in van der Vaart and Wellner (1996). ∎

Lemma A.2. *Let Assumptions 2.1(i), 2.2 and 3.1 hold, and $\mathbb{G}_1, \mathbb{G}_2 \in \mathbb{D}$ be independent random variables with the same law as $\mathbb{G}_0$. Then, it follows that on $\mathbb{D} \times \mathbb{D}$:*

$$
(r_n\{\hat{\theta}_n - \theta_0\}, r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}) \xrightarrow{L} (\mathbb{G}_1, \mathbb{G}_2) . \quad (A.67)
$$

PROOF: In these arguments we need to distinguish between outer and inner expectations, and we therefore employ the notation E^* and E_* respectively. For notational convenience we also let $\mathbb{G}_n \equiv r_n\{\hat{\theta}_n - \theta_0\}$ and $\mathbb{G}_n^* \equiv r_n\{\hat{\theta}_n^* - \hat{\theta}_n\}$. Then, note that Assumptions 2.2(i)-(ii), Lemma A.1, and Lemma 1.3.8 in van der Vaart and Wellner (1996) imply that both $\mathbb{G}_n$ and $\mathbb{G}_n^*$ are asymptotically measurable, and asymptotically tight in $\mathbb{D}$. Therefore, by Lemma 1.4.3 in van der Vaart and Wellner (1996) $(\mathbb{G}_n, \mathbb{G}_n^*)$ is asymptotically tight in $\mathbb{D} \times \mathbb{D}$ and asymptotically measurable as well. Thus, by Prohorov's theorem (Theorem 1.3.9 in van der Vaart and Wellner (1996)), each subsequence $\{(\mathbb{G}_{n_k}, \mathbb{G}_{n_k}^*)\}$ has an additional subsequence $\{(\mathbb{G}_{n_{k_j}}, \mathbb{G}_{n_{k_j}}^*)\}$ such that:

$$
(\mathbb{G}_{n_{k_j}}, \mathbb{G}_{n_{k_j}}^*) \xrightarrow{L} (\mathbb{Z}_1, \mathbb{Z}_2) \quad (A.68)
$$

for a tight Borel random variable $\mathbb{Z} \equiv (\mathbb{Z}_1, \mathbb{Z}_2) \in \mathbb{D} \times \mathbb{D}$. Since the sequence $\{(\mathbb{G}_{n_k}, \mathbb{G}^*_{n_k})\}$ was arbitrary, the Lemma follows if show the law of $\mathbb{Z}$ equals that of $(\mathbb{G}_1, \mathbb{G}_2)$.

Towards this end, let $f_1, f_2 \in \mathrm{BL}_1(\mathbb{D})$ satisfy $f_1(h) \geq 0$ and $f_2(h) \geq 0$ for all $h \in \mathbb{D}$. Then note that by result (A.68) it follows that:

$$\lim_{j \to \infty} E^*[f_1(\mathbb{G}_{n_{k_j}}) f_2(\mathbb{G}^*_{n_{k_j}})] = E[f_1(\mathbb{Z}_1) f_2(\mathbb{Z}_2)] . \tag{A.69}$$

However, $f_1, f_2 \in \mathrm{BL}_1(\mathbb{D})$ satisfying $f_1(h) \geq 0$ and $f_2(h) \geq 0$ for all $h \in \mathbb{D}$, Lemma 1.2.6 in van der Vaart and Wellner (1996), and Lemma A.5(iii) imply that:

$$\lim_{j \to \infty} E^*[f_1(\mathbb{G}_{n_{k_j}}) f_2(\mathbb{G}^*_{n_{k_j}})] - E^*[f_1(\mathbb{G}_{n_{k_j}}) E[f_2(\mathbb{G}_0)]]$$

$$\geq \lim_{j \to \infty} E^*[f_1(\mathbb{G}_{n_{k_j}}) E^*[f_2(\mathbb{G}^*_{n_{k_j}})|\{X_i\}_{i=1}^n]] - E^*[f_1(\mathbb{G}_{n_{k_j}}) E[f_2(\mathbb{G}_0)]]$$

$$\geq - \lim_{j \to \infty} E^*[f_1(\mathbb{G}_{n_{k_j}})|E^*[f_2(\mathbb{G}^*_{n_{k_j}})|\{X_i\}_{i=1}^n] - E[f_2(\mathbb{G}_0)]|]$$

$$\geq - \lim_{j \to \infty} E^*[\sup_{f \in \mathrm{BL}_1(\mathbb{D})} |E^*[f(\mathbb{G}^*_{n_{k_j}})|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|] , \tag{A.70}$$

where in the final inequality we exploited that $f_1 \in \mathrm{BL}_1(\mathbb{D})$. Similarly, Lemma 1.2.6 in van der Vaart and Wellner (1996), Lemma A.5(iv), and $f_1, f_2 \in \mathrm{BL}_1(\mathbb{D})$ also imply that:

$$\lim_{j \to \infty} E_*[f_1(\mathbb{G}_{n_{k_j}}) f_2(\mathbb{G}^*_{n_{k_j}})] - E_*[f_1(\mathbb{G}_{n_{k_j}}) E[f_2(\mathbb{G}_0)]]$$

$$\leq \lim_{j \to \infty} E_*[f_1(\mathbb{G}_{n_{k_j}}) E^*[f_2(\mathbb{G}^*_{n_{k_j}})|\{X_i\}_{i=1}^n]] - E_*[f_1(\mathbb{G}_{n_{k_j}}) E[f_2(\mathbb{G}_0)]]$$

$$\leq \lim_{j \to \infty} E^*[f_1(\mathbb{G}_{n_{k_j}})|E^*[f_2(\mathbb{G}^*_{n_{k_j}})|\{X_i\}_{i=1}^n] - E[f_2(\mathbb{G}_0)]|]$$

$$\leq \lim_{j \to \infty} E^*[\sup_{f \in \mathrm{BL}_1(\mathbb{D})} |E^*[f(\mathbb{G}^*_{n_{k_j}})|\{X_i\}_{i=1}^n] - E[f(\mathbb{G}_0)]|] . \tag{A.71}$$

Thus, combining result (A.65) together with (A.70) and (A.71), and the fact that $(\mathbb{G}_n, \mathbb{G}^*_n)$ and $\mathbb{G}_n$ are asymptotically measurable, we can conclude that:

$$\lim_{j \to \infty} E^*[f_1(\mathbb{G}_{n_{k_j}}) f_2(\mathbb{G}^*_{n_{k_j}})] = \lim_{j \to \infty} E^*[f_1(\mathbb{G}_{n_{k_j}}) E[f_2(\mathbb{G}_0)]]$$

$$= E[f_1(\mathbb{G}_0)] E[f_2(\mathbb{G}_0)] , \tag{A.72}$$

where the final result follows from $\mathbb{G}_n \overset{L}{\to} \mathbb{G}_0$ in $\mathbb{D}$. Hence, (A.69) and (A.72) imply

$$E[f_1(\mathbb{Z}_1) f_2(\mathbb{Z}_2)] = E[f_1(\mathbb{G}_0)] E[f_2(\mathbb{G}_0)] \tag{A.73}$$

for all $f_1, f_2 \in \mathrm{BL}_1(\mathbb{D})$ satisfying $f_1(h) \geq 0$ and $f_2(h) \geq 0$ for all $h \in \mathbb{D}$. Since $\mathbb{Z}$ is tight on $\mathbb{D} \times \mathbb{D}$ it is also separable by Lemma 1.3.2 in van der Vaart and Wellner (1996) and Assumption 2.1(i), and hence result (A.73) and Lemma 1.4.2 in van der Vaart and Wellner (1996) imply the law of $\mathbb{Z}$ equals that of $(\mathbb{G}_1, \mathbb{G}_2)$. In view of (A.68), the claim

49

of the Lemma then follows. ■

Lemma A.3. *Let Assumptions 2.1, 2.2(ii), and 2.3 hold, $\mathbb{D}_L$ denote the support of $\mathbb{G}_0$ and suppose $0 \in \mathbb{D}_L$. If $\phi'_{\theta_0} : \mathbb{D} \to \mathbb{E}$ is $\mathbb{G}_0$-translation invariant, then for any $a_0 \in \mathbb{D}_L$ and bounded continuous function $f : \mathbb{E} \to \mathbf{R}$, it follows that:*

$$E[f(\phi'_{\theta_0}(\mathbb{G}_0))] = E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_0) - \phi'_{\theta_0}(a_0))] . \tag{A.74}$$

PROOF: For any $a_0 \in \mathbb{D}$ and sequence $\{a_n\} \in \mathbb{D}$ with $\|a_0 - a_n\|_{\mathbb{D}} = o(1)$, continuity of ϕ'_{θ_0} and f, f being bounded, and the dominated convergence theorem imply:

$$\lim_{n \to \infty} E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_n) - \phi'_{\theta_0}(a_n))] = E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_0) - \phi'_{\theta_0}(a_0))] . \tag{A.75}$$

Next, let $B_\epsilon(a_0) \equiv \{a \in \mathbb{D} : \|a_0 - a\|_{\mathbb{D}} < \epsilon\}$, and observe that result (A.75) implies:

$$E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_0) - \phi'_{\theta_0}(a_0))] = \liminf_{\epsilon \downarrow 0} \inf_{a \in B_\epsilon(a_0)} E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a) - \phi'_{\theta_0}(a))]$$

$$\leq \limsup_{\epsilon \downarrow 0} \sup_{a \in B_\epsilon(a_0)} E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a) - \phi'_{\theta_0}(a))] = E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_0) - \phi'_{\theta_0}(a_0))] . \tag{A.76}$$

Letting L denote the law of $\mathbb{G}_0$, and for $\mathbb{G}_1$ and $\mathbb{G}_2$ independent copies of $\mathbb{G}_0$, we have:

$$\inf_{a \in B_\epsilon(a_0)} E[f(\phi'_{\theta_0}(\mathbb{G}_1 + a) - \phi'_{\theta_0}(a))] P(\mathbb{G}_2 \in B_\epsilon(a_0))$$

$$\leq \int_{B_\epsilon(a_0)} \int_{\mathbb{D}_L} f(\phi'_{\theta_0}(z_1 + z_2) - \phi'_{\theta_0}(z_2)) dL(z_1) dL(z_2)$$

$$\leq \sup_{a \in B_\epsilon(a_0)} E[f(\phi'_{\theta_0}(\mathbb{G}_1 + a) - \phi'_{\theta_0}(a))] P(\mathbb{G}_2 \in B_\epsilon(a_0)) . \tag{A.77}$$

In particular, if $a_0 \in \mathbb{D}_L$, then $P(\mathbb{G}_2 \in B_\epsilon(a_0)) > 0$ for all $\epsilon > 0$, and thus we conclude:

$$E[f(\phi'_{\theta_0}(\mathbb{G}_0 + a_0) - \phi'_{\theta_0}(a_0))] = \lim_{\epsilon \downarrow 0} E[f(\phi'_{\theta_0}(\mathbb{G}_1 + \mathbb{G}_2) - \phi'_{\theta_0}(\mathbb{G}_2)) | \mathbb{G}_2 \in B_\epsilon(a_0)]$$

$$= \lim_{\epsilon \downarrow 0} E[f(\phi'_{\theta_0}(\mathbb{G}_1 + \mathbb{G}_2) - \phi'_{\theta_0}(\mathbb{G}_2)) | \mathbb{G}_2 \in B_\epsilon(0)] = E[f(\phi'_{\theta_0}(\mathbb{G}_0))] , \tag{A.78}$$

where the first equality follows from (A.76) and (A.77), the second by ϕ'_{θ_0} being $\mathbb{G}_0$-translation invariant and $0 \in \mathbb{D}_L$, while the final equality follows by results (A.76), (A.77), and $\phi'_{\theta_0}(0) = 0$ due to ϕ'_{θ_0} being homogenous of degree one. ■

Lemma A.4. *Let Assumption 2.1(i) hold, $\psi : \mathbb{D} \to \mathbb{E}$ be continuous, and $K \subset \mathbb{D}$ be compact. It then follows that for every $\epsilon > 0$ there exist $\delta > 0, \eta > 0$ such that:*

$$\sup_{(a,b) \in K^\delta \times K^\delta : \|a-b\|_{\mathbb{D}} < \eta} \|\psi(a) - \psi(b)\|_{\mathbb{E}} < \epsilon . \tag{A.79}$$

PROOF: Fix $\epsilon > 0$ and note that since $\psi : \mathbb{D} \to \mathbb{E}$ is continuous, it follows that for every $a \in \mathbb{D}$ there exists a ζ_a such that $\|\psi(a) - \psi(b)\|_{\mathbb{E}} < \epsilon/2$ for all $b \in \mathbb{D}$ with $\|a - b\|_{\mathbb{D}} < \zeta_a$. Letting $B_{\zeta_a/4}(a) \equiv \{b \in \mathbb{D} : \|a - b\|_{\mathbb{D}} < \zeta_a/4\}$, then observe that $\{B_{\zeta_a/4}(a)\}_{a \in K}$ forms an open cover of K and hence, by compactness of K, there exists a finite subcover $\{B_{\zeta_{a_j}/4}(a_j)\}_{j=1}^{J}$ for some $J < \infty$. To establish the Lemma, we then let

$$\eta \equiv \min_{1 \le j \le J} \frac{\zeta_{a_j}}{4} \qquad \delta \equiv \min_{1 \le j \le J} \frac{\zeta_{a_j}}{4} \ . \tag{A.80}$$

For any $a \in K^{\delta}$, there then exists a $\Pi a \in K$ such that $\|a - \Pi a\|_{\mathbb{D}} < \delta$, and since $\{B_{\zeta_{a_j}/4}(a_j)\}_{j=1}^{J}$ covers K, there also is a $\bar{j}$ such that $\Pi a \in B_{\zeta_{a_{\bar{j}}}/4}(a_{\bar{j}})$. Thus, we have

$$\|a - a_{\bar{j}}\|_{\mathbb{D}} \le \|a - \Pi a\|_{\mathbb{D}} + \|\Pi a - a_{\bar{j}}\|_{\mathbb{D}} < \delta + \frac{\zeta_{a_{\bar{j}}}}{4} \le \frac{\zeta_{a_{\bar{j}}}}{2} \ , \tag{A.81}$$

due to the choice of δ in (A.80). Moreover, if $b \in \mathbb{D}$ satisfies $\|a - b\|_{\mathbb{D}} < \eta$, then:

$$\|b - a_{\bar{j}}\|_{\mathbb{D}} \le \|a - b\|_{\mathbb{D}} + \|a - a_{\bar{j}}\|_{\mathbb{D}} < \eta + \frac{\zeta_{a_{\bar{j}}}}{2} \le \zeta_{a_{\bar{j}}} \ , \tag{A.82}$$

by the choice of η in (A.80). We conclude from (A.81), (A.82) that $a, b \in B_{\zeta_{a_{\bar{j}}}}(a_{\bar{j}})$, and

$$\|\psi(a) - \psi(b)\|_{\mathbb{E}} \le \|\psi(a) - \psi(a_{\bar{j}})\|_{\mathbb{E}} + \|\psi(b) - \psi(a_{\bar{j}})\|_{\mathbb{E}} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \tag{A.83}$$

by our choice of $\zeta_{a_{\bar{j}}}$. Thus, the Lemma follows from result (A.83). ∎

Lemma A.5. *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space,* $c \in \mathbf{R}_{+}$ *and* $U : \Omega \to \mathbf{R}$ *and* $V : \Omega \to \mathbf{R}$ *be arbitrary maps satisfying* $U(\omega) \ge 0$ *and* $V(\omega) \ge 0$ *for all* $\omega \in \Omega$. *If* E^* *and* E_* *denote outer and inner expectations respectively, then it follows that:*

(i) $E^*[U] - c \ge -E^*[|U - c|]$.
(ii) $E_*[U] - c \le E^*[|U - c|]$.
(iii) $E^*[UV] - E^*[Uc] \ge -E^*[U|V - c|]$ *whenever* $\min\{E^*[UV], E^*[Uc]\} < \infty$.
(iv) $E_*[UV] - E_*[Uc] \le E^*[U|V - c|]$ *whenever* $\min\{E_*[UV], E_*[Uc]\} < \infty$.
(v) $|E^*[UV] - E^*[Uc]| \le E^*[U|V - c|]$ *whenever* $\min\{E_*[UV], E_*[Uc]\} < \infty$.

PROOF: The arguments are simple and tedious, but unfortunately necessary to address the possible nonlinearity of inner and outer expectations. Throughout, for a map $T : \Omega \to \mathbf{R}$, we let T^* and T_* denote the minimal measurable majorant and the maximal measurable minorant of T respectively. We will also exploit the fact that:

$$E_*[T] = -E^*[-T] \ , \tag{A.84}$$

and that $E^*[T] = E[T^*]$ whenever $E[T^*]$ exists, which in the context of this Lemma is always satisfied since all variables are positive.

To establish the first claim of the Lemma, note that Lemma 1.2.2(i) in van der Vaart and Wellner (1996) implies $U^* - c = (U - c)^*$. Therefore, (A.84) and $E_* \leq E^*$ yield:

$$E^*[U] - c = E[U^* - c] = E[(U - c)^*] = E^*[U - c]$$
$$\geq E^*[-|U - c|] = -E_*[|U - c|] \geq -E^*[|U - c|] . \quad \text{(A.85)}$$

Similarly, for the second claim of the Lemma, exploit that $E_* \leq E^*$, and once again employ Lemma 1.2.2(i) in van der Vaart and Wellner (1996) to conclude that:

$$E_*[U] - c \leq E^*[U] - c = E[U^* - c] = E[(U - c)^*] \leq E^*[|U - c|] . \quad \text{(A.86)}$$

For the third claim, note that Lemma 1.2.2(iii) in van der Vaart and Wellner (1996) implies $|(UV)^* - (Uc)^*| \leq |UV - Uc|^*$. Thus, since $|U(V - c)| = U|V - c|$ as a result of $U(\omega) \geq 0$ for all $\omega \in \Omega$, we obtain from relationship (A.84) and $E_* \leq E^*$ that:

$$E^*[UV] - E^*[Uc] = E[(UV)^* - (Uc)^*] \geq E[-|(UV)^* - (Uc)^*|]$$
$$\geq E[-|UV - Uc|^*] = -E_*[U|V - c|] \geq -E^*[U|V - c|] . \quad \text{(A.87)}$$

Similarly, for the fourth claim of the Lemma, employ (A.84), that $|(-Uc)^* - (-UV)^*| \leq |(-Uc) - (-UV)|^*$ by Lemma 1.2.2(iii) in van der Vaart and Wellner (1996), and that $|UV - Uc| = U|V - c|$ due to $U(\omega) \geq 0$ for all $\omega \in \Omega$ to obtain that:

$$E_*[UV] - E_*[Uc] = E[(-Uc)^* - (-UV)^*] \leq E[|(-Uc)^* - (-UV)^*|]$$
$$\leq E[|(-Uc) - (-UV)|^*] = E^*[U|V - c|] . \quad \text{(A.88)}$$

Finally, for the fifth claim of the Lemma, note the same arguments as in (A.88) yield

$$E^*[UV] - E^*[Uc] = E[(Uc)^* - (UV)^*] \leq E[|(Uc)^* - (UV)^*|]$$
$$\leq E[|(Uc) - (UV)|^*] = E^*[U|V - c|] . \quad \text{(A.89)}$$

Thus, part (v) of the Lemma follows from part (iii) and (A.89). ∎

Lemma A.6. *Let Assumptions 2.1, 2.3(i) hold, and suppose that for some $\kappa > 0$ and $C_0 < \infty$ we have $\|\hat{\phi}_n'(h_1) - \hat{\phi}_n'(h_2)\|_{\mathbb{E}} \leq C_0\|h_1 - h_2\|_{\mathbb{D}}^\kappa$ for all $h_1, h_2 \in \mathbb{D}$ outer almost surely. Then, Assumption 3.3 holds provided that for all $h \in \mathbb{D}_0$ we have:*

$$\|\hat{\phi}_n'(h) - \phi_{\theta_0}'(h)\|_{\mathbb{E}} = o_p(1) . \quad \text{(A.90)}$$

PROOF: Fix $\epsilon > 0$, let $K_0 \subseteq \mathbb{D}_0$ be compact, and for any $h \in \mathbb{D}$ let $\Pi : \mathbb{D} \to K_0$ satisfy $\|h - \Pi h\|_{\mathbb{D}} = \inf_{a \in K_0} \|h - a\|_{\mathbb{D}}$ – here attainment is guaranteed by compactness. Since

$\phi'_{\theta_0} : \mathbb{D} \to \mathbb{E}$ is continuous, Lemma A.4 implies there exists a $\delta_1 > 0$ such that:

$$\sup_{h \in K_0^{\delta_1}} \|\phi'_{\theta_0}(h) - \phi'_{\theta_0}(\Pi h)\|_{\mathbb{E}} < \epsilon . \tag{A.91}$$

Next, set $\delta_2 < (\epsilon/C_0)^{1/\kappa}$ and note that by hypothesis we have outer almost surely that:

$$\sup_{h \in K_0^{\delta_2}} \|\hat{\phi}'_n(h) - \hat{\phi}'_n(\Pi h)\|_{\mathbb{E}} \leq \sup_{h \in K_0^{\delta_2}} C_0 \|h - \Pi h\|_{\mathbb{E}}^{\kappa} \leq C_0 \delta_2^{\kappa} < \epsilon . \tag{A.92}$$

Defining $\delta_3 \equiv \min\{\delta_1, \delta_2\}$, exploiting (A.91), (A.92), and $\Pi h \in K_0$ we then conclude:

$$\sup_{h \in K_0^{\delta_3}} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}}$$
$$\leq \sup_{h \in K_0^{\delta_3}} \{\|\hat{\phi}'_n(h) - \hat{\phi}'_n(\Pi h)\|_{\mathbb{E}} + \|\phi'_{\theta_0}(h) - \phi'_{\theta_0}(\Pi h)\|_{\mathbb{E}} + \|\hat{\phi}'_n(\Pi h) - \phi'_{\theta_0}(\Pi h)\|_{\mathbb{E}}\}$$
$$\leq \sup_{h \in K_0} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} + 2\epsilon \tag{A.93}$$

outer almost surely. Thus, since $K_0^{\delta} \subseteq K_0^{\delta_3}$ for all $\delta \leq \delta_3$ we obtain from (A.93) that:

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} P(\sup_{h \in K_0^{\delta}} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} > 5\epsilon)$$
$$\leq \limsup_{n \to \infty} P(\sup_{h \in K_0} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} > 3\epsilon) . \tag{A.94}$$

Next note that since K_0 is compact, ϕ'_{θ_0} is uniformly continuous on K_0, and thus we can find a finite collection $\{h_j\}_{j=1}^{J}$ with $J < \infty$ such that $h_j \in K_0$ for all j and:

$$\sup_{h \in K_0} \min_{1 \leq j \leq J} \max\{C_0 \|h - h_j\|_{\mathbb{D}}^{\kappa}, \|\phi'_{\theta_0}(h) - \phi'_{\theta_0}(h_j)\|_{\mathbb{E}}\} < \epsilon . \tag{A.95}$$

In particular, since $\|\hat{\phi}'_{\theta_0}(h) - \hat{\phi}'_{\theta_0}(h_j)\|_{\mathbb{E}} \leq C_0 \|h - h_j\|_{\mathbb{D}}^{\kappa}$, we conclude from (A.95) that:

$$\sup_{h \in K_0} \|\hat{\phi}'_{\theta_0}(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} \leq \max_{1 \leq j \leq J} \|\hat{\phi}'_{\theta_0}(h_j) - \phi'_{\theta_0}(h_j)\|_{\mathbb{E}} + 2\epsilon . \tag{A.96}$$

Thus, we can conclude from (A.96) and $\hat{\phi}'_{\theta_0}$ satisfying (A.90) for any $h \in \mathbb{D}_0$ that:

$$\limsup_{n \to \infty} P(\sup_{h \in K_0} \|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)\|_{\mathbb{E}} > 3\epsilon)$$
$$\leq \limsup_{n \to \infty} P(\max_{1 \leq j \leq J} \|\hat{\phi}'_n(h_j) - \phi'_{\theta_0}(h_j)\|_{\mathbb{E}} > \epsilon) = 0 . \tag{A.97}$$

Since ϵ and K_0 were arbitrary, the Lemma follows from (A.94) and (A.95). ■

Lemma A.7. *Let Assumptions 2.1, 2.2(ii) hold, and $\mathbb{G}_0$ be a centered Gaussian measure. Then, it follows that the support of $\mathbb{G}_0$ is a separable Banach space under $\|\cdot\|_{\mathbb{D}}$.*

PROOF: Let τ and τ_w denote the strong and weak topologies on $\mathbb{D}$ respectively, and $\mathcal{B}(\tau)$ and $\mathcal{B}(\tau_w)$ the corresponding σ-algebras generated by them. Further let P denote the distribution of $\mathbb{G}_0$ on $\mathbb{D}$, and note that by Assumption 2.2(ii) and Lemma 1.3.2 in van der Vaart and Wellner (1996), P is τ-separable. Let $\mathbb{S}(\tau)$ denote the support of P under τ, formally the smallest τ-closed set $\mathbb{S}(\tau) \subseteq \mathbb{D}$ such that $P(\mathbb{S}(\tau)) = 1$, and let

$$\mathbb{P} \equiv \overline{\text{span}\{\mathbb{S}(\tau)\}}^{\tau} \tag{A.98}$$

denote the τ-closed linear span of $\mathbb{S}(\tau)$. Since P is separable and $\mathbb{S}(\tau) \subseteq \mathbb{D}$, it follows that $\mathbb{P}$ is a separable Banach space under $\|\cdot\|_{\mathbb{D}}$.

In what follows, we aim to show $\mathbb{P} = \mathbb{S}(\tau)$ in order to establish the Lemma. To this end, first note that $\mathbb{P}$ being separable, and Theorem 7.1.7 in Bogachev (2007) imply that P is Radon with respect to $\mathcal{B}(\tau)$. Since $\mathcal{B}(\tau_w) \subseteq \mathcal{B}(\tau)$ and τ-compact sets are also τ_w-compact, it follows that P is also Radon on $\mathcal{B}(\tau_w)$ when $\mathbb{D}$ is equipped with τ_w instead. Letting $\mathfrak{C}$ denote the cylindrical σ-algebra, we then conclude from $\mathfrak{C} \subseteq \mathcal{B}(\tau_w)$ that P is also Radon on $\mathfrak{C}$ with $\mathbb{D}$ equipped with τ_w. Hence, for $\mathbb{N}_P(\tau_w)$ the minimal closed affine subspace of $\mathbb{D}$ for which $P(\mathbb{N}_P(\tau_w)) = 1$, we obtain from P being Radon on $\mathfrak{C}$ and Proposition 7.4(i) in Davydov et al. (1998) that

$$\mathbb{N}_P(\tau_w) = \mathbb{S}(\tau_w) . \tag{A.99}$$

Moreover, since affine spaces are convex, Theorem 5.98 in Aliprantis and Border (2006) implies $\mathbb{N}_P(\tau) = \mathbb{N}_P(\tau_w)$. Thus, since $\mathbb{S}(\tau)$ is τ_w-closed, we have by (A.99):

$$\mathbb{S}(\tau) \subseteq \mathbb{N}_P(\tau) = \mathbb{N}_P(\tau_w) = \mathbb{S}(\tau_w) \subseteq \mathbb{S}(\tau) . \tag{A.100}$$

However, by Proposition 7.4(ii) in Davydov et al. (1998), $0 \in \mathbb{N}_P(\tau)$ and hence $\mathbb{N}_P(\tau)$ must be a vector space. Combining (A.98) and (A.99) we thus conclude $\mathbb{S}(\tau) = \mathbb{P}$ and the claim of the Lemma follows. $\blacksquare$

APPENDIX B - Results for Examples 2.1-2.6

Lemma B.1. *Let $\mathbf{A}$ be totally bounded under a norm $\|\cdot\|_{\mathbf{A}}$, and $\bar{\mathbf{A}}$ denote its closure under $\|\cdot\|_{\mathbf{A}}$. Further let $\phi : \ell^{\infty}(\mathbf{A}) \to \mathbf{R}$ be given by $\phi(\theta) = \sup_{a \in \mathbf{A}} \theta(a)$, and define $\Psi_{\bar{\mathbf{A}}}(\theta) \equiv \arg\max_{a \in \bar{\mathbf{A}}} \theta(a)$ for any $\theta \in \mathcal{C}(\bar{\mathbf{A}})$. Then, ϕ is Hadamard directionally differentiable tangentially to $\mathcal{C}(\bar{\mathbf{A}})$ at any $\theta \in \mathcal{C}(\bar{\mathbf{A}})$, and $\phi'_{\theta} : \mathcal{C}(\bar{\mathbf{A}}) \to \mathbf{R}$ satisfies:*

$$\phi'_{\theta}(h) = \sup_{a \in \Psi_{\bar{\mathbf{A}}}(\theta)} h(a) \qquad h \in \mathcal{C}(\bar{\mathbf{A}}) .$$

PROOF: First note Corollary 3.29 in Aliprantis and Border (2006) implies $\bar{\mathbf{A}}$ is compact under $\|\cdot\|_{\mathbf{A}}$. Next, let $\{t_n\}$ and $\{h_n\}$ be sequence with $t_n \in \mathbf{R}$, $h_n \in \ell^{\infty}(\mathbf{A})$ for all n

and $\|h_n - h\|_\infty = o(1)$ for some $h \in \mathcal{C}(\bar{\mathbf{A}})$. Then note that for any $\theta \in \mathcal{C}(\bar{\mathbf{A}})$ we have:

$$| \sup_{a \in \mathbf{A}}\{\theta(a) + t_n h_n(a)\} - \sup_{a \in \mathbf{A}}\{\theta(a) + t_n h(a)\}| \leq t_n \|h_n - h\|_\infty = o(t_n) \ . \qquad \text{(B.1)}$$

Further note that since $\bar{\mathbf{A}}$ is compact, $\Psi_{\bar{\mathbf{A}}}(\theta)$ is well defined for any $\theta \in \mathcal{C}(\bar{\mathbf{A}})$. Defining $\Gamma_\theta : \mathcal{C}(\bar{\mathbf{A}}) \to \mathcal{C}(\bar{\mathbf{A}})$ to be given by $\Gamma_\theta(g) = \theta + g$, then note that Γ_θ is trivially continuous. Therefore, Theorem 17.31 in Aliprantis and Border (2006) and the relation

$$\Psi_{\bar{\mathbf{A}}}(\theta + g) = \arg\max_{a \in \bar{\mathbf{A}}} \Gamma_\theta(g)(a) \qquad \text{(B.2)}$$

imply that $\Psi_{\bar{\mathbf{A}}}(\theta + g)$ is upper hemicontinuous in g. In particular, for $\Psi_{\bar{\mathbf{A}}}(\theta)^\epsilon \equiv \{a \in \bar{\mathbf{A}} : \inf_{a_0 \in \Psi_{\bar{\mathbf{A}}}(\theta)} \|a - a_0\|_{\mathbf{A}} \leq \epsilon\}$, it follows from $\|t_n h\|_\infty = o(1)$ that $\Psi_{\bar{\mathbf{A}}}(\theta + t_n h) \subseteq \Psi_{\bar{\mathbf{A}}}(\theta)^{\delta_n}$ for some $\delta_n \downarrow 0$. Thus, since $\Psi_{\bar{\mathbf{A}}}(\theta) \subseteq \Psi_{\bar{\mathbf{A}}}(\theta)^{\delta_n}$ we can conclude that

$$| \sup_{a \in \bar{\mathbf{A}}}\{\theta(a) + t_n h(a)\} - \sup_{a \in \Psi_{\bar{\mathbf{A}}}(\theta)}\{\theta(a) + t_n h(a)\}|$$
$$= \sup_{a \in \Psi_{\bar{\mathbf{A}}}(\theta)^{\delta_n}}\{\theta(a) + t_n h(a)\} - \sup_{a \in \Psi_{\bar{\mathbf{A}}}(\theta)}\{\theta(a) + t_n h(a)\}$$
$$\leq \sup_{a_0, a_1 \in \bar{\mathbf{A}} : \|a_0 - a_1\|_{\mathbf{A}} \leq \delta_n} t_n |h(a_0) - h(a_1)|$$
$$= o(t_n) \ , \qquad \text{(B.3)}$$

where the final result follows from h being uniformly continuous by compactness of $\bar{\mathbf{A}}$. Therefore, exploiting (B.1), (B.3) and θ being constant on $\Psi_{\bar{\mathbf{A}}}(\theta)$ yields

$$| \sup_{a \in \mathbf{A}}\{\theta(a) + t_n h_n(a)\} - \sup_{a \in \mathbf{A}} \theta(a) - t_n \sup_{a \in \Psi_{\bar{\mathbf{A}}}(\theta)} h(a)|$$
$$\leq | \sup_{a \in \Psi_{\bar{\mathbf{A}}}(\theta)}\{\theta(a) + t_n h(a)\} - \sup_{a \in \Psi_{\bar{\mathbf{A}}}(\theta)} \theta(a) - t_n \sup_{a \in \Psi_{\bar{\mathbf{A}}}(\theta)} h(a)| + o(t_n) = o(t_n) \ , \quad \text{(B.4)}$$

which verifies the claim of the Lemma. ∎

Lemma B.2. *Let $w : \mathbf{R} \to \mathbf{R}_+$ satisfy $\int_{\mathbf{R}} w(u)du < \infty$ and $\phi : \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R}) \to \mathbf{R}$ be given by $\phi(\theta) = \int_{\mathbf{R}} \max\{\theta^{(1)}(u) - \theta^{(2)}(u), 0\}w(u)du$ for any $\theta = (\theta^{(1)}, \theta^{(2)}) \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$. Then, ϕ is Hadamard directionally differentiable at any $\theta \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$ with $\phi'_\theta : \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R}) \to \mathbf{R}$ satisfying for any $h = (h^{(1)}, h^{(2)}) \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$*

$$\phi'_\theta(h) = \int_{B_0(\theta)} \max\{h^{(1)}(u) - h^{(2)}(u), 0\}w(u)du + \int_{B_+(\theta)} (h^{(1)}(u) - h^{(2)}(u))w(u)du \ ,$$

where $B_+(\theta) \equiv \{u \in \mathbf{R} : \theta^{(1)}(u) > \theta^{(2)}(u)\}$ and $B_0(\theta) \equiv \{u \in \mathbf{R} : \theta^{(1)}(u) = \theta^{(2)}(u)\}$.

PROOF: Let $\{h_n\} = \{(h_n^{(1)}, h_n^{(2)})\}$ be a sequence in $\ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$ satisfying $\|h_n^{(1)} - h^{(1)}\|_\infty \vee \|h_n^{(2)} - h^{(2)}\|_\infty = o(1)$ for some $h = (h^{(1)}, h^{(2)}) \in \ell^\infty(\mathbf{R}) \times \ell^\infty(\mathbf{R})$. Further let:

$$B_-(\theta) \equiv \{u \in \mathbf{R} : \theta^{(1)}(u) < \theta^{(2)}(u)\} . \tag{B.5}$$

Next, observe that since $\theta^{(1)}(u) - \theta^{(2)}(u) < 0$ for all $u \in B_-(\theta)$, and $\|h_n^{(1)} - h_n^{(2)}\|_\infty = O(1)$ due to $\|h^{(1)} - h^{(2)}\|_\infty < \infty$, the dominated convergence theorem yields that:

$$\int_{B_-(\theta)} \max\{(\theta^{(1)}(u) - \theta^{(2)}(u)) + t_n(h_n^{(1)}(u) - h_n^{(2)}(u)), 0\} w(u) du$$
$$\lesssim t_n \int_{B_-(\theta)} 1\{t_n(h_n^{(1)}(u) - h_n^{(2)}(u)) \geq -(\theta^{(1)}(u) - \theta^{(2)}(u))\} w(u) du = o(t_n) . \tag{B.6}$$

Thus, (B.6), $B_-(\theta)^c = B_+(\theta) \cup B_0(\theta)$ and the dominated convergence theorem imply

$$\frac{1}{t_n}\{\phi(\theta + t_n h_n) - \phi(\theta)\}$$
$$= \int_{B_-(\theta)^c} \max\{h_n^{(1)}(u) - h_n^{(2)}(u), -\frac{\theta(u)^{(1)} - \theta^{(2)}(u)}{t_n}\} w(u) du + o(1) = \phi'_\theta(h) + o(1)$$

which establishes the claim of the Lemma. ∎

Lemma B.3. *Let Assumptions 2.1, 2.3 hold, and* $\mathbf{A}$ *be compact under* $\|\cdot\|_{\mathbf{A}}$. *Further suppose* $\phi : \ell^\infty(\mathbf{A}) \to \mathbf{R}$ *is Hadamard directionally differentiable tangentially to* $\mathcal{C}(\mathbf{A})$ *at* $\theta_0 \in \mathcal{C}(\mathbf{A})$, *and that for some* $A_0 \subseteq \mathbf{A}$, *its derivative* $\phi'_{\theta_0} : \mathcal{C}(\mathbf{A}) \to \mathbf{R}$ *is given by:*

$$\phi'_{\theta_0}(h) = \sup_{a \in A_0} h(a) . \tag{B.7}$$

If $\hat{A}_0 \subseteq \mathbf{A}$ *outer almost surely, and* $d_H(\hat{A}_0, A_0, \|\cdot\|_{\mathbf{A}}) = o_p(1)$, *then it follows that* $\hat{\phi}'_n : \ell^\infty(\mathbf{A}) \to \mathbf{R}$ *given by* $\hat{\phi}'_n(h) = \sup_{a \in \hat{A}_0} h(a)$ *for any* $h \in \ell^\infty(\mathbf{A})$ *satisfies* (41).

PROOF: First note that $\hat{\phi}'_n$ is outer almost surely Lipschitz since $|\hat{\phi}'_n(h_1) - \hat{\phi}'_n(h_2)| \leq \|h_1 - h_2\|_\infty$ for all $h_1, h_2 \in \ell^\infty(\mathbf{A})$ due to $\hat{A}_0 \subseteq \mathbf{A}$ outer almost surely. Therefore, by Lemma A.6 it suffices to verify that for any $h \in \mathcal{C}(\mathbf{A})$, $\hat{\phi}'_n$ satisfies

$$|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)| = o_p(1) . \tag{B.8}$$

Towards this end, fix an arbitrary $\epsilon_0 > 0$ and note h is uniformly continuous on $\mathbf{A}$ due to $\mathbf{A}$ being compact. Hence, we conclude there exists an $\eta > 0$ such that

$$\sup_{\|a_1 - a_2\|_{\mathbf{A}} < \eta} |h(a_1) - h(a_2)| < \epsilon_0 . \tag{B.9}$$

Moreover, given the definitions of $\hat{\phi}'_n$ and ϕ'_{θ_0} it also follows that for any $h \in \ell^\infty(\mathbf{A})$:

$$|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)| \leq \sup_{\|a_1 - a_2\|_\mathbf{A} \leq d_H(\hat{A}_0, A_0, \|\cdot\|_\mathbf{A})} |h(a_1) - h(a_2)| . \tag{B.10}$$

Thus, by results (B.9) and (B.10), and the Hausdorff consistency of $\hat{A}_0$, we obtain:

$$\limsup_{n \to \infty} P(|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)| > \epsilon_0) \leq \limsup_{n \to \infty} P(d_H(\hat{A}_0, A_0, \|\cdot\|_\mathbf{A}) > \eta) = 0 . \tag{B.11}$$

It follows that (B.8) indeed holds, and the claim of the Lemma follows. ∎

APPENDIX C - Results for Section 4

PROOF OF PROPOSITION 4.1: We proceed by verifying Assumptions 2.1, 2.2, and 2.3, and then employing Theorem 2.1 to obtain (73). To this end, define the maps $\phi_1 : \mathbb{H} \to \mathbb{H}$ to be given by $\phi_1(\theta) = \theta - \Pi_\Lambda \theta$, and $\phi_2 : \mathbb{H} \to \mathbf{R}$ by $\phi_2(\theta) \equiv \|\theta\|_\mathbb{H}$. Letting $\phi \equiv \phi_2 \circ \phi_1$ and noting $\phi_1(\theta_0) = 0$ due to $\theta_0 \in \Lambda$, we then obtain the equality:

$$r_n \|\hat{\theta}_n - \Pi_\Lambda \hat{\theta}_n\|_\mathbb{H} = r_n \{\phi(\hat{\theta}_n) - \phi(\theta_0)\} . \tag{C.1}$$

By Lemma 4.6 in Zaranotello (1971), ϕ_1 is then Hadamard directionally differentiable at θ_0 with derivative $\phi'_{1,\theta_0} : \mathbb{H} \to \mathbb{H}$ given by $\phi'_{1,\theta_0}(h) = h - \Pi_{T_{\theta_0}} h$; see also (Shapiro, 1994, p. 135). Moreover, since ϕ_2 is Hadamard directionally differentiable at $0 \in \mathbb{H}$ with derivative $\phi'_{2,0}(h) = \|h\|_\mathbb{H}$, Proposition 3.6 in Shapiro (1990) implies ϕ is Hadamard directionally differentiable at θ_0 with $\phi'_{\theta_0} = \phi'_{2,0} \circ \phi'_{1,\theta_0}$. In particular, we have

$$\phi'_{\theta_0}(h) = \|h - \Pi_{T_{\theta_0}} h\|_\mathbb{H} , \tag{C.2}$$

for any $h \in \mathbb{H}$. Thus, (C.2) verifies Assumption 2.1 and, because in this case $\mathbb{D} = \mathbb{D}_0 = \mathbb{H}$, we conclude Assumption 2.3 holds as well. Since Assumption 2.2 was directly imposed, the Proposition then follows form Theorem 2.1. ∎

PROOF OF PROPOSITION 4.2: In order to establish the first claim of the Proposition, we first observe that for any $h_1, h_2 \in \mathbb{H}$ we must have that:

$$\hat{\phi}'_n(h_1) - \hat{\phi}'_n(h_2) \leq \sup_{\theta \in \Lambda : \|\theta - \Pi_\Lambda \hat{\theta}_n\|_\mathbb{H} \leq \epsilon_n} \{\|h_1 - \Pi_{T_\theta} h_1\|_\mathbb{H} - \|h_2 - \Pi_{T_\theta} h_2\|_\mathbb{H}\}$$

$$\leq \sup_{\theta \in \Lambda : \|\theta - \Pi_\Lambda \hat{\theta}_n\|_\mathbb{H} \leq \epsilon_n} \{\|h_1 - \Pi_{T_\theta} h_2\|_\mathbb{H} - \|h_2 - \Pi_{T_\theta} h_2\|_\mathbb{H}\} \leq \|h_1 - h_2\|_\mathbb{H} , \tag{C.3}$$

where the first inequality follows from the definition of $\hat{\phi}'_n(h)$, the second inequality is implied by $\|h_1 - \Pi_{T_\theta} h_1\|_\mathbb{H} \leq \|h_1 - \Pi_{T_\theta} h_2\|_\mathbb{H}$ for all $\theta \in \Lambda$, and the third inequality holds by the triangle inequality. Result (C.3) further implies $\hat{\phi}'_n(h_2) - \hat{\phi}'_n(h_1) \leq \|h_1 - h_2\|_\mathbb{H}$,

and hence we can conclude $\hat{\phi}'_n : \mathbb{H} \to \mathbf{R}$ is Lipschitz – i.e. for any $h_1, h_2 \in \mathbb{H}$:

$$|\hat{\phi}'_n(h_1) - \hat{\phi}'_n(h_2)| \leq \|h_1 - h_2\|_{\mathbb{H}} . \tag{C.4}$$

Thus, by Lemma A.6, in verifying $\hat{\phi}'_n$ satisfies Assumption 3.3 it suffices to show that:

$$|\hat{\phi}'_n(h) - \phi'_{\theta_0}(h)| = o_p(1) \tag{C.5}$$

for all $h \in \mathbb{H}$. To this end, note that convexity of Λ and Proposition 46.5(2) in Zeidler (1984) imply $\|\Pi_\Lambda \theta_0 - \Pi_\Lambda \theta\|_{\mathbb{H}} \leq \|\theta_0 - \theta\|_{\mathbb{H}}$ for any $\theta \in \mathbb{H}$. Thus, since $r_n\{\hat{\theta}_n - \theta_0\}$ is asymptotically tight by Assumption 2.2 and $r_n \epsilon_n \uparrow \infty$ by hypothesis, we conclude that:

$$\liminf_{n\to\infty} P(\|\Pi_\Lambda \theta_0 - \Pi_\Lambda \hat{\theta}_n\|_{\mathbb{H}} \leq \epsilon_n) \geq \liminf_{n\to\infty} P(r_n \|\theta_0 - \hat{\theta}_n\|_{\mathbb{H}} \leq r_n \epsilon_n) = 1 . \tag{C.6}$$

Moreover, the same arguments as in (C.6) and the triangle inequality further imply that:

$$\liminf_{n\to\infty} P(\|\theta - \Pi_\Lambda \theta_0\|_{\mathbb{H}} \leq 2\epsilon_n \text{ for all } \theta \in \Lambda \text{ s.t. } \|\theta - \Pi_\Lambda \hat{\theta}_n\|_{\mathbb{H}} \leq \epsilon_n)$$
$$\geq \liminf_{n\to\infty} P(\|\Pi_\Lambda \theta_0 - \Pi_\Lambda \hat{\theta}_n\|_{\mathbb{H}} \leq \epsilon_n) = 1 . \tag{C.7}$$

Hence, from the definition of $\hat{\phi}'_n$ and results (C.6) and (C.7) we obtain for any $h \in \mathbb{H}$:

$$\liminf_{n\to\infty} P(\|h - \Pi_{T_{\theta_0}} h\|_{\mathbb{H}} \leq \hat{\phi}'_n(h) \leq \sup_{\theta \in \Lambda : \|\theta - \Pi_\Lambda \theta_0\|_{\mathbb{H}} \leq 2\epsilon_n} \|h - \Pi_{T_\theta} h\|_{\mathbb{H}}) = 1 . \tag{C.8}$$

Next, select a sequence $\{\theta_n\}$ with $\theta_n \in \Lambda$ and $\|\theta_n - \Pi_\Lambda \theta_0\|_{\mathbb{H}} \leq 2\epsilon_n$ for all n, such that:

$$\limsup_{n\to\infty}\{\sup_{\theta \in \Lambda : \|\theta - \Pi_\Lambda \theta_0\|_{\mathbb{H}} \leq 2\epsilon_n} \|h - \Pi_{T_\theta} h\|_{\mathbb{H}}\} = \lim_{n\to\infty} \|h - \Pi_{T_{\theta_n}} h\|_{\mathbb{H}} . \tag{C.9}$$

By Theorem 4.2.2 in Aubin and Frankowska (1990), the cone valued map $\theta \mapsto T_\theta$ is lower semicontinuous on Λ and hence since $\|\theta_n - \Pi_\Lambda \theta_0\|_{\mathbb{H}} = o(1)$, it follows that there exists a sequence $\{\tilde{h}_n\}$ such that $\tilde{h}_n \in \Pi_{T_{\theta_n}}$ for all n and $\|\Pi_{T_{\theta_0}} h - \tilde{h}_n\|_{\mathbb{H}} = o(1)$. Thus,

$$\limsup_{n\to\infty}\{\sup_{\theta \in \Lambda : \|\theta - \Pi_\Lambda \theta_0\|_{\mathbb{H}} \leq 2\epsilon_n} \|h - \Pi_{T_\theta} h\|_{\mathbb{H}}\}$$
$$= \lim_{n\to\infty} \|h - \Pi_{T_{\theta_n}} h\|_{\mathbb{H}} \leq \lim_{n\to\infty} \|h - \tilde{h}_n\|_{\mathbb{H}} = \|h - \Pi_{T_{\theta_0}} h\|_{\mathbb{H}} , \tag{C.10}$$

where the first equality follows from (C.9), the inequality by $\tilde{h}_n \in T_{\theta_n}$, and the second equality by $\|\tilde{h}_n - \Pi_{T_{\theta_0}} h\|_{\mathbb{H}} = o(1)$. Hence, combining (C.8) and (C.10) we conclude that (C.5) holds, and by Lemma A.6 and (C.4) that $\hat{\phi}'_n$ satisfies Assumption 3.3.

For the second claim, first observe that Λ being convex implies T_{θ_0} is a closed convex cone. Hence, by Proposition 46.5(4) in Zeidler (1984), it follows that $\|\Pi_{T_{\theta_0}} h\|_{\mathbb{H}}^2 =$

$\langle h, \Pi_{T_{\theta_0}} h \rangle_{\mathbb{H}}$ for any $h \in \mathbb{H}$. In particular, for any $h_1, h_2 \in \mathbb{H}$ we must have:

$$\|h_1 + h_2 - \Pi_{T_{\theta_0}}(h_1 + h_2)\|_{\mathbb{H}}^2 = \langle h_1 + h_2, h_1 + h_2 - \Pi_{T_{\theta_0}}(h_1 + h_2) \rangle_{\mathbb{H}} . \qquad (C.11)$$

However, Proposition 46.5(4) in Zeidler (1984) further implies that $\langle c, h_1 + h_2 - \Pi_{T_{\theta_0}}(h_1 + h_2) \rangle \le 0$ for any $h_1, h_2 \in \mathbb{H}$ and $c \in T_{\theta_0}$. Therefore, since $\Pi_{T_{\theta_0}} h_1, \Pi_{T_{\theta_0}} h_2 \in T_{\theta_0}$, we can conclude from result (C.11) and the Cauchy Schwarz inequality that

$$\|h_1 + h_2 - \Pi_{T_{\theta_0}}(h_1 + h_2)\|_{\mathbb{H}}^2 \le \langle h_1 - \Pi_{T_{\theta_0}} h_1 + h_2 - \Pi_{T_{\theta_0}} h_2, h_1 + h_2 - \Pi_{T_{\theta_0}}(h_1 + h_2) \rangle_{\mathbb{H}}$$

$$\le \|h_1 + h_2 - \Pi_{T_{\theta_0}}(h_1 + h_2)\|_{\mathbb{H}} \times \|(h_1 - \Pi_{T_{\theta_0}} h_1) + (h_2 - \Pi_{T_{\theta_0}} h_2)\|_{\mathbb{H}} . \quad (C.12)$$

Thus, the Proposition follows from (C.12) and the triangle inequality. ∎

References

ABADIE, A., ANGRIST, J. and IMBENS, G. (2002). Instrumental variables estimates of the effects of subsidized trainining on the quantiles of trainee earnings. *Econometrica*, **70** 91–117.

ALIPRANTIS, C. D. and BORDER, K. C. (2006). *Infinite Dimensional Analysis – A Hitchhiker's Guide*. Springer-Verlag, Berlin.

ANDREWS, D. W. K. (2000). Inconsistency of the Bootstrap When a Parameter Is On the Boundary of the Parameter Space. *Econometrica*, **68** 399–405.

ANDREWS, D. W. K. and SHI, X. (2013). Inference based on conditional moment inequatlies. *Econometrica*, **81** 609–666.

ANDREWS, D. W. K. and SOARES, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, **78** 119–157.

ANGRIST, J., CHERNOZHUKOV, V. and FERNANDEZ-VAL, I. (2006). Quantile regression under misspecification, with an application to the u.s. wage structure. *Econometrica*, **74** 539–563.

ARELLANO, M., HANSEN, L. P. and SENTANA, E. (2012). Underidentification? *Journal of Econometrics*, **170** 256–280.

ARMSTRONG, T. B. and CHAN, H. P. (2012). Multiscale adaptive infernece on conditional moment inequalities. Working paper. Yale University.

AUBIN, J.-P. and FRANKOWSKA, H. (1990). *Set-Valued Analysis*. Birkhauser, Boston.

BEARE, B. and MOON, J. (2013). Nonparametric Tests of Density Ratio Ordering. Working paper, University of California - San Diego.

BEARE, B., SEO, J. and SHI, X. (2013). Power Improvements to Three Tests of Concavity. Working paper, University of California - San Diego.

BERAN, R. (1997). Diagnosing bootstrap success. *Annals of the Institute of Statistical Mathematics*, **49** 1–24.

BERESTEANU, A. and MOLINARI, F. (2008). Asymptotic properties for a class of partially identified models. *Econometrica*, **76** 763–814.

BICKEL, P., GÖTZE, F. and ZWET, W. (1997). Resampling Fewer Than n Observations: Gains, Losses, and Remedies for Losses. *Statistica Sinica*, **7** 1–31.

BOGACHEV, V. I. (1998). *Gaussian Measures*. American Mathematical Society, Providence.

BOGACHEV, V. I. (2007). *Measure Theory: Volume II*. Springer-Verlag, Berlin.

BONTEMPS, C., MAGNAC, T. and MAURIN, E. (2012). Set Identified Linear Models. *Econometrica*, **80** 1129–1155.

BUCHINSKY, M. (1994). Changes in the u.s. wage structure 1963-1987: Application of quantile regression. *Econometrica*, **62** 405–458.

BUGNI, F. A. (2010). Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set. *Econometrica*, **78** 735–753.

CANAY, I. A. (2010). El inference for partially identified models defined by moment inqequalities: Coverage of the identified set. *The Journal of Econometrics*, **156** 408–425.

CAROLAN, C. A. and TEBBS, J. M. (2005). Nonparametric tests for and against likelihood ratio ordering in the two-sample problem. *Biometrika*, **92** 159–171.

CHANDRASEKARH, A., CHERNOZHUKOV, V., MOLINARI, F. and SCHRIMPF, P. (2013). Working Paper, Massachusetts Institute of Technology.

CHEN, X., TAMER, E. and TORGOVITSKY, A. (2011). Cowles Foundation Discussion Paper No. 1836.

CHERNOZHUKOV, V. and HANSEN, C. (2005). An iv model of quantile treatment effects. *Econometrica*, **73** 245–261.

CHERNOZHUKOV, V., HONG, H. and TAMER, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, **75** 1243–1284.

CHERNOZHUKOV, V., LEE, S. S. and ROSEN, A. M. (2013). Intersection bounds: Estimation and inference. *Econometrica*, **81** 667–737.

CHETVERIKOV, D. (2012). Adaptive test of conditional moment inequalities. Working paper. University of California Los Angeles.

CILIBERTO, F. and TAMER, E. (2009). Market structure and multiple equilibria in airline markets. *Econometrica*, **77** 1791–1828.

DAVYDOV, Y. A., LIFSHITS, M. A. and SMORODINA, N. V. (1998). *Local Properties of Distribuions of Stochastic Functionals*. American Mathematical Society, Providence.

DEHLING, H. and PHILIPP, W. (2002). Empirical process techniques for dependent data. In *Empirical Process Techniques for Dependent Data* (H. Dehling, T. Mikosch and M. Sørensen, eds.). Birkhäuser.

DORFMAN, R. (1938). A note on the δ-method for finding variance formulae. *The Biometric Bulletin*, **1** 129–137.

DUGUNDJI, J. (1951). An extension of tietze's theorem. *Pacific Journal of Mathematics*, **1** 353–367.

DÜMBGEN, L. (1993). On Nondifferentiable Functions and the Bootstrap. *Probability Theory and Related Fields*, **95** 125–140.

EFRON, B. (1979). Bootstrap methods: Another look at the jacknife. *Annals of Statistics*, **7** 1–26.

ESCANCIANO, J. C. and ZHU, L. (2013). Working paper, Indiana University.

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, Berlin.

HANSEN, P. R. (2005). A test for superior predictive ability. *Journal of Business and Economic Statistics*, **23** 365–380.

HIRANO, K. and PORTER, J. R. (2012). Impossibility results for nondifferentiable functionals. *Econometrica*, **80** 1769–1790.

HOEF, J. M. V. (2012). Who invented the delta method? *The American Statistician*, **66** 124–127.

HOROWITZ, J. L. (2001). The bootstrap. In *Handbook of Econometrics V* (J. J. Heckman and E. Leamer, eds.). Elsevier, 3159–3228.

IMBENS, G. W. and MANSKI, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, **72** 1845–1857.

JACKWERTH, J. C. (2000). *Review of Financial Studies*, **13** 433–451.

KAIDO, H. (2013). A Dual Approach to Inference for Partially Identified Econometric Models. Working paper, Boston University.

Kaido, H. and Santos, A. (2013). Asymptotically Efficient Estimation of Models Defined by Convex Moment Inequalities. *Econometrica* forthcoming.

Kline, P. and Santos, A. (2013). Sensitivity to missing data assumptions: Theory and an evaluation of the u.s. wage structure. *Quantitative Economics*, **4** 231–267.

Kosorok, M. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer.

LeCam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag, New York.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces*. Springer-Verlag, Berlin.

Lehmann, E. and Romano, J. (2005). *Testing Statistical Hypotheses*. Springer Verlag.

Linton, O., Song, E., K. and Whang, Y.-J. (2010). An Improved Bootstrap Test of Stochastic Dominance. *Journal of Econometrics*, **154** 186 – 202.

Manski, C. F. (2003). *Partial Identification of Probability Distributions*. Springer-Verlag, New York.

Muralidharan, K. and Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from india. *Journal of Political Economy*, **9** 39–77.

Pakes, A., Porter, J., Ho, K. and Ishii, J. (2006). Moment inequalities and their application. Working Paper, Harvard University.

Romano, J. P. and Shaikh, A. M. (2008). Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference – Special Issue in Honor of Ted Anderson*, **138** 2786–2807.

Romano, J. P. and Shaikh, A. M. (2010). Inference for the Identified Set in Partially Identified Econometric Models. *Econometrica*, **78** 169–211.

Shapiro, A. (1990). On Concepts of Directional Differentiability. *Journal of Optimization Theory and Applications*, **66** 477–487.

Shapiro, A. (1991). Asymptotic Analysis of Stochastic Programs. *Annals of Operations Research*, **30** 169–186.

Shapiro, A. (1994). Existence and differentiability of metric projections in hilbert spaces. *Siam Journal of Optimization*, **4** 130–141.

Song, K. (2012). Local Asymptotic Minimax Estimation of Nonregular Parameters with Translation-Scale Equivariant Maps.

van der Vaart, A. (1991). On Differentiable Functionals. *The Annals of Statistics*, **19** pp. 178–204.

van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer Verlag.

White, H. (2000). A reality check for data snooping. *Econometrica*, **68** 1097–1126.

Wolak, F. A. (1988). Duality in testing multivariate hypotheses. *Biometrika*, **75** 611–615.

Zaranotello, E. H. (1971). Projections on convex sets and hilbert spaces and spectral theory. In *Contributions to Nonlinear Functional Analysis* (E. H. Zaranotello, ed.). Academic Press.

Zeidler, E. (1984). *Nonlinear Functional Analysis and its Applications III*. Springer-Verlag, New York.